MODELO DE PROCESO PARA PROYECTOS DE EXPLOTACIÓN DE INFORMACIÓN

Tesista

Lic. Sebastian MARTINS

Directores

Prof. Rodolfo BERTONE (UNLP)

Prof. Patricia PESADO (UNLP)

Prof. Hernán MERLINO (UNLA)

TESIS PRESENTADA PARA OBTENER EL GRADO
DE
DOCTOR EN CIENCIAS INFORMÁTICAS

FACULTAD DE INFORMÁTICA

UNIVERSIDAD NACIONAL DE LA PLATA

2020

RESUMEN

Hace ya más de dos décadas, se registra un esfuerzo sostenido en el tiempo por definir un modelo de proceso que guíe el desarrollo de proyectos de Ingeniería de Explotación de Información. Sin embargo, las propuestas existentes presentan una visión parcial e incompleta, conduciendo a una tasa de fracaso cercana al 60% [Gondar, 2005; Marbán et al., 2009], a partir de lo cual los siguientes autores [Kurgan y Musilek, 2006; Mariscal et al., 2010; Kdnuggets, 2014] señalan la necesidad de definir un modelo de proceso que resuelva las limitaciones existentes. En este contexto, la presente investigación tiene como objetivo desarrollar un modelo de proceso integral, el cual presente una visión unificada, integrando los procesos orientados al producto y a la gestión, completa y detallada, describiendo las actividades involucradas y sus dependencias.

Como resultado del trabajo, se propone MoProPEI, un modelo de proceso integrado por los subprocesos de Desarrollo y Gestión, y descompuesto en un mayor grado de detalle en fases y actividades, para las cuales se propone el uso de distintas técnicas y procedimiento que describen las tareas a realizar. La correcta integración de la propuesta fue verificada a partir de su aplicación en tres proyectos pertenecientes a las áreas de educación, salud y análisis web. Las características estáticas fueron evaluadas mediante el marco comparativo de metodologías para proyectos de explotación de información [Moine, 2013]. Finalmente, se implementa la validación mediante un experimento controlado, replicando el único experimento reproducible identificado en la disciplina [Sharma, 2008]. De los resultados derivados de las estrategias de evaluación utilizadas, se observa que MoProPEI presenta una diferencia significativa con respecto a las propuestas antecesoras.

ABSTRACT

For more than two decades, there has been a sustained effort over time to define a process model that guides the development of Information Mining Engineering projects. However, the existing proposals present a partial and incomplete vision, leading to a failure rate close to 60% [Gondar, 2005; Marbán et al., 2009]. Several authors [Kurgan and Musilek, 2006; Mariscal et al., 2010; Kdnuggets, 2014] pointed out the need to define a process model that resolves existing limitations. In this context, this research aims to develop a comprehensive process model, which presents a unified vision, integrating the product-oriented and management processes, complete and detailed, describing the activities involved and their dependencies.

This thesis proposes MoProPEI, a process model integrated by the Development and Management subprocesses, and decomposed in a greater degree of detail into phases and activities. For each activity, at least one technique or procedure is selected, describing the tasks to be performed. The correct integration of the proposal was verified from its application in three projects belonging to the areas of: education, health and web analysis. We evaluate the static characteristics through the comparative framework of methodologies for information mining projects [Moine, 2013]. Finally, we carried out a controlled experiment to validate the proposal, replicating the only reproducible experiment identified in the discipline [Sharma, 2008]. From the results derived from the evaluation strategies used, we observed that MoProPEI presents a significant difference with respect to its predecessors.

En honor a
Dr. Ramón García-Martínez

"El cielo no es un lugar, ni un tiempo.
El cielo consiste en ser perfecto"

Richard Bach, "Juan Salvador Gaviota"

De tantos profesores que me he cruzado, agradezco que la vida me pusiera un maestro delante. Entre tantos que van a cumplir con su trabajo, solo aquellos que la educación es su vida se detienen a brindarle un desafío extra a un estudiante con ganas de salir de la monotonía que a veces presentan los espacios de formación.

Gracias por la oportunidad que me brindaste de formar parte de un grupo que no solo colaboran como profesionales, sino que se acompañan como familia.

Gracias por tu confianza, por cada consejo, por cada palabra de apoyo a lo largo de estos años.

Gracias por cada frase, por cada almuerzo compartido, por cada reunión sin importar la hora, por cada esfuerzo extra, por cada pelea por defender lo que con tanta pasión perseguimos.

Gracias por hacerme mejor persona.

Eternamente agradecido de cada minuto que la vida me permitió compartir contigo.

"La diferencia entre algo difícil y algo imposible,
es que lo segundo requiere un poco más de esfuerzo"

R.G.M.

DEDICATORIA

A mi padres Silvia y Juan, por su sacrificio a lo largo de todos estos años, remándola en los momentos difíciles para que nada nos falte. Sin ellos nada de esto hubiese sido posible.

A mis abuelos Adela y Juan, mis segundos padres, siempre conmigo, siempre apoyándome.

A mis hermanos Amílcar, Nayla y Melina por sus molestias y apoyo incondicional, haciendo este desafío más divertido.

A mi esposa Victoria, por cambiarme la vida. Su amor y acompañamiento durante los momentos más difíciles forman parte oculta de esta tesis.

A mis amigos de la vida Lucas, Rodrigo, Ignacio D., Ignacio A., Martín, Germán y Gastón, que siempre estuvieron en los momentos intensos donde un poco de distracción era necesaria.

A Hernán M. por acompañarme en el momento más difícil, siempre con palabras de aliento y apoyo, brindándome la fuerza y el estímulo para dar lo mejor.

A mis colegas y amigos del "cuartel de investigación", Santiago, Hernán, Rodolfo, Gero y Darío, por cada debate, cada consejo, cada aguante a lo largo de todos estos años.

A Dante BB, por acompañarme en las largas horas de escritura.

Gracias a cada uno de ellos, indispensables en cada momento del trayecto.

AGRADECIMIENTOS

A la Facultad de Informática de la Universidad Nacional de la Plata por acogerme con generosidad de "*alma mater*" para que pudiera llevar a cabo mis estudios de Doctorado en Ciencias Informáticas.

Al Grupo de Investigación en Sistemas de Información del Departamento de Desarrollo Productivo y Tecnológico de la Universidad Nacional de Lanús, por proveerme un estimulante ambiente de intercambio de ideas y de soporte en todas las instancias del proceso para obtener el grado de Doctor.

A mis directores de tesis: Prof. Rodolfo Bertone, Prof. Patricia Pesado y Prof. Hernán Merlino por su incondicional apoyo y orientación en todo momento.

A Ramón García-Martínez por su apoyo y guía.

A mis colegas y amigos: Santiago Bianco, Hernán Amatriain y Darío Rodríguez por los consejos, debates y numerosos momentos compartidos.

A todos los numerosos colegas que generosamente compartieron su tiempo y conocimiento ayudándome a lo largo del proceso.

ÍNDICE

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

ÍNDICE DE FORMULAS

ÍNDICE DE FUENTES DE INFORMACIÓN

NOMENCLATURA

ADS	Ciencia de Datos Ágil (del inglés, Agil Data Science)
AREP	Cantidad y tipo de los repositorios de datos disponibles
ASD-BI	Desarrollo de Software Adaptativo para Inteligencia de Negocios (del inglés Adaptive Software Development – Business Intelligence)
BPMN	modelo y notación de procesos de negocio (del inglés, Business Process Model and Notation)
CRISP-DM	Proceso Estándar para Minería de Datos Multi Industria (del inglés, CRoss-Industry Standard Process for Data Mining)
DER	Diagrama Entidad-Relación
DMLC	Ciclo de Vida Genérico para Minería de Datos (del inglés, Generic Data Mining Life Cycle)
DRPY	Tiempo total requerido para el desarrollo del proyecto
EE	Esfuerzo Estimado
ENPreCoSP	Encuesta Nacional sobre Prevalencias de Consumo de Sustancias Psicoactivas
ER	Esfuerzo Real
ETL	Extraer, Transformar y Cargar (del inglés, Extract, Transform and Load)
FEF	Fecha Estimada de Finalización
FEI	Fecha Estimada de Inicio
FMDS	Metodología Fundamental para la Ciencia de Datos (del inglés, Foundational Methodology for Data Science)
FRF	Fecha Real de Finalización
FRI	Fecha Real de Inicio
FuTD	Fuente Temporaria de Datos
GUA	Grado de Utilidad de Atributos
HAC	Algoritmo de agrupamiento de aglomeración jerárquica (del inglés, Hierarchical Agglomerative Clustering)
ID3	Dicotomizador Iterativo 3 (del inglés, Iterative Dichotomiser 3)
IEEE	Instituto de Ingeniería Eléctrica y Electrónica (del inglés, Institute of Electrical and Electronics Engineers)
IKDDM	Proceso Integrado de Descubrimiento de Conocimiento en Bases de Datos y Minería de Datos (del inglés, Integrated Knowledge Discovery and Data Mining Process Model)
INDEC	Instituto Nacional de Estadísticas y Censos
ISO/IEC	Organización Internacional de Normalización / Comisión Electrotécnica Internacional (del inglés, International Organization for Standardization / International Electrotechnical Commission)
KDD	Descubrimiento de Conocimiento en Bases de Datos (del inglés, Knowledge Discovery in Databases)
KEXT	Nivel de conocimiento y experiencia del equipo de trabajo
KLDS	Nivel de conocimiento sobre los datos
LECO	Grado de apoyo de los miembros de la organización
MCV	Modelo de Ciclo de Vida
MoProPEI	Modelo de Proceso para Proyectos de Explotación de Información

MPIMD	Modelo de Proceso para Ingeniería de Minería de Datos
NANC	Número de Atributos No Correctos
NANS	Número de Atributos No Significativos
NASE	Número de Atributos Útiles Sin Errores
NAUD	Número de Atributos Útiles con Defectos
OBTY	Tipo de objetivo de explotación de información
PePR	Porcentaje entre el tiempo Planificado y Real
PMBOK	Libro del conocimiento sobre gestión de proyectos (del inglés, Project Management Book of Knowledge)
PMI	Instituto de Gestión de Proyectos (del inglés, Project Management Institute)
QTUA	Cantidad de tuplas disponibles en tablas auxiliares
QTUM	Cantidad de tuplas disponibles en la tabla principal
RAE	Real Academia Española
RAM	memoria de acceso aleatorio (del inglés, Random Access Memory)
ROC	Característica Operativa del Receptor (del inglés Receiver Operating Characteristic)
SEMMA	Muestra, Exploración, Modificación, Modelado y Evaluación (del inglés, Sample, Explore, Modify, Model and Assess)
SOM	Mapas Auto-Organizados (del inglés, Self-Organizing Maps)
SWEBOK	Cuerpo de Conocimientos de Ingeniería de Software (del inglés Software Engineering Body of Knowledge)
TDIDT	Inducción descendente de Árboles de Decisión (del inglés, Top-Down Induction of Decision Trees)
TDSP	Proceso de Ciencia de Datos en Equipo (del inglés Team Data Science Process)
TOOL	Funcionalidad de las herramientas disponibles
VAC	Valor Asignado a la Característica

1. INTRODUCCIÓN

En este capítulo, se contextualizan los antecedentes conceptuales de ingeniería de explotación de información (sección 1.1), se define el objetivo de la tesis (sección 1.2), se presenta una visión general de la estructura de la misma (sección 1.3), y se relaciona la producción científica derivada de los trabajos de investigación desarrollados (sección 1.4).

1.1. ANTECEDENTES CONCEPTUALES DE INGENIERÍA DE EXPLOTACIÓN DE INFORMACIÓN

Hace ya una década, un estudio de la Universidad de California en Berkeley [Lyman y Varian, 2003] señaló que la información disponible en Internet crecía a razón de 92 petabytes (10^{15} bytes) por año, mientras que estudios recientes señalan que el volumen de información crece de manera exponencial cada año [Manyika et al., 2011], destacándose que la misma está disponible para procesos de descubrimiento de conocimiento [Maimon y Rokach, 2005] con independencia que se encuentre en fuentes estructuradas [Rudin y Cressy, 2003; Moss, 2003] o desestructuradas [Vuori, 2006].

La Inteligencia de Negocio propone un abordaje interdisciplinario (dentro del que se encuentra la Informática), que tomando todos los recursos de información disponibles y el uso de herramientas analíticas y de síntesis con capacidad de transformar la información en conocimiento, se centra en generar a partir de éstos, conocimiento que contribuya con la toma de decisiones de gestión y generación de planes estratégicos en las organizaciones [Thomsen, 2003; Negash y Gray, 2008].

A partir del estudio de la historia de la disciplina, se identifica que el término Minería de Datos (Data Mining) es utilizado de manera ambigua haciendo referencia al conjunto de actividades requeridas para extraer conocimiento novedoso y de interés para dar soporte el proceso de toma de decisiones, y a la implementación de algoritmos de búsqueda de patrones en bases de datos. En [Fayyad et al., 1996], se propone diferenciar dichos conceptos haciendo uso de los términos Extracción de Conocimiento en Base de datos (Knowledge Discovery in Databases) y Minería de datos, respectivamente. Sin embargo, se ha señalado la existencia de líneas de investigación en campos tales como: minería de textos [Tan, 1999], minería de imágenes [Hsu et al., 2002], minería de patrones en flujos de información [Gaber et al., 2010], minería en la web [Kosala y Blockeel, 2000]; conviniéndose el uso del término explotación de información [Kruse y Borgelt, 2003; Gopal et al., 2011] como referencia genérica a cualquiera de los tipos de minería precitados. Además se

hace énfasis en que los resultados deben ser comprensibles [Kruse y Borgelt, 2003], validables y que brinden valor al proceso de toma de decisiones. En este sentido, se define a la Explotación de Información como la sub-disciplina de los Sistemas de Información que aporta a la Inteligencia de Negocio [Langseth y Vivatrat, 2003] las herramientas para la transformación de información en conocimiento [Srivastava et al., 2000], con el propósito de extraer patrones interesantes y de regularidades importantes en grandes masas de información [Abraham, 2003; Cooley, 2003].

A partir de la concepción de proceso integral (KDD) descripto en el párrafo previo, es que desde 1993 la disciplina ha realizado un esfuerzo sostenido en el tiempo para definir una propuesta que guíe el desarrollo de un proyecto de explotación de información (Adriaans y Zantinge [1996], Descubrimiento de Conocimiento en Bases de Datos (KDD) [Fayyad et al., 1996], Berry and Gordon [1997], Muestreo, Exploración, Modificación, Modelado y Evaluación (SEMMA) [SAS Institute Inc, 1997], Cabena et al. [1998], Edelstein [1998], Feldens et al. [1998], Kopanakis and Theodoulidis [1999], Reinartz [1999], Cios et al. [2000], Proceso Estándar para Minería de Datos Multi-Industria (CRISP-DM) [Chapman et al., 2000], Han and Cercone [2000], Catalyst [Pyle, 2003], Marbán et al., [2007; 2009], Proceso Integrado de Descubrimiento de Conocimiento y Minería de Datos (IKDDM) [Sharma, 2008], Desarrollo de Software Adaptativo para Minería de Datos (ASD-DM) [Alnoukari, et al., 2008], Desarrollo de Software Adaptativo para Inteligencia de Negocios (ASD-BI) [Alnoukari, 2010], AgileKDD [do Nascimento y de Oliveira, 2012], Metodología Fundacional para la Ciencia de Datos (FMDS) [Rollins, J. B., 2015], Proceso de Ciencia de Datos en Equipo (TDSP) [Microsoft, 2016], entre otros). Sin embargo, en los últimos años los proyectos de explotación de información han ido creciendo en lo que se refiere a su complejidad [Mariscal et al., 2010]. En este sentido, el incremento de fuentes de información accesibles (en cantidad y tamaño) para un proyecto, la amplia posibilidad de satisfacción de necesidades y las características intrínsecas de este tipo de proyectos (interdisciplinarios [Fayyad et al., 1996], complejos [Kurgan y Musilek, 2006; Gallardo, 2009] y dinámicos [Brachman & Anand, 1996]) hacen evidente la necesidad de definir una visión ingenieril del proceso de extracción de conocimiento [Marbán et al., 2009]. Esta visión incorpora aquellas actividades asociadas con la gestión de proyectos (planificar, administrar y controlar) [Kurgan y Musilek, 2006; Mariscal et al., 2010; do Nascimento y de Oliveira, 2012].

En este contexto, se estima de interés señalar la siguiente definición de la Ingeniería de Software proporcionada por el SWEBOK [Abran et al., 2004]: “la aplicación de un enfoque sistemático, disciplinado y cuantificable al desarrollo, operación y mantenimiento de software, y el estudio de estos enfoques, es decir, la aplicación de la ingeniería al software”. A partir de la definición previa, se conviene el término Ingeniería de Explotación de Información como la aplicación de un enfoque

sistemático, disciplinado y cuantificable al desarrollo de proyectos de explotación de información, y el estudio de este enfoque, es decir, la aplicación de la ingeniería a la explotación de información [García-Martínez et al. 2011]. Este enfoque entiende en los procesos y las metodologías utilizadas para: ordenar, controlar y gestionar la tarea de encontrar patrones de conocimiento en masas de información [García-Martínez et al. 2011].

En relación a lo previamente expuesto, se han identificado en las propuestas existentes una serie de deficiencias asociadas con la dificultad de aplicar los procesos vigentes, conduciendo a una tasa de fracaso cercana al 60% [Gondar, 2005; Marbán et al., 2009]. A continuación se presentan las principales carencias: a) falta de procesos completos, detallados y estandarizados que incorporen un enfoque ingenieril, b) inexistente o inadecuado soporte al usuario (realizando un énfasis en las etapas de comprensión de las necesidades del cliente y definición de los alcances del proyecto, y en su vinculación con la solución propuesta), c) deficiencias en la estimación del proyecto (a causa de herramientas y guías inexistentes) y d) necesidades insatisfechas asociadas con la mala calidad de los resultados.

La definición de un proceso que guíe el desarrollo de este tipo de proyectos es necesaria para reducir los riesgos y problemas asociados con la gestión de proyectos y el aseguramiento de la calidad de los resultados. De esta forma, es posible mejorar las prácticas y la productividad de los equipos de trabajo, evitando obtener resultados indeseados que perjudiquen al proceso de toma de decisiones asociado [Berry y Linoff, 2004]; e ingresar en periodos de crisis similares a los acontecidos en otras disciplinas como la ingeniería de software [Marbán et al., 2007].

1.2. OBJETIVO DE LA TESIS

El objetivo general de la tesis es sistematizar el conocimiento existente sobre ingeniería de explotación de información y formular una propuesta de modelo de proceso, la cual presente una visión unificada de los procesos orientado al producto y a la gestión, incorporando los métodos y técnicas para el desarrollo de cada una de sus actividades.

El objetivo previamente definido, se sustenta en las siguientes hipótesis:

Hipótesis I: Las metodologías de desarrollo de proyectos de explotación de información existentes, se centran en el proceso orientado al producto, es decir, las actividades enfocadas en el descubrimiento de patrones de conocimiento en masas de información. Sin embargo, no existe una visión de proceso que identifique fases,

tareas, dependencias, técnicas de representación y procedimientos de ejecución de la tarea; que permitan sistematizar la concreción del proyecto.

Hipótesis II: La actividad de descubrimiento de patrones en masas de datos utilizando algoritmos de descubrimiento de conocimiento conocidos hoy como algoritmos de minería de datos, acredita un desarrollo consolidado a lo largo de las últimas tres décadas. La sistematización de la aplicación de estos algoritmos está sustentada en metodologías orientadas a transformar datos en conocimiento. Sin embargo, no existe una visión generalizada que permita a partir del análisis de las necesidades del cliente, determinar los modelos de extracción de patrones de conocimiento a utilizar para satisfacer las problemáticas identificadas.

Hipótesis III: Trabajos de investigación recientes han consolidado resultados en el área de gestión de proyectos de explotación de información tales como: métodos para la determinación de la viabilidad de proyecto y la estimación de recursos necesarios, métricas del proyecto, entre otros. Sin embargo, las metodologías usuales de explotación de información no contemplan el nivel de administración del control y la gestión de proyecto que los nuevos modelos de proceso de software prevén.

A partir de las hipótesis y el objetivo general previamente mencionados, se definen los siguientes objetivos específicos:

Objetivo particular vinculados a la Hipótesis I y II:

Desarrollar un proceso orientado al producto para proyectos de ingeniería de explotación de información que identifique fases, actividades, dependencias, técnicas de representación y procedimientos para la ejecución de las tareas. Permitiendo a partir del marco metodológico sistematizar el desarrollo de proyectos en el área y mapear las necesidades definidas por las partes interesadas con los procesos o modelos a aplicar para dar respuesta a las problemáticas identificadas.

Objetivo particular vinculado a la Hipótesis III:

Desarrollar un proceso de gestión de proyectos de ingeniería de explotación de información que identifique fases, tareas, dependencias, técnicas de

representación y procedimientos de ejecución de la tarea; que permita sistematizar el control y la gestión de proyectos en el área.

Objetivo particular vinculado a las Hipótesis I y III:

Desarrollar un modelo de proceso integrado de ingeniería de explotación de información que articule los procesos orientado al producto y gestión de proyectos de ingeniería de explotación de información.

1.3. VISIÓN GENERAL DE LA TESIS

La tesis se estructura en siete capítulos: “Introducción”, “Estado de la Cuestión”, “Delimitación del Problema”, “Solución”, “Validación”, “Conclusiones” y “Referencias”, a los que se agregan tres anexos en los cuales se presentan los formalismos parciales de los proyectos desarrollados en la presente tesis, los resultados del análisis sistemático de la literatura y el material complementario (plantilla de la encuesta y datos obtenidos) del experimento realizado.

En el capítulo 1 “Introducción” se contextualizan los antecedentes conceptuales de la ingeniería de explotación de información, se presenta el objetivo de la tesis, se proporciona una visión general de la misma, y se relaciona la producción científica derivada de los trabajos de investigación desarrollados.

En el capítulo 2 “Estado de la Cuestión” se introducen conceptos relativos a la ingeniería de explotación de información. Se presenta una revisión histórica de las propuestas metodológicas existentes en la disciplina. Se detallan las más relevantes y se resumen las principales críticas. Se presentan aquellas técnicas (desarrolladas ad-hoc y de otras disciplinas) de interés para la definición de un modelo de proceso para proyectos de ingeniería de explotación de información. Se concluye el capítulo, introduciendo los conceptos de proceso, metodología y gestión de proyectos adoptados en la tesis.

En el capítulo 3 “Delimitación del Problema” se introduce el contexto disciplinar del problema abordado en la tesis, se formula una discusión de los abordajes metodológicos para la ingeniería de explotación de información (en donde se definen las propuestas y las dimensiones a analizar); y se define el problema abierto considerado en la tesis, describiendo las preguntas de investigación formuladas.

En el capítulo 4 “Solución” se introduce el modelo de proceso propuesto, describiendo la estructura y composición (fases, tareas, dependencias y técnicas) de los subprocesos (orientado al producto y gestión) que integran la propuesta. De forma complementaria, se presenta un proyecto perteneciente al área de la salud como prueba de concepto, con el objetivo de facilitar al lector en la comprensión de la solución.

En el capítulo 5 “Validación” se aplican tres métodos de evaluación acorde a este tipo de artefactos de tecnologías de la información [Hevner et al., 2004]: O*bservacional* a partir de dos casos de validación. El primero, perteneciente al análisis de un sitio web, cuyo objetivo es comprender el comportamiento e intereses de los usuarios, a partir del registro de los registros de navegación. El segundo caso, perteneciente al ámbito de la educación superior, cuyo objetivo es brindar soporte al proceso de toma de decisiones en cuestiones de políticas públicas universitarias vinculadas con la mejora de la calidad educativa, estudiando el comportamiento de los estudiantes en cuestiones de masividad. El caso se enmarca en un proyecto de investigación realizado por docentes de la Facultad de Ciencias Exactas, Físicas y Naturales (FCEFyN) en la Universidad Nacional de Córdoba; A*nalítico* mediante la comparación de la propuestas más significativas descritas en el capítulo 2 (Estado de la Cuestión) y analizadas en el capítulo 3 (Delimitación del Problema), haciendo uso del marco para evaluar metodologías de explotación de información definido en [Moine, 2013]; y *Experimental*, replicando el experimento definido en [Sharma, 2008] comparando a la propuesta superadora en dicho experimento con el modelo de proceso propuesto en esta investigación.

En el capítulo 6 “Conclusiones” se presentan las aportaciones de la tesis doctoral y se destacan las futuras líneas de investigación de interés, a partir del problema abierto abordado.

En el capítulo 7 “Referencias” se listan todas las publicaciones consultadas para el desarrollo de la tesis.

En “Anexo A” se presentan los formalismos secundarios (intermedios o complementarios) obtenidos para los tres proyectos presentados en la tesis (prueba de concepto y casos de validación).

En el capítulo “Anexo B” se presenta en detalle los pasos realizados para el estudio sistemático de la literatura.

En “Anexo C” se presentan los materiales complementarios del desarrollo experimental realizado para la validación de la propuesta y su posible replicación (material utilizado para la encuesta y los datos crudos y transformados).

1.4. PRODUCCIÓN CIENTÍFICA DERIVADAS DE LA INVESTIGACIÓN DE LA TESIS

Durante el desarrollo de esta tesis se han comunicado resultados parciales a través de diversas publicaciones que a continuación se detallan:

Libros:

1. García-Martínez, R., Britos, P., Martins, S., Baldizzoni, E. 2015. Explotación de Información. Ingeniería de Proyectos. Editorial Nueva Librería ISBN 978-987-1871-34-6.

Capítulos de libro:

1. Martins, S., Rodríguez, D., García-Martínez, R. 2014. Derivación del Proceso de Explotación de Información desde el Dominio de Negocio. Capítulo X en "Ingeniería del Software e Ingeniería del Conocimiento: Dos Disciplinas Interrelacionadas". Pág. 159-177. Sello Editorial de la Universidad de Medellín. ISBN 978-958-8815-31-2.

Artículos en revistas y series con referato Internacionales:

1. Martins, S., Pesado, P., García-Martínez, R., 2016. Intelligent Systems in Modeling Phase of Information Mining Development Process. Lecture Notes on Artificial Intelligence 9799: 3–15. ISSN 0302-9743.

2. Martins, S., Rodríguez, D., García-Martínez, R. 2014. Deriving Processes of Information Mining Based on Semantic Nets and Frames. Lecture Notes on Artificial Intelligence, 8482: 150-159. ISBN 978-3-319-07466-5.

Artículos en revistas y series con referato Nacionales:

1. Martins, S., Pesado, P., García-Martínez, R. 2014. Propuesta de Modelo de Procesos para una Ingeniería de Explotación de Información: MoProPEI. Revista Latinoamericana de Ingeniería de Software, 2(5): 313-332, ISSN 2314-2642.

2. Martins, S.2014. Derivación del Proceso de Explotación de Información Desde el Modelado del Negocio. Revista Latinoamericana de Ingeniería de Software, 2(1): 53-76, ISSN 2314-2642.

Congresos Internacionales:

1. García-Martínez, R., Martins, S., Bianco, S., & Navas, H. (2017). Discovery of Psychoactive Substance Addiction Patterns Based on Information Mining Engineering. Studies in health technology and informatics, 245, 1282-1282.

2. Martins, S., Pesado, P., García-Martínez, P. 2016. Information Mining Projects Management Process. Proceedings 28th International Conference on Software Engineering & Knowledge Engineering. Pág. 504-509. ISBN 1-891706-39-X.

Congresos Nacionales:

1. Flores, L., Mariño, S. I., & Martins, S. (2018). Propuesta de procedimiento para el análisis delictivo basado en la explotación de la información. In XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018, Universidad Nacional del Nordeste).

2. Bianco, S., Martins, S., Rodríguez, D., & García Martínez, R. (2017). Ingeniería de explotación de información aplicada a la gestión universitaria: caso licenciatura en sistema Universidad Nacional de Lanús. In XII Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET, La Matanza 2017).

3. Martins, S., Pesado, P., García-Martínez, R. 2016. Propuesta de Artefactos para el Subproceso de Gestión del Modelo de Proceso de Proyectos de Explotación de Información (G-MoProPEI). Libro de Actas del XXII Congreso Argentino de Ciencias de la Computación. Pág. 556-565. ISBN 978-987-733-072-4. Universidad Nacional del San Luis.

4. Díaz, L., Martins, S., Las Heras, J., García-Martínez, R. 2016. Explotación de Información Aplicada a la Caracterización de Patrones Socio-Económicos de la Población Estudiantil de Carreras de Ciencias Económicas. Proceedings del XI Congreso de Tecnología en Educación y Educación en Tecnología. Pág. 449-457. ISBN 978-987-3977-30-5.

5. Díaz, L., Martins, S., Garcia-Martinez, R. 2015. Descubrimiento de Patrones Socio-económicos de Población Estudiantil de Carreras de Ingeniería Basado en Tecnologías de Explotación de Información. Proceedings X Congreso de Tecnología en Educación y Educación en Tecnología. Pág. 306-315. ISBN 978-950-656-154-3.

6. Martins, S., Pesado, P., García-Martínez, R. 2014. Propuesta de Proceso de Ingeniería de Explotación de Información Centrado en Control y Gestión del Proyecto. XI Workshop de

Bases de Datos y Minería de Datos. Proceedings XX Congreso Argentino de Ciencias de la Computación. Universidad Nacional de la Matanza. ISBN 978-987-3806-05-6.

Simposios Doctorales Internacionales

1. Martins, S. (2018). Modelo de Proceso para Proyectos de Ingeniería de Explotación de Información. XXI Congreso Iberoamericano en Ingeniería de Software.

En el contexto de las investigaciones asociadas a la tesis, se realizaron actividades concurrentes con su temática:

Dirección de tesis de maestría

1. Ciciliani Gabriel. (2019). *Estudio de pertinencia de algoritmos en procesos de descubrimiento de reglas de pertenencia a grupos*. Maestría en Ingeniería en Sistemas de Información. Universidad Tecnológica Nacional. Facultad Regional Buenos Aires. Escuela de Posgrado.

2. Flores Lorena. (2019). *Integración de procesos de explotación de información y tecnologías GIS: Aplicación para el hallazgo de patrones de robos y hurtos de la Ciudad de Corrientes*. Maestría en Tecnologías de la Información. Universidad Nacional del Nordeste. Facultad de Ciencias Exactas y Naturales y Agrimensura.

Dirección de tesis de grado

1. Bedoya Tafur Julian David. (2017). Descubrimiento De Patrones En Causales De Desgranamiento Estudiantil Por Medio De Ingeniería De Explotación De Información. Ingeniería De Sistemas Y Telecomunicaciones. Departamento De Ciencias Básicas E Ingeniería. Universidad Católica De Pereira. Colombia.

2. ESTADO DE LA CUESTIÓN

En este capítulo se introduce el marco teórico relacionado con el problema abordado en esta tesis. Se presentan los conceptos de ingeniería de la explotación de información (sección 2.1) y la evolución de las propuestas de modelo de proceso / metodología para proyectos de explotación de información (sección 2.2), describiéndose en detalle las más relevantes. En la sección 2.3, se analizan las principales debilidades de cada una de las propuestas seleccionadas. En la sección 2.4, se detallan las técnicas o procedimientos específicamente desarrollados para proyectos de ingeniería de explotación de información. Finalmente, se describen las concepciones de modelo de proceso, metodología, ciclo de vida e ingeniería de proyectos sobre las cuales la propuesta se funda (sección 2.5).

2.1. INGENIERÍA DE EXPLOTACIÓN DE INFORMACIÓN

La Inteligencia de Negocio propone un abordaje interdisciplinario (dentro del que se encuentra la Informática), que tomando todos los recursos de información disponibles y el uso de herramientas analíticas y de síntesis con capacidad de transformar la información en conocimiento, se centra en generar a partir de éstos, conocimiento que contribuya con la toma de decisiones de gestión y generación de planes estratégicos en las organizaciones [Thomsen, 2003; Negash y Gray, 2008].

La Explotación de Información es la sub-disciplina de los Sistemas de Información que aporta a la Inteligencia de Negocio [Langseth y Vivatrat, 2003] las herramientas para la transformación de información en conocimiento [Srivastava et al., 2000]. Ha sido definida como la búsqueda de patrones interesantes y de regularidades importantes en grandes masas de información [Abraham, 2003; Cooley, 2003].

En este trabajo, se diferencian los conceptos minería de datos y explotación de información. El primero de ellos se entiende como la implementación de algoritmos de búsqueda de patrones en bases de datos [Fayyad et al., 1996], mientras que en la actualidad existen líneas de investigación en campos tales como: minería de textos [Tan, 1999], minería de imágenes [Hsu et al., 2002], minería de patrones en flujos de información [Gaber et al., 2010], minería en la web [Kosala y Blockeel, 2000], entre otras. En este contexto, se conviene utilizar al término explotación de información [Kruse y Borgelt, 2003; Gopal et al., 2011] como referencia genérica a cualquiera de los tipos de minería precitados, así como enfatizar que los resultados deben ser comprensibles [Kruse y Borgelt, 2003], validables y que provean valor al proceso de toma de decisiones.

Un Proceso de Explotación de Información se define, como un grupo de tareas relacionadas lógicamente [Curtis et al., 1992] que, a partir de un conjunto de información con un cierto grado de valor para la organización, se ejecuta para lograr otro, con un grado de valor mayor que el inicial [Ferreira et al., 2005; Kanungo, 2005]. Adicionalmente, existe una variedad de técnicas de minería de datos, en su mayoría provenientes del campo del Aprendizaje Automático [García-Martínez, 1997; García-Martínez et al., 2003], susceptibles de ser utilizadas en cada uno de estos procesos.

En el SWEBOK [Abran et al., 2004] se ha definido al término Ingeniería de Software como: "la aplicación de un enfoque sistemático, disciplinado y cuantificable al desarrollo, operación y mantenimiento de software, y el estudio de estos enfoques, es decir, la aplicación de la ingeniería al software". En este contexto, se conviene definir a la Ingeniería de Explotación de Información como la aplicación de un enfoque sistemático, disciplinado y cuantificable al desarrollo de proyectos de explotación de información, y el estudio de este enfoque, es decir, la aplicación de la ingeniería a la explotación de información. La ingeniería de explotación de información entiende en los procesos y las metodologías utilizadas para: ordenar, controlar y gestionar la tarea de encontrar patrones de conocimiento en masas de información [García-Martínez et al. 2011].

2.2. PROPUESTAS METODOLÓGICAS

En esta sección se presenta una descripción histórica de las metodologías o modelos de proceso más relevantes en la disciplina, detallando los aspectos fundamentales de cada uno de ellos. Los resultados se derivan del estudio sistemático de la disciplina detallado en el Anexo B (pág. 471). En el análisis solo fueron contempladas aquellas propuestas generales, omitiendo aquellas definidas para un dominio específico.

La figura 2.1, adaptada de [Mariscal et al., 2010; do Nascimento y de Oliveira, 2013], resume la evolución en el tiempo de las propuestas, señalando los antecedentes sobre los cuales estas fueron definidas, observándose a las propuestas KDD y CRISP-DM como bases a partir de las cuales la mayoría de las propuestas restantes se originan [Kurgan y Musilek, 2006; Mariscal et al., 2010; Alnoukari, M., & El Sheikh, 2012; do Nascimento y de Oliveira, 2013]. Adicionalmente, Gregory Piatetsky (presidente de kdnuggets.com) ha desarrollado entre los años 2002 y 2014 [Kdnuggets, 2002; 2004; 2007; 2014] cuatro encuestas respecto al uso de metodologías (o procesos) para el desarrollo de proyectos en la disciplina. En la figura 2.2 (adaptada de [Kdnuggets, 2014]) se ilustran los resultados obtenidos en las últimas dos encuestas, proporcionando información sobre las metodologías más utilizadas. Como resultado, se desprende como principales modelos de proceso, las propuestas bases: CRISP-DM (confirmando su predominio con más de un 40% de los

encuestados) y KDD (con el 7% aproximadamente), junto con SEMMA (que si bien ha disminuido un 5% aproximadamente en su representación, aún se conserva como la segunda propuesta más utilizada).

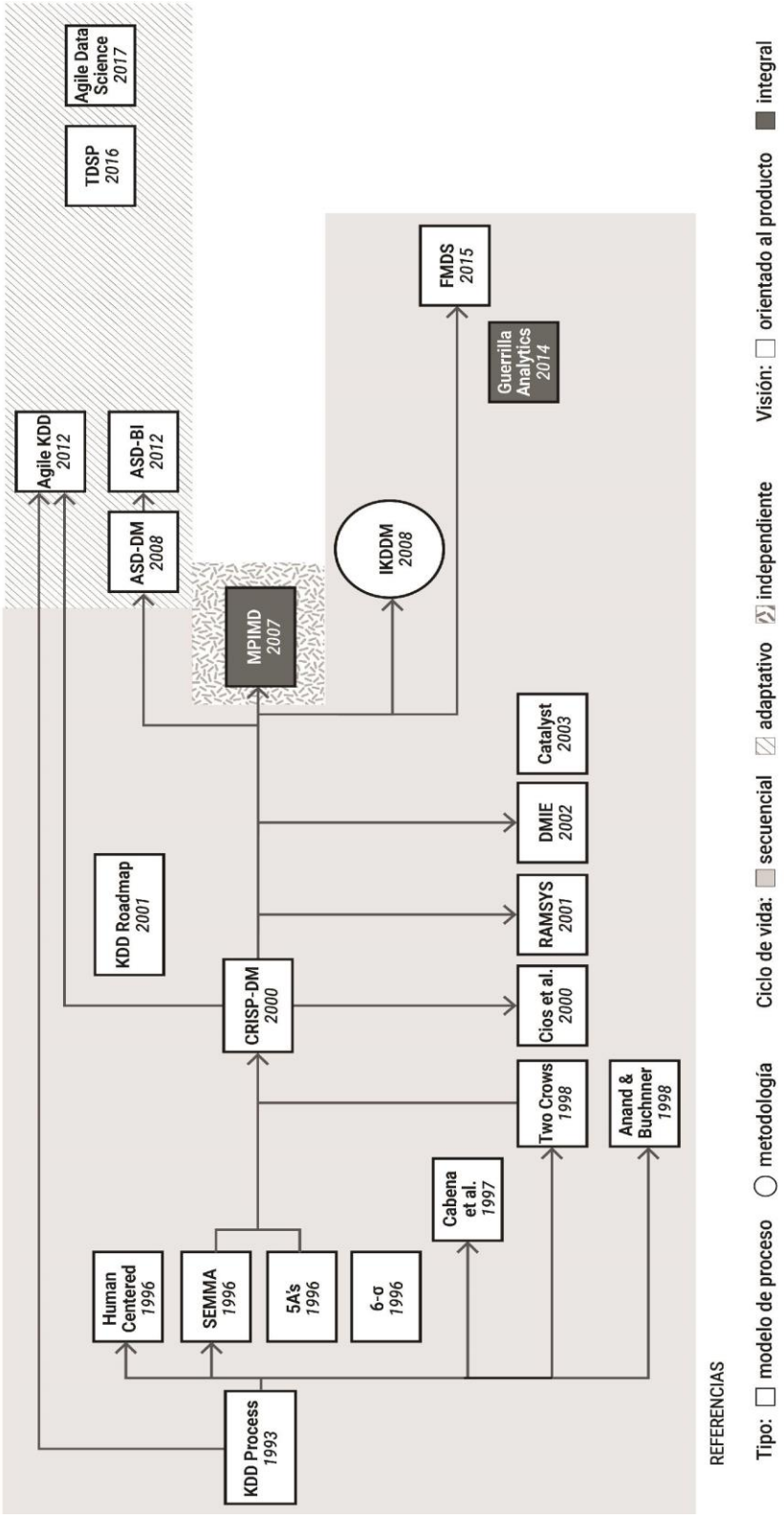


Figura 2.1. Evolución de los Procesos

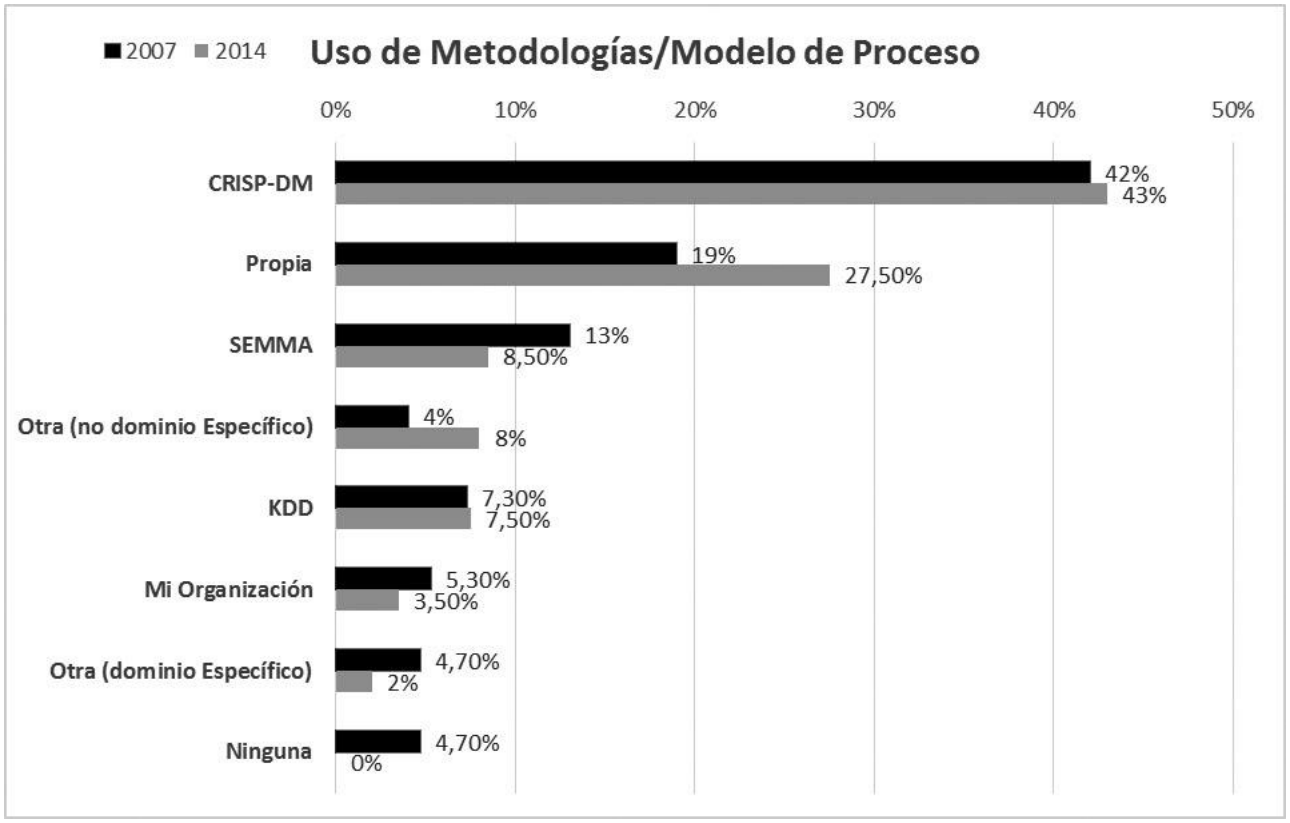


Figura 2.2. Uso de Modelo de Procesos y Metodologías - Resultados Encuestas 2007 y 2014

En el periodo de tiempo entre KDD y CRISP-DM (1993-2000), se identifica un auge en la proposición de procesos, en su mayoría basados en los abordajes precitados [Adriaans y Zantinge, 1996; Brachman y Anand, 1996; Berry y Gordon, 1997; Anand y Buchner, 1998; Cabena et al., 1998; Edelstein, 1998; Feldens et al., 1998; Two Crows Corporation, 1998; Harry y Schroeder, 1999; Kopanakis y Theodoulidis, 1999; Reinartz, 1999; Cios et al., 2000; Han y Cercone, 2000]. Dichas aportaciones han sido evaluadas y catalogadas como de menor impacto y relevancia [Kurgan y Musilek, 2006; Alnoukari, 2010; Alnoukari y El Sheikh, 2012], por lo cual no se considerarán en el presente trabajo.

A partir de la evolución de la disciplina, la metodología CRISP-DM se instala como el *estandar de facto* [Marbán et al., 2007; Kdnuggets, 2014]. En el 2003, se propone Catalyst, una metodología orientada a guiar el desarrollo de los proyectos mediante cajas que determinan el conjunto de pasos a realizar a partir del interés subyacente al proyecto. En [Moine, 2013], se determina la superioridad de dicha propuesta con respecto al *estándar de facto*. Entre los años 2007 y 2009, se definieron una serie de propuestas que surgen a partir de las críticas realizadas al principal modelo de proceso (CRISP-DM), en el cual están inspiradas. En [Marbán et al., 2007; 2009] se provee un abordaje inicial a un Modelo de Proceso para Ingeniería de Minería de Datos (MPIMD), el cual se destaca por incorporar al *estándar de facto*, una visión ingenieril de desarrollo de proyectos, basados en la

Ingeniería de Software, la cual incorpora los conceptos de gestión de proyectos. En [Sharma, 2008], se define un Proceso Integrado de Descubrimiento del Conocimiento y Minería de Datos (IKDDM, de sus siglas en inglés Integrated Knowledge Discovery and Data Mining), una metodología que proporciona una concepción detallada e integrada de la propuesta en la cual se basa, mediante la identificación de las vinculaciones entre las actividades (y sus dependencias) y las técnicas o métodos a utilizar en cada una de ellas.

En el mismo año (2008), inspirado en los métodos de gestión de proyectos ágiles utilizados para la ingeniería de software y su tendencia en dicha disciplina, se propone en [Alnoukari, et al., 2008] un modelo de proceso de Desarrollo de Software Adaptativo para Minería de Datos (ASD-DM, de sus siglas en inglés Adaptive Software Development – Data Mining), el cual fue mejorado dos años después mediante la propuesta Desarrollo de Software Adaptativo para Inteligencia de Negocios [Alnoukari, 2010] (ASD-BI, de sus siglas en inglés Adaptive Software Development – Business Intelligence). En 2012, en [do Nascimento y de Oliveira, 2012] se propone Agile KDD, un modelo de proceso basado en CRISP-DM y KDD, estructurado a partir del método ágil Proceso Unificado Abierto (OpenUP, del inglés Open Unified Process), el cual define una estructura de ciclo de vida iterativa e incremental, conformada por cuatro etapas generales: Inicio, Elaboración, Construcción y Transición.

Desde los inicios del presente trabajo de investigación, se identificaron cuatro nuevas propuestas: en [Ridge, E., 2014] se propone Guerrilla Analytics, un modelo de proceso de ciclo de vida en cascada que integra a las fases del proceso orientado al producto (Extracción, Recepción, Carga, Analítica, Consolidar, productos del trabajo y comunicar), con elementos de gestión de proyectos (control de versiones, gestión del proceso, análisis de la comunicación y evaluación del proyecto). En 2015, IBM propone la Metodología Fundamental para la Ciencia de Datos [Rollins, 2015] (FMDS, del inglés, Foundational Methodology for Data Science) basada en KDD y CRISP-DM. La propuesta incorpora nuevas prácticas de procesamiento de grandes volúmenes de datos, analítica de texto e imagen, inteligencia artificial, deep learning y procesamiento de lenguajes. En 2016, Microsoft define la metodología Proceso de Ciencia de Datos en Equipo [Microsoft, 2016] (TDSP, del inglés Team Data Science Process), centrada en mejorar el aprendizaje y la colaboración del equipo, estructurando el proceso a partir de un ciclo de vida ágil. Finalmente, en 2017 se presenta Ciencia de Datos Ágil [Jurney, 2017] (ADS, del inglés Agil Data Science) una propuesta basada en los modelos de ciclo de vida ágiles, la cual fomenta el desarrollo iterativo e incremental, y para cada fase del proceso define lineamientos a seguir para el desarrollo del proyecto.

En [Kurgan y Musilek, 2006; Alnoukari, 2010; Alnoukari y El Sheikh, 2012] se listan aquellas propuestas posteriores a CRISP-DM, cuyas aportaciones han sido evaluadas como poco significativas: [Debuse et al., 2001; Han y Kamber, 2001; Moyle y Jorge, 2001; Klosgen y Zytkow, 2002; Solarte, 2002; Haglin et al., 2005; Li y Ruan, 2007; Rennolls y ALShawabkeh, 2008].

En resumen, a partir de lo previamente expuestos y las aportaciones que las propuestas realizan, la tabla 2.1 resume los abordajes metodológicos relevantes seleccionados para su análisis.

Nombre	***Año***	***Característica***
KDD	1993	Modelo de Proceso, con MCV secuencial y visión orientada al producto.
SEMMA	1996	Modelo de Proceso, con MCV secuencial y visión orientada al producto.
CRISP-DM	2000	Modelo de Proceso, con MCV secuencial y visión orientada al producto.
Catalyst	2003	Modelo de Proceso, con MCV secuencial y visión orientada al producto.
MPIMD	2007	Modelo de Proceso, con MCV independiente y visión integral.
IKDDM	2008	Metodología, con MCV secuencial y visión orientada al producto.
ASD-BI	2012	Modelo de Proceso, con MCV adaptativo y visión orientada al producto.
FMDS	2015	Modelo de Proceso, con MCV secuencial y visión orientada al producto.
TDSP	2016	Modelo de Proceso, con MCV adaptativo y visión orientada al producto.

Tabla 2.1. Listado de abordajes seleccionados.

2.2.1. Descubrimiento de Conocimiento en Bases de Datos (KDD)

En [Fayyad et al., 1996] se presenta la primera aproximación a un modelo de procesos, definiendo las principales actividades técnicas orientadas a estructurar, ordenar y guiar el desarrollo de las tareas necesarias para producir conocimiento a partir de los datos almacenados por una organización. El objetivo principal es proveer un conjunto de herramientas para automatizar el proceso de análisis de datos y el artesanal proceso estadístico de selección de hipótesis. El objetivo subyacente al concepto de KDD, es el de diferenciar a la minería de datos, entendiendo a ésta como la actividad de aplicar distintos algoritmos en los datos para obtener patrones, del proceso necesario para generar conocimiento a partir de los datos. Es decir, entender a la minería de datos como una etapa que integra a un proceso general destinado a obtener patrones de conocimiento. Mediante la incorporación de KDD, se agregan una serie de pasos destinados a favorecer y garantizar el éxito de los resultados obtenidos en la aplicación de minería de datos. No basta con obtener patrones a partir de la ejecución de algoritmos de Minería de Datos, sino que se debe preparar los datos, así como el personal para una correcta aplicación de los algoritmos y posteriormente analizar los patrones obtenidos para generar conocimiento (el cual pueda ser comprendido), utilizándose para retroalimentar el proceso y para que los interesados puedan hacer uso del mismo. “Los pasos

adicionales a KDD, tales como Preparación de los Datos, Selección de los Datos, Limpieza de los Datos, Incorporación del Conocimiento Previo Adecuado y Correcta Interpretación de los Resultados de la Minería de Datos, son esenciales para garantizar la extracción de conocimiento útil desde los datos. La implementación de los métodos de Minería de Datos a ciegas, puede ser una actividad peligrosa, la cual fácilmente puede concluir en el descubrimiento de patrones inválidos o sin sentido." [Fayyad et al., 1996]. KDD fue definido, por sus autores, como un conjunto de pasos interactivos e iterativos en los cuales el usuario debe tomar una serie de decisiones. La figura 2.3 (adaptada de [Fayyad et al., 1996]), ilustra los pasos asociados (y las salidas de los mismos), los cuales se listan a continuación:

1. **Comprender el dominio de negocio**, los conocimientos previos relevantes e identificar las metas del proyecto desde el punto de vista del cliente.
2. **Seleccionar el conjunto de datos**, centrándose en el subconjunto de variables o registros sobre los cuales se realizará el descubrimiento.
3. **Limpiar y Pre-procesar los datos**, con el objetivo de eliminar los ruidos y campos inexistentes para favorecer al proceso de descubrimiento.
4. **Reducción y proyección de los datos**, donde se analiza la posibilidad de reducir la cantidad de variables bajo consideración.
5. **Relacionar el objetivo del proceso**, con un método particular de Minería de Datos.
6. **Análisis exploratorio y selección del modelo e hipótesis**. En este paso se eligen los algoritmos y se seleccionan los métodos a utilizar para encontrar patrones.
7. **Implementar Minería de Datos**: A partir de la implementación correcta de los pasos previos, se asiste significativamente a la obtención de mejores resultados en este paso.
8. **Interpretar los resultados**. Dicho paso puede generar la necesidad de iterar con alguno de los pasos previos.
9. **Actuar sobre el conocimiento descubierto**, es decir implementar el conocimiento en algún sistema, o simplemente documentarlo para su uso. Además incluye la confrontación del conocimiento identificado con previos conocimientos o creencias.

2.2.2. Muestreo, Exploración, Modificación, Modelado y Evaluación (SEMMA)

El instituto SAS en 1997 definió SEMMA [SAS Institute Inc, 1997], cuyo nombre proviene de las fases que la integran: Muestreo, Exploración, Modificación, Modelado y Evaluación (del inglés Sample, Explore, Modify, Model y Assess). Los autores describen la propuesta como el proceso utilizado para extraer información de valor y relaciones complejas existentes en grandes fuentes de

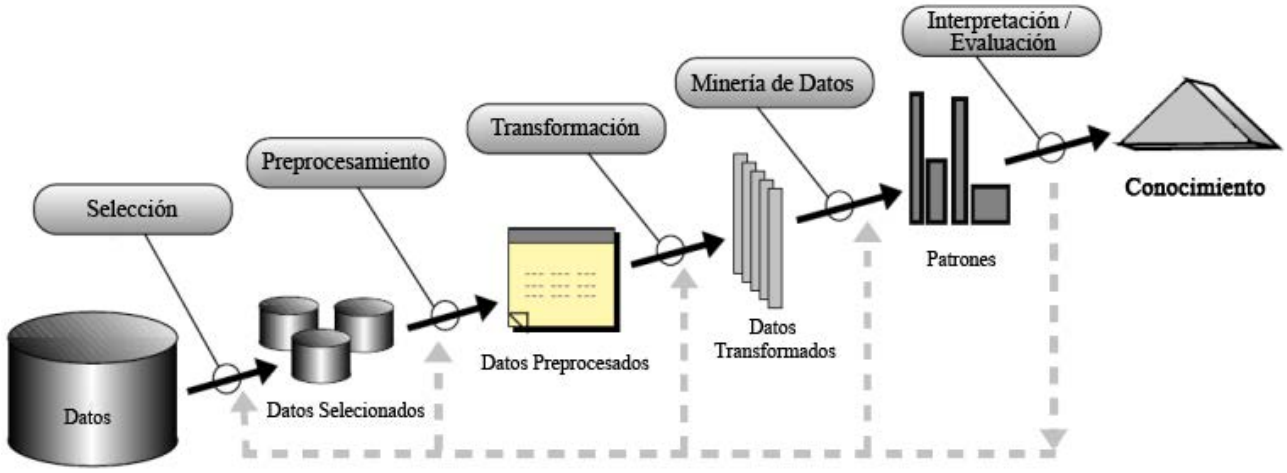


Figura 2.3. Modelo de Proceso: KDD.

datos, el cual facilite a los analistas de negocios en el proceso de toma de decisiones. A continuación se describen los alcances de cada una de sus fases:

1. **Muestreo**: se crean una o más tablas de datos las cuales serán utilizadas para extraer los patrones ocultos, realizando una muestra lo suficientemente grande para poder extraer información significante, pero a su vez suficientemente chica para que pueda ser procesada. En adición, se hace explícita la necesidad de dividir el set de datos en entrenamiento y prueba, para poder evaluar la precisión de los patrones encontrados.
2. **Exploración**: se evalúan los datos en busca de relaciones anticipadas, tendencias imprevistas y anomalías para obtener una mayor comprensión. Además, se manifiesta la posibilidad de aplicar técnicas de agrupamiento, clasificación, asociación y regresión.
3. **Modificación**: se realiza la creación, selección y transformación de las variables a utilizar en el modelo seleccionado. A partir de la fase previa, puede identificarse la necesidad de introducir nuevas variables, imputar valores faltantes o eliminar variables o registros anómalos.
4. **Modelado**: se utilizan técnicas analíticas para buscar patrones en los datos que permitan predecir de manera confiable el resultado deseado.
5. **Evaluación**: se analizan los datos y el modelo mediante la evaluación de la utilidad y fiabilidad de los resultados del proceso de minería de datos. Esta evaluación incluye la desestimación de patrones que no son válidos (los cuales surgen a partir de comparar los resultados obtenidos en los set de entrenamiento y prueba).

En [SAS Institute Inc, 1997], se describe a SEMMA como una estructura cíclica, en donde los pasos internos pueden realizarse iterativamente según sea necesario. La Figura 2.4 (adaptada de

[SAS Institute Inc, 1997]) ilustra las tareas de un proyecto de minería de datos y asigna esas tareas a sus cinco etapas.

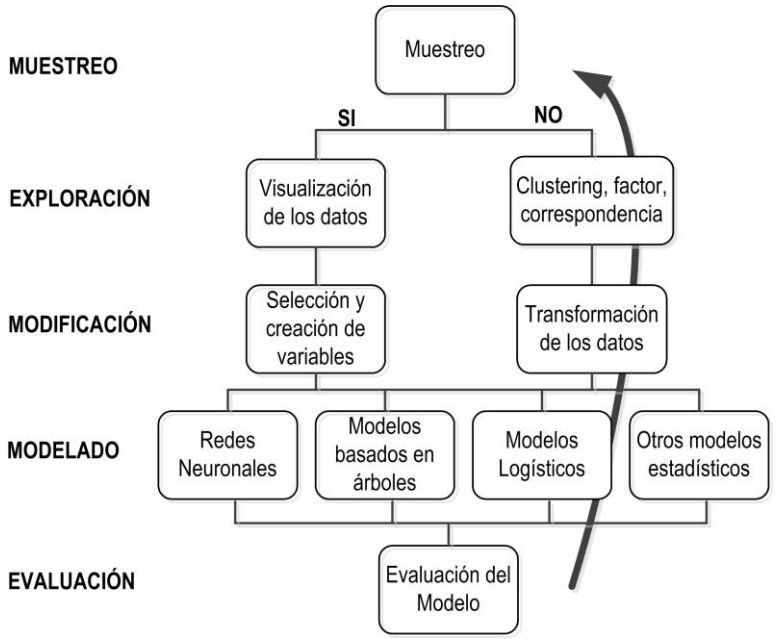


Figura 2.4. Modelo de Proceso: SEMMA.

2.2.3. Proceso Estándar para Minería de Datos Multi-Industria (CRISP-DM)

En [Chapman et al., 2000] se define al Proceso Estándar para Minería de Datos Multi-Industria, CRISP-DM (del inglés, CRoss-Industry Standard Process for Data Mining) como una metodología descrita en términos de modelo de proceso jerárquico, la cual se conforma por un conjunto de tareas descritas en cuatro niveles de abstracción (fases, tareas generales, tareas específicas e instancia del proceso), las cuales se ejecutan en base al modelo de ciclo de vida definido.

Este modelo de proceso, ampliamente utilizado para el desarrollo de proyectos de explotación de información, se centra en los pasos requeridos para la implementación del producto, considerando levemente algunas actividades vinculadas con la administración del proyecto. Dicha metodología se desarrolla a partir de un ciclo de vida, compuesto por seis fases (cada una compuesta por varias actividades), las cuales interactúan y se retroalimentan entre ellas (Figura 2.5, adaptada de [Chapman et al., 2000]).

El círculo exterior simboliza la naturaleza de la Minería de Datos, la retroalimentación de los resultados encadenando nuevas preguntas mejor enfocadas al negocio, y por consiguiente nuevas ejecuciones del ciclo de vida. Como se observa en la figura 2.5, las fases que componen al modelo de proceso son:

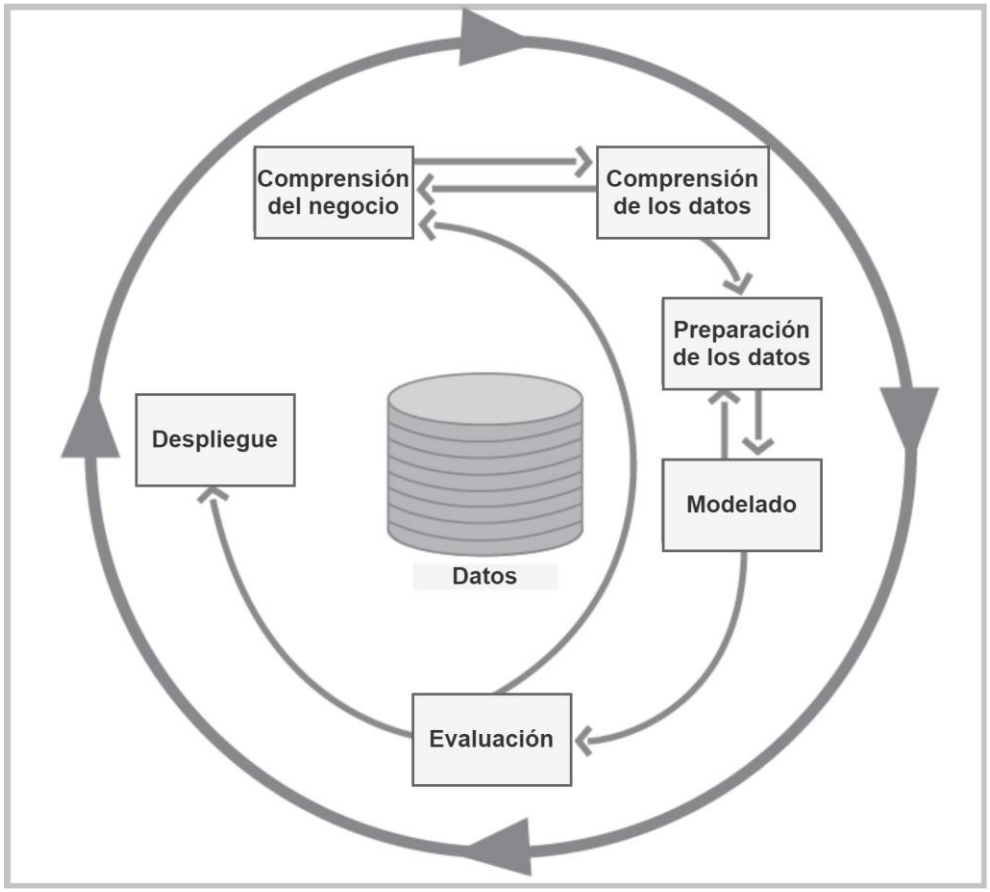


Figura 2.5. Modelo de Proceso: CRISP-DM.

1. **Comprensión del Negocio,** cuyas metas son entender los objetivos y requerimientos del proyecto desde la perspectiva del negocio, así como identificar el problema de minería de datos y realizar la planificación del proyecto. Las actividades asociadas son: Determinar los Objetivos del Negocio, Evaluar la Situación, Determinar las Metas de Minería de Datos y Producir el Plan del Proyecto.
2. **Comprensión de los Datos,** donde se realiza una recolección y análisis inicial de los datos, con el objetivo de identificar problemas de calidad y posibles subconjuntos de interés para distintas hipótesis. Las actividades generales que lo integran son: Recolección Inicial de los Datos, Descripción de los Datos, Exploración de los Datos y Verificación de la Calidad de los Datos.
3. **Preparación de los Datos,** conlleva la ejecución de todas las actividades necesarias para favorecer la calidad de los resultados. Las actividades generales que lo integran son: Selección de los Datos, Limpieza de los Datos, Construcción de los Datos, Integración de los Datos y Formateo de los Datos.
4. **Modelado,** se seleccionan y configuran las técnicas de modelado a utilizar. Las actividades generales que componen dicha fase son: Seleccionar de las Técnicas de Modelado, Generar el Diseño de las Pruebas, Construir el Modelo y Evaluar el Modelo.

5. **Evaluación,** se analiza el modelo generado para garantizar que este cumpla con los objetivos del negocio. Las actividades generales que lo integran son: Evaluar los Resultados, Revisar el Proceso y Determinar Próximos Pasos.
6. **Despliegue,** la cual abarca las actividades de integrar el conocimiento obtenido a algún sistema, o documentarlo para su posterior uso. Las actividades generales que conforman la última fase son: Planificar la Implementación, Planificar el Monitoreo y Mantenimiento, Producir el Reporte Final y Revisión del proyecto.

2.2.4. Catalyst

En [Pyle, 2003] se presenta Catalyst, como herramienta para guiar el desarrollo de proyectos de extracción del conocimiento para dar soporte al proceso de toma de decisiones. Su estructura se ilustra en la figura 2.6 (obtenida de [Moine, 2013]). El autor propone 5 posibles situaciones iniciales, que dan origen a un proyecto de explotación de información: 1) la existencia de un set de datos y el deseo de identificar patrones a partir de ellos; 2) una oportunidad de negocio; 3) como oportunidad para introducir una nueva herramienta en el negocio; 4) la necesidad de generar un modelo y 5) la necesidad de comprender una situación específica.

A partir de estas necesidades, se proponen un conjunto de métodos (descriptos a partir de 4 tipos de cajas: acción, descubrimiento, técnica y ejemplo) con el objetivo de definir dos aspectos claves: los datos requeridos y la necesidad o problemática real de los interesados.

Para alcanzar estos objetivos, se definen 5 pasos esenciales:

1. **Preparación de los datos,** donde se analizan los datos disponibles, realizando controles sobre la calidad y adecuación con respecto a los objetivos del proyecto. Como resultado de la fase, se limpia y transforma el set de datos a las necesidades del proceso.
2. **Selección de las herramientas,** se determinan los algoritmos y herramientas a utilizar para dar respuesta a las necesidades del cliente, de acuerdo a las 5 posibles situaciones iniciales.
3. **Minería de datos,** se implementan los algoritmos de minería de datos, determinando las variables de entrada y la clase, la adaptación de los valores faltantes según el modelo y la división del set de datos para validar la correcta generalización del patrón extraído.
4. **Refinamiento,** se evalúa y optimiza el modelo resultante, resolviéndose distintas problemáticas asociadas con la performance del modelo y los datos (parcialidad, falta de datos, sobre entrenamiento, etc.).
5. **Despliegue,** se informa a los interesados de los pasos desarrollados y las piezas de conocimiento obtenidas.

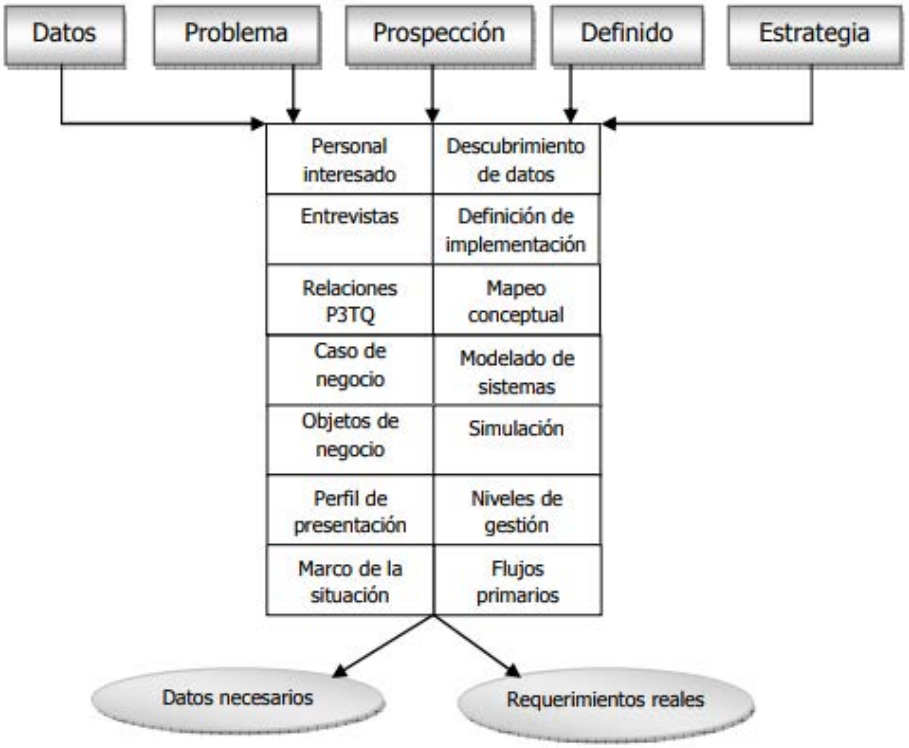


Figura 2.6. Modelo de proceso: Catalyst.

2.2.5. Un Modelo de Proceso para Ingeniería de Minería de Datos

La propuesta realizada en [Marbán et al., 2007] (de ahora en adelante mencionada como MPIMD), consiste en integrar a las propuestas KDD y CRISP-DM elementos de los estándares para el desarrollo de sistemas de información convencionales IEEE estándar 1074, ISO/IEC 12207 e ISO/IEC 15504. Su objetivo es cubrir la vacancia de gestión de proyectos señalada al estándar de facto para proyectos de explotación de información (CRISP-DM) [Marbán et al., 2009].

En este contexto, la solución se conforma de cuatro procesos (Figura 2.7, adaptada de [Marbán et al., 2009]) Organización, Administración del Proyecto, Desarrollo (integrado por tres sub-procesos: Pre-Desarrollo, Desarrollo y Post-Desarrollo) e Integral, cada uno de los cuales se encuentra integrado por actividades. El proceso Desarrollo, se estructura a partir del proceso KDD, mientras que los restantes se basan en los estándares de ingeniería de software. En adición, se utilizan los estándares IEEE e ISO/IEC y CRISP-DM para cubrir los alcances de las tareas a realizar o señalarlas como base para su adaptación. A continuación se resumen el alcance de cada uno de los procesos:

1. **Organización**: se llevan a cabo el conjunto de actividades orientadas a una organización más efectiva del proceso, la definición de los objetivos del negocio y la mejora del proceso,

producto y recursos de la organización. Se propone la adaptación de las tareas propuestas en el estándar ISO/IEC 15504.

2. **Administración del Proyecto**: se establece la estructura del proyecto, la coordinación y gestión de los recursos durante el desarrollo del mismo. Esta fase cubre la definición de un ciclo de vida, subcontratación, estimación y asignación de recursos y la utilización de métricas. En este proceso, se identifican técnicas para estimación de esfuerzo y posibles métricas a utilizar, mientras que las demás actividades no se especifica su abordaje (selección del ciclo de vida, planificación del proyecto y control y monitoreo del proyecto) o se propone su adaptación del estándar ISO/IEC 12207 (Adquisición y Suministro).
3. **Pre-Desarrollo**: cubre todas las tareas a realizar antes que el proyecto inicie, esto es, identificar las necesidades o ideas, problemas, soluciones potenciales y realizar estudios de viabilidad del proyecto, determinar los criterios de éxito, evaluar las técnicas y herramientas, modelar el negocio y reutilizar el conocimiento o modelos existentes. Sin embargo, algunas de los alcances mencionados no son descriptos respecto a su desarrollo (exploración de conceptos, Modelado del Negocio e importación del conocimiento).
4. **Desarrollo**: basado en el proceso KDD, abarca las actividades de selección, pre-procesamiento (entender los datos e identificar problemas de calidad) y transformación de los datos, minería de datos y análisis de los resultados. Adicionalmente, se realiza un mapeo entre las fases de KDD y CRISP-DM, aunque no se detalla el alcance de las tareas.
5. **Post-Desarrollo**: abarca aquellas actividades a realizar luego de que el conocimiento fue extraído. En [Marbán et al., 2009] los autores señalan como alcance del proceso, realizar la transferencia del conocimiento obtenido como resultado del proceso para dar soporte a futuras decisiones “o puede necesitarse desarrollar algún software”, requiriendo la instalación y aceptación del producto, la validación de los resultados, el soporte técnico al cliente, mantenimiento y retiro. Este elemento presenta contradicciones con la concepción del modelo que realizan los autores respecto a la producción de software como resultado del proceso, señalando que un proyecto de minería de datos no produce software. En este contexto, las actividades de instalación (excluyendo la presentación del conocimiento), mantenimiento, operación y soporte y retiro, no corresponden con los objetivos del proyecto, entendiendo que una vez aplicados los resultados al dominio analizado, el estado del mismo se ve alterado por un nuevo accionar.
6. **Integral**: estas actividades se realizan en paralelo al proceso de desarrollo y tienen como objetivo garantizar la integridad y calidad del proyecto. Estas abarcan la realización de controles para detectar defectos en el producto o el proceso, gestión de cambios y documentación, y entrenamiento del usuario. Los autores señalan que las actividades para la

evaluación y gestión de la configuración pueden ser adaptadas de los estándares de ingeniería de software, mientras que la actividad de entrenamiento del usuario no se encuentra definida y no es clara la necesidad o alcance de dicha tarea (teniendo en consideración que en el proceso de post-desarrollo se señala como objetivo la presentación de los resultados obtenidos, garantizando la correcta transferencia de los mismos).

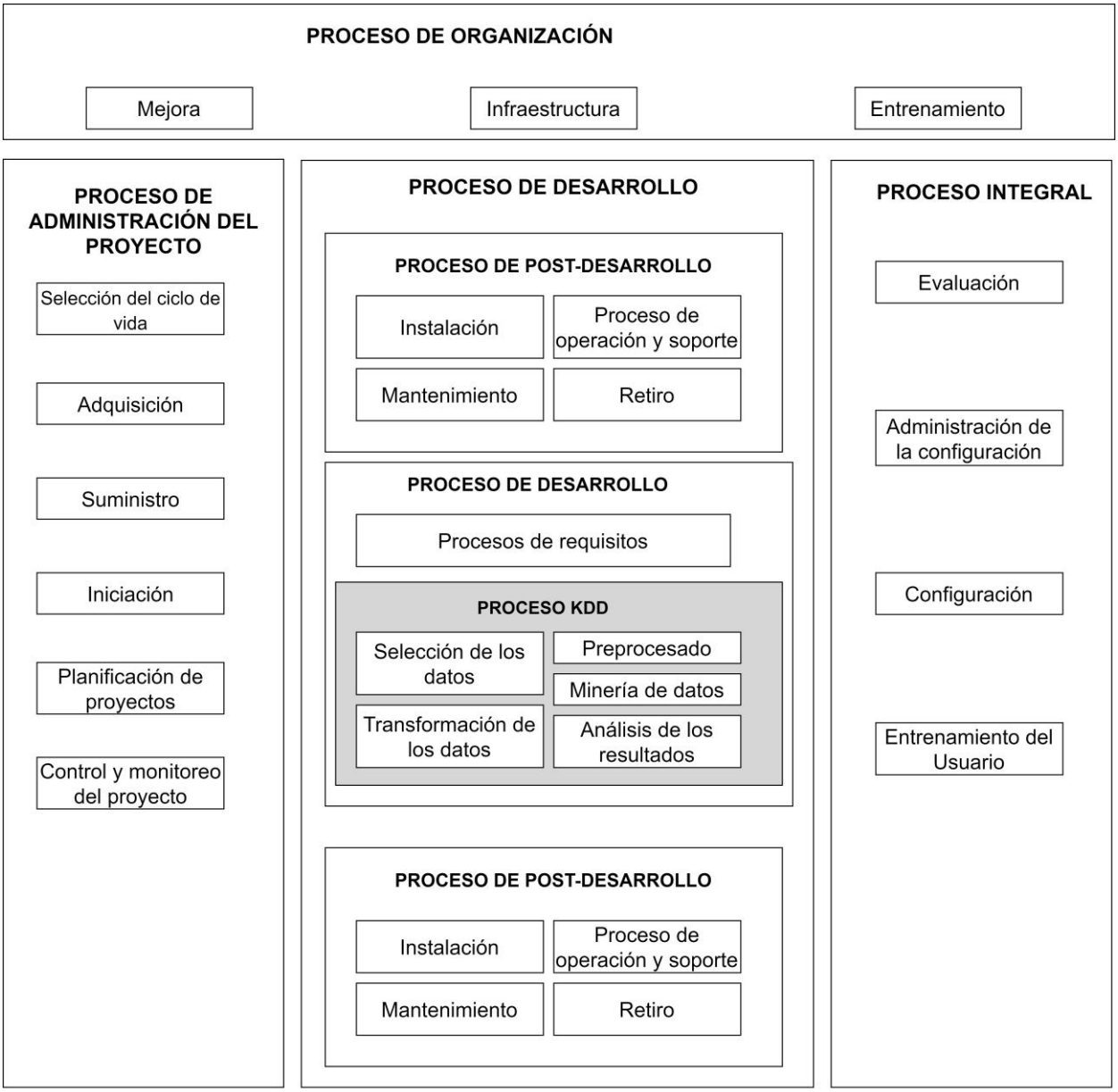


Figura 2.7. Modelo de Proceso: MPIMD.

2.2.6. Proceso Integrado de Descubrimiento de Conocimiento y Minería de Datos (IKDDM)

En [Sharma, 2008] se identifican una serie de limitaciones al *estándar de facto* (CRISP-DM), las

cuales motivaron la definición de la metodología: Proceso Integrado de Descubrimiento de Conocimiento y Minería de Datos (IKDDM, de sus siglas en inglés Integrated Knowledge Discovery and Data Mining). La misma fue definida orientada al usuario, proporcionando una visión detallada de las tareas y las dependencias del proceso, proveyendo al usuario de técnicas y procedimientos que lo guíen en el desarrollo de cada una de las actividades. La concepción de la propuesta se realiza desde una punto de vista integrado (y no fragmentado como las propuestas previas [Sharma et al., 2012]) en el cual se prevén las relaciones (o dependencias) entre las fases del proceso con el objetivo de minimizar la necesidad de iteraciones durante el desarrollo del proyecto.

La figura 2.8 (adaptada de [Sharma et al., 2012]), ilustra las fases que conforman la metodología, las cuales respetan la estructura de su proceso base. Estas son:

1. **Comprensión del Negocio,** se llevan a cabo tareas de entendimiento de los objetivos y requerimientos del proyecto desde la perspectiva del negocio, definiendo a partir de ellos los problemas de minería de datos y el diseño del plan del proyecto.
2. **Comprensión de los Datos,** se inicia con una recolección inicial de los datos, para luego realizar una evaluación de su calidad y detectar posibles subconjuntos de interés para desarrollar hipótesis de la información oculta.
3. **Preparación de los Datos,** abarca la construcción del set de datos final (el cual será utilizado en herramientas de modelado), incluyendo la transformación y limpieza.
4. **Modelado (o Minería de Datos),** se seleccionan las técnicas y se calibran sus parámetros. Los autores señalan la dependencia respecto a estructuras específicas de datos por algunos algoritmos, por lo cual la realización de tareas de transformación de los datos pueden volver a ser necesarias [Sharma et al., 2012].
5. **Evaluación,** se analiza el modelo generado para garantizar que este cumpla con los objetivos del negocio. Al finalizar esta etapa se debe determinar el uso de los resultados obtenidos.
6. **Despliegue,** la creación del modelo no es el último paso, sino que el conocimiento obtenido debe ser organizado y presentado de manera que los clientes puedan utilizarlo.

Si bien, la fase de despliegue es contemplada en la estructura de la propuesta, la autora señala que la misma no es cubierta en la metodología definida. En este contexto, se considera relevante destacar que el aporte de la propuesta se encuentra en la descripción detallada de los alcances de las fases y actividades, junto con la introducción de técnicas y procedimientos que norman las tareas a realizar. Adicionalmente, se señala que si bien los autores realizan un análisis detallado de las dependencias entre las entradas y salidas de las actividades, no se realiza un ajuste al proceso a partir de las conclusiones derivadas.

2.2.7. Desarrollo de Software Adaptativo para Inteligencia de Negocios (ASD-BI)

En [Alnoukari, 2010] se define la propuesta Desarrollo de Software Adaptativo para Inteligencia de Negocios [Alnoukari, 2010] (ASD-BI, de sus siglas en inglés Adaptive Software Development –

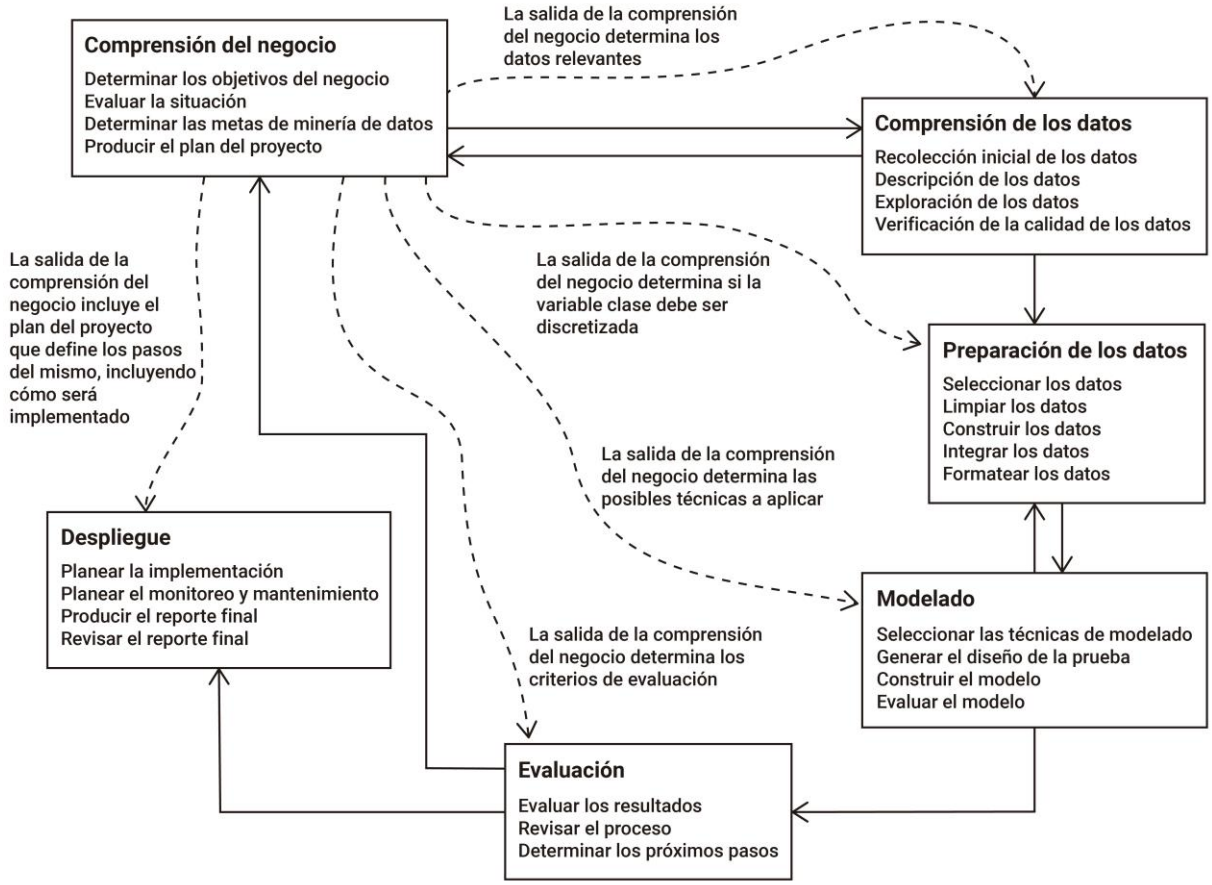


Figura 2.8. Metodología: IKDDM

Business Intelligence), la cual se basa en el método ágil Desarrollo de Software Adaptativo. Esta propuesta presenta un nuevo marco de desarrollo de procesos (o ciclo de vida) alternativo a la tradicional estructura secuencial. Este se encuentra conformado por tres etapas generales: Especulación, Colaboración y Aprendizaje. La primera de ellas conlleva la planificación adaptativa del proyecto, en la cual se determinan los objetivos y misiones de cada iteración del ciclo, así como una estimación preliminar del tamaño y alcances, y la identificación de posibles riesgos. La segunda incorpora el rol de las personas en el procedimiento constructivo, abordando las tareas asociadas a la comunicación del equipo y las cuestiones relativas a la concurrencia de la entrega de componentes. La última fase, reconoce el incremento del conocimiento mediante la experiencia, la cual se centra en las revisiones de calidad desde distintas perspectivas: cliente, técnica, prácticas y estado del proyecto.

Este modelo de proceso, incluye a las etapas definidas en CRISP-DM, la etapa de *Configuración de los objetivos/hipótesis*. La figura 2.9 (adaptada de [Alnoukari y El Sheikh, 2012]), ilustra la estructura de la propuesta (en la cual se omitieron los roles asociados por simplicidad), donde se visualizan las tres fases del método de gestión de proyectos: Especulación, Colaboración y Aprendizaje, y en cada una de ellas, se identifican las etapas a realizar. Adicionalmente, se señalan las fuentes de información de interés para cada una de las fases. A continuación se presentan los alcances de cada una de ellas definidos en [Alnoukari y El Sheikh, 2012]:

1. **Entendimiento del Negocio:** se enfoca principalmente en el entendimiento de los objetivos y requerimientos del proyecto desde las perspectivas de negocio. En esta etapa se realizan distintas actividades: definir los objetivos y criterios de éxito del negocio, evaluar la situación, delimitar los objetivos desde la perspectiva de la inteligencia de negocios y determinar las reglas básicas del negocio.
2. **Entendimiento de los Datos:** se centra en ayudar a los usuarios a entender los datos para abordar los problemas que puedan ocurrir en la siguiente etapa. En ella se realizan tareas de recolección inicial, exploración, descripción y verificación de la calidad de los datos.
3. **Configuración de los Objetivos/Hipótesis:** esta etapa utiliza las salidas de las dos anteriores para formular las metas e hipótesis de la aplicación de la inteligencia de negocio, a partir de los cuales puede resaltar nuevas ideas de negocio.
4. **Preparación de los Datos (o ETL):** se realizan todas las actividades requeridas para construir el set de datos requerido para la etapa de modelado. Las tareas involucradas son: seleccionar, limpiar, construir, integrar y formatear los datos.
5. **Modelado (o Minería de Datos):** se centra en la selección de los métodos o algoritmos de minería de datos apropiados para los problemas de inteligencia de negocio del proyecto. El objetivo de esta etapa es llevar a cabo tareas de análisis utilizando diferentes modelos de minería de datos o una combinación de ellos.
6. **Evaluación:** se asegura que los métodos o modelos de minería de datos seleccionados sean apropiados para alcanzar los objetivos del proyecto, almacenando los resultados en el repositorio de conocimiento para que puedan ser utilizados en proyectos futuros.
7. **Implementación:** se organizan y presentan los resultados obtenidos en la etapa de modelado, así como se realizan actividades de revisión del estado del proyecto, en orden a evaluar los resultados del ciclo actual y preparar las actividades de la próxima iteración.

2.2.8. Metodología Fundacional para la Ciencia de Datos (FMDS)

Metodología Fundacional para la Ciencia de Datos (FMDS, de sus siglas en inglés Foundational Methodology for Data Science) [Rollins, 2015] es una propuesta realizada por IBM la cual adopta muchas de las características de los modelos de proceso KDD y CRISP-DM, incorporando nuevas prácticas de procesamiento de grandes volúmenes de datos, analítica texto and imagen, inteligencia artificial, Aprendizaje Profundo (Deep Learning) y procesamiento de lenguajes. El modelo de proceso está conformado por 10 etapas (Figura 2.10, adaptada de [Rollins, 2015]):

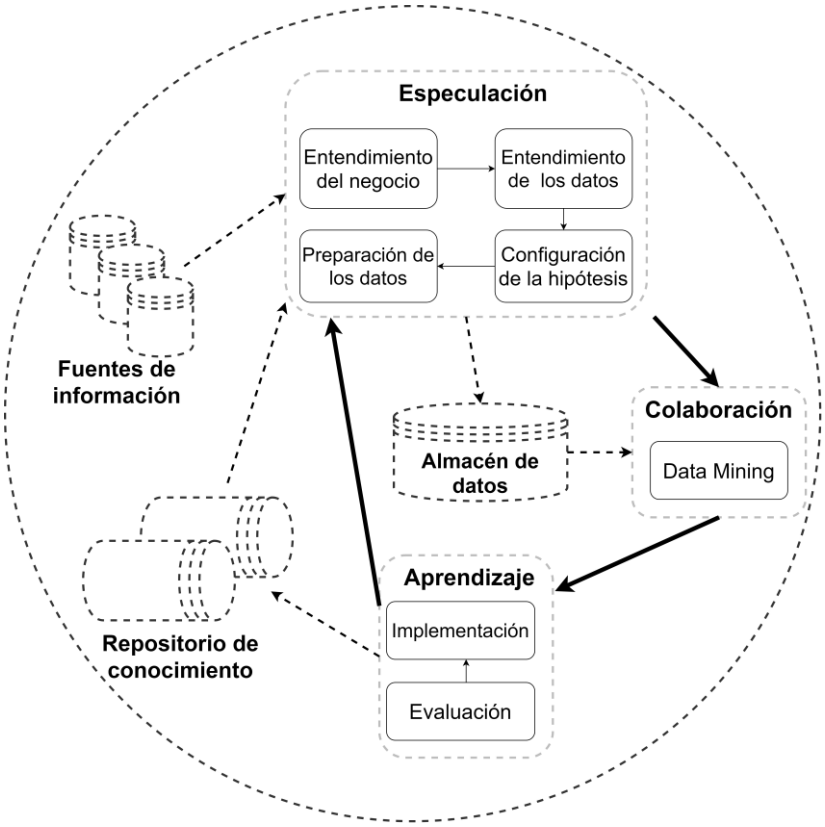


Figura 2.9. Modelo de Proceso: ASD-BI.

1. **Entendimiento del Negocio**: Esta primera fase define las bases para una solución rentable y eficaz de un problema empresarial.
2. **Enfoque analítico**: se determina un enfoque analítico mediante la identificación de una técnica de aprendizaje automático adecuada para resolver el problema identificado.
3. **Requisitos de datos**: a partir de lo definido en la etapa previa, se definen los requisitos en los datos para resolver el problema.
4. **Recopilación de datos**: se identifican y recopilan los recursos de datos disponibles que están relacionados y de interés al problema.
5. **Entendimiento de los datos**: se aplican técnicas de estadística descriptiva y visualizaciones para comprender los datos. Esta etapa puede requerir revisar las fases previas.

6. **Preparación de los datos**: se realizan acciones para construir el set de datos de acuerdo a las necesidades del modelo a aplicar en la fase de modelado.
7. **Modelado**: en esta fase se desarrolla el modelo descriptivo o predictivo a partir de los aspectos elaborados en las etapas anteriores.
8. **Evaluación**: se evalúa la calidad y eficacia del modelo desarrollado, determinando si satisface la problemática definida.
9. **Implementación**: en esta etapa se realiza la transferencia de los resultados (instalación en producción o un reporte).
10. **Retroalimentación**: se recolectan los resultados del modelo implementado para analizar la satisfacción y eficiencia de la solución.

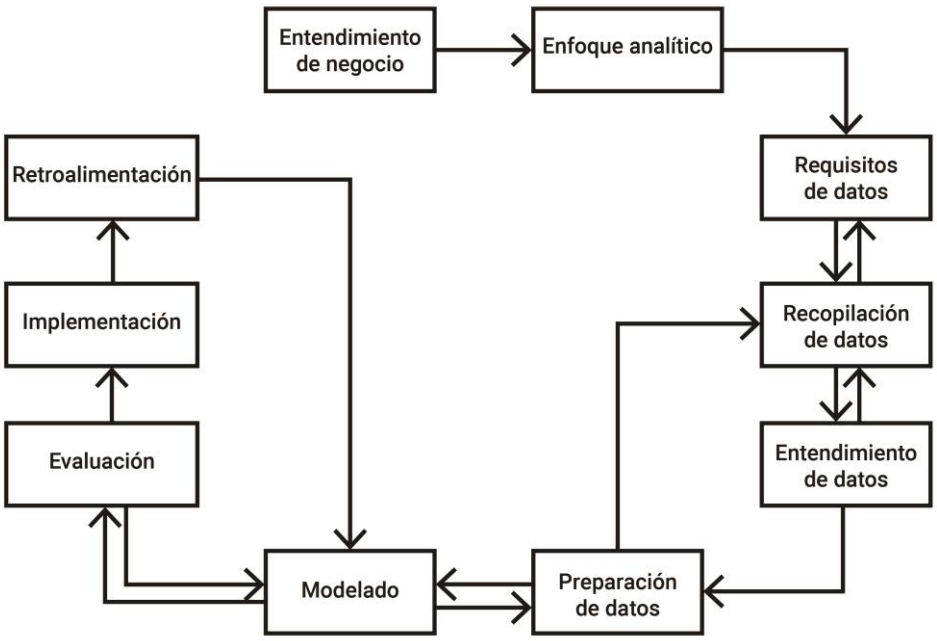


Figura 2.10. Modelo de Proceso: FMDS.

2.2.9. Proceso de Ciencia de Datos en Equipo (TDSP)

En 2016, Microsoft define el Proceso de Ciencia de Datos en Equipo (TDSP, de sus siglas en inglés Team Data Science Process) [Microsoft, 2016]. Una propuesta basada en un ciclo de vida ágil, enfocada en mejorar el aprendizaje y la colaboración del equipo. La propuesta establece una estructura estandarizada del proyecto, definiendo las herramientas y plantillas para el desarrollo de las tareas. La metodología está compuesta por 5 fases (Figura 2.11, adaptada de [Microsoft, 2016]):

1. **Entendimiento del negocio**: se identifica la pregunta que define el problema/objetivo, y se identifican las fuentes de datos y los modelos predictivos relevantes.

2. **Adquisición y entendimiento de los datos**: se realiza la recolección, análisis, limpieza y adecuación de los datos para implementar los modelos.
3. **Modelado**: se transforman los datos para aplicar el entrenamiento del modelo. Además, se realiza la selección, creación y evaluación del modelo.
4. **Implementación**: Se realiza la puesta en producción del modelo y el pipeline de datos.
5. **Validación del Cliente**: El objetivo de esta fase es obtener la aceptación de la satisfacción del cliente.

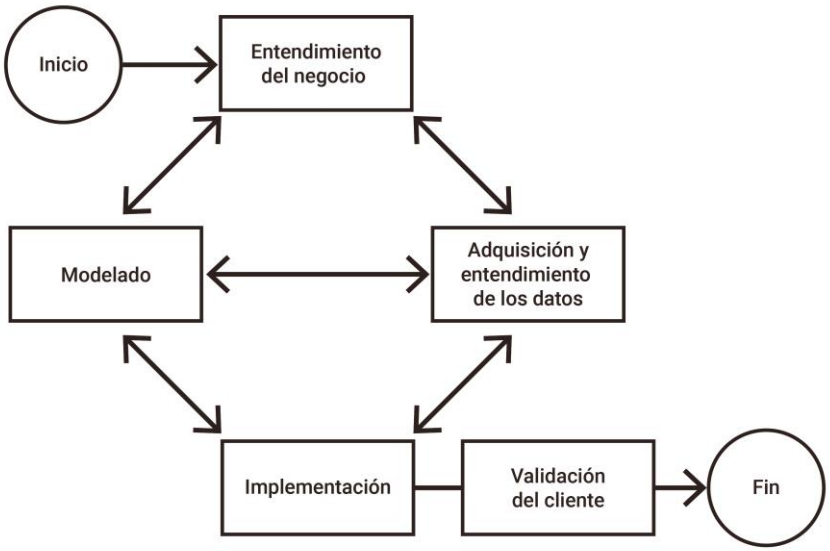


Figura 2.11. Modelo de Proceso: TDSP.

2.3. DISCUSIÓN DE LAS PROPUESTAS METODOLÓGICAS

En los siguientes párrafos se presentan las principales críticas identificadas en la disciplina a las propuestas descritas en la sección previa. En la sección 3.2 (página 57), se amplía el análisis existente en la literatura, realizando un análisis detallado de la estructura de las propuestas y sus principales características, concluyéndose las vacancias existentes en cada propuesta.

KDD es el primer abordaje en identificar la necesidad de un proceso que guíe el desarrollo de proyectos de explotación de información, siendo señalada como la piedra angular de las propuestas [Alnoukari y El Sheikh, 2012]. SEMMA, es otra propuesta ampliamente utilizada, gracias a su fuerte vinculación con las herramientas de explotación de información desarrolladas por la empresa SAS. Ambas propuestas son unas de las más utilizadas en la disciplina y comparten sus principales críticas: falta de perspectiva del negocio [Hofmann, 2003; Cios et al., 2007], falta de integridad y detalle para llevar a cabo un proyecto [Hofmann, 2003; Sharma, 2008; Sharma et al., 2012], carencia de la fase de implementación dificultando la evaluación de los resultados, restricciones en

el ciclo de vida, no se almacena el conocimiento para mejorar procesos futuros e ignora la participación de los recursos humanos en el proceso [Alnoukari y El Sheikh, 2012].

CRISP-DM es la propuesta más utilizada en la disciplina, siendo el primer modelo de proceso en identificar las fases de entendimiento del negocio y de los datos, etapas relevantes en el desarrollo del proyecto [Alnoukari y El Sheikh, 2012]. Sin embargo, se han señalado varias deficiencias a la misma: el conocimiento extraído no es almacenado ni utilizado [Alnoukari y El Sheikh, 2012], ignora la participación de los recursos humanos en el proceso [Hofmann, 2003], falta de integridad y detalle para guiar el desarrollo de un proyecto [Gartner, 2000; Hofmann, 2003; Marbán et al., 2007; Sharma, 2008; Sharma et al., 2012], limitaciones y restricciones en el ciclo de vida integrado a la propuesta [Rennolls y AL-Shawabkeh, 2008; Alnoukari y El Sheikh, 2012] con respecto a la secuencialidad en la naturaleza del mismo, carencia en las iteraciones, la última fase del proceso es un punto muerto, y falta de un método o proceso de selección de modelos [Rennolls y AL-Shawabkeh, 2008; Martins et al., 2014]. En adición, en [Marbán et al., 2007] se señala que las tres propuestas previamente mencionadas, carecen de gestión del proyecto y presentan limitaciones con respecto a la posibilidad de adaptación del desarrollo del proyecto a las necesidades del mismo.

MPIMD [Marbán et al., 2007], se destaca por ser el primer proceso en señalar la necesidad de una visión ingenieril, a partir de la creciente complejidad de los proyectos y los problemas asociados. Sin embargo, los autores señalan en [Marbán et al., 2009] que el modelo no está completo, limitándose a señalar la necesidad de las actividades faltantes en CRISP-DM, por lo cual la propuesta puede ser identificada como bosquejo de proceso para proyectos de ingeniería de explotación de información. Adicionalmente, presenta dificultades en su comprensión e implementación a causa de *contradicciones en su concepción*: en [Marbán et al., 2009], los autores señalan la diferencia en el objetivo de la disciplina (con respecto a la ingeniería de software) en que su resultado es la extracción de conocimiento de los datos y no el desarrollo de un producto software, sin embargo la propuesta incluye actividades como instalación, mantenimiento y retiro señalando que "puede necesitarse desarrollar algún software", pero no se incluye una paso de programación, *falta de claridad en la definición de los alcances de las actividades*: dejando sin definir o sin describir cómo se adaptan aquellas actividades adoptadas de la ingeniería de software (por ejemplo, del proceso de administración del proyecto: selección del ciclo de vida, planificación del proyecto y control y monitoreo del proyecto, del proceso Pre-Desarrollo: exploración de conceptos, Modelado del Negocio, entre otras), así como actividades adoptadas de CRISP-DM cuyo alcance se sobrepone con actividades incorporadas de los modelos de procesos de Ingeniería de Software (por ejemplo: las actividades de instalación, y operación y soporte pertenecientes al proceso post-desarrollo y entrenamiento del usuario perteneciente al proceso integral tienen como

objetivo la presentación de los resultados obtenidos, garantizando la correcta transferencia de los mismos, aspectos ya contemplados en la propuesta base) y *persistencia en las críticas realizadas al modelo de proceso de explotación de información base*: falta de integridad y detalle para guiar el desarrollo de un proyecto, e ignora la participación de los recursos humanos en el proceso.

IKDDM [Sharma, 2008], busca solucionar la crítica realizada a la propuesta base (CRISP-DM) respecto a la dificultad en su implementación (falta de integridad y detalle para llevar a cabo un proyecto), proponiendo una metodología que detalla los pasos a realizar, sin embargo, las demás críticas realizadas a la propuesta base aún persisten.

ASD-BI [Alnoukari, 2010], pretende resolver las críticas realizadas a CRISP-DM respecto a las limitaciones y restricciones en el ciclo de vida y su secuencialidad, proponiendo un modelo de proceso basado en los métodos ágiles, sin embargo, las demás críticas realizadas a la propuesta base aún persisten.

FMDS [Rollins, 2015], pretende resolver la limitación del ciclo de vida respecto a la falta de pasos siguientes en la última etapa del proceso, sin embargo, la posibilidad de iteración está limitada a las últimas fases. Además críticas realizadas a la propuesta base aún persisten.

TDSP [Microsoft, 2016] se enfoca en proveer un mayor nivel de detalle de las actividades, incorporando plantillas para los elementos de salida de cada etapa y la contemplación del recurso humano en el proceso, sin embargo, las restantes críticas mencionadas a CRISP-DM persisten.

2.4. TÉCNICAS

En esta sección se describen las técnicas *ad hoc* desarrolladas para proyectos de ingeniería de explotación de información (sección 2.4.1) y aquellas de gestión de proyectos aplicables a la disciplina (sección 2.4.2).

2.4.1. Técnicas ad hoc para Ingeniería de Explotación de Información

En los últimos años, el grupo de investigación en el cual se enmarca esta propuesta, ha estado trabajando en un conjunto de técnicas *ad hoc* para el desarrollo de proyectos de ingeniería de explotación de información con el objetivo de cubrir las vacancias identificadas en las distintas etapas del proceso. En este contexto, se describen brevemente cada una de los métodos utilizados en la propuesta: evaluación de herramientas de explotación de información (sección 2.4.1.1), educción de requerimientos para proyectos de explotación de información (sección 2.4.1.2), procesos de explotación de información (sección 2.4.1.3), derivación del proceso de explotación de información

(sección 2.4.1.4), modelo de viabilidad (sección 2.4.1.5), modelo de estimación (sección 2.4.1.6), métricas (sección 2.4.1.7) y ciclo de vida (sección 2.4.1.8).

2.4.1.1. Evaluación de Herramientas de Explotación de Información

En [Britos, et. Al., 2006], se propone un método de selección de herramientas de explotación de información, el cual evalúa cuatro características generales del producto: técnico/funcionales, del proveedor, del servicio y económicas. Cada una de estas características posee un peso (o ponderación), el cual determina el impacto de la misma respecto a la valoración total de la herramienta. A su vez, cada una de ellas está compuesta por un conjunto de aspectos ponderados (los cuales pueden asignarse un valor del 1 al 4), a partir de los cuales se obtiene el valor de las características generales. Las tablas 2.2.a y 2.2.b (adaptadas de [Britos et al., 2006]) ilustran la estructura, los criterios de evaluación utilizados y sus pesos.

Criterios de selección de Herramienta		Peso	Herramienta 1	…	Herramienta N
1. Funcional - Características Técnicas					
Soporte de Metodología / Ciclo de vida	Soporte del proceso	3	Valor asignado al criterio	…	Valor asignado al criterio
Compatibilidad con fuentes de datos	Base de datos	8	Valor asignado al criterio	…	Valor asignado al criterio
	Otras fuentes (word, excel, etc.)	8	Valor asignado al criterio	…	Valor asignado al criterio
Integración	Soporte de distintas técnicas asociadas al proceso de explotación de Información	5	Valor asignado al criterio	…	Valor asignado al criterio
Multilenguaje	Soporta distintas idiomas	2	Valor asignado al criterio	…	Valor asignado al criterio
Técnicas	Variedad de técnicas que provee	18	Valor asignado al criterio	…	Valor asignado al criterio
Reporte y visualización	Permite generar reportes y visualizaciones	12	Valor asignado al criterio	…	Valor asignado al criterio
Multiplataforma	Soporta múltiples plataformas	5	Valor asignado al criterio	…	Valor asignado al criterio
Instalación remota	La administración y mantenimiento son remotos	5	Valor asignado al criterio	…	Valor asignado al criterio
Usuarios Múltiples	Posee perfiles de usuarios	2	Valor asignado al criterio	…	Valor asignado al criterio
Seguridad	Provee seguridad de la información configurada por perfiles	2	Valor asignado al criterio	…	Valor asignado al criterio
Backup	Metodología de backup	2	Valor asignado al criterio	…	Valor asignado al criterio
Amigable	Interfaz de usuario	10	Valor asignado al criterio	…	Valor asignado al criterio
Configuraciones	Permite la configuración del perfil	8	Valor asignado al criterio	…	Valor asignado al criterio
Documentación	Servicio de soporte y ayuda	5	Valor asignado al criterio	…	Valor asignado al criterio
Conexión	Soporta conexión por: Internet, FTP, ERPs.	2	Valor asignado al criterio	…	Valor asignado al criterio
Soporte de sistemas de mensaje	Soporta compartir información (por mail u otro medio)	3	Valor asignado al criterio	…	Valor asignado al criterio
Total			Valor total sección 1	…	Valor total sección 1
	Peso del Grupo	40%	Valor total ponderado sección 1	…	Valor total ponderado sección 1

Tabla 2.2.a. Técnica de Evaluación de herramientas.

2. Características del Proveedor					
Características del proveedor	Historia	30	Valor asignado al criterio	...	Valor asignado al criterio
Crecimiento	Perspectiva a futuro	10	Valor asignado al criterio	...	Valor asignado al criterio
Ubicación Geográfica	Oficinas	30	Valor asignado al criterio	...	Valor asignado al criterio
Implementación	Otras implementaciones de la misma herramienta	5	Valor asignado al criterio	...	Valor asignado al criterio
	Contacto con otros clientes	5	Valor asignado al criterio	...	Valor asignado al criterio
Confidencialidad	Confidencialidad de la información	20	Valor asignado al criterio	...	Valor asignado al criterio
Total			Valor total sección 2	...	Valor total sección 2
	Peso del Grupo	25%	Valor total ponderado sección 2	...	Valor total ponderado sección 2
3. Características del Servicio					
Garantía del producto	Duración y Alcance	30	Valor asignado al criterio	...	Valor asignado al criterio
Mejora	Brinda soporte a versiones previas	20	Valor asignado al criterio	...	Valor asignado al criterio
Licencia	Costo, alcances y soporte postventa	30	Valor asignado al criterio	...	Valor asignado al criterio
Soporte	Tiempo de respuesta y disponibilidad	20	Valor asignado al criterio	...	Valor asignado al criterio
Total			Valor total sección 3	...	Valor total sección 3
	Peso del Grupo	20%	Valor total ponderado sección 3	...	Valor total ponderado sección 3
4. Características Económicas					
Costo del software	Costo de la herramienta	30	Valor asignado al criterio	...	Valor asignado al criterio
Costo del Hardware	Necesidad de mejorar o comprar nuevo hardware compatible con la herramienta	20	Valor asignado al criterio	...	Valor asignado al criterio
Otros costos software	Costos adicionales al producto (backup, web servers, bases de datos, etc.)	20	Valor asignado al criterio	...	Valor asignado al criterio
Licencias	Política de licencia	10	Valor asignado al criterio	...	Valor asignado al criterio
Financiamiento	Existencia	10	Valor asignado al criterio	...	Valor asignado al criterio
Mejoras	Costo promedio de la mejora del producto	10	Valor asignado al criterio	...	Valor asignado al criterio
Total			Valor total sección 4	...	Valor total sección 4
	Peso del Grupo	-15%	Valor total ponderado sección 4	...	Valor total ponderado sección 4
Final					
1. Funcional - Características Técnicas		40%	Valor total ponderado sección 1	...	Valor total ponderado sección 1
2. Características del Proveedor		25%	Valor total ponderado sección 2	...	Valor total ponderado sección 2
3. Características del Servicio		20%	Valor total ponderado sección 3	...	Valor total ponderado sección 3
4. Características Económicas		-15%	Valor total ponderado sección 4	...	Valor total ponderado sección 4
TOTAL			**Valoración final de la herramienta 1**	...	**Valoración final de la herramienta N**

Tabla 2.2.b. Técnica de Evaluación de herramientas.

2.4.1.2. Educción de Requerimientos para Proyectos de Explotación de Información

En [Britos et al., 2008] se identifica la necesidad de adaptar los procesos tradicionales de ingeniería de requisitos para proyectos de explotación de información, dada las diferencias sustanciales en los objetivos de este tipo de proyectos. A partir del estudio de las principales dificultades identificadas en la disciplina en la etapa de entendimiento del negocio, se identifica el conjunto de conceptos involucrados en esta etapa y sus dependencias, definiendo un proceso y un conjunto de formalismos que guían el desarrollo de la tarea. En la figura 2.12 (adaptada de [Britos et al., 2008]), se observan los conceptos identificados y sus dependencias, los cuales representan el conjunto de conocimientos de interés para el entendimiento de las necesidades del cliente.

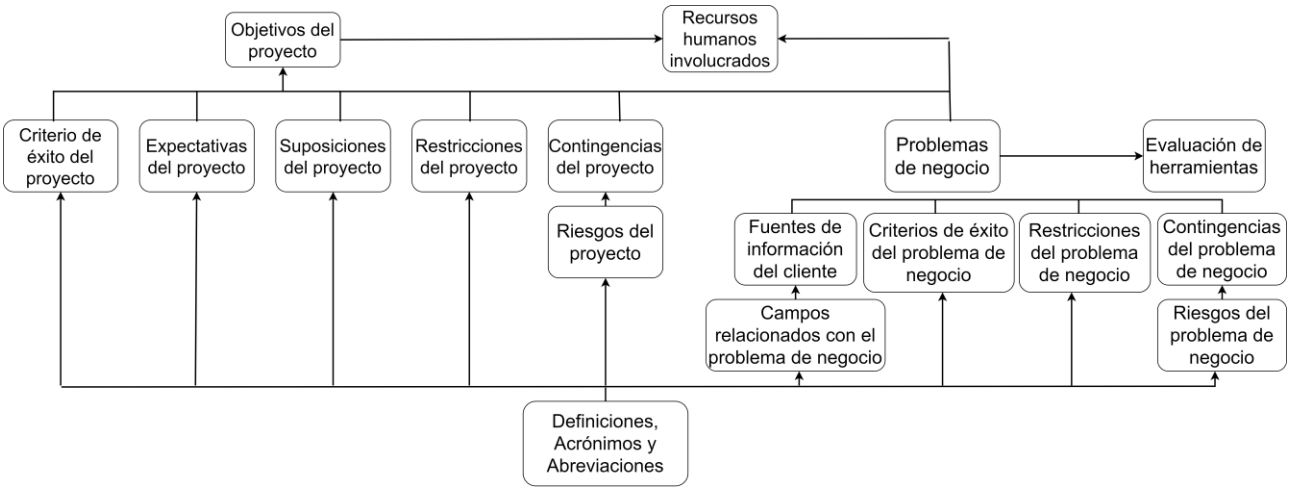


Figura 2.12. Educción de Requerimientos - Dependencia entre conceptos representados por formalismo.

2.4.1.3. Procesos de Explotación de Información

Los procesos de explotación de información definen las técnicas o algoritmos a utilizar en base a las características y necesidades del problema de explotación de información. En [Britos y García-Martínez, 2009; García-Martínez et al., 2013] se presentan cinco procesos de explotación de información: descubrimiento de reglas de comportamiento, descubrimiento de grupos, descubrimiento de atributos significativos, descubrimiento de reglas de pertenencia a grupos y ponderación de reglas de comportamiento o de pertenencia.

El proceso de descubrimiento de reglas de comportamiento (figura 2.13, adaptada de [García-Martínez et al., 2013]), aplica cuando se requiere identificar cuáles son las condiciones para obtener determinado resultado en el dominio del problema y se propone la utilización de algoritmos de inducción TDIDT para descubrir las reglas de comportamiento de cada atributo clase.

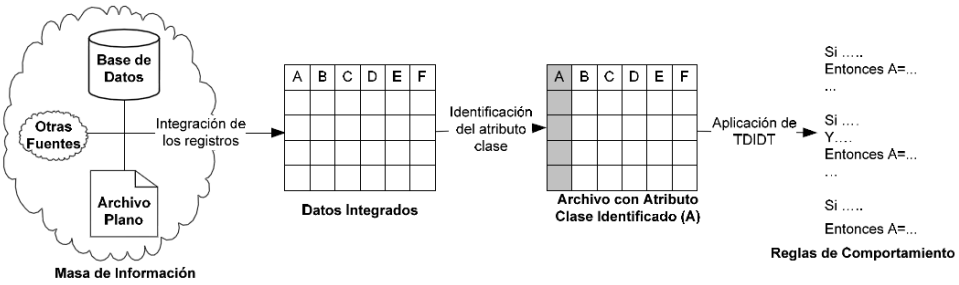


Figura 2.13. Proceso de descubrimiento de reglas de comportamiento.

El proceso de descubrimiento de grupos (figura 2.14, adaptada de [García-Martínez et al., 2013])), aplica cuando se requiere identificar una partición en la masa de información disponible sobre el dominio de problema y se propone el uso de algoritmos de agrupamiento o clustering (por ejemplo: mapas auto-organizados, SOM).

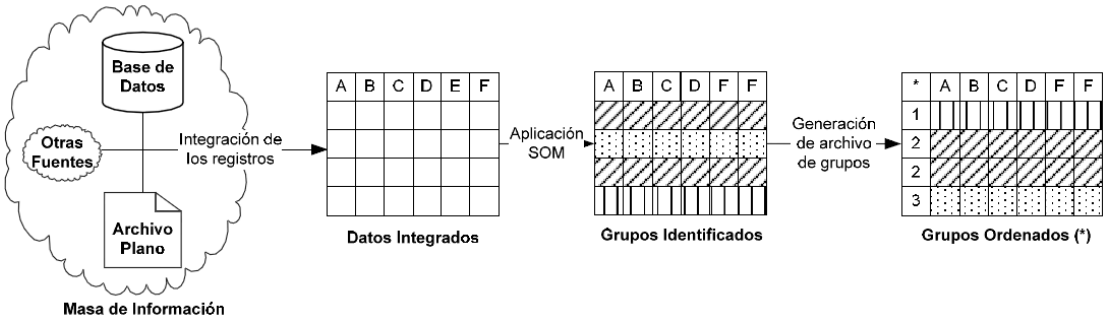


Figura 2.14. Proceso de descubrimiento de grupos

El proceso de ponderación de interdependencia de atributos (figura 2.15, adaptada de [García-Martínez et al., 2013])), aplica cuando se requiere identificar cuáles son los factores con mayor incidencia (o frecuencia de ocurrencia) sobre un determinado resultado del problema. Para ponderar en qué medida la variación de los valores de un atributo incide sobre la variación del valor de un atributo clase se propone la utilización de Redes Bayesianas.

El proceso de descubrimiento de reglas de pertenencia a grupos (figura 2.16, adaptada de [García-Martínez et al., 2013])), aplica cuando se requiere identificar cuáles son las condiciones de pertenencia a cada una de las clases en una partición desconocida “a priori”, pero presente en la masa de información disponible sobre el dominio de problema. Se propone la utilización de algoritmos de agrupamiento o clustering (por ejemplo: SOM) para el hallazgo de los grupos y; una vez identificados los mismos, la utilización de algoritmos de inducción (TDIDT) para establecer las reglas de pertenencia a cada uno.

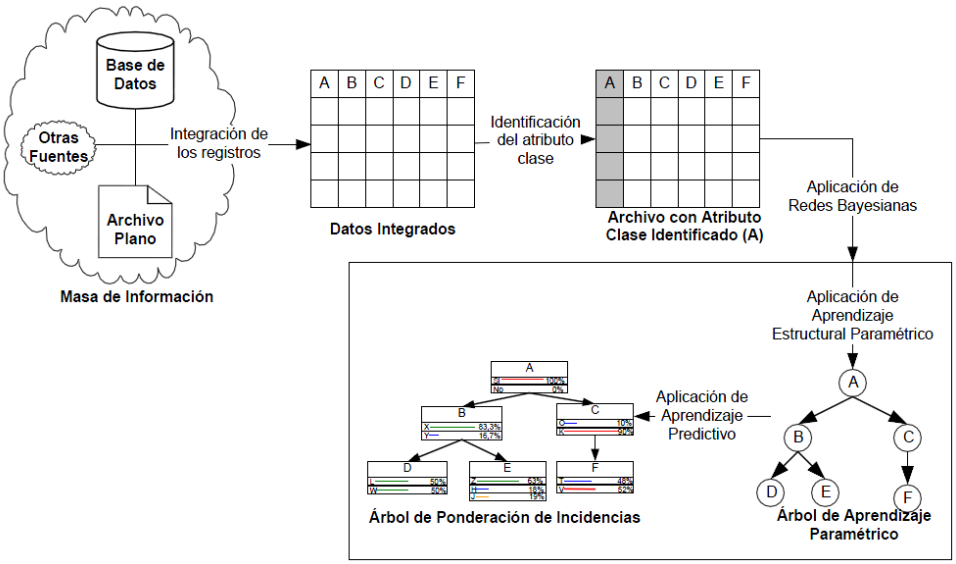


Figura 2.15. Proceso de ponderación de interdependencia de atributos.

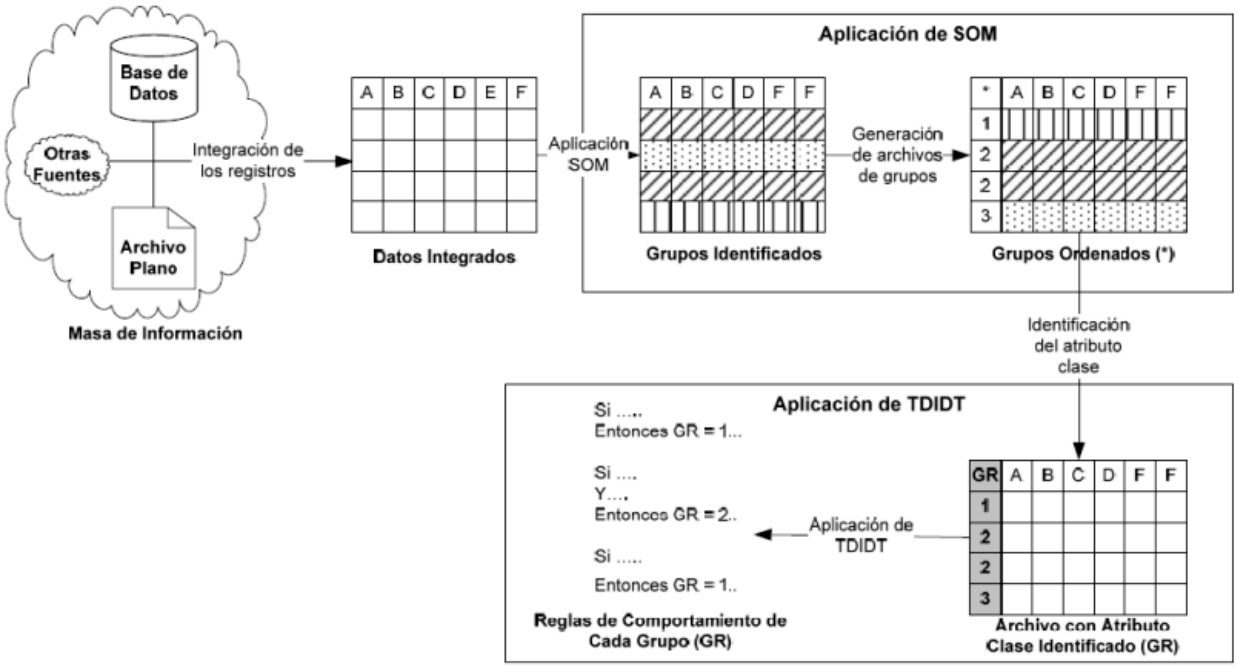


Figura 2.16. Proceso de descubrimiento de reglas de pertenencia a grupos.

El proceso de ponderación de reglas de comportamiento o de la pertenencia a grupos (figura 2.17, adaptada de [García-Martínez et al., 2013])), aplica cuando se requiere identificar cuáles son las condiciones con mayor incidencia (o frecuencia de ocurrencia) sobre la obtención de un determinado resultado en el dominio del problema, sean estas las que en mayor medida inciden sobre un comportamiento o las que mejor definen la pertenencia a un grupo. Para la ponderación de reglas de comportamiento o de pertenencia a grupos se propone la utilización de redes bayesianas. Esto puede hacerse a partir de dos procedimientos dependiendo de las características del problema a resolver: cuando no hay clases/grupos identificados; o cuando hay clases/grupos identificados. El

procedimiento a aplicar cuando hay clases/grupos identificados consiste en la utilización de algoritmos de inducción TDIDT para descubrir las reglas de comportamiento de cada atributo clase y posteriormente se utiliza redes bayesianas para descubrir cuál de los atributos establecidos como antecedentes de las reglas tiene mayor incidencia sobre el atributo establecido como consecuente. El procedimiento a aplicar cuando no hay clases/grupos identificados consiste en aplicar algoritmos de agrupamiento o clustering (por ejemplo: SOM) para el hallazgo de los grupos y; una vez identificados los mismos, la utilización de redes bayesianas para establecer la frecuencia (o incidencia) de cada atributos con respecto al atributo grupo.

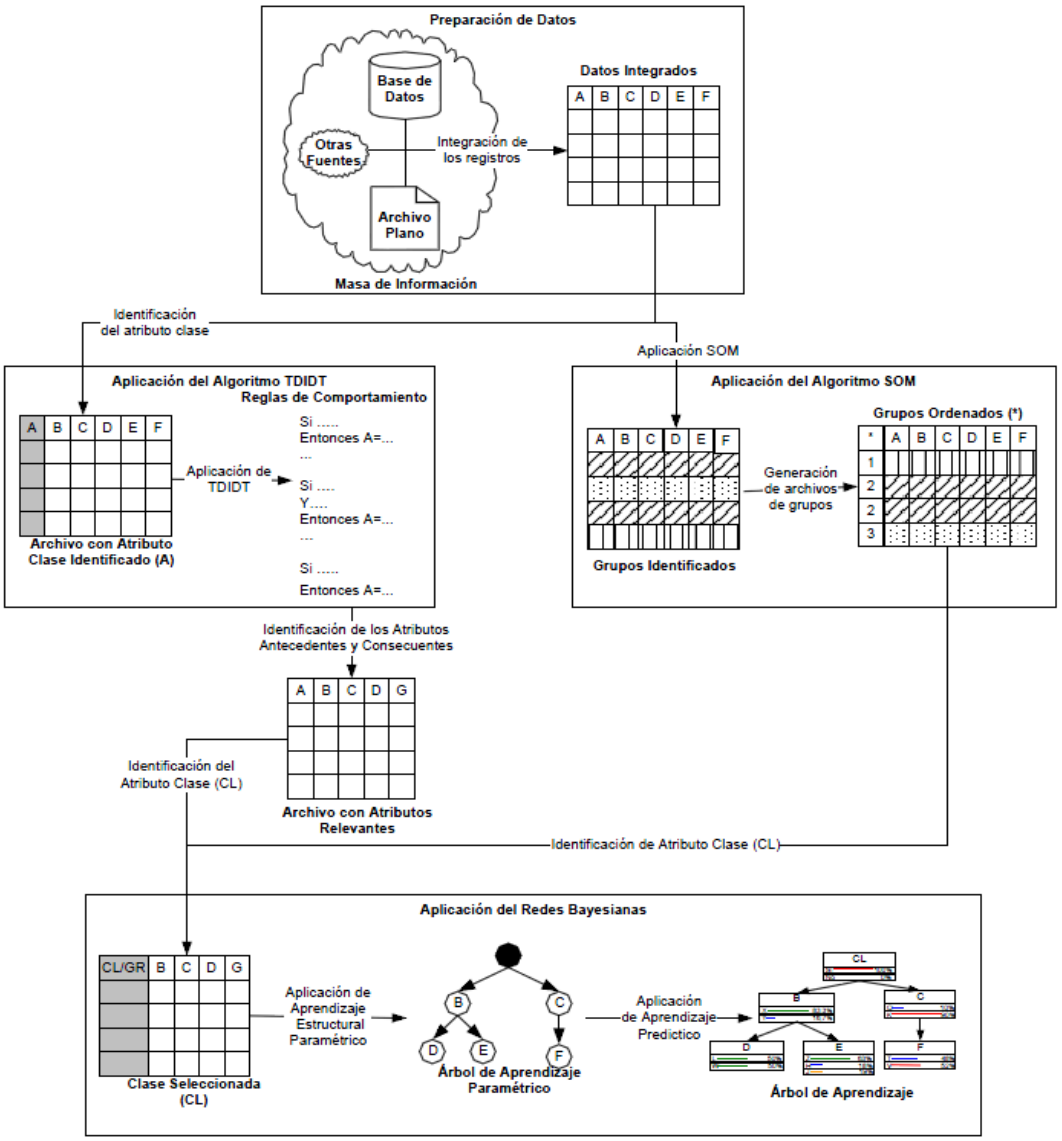


Figura 2.17. Proceso de ponderación de reglas de comportamiento o de la pertenencia a grupos.

2.4.1.4. Derivación de Procesos de Explotación de Información

En [Martins et al., 2014] se propone un proceso de modelado del negocio basados en marcos y redes semánticas, que permite a partir del conocimiento educido en la etapa de entendimiento del negocio y sus necesidades derivadas, determinar el modelo o proceso de explotación de información (sección 2.4.1.3) que da respuesta a dicha necesidad, y por consiguiente la familia de algoritmos a utilizar. Esta técnica se encuentra conformada por tres etapas: la primera destinada a modelar las características del dominio del negocio, la segunda a representar las necesidades o problemas del cliente, y la última proporciona un conjunto de pasos que permite identificar qué proceso de explotación de información utilizar para dar respuesta a dichas necesidades. La figura 2.18, ilustra los pasos de la técnica propuesta.

2.4.1.5. Modelo de Viabilidad para Proyectos de Explotación de Información

En [Pytel et al., 2015] se presenta un modelo de evaluación de la viabilidad de proyectos de explotación de información en el cual se evalúan cuatro aspectos generales del proyecto: Datos, Problema de Negocio, Proyecto y Equipo de trabajo. Cada uno de estos aspectos agrupa un conjunto de preguntas utilizadas para valorar las características del proyecto y por consiguiente la viabilidad del mismo. Dichas preguntas se encuentran clasificadas por dimensión: plausibilidad (P), adecuación (A) y éxito (E), las cuales proporcionan distintos indicadores en la viabilidad del proyecto. Cada una de las preguntas o características puede ser valorada mediante un valor de la escala lingüística (nada, poco, regular, mucho, todo) que se encuentra asociado a intervalos difusos (cuatro puntos críticos entre los valores 0 y 10).

Adicionalmente, cada característica tiene asociado un valor de umbral, que define la valoración mínima viable para el proyecto y un peso, que indica su ponderación con respecto a la dimensión a la cual pertenece. La tabla 2.3 [Pytel et al., 2015], detalla cada una de las características, con su umbral, peso, categoría y dimensión a la cual pertenece (nótese que la dimensión está indicada con la primera letra en mayúscula y un número de orden). A partir de las fórmulas 2.1 (pasos 1 y 2), y 2.2 [Pytel et al., 2015], se determinan los valores de viabilidad por dimensión y global (respectivamente), los cuales deben ser mayores a 5 para considerar al proyecto viable.

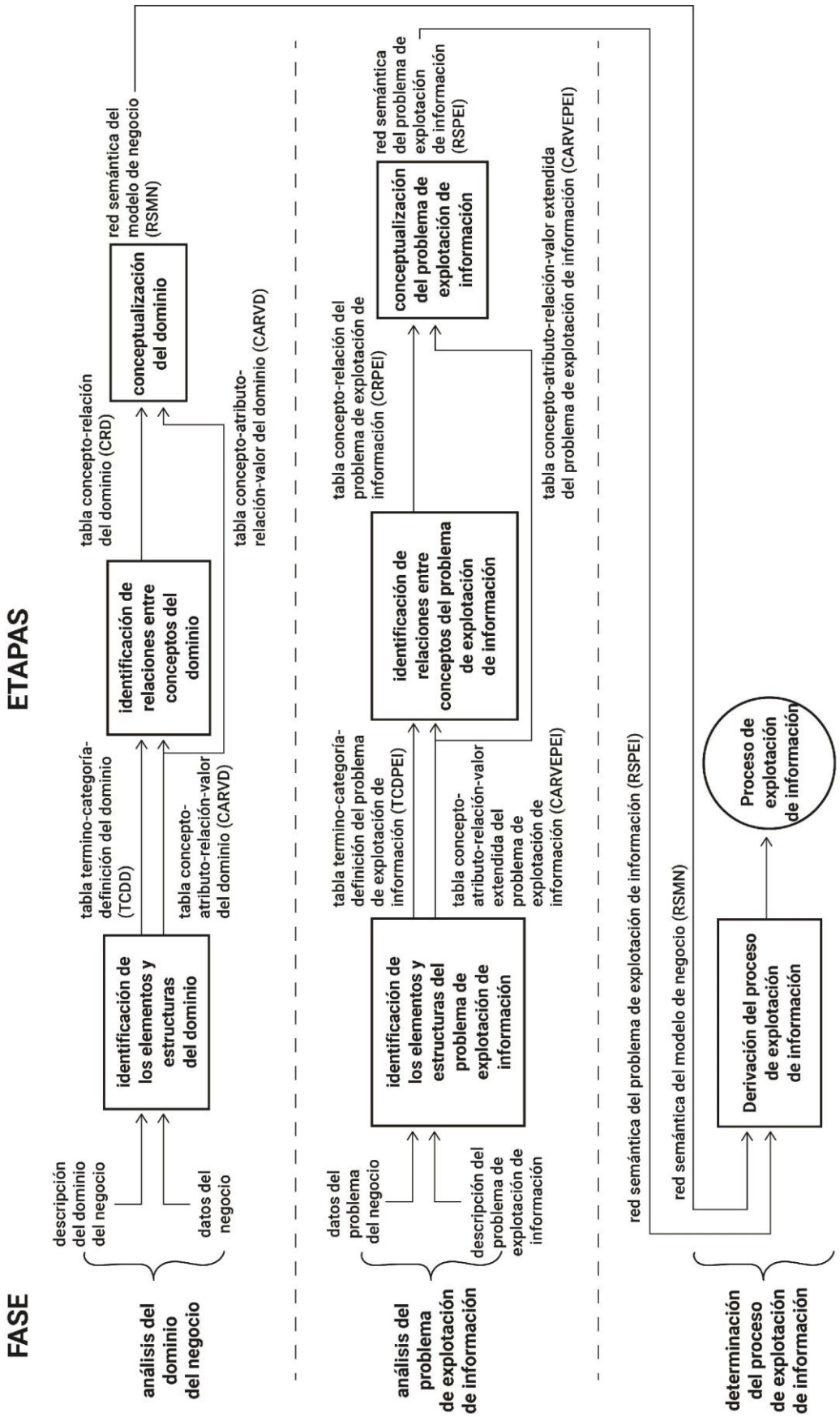


Figura 2.18. Técnica de Modelado – Derivación del Proceso de Explotación de Información.

$$I_d = \left(\frac{1}{2} \cdot \frac{\sum_{i=1}^{n_d} P_{d_i}}{\sum_{i=1}^{n_d} \left(\frac{P_{d_i}}{C_{d_i}} \right)} \right) + \left(\frac{1}{2} \cdot \frac{\sum_{i=1}^{n_d} (P_{d_i} \cdot C_{d_i})}{\sum_{i=1}^{n_d} P_{d_i}} \right) \quad (1)$$

Donde:

I_d: representa el intervalo difuso calculado para la dimensión d (usando como nomenclatura 'P' para plausibilidad, 'A' para adecuación y 'E' para criterio de éxito).

P_{di}: representa el peso de la característica i perteneciente a la dimensión d.

C_{di}: representa el intervalo difuso asignado a la característica i perteneciente a la dimensión d.

n_d: representa la cantidad de características asociada a la dimensión d.

$$V_d = \frac{\sum_{i=1}^{4} I_{d_i}}{4} \quad (2)$$

Donde:

V_d: representa el valor numérico calculado para la dimensión d.

I_{di}: representa el valor de la posición i del intervalo difuso calculado para la dimensión d .

Formula 2.1. Cálculo de valoración de viabilidad por dimensión.

$$EV = \frac{8 \cdot V_P + 8 \cdot V_A + 6 \cdot V_E}{22}$$

Donde:

EV: representa el valor global de la viabilidad del proyecto.

V_P: representa el valor para la dimensión plausibilidad.

V_A: representa el valor para la dimensión adecuación.

V_E: representa el valor para la dimensión criterio de éxito.

Formula 2.2. Cálculo de viabilidad global.

Categoría	ID	Pregunta asociada a la Característica	Peso	Umbral
Datos	P1	¿En qué medida los repositorios disponibles poseen datos actuales?	*8*	poco
	P2	¿Qué tan representativos son los datos de los repositorios disponibles para resolver el problema de negocio?	*9*	poco
	A1	¿En qué medida los repositorios se encuentran disponibles en formato digital?	*4*	poco
	A2	¿Qué cantidad de atributos y registros tienen los datos disponibles?	*7*	poco
	A3	¿Cuánta confianza se posee en la credibilidad de los datos disponibles?	*8*	poco
	E1	¿Cuánto facilita la tecnología de los repositorios disponibles las tareas de manipulación de los datos?	*6*	nada
Problema de Negocio	P3	¿Cuánto se entiende del problema de negocio?	*7*	poco
	A4	¿En qué medida el problema de negocio no puede ser resuelto aplicando técnicas estadísticas tradicionales?	*10*	poco
	A5	¿Qué tan estable es el problema de negocio durante el desarrollo del proyecto?	*9*	poco
Proyecto	E2	¿Cuánto apoyan los interesados (stakeholders) al proyecto?	*8*	nada
	E3	¿En qué medida la planificación del proyecto permite considerar la realización de buenas prácticas ingenieriles con el tiempo adecuado?	*7*	nada
Equipo de Trabajo	P4	¿Qué nivel de conocimientos posee el equipo de trabajo sobre explotación de información?	*6*	poco
	E4	¿Qué nivel de experiencia posee el equipo de trabajo en proyectos similares?	*6*	nada

Tabla 2.3. Características del modelo de viabilidad para proyectos de explotación de información.

2.4.1.6. Modelo de Estimación para Proyectos de Explotación de Información

En [Pytel et al., 2015], se presenta un modelo de estimación de esfuerzo para proyectos de explotación de información a partir de tres aspectos (proyecto, datos y recursos) del proyecto detallados en ocho factores, cada uno de los cuales posee un valor ponderado para determinar el esfuerzo total estimado para el desarrollo del producto del proyecto. Los factores de costo contemplados son: Tipo de objetivo de explotación de información (OBTY), Grado de apoyo de los miembros de la organización (LECO), Cantidad y tipo de los repositorios de datos disponibles (AREP), Cantidad de tuplas disponibles en la tabla principal (QTUM), Cantidad de tuplas disponibles en tablas auxiliares (QTUA), Nivel de conocimiento sobre los datos (KLDS), Nivel de conocimiento y experiencia del equipo de trabajo (KEXT) y Funcionalidad de las herramientas disponibles (TOOL). A partir de sus valores asignados, mediante la fórmula 2.3 [Pytel et al., 2015], se obtiene el esfuerzo estimado (PEM) en mes/hombre.

$$PEM = 0.80\times OBTY+1.10\times LECO-1.20\times AREP-0.30\times QTUM-0.70\times QTUA+1.80\times KLDS-0.90\times KEXT+1.86\times TOOL-3.30$$

Formula 2.3. Cálculo de estimación de esfuerzo.

2.4.1.7. Métricas para Proyectos de Explotación de Información

En [Basso et al., 2014], se presentan un conjunto de métricas para proyectos de explotación de información, enfocadas en las distintas etapas del proceso de extracción de conocimiento. Las métricas propuestas cubren las fases de *entendimiento de los datos*, por ejemplo: número total de fuentes de datos, densidad de atributos nulos, número de atributos a normalizar, entre otras, *preparación de los datos*, por ejemplo: grado de utilidad de la tabla, número de atributos a agregar en la integración, grado de utilidad total de los atributos para el proyecto, etc., *Modelado*, distinguiendo respecto al objetivo del modelo (y los procesos de explotación de información correspondientes), por ejemplo: número de elementos a utilizar para el entrenamiento de un modelo, índice de pertenencia de un elemento al clúster, sensibilidad del modelo de clasificación, entre otros, y otros generales del *proyecto*, las cuales cubren los criterios de evaluación de los resultados obtenidos para lograr un resultado exitoso del proyecto, por ejemplo: nivel de exactitud del modelo, grado de documentación a entregar, etc.

2.4.1.8. Ciclo de vida para Proyectos de Explotación de Información

El ciclo de vida son las fases que abarca un proyecto desde el inicio hasta su finalización. Este define el orden en el cual las actividades se realizan [McConnell, 1997], estableciendo las posibles transiciones para cada una de las fases que lo integran (posibles estados futuros a partir de la fase actual), y el criterio para determinar la transición entre las mismas [Mariscal et al., 2010]. En este contexto, la estrategia de modelo de ciclo de vida seleccionada debe analizarse a partir de las características específicas del proyecto: objetivos, necesidades y expectativas del cliente, cuan bien se comprenden las problemáticas y la posibilidad de resolverlas mediante explotación de información, el soporte y disponibilidad de los clientes/expertos de la organización contratante, la cultura de la organización desarrolladora, la experiencia de sus miembros y el deseo de correr riesgos. En [Project Management Institute, Inc., 2013a] se definen tres tipos de ciclo de vida:

Predictivos (o secuenciales): aquellos en los que el alcance del proyecto, el tiempo y el costo requeridos para entregar ese alcance, se determinan tan pronto como sea prácticamente posible en el ciclo de vida del proyecto. Cada fase se realiza de manera secuencial o superpuesta, y la finalización del proyecto se alcanza una vez alcanzada el último estado posible (última fase).

Iterativos (o incrementales): las fases del proyecto se repiten una o más veces a medida que aumenta la comprensión del producto, en el que cada iteración incrementa el cumplimiento de los objetivos del proyecto (mayor cubrimiento de necesidades vigentes o resolución de nuevas). Las iteraciones pueden realizarse de manera secuencial o superpuesta.

Adaptativos (o ágiles): están diseñados para responder a altos niveles de cambio y a la participación continua de los interesados. Estos son también iterativos, pero difieren de los anteriores en que las iteraciones son muy rápidas, con tiempos y/o costos prefijados. Varias iteraciones pueden ser realizadas de manera simultánea.

Los ciclos de vida definidos en la disciplina de la explotación de información, están asociados a los modelos de procesos o metodologías existentes, siendo en su mayoría del tipo secuencial. A continuación se presentan ciclos de vida pertenecientes a cada una de las categorías previamente mencionadas:

- **Basados en CRISP-DM**: el ciclo de vida definido en CRISP-DM (figura 2.5), es el más común en las propuestas modernas y presenta una visión secuencial de las actividades. Adicionalmente, en [Hofmann, 2003] se define un ciclo de vida genérico para proyectos de explotación de información (figura 2.19, adaptado de [Hofmann, 2003]), basado en CRISP-

DM, el cual incorpora la participación de los recursos humanos y las fuentes de información en el proceso. Estas propuestas, presentan una visión secuencial de las actividades, con un control de mejora continua del proceso durante la finalización de cada iteración. Son ideales para proyectos en los cuales las necesidades y alcances del proyecto son conocidas y pueden ser definidos en etapas tempranas del proyecto, y/o se posee experiencia en proyectos similares.

- **Espiral**: propuesto en [Arboleya, 2013], es una propuesta basada en CRISP-DM, la cual presenta un enfoque dirigido por el riesgo para el análisis y estructuración del proceso, mediante una estrategia de desarrollo del proyecto evolutiva (o incremental). La figura 2.20 [Arboleya, 2013] ilustra su estructura. Dada las características de la propuesta, su uso es recomendado para proyectos complejos, en los cuales se identifiquen riesgos asociados con la viabilidad del proyecto, el presupuesto y/o los plazos, y no se posea experiencia en proyectos similares.

- **ASD-BI**: es un ciclo de vida ágil basado en el método Desarrollo de Software Adaptativo (figura 2.9), propuesto en [Alnoukari, 2010], el cual brinda una visión flexible del proceso, centrada en las personas. Este modelo es ideal para entornos cambiantes en los cuales no se conocen con claridad las necesidades y alcances del proyecto en sus etapas tempranas, y se posee experiencia en proyectos similares.

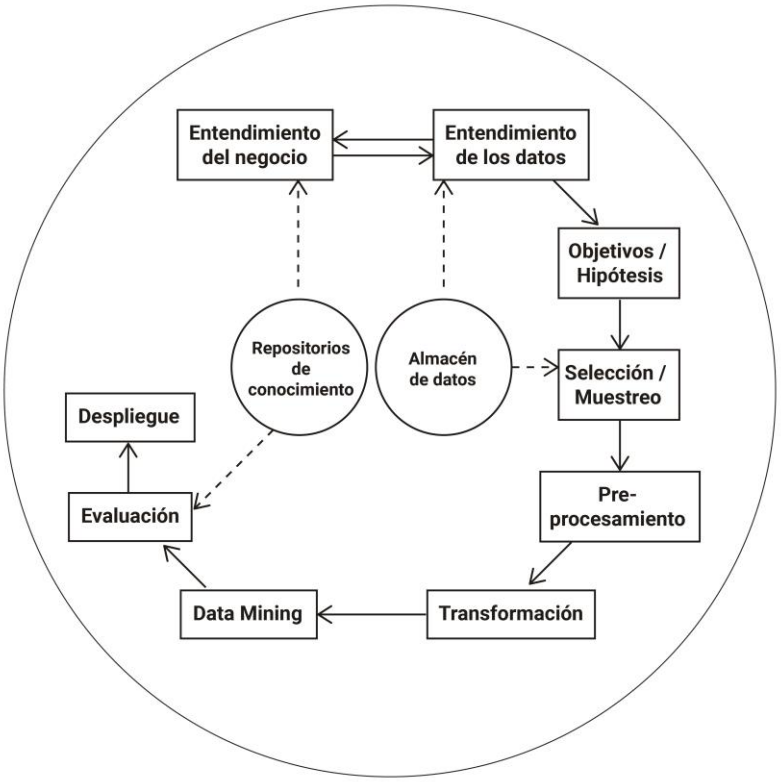


Figura 2.19. Ciclo de vida predictivo: DMLC (RRHH omitidos)

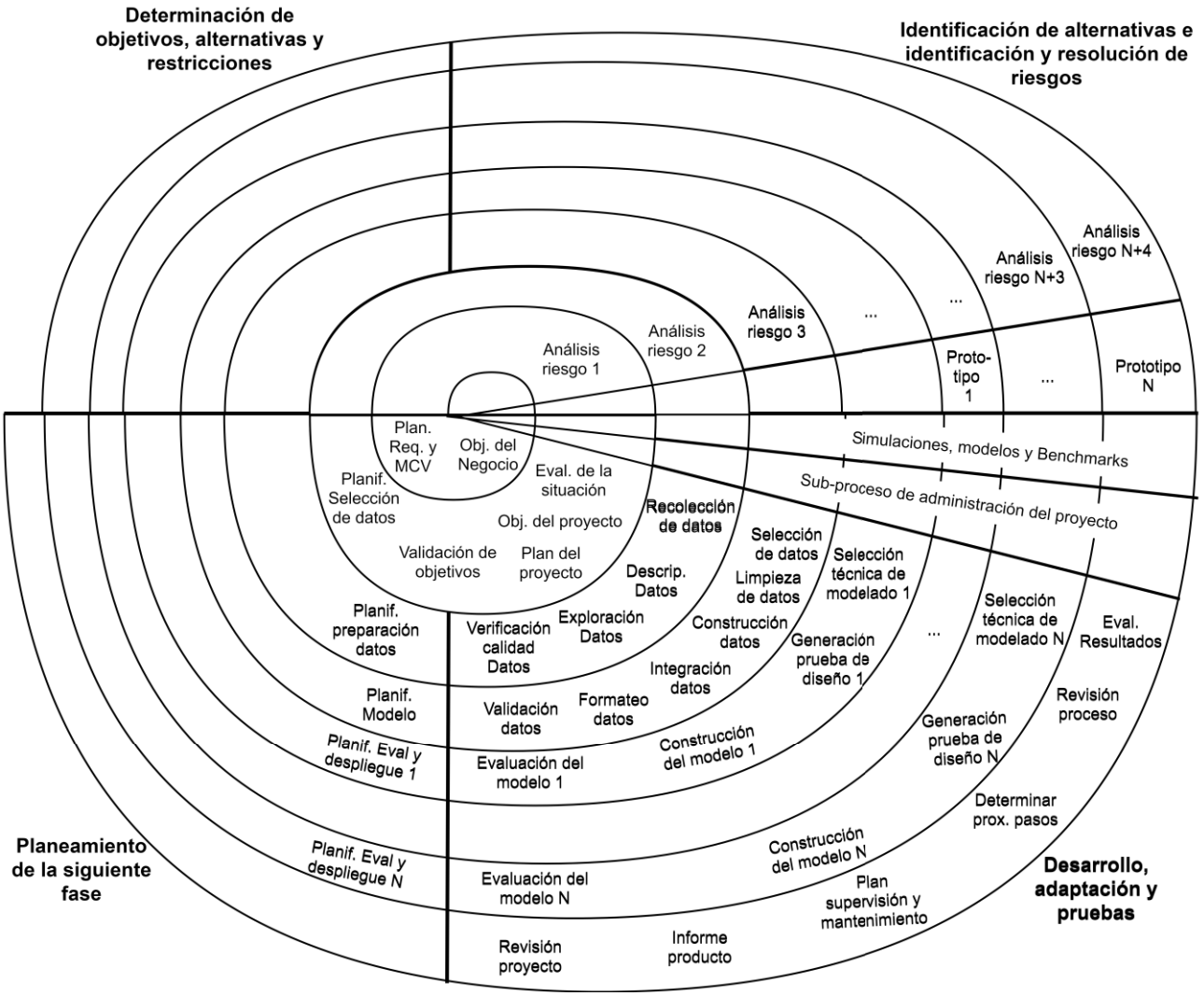


Figura 2.20. Ciclo de vida incremental: Espiral.

2.4.2. Técnicas Aplicables a Ingeniería de Explotación de Información

Existen un grupo de técnicas de uso general para gestión de proyectos, que han sido relevadas en [Verzuh, E., 2015] las cuales pueden ser aplicables a la ingeniería de explotación de información. Estas técnicas son: plan de comunicación (sección 2.4.2.1), plan de acción (sección 2.4.2.2), matriz de responsabilidades (sección 2.4.2.3), reporte de estado (sección 2.4.2.4) y reporte de cierre (sección 2.4.2.5).

2.4.2.1. Plan de Comunicación

Se define al plan de comunicación como la estrategia escrita para obtener la información correcta para la persona adecuada en el tiempo preciso [Verzuh, E., 2015]. En [Project Management Institute, Inc., 2013a] se define que una comunicación es efectiva cuando se provee en el formato correcto, a la audiencia y tiempo adecuados y con el impacto debido, es decir, proveyendo

únicamente la información que se necesita. Las personas involucradas en un proyecto son muchas y con variadas necesidades/intenciones, siendo necesario que el plan de comunicación contemple cada una de los requerimientos particulares, proporcionando la información que cada uno necesita para ser más productivos.

En [Project Management Institute, Inc., 2013a] se señala que el plan de comunicación es significativamente importante en equipos virtuales de trabajo y/o culturalmente diferentes, donde un esfuerzo adicional debe ser realizado para mantener expectativas claras, facilitar la comunicación y desarrollar protocolos para la resolución de conflictos en los cuales se contemple a las personas en el proceso de toma de decisiones, entendiendo sus diferencias culturales y compartiendo el crédito en éxitos. Además, se señala dicha técnica como una herramienta necesaria para reducir la cantidad de conflictos, los cuales son inevitables en un entorno de proyecto.

En dicho formalismo debe darse respuesta a las preguntas: ¿Quién necesita información?, ¿Qué información se necesita?, ¿Qué información tiene autorizada?, ¿Con qué frecuencia? Y ¿De qué forma/medio la misma será informada?

2.4.2.2. Plan de Acción

En [Clark et al., 1922] se señala que la gestión de proyectos se centra casi enteramente en el futuro, y en tomar decisiones a partir de lo ocurrido en el pasado, para alcanzar las condiciones deseadas. En este contexto, es relevante conocer en todo momento cuando los eventos de un proyecto dieron lugar y el porcentaje del trabajo que está realizado. Siendo el plan de acción la técnica que proporciona dicho conocimiento. Este puede ser definido como el programa que establece la secuencia de actividades o acciones que deben realizarse para desarrollar de manera exitosa una estrategia que permita alcanzar las metas definidas.

El diagrama Gantt, es la herramienta gráfica más notable para este tipo de actividades [Clark et al., 1922], dado que brinda una visión sencilla de lo hecho con respecto a lo estipulado.

2.4.2.3. Matriz de Responsabilidades

En [Project Management Institute, Inc., 2013a] se define a la matriz de responsabilidad como una grilla que muestra los recursos asignados a cada agrupación de tareas, la cual se utiliza para ilustrar las conexiones entre las actividades y las partes interesadas del proyecto. Esta técnica, ilustra todas las actividades asociadas con cada persona o grupo (por ejemplo: por roles) involucrados en el proyecto, describiendo su rol y nivel de participación.

En proyectos grandes, la matriz de responsabilidades puede ser desarrollada en varios niveles, por ejemplo, un nivel general que indique los roles de cada miembro en los grupos de tareas y un nivel inferior en el cual se establecen los roles, responsabilidades y niveles de autoridad para cada actividad específica.

2.4.2.4. Reporte de Estado

El reporte de estado presenta un resumen de la situación del proyecto en un hito o periodo de tiempo determinado. Dicha técnica proporciona un resumen del desempeño del proyecto para comunicar de manera eficiente el progreso del mismo [Project Management Institute, Inc., 2013a], en el cual se resaltan distintas características como los logros principales en el periodo, la comparación de progreso actual con respecto a lo previsto, el porcentaje del proyecto completado, así como otros aspectos relevantes como la identificación de riesgos, situaciones críticas, solicitudes de cambios significativos para el proyecto y sus alcances, entre otros. En [Verzuh, E., 2015] se señala la necesidad de mantener el reporte de estado lo más breve posible, simplificando la lectura del mismo, facilitando la identificación de la necesidad de tomar de acciones con respecto al desarrollo del proyecto.

2.4.2.5. Reporte de Cierre

El reporte de cierre resume los resultados del proyecto, comparando las metas originales con las reales [Verzuh, E., 2015]. En este, el equipo de trabajo evalúa el desarrollo del proyecto (y su progreso en los distintos hitos) y las mediciones tomadas, con respecto a lo proyectado, con el objetivo de identificar y documentar las razones y los resultados de las acciones definidas durante la evolución del proyecto. En orden a alcanzar satisfactoriamente las metas de esta técnica, debe participar en el proceso todos los involucrados en el proyecto posibles [Project Management Institute, Inc., 2013a], con el propósito de obtener la visión de los distintos roles participes en el proceso, favoreciendo la identificación de los aspectos positivos y negativos, así como las dificultades enfrentadas. El objetivo de la información recabada es que sirva como base histórica para futuros proyectos.

2.5. PROCESO, METODOLOGÍA e INGENIERÍA DE PROYECTOS

En este apartado se describen las concepciones adoptadas para este trabajo: se diferencian los conceptos proceso y metodología (abordando también el concepto de ciclo de vida), y se detalla la estructura general contemplada en la ingeniería de proyectos.

En [Rodríguez, 2015], a partir del estudio de distintos autores en variadas disciplinas y el consenso identificado en ellas, se señala: „El Diccionario de la Lengua Española [RAE, 2015] define el término "proceso" como "…conjunto de las fases sucesivas de un fenómeno natural o de una operación artificial…", y al término "metodología" como "…conjunto de métodos que se siguen en una investigación científica o en una exposición doctrinal…."; por otra parte define el término "método" como "…procedimiento que se sigue en las ciencias para hallar la verdad y enseñarla…", y "procedimiento" como "…método de ejecutar algunas cosas …". Sin embargo en las Ciencias Informáticas es frecuente encontrar referenciados ambos términos proceso y metodología, como si fueran equivalentes. En [Marbán et al., 2007] se propone diferenciar ambos términos de la siguiente manera: un proceso (o modelo de proceso) se define como el conjunto de tareas (agrupadas en fases) a realizar para desarrollar un elemento en particular, así como los elementos que se producen en cada tarea (salidas) y los elementos que son necesarios para hacer una tarea (entradas). El objetivo de un proceso es hacer que la construcción de cada elemento sea repetible, manejable y medible. Por otra parte, en [Marbán et al., 2007] se define metodología como la instancia de un proceso que además de definir las tareas, las entradas y las salidas, especifica cómo hacer las tareas. Las tareas se realizan con técnicas/procedimientos que estipulan cómo se deben hacer. En suma se conviene que: una metodología es un proceso en el que se ha identificado (instanciado) con que técnica se desarrolla cada tarea de cada fase del proceso. Esto se resume en la convención: "metodología = proceso + técnicas"."

Adicionalmente, se considera relevante definir el concepto de ciclo de vida. El ciclo de vida es el elemento que define el flujo u orden en el cual las actividades se realizarán a lo largo del proyecto (incluyendo posibles iteraciones) [McConnell, 1997; Sommerville, 2011; Mariscal et al., 2010; ISO, 2012; Project Management Institute, Inc., 2013a], mientras que en [Mariscal et al., 2010] se señala como alcance adicional la definición del criterio para pasar a la siguiente fase y los posibles futuros estados para cada una de las fases. Este tiene una fuerte relación con las necesidades del proyecto, la posibilidad de adaptarse a las características del mismo y su dependencia con respecto al éxito del

proyecto, demoras y ejecución de esfuerzo innecesario [Sommerville, 2011; Project Management Institute, Inc., 2013a].

A partir de los conceptos previamente descritos, se considera relevante señalar cómo está conformado un proceso y cuáles son las ventajas que provee.

En la guía de los fundamentos para la dirección de proyectos (guía PMBOK) [Project Management Institute, Inc., 2013a] se destaca que un proceso está conformado por dos subprocesos generales: uno orientado al producto y otro orientado a la gestión del proyecto. El primero abarca aquellas actividades específicas para la creación del producto. Se encuentra definido por el ciclo de vida del proyecto y su estructura es específica de la disciplina en cuestión, mientras que el segundo tiene como objetivo garantizar el desarrollo efectivo del proyecto (y por consiguiente el ciclo de vida) de acuerdo a lo previsto. Su estructura es más general para distintos proyectos, sin embargo, este debe de ajustarse a las necesidades y características específicas de los mismos.

En [Project Management Institute, Inc., 2013a], se definen como estructura del subproceso de gestión de proyectos, las siguientes fases:

- **Iniciación**: consiste en aquellas actividades necesarias para definir un nuevo proyecto o nueva iteración. En esta fase, se define el alcance inicial del proyecto y se determinan los recursos comprometidos al mismo. Adicionalmente, se identifican aquellos interesados (internos y externos) a interactuar cuya opinión tenga influencia en la definición de los resultados del proyecto, y se aplica un proceso para evaluar las posibles alternativas y determinar la viabilidad de la solución. En proyectos grandes, las tareas de iniciación deben ser realizadas de manera transversal al desarrollo del proyecto, tomando decisiones acerca de la estrategia utilizada y los recursos involucrados.
- **Planificación**: abarca aquellas actividades vinculadas con la definición del esfuerzo total del proyecto, la definición del curso de acciones requerido para alcanzar los objetivos y su ajuste (en caso de ser necesario), la planificación de las actividades y documentos involucrados en el proceso. En esta fase se evalúan los aspectos del proyecto (tiempo, costo, calidad, recursos, alcances, etc.). La complejidad natural del proyecto y los cambios significativos durante el ciclo de vida, pueden requerir la reiteración de la fase, profundizando en el análisis del proyecto. El principal beneficio de esta fase es la estructuración del flujo de trabajo a realizar para completar exitosamente el proyecto, facilitando la participación del equipo de trabajo, el control de los distintos aspectos del proyecto y mejorando el compromiso de los interesados. Los cambios aprobados durante el

desarrollo del proyecto, pueden tener un impacto significativo en las actividades de la fase, requiriendo la actualización de las decisiones realizadas.

- **Ejecución**: consiste en la realización de las actividades definidas para satisfacer el plan de trabajo definido. Este proceso involucra la coordinación del equipo de trabajo, recursos, expectativas de los interesados, la evolución del proyecto y el registro de los aspectos del proyecto, incluyendo la definición de estándares que permitan identificar los cambios y la necesidad de ajustar el plan y la línea base. A partir de las características del proyecto registradas en esta etapa, puede identificarse la necesidad de ajustar los planes y otros documentos del proyecto.
- **Monitoreo y Control**: cubre aquellas actividades vinculadas con el seguimiento, revisión y análisis del progreso y rendimiento del proyecto de acuerdo al plan. En esta actividad pueden identificarse desvíos o riesgos que requieren de la implementación de acciones correctivas o preventivas. Adicionalmente, cualquier petición de cambio es evaluada de manera formal, analizando el impacto de la misma sobre lo planificado y desarrollado.
- **Cierre**: abarca aquellas actividades requeridas para la finalización formal del proyecto o de la iteración del ciclo de vida. Validar que se han satisfecho las obligaciones contractuales, estableciendo formalmente la terminación del proyecto, realizar revisiones del proceso/iteración, documentar las lecciones aprendidas y finalizar formalmente las obligaciones mediante la aceptación de las mismas, son tareas a realizar en dicha fase. Adicionalmente, el cierre prematuro del proyecto (por ejemplo: a causa de su cancelación, problemas críticos, etc.) es documentado en esta etapa.

Las principales ventajas que provee el uso de un proceso para el desarrollo de proyectos, son:

- Define un marco común de trabajo, estableciendo las tareas y alcances de cada una de las actividades, permitiendo a los integrantes del proyecto comprender lo que se espera en cada etapa, resultando en un desarrollo más rápido, barato, confiable [Kurgan y Musilek, 2006], repetible, manejable y medible [Marbán et al., 2007].
- Proporciona estabilidad, control y organización a una tarea que puede, si no se controla, volverse caótica [Pressman, 2005], además de establecer un marco de comunicación eficiente [Verzuh, 2015],
- Permite la comparación entre proyectos, a partir de la estructura en común que poseen, pudiendo utilizarse de referencia para la evaluación y estimación de distintos aspectos, así como extraer conocimiento del proceso [Verzuh, 2015],
- Facilita la trazabilidad de las actividades y la repetición de las mismas [Verzuh, 2015], y

- Facilita la comprensión de los proyectos y su integración, mediante la estandarización de los resultados, reduciendo los riesgos asociados con el desarrollo de las actividades [Verzuh, 2015].

3. DELIMITACIÓN DEL PROBLEMA

En este capítulo se introduce el contexto disciplinar del problema abordado en la tesis (sección 3.1), se formula una discusión de los abordajes metodológicos para Proyectos de Ingeniería de Explotación de Información (sección 3.2), y se identifica el problema abierto considerado en la tesis, formulando las preguntas de investigación asociadas (sección 3.3).

3.1. CONTEXTO DISCIPLINAR DEL PROBLEMA

Como se definió previamente, la ingeniería de Explotación de Información (en ocasiones también referida como Minería de Datos o Extracción de Conocimiento) es la sub-disciplina de los Sistemas de Información que aporta a la Inteligencia de Negocio [Langseth y Vivatrat, 2003] las herramientas para la transformación de información en conocimiento [Srivastava et al., 2000]. Esto es, la búsqueda de patrones relevantes en masas de información [Abraham, 2003; Cooley, 2003], las cuales pueden estar almacenadas en distintos medios y formatos.

En los últimos 15 años, numerosos autores han señalado dificultades y carencias identificadas a partir del desarrollo de distintos proyectos, afirmándose en [Gondar, 2005; Marbán et al., 2009], que la tasa de proyectos que fracasan es aproximadamente del 60% (señalándose como la principal metodología utilizada CRISP-DM, la cual se mantiene como el estándar de facto [Marbán et al., 2007; Kdnuggets, 2014]). En adición, en [Berry y Linoff, 2004] se señala que a partir del incremento en la complejidad de los proyectos, la necesidad de un enfoque riguroso se hace notable.

Sin embargo, en [Gallardo, 2009] se indica que: “históricamente, el principal foco de las investigaciones en Data Mining, ha estado centrado en el desarrollo de algoritmos y herramientas, sin considerar que para garantizar el éxito de un proyecto de Data Mining, complejo en su esencia, la atención no sólo debiera centrarse en el desarrollo de modelos y algoritmos, sino también en un enfoque metodológico que permita el desarrollo entre otros, de una definición y especificación sistemática y organizada de las necesidades estratégicas que deberá satisfacer el proyecto y las restricciones de confiabilidad inherente de los resultados.”.

Dicha situación nos conlleva a realizarnos las siguientes preguntas, las cuales intentaremos dar respuesta en los siguientes párrafos:

a) ¿Por qué es necesario contar con un proceso común que guie el desarrollo de proyectos de explotación de información?, Y

b) ¿Cuáles son las insuficiencias que presentan las propuestas existentes?

a) ¿Por qué es necesario contar con un proceso común que guie el desarrollo de proyectos de explotación de información?

Para entender las ventajas que un proceso brinda, es relevante comprender dicho concepto y los alcances del mismo. Sin adentrarse en la discusión técnica acerca de la diferencia entre los conceptos modelo de proceso y metodología, se infiere a partir del estudio de la literatura, y de las concepciones adoptadas en esta tesis (sección 2.5, pág. 48), que ambos conceptos expresan con detalle la estructura que debe tener el procedimiento requerido para ser llevado a cabo en función de las necesidades del cliente. Además, debe brindar una guía sobre las acciones a realizar las cuales deriven en un resultado que satisfaga dichas necesidades; es decir, propiciar la ejecución de las tareas necesarias para lograr el objetivo requerido, junto con los elementos de entrada que deben utilizarse en cada tarea y las salidas esperadas [Pressman, 2005; Mariscal et al., 2010].

En [Marbán et al., 2007] se evalúa el estado actual de la disciplina, comparándola con la ingeniería de software. En él se señalan las dificultades que enfrentaban las empresas de desarrollo de software en 1968, lo cual desembocó en la llamada crisis del software; y donde se menciona como principal causa, la falta de procesos o metodologías formales que den soporte al desarrollo y la gestión del proyecto. Los autores señalan que la mejora en la disciplina, fue resultado de la adopción de conceptos pertenecientes a otros campos de la ingeniería y el surgimiento de nuevas metodologías, resolviendo problemas como la gestión de proyectos y el aseguramiento de la calidad de los resultados, además de mejorar la productividad y el mantenimiento del software. De la comparación realizada entre las disciplinas, se señala que la minería de datos (o la explotación de información) se encuentra en un estado similar a los que dieron inicio a la crisis del software, donde el esfuerzo se enfoca en los algoritmos y herramientas, y no un proceso robusto.

Siendo relevante destacar que los proyectos de ingeniería de explotación de información poseen características muy distintas a los proyectos de desarrollo de software tradicional [Marbán et al., 2007; Vanrell et al., 2010], sobre todo en la parte operativa del proyecto. La diferencia se presenta en los procesos de desarrollo y mantenimiento en los cuales el ciclo de fases de un proyecto de software tradicional: inicio, requisitos, análisis, diseño, construcción, integración y pruebas no resultan naturales en un proyecto de explotación de información [Vanrell et al., 2012].

Mientras que en [Berry y Linoff, 2004] se define la necesidad de aplicar un enfoque para el desarrollo de proyectos de explotación de información, para evitar obtener dos resultados indeseados: aprender conocimientos que no son verdaderos y aprender conocimientos poco útil. Dado que el objetivo para el cual se realiza un proyecto de explotación de información, es la

generación de conocimiento para dar soporte a la toma de decisiones, el riesgo que conlleva basar dichas decisiones sobre conocimientos no válidos, el primero de los resultados no deseados, es mayor que el segundo tipo de resultados indeseados y puede ocasionar grandes pérdidas a la organización.

Finalmente, se destaca que en los últimos 20 años, se ha llevado a cabo un esfuerzo sostenido en el tiempo en la generación de propuestas que guíen el desarrollo de un proyecto de explotación de información (las cuales se describen en la sección 2.2, pág. 12), y a partir de los resultados de las últimas dos encuestas realizadas en los años 2007 y 2014 (figura 2.2, página 14), se observa una tendencia favorable a la adopción de un proceso para el desarrollo de proyectos de explotación de información, pero se observa una disminución en el uso de las metodologías tradicionales incrementándose el uso de una versión propia del equipo de trabajo, indicando los autores de la encuesta como motivo de dicho comportamiento las carencias en las propuestas existentes y la falta de actualización de la metodología más utilizada (CRISP-DM) a las necesidades actuales de los proyectos. Este resultado, nos transmite a la segunda pregunta previamente planteada.

b) ¿Cuáles son las insuficiencias que presentan las propuestas existentes?

Como se mencionó previamente en esta sección, en los últimos 15 años distintos autores han señalado una serie de limitaciones y dificultades en las distintas propuestas, las cuales se abordan a continuación desde una perspectiva general (realizándose por propuesta en detalle en la sección 3.2.3).

Carecen de una visión completa de las actividades: en los últimos años varios autores [Clifton y Thuraisingham, 2001; Charest y Delisle, 2006; Gallardo, 2009; Sharma y Osei-Bryson, 2009; El Sheikh y Alnoukari, 2012; Kdnuggets, 2014] han señalado la carencia de métodos o técnicas que brinden al usuario una guía detallada de las acciones a realizar en cada una de las tareas que los procesos existentes contemplan, dificultando la comprensión de cómo se produce el resultado esperado. En [Gallardo, 2009], se resalta respecto de las propuestas existentes: “Todos estos procesos o guías de desarrollo sin embargo, adolecen de métodos o técnicas que permitan educir adecuadamente los requisitos del proyecto. Más concretamente, aún no existe un proceso maduro que pueda calificarse como una metodología sólida, pues si bien por ejemplo, CRISP-DM, establece un conjunto de tareas y actividades que deben ser ejecutadas en el proyecto, no establece con qué técnicas o modelos se debe implementar cada actividad.”. En adición, en [Sharma y Osei-Bryson, 2009] se menciona que no se indican los objetivos para las tareas y actividades, y no se describe o se presentan algunos lineamientos de los pasos requeridos, dificultando su implementación,

pudiendo esto explicar por qué las tareas indicadas no siempre se implementan formalmente. Además, a menudo hay poca orientación sobre cómo implementar una tarea en particular, dando como ejemplo la fase de comprensión del problema o dominio del negocio.

No se contempla al recurso humano: los procesos de explotación de información existentes no consideran al recurso humano involucrado [El Sheikh y Alnoukari, 2012], ignorando su participación en las distintas etapas del proyecto, y sus necesidades vinculadas con la organización y desarrollo de las mismas.

Vista fragmentada del proceso: el incorrecto desarrollo del proceso y los desvíos de las metas establecidas en los proyectos, surgen a causa de decisiones deficientes o inexistentes originadas por el desconocimiento de las dependencias entre las numerosas actividades que conforman al proceso, las cuales deben ser claramente identificadas en orden a implementar de forma eficiente las tareas [Sharma, 2008].

Carecen de una visión completa del proceso: las propuestas actuales se centran únicamente el proceso orientado al producto (o desarrollo). Sin embargo, numerosos autores [Kurgan y Musilek, 2006; Mariscal et al., 2010; do Nascimento y de Oliveira, 2012] han señalado que los procesos existentes no contemplan las actividades asociadas con la gestión del proyecto. La necesidad de planificar, administrar y controlar un proyecto interdisciplinario [Fayyad et al., 1996], complejo [Kurgan y Musilek, 2006; Gallardo, 2009] y dinámico [Brachman & Anand, 1996], el cual presenta una creciente complejidad [Mariscal et al., 2010] a partir del incremento de las fuentes de información accesibles para un proyecto (en cantidad y tamaño) y la amplia posibilidad de satisfacción a necesidades del cliente que las mismas brindan.

No se adaptan a las necesidades del proyecto: las propuestas existentes poseen en su estructura un modelo de ciclo de vida implícito (en su mayoría del tipo cascada o secuencial), lo cual limita al equipo de trabajo con respecto al soporte que el proceso brinda en relación a las necesidades específicas del proyecto [Marbán, et al., 2007; Gallardo, 2009; Mariscal et al., 2010].

Finalmente, y como consecuencia de los aspectos anteriormente cubiertos, se señala [Kurgan y Musilek, 2006; Mariscal et al., 2010; Kdnuggets, 2014] la necesidad de definir un modelo de proceso, el cual brinde una visión completa y detallada los pasos asociados al desarrollo de un proyecto de explotación de información. Pero además, debe incluir una visión ingenieril [Marbán et al., 2009] de la disciplina, esto es, no solo detallar la estructura y elementos del proceso, sino también incluir las técnicas y/o métodos que él mismo comprende [Pressman, 2005].

3.2. DISCUSIÓN DE LOS ABORDAJES METODOLÓGICOS PARA PROYECTOS DE EXPLOTACIÓN DE INFORMACIÓN

En la sección previa, se han resumido las deficiencias frecuentemente señaladas en los últimos años a los modelos de proceso o metodologías existentes. A partir de las mismas, se destaca la necesidad de definir un modelo de proceso que satisfaga dichas carencias. En esta sección, se presenta una discusión de las propuestas más relevantes (sección 3.2.1), detallando los motivos por los cuales dichas propuestas se destacan, y las dimensiones de análisis a partir de las cuales cada una de ellas es evaluada en la sección 3.2.2. Finalmente, se presentan los resultados obtenidos de la evaluación de los abordajes seleccionados.

3.2.1. Abordajes Metodológicos Considerados

En esta sección se listan las propuestas seleccionadas a partir de su importancia en la disciplina, justificando los motivos de su elección. Una visión detallada de cada uno de los modelos de proceso o metodologías existentes puede obtenerse en la sección 2.2, pág. 12.

Como previamente fue mencionado, en los últimos 20 años se han desarrollado distintas metodologías o modelos de proceso con el objetivo de normalizar el vocabulario y los pasos requeridos para el desarrollo de proyectos [Fayyad et al., 1996; Clifton y Thuraisingham, 2001]. Varias revisiones sobre el estado actual de la disciplina se han realizado desde esta perspectiva [Kurgan y Musilek, 2006; Mariscal et al., 2010; Alnoukari y El Sheikh, 2012], señalándose la evolución de las propuestas relevadas (figura 2.1, página 13). En [Mariscal et al., 2010] se señala a KDD y a CRISP-DM como las propuestas iniciales, a partir de las cuales la mayoría de las propuestas están basadas. A partir de lo previamente expuesto, y en concordancia con lo descrito en la sección 2.2 (pág. 12), se determina la inclusión de CRISP-DM, como estándar *de facto* de la industria, KDD como propuesta que se mantiene en la industria y por su relevancia en la disciplina [Mariscal et al., 2010] como base para otras propuestas, y SEMMA, que si bien se evidencia una reducción considerable en su adopción, aún se mantiene como el segundo abordaje más utilizado [Kdnuggets, 2014].

De forma complementaria, a partir del análisis de las propuestas restantes y las críticas resumidas previamente, se considera relevante incluir las propuestas MPIMD, IKDDM y FMDS (secciones 2.2.5, 2.2.6 y 2.2.8, respectivamente), que si bien se encuentran basadas en CRISP-DM y KDD, introducen conceptos significativos no considerados por sus predecesoras. La primera, introduce la necesidad de una visión ingenieril en la cual se contemple las actividades de gestión en el proceso.

IKDDM se destaca como la única metodología que profundiza en la descripción e identificación de técnicas o métodos a utilizar en las fases del proyecto, mientras que la FMDS introduce aspectos vinculados con el procesamiento de grandes volúmenes de datos.

En los últimos años (inspirado en las adopciones realizadas en la disciplina de ingeniería de software) se han introducido propuestas basadas en los métodos ágiles. El enfoque ágil define una aproximación distinta al tradicional proceso de gestión de proyectos (junto con una filosofía [Highsmith, 2001]), es decir, una opción alternativa al histórico enfoque de desarrollo secuencial [Sommerville, 2011; Rubin, 2012; Project Management Institute, Inc., 2013a] enfocado a dar respuesta a entornos altamente cambiantes y la participación continua de los interesados.

Bajo este enfoque de desarrollo de proyectos, se identifican cinco propuestas: ASD-DM [Alnoukari, et al., 2008], Agile KDD [do Nascimento y de Oliveira, 2012], ASD-BI [Alnoukari, 2012], TDSP [Microsoft, 2016] y Agile Data Science [Jurney, 2017]. En esta categoría se seleccionan ASD-BI, siendo la propuesta sucesora de ASD-DM sobre la cual se dispone de mayor nivel de información y cubrimiento de las actividades de explotación de información y TDSP como primer propuesta ágil en introducir los formalismos de los artefactos de salida para cada etapa del proceso.

Finalmente, entre las propuestas que no se encuentran basadas en KDD o CRISP-DM (figura 2.1, pág. 13), se selecciona a la propuesta Catalyst (sección 2.2.4, pág. 21), la cual cubre aspectos superadores con respecto al *estándar de facto* (CRISP-DM) [Moine, 2013].

3.2.2. Selección de las Dimensiones de Análisis

Las dimensiones de análisis se definen como los elementos de evaluación que permiten entender los aspectos que cubre cada una de las propuestas seleccionadas, con el objetivo de identificar las carencias y deficiencias señaladas en los últimos años en la disciplina, y así poder comprender los alcances y las dificultades presentes en las mismas.

Para ello, es necesario establecer los aspectos que una propuesta de modelo de proceso (o metodología) debe comprender. Si bien se ha señalado que en la disciplina informática los términos modelo de proceso y metodología se han utilizado como equivalentes [Rodríguez, 2015], diversos autores [Marbán et al., 2007; Rodríguez, 2015] han adoptado como diferenciador del significado de ambos conceptos la siguiente convención: un modelo de proceso define un conjunto de actividades para realizar un trabajo, el cual incluye los elementos de entrada y salida para llevar a cabo cada uno de ellas, mientras que una metodología incorpora como dicha actividad debe ser realizada, es decir, Metodología = Proceso + Técnicas [Rodríguez, 2015] (abordado con mayor detalle en la

sección 2.5, pág. 48). En [Mariscal et al., 2010; Kerzner, 2013] se señala la necesidad de contar con una descripción detallada de los pasos a realizar en cada actividad del proceso, mientras que varios autores señalan como carencia la necesidad de incluir técnicas o procedimientos para las actividades contempladas en el proceso [Clifton y Thuraisingham, 2001; Charest y Delisle, 2006; Sharma y Osei-Bryson, 2009; Gallardo, 2009; El Sheikh y Alnoukari, 2012; Kdnuggets, 2014]. Complementariamente a la definición previa, en la guía de los fundamentos para la dirección de proyectos (guía PMBOK) [Project Management Institute, Inc., 2013a] se agrega que dicho proceso contempla dos categorías generales: un proceso orientado al producto y un proceso orientado a la gestión (sección 2.5, pág. 48). Mientras que en [Mariscal et al., 2010; Kerzner, 2013; Project Management Institute, Inc., 2013a] se detallan como aspectos deseados en un proceso: la estandarización de productos, trazabilidad, planificación, repetibilidad, adaptabilidad a las necesidades del proyecto, entre otros. Características señaladas en [Kurgan y Musilek, 2006; Mariscal et al., 2010; do Nascimento y de Oliveira, 2012] como deficiencias en las propuestas existentes.

A partir del análisis de las definiciones provistas en el párrafo previo, se destacan como conceptos necesarios: la identificación de la estructura del proceso (o actividades requeridas) incorporando los aspectos vinculados con el producto y con la gestión del proyecto, las dependencias entre las mismas (elementos de entrada y salida), las técnicas o procedimientos a realizar (los pasos asociados a la actividad) y productos estandarizados (modelos definidos para las salidas de las actividades).

Por último, es necesario definir la estructura para ambas categorías generales de actividades. En [Mariscal et al., 2010] se realiza un análisis de la estructura de las distintas propuestas existentes. Se identifica la estructura propuesta en CRISP-DM para las actividades orientadas al producto, como la forma de agrupar las actividades en etapas de acuerdo a sus objetivos, más aceptada por las propuestas existentes, siendo además la propuesta más utilizada en la industria. A continuación se listan sus agrupaciones de actividades (o fases), las cuales se describen en detalle en la sección 2.2.3 (pág. 19):

- **Comprensión del Negocio**: se centra en el entendimiento de los objetivos, las necesidades del proyecto (definiendo los problemas que el mismo abarca) y sus criterios de éxito.
- **Comprensión de los Datos**: se realiza la recolección inicial de los datos y aquellas actividades asociadas con el entendimiento de los datos y la identificación de problemas en la calidad de los mismos.

- **Preparación de los Datos**: cubre todas las actividades necesarias para construir el set de datos final (sobre los cuales se aplican los modelos). Abarca la preparación, selección, transformación y limpieza de los datos.
- **Modelado**: se identifican, seleccionan y aplican los modelos, de acuerdo a las estrategias de evaluación y optimización de parámetros elegidos. Adicionalmente, se incorpora la necesidad de representar el conocimiento del dominio de forma que se pueda sistematizar el proceso de selección del modelo [Marbán et al., 2007; García-Martínez et al., 2013; Martins et al., 2014].
- **Evaluación**: se analizan los resultados del modelo con respecto a los objetivos del negocio, determinando si los mismos satisfacen las necesidades planteadas.
- **Despliegue**: se organizan y presentan los resultados obtenidos de forma que el cliente pueda utilizarlos.

La estructura para el proceso de gestión, se define a partir de la guía estándar para la gestión de proyectos: guía PMBOK [Project Management Institute, Inc., 2013a], en la cual se detallan los elementos generales que agrupan las actividades a realizar según sus objetivos para el proceso orientado a la gestión del proyecto (las cuales se detallan en la sección 2.5, pág. 48):

- **Iniciación**: consiste en aquellas actividades necesarias para definir un nuevo proyecto o nueva iteración, en la cual se determinan los recursos involucrados, la aprobación de realización del proyecto, la interacción con los interesados, estándares específicos de la organización y ciclo de vida del proyecto.
- **Planificación**: se define el esfuerzo total del proyecto y se establece el curso de las acciones y recursos requeridos para alcanzar los objetivos definidos.
- **Ejecución**: supone la coordinación de las tareas del proyecto de acuerdo con el plan realizado, incluyendo la definición de estándares que permitan identificar los cambios, pudiendo determinar la necesidad de ajustar el plan y la línea base.
- **Monitoreo y Control**: se realiza el seguimiento y control del plan del proyecto, evaluando las peticiones de cambio y los procedimientos de control de riesgo.
- **Cierre**: abarca aquellas tareas asociadas con la finalización formal del proyecto o iteración. Incluye validar que se han satisfecho las obligaciones contractuales estableciendo formalmente la terminación del proyecto, y realizar revisiones del proceso.

A modo de resumen, se listan los conceptos de interés a identificar como dimensiones de análisis de las propuestas:

- ¿En qué medida se identifican los elementos trascendentales de la estructura de desarrollo? Es decir, las actividades orientadas al producto: Entendimiento del negocio, Entendimiento de los Datos, Preparación de los Datos, Modelado, Evaluación y Despliegue.
- ¿En qué medida se identifican los elementos trascendentales de la estructura de gestión? Es decir, las actividades orientadas al proyecto: Inicial, Planificación, Ejecución, Monitoreo y Control, y Cierre.
- ¿En qué medida se identifican los elementos de entrada y salida de las actividades? Es decir las dependencias entre los elementos del proceso.
- ¿En qué medida se identifican las técnicas o procedimientos a implementar en cada actividad?
- ¿En qué medida se definen productos estandarizados? Es decir, formalismos que normalicen las salidas de las actividades.

3.2.3. Análisis de los Abordajes Metodológicos

En esta sección se realiza la evaluación de las dimensiones de análisis en las propuestas seleccionadas, la cual se resume en la tabla 3.1. Para indicar la gradualidad de los criterios, se utilizan tres posibles valores:

- *No contemplado* (□), la propuesta no considera ninguno de los objetivos o actividades correspondientes,
- *Parcialmente contemplado* (▒), cubre algunas de las actividades, y
- *Contemplado* (■), considera todas las actividades.

El análisis realizado en la tabla 3.1, deja de manifiesto la visión parcial y fragmentada que poseen las propuestas existentes con respecto a las necesidades requeridas para un proceso que guíe el desarrollo de proyectos de ingeniería de explotación de información.

De manera global, puede observarse que la totalidad de las propuestas se centran en el proceso orientado al producto (siendo CRISP-DM y sus sucesores los abordajes que contemplan la mayoría de las tareas requeridas), mientras que para el subproceso de gestión de proyectos, la mayoría de las propuestas cubren menos de la mitad de las actividades, siendo las tareas de definir el plan del proyecto y la revisión del proceso (lecciones aprendidas) las comúnmente señaladas. Adicionalmente, solo una de las propuestas presenta una visión detallada en la cual se incluyen técnicas o procedimientos claros que guían el desarrollo de las actividades, mientras que en la totalidad restante (con excepción de CRISP-DM que brinda lineamientos a considerar) sólo

DIMENSIONES DE ANÁLISIS		KDD	SEMMA	CRISP-DM	Catalyst	IKDDM	MPIMD	ASD-BI	FMDS	TDSP
Subproceso Gestión	Iniciación	☐	☐	▒	☐	▒	▒	☐	☐	☐
	Planificación	☐	☐	▒	☐	▒	■	☐	☐	☐
	Ejecución	☐	☐	☐	☐	☐	■	☐	☐	☐
	Monitoreo y Control	☐	☐	☐	☐	☐	▒	☐	☐	☐
	Cierre	☐	☐	▒	☐	▒	▒	▒	☐	▒
Subproceso Desarrollo	Entendimiento del Negocio	▒	☐	■	■	■	■	■	■	■
	Entendimiento de los Datos	☐	▒	■	■	■	■	■	■	■
	Preparación de los Datos	■	▒	■	■	■	■	■	■	■
	Modelado	▒	▒	▒	▒	▒	▒	▒	▒	▒
	Evaluación	☐	■	■	▒	■	■	■	■	■
	Despliegue	■	☐	■	■	■	■	■	■	■
Elementos de entrada y salida		☐	☐	☐	☐	■	☐	☐	☐	▒
Técnicas / Procedimientos		☐	☐	▒	■	■	☐	☐	☐	▒
Productos estandarizados		☐	☐	☐	☐	☐	☐	☐	☐	■

Tabla 3.1. Análisis comparativo de las propuestas existentes

mencionan de manera global los alcances de las mismas. Finalmente, solo TDSP provee al usuario de productos estandarizados que normalicen los resultados de las actividades. A continuación se brinda una descripción detallada de la evaluación de cada abordaje analizado:

- **KDD**: su importancia como pionera en identificar la necesidad de un proceso general que se diferencia de la actividad de minería de datos es innegable, sin embargo se presenta como la segunda opción con menor cubrimiento de los aspectos evaluados. En ella se identifican de manera completa las dos fases de desarrollo (Preparación de los datos y Despliegue), mientras que en las fases de entendimiento del negocio y modelado hay actividades faltantes (no se identifican criterios de éxito y no se definen los criterios a utilizar para la optimización de los parámetros del modelo).
- **SEMMA**: es el enfoque con menor cubrimiento, indicando de manera completa únicamente la etapa de preparación de los datos, mientras que las fases de Entendimiento y Preparación de los Datos y modelado se definen de manera parcial. En las primeras dos fases se omite la necesidad de evaluar la calidad de los datos y la limpieza de los mismos, mientras que en la última, únicamente se menciona la necesidad de seleccionar y aplicar el modelo.

- **CRISP-DM**: presenta una propuesta del proceso orientado al producto que cubre los aspectos de todas las fases, con excepción del emparejamiento de los modelos a partir de las necesidades del cliente [Marbán et al., 2007; 2009; Rennolls y AL-Shawabkeh, 2008], brindando lineamientos respecto a los alcances de las actividades. Además, identifica cuatro actividades pertenecientes al proceso de gestión: identificar los recursos y evaluar las herramientas (fase iniciación), diseñar el plan del proyecto (Planificación) y realizar una revisión del proyecto (Cierre).
- **Catalyst**: es la primera propuesta que incorpora una descripción detallada de los pasos a realizar para la resolución del problema (a través de las cajas), introduciendo un mapeo desde la identificación de las necesidades de los interesados hasta la presentación de los patrones obtenidos, aunque no se indican las dependencias entre las actividades.
- **IKDDM**: adopta los elementos de CRISP-DM, incorporando la identificación de los elementos de entrada y salida, y técnicas y procedimientos aplicables a cada actividad, en respuesta a la necesidad expresada por múltiples autores de una propuesta que detalle las tareas a realizar, siendo la única en abordar dicha problemática.
- **MPIMD**: está basada en CRISP-DM, ampliando los aspectos correspondientes al subproceso de gestión de proyectos. En esta no solo se cubren los alcances de las fases de planificación y ejecución, sino que además se incorporan las tareas de selección del ciclo de vida, estudios de viabilidad pertenecientes a la fase Iniciación (quedando sin identificar los interesados y miembros del proyecto, y su interacción) y de seguimiento y evaluación de riesgos del proyecto correspondiente a la fase monitoreo y control (pero no se consideran las actividades asociadas con las peticiones de cambio). Si bien esta aproximación presenta la visión más completa entre los modelos de proceso existentes, la misma presenta una visión incompleta [Marbán et al., 2009] la cual posee inconsistencias en su construcción, dificultando su comprensión e implementación (ver sección 2.3, pág. 30 para más detalles).
- **ASD-BI**: adopta los elementos de desarrollo de CRISP-DM, resolviendo la carencia de su antecesora (ASD-DM). Además, altera el ciclo de vida tradicional embebido en el estándar de facto [Rennolls y AL-Shawabkeh, 2008; Alnoukari, 2012] por un enfoque ágil. De las actividades asociadas con la gestión de proyectos, únicamente considera la evaluación del proyecto (o lecciones aprendidas) perteneciente a la fase Cierre.
- **FMDS**: define la estructura del proceso en base a CRISP-DM y KDD, e incorpora aspectos de procesamiento de grandes volúmenes de datos.
- **TDSP**: presenta un ciclo de vida ágil basado en la estructura propuesta en CRISP-DM enfocado en mejorar el aprendizaje y la colaboración del equipo. Identifica los elementos de

salida esperados para cada etapa del proceso y define documentos estandarizados para cada uno de las salidas.

A partir de lo previamente expuesto, se desprende la necesidad de definir un modelo de proceso que proporcione un abordaje *completo*, cubriendo los aspectos vinculados con el proceso orientado al producto y de gestión, *integral*, el cual considere los elementos de la solución como un sistema (contemplando las dependencias entre los elementos de entrada y salida), *detallado*, indicando el conjunto de técnicas o procedimientos aplicables para cada tarea (especificando sus acciones requeridas) y *estandarizado*, mediante el uso de productos definidos los cuales faciliten la comprensión, implementación y calidad de los resultados.

3.3. IDENTIFICACIÓN DEL PROBLEMA ABIERTO Y FORMULACIÓN DE LAS PREGUNTAS DE INVESTIGACIÓN

Los modelos de proceso proporcionan al equipo de trabajo una guía detallada de los pasos a realizar para el desarrollo de un proyecto de ingeniería de explotación de información, proveyendo los elementos de entrada y salida a utilizar para cada una de las actividades requeridas. Para que dicho proceso pueda ser correctamente implementado, es importante definir como se deben llevar a cabo cada uno de los pasos (mediante técnicas o procedimientos) estableciendo productos estandarizados que faciliten la comprensión y ejecución de las actividades. Por otra parte, y habida cuenta de la complejidad incremental que revisten los proyectos [Kurgan y Musilek, 2006; Gallardo, 2009; Mariscal et al., 2010], así como también su interdisciplinaridad [Fayyad et al., 1996] y su dinamismo [Brachman & Anand, 1996], dicho proceso debe proporcionar una visión ingenieril que favorezca el seguimiento, trazabilidad, reproducción, control y mejora del mismo.

Sin embargo, en los últimos años se han indicado una serie de deficiencias en las propuestas existentes (detalladas en las secciones previas), asociadas con la dificultad de aplicar los procesos vigentes y las causas de fracaso de los proyectos (con una tasa cercana al 60% [Gondar, 2005; Marbán et al., 2009]). Las principales carencias son resumidas a continuación: a) falta de procesos completos, detallados y estandarizados que incorporen un enfoque ingenieril, b) inexistente o inadecuado soporte al usuario (realizando un énfasis en las etapas de comprensión de las necesidades del cliente y definición de los alcances del proyecto, y en su vinculación con la solución propuesta), c) deficiencias en la estimación del proyecto (a causa de herramientas y guías inexistentes) y d) necesidades insatisfechas asociadas con la mala calidad de los resultados.

En este contexto, el problema abierto de la tesis se formula como la necesidad de definir un modelo de proceso para proyectos de ingeniería de explotación de información [Kurgan y Musilek, 2006; Mariscal et al., 2010; Kdnuggets, 2014] el cual contemple los procesos orientado al producto y a la

gestión de proyectos. Este problema se enmarca en las dificultades señaladas por varios autores para el desarrollo de proyectos de ingeniería de explotación de información [Clifton y Thuraisingham, 2001; Charest y Delisle, 2006; Kurgan y Musilek, 2006; Marbán et al., 2007; Sharma y Osei-Bryson, 2009; Gallardo, 2009; Mariscal et al., 2010; do Nascimento y de Oliveira, 2012; El Sheikh y Alnoukari, 2012; Kdnuggets, 2014].

El problema abierto identificado [Sabato y Mackenzie, 1982; Riveros y Rosas, 1985] se puede caracterizar a partir de las siguientes preguntas de investigación [Creswell, 2002]:

Pregunta 1: ¿Es posible desarrollar un modelo de proceso integral el cual defina las actividades y dependencias para proyectos de explotación de información desde una visión ingenieril que incorpore los conceptos de planificación, administración y control requeridos en todo proceso?

Pregunta 2: De ser viable, ¿Es posible definir el modelo de acuerdo a las prácticas y propuestas vigentes?

Pregunta 3: ¿Existen indicios que deriven en la necesidad de adaptar o redefinir las prácticas existentes?

Pregunta 4: ¿Es posible articular técnicas o procedimientos de ingeniería de explotación de información desarrollados *ad hoc* en dicho modelo de proceso?

Pregunta 5: En caso de poder articular las técnicas con el modelo de proceso propuesto: ¿Existen técnicas faltantes para las actividades? De existir, ¿pueden dichas técnicas ser adaptadas de otras disciplinas o desarrolladas ad hoc?

4. SOLUCIÓN

La solución propuesta se enmarca en el concepto de Ingeniería de Explotación de Información, entendiéndose a este como el conjunto de pasos a realizar para ordenar, controlar y gestionar la tarea de encontrar patrones de conocimiento en masas de información [García-Martínez et al. 2011], las cuales pueden hallarse en distintas estructuras de almacenamiento [Kruse y Borgelt, 2003; Gopal et al., 2011], bases de datos, imágenes, documentos, grafos, entre otras, y en la convención "metodología = proceso + técnicas" [Rodriguez, 2015] (sección 2.5, pág. 48), en la cual se define a la metodología como la instanciación (o selección) de las técnicas a utilizar para cada fase del proceso.

En el contexto descripto en los capítulo previos, y brevemente resumido en el párrafo anterior, se propone un Proceso para Proyectos de Explotación de Información (MoProPEI), el cual se centra en tres aspectos: entender a las piezas de conocimiento como eje principal del proyecto, entender a la explotación de información como una actividad que integra un proceso mayor (cuyas otras actividades poseen similar relevancia respecto a la calidad final del producto [Fayyad et al., 1996]) e integrar al proceso tradicional orientado al producto con la visión de gestión de proyectos, los cuales son cada vez más grande [Mariscal et al., 2010], complejos [Kurgan y Musilek, 2006; Gallardo, 2009], interdisciplinarios [Fayyad et al., 1996] y dinámicos [Brachman & Anand, 1996].

Para facilitar al lector en la comprensión de la propuesta, se estructura el capítulo de la siguiente forma: en la sección 4.1, se presenta la visión general de la propuesta, brindando una comprensión global de cómo se estructura el proceso de manera conjunta con una breve descripción de cada uno de los elementos que lo integran. La importancia de esta visión global, está dada por la fuerte dependencia entre los elementos de las partes que conforman la propuesta, y que la secuencia de implementación de las actividades no es lineal, es decir, la ejecución de todas las actividades que forman el proceso no son secuenciales, y no siempre el orden de ejecución de las mismas es idéntico (debido a la estrategia de desarrollo o ciclo de vida seleccionado). En la sección 4.2, se introduce el proyecto "ENPreCoSP-2011", seleccionado como prueba de concepto, mediante el cual se ilustra la implementación del proceso, con el objetivo de facilitar la comprensión de la aplicación de las actividades a partir de la puesta en práctica de las mismas en el caso mencionado. Finalmente, se describe en detalle la propuesta dividida por los subprocesos (primer nivel estructural de la propuesta) *Gestión* y *Desarrollo (*secciones 4.3 y 4.4 respectivamente), incluyendo la implementación de cada elemento en el proyecto.

4.1. PROPUESTA DE PROCESO: MoProPEI

MoProPEI es un proceso para proyectos de explotación de información, que surge a partir del análisis de las problemáticas identificadas en las propuestas existentes y las dificultades experimentadas en la implementación de las mismas, el cual introduce la visión de control y gestión de proyectos en el proceso de extracción de piezas de conocimiento para dar soporte al proceso de toma de decisiones.

Este modelo se concibe desde la perspectiva de la compleja dinámica de los proyectos, a partir de la cual se define la incorporación de una capa transversal al tradicional proceso orientado al producto, la cual provea una visión de planificación, seguimiento, control y revisión de las tareas, así como la posibilidad de adaptar el flujo del proyecto a las necesidades del mismo. Adicionalmente, se introducen ajustes al proceso orientado al producto, identificados a partir de las dependencias entre las distintas actividades.

MoProPEI se estructura de forma jerárquica mediante tres niveles, cada uno de los cuales presenta un mayor grado de especificidad en los objetivos de sus elementos. Estos niveles son:

- **Subproceso**: Es la división de mayor nivel de generalidad, la cual está integrada por fases. Brinda una capa de separación entre las actividades dependientes de la estructura y dinámica del progreso del proyecto y aquellas transversales, que dan soporte a lo largo del mismo.
- **Fase**: Conformada por actividades cuyos objetivos se encuentran asociados a un propósito general en común.
- **Actividad**: Conjunto de tareas asociadas por el alcance de sus objetivos específicos y las vinculaciones entre sus conocimientos. Las tareas a realizar tienen asociado una o más técnicas, las cuales brindan un conjunto de pasos o procedimientos que permiten generar a partir de un conjunto de elementos de entrada, los elementos de salida deseados.

La propuesta está compuesto por dos subprocesos: *Desarrollo* (MoProPEI-D), enfocado a las actividades del producto, esto es, aquellas directamente asociadas con el proceso de generación de un modelo que extraiga piezas de conocimiento que den soporte al proceso de tomas de decisiones, y *Gestión* (MoProPEI-G), orientado a la planificación, control y administración de las tareas asociadas al proyecto, con el objetivo de garantizar su desarrollo de acuerdo a lo planificado. Cada subproceso está compuesto por un conjunto de fases, las cuales agrupan a las actividades según las finalidades de las mismas.

El subproceso Gestión contiene cinco fases:

- *Iniciación*: se realizan los esfuerzos iniciales para entender distintos aspectos del proyecto que permitan identificar la posibilidad de realizar el mismo de forma exitosa.
- *Planificación*: se evalúan y prevén las actividades y recursos que estarán involucrados a lo largo del proyecto.
- *Soporte*: actividades que apoyan la estructuración y progreso del proyecto. El nombre de la fase fue alterado respecto a la propuesta original (sección 2.5, pág. 48), con el propósito de eliminar ambigüedades con respecto a la fase de implementación del subproceso Desarrollo, sin embargo, sus objetivos y alcances se mantienen sin alterar.
- *Control*: se analizan y evalúan los distintos aspectos del proyecto con el objetivo de detectar posibles desvíos y realizar los ajustes correspondientes.
- *Cierre*: se realizan las tareas formales de finalización del proyecto, analizando y evaluando los resultados obtenidos durante el mismo.

El subproceso Desarrollo está integrado por seis fases:

- *Entendimiento del Dominio*: en donde se comprende en detalle las características del dominio del negocio, el proyecto y los recursos disponibles para el desarrollo del mismo.
- *Entendimiento de los Datos*: se analiza y evalúan las distintas fuentes de datos disponibles para generar conocimiento a partir de los mismos. Esto incluye la investigación, comprensión y evaluación de la calidad de los datos.
- *Modelado*: en donde se identifican las distintas soluciones y el conjunto de técnicas y algoritmos de explotación de información a utilizar. En [Britos et al., 2008], se propuso y demostró la identificación (a partir de las necesidades del negocio) del proceso de explotación de información (sección 2.4.1.3, pág. 35), junto con los algoritmos a utilizar, procesando los datos de manera posterior a partir de las necesidades intrínsecas de los algoritmos seleccionados (debido a la dependencia de la preparación de los datos con el proceso de explotación de información seleccionado). En este contexto, se introduce una nueva fase destinada a identificar los modelos a utilizar de manera previa a la preparación de los datos, modificando la tradicional estructura, evitando de esta forma realizar un esfuerzo adicional en la adecuación de los datos.
- *Preparación de los Datos*: en esta etapa se realizan las adecuaciones de los datos disponibles de acuerdo a las necesidades del modelo, preparando los datos para la aplicación de los mismos. Esto incluye la limpieza, formateo y construcción de los campos.
- *Implementación*: se aplican los modelos y las estrategias definidas, con el objetivo de obtener la mejor configuración de los modelos los cuales brinden resultados relevantes y de interés para los distintos problemas de negocio identificados. El nombre de esta fase fue

adecuado, entendiendo que explotación de información (o minería de datos como se menciona en las propuestas predecesoras) involucra una actividad de un conjunto mayor.

- *Evaluación y Presentación*: se controlan que los resultados obtenidos cumplan con los requerimientos de los interesados, garantizando la correcta comprensión e inserción de los mismos en el dominio del cliente. El término Presentación se utiliza en lugar de Despliegue, con el propósito de acentuar como objetivo del proceso el descubrimiento de patrones comprensibles, que permitan interpretar conocimiento interesante y útil para dar soporte al proceso de toma de decisiones [Fayyad, et al., 1996].

Se considera relevante destacar, que la ejecución de los subprocesos no es secuencial, sino que la misma debe realizarse de forma paralela, debido que el subproceso de gestión provee de soporte a las actividades pertenecientes al subproceso de Desarrollo (aquellas destinadas específicamente a la construcción del producto final).

Para facilitar la implementación de la propuesta, se brinda una visión integrada del proceso (subprocesos, fases y actividades), detallando las relaciones y dependencias entre los elementos (productos de entrada y salida), así como las posibles técnicas a utilizar para la ejecución de un proyecto.

La figura 4.1, presenta una primera visualización de los componentes generales que forman parte del proceso, identificando los elementos de mayor jerarquía previamente descriptos (subprocesos y fases), y las vinculaciones externas entre los mismos, es decir, las dependencias de conceptos existentes entre las fases que integran a los subprocesos. Dichas dependencias son identificadas con una flecha apuntando en su extremo con un triángulo relleno a la fase que insume al menos uno de los elementos generados por la fase donde se origina la flecha. En este contexto, la primera fase del subproceso Gestión (Iniciación) provee de información para las fases Entendimiento de los Datos, Modelado, Preparación de los Datos e Implementación (pertenecientes al subproceso Desarrollo), mientras que en el subproceso Desarrollo, la fase Entendimiento del Negocio aprovisiona las fases de Iniciación y Planificación, y Evaluación y Presentación abastece a la última fase de Gestión (Cierre).

Se considera relevante destacar que las flechas reflejan únicamente aquellos productos internos de las fases los cuales se utilizan como elementos de entrada por una fase que no pertenece al mismo subproceso, es decir, no se incluyen las dependencias entre las fases del mismo subproceso, ni aquellas que determinan la ejecución de las actividades, como por ejemplo: la planificación de las actividades y la gestión de las peticiones de cambios, que si bien tienen un impacto en otras fases, no son elementos de entrada para la preparación de un producto interno específico.

Las fases presentadas en la figura 4.1, agrupan un conjunto de actividades por sus objetivos en común. Dichas actividades, representan el desarrollo de una tarea específica las cuales tienen asociadas una serie técnicas que describen los pasos que permiten a partir de un conjunto de elementos de entrada, producir elementos de salida que favorezcan el progreso del proyecto. En la tabla 4.1(a y b) y 4.2 (a y b) se resume la estructura de la propuesta (subprocesos de gestión y desarrollo, respectivamente), presentando las fases y actividades que conforman cada subproceso y sus objetivos asociados.

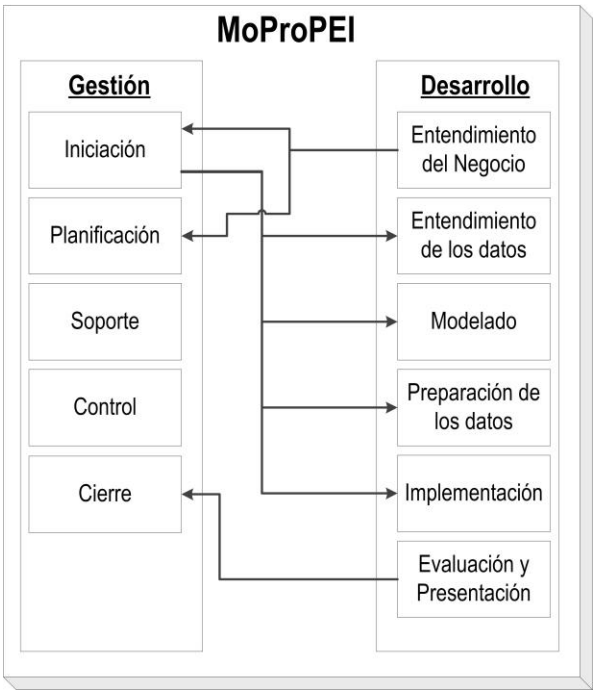


Figura 4.1. MoProPEI : Estructura General (subprocesos y fases)

FASE	ACTIVIDAD	OBJETIVO
Iniciación	Exploración Inicial del Proyecto	Se identifican los miembros de interés para el proyecto y las posibles situaciones de riesgo
	Definición de la Comunicación	Se definen las necesidades y canales de comunicación durante el desarrollo del proyecto
	Evaluación de la Situación	Se analizan las herramientas de utilidad para el desarrollo del proyecto, determinando la viabilidad del mismo
	Definición del Ciclo de Vida	Se establece de acuerdo a las características del proyecto, el flujo mediante el cual se llevarán a cabo las tareas de desarrollo
Planificación	Planificación de la Mediciones	Se definen las mediciones que se llevarán a cabo durante el proyecto, realizando una estimación inicial del esfuerzo requerido

Tabla 4.1.a. MoProPEI: Estructura subproceso Gestión

FASE	ACTIVIDAD	OBJETIVO
Planificación	Planificación de las Actividades	Se definen las tareas a realizar y sus alcances en el transcurso del tiempo para el desarrollo del proyecto
	Planificación de los Recursos	Se prevén los recursos (humanos y materiales) necesarios para el desarrollo de las actividades en el tiempo
	Planificación de las Responsabilidades	Se deja registro formal de las responsabilidades y obligaciones de las partes involucradas
Soporte	Mediciones del Proyecto	Se calculan las métricas durante el desarrollo del proyecto
	Gestión de la Configuración	Se mantiene registro de la evolución de los productos, garantizando su trazabilidad
Control	Gestión del Desarrollo	Se evalúa el desarrollo del proyecto, verificando que el mismo se efectúe de acuerdo a lo planificado y pactado con el cliente, y dejando registro formal de cualquier desvió, cambio o posible evento riesgoso que aconteciese
	Control de las Actividades	Se evalúan las situaciones potencialmente peligrosas para el desarrollo del proyecto, realizando un seguimiento, control y registro de acontecimientos, así como de las acciones realizadas
	Gestión del Cambio	Se realiza un proceso de evaluación formal de las peticiones de cambio, determinando como resultado la procedencia o no del mismo y sus efectos asociados
Cierre	Formalización Externa del Cierre del Proyecto	Se obtiene la conformidad del cliente, respecto a los compromisos asumidos en la propuesta del proyecto, dejando registro formal de la finalización del mismo
	Formalización Interna del Cierre del Proyecto	Se evalúa el desarrollo del proyecto, dejando registro de aquellos aspectos que sean de valor para proyectos futuros

Tabla 4.1.b. MoProPEI: Estructura subproceso Gestión

FASE	ACTIVIDAD	OBJETIVO
Entendimiento del Negocio	Análisis del Negocio	Identificar y comprender las metas del proyecto, en base a las necesidades del requirente y los interesados
	Comprensión del Problema de Negocio	Se definen las problemáticas específicas del negocio a analizar
Entendimiento de los Datos	Análisis de los Datos	Se evalúan las variables disponibles en las distintas fuentes de información
	Exploración de los Datos	Se analizan los valores de los campos identificados de interés para los distintos problemas de negocio, con el objetivo de comprender las características de la población o muestra de estudio, identificando relaciones iniciales entre las distintas variables estudiadas
	Evaluación de los Datos	Se analizan los campos identificados de interés para los distintos problemas de negocio, identificando aquellas características que puedan afectar la calidad del modelo
Modelado	Modelado del Problema	Se realiza un modelado de representación de los problemas de negocio, identificando los métodos de explotación de información a utilizar
	Configuración del Modelo	Se definen los elementos que conforman la estrategia de implementación y evaluación de los distintos modelos para la extracción de patrones vinculados con el problema de negocio

Tabla 4.2.a. MoProPEI: Estructura subproceso Desarrollo

FASE	ACTIVIDAD	OBJETIVO
Preparación de los Datos	Construcción de la Fuente Temporaria de Datos	Se realizan las tareas finales para la generación de las fuentes de datos requeridas para las distintas etapas de implementación del modelo
	Adecuación de la Fuente Temporaria de Datos	Se transforman los datos de acuerdo a las necesidades del modelo
Implementación	Selección del Modelo	Se define el proceso mediante el cual se evalúa la calidad de los modelos y el criterio de selección del mismo
	Explotación de Información	Se aplican los algoritmos de explotación de información, documentando los resultados obtenidos
Evaluación y Presentación	Evaluación de los Resultados	se evalúa la validez de los patrones de conocimiento para los problemas de negocio
	Presentación de los Resultados	Se garantiza la correcta transferencia del conocimiento extraído para dar soporte al proceso de toma de decisiones

Tabla 4.2.b. MoProPEI: Estructura subproceso Desarrollo

Con el objetivo de facilitar al lector en la comprensión de la propuesta y la identificación de la pertenencia de los elementos presentados con respecto la estructura del proceso, se utiliza una regla nemotécnica, indicada entre paréntesis a continuación de cada elemento (subproceso, fase, actividad y productos de salida) del proceso.

La estructura de la regla nemotécnica presenta la siguiente lógica: Cada jerarquía de elementos estará separada por punto ".", indicadas de manera decreciente, es decir, el elemento más general primero a la izquierda, disminuyendo hacia la derecha. Debido a que la propuesta está conformada por 3 tipos de jerarquías de elementos (subproceso, fase, actividad), y los productos de salida, el formato de la estructura de la regla es: "<**sigla subproceso**>.<**sigla fase**>.<**sigla actividad**>.<**sigla producto salida**>". Por simplicidad, se omiten aquellos puntos sin sigla en las instancias superiores, por ejemplo: para indicar la fase de entendimiento del negocio (perteneciente al subproceso de desarrollo) la sigla identificando los 4 niveles jerárquicos es: "D.EN.." (**D**esarrollo.**E**ntendimiento del **N**egocio.<vacío>.<vacío>) por convención serán eliminados los puntos vacíos quedando: "D.EN".

Por último, debido a la cantidad de elementos en cada jerarquía, el número de letras a utilizar para la sigla que identifique al elemento en cada nivel se ve incrementado en 1, es decir, que para el primer elemento jerárquico (**S**ubproceso) se utiliza 1 letra, para el segundo (**F**ase) 2 letras, **A**ctividad 3 letras y **E**lementos de salida 4 letras, siendo el patrón de la estructura de la expresión: "S.FF.AAA.EEEE".

4.2. PRUEBA DE CONCEPTO: ENPreCoSP-2011

El caso seleccionado como prueba de concepto, se centra en el análisis de la Encuesta Nacional sobre Prevalencias de Consumo de Sustancias Psicoactivas [Instituto Nacional de Estadística y Censos, 2016], realizada por el Instituto Nacional de Estadísticas y Censos (INDEC) en noviembre 2011, en las localidades de 5000 y más habitantes, en la totalidad de Argentina, a personas de 16 a 65 años de edad. El objetivo general de esta encuesta fue contribuir a actualizar el sistema de información sobre el consumo de sustancias psicoactivas a nivel nacional y, de esa manera, al diseño de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población. La misma cuenta con información de personas que declararon haber consumido sustancias psicoactivas en diferentes períodos de referencia (prevalencias). Las sustancias psicoactivas contempladas son: drogas legales o sociales (tabaco, bebidas alcohólicas), ilegales (marihuana, cocaína, pasta base, éxtasis, opiáceos y anestésicos, crack, alucinógenos, inhalables y otras drogas) y fármacos (estimulantes, tranquilizantes, anorexígenos). Adicionalmente, se cuenta con un conjunto de datos del encuestado relacionados con los aspectos sociodemográficos, socioeconómico, educativo y de su entorno familiar social.

Se dispone del registro de más de 34000 encuestados, con información de 292 preguntas por persona (pudiendo algunas de ellas no ser respondidas). Estas cubren los siguientes aspectos del individuo: Dominios de estimación geográfica, Características de la vivienda donde reside, Características del hogar al que pertenece, Características del jefe del hogar, Características de la persona seleccionada, Autopercepción de salud y entorno, Consumo de sustancias psicoactivas, Impacto del consumo en las actividades habituales y Demandas de tratamiento de la persona seleccionada. Adicionalmente, se cuenta con una muestra del cuestionario y una guía para la utilización de la base de datos, donde se describen consideraciones técnicas para el uso de la información recabada. A continuación se listan algunas de las características más relevantes del set de datos:

- **ABU_C:** Abuso de cerveza.
- **ABU_V:** Abuso de vino.
- **BHCH04:** sexo.
- **BHCH05:** ¿Cuál es su edad en años cumplidos?
- **BIAC01:** ¿Conoce personas cercanas a usted que en la actualidad consuman alguna sustancia como marihuana, cocaína, éxtasis, etc.?
- **BIES04:** ¿Cuándo fue la primera vez que probó estimulantes sin indicación médica?
- **CAT_OCUP:** Categoría ocupacional.
- **CONDACT:** Condición de actividad laboral.
- **GRUPEDAD:** Edad separada por rangos.
- **NBI_TOTAL:** Indicador de necesidades básicas insatisfechas de hogar.
- **NIVINSTR:** Nivel de máximo estudios alcanzado.
- **POB_URB:** Agrupamiento de poblaciones urbanas.
- **PROV:** Provincia de residencia.
- **PV_MA:** Prevalencia de vida de consumo de marihuana.

- **RANGOING:** rango de ingreso.

El objetivo general del proyecto, de acuerdo a lo expuesto por el cliente, es la identificación de un proceso que permita automatizar o semi-automatizar el análisis de la población de consumo de sustancias psicoactivas y caracterizar el comportamiento de la misma.

4.3. MoProPEI-G: Subproceso Gestión (G)

El subproceso de gestión, se concibe de forma transversal a las actividades de desarrollo, cuya ejecución no es de forma lineal (definidas por el modelo de ciclo de vida del proyecto), sino que dichas tareas se realizan durante todo el progreso del proyecto. Dicho subproceso se centra en controlar que se cumplan las expectativas del proyecto, en cuanto a objetivos, recursos y tiempo, balanceando los mismos respecto a la calidad de los elementos con el fin de satisfacer exitosamente las necesidades del cliente. En este, se realizan las tareas de administración del proyecto, comprensión la situación del cliente, identificación, planificación y control de los recursos, identificación del modelo de ciclo de vida del proyecto, control de la ejecución de las actividades, realización las mediciones, determinación de la viabilidad y formalización del cierre del proyecto.

El subproceso está conformado por cinco fases: Iniciación, Planificación, Soporte, Control y Cierre. Cada una de las cuales está compuesta por actividades, que describen las acciones a realizar para obtener los productos internos del proyecto. Como se mencionó previamente, la figura 4.1, presenta las relaciones y dependencias desde la perspectiva de las fases, abstrayéndose del elemento específicamente vinculado y las asociaciones internas entre las fases y actividades pertenecientes a un mismo subproceso. La figura 4.2, amplía dicho concepto ilustrando las relaciones entre las fases y las distintas actividades que componen al subproceso Gestión, haciendo uso del formalismo de representación propuesto en [Hossian, 2012; Rodriguez, 2015]. Este formalismo fue seleccionado ante otras posibles opciones como el diagrama de actividades con calles o el modelo y notación de procesos de negocio (BPMN), por su énfasis y simplicidad en describir las dependencias entre los elementos.

En los extremos superior e inferior de la figura, utilizando elementos en forma de elipse divididos en dos partes, se presenta el nombre de la actividad y la fase (en la parte inferior y superior, respectivamente) de los elementos pertenecientes al subproceso Desarrollo. Las flechas representan dependencias entre las actividades (indicándose las entradas del lado izquierdo de las actividades y las salidas del lado derecho). El subproceso Gestión, se encuentra conformado por cinco fases, las cuales se detallan en las siguientes secciones: Iniciación (sección 4.3.1), Planificación (sección 4.3.2), Soporte (sección 4.3.3), Control (sección 4.3.4) y Cierre (sección 4.3.5).

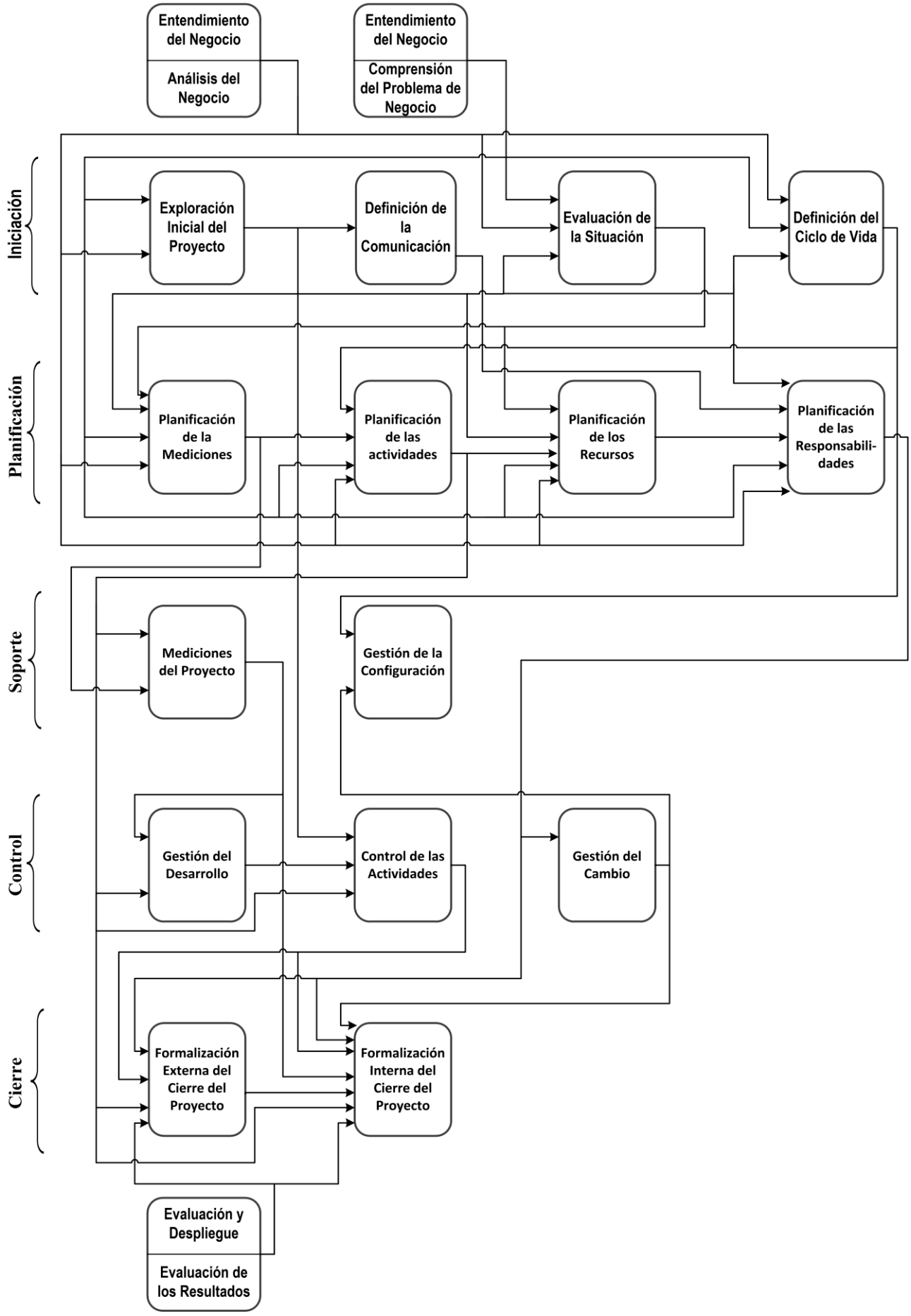


Figura 4.2. MoProPEI-G: Subproceso de Gestión

4.3.1. Fase: Iniciación (G.IN)

Durante la fase Iniciación se realizan, a partir de las primeras interacciones con el cliente y los expertos del dominio, evaluaciones respecto a las características del proyecto a abordar, los recursos humanos involucrados y la valoración de su posibilidad de éxito. El análisis se realiza desde la perspectiva de la organización que llevará a cabo el proyecto, evaluando los riesgos y beneficios del mismo.

Esta fase se encuentra integrada por cuatro actividades: *Exploración Inicial del Proyecto*: donde se identifican los miembros de interés para el proyecto y las posibles situaciones de riesgo durante el desarrollo del mismo (sección 4.3.1.1), *Definición de la Comunicación*: previendo las necesidades y canales de comunicación a lo largo del proceso (sección 4.3.1.2), *Evaluación de la Situación*: se analizan las herramientas de utilidad para el desarrollo del proyecto, determinando la viabilidad del mismo (sección 4.3.1.3) y *Definición del Ciclo de Vida*: donde se establece de acuerdo a las características del proyecto, el flujo mediante el cual se llevarán a cabo las tareas de desarrollo del mismo (sección 4.3.1.4). La figura 4.3, presenta una visión resumida de las actividades que integran la fase y sus elementos de entrada y salida (las imágenes de cada formalismo son representaciones miniatura de los mismos).

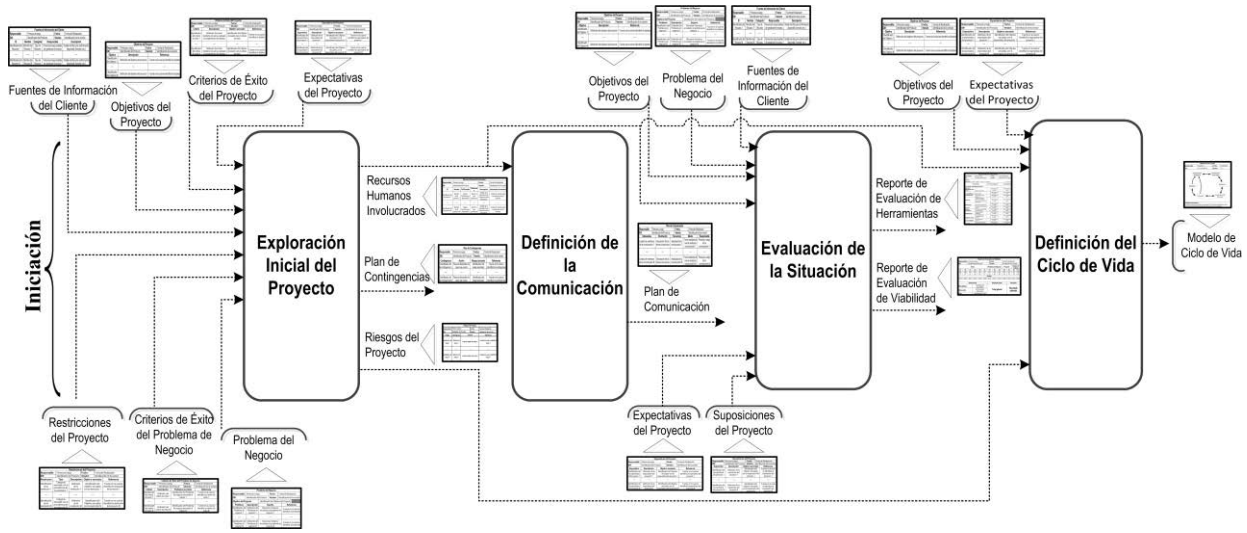


Figura 4.3. Fase: Iniciación

4.3.1.1. Actividad: Exploración Inicial del Proyecto (G.In.EIP)

En esta actividad se realizan las tareas necesarias para definir un nuevo proyecto o iteración, identificando aquellas personas relevantes para su desarrollo, analizando aquellos aspectos vinculados con los posibles eventos durante el progreso del proyecto que afecten el desarrollo del mismo y las acciones a realizar en caso de contingencia.

Información de Entrada

- Fuentes de Información del Cliente (D.EN.AnN.FuIC)
- Objetivos del Proyecto (D.EN.AnN.ObPr)
- Criterios de Éxito del Proyecto (D.EN.AnN.CrEP)
- Expectativas del Proyecto (D.EN.AnN.ExPr)
- Restricciones del Proyecto (D.EN.AnN.RePr)
- Problema del Negocio (D.EN.CPN.PrNe)
- Criterios de Éxito del Problema de Negocio (D.EN.CPN.CEPN)

Información de Salida

- Recursos Humanos Involucrados (G.In.EIP.ReHI)
- Riesgos del Proyecto (G.In.EIP.RiPr)
- Plan de Contingencias (G.In.EIP.PCon)

4.3.1.1.1. Formalismos Identificados

A continuación se listan los formalismos utilizados para el registro formal de los productos internos de la actividad.

Recursos Humanos Involucrados (G.In.EIP.ReHI): el formalismo propuesto en [Britos et al., 2008], identifica a los miembros de interés de la organización requirente, cuyo conocimiento u opinión sea relevante para el desarrollo del proyecto, así como aquellas personas que integran al equipo que llevarán a cabo el proyecto (tanto pertenecientes a la organización desarrolladora, como externos contratados). Para cada individuo identificado se incorpora una descripción del mismo (rol, área a la que pertenece e información descriptiva del área de pericia y de interés para el proyecto) y su información de contacto. La tabla 4.3 ilustra el formalismo descripto.

Recursos Humanos Involucrados					
Responsable:	Persona a cargo			**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto			**Versión:**	Identificación de la versión
ID	**Nombre**	**Rol/Posición**	**Pertenece a**	**Descripción**	**Información de contacto**
Identificador de la persona 1	Nombre de la persona 1	Cargo o Posición que desempeña	Sector al que pertenece	Detalle de la importancia de la persona para el proyecto	Listado de medios para comunicarse con la persona
…	…	…	…	…	…
Identificador de la persona N	Nombre de la persona N	Cargo o Posición que desempeña	Sector al que pertenece	Detalle de la importancia de la persona para el proyecto	Listado de medios para comunicarse con la persona

Tabla 4.3. Formalismo: Recursos Humanos Involucrados

Riesgos del Proyecto (G.In.EIP.RiPr): el formalismo propuesto en [Britos et al., 2008], provee un registro formal de los riesgos que puedan afectar, demorar o imposibilitar la exitosa finalización del

proyecto, proveyendo un identificador del mismo, su descripción (en las columnas homónimas), el conjunto de elementos que están vinculados con dicho alcance (objetivos, problemas de negocio, o el proyecto en general), registrándose dicha información en la columna "Alcance", y la referencia al elemento a partir del cual fue identificado. En la tabla 4.4 se ilustra el formalismo.

Riesgos del Proyecto			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Riesgo	**Descripción**	**Alcance**	**Referencia**
Identificador del riesgo 1	Definición del riesgo 1	Área de impacto del riesgo 1	Fuente en la cual se identificó el riesgo 1
…	…	…	…
Identificador del riesgo N	Definición del riesgo N	Área de impacto del riesgo N	Fuente en la cual se identificó el riesgo N

Tabla 4.4. Formalismo: Riesgos del Proyecto

Plan de Contingencias (G.In.EIP.PCon): en el formalismo propuesto en [Britos et al., 2008], se describen las acciones a realizar en caso que un riesgo acontezca, registrando un identificador y la acciones a realizar (columnas "Contingencia" y "Acción" respectivamente), el riesgo asociado a dicha acción (en la columna homónima) y la referencia al elemento a partir del cual fue identificado. La tabla 4.5 presenta el formalismo descripto.

Plan de Contingencias			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Contingencia	**Acción**	**Riesgo asociado**	**Referencia**
Identificador de la contingencia 1	Pasos a desarrollar en caso que ocurra	Identificador del riesgo asociado	Fuente en la cual se identificó la contingencia
…	…	…	…
Identificador de la contingencia N	Pasos a desarrollar en caso que ocurra	Identificador del riesgo asociado	Fuente en la cual se identificó la contingencia

Tabla 4.5. Formalismo: Plan de Contingencias

4.3.1.1.2. Técnica Identificada

Para el desarrollo de esta actividad se identifica la técnica propuesta en [Britos et al., 2008], denominada **"Caracterización del desarrollo del proyecto"** perteneciente a la Metodología para la educción de requerimientos para proyectos de explotación de información (sección 2.4.1.2, pág. 35). De su aplicación, se obtienen los formalismos Recursos Humanos Involucrados, Riesgos del Proyecto y Plan de Contingencias, descriptos en la sección anterior.

En primera instancia mediante la interacción con el cliente y el análisis del entorno del negocio, se identifican aquellas personas de la organización contratante que sean de interés en el desarrollo del proyecto, las cuales serán de utilidad para ampliar los conocimientos, así como validar resultados. Adicionalmente, se deja registro del personal (interno y externo) de la organización que se incorpore para el desarrollo del proyecto (es posible que a partir de los esfuerzos iniciales del proyecto, se identifiquen nuevos recursos humanos). Para cada persona identificada, se detalla en el formalismo Recursos Humanos Involucrados, el rol que ocupa, el lugar y área en la cual se desempeña, una descripción que permita comprender la relevancia de la persona en el proyecto (permitiendo comprender su importancia en las distintas etapas), y su información de contacto.

A partir de las distintas interacciones con los clientes y expertos de la organización contratante, así como el análisis del dominio realizado durante la etapa inicial de entendimiento del negocio, se identifican aquellas situaciones potencialmente peligrosas para el correcto progreso del proyecto y de los objetivos definidos. En caso de identificar dichos riesgos, se debe determinar cuál es el ámbito en el cual este tiene impacto (es decir, si el riesgo es específico de un objetivo, de un problema de negocio o del proyecto en general) y a partir de ello, identificar posibles acciones a realizar del tipo preventiva o correctiva. Dicha información será registrada en los formalismos Riesgos del Proyecto y Plan de Contingencias.

En la siguiente sección se presenta la aplicación de las técnicas y el registro de los formalismos asociados en el proyecto seleccionado como prueba de concepto.

4.3.1.1.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica caracterización del desarrollo del proyecto, perteneciente a la metodología para la educción de requerimientos para proyectos de explotación de información, la cual utiliza como insumos los formalismos: Fuentes de Información del Cliente (Tabla 4.55), Objetivos del Proyecto (Tabla 4.57), Criterios de Éxito del Proyecto (Tabla 4.58), Expectativas del Proyecto (Tabla 4.59), Restricciones del Proyecto (Tabla 4.61), Problema del Negocio (Tabla 4.64) y Criterios de Éxito del Problema de Negocio (Tabla 4.65). Si bien estos formalismos se presentan de forma posterior en el documento (sección 4.4.1), estos han sido desarrollados en paralelo durante las etapas iniciales del proyecto, transcribiéndolos con el mismo número de tabla, para facilitar al lector en la comprensión de la aplicación de la técnica.

Fuentes de Información del Cliente					
Responsable:		Esposito E.		**Fecha:**	05/04/2016
ID#:		D.EN.ANN.FUIC		**Versión:**	1.0
ID	**Nombre**	**Categoría**	**Responsable**	**Descripción**	
fuic.1	Documento para la utilización de la base de datos usuario 2011	Documento	-	Describe distintas consideraciones acerca de la base de datos obtenida a partir de la encuesta ENPreCoSP 2011, indicando los métodos de recolección aplicados, los objetivos de la encuesta y la descripción de los datos.	
fuic.2	Cuestionario ENPreCoSP 2011	Planilla	-	Ejemplo de cuestionario ENPreCoSP 2011	
fuic.3	Base ENPreCoSP 2011	Almacén de datos	-	Almacén de registros de respuestas del cuestionario ENPreCoSP 2011 en formato txt (separado por el carácter "|") conformado por 34343 personas que respondieron 292 preguntas. El primer renglón contiene los nombres de los campos. Decimales separados por punto (.)	

Tabla 4.55 (Transcripta). Prueba de Concepto - Fuentes de Información del Cliente

Objetivos del Proyecto			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.AnN.ObPr	**Versión:**	1.0
Objetivo	**Descripción**		**Referencia**
obpr.1	Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos		Entrevista 1

Tabla 4.57 (Transcripta). Prueba de Concepto - Objetivos del Proyecto

Criterios de Éxito del Proyecto				
Responsable:	Esposito E.	**Fecha:**	15/04/2016	
ID#:	D.EN.AnN.CrEP	**Versión:**	1.0	
Criterio	**Descripción**	**Objetivo asociado**		**Referencia**
crexpr.1	obtener piezas de conocimiento que favorezcan la comprensión del comportamiento de grupos masivos de personas, brindando indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales serán validadas por el cliente	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos		Entrevista 1

Tabla 4.58 (Transcripta). Prueba de Concepto - Criterios de Éxito del Proyecto

Expectativas del Proyecto				
Responsable:	Esposito E.	**Fecha:**	15/04/2016	
ID#:	D.EN.AnN.ExPr	**Versión:**	1.0	
Expectativa	**Descripción**	**Objetivo asociado**		**Referencia**
expr.1	Definir un proceso que automatice o semi-automatice el análisis de los datos, reduciendo el costo y tiempo asociado con la generación de resultados y el correspondiente accionar a partir de los mismos, que brinde indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales serán validadas por el cliente	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos		Entrevista 1

Tabla 4.59 (Transcripta). Prueba de Concepto - Expectativas del Proyecto

Restricciones del Proyecto				
Responsable:		Esposito E.	**Fecha:**	20/04/2016
ID#:		D.EN.AnN.RePr	**Versión:**	1.0
Restricción	**Tipo**	**Descripción**	**Objetivo asociado**	**Referencia**
repr.1	datos	Se identifica un desbalance entre la cantidad de registros que han consumido distintas Sustancias Psicoactivas y quienes no han consumido	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1
repr.2	datos	Se carece de información detallada respecto al grado o frecuencia de consumo de las sustancias psicoactivas	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1

Tabla 4.61 (Transcripta). Prueba de Concepto - Restricciones del Proyecto

Problema del Negocio				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.PRNE	**Versión:**	1.0	
Objetivo del Proyecto	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos			
Problema	**Descripción**		**Experto**	**Referencia**
prne.1	Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		(rehi.3) Silva H.	Entrevista 3

Tabla 4.64 (Transcripta). Prueba de Concepto - Problema del Negocio

Criterios de Éxito del Problema de Negocio				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.CEPN	**Versión:**	1.0	
Criterio	**Descripción**	**Problema asociado**		**Referencia**
cepn.1	Identificar aspectos que permitan comprender el comportamiento de grupos masivos de personas de acuerdo a la evaluación de Silva H. (rehi.3)	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		Entrevista 3

Tabla 4.65 (Transcripta). Prueba de Concepto - Criterios de Éxito del Problema de Negocio

Recursos Humanos Involucrados (G.In.EIP.ReHI): se deja registro de la información de los miembros de la organización que desarrolla el proyecto, así como de la organización cliente que realizó el contacto con el equipo de trabajo, siendo el único individuo externo que participe en el proyecto. Por cuestiones de privacidad, no se presenta la información de contacto de las personas. En primer lugar se procede a dar registro, en sus columnas homónimas, al nombre del individuo asignando un identificador, para evitar inconvenientes en caso que haya más de una persona que coincida en dicho campo. Luego, se detalla el rol/posición que dicha persona tiene en el desarrollo del proyecto, si la persona es quién realiza el pedido, si es un experto en algún tópico de interés para

el proyecto, el responsable del proyecto, etc. y se registra en el campo "pertenece a" la organización a la cual pertenece o si es un miembro externo a la misma (ya sea un asesor, técnico/proveedor de servicios, etc.), detallando aquellas consideraciones que sean relevantes y permitan identificar cuáles son los intereses y conocimientos que la persona posee (por ejemplo el área a la cual pertenece), los cuales serán de utilidad al momento de definir las interacciones (necesidad de información a obtener) que los mismos tendrán en el transcurso del proyecto. Por último, en el campo "información de contacto" se deja constancia de los medios de comunicación a utilizar con dicho individuo, así como cualquier aspecto relevante asociado al mismo (por ejemplo: restricciones en la comunicación). En la tabla 4.6, se muestra la información registrada, teniendo en consideración las salvedades previamente mencionada.

Recursos Humanos Involucrados					
Responsable:	Rodriguez H.			**Fecha:**	04/04/2016
ID#:	G.In.EIP.ReHI			**Versión:**	1.0
ID	**Nombre**	**Rol/Posición**	**Pertenece a**	**Descripción**	**Información de contacto**
rehi.1	Rodriguez H.	Líder de Proyecto	recurso interno	Persona encargada de la dirección del proyecto	Skype: XXXX
rehi.2	Esposito E.	Ingeniero de Explotación de información	recurso interno	Ingeniero de Explotación de información Senior	Skype: XXXXX
rehi.3	Silva H.	Cliente	Organización Contratante	Cliente Experto en el área	Correo: xxxx@gmail.com Skype: xxxxxxx

Tabla 4.6. Prueba de Concepto - Recursos Humanos Involucrados

Riesgos del Proyecto (G.In.EIP.RiPr): a partir de los recursos humanos involucrados, identificados en esta actividad y la dependencia del cliente/experto con la obtención del conocimiento y la validación de los resultados derivados (a partir del formalismo Criterios de Éxito del Proyecto), se identifica como posible evento crítico para el desarrollo del plan del proyecto, la ausencia del único experto del dominio durante etapas en las cuales se requiere de su intervención para el progreso del proyecto (entendimiento del negocio y los datos, evaluación de los resultados). El alcance del riesgo es transversal al desarrollo del proyecto y se asigna el identificador "risk.1". En la tabla 4.7, se ilustra el formalismo resultante.

Riesgos del Proyecto			
Responsable:	Rodriguez H.	**Fecha:**	05/04/2016
ID#:	G.In.EIP.RiPr	**Versión:**	1.0
Riesgo	**Descripción**	**Alcance**	**Referencia**
risk.1	No contar con la presencia del experto en etapas críticas de necesidad de interacción con el mismo	proyecto	

Tabla 4.7. Prueba de Concepto - Riesgos del Proyecto

Plan de Contingencias (G.In.EIP.Pcon): se registra a partir del riesgo previamente identificado, las posibles acciones a realizar de forma de prevenir, minimizar o ajustar el impacto de dicha contingencia. Para el riesgo "(risk.1) No contar con la presencia del experto en etapas críticas de necesidad de interacción con el mismo" se identifica como única acción el ajuste de los plazos del proyecto, asignando como identificador "cont.1". En la tabla 4.8, se ilustra la información previamente descripta registrada en el formalismo.

Plan de Contingencias			
Responsable:	Rodriguez H.	**Fecha:**	05/04/2016
ID#:	G.In.EIP.PCon	**Versión:**	1.0

Contingencia	**Acción**	**Riesgo asociado**	**Referencia**
cont.1	Ajustes en los plazos del proyecto	(risk.1) No contar con la presencia del experto en etapas críticas de necesidad de interacción con el mismo	

Tabla 4.8. Prueba de Concepto - Plan de Contingencias

4.3.1.2. Actividad: Definición de la Comunicación (G.In.DeC)

La falta o ineficiencia en las comunicaciones es una de las principales causas de demoras y fracasos en proyectos [Project Management Institute, Inc., 2013b]. Adicionalmente, para que la información sea útil, esta debe ser en tiempo y forma. Por dicho motivo, esta no puede ser realizada de manera improvisada o centrada únicamente en comunicaciones informales, sino que es necesario definir un plan o estrategia formal de comunicación de acuerdo a las necesidades e intereses de los miembros (tanto del equipo, como externos), el cual identifique que persona necesita qué información, en que tiempo y cuál es el mejor medio para comunicarse.

En este contexto, se propone la actividad *definición de la comunicación*, en la cual se establecen las estrategias formales de comunicación a partir de la necesidad e intereses de las partes involucradas en el proyecto.

Información de Entrada

- Recursos Humanos Involucrados (G.In.EIP.ReHI)

Información de Salida

- Plan de Comunicación (G.In.DeC.PCom)

4.3.1.2.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se utiliza el Plan de Comunicación (sección 2.4.2.1, pág. 45), presentado en [Verzuh, 2015], el cual se describe a continuación.

Plan de Comunicación (G.In.DeC.PCom): se formalizan las comunicaciones a realizar en el proyecto, indicando los interesados que formarán parte de la misma, la información general a tratar en dicha comunicación, la frecuencia con la cual será realizada, el medio de comunicación a utilizar y la persona responsable de llevar a cabo la misma. Cada información es registrada en su columna homónima. La tabla 4.9 ilustra el formalismo descripto.

Plan de Comunicación				
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización	
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión	
Interesados	**Información**	**Frecuencia**	**Medio**	**Responsable**
Listado de miembros de la comunicación 1	Descripción de los temas a comunicar	Asiduidad de la comunicación 1	Forma mediante la cual se realiza la comunicación 1	Persona a cargo de la comunicación 1
…	…	…	…	…
Listado de miembros de la comunicación N	Descripción de los temas a comunicar	Asiduidad de la comunicación N	Forma mediante la cual se realiza la comunicación N	Persona a cargo de la comunicación N

Tabla 4.9. Formalismo: Plan de Comunicación

4.3.1.2.2. Técnica Identificada

Para el desarrollo de esta actividad se identifica la técnica **"Definición de la Comunicación"** [Verzuh, 2015], en la cual se define el plan de comunicación, detallando la necesidad de información para los distintos interesados, así como el medio indicado para que dicha comunicación sea realizada con éxito. Su objetivo es definir el flujo de información requerida a lo largo del proyecto, para garantizar que todas las partes involucradas estén en constante conocimiento de las necesidades y/o progresos. Este análisis debe realizarse tanto para el personal de la organización que desarrolla el proyecto, garantizando que en todo momento estén informados de los cambios y evoluciones del proyecto, así como del cliente y los expertos, manteniendo el apoyo e interés de los mismos durante el proceso. Como resultado de esta técnica, quedan definidos los recursos humanos y/o roles que intervienen, la información que debe comunicarse, la frecuencia, el medio y el responsable de la misma. En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo para la prueba de concepto.

4.3.1.2.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica Definición de la Comunicación, la cual utiliza como insumo los Recursos Humanos Involucrados (Tabla 4.6). El formalismo previamente generado, identificado como elemento de entrada, será transcripto con el mismo número de tabla, para facilitar al lector en la comprensión de la aplicación de la técnica.

Recursos Humanos Involucrados					
Responsable:	Rodriguez H.			**Fecha:**	04/04/2016
ID#:	G.In.EIP.ReHI			**Versión:**	1.0
ID	**Nombre**	**Rol/Posición**	**Pertenece a**	**Descripción**	**Información de contacto**
rehi.1	Rodriguez H.	Líder de Proyecto	recurso interno	Persona encargada de la dirección del proyecto	Skype: XXXX
rehi.2	Esposito E.	Ingeniero de Explotación de información	recurso interno	Ingeniero de Explotación de información Senior	Skype: XXXXX
rehi.3	Silva H.	Cliente	Organización Contratante	Cliente Experto en el área	Correo: xxxx@gmail.com Skype: xxxxxxx

Tabla 4.6 (Transcripta). Prueba de Concepto - Recursos Humanos Involucrados

Plan de Comunicación (G.In.DeC.PCom): previamente fueron identificados tres recursos asociados al proyecto, dos miembros del equipo de trabajo y un experto/cliente, previendo tres tipos de comunicaciones: de comprensión del proyecto, de reporte de avances y de estado interno del proyecto. La primera de ellas, tiene como objetivo mantener un continuo vínculo con el experto e interesado del negocio, analizando los alcances y restricciones del proyecto, formando parte los tres miembros del proyecto. Dicha comunicación se realiza de manera semanal durante el periodo de entendimiento del negocio.

El segundo tipo de comunicación, es para mantener al cliente constantemente informado durante el desarrollo del proyecto, pudiendo identificar posibles cambios en los intereses o nuevas necesidades. La información se brinda de manera mensual y participan los tres involucrados en el proyecto. Finalmente, el último tipo de comunicación tiene como objetivo mantener al equipo de trabajo informado respecto al estado del proyecto y los problemáticas que pudiesen ocurrir durante el desarrollo del mismo. Participan los miembros del equipo de trabajo y serán realizadas de manera bisemanal. El modo en el cual las comunicaciones serán realizadas es por videollamada, mediante la herramienta Skype, siendo el responsable de las mismas el líder del proyecto. La tabla 4.10 ilustra el resultado obtenido de aplicar la técnica a la prueba de concepto.

Plan de Comunicación				
Responsable:	Rodriguez H.	**Fecha:**		04/04/2016
ID#:	G.In.DeC.PCom	**Versión:**		1.0
Interesados	**Información**	**Frecuencia**	**Medio**	**Responsable**
(rehi.1) Rodriguez H. (rehi.2) Esposito E. (rehi.3) Silva H.	Comprensión del Proyecto	semanal durante el periodo de entendimiento del negocio	Skype	(rehi.1) Rodriguez H.
(rehi.1) Rodriguez H. (rehi.2) Esposito E. (rehi.3) Silva H.	Avances del Proyecto	mensual	Skype	(rehi.1) Rodriguez H.
(rehi.1) Rodriguez H. (rehi.2) Esposito E.	Estado del Proyecto	bisemanal	Skype	(rehi.1) Rodriguez H.

Tabla 4.10. Prueba de Concepto - Plan de Comunicación

4.3.1.3. Actividad: Evaluación de la Situación (G.In.EvS)

En esta actividad se analiza la posibilidad de éxito del proyecto, teniendo en consideración los objetivos y las posibles soluciones que brindan las distintas herramientas de explotación de información existentes. Los objetivos de la actividad son: seleccionar las herramientas a utilizar y determinar la viabilidad del proyecto.

Como resultado de esta actividad se define si se realizará o no el proyecto, y que herramientas disponibles se encuentran para el desarrollo del mismo.

Información de Entrada

- Objetivos del Proyecto (D.EN.AnN.ObPr)
- Problema del Negocio (D.EN.CPN.PrNe)
- Fuentes de Información del Cliente (D.EN.AnN.FuIC)
- Recursos Humanos Involucrados (G.In.EIP.ReHI)
- Expectativas del Proyecto (D.EN.AnN.ExPr)
- Suposiciones del Proyecto (D.EN.AnN.SuPr)

Información de Salida

- Reporte de Evaluación de Herramientas (G.In.EvS.REHe)
- Reporte de Evaluación de Viabilidad (G.In.EvS.REVi)

4.3.1.3.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se proponen los reportes de Evaluación de Herramientas y Evaluación de Viabilidad, los cuales se presentan a continuación.

Reporte de Evaluación de Herramientas (G.In.EvS.REHe): en [Britos et al., 2006] se propone un formalismo para analizar y seleccionar herramientas para proyectos de explotación de información. El mismo está integrado por 4 secciones: 1) Funcional – Características técnicas, 2) Características del Proveedor, 3) Características del Servicio y 4) Características Económicas. Cada una de ellas tiene una serie de criterios ponderados, los cuales deben responderse para cada herramienta a analizar con la escala de valores de 1 a 4. En las columnas a la derecha de los criterios, se incorpora cada herramienta que se desea evaluar, y se puntúa cada elemento con la escala correspondiente (indicada en la parte superior del formalismo). En las tablas 4.11.a y 4.11.b se presenta el formalismo descripto. En los criterios que deben ser valorados por los expertos se incorporó el texto “Valor asignado al criterio”, mientras que para los valores totalizados por sección se utilizó el texto “Valor total sección X” y “Valor total ponderado sección X” para el totalizado y el ponderado respectivamente. Por último, la ponderación final de cada herramienta se señaló con el texto “Valoración final de la herramienta X”.

Reporte de Evaluación de Viabilidad (G.In.EvS.REVi): se propone el uso del formalismo, basado en la técnica de viabilidad propuesta en [Pytel et al., 2015], en el cual se registra para cada característica perteneciente a las categorías establecidas (Datos, Problema de Negocio, Proyecto y Equipo de trabajo), el valor lingüístico considerado. Cada una de las características tiene registrado su dimensión asociada (Plausibilidad, Adecuación y Éxito) y, debajo del valor a asignar, el valor de umbral. Por último, se detallan los totales por dimensión y el total global de viabilidad, junto con el resultado obtenido. En la tabla 4.12 se presenta el formalismo descripto, en el cual se utiliza la sigla VAC para indicar "Valor Asignado a la Característica".

Reporte de Evaluación de Herramientas					
Responsable:	Persona a cargo		**Fecha:**	Fecha de Realización	
ID#:	Identificador del Producto		**Versión:**	Identificación de la versión	
Criterios:					
Evaluación: 1 = Malo, 2 = débil, 3 = Bueno, 4 = Excelente				1 = No, 4 = SI	
Herramientas		**Peso**	**Herramienta 1**	**…**	**Herramienta N**
1. Funcional - Características Técnicas					
Soporte de Metodología / Ciclo de vida	Soporte del proceso	3	Valor asignado al criterio	...	Valor asignado al criterio
Compatibilidad con fuentes de datos	Base de datos	8	Valor asignado al criterio	...	Valor asignado al criterio
	Otras fuentes (word, excel, etc.)	8	Valor asignado al criterio	...	Valor asignado al criterio
Integración	Soporte de distintas técnicas asociadas al proceso de explotación de Información	5	Valor asignado al criterio	...	Valor asignado al criterio
Multilenguaje	Soporta distintas idiomas	2	Valor asignado al criterio	...	Valor asignado al criterio
Técnicas	Variedad de técnicas que provee	18	Valor asignado al criterio	...	Valor asignado al criterio
Reporte y visualización	Permite generar reportes y visualizaciones	12	Valor asignado al criterio	...	Valor asignado al criterio
Multiplataforma	Soporta múltiples plataformas	5	Valor asignado al criterio	...	Valor asignado al criterio
Instalación remota	La administración y mantenimiento son remotos	5	Valor asignado al criterio	...	Valor asignado al criterio
Usuarios Múltiples	Posee perfiles de usuarios	2	Valor asignado al criterio	...	Valor asignado al criterio
Seguridad	Provee seguridad de la información configurada por perfiles	2	Valor asignado al criterio	...	Valor asignado al criterio
Backup	Metodología de backup	2	Valor asignado al criterio	...	Valor asignado al criterio
Amigable	Interfaz de usuario	10	Valor asignado al criterio	...	Valor asignado al criterio
Configuraciones	Permite la configuración del perfil	8	Valor asignado al criterio	...	Valor asignado al criterio
Documentación	Servicio de soporte y ayuda	5	Valor asignado al criterio	...	Valor asignado al criterio
Conexión	Soporta conexión por: Internet, FTP, ERPs.	2	Valor asignado al criterio	...	Valor asignado al criterio
Soporte de sistemas de mensaje	Soporta compartir información (por mail u otro medio)	3	Valor asignado al criterio	...	Valor asignado al criterio
Total			Valor total sección 1	···	Valor total sección 1
	Peso del Grupo	40%	Valor total ponderado sección 1	...	Valor total ponderado sección 1

Tabla 4.11.a. Formalismo: Técnica de Evaluación de herramientas. Adaptado de [Britos et al., 2006].

2. Características del Proveedor					
Características del proveedor	Historia	30	Valor asignado al criterio	...	Valor asignado al criterio
Crecimiento	Perspectiva a futuro	10	Valor asignado al criterio	...	Valor asignado al criterio
Ubicación Geográfica	Oficinas	30	Valor asignado al criterio	...	Valor asignado al criterio
Implementación	Otras implementaciones de la misma herramienta	5	Valor asignado al criterio	...	Valor asignado al criterio
	Contacto con otros clientes	5	Valor asignado al criterio	...	Valor asignado al criterio
Confidencialidad	Confidencialidad de la información	20	Valor asignado al criterio	...	Valor asignado al criterio
Total			Valor total sección 2	...	Valor total sección 2
	Peso del Grupo	25%	Valor total ponderado sección 2	...	Valor total ponderado sección 2
3. Características del Servicio					
Garantía del producto	Duración y Alcance	30	Valor asignado al criterio	...	Valor asignado al criterio
Mejora	Brinda soporte a versiones previas	20	Valor asignado al criterio	...	Valor asignado al criterio
Licencia	Costo, alcances y soporte postventa	30	Valor asignado al criterio	...	Valor asignado al criterio
Soporte	Tiempo de respuesta y disponibilidad	20	Valor asignado al criterio	...	Valor asignado al criterio
Total			Valor total sección 3	...	Valor total sección 3
	Peso del Grupo	20%	Valor total ponderado sección 3	...	Valor total ponderado sección 3
4. Características Económicas					
Costo del software	Costo de la herramienta	30	Valor asignado al criterio	...	Valor asignado al criterio
Costo del Hardware	Necesidad de mejorar o comprar nuevo hardware compatible con la herramienta	20	Valor asignado al criterio	...	Valor asignado al criterio
Otros costos software	Costos adicionales al producto (backup, web servers, bases de datos, etc.)	20	Valor asignado al criterio	...	Valor asignado al criterio
Licencias	Política de licencia	10	Valor asignado al criterio	...	Valor asignado al criterio
Financiamiento	Existencia	10	Valor asignado al criterio	...	Valor asignado al criterio
Mejoras	Costo promedio de la mejora del producto	10	Valor asignado al criterio	...	Valor asignado al criterio
Total			Valor total sección 4	...	Valor total sección 4
	Peso del Grupo	-15%	Valor total ponderado sección 4	...	Valor total ponderado sección 4
Final					
1. Funcional - Características Técnicas		40%	Valor total ponderado sección 1	...	Valor total ponderado sección 1
2. Características del Proveedor		25%	Valor total ponderado sección 2	...	Valor total ponderado sección 2
3. Características del Servicio		20%	Valor total ponderado sección 3	...	Valor total ponderado sección 3
4. Características Económicas		-15%	Valor total ponderado sección 4	...	Valor total ponderado sección 4
TOTAL			**Valoración final de la herramienta 1**	...	**Valoración final de la herramienta N**

Tabla 4.11.b. Formalismo: Evaluación de herramientas. Adaptado de [Britos et al., 2006].

4.3.1.3.2. Técnica Identificada

Para el desarrollo de esta actividad se deben alcanzar dos objetivos fuertemente relacionados, la disponibilidad de herramientas a utilizar y las características que estas brindan, y la evaluación de la posibilidad de satisfacer las necesidades del cliente en tiempo y forma. Para ello, se utilizan las técnicas: **"Metodología para la selección de Herramientas de Explotación de Información"** [Britos et al., 2006] y **"Modelo de Evaluación de Viabilidad para Proyectos de Explotación de Información"** [Pytel et al., 2015], descriptas en la sección 2.4.1.1 (pág. 33) y 2.4.1.5 (pág. 39), respectivamente.

Reporte de Evaluación de Viabilidad												
Responsable:	Persona a cargo			**Fecha:**		Fecha de Realización						
ID#:	Identificador del Producto			**Versión:**		Identificación de la versión						
Datos						**Problema de Negocio**			**Proyecto**		**Equipo de Trabajo**	
P1	**P2**	**A1**	**A2**	**A3**	**E1**	**P3**	**A4**	**A5**	**E2**	**E3**	**P4**	**E4**
VAC P1	VAC P2	VAC A1	VAC A2	VAC A3	VAC E1	VAC P3	VAC A4	VAC A5	VAC E2	VAC E3	VAC P4	VAC E4
Umbral												
poco	poco	poco	poco	poco	nada	poco	poco	poco	nada	nada	poco	nada
Dimensiones						***Viabilidad global***				***Resultado***		
Plausibilidad			Total dimensión plausibilidad			***Total global***				***Resultado obtenido***		
Adecuación			Total dimensión Adecuación									
Éxito			Total dimensión Éxito									

Tabla 4.12. Formalismo: Evaluación de Viabilidad

En la primera de ellas, se analizan las distintas herramientas de explotación de información de interés con el objetivo de determinar según las necesidades del proyecto, cual brinda la mejor asistencia para el desarrollo del mismo. La evaluación de las mismas se realiza desde cuatro aspectos: técnico/funcional, del proveedor, del servicio y económico. Cada uno de ellos está conformado por una serie de características que tiene asignado un peso asociado y que deberán ser respondidas para cada una de las herramientas utilizando la escala provista, la cual varía según el tipo de respuesta: si es una respuesta de NO/SI, se usan los valores 1 y 4 respectivamente, mientras que si es una respuesta con graduaciones, se utiliza: 1 = Malo, 2 = débil, 3 = Bueno, 4 = Excelente. Mediante las valoraciones realizadas a cada característica, y utilizando las fórmulas descriptas en la propuesta, se obtienen los totales por aspecto (común y ponderado según el peso de los mismos). Finalmente, con los pesos ponderados de cada aspecto, se determina la valoración global de la herramienta. Como resultado de esta técnica se obtiene el reporte de evaluación de herramientas. La técnica se describe con mayor detalle en la sección 2.4.1.1 (pág. 33).

En la segunda técnica se realizan trece preguntas asociadas a distintas características del proyecto, puntuándose mediante una escala lingüística (nada, poco, regular, mucho, todo) asociada a intervalos difusos (cuatro puntos críticos entre los valores 0 y 10) que a partir de tres instancias de control (umbral, valoración por dimensión y valoración global) se determina la viabilidad o no del proyecto. Dichas preguntas están asociadas a cuatro categorías (o aspectos) del proyecto: datos, problema de negocio, proyecto y equipo de trabajo. Las respuestas a las preguntas asociadas a los datos, se derivan del formalismo Fuentes de Información del Cliente, Evaluación de herramientas, Suposiciones, y restricciones del Proyecto. Aquellas asociadas al problema de negocio, se obtienen a partir de los formalismos Objetivos del Proyecto y Problema del Negocio. Finalmente, las respuestas a las características proyecto y equipo de trabajo, se derivan del formalismo Recursos Humanos Involucrados.

En la sección 2.4.1.5 (pág. 39) se presenta una visión detallada de la técnica. En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo en la prueba de concepto.

4.3.1.3.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la metodología para la selección de herramientas de explotación de información y el modelo de evaluación de viabilidad para proyectos de explotación de información, la cual utiliza como insumos los formalismos: Objetivos del Proyecto (Tabla 4.57), Problema del Negocio (Tabla 4.64), Fuentes de Información del Cliente (Tabla 4.55), Recursos Humanos Involucrados (Tabla 4.6), Expectativas del Proyecto (Tabla 4.59) y Suposiciones del Proyecto (Tabla 4.60).

Los formalismos indicados como elementos de entrada, son transcriptos con el mismo número de tabla, para facilitar al lector en la comprensión de la aplicación de la técnica.

Objetivos del Proyecto			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.AnN.ObPr	**Versión:**	1.0
Objetivo	**Descripción**		**Referencia**
obpr.1	Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos		Entrevista 1

Tabla 4.57 (Transcripta). Prueba de Concepto - Objetivos del Proyecto

Problema del Negocio				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.PRNE	**Versión:**	1.0	
Objetivo del Proyecto	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos			
Problema	**Descripción**		**Experto**	**Referencia**
prne.1	Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		(rehc.1) Silva H.	Entrevista 3

Tabla 4.64 (Transcripta). Prueba de Concepto - Problema del Negocio

Fuentes de Información del Cliente					
Responsable:		Esposito E.		**Fecha:**	05/04/2016
ID#:		D.EN.ANN.FUIC		**Versión:**	1.0
ID	**Nombre**	**Categoría**	**Responsable**	**Descripción**	
fuic.1	Documento para la utilización de la base de datos usuario 2011	Documento	-	Describe distintas consideraciones acerca de la base de datos obtenida a partir de la encuesta ENPreCoSP 2011, indicando los métodos de recolección aplicados, los objetivos de la encuesta y la descripción de los datos.	
fuic.2	Cuestionario ENPreCoSP 2011	Planilla	-	Ejemplo de cuestionario ENPreCoSP 2011	
fuic.3	Base ENPreCoSP 2011	Almacén de datos	-	Almacén de registros de respuestas del cuestionario ENPreCoSP 2011 en formato txt (separado por el carácter "|") conformado por 34343 personas que respondieron 292 preguntas. El primer renglón contiene los nombres de los campos. Decimales separados por punto (.)	

Tabla 4.55 (Transcripta). Prueba de Concepto - Fuentes de Información del Cliente

Recursos Humanos Involucrados					
Responsable:		Rodriguez H.		**Fecha:**	04/04/2016
ID#:		G.In.EIP.ReHI		**Versión:**	1.0
ID	**Nombre**	**Rol/Posición**	**Pertenece a**	**Descripción**	**Información de contacto**
rehi.1	Rodriguez H.	Líder de Proyecto	recurso interno	Persona encargada de la dirección del proyecto	Skype: XXXX
rehi.2	Esposito E.	Ingeniero de Explotación de información	recurso interno	Ingeniero de Explotación de información Senior	Skype: XXXXX
rehi.3	Silva H.	Cliente	Organización Contratante	Cliente Experto en el área	Correo: xxxx@gmail.com Skype: xxxxxxx

Tabla 4.6 (Transcripta). Prueba de Concepto - Recursos Humanos Involucrados

Expectativas del Proyecto				
Responsable:	Esposito E.	**Fecha:**	15/04/2016	
ID#:	D.EN.AnN.ExPr	**Versión:**	1.0	
Expectativa	**Descripción**	**Objetivo asociado**		**Referencia**
expr.1	Definir un proceso que automatice o semi-automatice el análisis de los datos, reduciendo el costo y tiempo asociado con la generación de resultados y el correspondiente accionar a partir de los mismos, que brinde indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales serán validadas por el cliente	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos		Entrevista 1

Tabla 4.59 (Transcripta). Prueba de Concepto - Expectativas del Proyecto

Suposiciones del Proyecto			
Responsable:	Esposito E.	**Fecha:**	20/04/2016
ID#:	D.EN.AnN.SuPr	**Versión:**	1.1

Suposición	Descripción	Objetivo asociado	Referencia
supr.1	Los cuestionarios y la carga de la información se ha realizado de manera correcta	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1
supr.2	Las conductas de consumo se considerarán como análogas sin importar la gradualidad del mismo	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1
supr.3	Las variables vinculadas con autopercepción brindan información fiable respecto al entorno real del individuo	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 2
supr.4	El proceso de diseño de la muestra es representativo a nivel nacional y provincial	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 2 / fuic.1

Tabla 4.60 (Transcripta). Prueba de Concepto - Suposiciones del Proyecto

Reporte de Evaluación de Herramientas (G.In.EvS.REHe): para el desarrollo del proyecto, se identifican como posibles herramientas para utilizar de acuerdo a las necesidades del cliente y la experiencia de los miembros del proyecto: Tanagra Versión 1.4.50 , Weka Versión 3.7.11 y Orange Versión 2.7.8. A partir de ello, se puntúa cada una de las características de acuerdo a la escala de valores entre 1 y 4 de acuerdo al tipo de pregunta, señalándose con "--" aquellos aspectos que no han sido evaluados. Como resultado se identifica a la herramienta tanagra como la más adecuada para el proyecto. En las tablas 4.13.a y 4.13.b, se ilustran las valoraciones realizadas y los resultados obtenidos para cada una de las herramientas en las cuatro características generales.

Reporte de Evaluación de Viabilidad (G.In.EvS.REVi): a partir de la información recabada sobre la fuente de datos, el problema de negocio y los miembros que forman parte del mismo, se valúan las trece características a considerar, determinando las valoraciones por dimensión y global. De los resultados obtenidos, se verifica en primera instancia que las estimaciones individuales de cada característica sean superiores al umbral, y que las valoraciones de las dimensiones y global sean mayores a 5, determinando que el proyecto es viable. En la tabla 4.14 se ilustra las valoraciones realizadas y los resultados obtenidos para cada dimensión del proyecto.

Reporte de Evaluación de Herramientas					
Responsable:	Rodriguez H.			**Fecha:**	07/04/2016
ID#:	G.In.EvS.REHe			**Versión:**	1.0
Criterios:					
Evaluación: 1 = Malo, 2 = débil, 3 = Bueno, 4 = Excelente					1 = No, 4 = SI
Herramientas		**Peso**	**Tanagra V.1.4.50**	**Weka V.3.7.11**	**Orange V.2.7.8**
1. Funcional - Características Técnicas					
Soporte de Metodología / Ciclo de vida	Soporte del proceso	3	2	2	2
Compatibilidad con fuentes de datos	Base de datos	8	--	--	--
	Otras fuentes (word, excel, etc.)	8	3	2	3
Integración	Soporte de distintas técnicas asociadas al proceso de explotación de Información	5	4	4	4
Multilenguaje	Soporta distintas idiomas	2	1	1	1
Técnicas	Variedad de técnicas que provee	18	4	4	4
Reporte y visualización	Permite generar reportes y visualizaciones	12	2	2	2
Multiplataforma	Soporta múltiples plataformas	5	1	4	4
Instalación remota	La administración y mantenimiento son remotos	5	--	--	--
Usuarios Múltiples	Posee perfiles de usuarios	2	1	1	1
Seguridad	Provee seguridad de la información configurada por perfiles	2	1	1	1
Backup	Metodología de backup	2	1	1	1
Amigable	Interfaz de usuario	10	4	2	4
Configuraciones	Permite la configuración del perfil	8			
Documentación	Servicio de soporte y ayuda	5	4	1	3
Conexión	Soporta conexión por: Internet, FTP, ERPs.	2	1	1	1
Soporte de sistemas de mensaje	Soporta compartir información (por mail u otro medio)	3	1	1	1
Total			224	196	234
	Peso del Grupo	40%	89,6	78,4	93,6
2. Características del Proveedor					
Características del proveedor	Historia	30	3	3	1
Crecimiento	Perspectiva a futuro	10	2	3	2
Ubicación Geográfica	Oficinas	30	--	--	--
Implementación	Otras implementaciones de la misma herramienta	5	--	--	--
	Contacto con otros clientes	5	--	--	--
Confidencialidad	Confidencialidad de la información	20	--	--	--
Total			110	120	50
	Peso del Grupo	25%	27,5	30	12,5

Tabla 4.13.a. Prueba de Concepto – Reporte de Evaluación de herramientas

3. Características del Servicio					
Garantía del producto	Duración y Alcance	30	--	--	--
Mejora	Brinda soporte a versiones previas	20	1	1	1
Licencia	Costo, alcances y soporte postventa	30	--	--	--
Soporte	Tiempo de respuesta y disponibilidad	20	--	--	--
Total			20	20	20
	Peso del Grupo	20%	4	4	4
4. Características Económicas					
Costo del software	Costo de la herramienta	30	--	--	--
Costo del Hardware	Necesidad de mejorar o comprar nuevo hardware compatible con la herramienta	20	--	--	--
Otros costos software	Costos adicionales al producto (backup, web servers, bases de datos, etc.)	20	--	--	--
Licencias	Política de licencia	10	--	--	--
Financiamiento	Existencia	10	--	--	--
Mejoras	Costo promedio de la mejora del producto	10	--	--	--
Total			0	0	0
	Peso del Grupo	-15%	0	0	0
Final					
1. Funcional - Características Técnicas		40%	89,6	78,4	93,6
2. Características del Proveedor		25%	27,5	30	12,5
3. Características del Servicio		20%	4	4	4
4. Características Económicas		-15%	0	0	0
TOTAL			**121,1**	**112,4**	**110,1**

Tabla 4.13.b. Prueba de Concepto – Reporte de Evaluación de herramientas

Reporte de Evaluación de Viabilidad													
Responsable:	Rodriguez H.							**Fecha:**			07/04/2016		
ID#:	G.In.EvS.REVi							**Versión:**			1.0		
Datos							**Problema de Negocio**			**Proyecto**		**Equipo de Trabajo**	
P1	**P2**	**A1**	**A2**	**A3**	**E1**		**P3**	**A4**	**A5**	**E2**	**E3**	**P4**	**E4**
regular	mucho	mucho	mucho	mucho	regular		regular	regular	mucho	mucho	mucho	mucho	poco
Umbral													
poco	poco	poco	poco	poco	nada		poco	poco	poco	nada	nada	poco	nada

Dimensiones		***Viabilidad global***	***Resultado***
Plausibilidad	6,00	***6,00***	***Viable***
Adecuación	6,53		
Éxito	5,31		

Tabla 4.14. Prueba de Concepto - Evaluación de Viabilidad

4.3.1.4. Actividad: Definición del Ciclo de Vida (G.In.DCV)

Un modelo de ciclo de vida determina la forma mediante la cual se llevará a cabo el desarrollo de un proyecto, estableciendo el orden y los posibles flujos de acción entre las distintas etapas del proyecto. Dichas etapas pueden ser divididas por necesidades u objetivos parciales, entregables o hitos específicos para alcanzar los objetivos generales del proyecto. Todo proyecto tiene en común un principio y final definidos, mientras que las actividades y productos internos que se generan varían de un proyecto a otro, de acuerdo al modelo de ciclo de vida seleccionado y las estrategias de implementación que el mismo conlleva [Project Management Institute, Inc., 2013b].

Es en la actividad actual, en donde se analizan las características del proyecto con el objetivo de definir la estrategia de implementación más adecuada para su desarrollo. Del resultado de esta actividad se establece la estructura y el flujo de ejecución de las fases y actividades, así como la iteración entre las mismas (en caso que fuese necesario).

Información de Entrada

- Objetivos del Proyecto (D.EN.AnN.ObPr)
- Expectativas del Proyecto (D.EN.AnN.ExPr)
- Problema del Negocio (D.EN.CPN.PrNe)
- Riesgos del Proyecto (G.In.EIP.RiPr)
- Recursos Humanos Involucrados (G.In.EIP.ReHI)

Información de Salida

- Modelo de Ciclo de Vida (G.In.DCV.MoCV)

4.3.1.4.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se propone la plantilla Modelo de Ciclo de Vida, la cual se presenta a continuación.

Modelo de Ciclo de Vida (G.In.DCV.MoCV): Se formaliza la estructura y flujo de las fases y actividades del proyecto, mediante la descripción de la estrategia o ciclo de vida a utilizar para el desarrollo del mismo, utilizando un diagrama en el cual queda expresado el orden y las posibles iteraciones entre las fases del proyecto, junto con los criterios para las transiciones. La tabla 4.15 ilustra la estructura del formalismo previamente descripto.

Modelo de Ciclo de Vida			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
<Diagrama del Ciclo de Vida>			
Criterios de transición: Descripción de las reglas para pasar a la siguiente fase o iteración.			

Tabla 4.15. Formalismo: Modelo de Ciclo de Vida

4.3.1.4.2. Técnica Identificada

Para el desarrollo de esta actividad se propone la técnica **"Selección del Ciclo de Vida"**, mediante la cual se determina el orden de las fases del proceso, las interacciones o flujos posibles y los criterios de transición entre fases y/o iteraciones. Para ello, se evalúan las características del proyecto: los objetivos, necesidades y expectativas del cliente, cuán bien se comprenden las problemáticas y la posibilidad de resolverlas mediante explotación de información, el soporte y disponibilidad de los clientes/expertos de la organización contratante, la cultura de la organización desarrolladora y la experiencia de sus miembros.

A partir de las características identificadas, se evalúa y selecciona el ciclo de vida que mejor se adapte a dichas necesidades. A continuación se listan tres propuestas presentadas en la sección 2.4.1.8 (pág. 43):

- **DMLC**: presenta una visión secuencial de las actividades, con un control de mejora continua del proceso durante la finalización de cada iteración. Este modelo es ideal para proyectos en los cuales las necesidades y alcances del proyecto son conocidas y definidas en los inicios del proyecto, o se posee experiencia en proyectos similares.
- **ASD-BI**: brinda una visión flexible del proceso, centrada en las personas. Este modelo es ideal para entornos cambiantes en los cuales no se conocen con claridad las necesidades y alcances del proyecto en sus etapas tempranas, y se posee experiencia en proyectos similares.
- **Espiral**: presenta un enfoque dirigido por el riesgo para el análisis y estructuración del proceso, mediante una estrategia de desarrollo del proyecto evolutiva. Dada las características de la propuesta, se recomienda para proyectos complejos, en los cuales se identifiquen riesgos asociados con la viabilidad del proyecto, el presupuesto y/o los plazos, y no se posee experiencia en proyectos similares.

En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo para la prueba de concepto.

4.3.1.4.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica Selección del Ciclo de Vida. Esta utiliza como insumos: Objetivos del Proyecto (Tabla 4.57), Expectativas del Proyecto (Tabla 4.59), Problema del Negocio (Tabla 4.64), Riesgos del Proyecto (Tabla 4.7) y Recursos Humanos Involucrados (Tabla 4.6).

Los formalismos indicados como elementos de entrada, son transcriptos con el mismo número de referencia, para facilitar al lector en la comprensión de la aplicación de la técnica.

Objetivos del Proyecto			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.AnN.ObPr	**Versión:**	1.0
Objetivo	**Descripción**		**Referencia**
obpr.1	Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos		Entrevista 1

Tabla 4.57 (Transcripta). Prueba de Concepto - Objetivos del Proyecto

Expectativas del Proyecto			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.AnN.ExPr	**Versión:**	1.0

Expectativa	**Descripción**	**Objetivo asociado**	**Referencia**
expr.1	Definir un proceso que automatice o semi-automatice el análisis de los datos, reduciendo el costo y tiempo asociado con la generación de resultados y el correspondiente accionar a partir de los mismos, que brinde indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales serán validadas por el cliente	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1

Tabla 4.59 (Transcripta). Prueba de Concepto - Expectativas del Proyecto

Problema del Negocio			
Responsable:	Esposito E.	**Fecha:**	20/04/2016
ID#:	D.EN.CPN.PRNE	**Versión:**	1.0
Objetivo del Proyecto	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos		

Problema	**Descripción**	**Experto**	**Referencia**
prne.1	Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo	(rehc.1) Silva H.	Entrevista 3

Tabla 4.64 (Transcripta). Prueba de Concepto - Problema del Negocio

Riesgos del Proyecto			
Responsable:	Rodriguez H.	**Fecha:**	05/04/2016
ID#:	G.In.EIP.RiPr	**Versión:**	1.0
Riesgo	**Descripción**	**Alcance**	**Referencia**
risk.1	No contar con la presencia del experto en etapas críticas de necesidad de interacción con el mismo	proyecto	

Tabla 4.7 (Transcripta). Prueba de Concepto - Riesgos del Proyecto

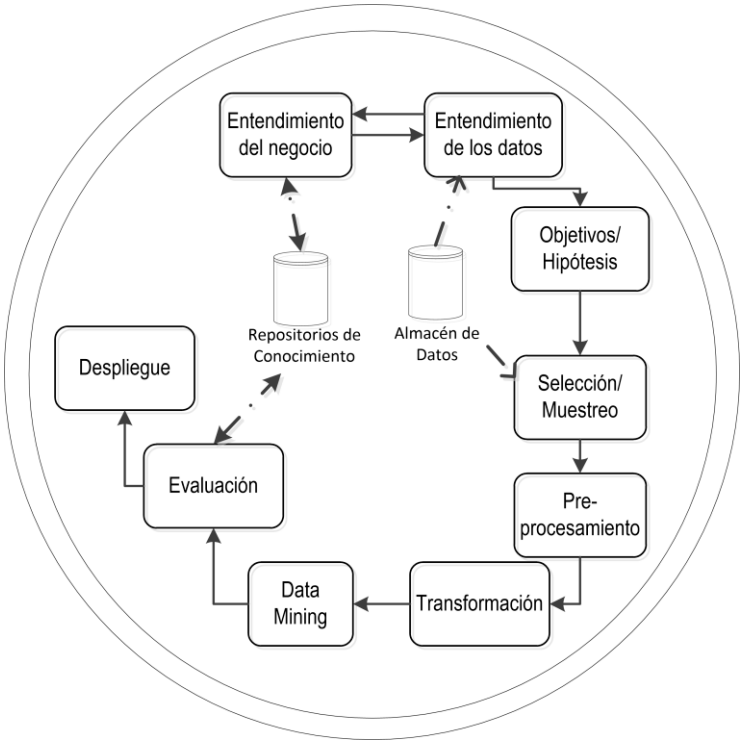
Recursos Humanos Involucrados					
Responsable:	Rodriguez H.			**Fecha:**	04/04/2016
ID#:	G.In.EIP.ReHI			**Versión:**	1.0
ID	**Nombre**	**Rol/Posición**	**Pertenece a**	**Descripción**	**Información de contacto**
rehi.1	Rodriguez H.	Líder de Proyecto	recurso interno	Persona encargada de la dirección del proyecto	Skype: XXXX
rehi.2	Esposito E.	Ingeniero de Explotación de información	recurso interno	Ingeniero de Explotación de información Senior	Skype: XXXXX
rehi.3	Silva H.	Cliente	Organización Contratante	Cliente Experto en el área	Correo: xxxx@gmail.com Skype: xxxxxxx

Tabla 4.6 (Transcripta). Prueba de Concepto - Recursos Humanos Involucrados

Modelo de Ciclo de Vida (G.In.DCV.MoCV): de acuerdo al análisis de las necesidades del cliente y la comprensión de los alcances de la explotación de información para satisfacer los problemas de la organización, conociéndose y comprendiéndose las problemáticas a abordar y los resultados esperados, junto con el conocimiento del equipo sobre las características del proyecto, habiendo trabajado en objetivos similares, y considerando los desafíos y riesgos identificados en el proyecto son de baja criticidad, se determina que un ciclo de vida predictivo (en el cual el alcance, tiempo y presupuesto es definido en una etapa temprano o inicial del proyecto), con un flujo lineal de actividades es el que mejor se adapta al proyecto, utilizando el ciclo de vida DMLC (sección 2.4.1.8, pág. 43), basado en el definido en CRISP-DM, debido a que el equipo conoce ampliamente dicha dinámica de trabajo, y las necesidades y alcances del proyecto son conocidos. Se establece como criterio de transición, el faltante del 10% en el tiempo de finalización de la fase.

La tabla 4.16 ilustra el formalismo generado, presentando la visualización del modelo de ciclo de vida y su criterio de transición.

Modelo de Ciclo de Vida			
Responsable:	Esposito E.	**Fecha:**	22/04/2016
ID#:	G.In.DCV.MoCV	**Versión:**	1.0



Criterios de transición:
Las fases posteriores podrán comenzarse 10% antes de la finalización de la misma (fecha estimada de finalización de las tareas involucradas).

Tabla 4.16. Prueba de Concepto - Modelo de Ciclo de Vida

4.3.2. Fase: Planificación (G.Pl)

En la fase Planificación se define el curso de acciones requeridos para alcanzar los objetivos del proyecto. En este se prevén la necesidad de recursos en el tiempo, en base a los objetivos y actividades a realizar, y finaliza con la propuesta del proyecto. Debido a la compleja naturaleza de los proyectos, donde es posible que ocurran cambios en los objetivos o ampliaciones a partir de la profundización sobre las necesidades del mismo, puede requerirse la reevaluación y/o ajuste de las planificaciones realizadas.

En este contexto la fase planificación se encuentra conformada por cuatro actividades: Planificación de la Mediciones (sección 4.3.2.1), Planificación de las Actividades (sección 4.3.2.2), Planificación de los Recursos (sección 4.3.2.3) y Planificación de las Responsabilidades (sección 4.3.2.4). La figura 4.4, resume las actividades que integran la fase y sus elementos de entrada y salida (las imágenes de cada formalismo son representaciones miniatura de los mismos).

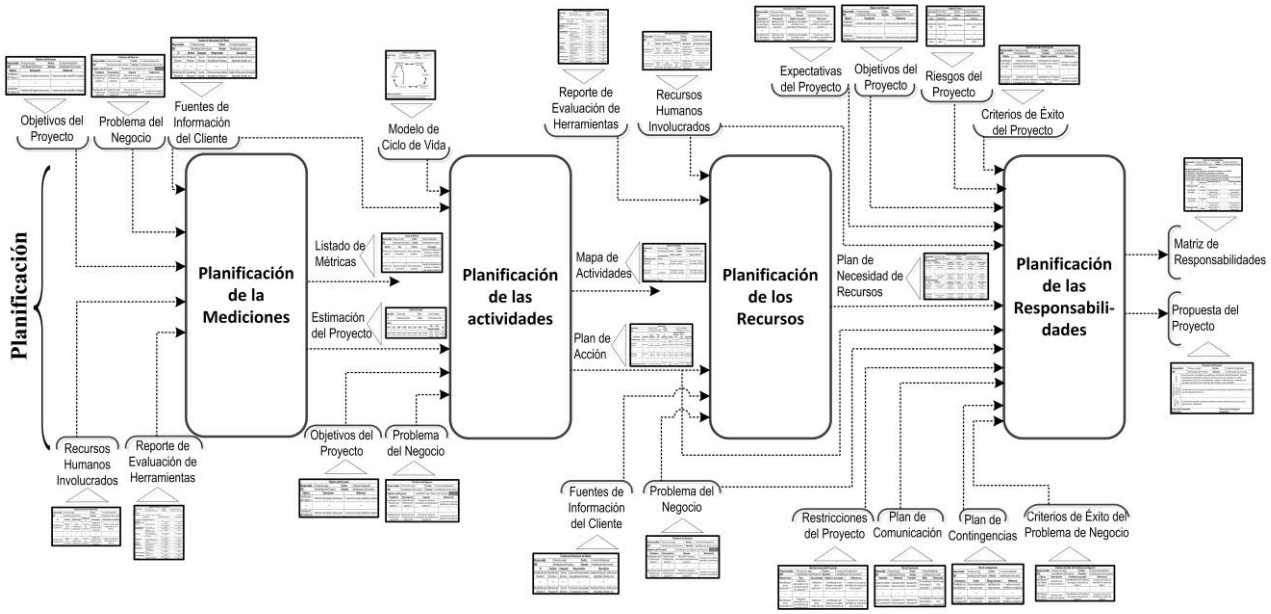


Figura 4.4. Fase Planificación

4.3.2.1. Actividad: Planificación de la Mediciones (G.Pl.PlM)

En esta actividad se realiza una estimación inicial del tiempo requerido para el desarrollo del programa del proyecto y se definen las mediciones que se llevarán a cabo durante el transcurso del mismo.

Información de Entrada

- Objetivos del Proyecto (D.EN.AnN.ObPr)
- Problema del Negocio (D.EN.CPN.PrNe)
- Fuentes de Información del Cliente (D.EN.AnN.FuIC)
- Recursos Humanos Involucrados (G.In.EIP.ReHI)
- Reporte de Evaluación de Herramientas (G.In.EvS.EvHe)

Información de Salida

- Listado de Métricas (G.Pl.PlM.LiMe)
- Estimación del Proyecto (G.Pl.PlM.EsPr)

4.3.2.1.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se proponen el Listado de Métricas y Estimación del Proyecto, las cuales se presentan a continuación.

Listado de Métricas (G.Pl.PlA.LiMe): se formaliza el conjunto de métricas a calcular durante el desarrollo del proyecto, registrando su nombre junto con la clasificación del tipo de aspecto (detallados en la sección 2.4.1.7, pág. 42) del proyecto que cubren (en las columnas homónimas), el método o procedimiento mediante el cual se calculan y cualquier detalle adicional necesario para

comprender la misma (registrándose en las columnas "Fórmula" y "Comentarios", respectivamente). La tabla 4.17 ilustra la estructura del formalismo previamente descripto.

Listado de Métricas			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Nombre	**Tipo**	**Fórmula**	**Comentarios**
Nombre de la métrica 1	Elemento asociado a la métrica 1	Método para calcular la métrica 1	Descripción detallada de los elementos a considerar para calcular la métrica 1
…	…	…	…
Nombre de la métrica N	Elemento asociado a la métrica N	Método para calcular la métrica N	Descripción detallada de los elementos a considerar para calcular la métrica N

Tabla 4.17. Formalismo: Listado de Métricas

Estimación del Proyecto (G.Pl.PlA.EsPr): se formaliza el esfuerzo requerido para el desarrollo del proyecto, basado en la propuesta de Modelo de Estimación para Proyectos de Explotación de Información publicada en [Pytel et al., 2015], la cual se describe brevemente en la sección próxima y en detalle en la sección 2.4.1.6 (pág. 42). En este, se registran las valoraciones para cada uno de los ocho factores de costo vinculados con los tres aspectos cubiertos por la propuesta: necesidades del proyecto, datos y recursos disponibles. Cada uno de los factores se completa con un valor de la escala correspondiente, obteniendo el esfuerzo total estimado para el proyecto en meses/hombre. La tabla 4.18 ilustra la estructura del formalismo.

Estimación del Proyecto										
Responsable:	Persona a cargo					**Fecha:**	Fecha de Realización			
ID#:	Identificador del Producto					**Versión:**	Identificación de la versión			
Esfuerzo										
OBTY	**LECO**	**AREP**	**QTUM**	**QTUA**	**KLDS**	**KEXT**	**TOOL**	**Total Desarrollo**	**Total Gestión**	**Total**
Valoración de OBTY	Valoración de LECO	Valoración de AREP	Valoración de QTUM	Valoración de QTUA	Valoración de KLDS	Valoración de KEXT	Valoración de TOOL	Esfuerzo estimado Desarrollo	Esfuerzo estimado Gestión	Esfuerzo estimado del proyecto

Tabla 4.18. Formalismo: Estimación del Proyecto

4.3.2.1.2. Técnicas Identificadas

Para el desarrollo de esta actividad se utilizan las propuestas de **"Métricas para Proyectos de Explotación de Información"** [Basso et al., 2013] y **"Modelo de Estimación para Proyectos de Explotación de Información"** [Pytel et al., 2015]. En la primera (descripta en detalle en la sección 2.4.1.7, pág. 42), se proponen distintas métricas centradas en tres aspectos: datos, modelos y

proyectos, las cuales permiten medir el avance del producto en cada etapa de su desarrollo y calidad del mismo. A partir de las características del proyecto (objetivos y problemas de negocio), los recursos y las necesidades del equipo de trabajo, se selecciona el conjunto de métricas a utilizar, dejando de manera clara y precisa la forma mediante la cual se calculará cada una de ellas.

En la segunda (detallada en la sección 2.4.1.6, pág. 42), se establecen ocho factores a evaluar con distintas escalas (1 a 4, 1 a 5 y 1 a 6), a partir de los cuales se determina el esfuerzo estimado (Fórmula 2.3) en meses/hombre que requerirá la ejecución del subproceso Desarrollo. Estos factores de costo abarcan tres categorías (o aspectos): alcances del *Proyecto*, los *Datos* y el *Equipo de Trabajo*. Las preguntas asociadas al proyecto, se derivan de los formalismos: Objetivos del Proyecto, Problema de Negocio y Recursos Humanos Involucrados, aquellas asociadas a la segunda categoría mediante Fuentes de Información del Cliente, y la tercera a partir de Recursos Humanos Involucrados y Reporte de Evaluación de Herramientas.

Debido a que el modelo de estimación utilizado, fue generado a partir de la experiencia recabada de 34 proyectos cuya metodología utilizada no consideraba los aspectos de gestión del proyecto, se ve necesario ajustar los resultados obtenidos para contemplar este aspecto. Dado la ausencia de un modelo que contemple estos aspectos para proyectos de explotación de información, y por consiguiente de estimaciones al respecto, se utiliza como parámetro los criterios definidos en [Mochal, T., 2006] para proyectos en general: incrementar un 15% el tiempo asociado al proceso orientado al producto. En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo para la prueba de concepto.

4.3.2.1.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar las técnicas Métricas para Proyectos de Explotación de Información y Modelo de Estimación para Proyectos de Explotación de Información, las cuales utilizan como insumo los siguientes formalismos: Objetivos del Proyecto (Tabla 4.57), Problema del Negocio (Tabla 4.64), Fuentes de Información del Cliente (Tabla 4.55), Recursos Humanos Involucrados (Tabla 4.6) y Reporte de Evaluación de Herramientas (Tabla 4.13).

Los formalismos indicados como elementos de entrada, son transcriptos con el mismo número de referencia, para facilitar la comprensión de las técnicas.

Objetivos del Proyecto			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.AnN.ObPr	**Versión:**	1.0
Objetivo	**Descripción**		**Referencia**
obpr.1	Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos		Entrevista 1

Tabla 4.57 (Transcripta). Prueba de Concepto - Objetivos del Proyecto

Problema del Negocio				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.PRNE	**Versión:**	1.0	
Objetivo del Proyecto	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos			
Problema	**Descripción**		**Experto**	**Referencia**
prne.1	Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		(rehc.1) Silva H.	Entrevista 3

Tabla 4.64 (Transcripta). Prueba de Concepto - Problema del Negocio

Fuentes de Información del Cliente					
Responsable:		Esposito E.		**Fecha:**	05/04/2016
ID#:		D.EN.ANN.FUIC		**Versión:**	1.0
ID	**Nombre**	**Categoría**	**Responsable**	**Descripción**	
fuic.1	Documento para la utilización de la base de datos usuario 2011	Documento	-	Describe distintas consideraciones acerca de la base de datos obtenida a partir de la encuesta ENPreCoSP 2011, indicando los métodos de recolección aplicados, los objetivos de la encuesta y la descripción de los datos.	
fuic.2	Cuestionario ENPreCoSP 2011	Planilla	-	Ejemplo de cuestionario ENPreCoSP 2011	
fuic.3	Base ENPreCoSP 2011	Almacén de datos	-	Almacén de registros de respuestas del cuestionario ENPreCoSP 2011 en formato txt (separado por el carácter "|") conformado por 34343 personas que respondieron 292 preguntas. El primer renglón contiene los nombres de los campos. Decimales separados por punto (.)	

Tabla 4.55 (Transcripta). Prueba de Concepto - Fuentes de Información del Cliente

Recursos Humanos Involucrados					
Responsable:	Rodriguez H.			**Fecha:**	04/04/2016
ID#:	G.In.EIP.ReHI			**Versión:**	1.0
ID	**Nombre**	**Rol/Posición**	**Pertenece a**	**Descripción**	**Información de contacto**
rehi.1	Rodriguez H.	Líder de Proyecto	recurso interno	Persona encargada de la dirección del proyecto	Skype: XXXX
rehi.2	Esposito E.	Ingeniero de Explotación de información	recurso interno	Ingeniero de Explotación de información Senior	Skype: XXXXX
rehi.3	Silva H.	Cliente	Organización Contratante	Cliente Experto en el área	Correo: xxxx@gmail.com Skype: xxxxxxx

Tabla 4.6 (Transcripta). Prueba de Concepto - Recursos Humanos Involucrados

Reporte de Evaluación de Herramientas					
Responsable:	Rodriguez H.			**Fecha:**	07/04/2016
ID#:	G.In.EvS.REHe			**Versión:**	1.0
Criterios:					
Evaluación: 1 = Malo, 2 = débil, 3 = Bueno, 4 = Excelente					1 = No, 4 = SI
Herramientas			**Tanagra V.1.4.50**	**Weka V.3.7.11**	**Orange V.2.7.8**
1. Funcional - Características Técnicas					
Soporte de Metodología / Ciclo de vida	Soporte del proceso	3	2	2	2
Compatibilidad con fuentes de datos	Base de datos	8	--	--	--
	Otras fuentes (word, excel, etc.)	8	3	2	3
Integración	Soporte de distintas técnicas asociadas al proceso de explotación de Información	5	4	4	4
Multilenguaje	Soporta distintas idiomas	2	1	1	1
Técnicas	Variedad de técnicas que provee	18	4	4	4
Reporte y visualización	Permite generar reportes y visualizaciones	12	2	2	2
Multiplataforma	Soporta múltiples plataformas	5	1	4	4
Instalación remota	La administración y mantenimiento son remotos	5	--	--	--
Usuarios Múltiples	Posee perfiles de usuarios	2	1	1	1
Seguridad	Provee seguridad de la información configurada por perfiles	2	1	1	1
Backup	Metodología de backup	2	1	1	1
Amigable	Interfaz de usuario	10	4	2	4
Configuraciones	Permite la configuración del perfil	8			
Documentación	Servicio de soporte y ayuda	5	4	1	3
Conexión	Soporta conexión por: Internet, FTP, ERPs.	2	1	1	1
Soporte de sistemas de mensaje	Soporta compartir información (por mail u otro medio)	3	1	1	1
Total			224	196	234
	Peso del Grupo	40%	89,6	78,4	93,6
2. Características del Proveedor					
Características del proveedor	Historia	30	3	3	1
Crecimiento	Perspectiva a futuro	10	2	3	2
Ubicación Geográfica	Oficinas	30	--	--	--
Implementación	Otras implementaciones de la misma herramienta	5	--	--	--
	Contacto con otros clientes	5	--	--	--
Confidencialidad	Confidencialidad de la información	20	--	--	--
Total			110	120	50
	Peso del Grupo	25%	27,5	30	12,5

Tabla 4.13.a (Transcripta). Prueba de Concepto – Reporte de Evaluación de herramientas

3. Características del Servicio					
Garantía del producto	Duración y Alcance	30	--	--	--
Mejora	Brinda soporte a versiones previas	20	1	1	1
Licencia	Costo, alcances y soporte postventa	30	--	--	--
Soporte	Tiempo de respuesta y disponibilidad	20	--	--	--
Total			20	20	20
	Peso del Grupo	20%	4	4	4
4. Características Económicas					
Costo del software	Costo de la herramienta	30	--	--	--
Costo del Hardware	Necesidad de mejorar o comprar nuevo hardware compatible con la herramienta	20	--	--	--
Otros costos software	Costos adicionales al producto (backup, web servers, bases de datos, etc.)	20	--	--	--
Licencias	Política de licencia	10	--	--	--
Financiamiento	Existencia	10	--	--	--
Mejoras	Costo promedio de la mejora del producto	10	--	--	--
Total			0	0	0
	Peso del Grupo	-15%	0	0	0
Final					
1. Funcional - Características Técnicas		40%	89,6	78,4	93,6
2. Características del Proveedor		25%	27,5	30	12,5
3. Características del Servicio		20%	4	4	4
4. Características Económicas		-15%	0	0	0
TOTAL			**121,1**	**112,4**	**110,1**

Tabla 4.13.b (Transcripta). Prueba de Concepto – Reporte de Evaluación de herramientas

Listado de Métricas (G.Pl.PlA.LiMe): para este proyecto se propone el uso de dos métricas, una enfocada en el proyecto y la otra en los datos. La primera de ellas denominada "Tiempo total requerido para el desarrollo del proyecto", se determina mediante la sumatoria de los tiempos reales del proyecto, mientras que la segunda denominada "Grado de Utilidad de Atributos", se calcula a partir de cuatro valores: número de atributos útiles sin errores [NASE (T)], número de atributos útiles con defectos [NAUD (T)], número de atributos no correctos [NANC (T)] y número de atributos no significativos [NANS (T)]. La tabla 4.19 ilustra el formalismo resultante para la prueba de concepto.

Estimación del Proyecto (G.Pl.PlA.EsPr): de acuerdo a las características del proyecto, se evalúan y fijan los factores de acuerdo a las escalas establecidas en la técnica. A partir de los formalismos objetivos del proyecto y problema de negocio, se identifica que las necesidades del cliente están asociadas con la comprensión y caracterización del comportamiento de una clase conocida, asignando el valor "1" al factor OBTY. A partir del formalismo fuente de información del

Listado de Métricas			
Responsable:	Rodriguez H.	**Fecha:**	25/04/16
ID#:	G.PI.PIA.LiMe	**Versión:**	1.0

Nombre	Tipo	Fórmula	Comentarios
Tiempo total requerido para el desarrollo del proyecto	Proyecto	DRPY = $\sum$ trA trA = tiempo requerido por actividad	Sumatoria de los tiempos requeridos para cada actividad del proyecto
Grado de Utilidad de Atributos	Datos	$GUA = \frac{NA(T) - (NO_{UTILES(T)} + 0{,}5 * NAUD(T)}{NA(T)}$ **NA(T)** = NASE(T) + NAUD(T) + NANC(T) + NANS(T) **NO _ UTILES(T)** = NANC(T) + NANS(T)	− Nro. de atributos útiles sin errores [NASE (T)] − Nro. de atributos útiles con defectos [NAUD (T)] − Nro. de atributos no correctos [NANC (T)] − Nro. de atributos no significativos [NANS (T)]

Tabla 4.19. Prueba de Concepto - Listado de Métricas

cliente, se identifica que solo se posee un repositorio de datos (asignando el valor "1" al factor AREP), la cantidad de tuplas existentes es de 34343 (asignando el valor "4" al factor QTUM), no existen tablas auxiliares (asignando el valor "1" al factor QTUA) y se posee un registro detallado de cada columna de la fuente de datos, sus significados y un ejemplar del elemento utilizado para recabar la información (asignando el valor "1" al factor KLDS). De acuerdo a los interesados del proyecto identificados (Recursos Humanos Involucrados), se definen las variables LECO y KEXT, con el valor "1". Finalmente, del Reporte de Evaluación de Herramientas se determina las funcionalidades que la herramienta seleccionada brinda, asignando al factor TOOL, el valor "3".

A partir de los valores asignados, se obtiene como esfuerzo total del subproceso desarrollo: 1.98 meses/hombre y de acuerdo al porcentaje estimado para tareas de gestión (15%), se determina como esfuerzo para dicha fase: 0.3 meses/hombre, siendo el esfuerzo total del proyecto igual a: 2.28 meses/hombre. La tabla 4.20 ilustra las valoraciones realizadas y el resultado obtenido.

Estimación del Proyecto										
Responsable:	Rodriguez H.				**Fecha:**			26/04/16		
ID#:	G.PI.PIM.EsPr				**Versión:**			1.0		
Esfuerzo										
OBTY	**LECO**	**AREP**	**QTUM**	**QTUA**	**KLDS**	**KEXT**	**TOOL**	**Total Desarrollo**	**Total Gestión**	**Total**
1	1	1	4	1	1	1	3	1,98	0,30	**2,28**

Tabla 4.20. Prueba de Concepto - Estimación del Proyecto

4.3.2.2. Actividad: Planificación de las Actividades (G.Pl.PlA)

En esta actividad se prevén las acciones a realizar durante el transcurso del proyecto y sus alcances, definiendo la ejecución de las actividades en el transcurso el tiempo. Como resultado de esta tarea, se define el programa de actividades del proyecto.

Información de Entrada

- Modelo de Ciclo de Vida (G.In.DCV.MoCV)
- Estimación del Proyecto (G.Pl.PlA.EsPr)
- Objetivos del Proyecto (D.EN.AnN.ObPr)
- Problema del Negocio (D.EN.CPN.PrNe)
- Fuentes de Información del Cliente (D.EN.AnN.FuIC)

Información de Salida

- Mapa de Actividades (G.Pl.PlA.MaAc)
- Plan de Acción (G.Pl.PlA.PlAc)

4.3.2.2.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se utilizan el Mapa de Actividades y el Plan de Acción, los cuales se presentan a continuación.

Mapa de Actividades (G.Pl.PlA.MaAc): En este formalismo, se listan en las filas las actividades a realizar (incluyendo las fases a las que pertenecen a modo de referencia en la estructura del proyecto) y en las columnas las etapas (e iteraciones) del modelo de ciclo de vida seleccionado (MCV). En los cuadrantes de intersección entre las actividades y las etapas del MCV, se indica con una equis (“X”), si la actividad se debe realizar durante dicha fase. El desarrollo completo de la actividad puede requerir más de una etapa. La tabla 4.21 ilustra la estructura del formalismo previamente descripto.

Mapa de Actividades				
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización	
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión	
ID	**Fase/Actividad**	**Etapa 1 del MCV**	**…**	**Etapa N del MCV**
Identificador fase 1	**Fase 1**			
Identificador actividad 1	Actividad 1	Actividad 1 a realizar en la etapa 1 del MCV	…	Actividad 1 a realizar en la etapa N del MCV
…	…	…	…	…
Identificador actividad N	Actividad N	Actividad N a realizar en la etapa 1 del MCV	…	Actividad N a realizar en la etapa N del MCV
…	…	…	…	…
Identificador fase N	**Fase N**			

Tabla 4.21. Formalismo: Mapa de Actividades

Plan de Acción (G.Pl.PlA.PlAc): en este formalismo se deja registro de las fechas estimadas de inicio (FEI) y fin (FEF) de cada actividad (y sus correspondientes fases), así como las fechas reales de inicio (FRI) y fin (FRF). Adicionalmente, se indica el esfuerzo estimado para cada actividad en horas (EE) y el real (ER) y se detalla en el campo comentarios cualquier información relevante respecto a la ejecución de las actividades. De forma complementaria, con el objetivo de facilitar la visualización y comprensión del progreso del proyecto, puede utilizarse el diagrama Gantt (sección 2.4.2.2, pág. 46) como herramienta complementaria. La tabla 4.22 ilustra la estructura del plan de acción y la figura 4.5 el diagrama Gantt.

Plan de Acción								
Responsable:		Persona a cargo		**Fecha:**		Fecha de Realización		
ID#:		Identificador del Producto		**Versión:**		Identificación de la versión		
ID Actividad	**Actividad**	**Inicio Estimado**	**Inicio Real**	**Fin Estimado**	**Fin Real**	**Esfuerzo Estimado (Hs)**	**Esfuerzo Real (Hs)**	**Comentarios**
Identificador fase 1	**Fase 1**	FEI fase 1	FRI fase 1	FEF fase 1	FRF fase 1	EE fase 1	ER fase 1	Detalles de la fase 1
Identificador actividad 1	Actividad 1	FEI Actividad 1	FRI Actividad 1	FEF Actividad 1	FRF Actividad 1	EE Actividad 1	ER Actividad 1	Detalles de la Actividad 1
…	…	…	…	…	…	…	…	…
Identificador actividad N	Actividad N	FEI Actividad N	FRI Actividad N	FEF Actividad N	FRF Actividad N	EE Actividad N	ER Actividad N	Detalles de la Actividad N
…	…	…	…	…	…	…	…	…
Identificador fase N	**Fase N**	FEI fase N	FRI fase N	FEF fase N	FRF fase N	EE fase N	ER fase N	Detalles de la fase N

Tabla 4.22. Formalismo: Plan de Acción

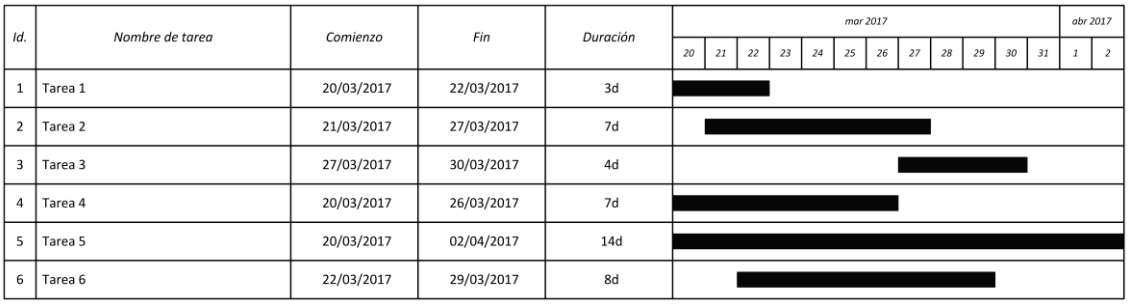


Figura 4.5. Formalismo: Plan de Acción - Diagrama de Gantt

4.3.2.2.2. Técnica Identificada

Para esta actividad se utiliza la técnica **"Definición del Programa del Proyecto"**, que a partir del modelo de ciclo de vida seleccionado, el tiempo estimado de duración del proyecto, las fuentes de información disponibles y las necesidades del cliente (objetivos y problemas de negocio), se

determina la serie de acciones a realizar para cada una de las etapas del proceso y sus alcances, así como su implementación en el tiempo.

En primera instancia, se determinan los alcances del proyecto de acuerdo a la estrategia de desarrollo seleccionada, esto es si el proyecto se desarrollará de manera lineal abarcando todos los intereses del cliente de forma completa en cada una de las fases, o si se establece alguna estrategia incremental o progresiva. A partir de ello, se identifican las actividades a realizar en cada una de las etapas del ciclo de vida seleccionado y en las iteraciones de los mismos (en caso que fuese necesario). Como resultado de este primer paso, se obtiene una visualización completa del desarrollo de las actividades durante el proyecto, la cual queda registrada en el Mapa de Actividades.

Una vez definida la estructura del proyecto, se procede a determinar de acuerdo a las valoraciones globales obtenidas (en la actividad Planificación de la Mediciones), las estimaciones empíricas de carga de trabajo realizadas en [Rodríguez et al., 2010], las características y necesidades del proyecto y los datos, el tiempo requerido para cada uno de los elementos del proceso (subprocesos, fases y actividades) según las iteraciones y ejecuciones identificadas en el mapa de actividades, debiendo quedar definido las fechas de inicio y finalización estimadas para cada elemento, y el esfuerzo estimado a dedicar durante dicho periodo, así como la descripción de cualquier detalle a considerar para el desarrollo de la actividad durante el rango de fechas estimado (por ejemplo: alcances de la actividad, dependencia de otras actividades para su implementación, entre otras), conformando el Plan de Acción del proyecto.

Es relevante destacar, que el plan de acción es un documento dinámico el cual es actualizado reiteradas veces durante el transcurso del proyecto, registrándose las fechas y esfuerzos reales aplicados a cada uno de los elementos del proyecto. Esta información permitirá durante su desarrollo y una vez finalizado el mismo, analizar desvíos y posibles acciones a realizar (o a mejorar).

En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo para la prueba de concepto.

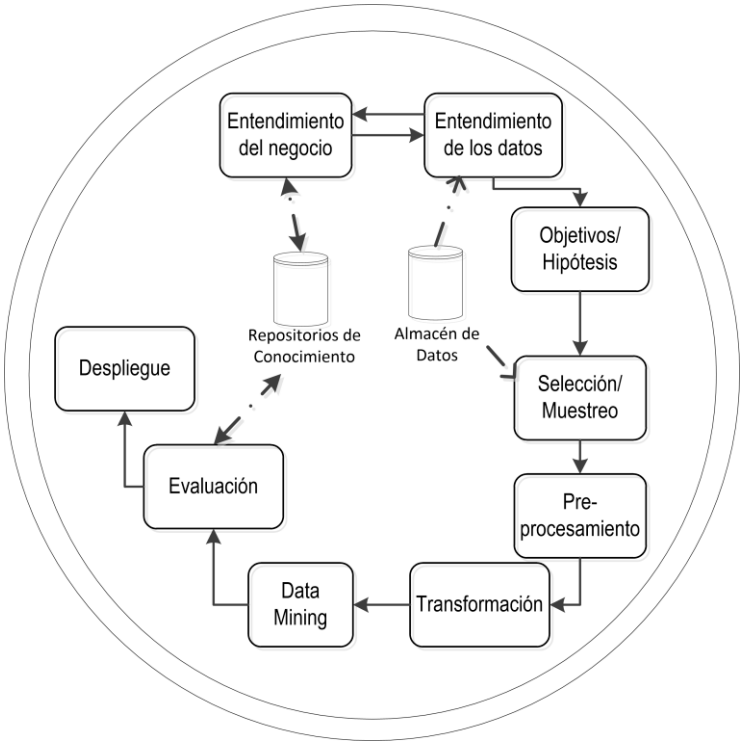
4.3.2.2.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica Definición del Programa del Proyecto. La técnica utiliza como insumos los formalismos: Modelo de Ciclo de Vida (Tabla

4.16), Estimación del Proyecto (Tabla 4.20), Objetivos del Proyecto (Tabla 4.57), Problema del Negocio (Tabla 4.64) y Fuentes de Información del Cliente (Tabla 4.55)

Los formalismos indicados como elementos de entrada, son transcriptos con el mismo número de tabla, para facilitar al lector en la comprensión de la aplicación de la técnica.

Modelo de Ciclo de Vida			
Responsable:	Esposito E.	**Fecha:**	22/04/2016
ID#:	G.In.DCV.MoCV	**Versión:**	1.0



Criterios de transición:
Las fases posteriores podrán comenzarse 10% antes de la finalización de la misma (fecha estimada de finalización de las tareas involucradas).

Tabla 4.16 (Transcripta). Prueba de Concepto - Modelo de Ciclo de Vida

Estimación del Proyecto										
Responsable:	Rodriguez H.			**Fecha:**				26/04/16		
ID#:	G.PI.PIM.EsPr			**Versión:**				1.0		
Esfuerzo										
OBTY	**LECO**	**AREP**	**QTUM**	**QTUA**	**KLDS**	**KEXT**	**TOOL**	**Total Desarrollo**	**Total Gestión**	**Total**
1	1	1	4	1	1	1	3	1,98	0,30	**2,28**

Tabla 4.20 (Transcripta). Prueba de Concepto - Estimación del Proyecto

Objetivos del Proyecto			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.AnN.ObPr	**Versión:**	1.0
Objetivo	**Descripción**		**Referencia**
obpr.1	Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos		Entrevista 1

Tabla 4.57 (Transcripta). Prueba de Concepto - Objetivos del Proyecto

Problema del Negocio			
Responsable:	Esposito E.	**Fecha:**	20/04/2016
ID#:	D.EN.CPN.PRNE	**Versión:**	1.0
Objetivo del Proyecto	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos		

Problema	Descripción	Experto	Referencia
prne.1	Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo	(rehc.1) Silva H.	Entrevista 3

Tabla 4.64 (Transcripta). Prueba de Concepto - Problema del Negocio

Fuentes de Información del Cliente					
Responsable:		Esposito E.		**Fecha:**	05/04/2016
ID#:		D.EN.ANN.FUIC		**Versión:**	1.0
ID	**Nombre**	**Categoría**	**Responsable**	**Descripción**	
fuic.1	Documento para la utilización de la base de datos usuario 2011	Documento	-	Describe distintas consideraciones acerca de la base de datos obtenida a partir de la encuesta ENPreCoSP 2011, indicando los métodos de recolección aplicados, los objetivos de la encuesta y la descripción de los datos.	
fuic.2	Cuestionario ENPreCoSP 2011	Planilla	-	Ejemplo de cuestionario ENPreCoSP 2011	
fuic.3	Base ENPreCoSP 2011	Almacén de datos	-	Almacén de registros de respuestas del cuestionario ENPreCoSP 2011 en formato txt (separado por el carácter "|") conformado por 34343 personas que respondieron 292 preguntas. El primer renglón contiene los nombres de los campos. Decimales separados por punto (.)	

Tabla 4.55 (Transcripta). Prueba de Concepto - Fuentes de Información del Cliente

Mapa de Actividades (G.Pl.PlA.MaAc): A partir del modelo de ciclo de vida seleccionado, se determina las etapas durante las cuales se desarrollarán las distintas actividades del proceso propuesto, ajustado a las necesidades propias del proyecto. La tabla 4.23, presentan la distribución de las actividades a través de las fases que integran el modelo de ciclo de vida, como puede observarse las actividades pertenecientes a la fase de soporte y control, se realizan de manera transversal al desarrollo del proyecto.

Mapa de Actividades										
Responsable:		Rodriguez H.				**Fecha:**		22/04/16		
ID#:		G.PI.PIA.MaAc				**Versión:**		1.0		
ID	**Fase/Actividad**	**E.N.**	**E.D.**	**H.**	**S.**	**Pre.P.**	**T.**	**D.M.**	**E.**	**D.**
G.In	**Iniciación**									
G.In.EIP	Exploración Inicial del Proyecto	x								
G.In.DeC	Definición de la Comunicación	x								
G.In.EvS	Evaluación de la Situación	x								
G.In.DCV	Definición del Ciclo de Vida	x								
G.PI	**Planificación**									
G.PI.PIM	Planificación de la Mediciones	x	x							
G.PI.PIA	Planificación de las Actividades	x	x							
G.PI.PIR	Planificación de los Recursos	x	x							
G.PI.PRe	Planificación de las Responsabilidades	x	x							
G.So	**Soporte**									
G.So.MeP	Mediciones del Proyecto	x	x	x	x	x	x	x	x	x
G.So.GeC	Gestión de la Configuración	x	x	x	x	x	x	x	x	x
G.Co	**Control**									
G.Co.GeD	Gestión del Desarrollo	x	x	x	x	x	x	x	x	x
G.Co.CoA	Control de las Actividades	x	x	x	x	x	x	x	x	x
G.Co.Gca	Gestión del Cambio	x	x	x	x	x	x	x	x	x
G.Ci	**Cierre**									
G.Ci.FEC	Formalización Externa del Cierre del Proyecto									x
G.Ci.FIC	Formalización Interna del Cierre del Proyecto									x
D.EN	**Entendimiento del Negocio**									
D.EN.AnN	Análisis del Negocio	x								
D.EN.CPN	Comprensión del Problema de Negocio	x								
D.ED	**Entendimiento de los Datos**									
D.ED.AnD	Análisis de los Datos		x							
D.ED.ExD	Exploración de los Datos		x							
D.ED.EvD	Evaluación de los Datos		x							
D.Mo	**Modelado**									
D.Mo.MoP	Modelado del problema			x						
D.PD	**Preparación de los Datos**			x						
D.PD.CFT	Construcción de la Fuente Temporaria de Datos				x	x				
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos					x	x			
D.Im	**Implementación**									
D.Im.SeM	Selección del Modelo							x		
D.Im.Exl	Explotación de Información							x		
D.EP	**Evaluación y Presentación**									
D.EP.EvR	Evaluación de los Resultados								x	x
D.EP.PrR	Presentación de los Resultados									x

Tabla 4.23. Prueba de Concepto - Mapa de Actividades

Plan de Acción (G.Pl.PlA.PlAc): a partir de la estimación de tiempos y la selección de las actividades a realizar en cada etapa del modelo de ciclo de vida (mapa de actividades), se asigna la duración y rango de fechas de ejecución de cada una de las actividades usando de base las mediciones de esfuerzo requeridas para proyectos de explotación de información [Rodríguez et al., 2010]. La asignación final del tiempo asociado al esfuerzo estimado para cada subproceso, fue determinado teniendo en cuenta la participación de la cantidad de personas en cada subproceso. Por ejemplo, para la actividad "Exploración Inicial del Proyecto" se prevé la ejecución de la misma durante el periodo del 04/04/2016 y 18/04/16, estimando un esfuerzo total estimado y dedicado a dicha tarea de cuatro horas.

Como fue mencionado previamente, el plan de acción se mantiene ajustado durante el desarrollo del proyecto, registrándose en los hitos de control y reporte de estado, los avances del proyecto (tiempos y fechas reales). El resultado obtenido al final del proyecto, se presenta en las tablas 4.24.a y 4.24.b, siendo relevante aclarar que dicho formalismo fue ajustado en tres instancias: la inicial (versión 1.0) y las asociadas a los reporte de estado G.Co.GeD.ReEs.1y G.Co.GeD.ReEs.2 (tablas 4.38 y 4.39, respectivamente), las cuales se presentan en las tablas A.1 y A.2 (sección A.1.1). Adicionalmente, en la figura A.1, se ilustra el diagrama de Gantt actualizado al cierre del proyecto.

Plan de Acción								
Responsable:		Rodriguez H.			**Fecha:**		15/06/16	
ID#:		G.Pl.PlA.PlAc			**Versión:**		1.3	
ID Actividad	**Actividad**	**Inicio Estimado**	**Inicio Real**	**Fin Estimado**	**Fin Real**	**Esfuerzo Estimado (Hs)**	**Esfuerzo Real (Hs)**	**Comentarios**
G.In	**Iniciación**	**04/04/16**	**04/04/16**	**22/04/16**	**22/04/16**	**12**	**11**	
G.In.EIP	Exploración Inicial del Proyecto	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DeC	Definición de la Comunicación	04/04/16	04/04/16	18/04/16	18/04/16	2	2	
G.In.EvS	Evaluación de la Situación	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DCV	Definición del Ciclo de Vida	19/04/16	19/04/16	22/04/16	22/04/16	2	1	
G.Pl	Planificación	04/04/16	04/04/16	29/04/16	29/04/16	10	9	
G.Pl.PlM	Planificación de la Mediciones	04/04/16	04/04/16	29/04/16	29/04/16	2	1	
G.Pl.PlA	Planificación de las Actividades	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.Pl.PlR	Planificación de los Recursos	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.Pl.PRe	Planificación de las Responsabilidades	04/04/16	04/04/16	29/04/16	29/04/16	4	4	
G.So	Soporte	20/04/16	20/04/16	15/06/16	15/06/16	8	12	
G.So.MeP	Mediciones del Proyecto	20/04/16	20/04/16	15/06/16	15/06/16	4	6	Se prevé el registro de las métricas de manera mensual
G.So.GeC	Gestión de la Configuración	20/04/16	20/04/16	15/06/16	15/06/16	4	6	

Tabla 4.24.a. Prueba de Concepto - Plan de Acción (fin del proyecto)

G.Co	**Control**	**20/04/16**	**20/04/16**	**13/06/16**	**13/06/16**	**12**	**15**	
G.Co.GeD	Gestión del Desarrollo	20/04/16	20/04/16	13/06/16	13/06/16	4	6	Se prevé la aplicación del reporte de estado de manera mensual
G.Co.CoA	Control de las Actividades	20/04/16	20/04/16	13/06/16	13/06/16	6	8	
G.Co.Gca	Gestión del Cambio	20/04/16	20/04/16	13/06/16	13/06/16	2	1	
G.Ci	**Cierre**	**06/06/16**	**06/06/16**	**15/06/16**	**15/06/16**	**4**	**4**	
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	06/06/16	06/06/16	13/06/16	13/06/16	2	2	
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	14/06/16	14/06/16	15/06/16	15/06/16	2	2	
D.EN	**Entendimiento del Negocio**	**04/04/16**	**04/04/16**	**22/04/16**	**20/04/16**	**44**	**32**	
D.EN.AnN	Análisis del Negocio	04/04/16	04/04/16	20/04/16	20/04/16	28	22	
D.EN.CPN	Comprensión del Problema de Negocio	18/04/16	18/04/16	22/04/16	20/04/16	16	10	
D.ED	**Entendimiento de los Datos**	**25/04/16**	**25/04/16**	**06/05/16**	**06/05/16**	**56**	**52**	
D.ED.AnD	Análisis de los Datos	25/04/16	25/04/16	29/04/16	29/04/16	22	20	
D.ED.ExD	Exploración de los Datos	28/04/16	28/04/16	06/05/16	06/05/16	22	22	
D.ED.EvD	Evaluación de los Datos	02/05/16	02/05/16	06/05/16	06/05/16	12	10	
D.Mo	**Modelado**	**09/05/16**	**09/05/16**	**13/05/16**	**13/05/16**	**26**	**16**	
D.Mo.MoP	Modelado del problema	09/05/16	09/05/16	10/05/16	10/05/16	14	8	
D.Mo.CoM	Configuración del Modelo	11/05/16	11/05/16	13/05/16	13/05/16	12	8	
D.PD	**Preparación de los Datos**	**16/05/16**	**16/05/16**	**20/05/16**	**20/05/16**	**36**	**28**	
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	16/05/16	16/05/16	18/05/16	18/05/16	20	16	
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos	17/05/16	17/05/16	20/05/16	20/05/16	16	12	
D.Im	**Implementación**	**23/05/16**	**23/05/16**	**27/05/16**	**27/05/16**	**30**	**26**	
D.Im.SeM	Selección del Modelo	23/05/16	23/05/16	24/05/16	24/05/16	8	6	
D.Im.ExI	Explotación de Información	24/05/16	24/05/16	27/05/16	27/05/16	22	20	
D.EP	**Evaluación y Presentación**	**30/05/16**	**30/05/16**	**13/06/16**	**13/06/16**	**26**	**18**	
D.EP.EvR	Evaluación de los Resultados	30/05/16	30/05/16	02/06/16	02/06/16	10	6	
D.EP.PrR	Presentación de los Resultados	01/06/16	01/06/16	13/06/16	13/06/16	16	12	

Tabla 4.24.b. Prueba de Concepto - Plan de Acción (fin del proyecto)

4.3.2.3. Actividad: Planificación de los Recursos (G.Pl.PlR)

En esta actividad se prevén los recursos (tanto humanos como materiales) necesarios para el desarrollo de las actividades en el tiempo.

Información de Entrada

- Recursos Humanos Involucrados (G.In.EIP.ReHI)
- Reporte de Evaluación de Herramientas (G.In.EvS.EvHe)
- Plan de Acción (G.Pl.PlA.PlAc)
- Problema del Negocio (D.EN.CPN.PrNe)
- Fuentes de Información del Cliente (D.EN.AnN.FuIC)

Información de Salida

- Plan de Necesidad de Recursos (G.Pl.PlR.PlNR)

4.3.2.3.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se propone el Plan de Necesidad de Recursos, el cual se presenta a continuación.

Plan de Necesidad de Recursos (G.Pl.PlR.PlNR): se registran los recursos requeridos humanos y materiales (en las secciones correspondientes), indicando su tipo, junto con un identificador específico para distinguir entre recursos similares que tengan distintas fechas requeridas, la cantidad de recursos y la fecha de inicio y fin en el que son requeridos, y una descripción detallada de los mismos. La tabla 4.25 ilustra la estructura del formalismo previamente descripto.

4.3.2.3.2. Técnica Identificada

Para el desarrollo de esta actividad se propone la técnica **"Planificación de los Recursos Necesarios"**, que a partir de las necesidades del cliente, las herramientas seleccionadas, los recursos vigentes, el programa del proyecto y la experiencia del equipo de trabajo, se determinan los recursos humanos y materiales que se requerirán en el transcurso del tiempo para el proyecto. En este contexto, se debe identificar el tipo de recurso y las cantidades requeridas para los rangos de tiempo, detallando cualquier información relevante del mismo (como por ejemplo para recursos humanos: experiencia en el dominio del negocio, conocimientos sobre ciertas herramientas y/o tecnologías, entre otros y para recursos materiales: capacidad de procesamiento, memoria, etc.).

Plan de Necesidad de Recursos					
Responsable:	Persona a cargo		**Fecha:**	Fecha de Realización	
ID#:	Identificador del Producto		**Versión:**	Identificación de la versión	
Recursos Humanos					
ID	**Recurso**	**Cantidad**	**F. Inicio**	**F. Fin**	**Descripción**
Identificador del RH 1	Tipo de recurso requerido	Cantidad de recursos requeridos	Fecha desde que se lo requiere	Fecha hasta que se lo requiere	Detalles adicionales del recurso
…	…	…	…	…	…
Identificador del RH N	Tipo de recurso requerido	Cantidad de recursos requeridos	Fecha desde que se lo requiere	Fecha hasta que se lo requiere	Detalles adicionales del recurso
Recursos Materiales					
ID	**Recurso**	**Cantidad**	**F. Inicio**	**F. Fin**	**Descripción**
Identificador del recurso material 1	Tipo de recurso requerido	Cantidad de recursos requeridos	Fecha desde que se lo requiere	Fecha hasta que se lo requiere	Detalles adicionales del recurso
…	…	…	…	…	…
Identificador del recurso material N	Tipo de recurso requerido	Cantidad de recursos requeridos	Fecha desde que se lo requiere	Fecha hasta que se lo requiere	Detalles adicionales del recurso

Tabla 4.25. Formalismo: Plan de Necesidad de Recursos

En [El Sheikh y Alnoukari, 2012; Microsoft, 2016] se listan los roles de recursos humanos diferenciados según las capacidades y responsabilidades en el proyecto. En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo para la prueba de concepto.

4.3.2.3.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se introducen los resultados obtenidos de aplicar la técnica Planificación de los Recursos Necesarios. Esta técnica utiliza como insumos los formalismos: Recursos Humanos Involucrados (Tabla 4.6), Fuentes de Información del Cliente (Tabla 4.55), Reporte de Evaluación de Herramientas (Tabla 4.13.a y 4.13.b), Plan de Acción (Tabla 4.24.a y 4.24.b) y Problema del Negocio (Tabla 4.64). Los formalismos indicados como elementos de entrada, son transcriptos con el mismo número de tabla, para facilitar al lector en la comprensión de la aplicación de la técnica.

Recursos Humanos Involucrados					
Responsable:	Rodriguez H.			**Fecha:**	04/04/2016
ID#:	G.In.EIP.ReHI			**Versión:**	1.0
ID	**Nombre**	**Rol/Posición**	**Pertenece a**	**Descripción**	**Información de contacto**
rehi.1	Rodriguez H.	Líder de Proyecto	recurso interno	Persona encargada de la dirección del proyecto	Skype: XXXX
rehi.2	Esposito E.	Ingeniero de Explotación de información	recurso interno	Ingeniero de Explotación de información Senior	Skype: XXXXX
rehi.3	Silva H.	Cliente	Organización Contratante	Cliente Experto en el área	Correo: xxxx@gmail.com Skype: xxxxxxx

Tabla 4.6 (Transcripta). Prueba de Concepto - Recursos Humanos Involucrados

Fuentes de Información del Cliente					
Responsable:		Esposito E.		**Fecha:**	05/04/2016
ID#:		D.EN.ANN.FUIC		**Versión:**	1.0
ID	**Nombre**	**Categoría**	**Responsable**	**Descripción**	
fuic.1	Documento para la utilización de la base de datos usuario 2011	Documento	-	Describe distintas consideraciones acerca de la base de datos obtenida a partir de la encuesta ENPreCoSP 2011, indicando los métodos de recolección aplicados, los objetivos de la encuesta y la descripción de los datos.	
fuic.2	Cuestionario ENPreCoSP 2011	Planilla	-	Ejemplo de cuestionario ENPreCoSP 2011	
fuic.3	Base ENPreCoSP 2011	Almacén de datos	-	Almacén de registros de respuestas del cuestionario ENPreCoSP 2011 en formato txt (separado por el carácter "|") conformado por 34343 personas que respondieron 292 preguntas. El primer renglón contiene los nombres de los campos. Decimales separados por punto (.)	

Tabla 4.55 (Transcripta). Prueba de Concepto - Fuentes de Información del Cliente

Reporte de Evaluación de Herramientas					
Responsable:	Rodriguez H.			**Fecha:**	07/04/2016
ID#:	G.In.EvS.REHe			**Versión:**	1.0
Criterios:					
Evaluación: 1 = Malo, 2 = débil, 3 = Bueno, 4 = Excelente					1 = No, 4 = SI
Herramientas			**Tanagra V.1.4.50**	**Weka V.3.7.11**	**Orange V.2.7.8**
1. Funcional - Características Técnicas					
Soporte de Metodología / Ciclo de vida	Soporte del proceso	3	2	2	2
Compatibilidad con fuentes de datos	Base de datos	8	--	--	--
	Otras fuentes (word, excel, etc.)	8	3	2	3
Integración	Soporte de distintas técnicas asociadas al proceso de explotación de Información	5	4	4	4
Multilenguaje	Soporta distintas idiomas	2	1	1	1
Técnicas	Variedad de técnicas que provee	18	4	4	4
Reporte y visualización	Permite generar reportes y visualizaciones	12	2	2	2
Multiplataforma	Soporta múltiples plataformas	5	1	4	4
Instalación remota	La administración y mantenimiento son remotos	5	--	--	--
Usuarios Múltiples	Posee perfiles de usuarios	2	1	1	1
Seguridad	Provee seguridad de la información configurada por perfiles	2	1	1	1
Backup	Metodología de backup	2	1	1	1
Amigable	Interfaz de usuario	10	4	2	4
Configuraciones	Permite la configuración del perfil	8			
Documentación	Servicio de soporte y ayuda	5	4	1	3
Conexión	Soporta conexión por: Internet, FTP, ERPs.	2	1	1	1
Soporte de sistemas de mensaje	Soporta compartir información (por mail u otro medio)	3	1	1	1
Total			224	196	234
	Peso del Grupo	40%	89,6	78,4	93,6
2. Características del Proveedor					
Características del proveedor	Historia	30	3	3	1
Crecimiento	Perspectiva a futuro	10	2	3	2
Ubicación Geográfica	Oficinas	30	--	--	--
Implementación	Otras implementaciones de la misma herramienta	5	--	--	--
	Contacto con otros clientes	5	--	--	--
Confidencialidad	Confidencialidad de la información	20	--	--	--
Total			110	120	50
	Peso del Grupo	25%	27,5	30	12,5

Tabla 4.13.a (Transcripta). Prueba de Concepto - Reporte de Evaluación de Herramientas

3. Características del Servicio					
Garantía del producto	Duración y Alcance	30	--	--	--
Mejora	Brinda soporte a versiones previas	20	1	1	1
Licencia	Costo, alcances y soporte postventa	30	--	--	--
Soporte	Tiempo de respuesta y disponibilidad	20	--	--	--
Total			20	20	20
	Peso del Grupo	20%	4	4	4
4. Características Económicas					
Costo del software	Costo de la herramienta	30	--	--	--
Costo del Hardware	Necesidad de mejorar o comprar nuevo hardware compatible con la herramienta	20	--	--	--
Otros costos software	Costos adicionales al producto (backup, web servers, bases de datos, etc.)	20	--	--	--
Licencias	Política de licencia	10	--	--	--
Financiamiento	Existencia	10	--	--	--
Mejoras	Costo promedio de la mejora del producto	10	--	--	--
Total			0	0	0
	Peso del Grupo	-15%	0	0	0
Final					
1. Funcional - Características Técnicas		40%	89,6	78,4	93,6
2. Características del Proveedor		25%	27,5	30	12,5
3. Características del Servicio		20%	4	4	4
4. Características Económicas		-15%	0	0	0
TOTAL			**121,1**	**112,4**	**110,1**

Tabla 4.13.b (Transcripta). Prueba de Concepto - Reporte de Evaluación de Herramientas

Plan de Acción								
Responsable:		Rodriguez H.			**Fecha:**		15/06/16	
ID#:		G.PI.PIA.PlAc			**Versión:**		1.3	
ID Actividad	**Actividad**	**Inicio Estimado**	**Inicio Real**	**Fin Estimado**	**Fin Real**	**Esfuerzo Estimado (Hs)**	**Esfuerzo Real (Hs)**	**Comentarios**
G.In	**Iniciación**	**04/04/16**	**04/04/16**	**22/04/16**	**22/04/16**	**12**	**11**	
G.In.EIP	Exploración Inicial del Proyecto	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DeC	Definición de la Comunicación	04/04/16	04/04/16	18/04/16	18/04/16	2	2	
G.In.EvS	Evaluación de la Situación	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DCV	Definición del Ciclo de Vida	19/04/16	19/04/16	22/04/16	22/04/16	2	1	
G.Pl	**Planificación**	**04/04/16**	**04/04/16**	**29/04/16**	**29/04/16**	**10**	**9**	
G.Pl.PIM	Planificación de la Mediciones	04/04/16	04/04/16	29/04/16	29/04/16	2	1	
G.Pl.PIA	Planificación de las Actividades	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.Pl.PIR	Planificación de los Recursos	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.Pl.PRe	Planificación de las Responsabilidades	04/04/16	04/04/16	29/04/16	29/04/16	4	4	

Tabla 4.24.a (Transcripta). Prueba de Concepto - Plan de Acción (fin del proyecto)

G.So	**Soporte**	**20/04/16**	**20/04/16**	**15/06/16**	**15/06/16**	**8**	**12**	
G.So.MeP	Mediciones del Proyecto	20/04/16	20/04/16	15/06/16	15/06/16	4	6	Se prevé el registro de las métricas de manera mensual
G.So.GeC	Gestión de la Configuración	20/04/16	20/04/16	15/06/16	15/06/16	4	6	
G.Co	**Control**	**20/04/16**	**20/04/16**	**13/06/16**	**13/06/16**	**12**	**15**	
G.Co.GeD	Gestión del Desarrollo	20/04/16	20/04/16	13/06/16	13/06/16	4	6	Se prevé la aplicación del reporte de estado de manera mensual
G.Co.CoA	Control de las Actividades	20/04/16	20/04/16	13/06/16	13/06/16	6	8	
G.Co.Gca	Gestión del Cambio	20/04/16	20/04/16	13/06/16	13/06/16	2	1	
G.Ci	**Cierre**	**06/06/16**	**06/06/16**	**15/06/16**	**15/06/16**	**4**	**4**	
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	06/06/16	06/06/16	13/06/16	13/06/16	2	2	
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	14/06/16	14/06/16	15/06/16	15/06/16	2	2	
D.EN	**Entendimiento del Negocio**	**04/04/16**	**04/04/16**	**22/04/16**	**20/04/16**	**44**	**32**	
D.EN.AnN	Análisis del Negocio	04/04/16	04/04/16	20/04/16	20/04/16	28	22	
D.EN.CPN	Comprensión del Problema de Negocio	18/04/16	18/04/16	22/04/16	20/04/16	16	10	
D.ED	**Entendimiento de los Datos**	**25/04/16**	**25/04/16**	**06/05/16**	**06/05/16**	**56**	**52**	
D.ED.AnD	Análisis de los Datos	25/04/16	25/04/16	29/04/16	29/04/16	22	20	
D.ED.ExD	Exploración de los Datos	28/04/16	28/04/16	06/05/16	06/05/16	22	22	
D.ED.EvD	Evaluación de los Datos	02/05/16	02/05/16	06/05/16	06/05/16	12	10	
D.Mo	**Modelado**	**09/05/16**	**09/05/16**	**13/05/16**	**13/05/16**	**26**	**16**	
D.Mo.MoP	Modelado del problema	09/05/16	09/05/16	10/05/16	10/05/16	14	8	
D.Mo.CoM	Configuración del Modelo	11/05/16	11/05/16	13/05/16	13/05/16	12	8	
D.PD	**Preparación de los Datos**	**16/05/16**	**16/05/16**	**20/05/16**	**20/05/16**	**36**	**28**	
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	16/05/16	16/05/16	18/05/16	18/05/16	20	16	
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos	17/05/16	17/05/16	20/05/16	20/05/16	16	12	
D.Im	**Implementación**	**23/05/16**	**23/05/16**	**27/05/16**	**27/05/16**	**30**	**26**	
D.Im.SeM	Selección del Modelo	23/05/16	23/05/16	24/05/16	24/05/16	8	6	
D.Im.ExI	Explotación de Información	24/05/16	24/05/16	27/05/16	27/05/16	22	20	
D.EP	**Evaluación y Presentación**	**30/05/16**	**30/05/16**	**13/06/16**	**13/06/16**	**26**	**18**	
D.EP.EvR	Evaluación de los Resultados	30/05/16	30/05/16	02/06/16	02/06/16	10	6	
D.EP.PrR	Presentación de los Resultados	01/06/16	01/06/16	13/06/16	13/06/16	16	12	

Tabla 4.24.b (Transcripta). Prueba de Concepto - Plan de Acción (fin del proyecto)

Problema del Negocio				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.PRNE	**Versión:**	1.0	
Objetivo del Proyecto	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos			
Problema	**Descripción**		**Experto**	**Referencia**
prne.1	Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		(rehc.1) Silva H.	Entrevista 3

Tabla 4.64 (Transcripta). Prueba de Concepto - Problema del Negocio

Plan de Necesidad de Recursos (G.Pl.PlR.PlNR): para este proyecto, se planifica la participación de dos personas: un líder del proyecto y un ingeniero de explotación de información (Junior), los cuales serán requeridos durante todo el proyecto (con tareas realizadas en concurrencia de ambos miembros para la capacitación del segundo de ellos). Adicionalmente, se detalla la necesidad de 2 computadoras para cada uno de los miembros con las siguientes características: SO Windows 7 (en adelante), RAM 4GB o más y 10GB o más espacio en disco, de acuerdo con las necesidades requeridas por la herramienta seleccionada. En la tabla 4.26 se ilustra el resultado derivado de la prueba de concepto. Se considera relevante destacar que el mismo es la versión final del proyecto, la cual fue actualizada a partir de la necesidad de cambio del plan de acción ("ReCa.1 - Ajuste de plazos de hitos del proyecto", tabla 4.43). La versión previa del formalismo puede visualizarse en la tabla A.3.

Plan de Necesidad de Recursos					
Responsable:	Rodriguez H.		**Fecha:**	03/06/2016	
ID#:	G.Pl.PIR.PINR		**Versión:**	1.1	
Recursos Humanos					
ID	**Recurso**	**Cantidad**	**F. Inicio**	**F. Fin**	**Descripción**
rhr.1	Líder de Proyecto	1	04/04/16	17/06/16	
rhr.2	Ingeniero de Explotación de Información Junior	1	04/04/16	17/06/16	
Recursos Materiales					
ID	**Recurso**	**Cantidad**	**F. Inicio**	**F. Fin**	**Descripción**
rmr.1	Computadora Personal	2	04/04/16	17/06/16	SO windows (7 en adelante) RAM 4 gb o más 10GB o más espacio en disco

Tabla 4.26. Prueba de Concepto - Plan de Necesidad de Recursos

4.3.2.4. Actividad: Planificación de las Responsabilidades (G.Pl.PRe)

En esta actividad se definen las responsabilidades y obligaciones de las partes involucradas en el proyecto, tanto entre el cliente y la organización que desarrolla el proyecto, así como entre los miembros que integran esta última. Como resultado debe formalizarse quién es el encargado de

realizar cada tarea y los compromisos asumidos por cada una de las partes intervinientes en el acuerdo.

Información de Entrada

- Recursos Humanos Involucrados (G.In.EIP.ReHI)
- Plan de Comunicación (G.In.DeC.PCom)
- Plan de Acción (G.Pl.PlA.PlAc)
- Plan de Necesidad de Recursos (G.Pl.PlR.PlNR)
- Objetivos del Proyecto (D.EN.AnN.ObPr)
- Criterios de Éxito del Proyecto (D.EN.AnN.CrEP)
- Expectativas del Proyecto (D.EN.AnN.ExPr)
- Restricciones del Proyecto (D.EN.AnN.RePr)
- Problema del Negocio (D.EN.CPN.PrNe)
- Criterios de Éxito del Problema de Negocio (D.EN.CPN.CEPN)
- Riesgos del Proyecto (G.In.EIP.RiPr)
- Plan de Contingencias (G.In.EIP.PCon)

Información de Salida

- Matriz de Responsabilidades (G.Pl.PRe.MaRe)
- Propuesta del Proyecto (G.Pl.PRe.PrPr)

4.3.2.4.1. Formalismos Identificados

La información de salida asociada con la actividad se formaliza mediante los elementos: Matriz de Responsabilidades y Propuesta del Proyecto, los cuales se presentan a continuación.

Matriz de Responsabilidades (G.Pl.PlR.MaRe): el formalismo propuesto en [Project Management Institute, Inc., 2013a] está conformado por una cabecera donde se describen los niveles de participación del proyecto, filas las cuales listan las actividades (agrupadas por fase por simplicidad) que se desarrollarán en el proyecto, y columnas que representan los distintos recursos humanos (tanto internos como externos a la organización que lleva a cabo el proyecto) en los cuales se indica por cada actividad el nivel de participación del miembro (en caso que tuviese). La tabla 4.27 ilustra la estructura del formalismo previamente descripto.

Propuesta del Proyecto (G.Pl.PlR.PrPr): se registran aquellos aspectos a acordar entre las partes para dar inicio formal al desarrollo del proyecto. El formalismo está integrado por tres secciones: en la primera de ellas, se definen los objetivos y los criterios de éxito, junto con las expectativas del cliente, el programa del proyecto, los riesgos y acciones a realizar en caso de contingencia. En la segunda, se definen las obligaciones y responsabilidades del proyecto, mientras que en la tercera se

determinan aquellos aspectos legales vinculados con el uso y divulgación de la información y los resultados derivados. La tabla 4.28 ilustra la estructura del formalismo previamente descripto.

Matriz de Responsabilidades					
Responsable:	Persona a cargo		**Fecha:**	Fecha de Realización	
ID#:	Identificador del Producto		**Versión:**	Identificación de la versión	
Descripción **Niveles de participación:** **(R) Responsable: encargado de las tareas asociadas a la actividad.** **(E) Ejecución: asignado tareas asociadas a la actividad.** **(A) Aprobación: aceptación Final del resultado de la actividad.** **(C) Consultado: posee conocimiento relevante para el desarrollo de la actividad.** **(I) Informado: requiere estar alerta del progreso de la actividad.**					
ID	**Actividad**	**Recurso Humano 1**	**…**	**Recurso Humano N**	
Identificador fase 1	**Fase 1**				
Identificador actividad 1	Actividad 1	Valor de nivel de participación	**…**	Valor de nivel de participación	
…	**…**	Valor de nivel de participación	**…**	Valor de nivel de participación	
Identificador actividad N	Actividad N				
…	**…**	Valor de nivel de participación	**…**	Valor de nivel de participación	
Identificador fase N	**Fase N**	Valor de nivel de participación	**…**	Valor de nivel de participación	

Tabla 4.27. Formalismo: Matriz de Responsabilidades

Propuesta del Proyecto			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Alcance	En esta sección se detallan los objetivos y criterios de éxito del proyecto, dejando formalmente registrado los alcances del proyecto (y que aspectos no están contemplados), así como los obstáculos, riesgos y contingencias. Finalmente, se describen los alcances de cada hito del programa o plan acordado.		
Obligaciones y responsa-bilidades	Se describen los compromisos asumidos por las partes intervinientes del proyecto, junto con las cláusulas del mismo.		
Aspectos Legales	Se describen aquellas cuestiones legales y de privacidad de la información de la organización contratante		
Firma del Contratante: **Aclaración:**		**Firma de la Contraparte:** **Aclaración:**	

Tabla 4.28. Formalismo: Propuesta del Proyecto

4.3.2.4.2. Técnica Identificada

Para el desarrollo de esta actividad se propone la técnica **"Designación de Responsabilidades"**, propuesta en [Project Management Institute, Inc., 2013a] (sección 2.4.2.3, pág. 46), mediante la

cual se definen las obligaciones y responsabilidades de las partes involucradas en el proyecto, desde dos perspectivas: individual (entre los miembros de ambas organizaciones) y general (entre la organización que llevará a cabo el proyecto y la contratante).

En la perspectiva individual, se determina a partir de las necesidades del cliente (objetivos y problema de negocio), del plan de acción del proyecto, de los interesados, y de manera vinculada con la planificación de los recursos en el tiempo, el nivel de participación de los miembros del proyecto en cada una de las actividades del mismo, teniendo en consideración sus conocimientos e intereses. Las asignaciones pueden realizarse por actividad o por tarea/entregable. Como resultado de este paso, queda diseñada la matriz de responsabilidades.

En la perspectiva global, se definen las obligaciones y responsabilidades de las organizaciones involucradas dejando constancia de los alcances del proyecto (lo que se va a realizar y lo que no será contemplado), los plazos previstos para el cumplimiento de los hitos y los compromisos que la organización cliente debe realizar como contraparte, así como aquellos aspectos legales vinculados con la información a utilizar. Para su desarrollo, se debe contemplar los conceptos definidos en el plan de acción (respecto a tiempos de desarrollo de los hitos), los objetivos, expectativas y restricciones del proyecto, junto con el problema de negocio y los criterios de éxito asociados a ambos formalismos, identificando a partir de estos, las obligaciones de las partes (alcances y limitaciones del proyecto), así como la definición de los compromisos de los interesados. Adicionalmente, los costos del proyecto pueden ser determinados a partir del esfuerzo estimado, los recursos y su necesidad durante el proyecto. Como resultado final de la tarea, se crea la Propuesta del Proyecto. En la siguiente sección se presenta la aplicación de la técnica y los formalismos derivados de implementar la prueba de concepto.

4.3.2.4.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica Designación de Responsabilidades. Dicha técnica utiliza como insumos los formalismos: Plan de Acción (Tabla 4.24), Recursos Humanos Involucrados (Tabla 4.6), Plan de Comunicación (Tabla 4.10), Plan de Necesidad de Recursos (Tabla 4.26), Objetivos del Proyecto (Tabla 4.57), Criterios de Éxito del Proyecto (Tabla 4.58), Expectativas del Proyecto (Tabla 4.59), Restricciones del Proyecto (Tabla 4.61), Problema del Negocio (Tabla 4.64), Criterios de Éxito del Problema de Negocio (Tabla 4.65), Riesgos del Proyecto (Tabla 4.7) y Plan de Contingencias (Tabla 4.8).

Los formalismos indicados como elementos de entrada, son transcriptos con el mismo número de tabla, para facilitar la comprensión de la técnica.

Plan de Acción								
Responsable:		Rodriguez H.			**Fecha:**		15/06/16	
ID#:		G.PI.PIA.PIAc			**Versión:**		1.3	
ID Actividad	**Actividad**	**Inicio Estimado**	**Inicio Real**	**Fin Estimado**	**Fin Real**	**Esfuerzo Estimado (Hs)**	**Esfuerzo Real (Hs)**	**Comentarios**
G.In	**Iniciación**	**04/04/16**	**04/04/16**	**22/04/16**	**22/04/16**	**12**	**11**	
G.In.EIP	Exploración Inicial del Proyecto	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DeC	Definición de la Comunicación	04/04/16	04/04/16	18/04/16	18/04/16	2	2	
G.In.EvS	Evaluación de la Situación	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DCV	Definición del Ciclo de Vida	19/04/16	19/04/16	22/04/16	22/04/16	2	1	
G.Pl	**Planificación**	**04/04/16**	**04/04/16**	**29/04/16**	**29/04/16**	**10**	**9**	
G.Pl.PIM	Planificación de la Mediciones	04/04/16	04/04/16	29/04/16	29/04/16	2	1	
G.Pl.PIA	Planificación de las Actividades	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.Pl.PIR	Planificación de los Recursos	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.Pl.PRe	Planificación de las Responsabilidades	04/04/16	04/04/16	29/04/16	29/04/16	4	4	
G.So	**Soporte**	**20/04/16**	**20/04/16**	**15/06/16**	**15/06/16**	**8**	**12**	
G.So.MeP	Mediciones del Proyecto	20/04/16	20/04/16	15/06/16	15/06/16	4	6	Se prevé el registro de las métricas de manera mensual
G.So.GeC	Gestión de la Configuración	20/04/16	20/04/16	15/06/16	15/06/16	4	6	
G.Co	**Control**	**20/04/16**	**20/04/16**	**13/06/16**	**13/06/16**	**12**	**15**	
G.Co.GeD	Gestión del Desarrollo	20/04/16	20/04/16	13/06/16	13/06/16	4	6	Se prevé la aplicación del reporte de estado de manera mensual
G.Co.CoA	Control de las Actividades	20/04/16	20/04/16	13/06/16	13/06/16	6	8	
G.Co.Gca	Gestión del Cambio	20/04/16	20/04/16	13/06/16	13/06/16	2	1	
G.Ci	**Cierre**	**06/06/16**	**06/06/16**	**15/06/16**	**15/06/16**	**4**	**4**	
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	06/06/16	06/06/16	13/06/16	13/06/16	2	2	
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	14/06/16	14/06/16	15/06/16	15/06/16	2	2	
D.EN	**Entendimiento del Negocio**	**04/04/16**	**04/04/16**	**22/04/16**	**20/04/16**	**44**	**32**	
D.EN.AnN	Análisis del Negocio	04/04/16	04/04/16	20/04/16	20/04/16	28	22	
D.EN.CPN	Comprensión del Problema de Negocio	18/04/16	18/04/16	22/04/16	20/04/16	16	10	

Tabla 4.24.a (Transcripta). Prueba de Concepto - Plan de Acción (fin del proyecto)

D.ED	**Entendimiento de los Datos**	**25/04/16**	**25/04/16**	**06/05/16**	**06/05/16**	**56**	**52**	
D.ED.AnD	Análisis de los Datos	25/04/16	25/04/16	29/04/16	29/04/16	22	20	
D.ED.ExD	Exploración de los Datos	28/04/16	28/04/16	06/05/16	06/05/16	22	22	
D.ED.EvD	Evaluación de los Datos	02/05/16	02/05/16	06/05/16	06/05/16	12	10	
D.Mo	**Modelado**	**09/05/16**	**09/05/16**	**13/05/16**	**13/05/16**	**26**	**16**	
D.Mo.MoP	Modelado del problema	09/05/16	09/05/16	10/05/16	10/05/16	14	8	
D.Mo.CoM	Configuración del Modelo	11/05/16	11/05/16	13/05/16	13/05/16	12	8	
D.PD	**Preparación de los Datos**	**16/05/16**	**16/05/16**	**20/05/16**	**20/05/16**	**36**	**28**	
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	16/05/16	16/05/16	18/05/16	18/05/16	20	16	
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos	17/05/16	17/05/16	20/05/16	20/05/16	16	12	
D.Im	**Implementación**	**23/05/16**	**23/05/16**	**27/05/16**	**27/05/16**	**30**	**26**	
D.Im.SeM	Selección del Modelo	23/05/16	23/05/16	24/05/16	24/05/16	8	6	
D.Im.ExI	Explotación de Información	24/05/16	24/05/16	27/05/16	27/05/16	22	20	
D.EP	**Evaluación y Presentación**	**30/05/16**	**30/05/16**	**13/06/16**	**13/06/16**	**26**	**18**	
D.EP.EvR	Evaluación de los Resultados	30/05/16	30/05/16	02/06/16	02/06/16	10	6	
D.EP.PrR	Presentación de los Resultados	01/06/16	01/06/16	13/06/16	13/06/16	16	12	

Tabla 4.24.b (Transcripta). Prueba de Concepto - Plan de Acción (fin del proyecto)

Recursos Humanos Involucrados					
Responsable:	Rodriguez H.			**Fecha:**	04/04/2016
ID#:	G.In.EIP.ReHI			**Versión:**	1.0
ID	**Nombre**	**Rol/Posición**	**Pertenece a**	**Descripción**	**Información de contacto**
rehi.1	Rodriguez H.	Líder de Proyecto	recurso interno	Persona encargada de la dirección del proyecto	Skype: XXXX
rehi.2	Esposito E.	Ingeniero de Explotación de información	recurso interno	Ingeniero de Explotación de información Senior	Skype: XXXXX
rehi.3	Silva H.	Cliente	Organización Contratante	Cliente Experto en el área	Correo: xxxx@gmail.com Skype: xxxxxxx

Tabla 4.6 (Transcripta). Prueba de Concepto - Recursos Humanos Involucrados

Plan de Comunicación				
Responsable:	Rodriguez H.		**Fecha:**	04/04/2016
ID#:	G.In.DeC.PCom		**Versión:**	1.0
Interesados	**Información**	**Frecuencia**	**Medio**	**Responsable**
(rehi.1) Rodriguez H. (rehi.2) Esposito E. (rehi.3) Silva H.	Comprensión del Proyecto	semanal durante el periodo de entendimiento del negocio	Skype	(rehi.1) Rodriguez H.
(rehi.1) Rodriguez H. (rehi.2) Esposito E. (rehi.3) Silva H.	Avances del Proyecto	mensual	Skype	(rehi.1) Rodriguez H.
(rehi.1) Rodriguez H. (rehi.2) Esposito E.	Estado del Proyecto	bisemanal	Skype	(rehi.1) Rodriguez H.

Tabla 4.10 (Transcripta). Prueba de Concepto - Plan de Comunicación

Plan de Necesidad de Recursos					
Responsable:	Rodriguez H.			**Fecha:**	03/06/2016
ID#:	G.Pl.PIR.PINR			**Versión:**	1.1
Recursos Humanos					
ID	**Recurso**	**Cantidad**	**F. Inicio**	**F. Fin**	**Descripción**
rhr.1	Líder de Proyecto	1	04/04/16	17/06/16	
rhr.2	Ingeniero de Explotación de Información Junior	1	04/04/16	17/06/16	
Recursos Materiales					
ID	**Recurso**	**Cantidad**	**F. Inicio**	**F. Fin**	**Descripción**
rmr.1	Computadora Personal	2	04/04/16	17/06/16	SO windows (7 en adelante) RAM 4 gb o más 10GB o más espacio en disco

Tabla 4.26 (Transcripta). Prueba de Concepto - Plan de Necesidad de Recursos

Objetivos del Proyecto			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.AnN.ObPr	**Versión:**	1.0
Objetivo	**Descripción**		**Referencia**
obpr.1	Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos		Entrevista 1

Tabla 4.57 (Transcripta). Prueba de Concepto - Objetivos del Proyecto

Criterios de Éxito del Proyecto			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.AnN.CrEP	**Versión:**	1.0

Criterio	**Descripción**	**Objetivo asociado**	**Referencia**
crexpr.1	obtener piezas de conocimiento que favorezcan la comprensión del comportamiento de grupos masivos de personas, brindando indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales serán validadas por el cliente	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1

Tabla 4.58 (Transcripta). Prueba de Concepto - Criterios de Éxito del Proyecto

Expectativas del Proyecto			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.AnN.ExPr	**Versión:**	1.0

Expectativa	**Descripción**	**Objetivo asociado**	**Referencia**
expr.1	Definir un proceso que automatice o semi-automatice el análisis de los datos, reduciendo el costo y tiempo asociado con la generación de resultados y el correspondiente accionar a partir de los mismos, que brinde indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales serán validadas por el cliente	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1

Tabla 4.59 (Transcripta). Prueba de Concepto - Expectativas del Proyecto

Restricciones del Proyecto				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.AnN.RePr	**Versión:**	1.0	
Restricción	**Tipo**	**Descripción**	**Objetivo asociado**	**Referencia**
repr.1	datos	Se identifica un desbalance entre la cantidad de registros que han consumido distintas Sustancias Psicoactivas y quienes no han consumido	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1
repr.2	datos	Se carece de información detallada respecto al grado o frecuencia de consumo de las sustancias psicoactivas	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1

Tabla 4.61 (Transcripta). Prueba de Concepto - Restricciones del Proyecto

Problema del Negocio				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.PRNE	**Versión:**	1.0	
Objetivo del Proyecto	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos			
Problema	**Descripción**		**Experto**	**Referencia**
prne.1	Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		(rehc.1) Silva H.	Entrevista 3

Tabla 4.64 (Transcripta). Prueba de Concepto - Problema del Negocio

Criterios de Éxito del Problema de Negocio				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.CEPN	**Versión:**	1.0	
Criterio	**Descripción**	**Problema asociado**		**Referencia**
cepn.1	Identificar aspectos que permitan comprender el comportamiento de grupos masivos de personas de acuerdo a la evaluación de Silva H. (rehi.3)	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		Entrevista 3

Tabla 4.65 (Transcripta). Prueba de Concepto - Criterios de Éxito del Problema de Negocio

Riesgos del Proyecto			
Responsable:	Rodriguez H.	**Fecha:**	05/04/2016
ID#:	G.In.EIP.RiPr	**Versión:**	1.0
Riesgo	**Descripción**	**Alcance**	**Referencia**
risk.1	No contar con la presencia del experto en etapas críticas de necesidad de interacción con el mismo	proyecto	

Tabla 4.7 (Transcripta). Prueba de Concepto - Riesgos del Proyecto

Plan de Contingencias			
Responsable:	Rodriguez H.	**Fecha:**	05/04/2016
ID#:	G.In.EIP.PCon	**Versión:**	1.0

Contingencia	Acción	Riesgo asociado	Referencia
cont.1	Ajustes en los plazos del proyecto	(risk.1) No contar con la presencia del experto en etapas críticas de necesidad de interacción con el mismo	

Tabla 4.8 (Transcripta). Prueba de Concepto - Plan de Contingencias

Matriz de Responsabilidades (G.Pl.PlR.MaRe): en la prueba de concepto se identifican 3 interesados (dos miembros del equipo y un cliente/experto) introduciendo a cada uno de ellos en una columna, y asignando el nivel de participación en cada una de las actividades de acuerdo los intereses de información y el conocimiento de los mismos. Por ejemplo, Rodriguez H. (rehi.1) en la actividad "Exploración Inicial del Proyecto" tiene asignado el nivel de participación "R" (Responsable), mientras que el Esposito E. (rehi.2) debe estar informado ("I") de las acciones y conocimientos identificados en dicha actividad. En la tabla 4.29 se ilustran las asignaciones de responsabilidades de los miembros interesados del proyecto.

Propuesta del Proyecto (G.Pl.PlR.PrPr): de acuerdo a las necesidades del proyecto identificadas en los formalismos Objetivos del Proyecto, Criterios de Éxito del Proyecto, Expectativas del Proyecto, Restricciones del Proyecto, Problema del Negocio y Criterios de Éxito del Problema de Negocio, se describe el alcance del proyecto, dejando constancia de cuáles son los aspectos que se cubrirán en su desarrollo (y cuales no serán considerados). Adicionalmente, se describen aquellos riesgos asociados con las necesidades del cliente y sus acciones asociadas (Riesgos del Proyecto y Plan de Contingencias).

De ser necesario, se incorporan aquellos entregables parciales de acuerdo a los hitos y estructura del proyecto (según lo definido en el Plan de Acción). Derivándose el siguiente contenido para la prueba de concepto: "*Se establece como objetivo del proyecto analizar la población de encuestados de Encuesta Nacional sobre Prevalencias de Consumo de Sustancias Psicoactivas del 2011, con el fin de comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos. Se espera definir un proceso que automatice o semi-automatice el análisis de los datos, reduciendo el costo y tiempo asociado con la generación de resultados y el correspondiente accionar a partir de los mismos, que brinde indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales serán validadas por el cliente. Específicamente, se requiere identificar las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la*

Matriz de Responsabilidades			
Responsable:	Rodriguez H.	**Fecha:**	25/04/2016
ID#:	G.PI.PIR.MaRe	**Versión:**	1.0

Descripción

Niveles de participación:
(R) Responsable: encargado de las tareas asociadas a la actividad.
(E) Ejecución: asignado tareas asociadas a la actividad.
(A) Aprobación: aceptación Final del resultado de la actividad.
(C) Consultado: posee conocimiento relevante para el desarrollo de la actividad.
(I) Informado: requiere estar alerta del progreso de la actividad.

ID Actividad	Actividad	Rodriguez H. (rehi.1)	Esposito E. (rehi.2)	Silva H. (rehi.3)
G.In	**Iniciación**			
G.In.EIP	Exploración Inicial del Proyecto	R	I	
G.In.DeC	Definición de la Comunicación	R	I	I
G.In.EvS	Evaluación de la Situación	R	C	I
G.In.DCV	Definición del Ciclo de Vida	R	I	
G.PI	**Planificación**			
G.PI.PIM	Planificación de la Mediciones	R	I	
G.PI.PIA	Planificación de las Actividades	R	I	
G.PI.PIR	Planificación de los Recursos	R	I	
G.PI.PRe	Planificación de las Responsabilidades	R	I	A
G.So	**Soporte**			
G.So.MeP	Mediciones del Proyecto	R	I	
G.So.GeC	Gestión de la Configuración	R	I	
G.Co	**Control**			
G.Co.GeD	Gestión del Desarrollo	R	I	I
G.Co.CoA	Control de las Actividades	R	I	I
G.Co.Gca	Gestión del Cambio	A	R	I
G.Ci	**Cierre**			
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	R		A
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	R	C	
D.EN	**Entendimiento del Negocio**			
D.EN.AnN	Análisis del Negocio	E	R	C
D.EN.CPN	Comprensión del Problema de Negocio	E	R	C
D.ED	**Entendimiento de los Datos**			
D.ED.AnD	Análisis de los Datos	E	R	C
D.ED.ExD	Exploración de los Datos	E	R	
D.ED.EvD	Evaluación de los Datos	E	R	C
D.Mo	**Modelado**			
D.Mo.MoP	Modelado del problema	E	R	I
D.Mo.CoM	Configuración del Modelo	C/A	R	I
D.PD	**Preparación de los Datos**			
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	I	R	
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos		R	
D.Im	**Implementación**			
D.Im.SeM	Selección del Modelo	C	R	
D.Im.Exl	Explotación de Información	C	R	I
D.EP	**Evaluación y Presentación**			
D.EP.EvR	Evaluación de los Resultados	R	E	C
D.EP.PrR	Presentación de los Resultados	R	E	I

Tabla 4.29. Prueba de Concepto - Matriz de Responsabilidades

autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo."

En la sección de Obligaciones y Responsabilidades, se determinan los compromisos asumidos por las partes intervinientes (plazo, costo, recursos), los cuales se derivan de los formalismos Matriz de Responsabilidades, Plan de Comunicación y el Plan de Acción. De los contenidos generados en las etapas previas del proyecto, se genera el siguiente contenido: "*La parte contratante se compromete a brindar disposición de todos los recursos requeridos en tiempo y forma, informando con una antelación no menor a 72hs en caso que estos no pudiesen ser entregados. Cualquier demora fuera del plazo estipulado, requerirá del ajuste de los plazos del proyecto. La contraparte se compromete a dar solución a las problemáticas requeridas por el cliente (ver sección Alcance), así como la veracidad de los resultados presentados. Asimismo, la organización se compromete a informar de forma mensual los avances del proyecto. Las partes acuerdan como fecha de finalización del proyecto el 06/06/2016.*"

No se identifican aspectos vinculados con el uso de la información en los formalismos Restricciones del Proyecto, los Riesgos del Proyecto y Plan de Contingencias identificadas, siendo la sección "Aspectos Legales" omitida. Por último, se registra la conformidad de las partes interesadas. La tabla 4.30 ilustra el resultado derivado de aplicar la técnica en la prueba de concepto.

<table>
<tr><th colspan="4">Propuesta del Proyecto</th></tr>
<tr><td colspan="2">Responsable:</td><td>Rodriguez H.</td><td>Fecha:</td><td>29/04/2016</td></tr>
<tr><td colspan="2">ID#:</td><td>G.PI.PIR.PrPr</td><td>Versión:</td><td>1.0</td></tr>
<tr><td>Alcance</td><td colspan="4">Se establece como objetivo del proyecto analizar la población de encuestados de Encuesta Nacional sobre Prevalencias de Consumo de Sustancias Psicoactivas del 2011, con el fin de comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos. Se espera definir un proceso que automatice o semi-automatice el análisis de los datos, reduciendo el costo y tiempo asociado con la generación de resultados y el correspondiente accionar a partir de los mismos, que brinde indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales serán validadas por el cliente. Específicamente, se requiere identificar las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo.</td></tr>
<tr><td>Obligaciones y responsabilidades</td><td colspan="4">La parte contratante se compromete a brindar disposición de todos los recursos requeridos en tiempo y forma, informando con una antelación no menor a 72hs en caso que estos no pudiesen ser entregados. Cualquier demora fuera del plazo estipulado, requerirá del ajuste de los plazos del proyecto.
La contraparte se compromete a dar solución a las problemáticas requeridas por el cliente (ver sección Alcance), así como la veracidad de los resultados presentados. Asimismo, la organización se compromete a informar de forma mensual los avances del proyecto.

Las partes acuerdan como fecha de finalización del proyecto el 06/06/2016.</td></tr>
<tr><td colspan="3">Firma del Contratante: Silva H.</td><td colspan="2">Firma de la Contraparte: Rodriguez H.</td></tr>
<tr><td colspan="3">Aclaración: Silva H.</td><td colspan="2">Aclaración: Rodriguez H.</td></tr>
</table>

Tabla 4.30. Prueba de Concepto - propuesta del Proyecto

4.3.3. Fase: Soporte (G.So)

Durante la fase Soporte se realiza el seguimiento del proyecto dejando registro formal del estado actual del mismo, así como sus estadios previos. Las actividades de esta fase se realizan de manera transversal al ciclo de vida y sirven de apoyo para los miembros del proyecto durante el desarrollo del mismo. En este contexto, la fase Soporte se encuentra conformada por dos actividades: Mediciones del Proyecto (sección 4.3.3.1) y Gestión de la Configuración (sección 4.3.3.2). La figura 4.6, resume las actividades que integran la fase y sus elementos de entrada y salida (las imágenes de cada formalismo son representaciones miniatura de los mismos).

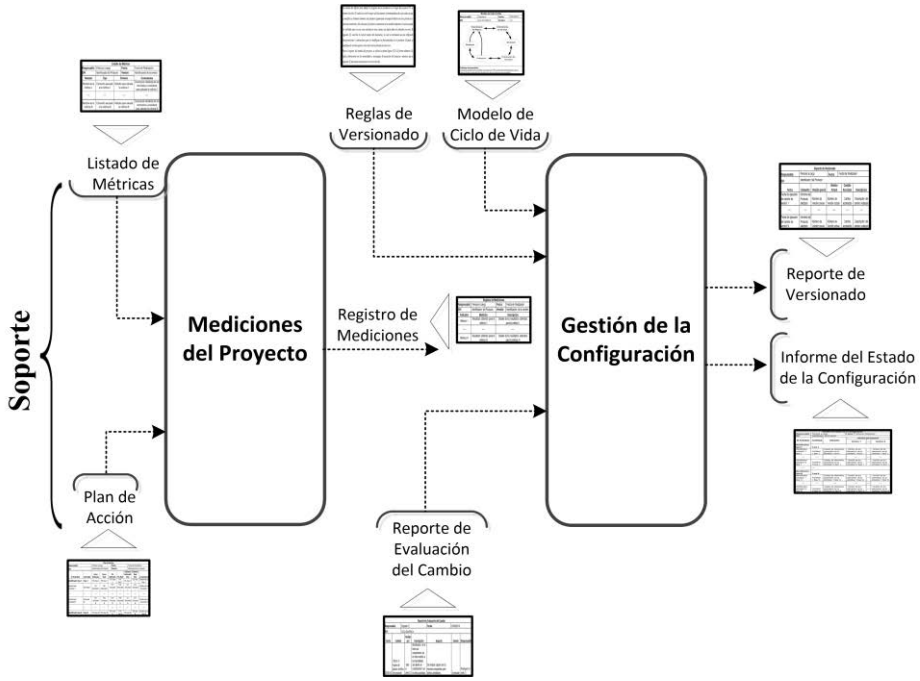


Figura 4.6. Fase Soporte

4.3.3.1. Actividad: Mediciones del Proyecto (G.So.MeP)

En esta actividad se calculan las métricas durante el desarrollo del proyecto, dejando registro formal del progreso de los indicadores. El resultado de esta actividad contribuye en la toma de decisiones del líder del proyecto, así como en la evaluación de la calidad del proceso y/o del producto.

Información de Entrada

- Listado de Métricas (G.Pl.PlA.LiMe)
- Plan de Acción (G.Pl.PlA.PlAc)

Información de Salida

- Registro de Mediciones (G.So.MeP.ReMe)

4.3.3.1.1. Formalismo Identificado

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se propone el Registro de Mediciones, el cual se presenta a continuación.

Registro de Mediciones (G.So.MeP.ReMe): se asienta formalmente las mediciones realizadas para cada una de las métricas seleccionadas, registrando la métrica (columna "Indicador"), el resultado obtenido hasta la fecha (columna "Medición") y la descripción de los resultados obtenidos (en la columna homónima). La tabla 4.31 ilustra la estructura del formalismo previamente descripto.

Registro de Mediciones			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Indicador	**Medición**	**Descripción**	
Métrica 1	Resultado obtenido para la métrica 1	Detalle de los resultados obtenidos para la métrica 1	
…	…	…	
Métrica N	Resultado obtenido para la métrica N	Detalle de los resultados obtenidos para la métrica N	

Tabla 4.31. Formalismo: Registro de Mediciones

4.3.3.1.2. Técnica Identificada

Para el desarrollo de esta actividad se realiza el **"Cálculo de Métricas"**, a partir de lo propuesto en [Basso et al., 2013], mediante la cual se miden las métricas seleccionadas para el proyecto, dejando registrado las variables dependientes que estas poseen mediante las cuales se obtiene el valor del indicador. El proceso de medición debe ser planificado y preciso, estableciendo las condiciones sobre las cuales se realizarán las mediciones y qué aspectos serán considerados para las mismas.

En la siguiente sección se presenta la aplicación de la técnica y el registro de los formalismos para la prueba de concepto.

4.3.3.1.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica Cálculo de Métricas, que utiliza como insumos el Listado de Métricas (Tabla 4.19) y Plan de Acción (Tabla 4.24), los cuales son transcriptos con el mismo número de tabla asignados al momento de su presentación, para facilitar al lector en la comprensión de la técnica.

Listado de Métricas				
Responsable:	Rodriguez H.		**Fecha:**	25/04/16
ID#:	G.PI.PIA.LiMe		**Versión:**	1.0
Nombre	**Tipo**	**Fórmula**	**Comentarios**	
Tiempo total requerido para el desarrollo del proyecto	Proyecto	DRPY = $\sum$ trA trA = tiempo requerido por actividad	Sumatoria de los tiempos requeridos para cada actividad del proyecto	
Grado de Utilidad de Atributos	Datos	$GUA = \frac{NA(T) - (NO_{UTILES(T)} + 0,5 * NAUD(T)}{NA(T)}$ **NA(T)** = NASE(T) + NAUD(T) + NANC(T) + NANS(T) **NO _ UTILES(T)** = NANC(T) + NANS(T)	− Nro. de atributos útiles sin errores [NASE (T)] − Nro. de atributos útiles con defectos [NAUD (T)] − Nro. de atributos no correctos [NANC (T)] − Nro. de atributos no significativos [NANS (T)]	

Tabla 4.19 (Transcripta). Prueba de Concepto – Listado de Métricas

Plan de Acción								
Responsable:		Rodriguez H.			**Fecha:**		15/06/16	
ID#:		G.PI.PIA.PlAc			**Versión:**		1.3	
ID Actividad	**Actividad**	**Inicio Estimado**	**Inicio Real**	**Fin Estimado**	**Fin Real**	**Esfuerzo Estimado (Hs)**	**Esfuerzo Real (Hs)**	**Comentarios**
G.In	**Iniciación**	**04/04/16**	**04/04/16**	**22/04/16**	**22/04/16**	**12**	**11**	
G.In.EIP	Exploración Inicial del Proyecto	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DeC	Definición de la Comunicación	04/04/16	04/04/16	18/04/16	18/04/16	2	2	
G.In.EvS	Evaluación de la Situación	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DCV	Definición del Ciclo de Vida	19/04/16	19/04/16	22/04/16	22/04/16	2	1	
G.PI	**Planificación**	**04/04/16**	**04/04/16**	**29/04/16**	**29/04/16**	**10**	**9**	
G.PI.PIM	Planificación de la Mediciones	04/04/16	04/04/16	29/04/16	29/04/16	2	1	
G.PI.PIA	Planificación de las Actividades	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.PI.PIR	Planificación de los Recursos	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.PI.PRe	Planificación de las Responsabilidades	04/04/16	04/04/16	29/04/16	29/04/16	4	4	
G.So	**Soporte**	**20/04/16**	**20/04/16**	**15/06/16**	**15/06/16**	**8**	**12**	
G.So.MeP	Mediciones del Proyecto	20/04/16	20/04/16	15/06/16	15/06/16	4	6	Se prevé el registro de las métricas de manera mensual
G.So.GeC	Gestión de la Configuración	20/04/16	20/04/16	15/06/16	15/06/16	4	6	
G.Co	**Control**	**20/04/16**	**20/04/16**	**13/06/16**	**13/06/16**	**12**	**15**	
G.Co.GeD	Gestión del Desarrollo	20/04/16	20/04/16	13/06/16	13/06/16	4	6	Se prevé la aplicación del reporte de estado de manera mensual
G.Co.CoA	Control de las Actividades	20/04/16	20/04/16	13/06/16	13/06/16	6	8	
G.Co.Gca	Gestión del Cambio	20/04/16	20/04/16	13/06/16	13/06/16	2	1	

Tabla 4.24.a (Transcripta). Prueba de Concepto - Plan de Acción (fin del proyecto)

G.Ci	**Cierre**	**06/06/16**	**06/06/16**	**15/06/16**	**15/06/16**	**4**	**4**	
G.Ci.FEC	**Formalización Externa del Cierre del Proyecto**	**06/06/16**	**06/06/16**	**13/06/16**	**13/06/16**	**2**	**2**	
G.Ci.FIC	**Formalización Interna del Cierre del Proyecto**	**14/06/16**	**14/06/16**	**15/06/16**	**15/06/16**	**2**	**2**	
D.EN	**Entendimiento del Negocio**	**04/04/16**	**04/04/16**	**22/04/16**	**20/04/16**	**44**	**32**	
D.EN.AnN	**Análisis del Negocio**	**04/04/16**	**04/04/16**	**20/04/16**	**20/04/16**	**28**	**22**	
D.EN.CPN	**Comprensión del Problema de Negocio**	**18/04/16**	**18/04/16**	**22/04/16**	**20/04/16**	**16**	**10**	
D.ED	**Entendimiento de los Datos**	**25/04/16**	**25/04/16**	**06/05/16**	**06/05/16**	**56**	**52**	
D.ED.AnD	**Análisis de los Datos**	**25/04/16**	**25/04/16**	**29/04/16**	**29/04/16**	**22**	**20**	
D.ED.ExD	**Exploración de los Datos**	**28/04/16**	**28/04/16**	**06/05/16**	**06/05/16**	**22**	**22**	
D.ED.EvD	**Evaluación de los Datos**	**02/05/16**	**02/05/16**	**06/05/16**	**06/05/16**	**12**	**10**	
D.Mo	**Modelado**	**09/05/16**	**09/05/16**	**13/05/16**	**13/05/16**	**26**	**16**	
D.Mo.MoP	Modelado del problema	09/05/16	09/05/16	10/05/16	10/05/16	14	8	
D.Mo.CoM	Configuración del Modelo	11/05/16	11/05/16	13/05/16	13/05/16	12	8	
D.PD	**Preparación de los Datos**	**16/05/16**	**16/05/16**	**20/05/16**	**20/05/16**	**36**	**28**	
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	16/05/16	16/05/16	18/05/16	18/05/16	20	16	
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos	17/05/16	17/05/16	20/05/16	20/05/16	16	12	
D.Im	**Implementación**	**23/05/16**	**23/05/16**	**27/05/16**	**27/05/16**	**30**	**26**	
D.Im.SeM	Selección del Modelo	23/05/16	23/05/16	24/05/16	24/05/16	8	6	
D.Im.ExI	Explotación de Información	24/05/16	24/05/16	27/05/16	27/05/16	22	20	
D.EP	**Evaluación y Presentación**	**30/05/16**	**30/05/16**	**13/06/16**	**13/06/16**	**26**	**18**	
D.EP.EvR	Evaluación de los Resultados	30/05/16	30/05/16	02/06/16	02/06/16	10	6	
D.EP.PrR	Presentación de los Resultados	01/06/16	01/06/16	13/06/16	13/06/16	16	12	

Tabla 4.24.b (Transcripta). Prueba de Concepto - Plan de Acción (fin del proyecto)

Registro de Mediciones (G.So.MeP.ReMe): se establecieron dos métricas para evaluar durante el proyecto: "Tiempo total requerido para el desarrollo del proyecto" y "Grado de Utilidad de Atributos". Se determinó (en el plan de acción) que el registro formal de las métricas se realiza de manera mensual, a partir de lo cual se realizó el cálculo de las mismas en tres ocasiones (al cumplir un mes el proyecto, al cumplir dos meses y en el cierre del proyecto), obteniéndose los siguientes resultados: Tiempo total requerido para el desarrollo del proyecto (111; 216; 223) y Grado de Utilidad de Atributos igual a 6,71 (valor constante para todo el proyecto). Finalmente, se presenta en la tabla 4.32 los resultados registrados en la última versión del producto interno previamente descripto, existiendo dos versiones previas las cuales se ilustran en las tablas A.4 y A.5.

Registro de Mediciones			
Responsable:	Esposito E.	**Fecha:**	15/06/2016
ID#:	G.So.MeP.ReMe	**Versión:**	1.2

Indicador	Medición	Descripción
Tiempo total requerido para el desarrollo del proyecto	**DRPY** = 223	Tdesarrollo = 172 Tgestion = 51
Grado de Utilidad de Atributos	**GUA** = 6,71	NA = 392 NASE = 15 NAUD= 2 NO_UTILES = 275

Tabla 4.32. Prueba de Concepto – Registro de Mediciones (fin del proyecto)

4.3.3.2. Actividad: Gestión de la Configuración (G.So.GeC)

En esta actividad se realizan las tareas vinculadas con la trazabilidad de los productos generados durante el desarrollo del proyecto, garantizando que en todo momento los miembros del equipo estén informados de las versiones actuales de los resultados producidos en cada fase.

Información de Entrada

- Reglas de Versionado (Externo)
- Reporte de Evaluación del Cambio (G.Co.Gca.RECa)
- Modelo de Ciclo de Vida (G.In.DCV.MoCV)Plan de Acción (G.Pl.PlA.PlAc)

Información de Salida

- Reporte de Versionado (G.So.GeC.ReVe)
- Informe del Estado de la Configuración (G.So.GeC.InEC)

4.3.3.2.1. Formalismos Identificados

La información de salida de la actividad se formaliza mediante los siguientes elementos: el Reporte de Versionado y el Informe del Estado de la Configuración, los cuales se presentan a continuación.

Reporte de Versionado (G.So.GeC.ReVe): se deja registro de los cambios de versión en los distintos productos internos del proyecto, registrando la fecha en la cual se realiza la actualización de versión, el elemento afectado, el número de versión previa y actual (cada cual registrado en la columna homónima), junto con el registro del motivo del cambio y el detalle de las acciones realizadas en el mismo. La tabla 4.33 ilustra la estructura del formalismo previamente descripto.

Informe del Estado de la Configuración (G.So.GeC.InEC): se registra el estado global del proyecto, donde cada fila pertenece al conjunto de productos internos (o elementos) generados por cada actividad (agrupados por fase por simplicidad) y las columnas corresponden al estado o versión global del proyecto, identificándose el conjunto de versionado de cada elemento del proyecto. La tabla 4.34 ilustra la estructura del formalismo previamente descripto.

Reporte de Versionado					
Responsable:	Persona a cargo		**Fecha:**	Fecha de Realización	
ID#:	Identificador del Producto				
Fecha	**Elemento**	**Versión previa**	**Versión Actual**	**Cambio Asociado**	**Descripción**
Fecha de ejecución del cambio de versión 1	Nombre del Producto afectado	Número de Versión previa	Número de Versión actual	Cambio acontecido	Descripción del cambio realizado
…	…	…	…	…	…
Fecha de ejecución del cambio de versión N	Nombre del Producto afectado	Número de Versión previa	Número de Versión actual	Cambio acontecido	Descripción del cambio realizado

Tabla 4.33. Formalismo: Reporte de Versionado

Informe del Estado de la Configuración					
Responsable:	Persona a cargo		**Fecha:**	Fecha de Realización	
ID#:	Identificador del Producto				
ID Actividad	**Actividad**	**Elemento**	**Versión del Proyecto**		
			Versión 1	**…**	**Versión N**
Identificador fase 1	**Fase 1**				
Identificador actividad 1 fase 1	Actividad 1 fase 1	Listado de elementos generados en la actividad 1 fase 1	Versión de los elementos de la actividad 1 fase 1	…	Versión de los elementos de la actividad 1 fase 1
…	…	…	…	…	…
Identificador actividad N fase 1	Actividad N fase 1	Listado de elementos generados en la actividad N fase 1	Versión de los elementos de la actividad N fase 1	…	Versión de los elementos de la actividad N fase 1
…	…				
Identificador fase N	**Fase N**				
Identificador actividad 1 fase N	Actividad 1 fase N	Listado de elementos generados en la actividad 1 fase N	Versión de los elementos de la actividad 1 fase N	…	Versión de los elementos de la actividad 1 fase N
…	…	…	…	…	…
Identificador actividad N fase N	Actividad N fase N	Listado de elementos generados en la actividad N fase N	Versión de los elementos de la actividad N fase N	…	Versión de los elementos de la actividad N fase N

Tabla 4.34. Formalismo: Informe de Estado de la Configuración

4.3.3.2.2. Técnica Identificada

Para el desarrollo de esta actividad se propone la técnica **"Configuración del versionado"**, que tiene como objetivo manifestar el estado actual del proyecto para que todos los miembros del equipo de trabajo estén informados de los productos a utilizar y los cambios introducidos en el mismo. Mediante esta técnica se permite determinar el progreso del proyecto, dejando registro formal de los motivos por los cuales se llegó a la evolución actual, así como posibilitar el retorno a un estado anterior (en caso de haber detectado inconvenientes o errores en su desarrollo). Para su

implementación, la organización debe tener formalizada las reglas que utiliza para representar la evolución del producto y del proyecto, mediante una estructura estándar de representación de los cambios y su significado. Debiendo ser definida en caso que la organización no contase con una.

A partir de la información de cambios realizadas por los miembros del proyecto, se registra el impacto de los mismos en los productos internos, indicando: la fecha de acontecimiento, los productos internos afectados junto con el cambio de versión generado (de acuerdo a la reglas de versionado), se describe el tipo de cambio realizado y se resume su efecto sobre los elementos en cuestión, permitiendo trazar los motivos por los cuales el producto fue evolucionando hasta la versión actual. De forma complementaria, se actualiza el cambio de versión en el Informe del Estado de la Configuración. De acuerdo al tipo de estrategia a utilizar para el proyecto (según el modelo de ciclo de vida seleccionado), se evalúa su progreso general y el cambio de versión global según las iteraciones en las etapas del mismo y las modificaciones en sus objetivos. En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo para la prueba de concepto.

4.3.3.2.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica Configuración del versionado. Dicha técnica utiliza como insumos las Reglas de Versionado (Fuente de Información 4.1), existentes en la organización, Reporte de Evaluación del Cambio (Tabla 4.43) y Modelo de Ciclo de Vida (Tabla 4.16), los cuales se presentan a continuación. Los formalismos indicados como elementos de entrada, son transcriptos con el mismo número de tabla, para facilitar al lector en la comprensión de la técnica.

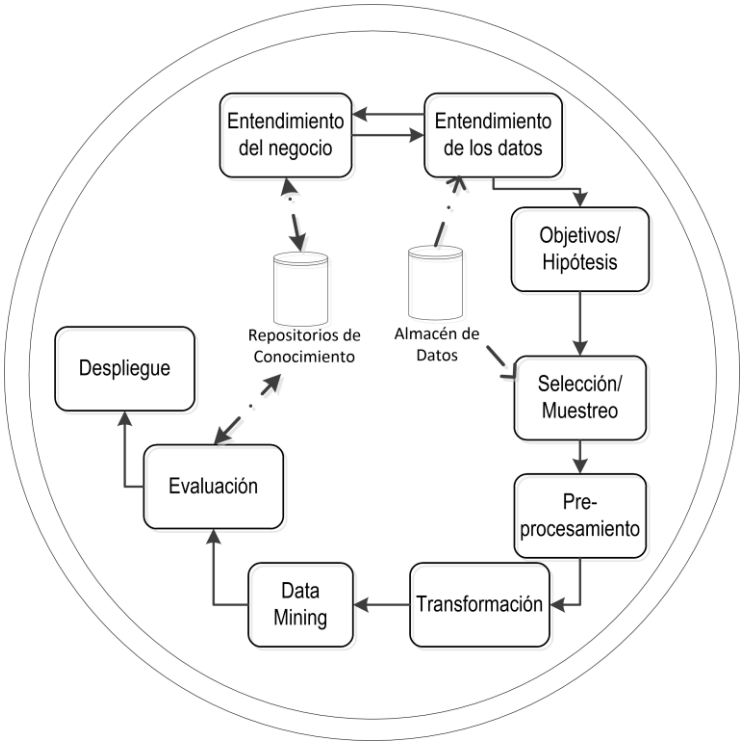
Se utilizan dos dígitos para reflejar el progreso de los productos a lo largo del proyecto X.Y: el primero de ellos (X) indica la versión mayor del documento, incrementándose de a uno cada vez que se modifican o eliminan elementos del producto (generando incompatibilidad con otros productos o versiones anteriores). En caso que el producto se encuentre en un estadio temprano, el cual no puede ser utilizado para su uso como entrada en otras tareas, este digito debe ser indicado con cero. El segundo (Y), describe la versión menor del documento, la cual se incrementa en uno reflejando incorporaciones o alteraciones que no modifiquen la funcionalidad en el producto. Cuando se modifique la versión superior, este valor será restituido al valor cero. *Para el registro del estado del proyecto, se utiliza la misma lógica (X.Y). El primer elemento (X) indica alteraciones en las necesidades o estrategias de ejecución del proyecto, mientras que el segundo (Y) representa iteraciones en el ciclo de vida.*

Fuente de Información 4.1. Prueba de Concepto - Reglas de Versionado

Reporte de Evaluación del Cambio						
Responsable:	Esposito E.		**Fecha :**	03/06/2016		
ID#:	G.Co.Gca.ReCa					
Fecha	**Cambio**	**Pedido por**	**Descripción**	**Impacto**	**Estado**	**Responsable**
03/06/16	(ReCa.1) Ajuste de plazos de hitos del proyecto	Silva H. (rehi.3)	Modificación de la fecha de cumplimiento de los hitos debido a la imposibilidad del cliente de cumplimentar con los hitos pactados	Modificación en la fecha de las actividades pendientes en una semana. La fecha inicial de Formalización Interna del Cierre del Proyecto, se cambia al 14/06/16. La fecha final de las siguientes actividades se posponen al 13/06/16: Gestión del Desarrollo, Control de las Actividades, Gestión del Cambio, Formalización Externa del Cierre del Proyecto y Presentación de los Resultados, y la fecha final de las siguientes se pospone al 15/06/16: Mediciones del Proyecto, Gestión de la Configuración y Formalización Interna del Cierre del Proyecto. Se realizan ajustes en los recursos requeridos para dichas actividades.	Aprobado	Rodriguez H. (rehi.1)

Tabla 4.43 (Transcripta). Prueba de Concepto – Reporte de Evaluación del Cambio

Modelo de Ciclo de Vida			
Responsable:	Esposito E.	**Fecha:**	22/04/2016
ID#:	G.In.DCV.MoCV	**Versión:**	1.0



Criterios de transición:
Las fases posteriores podrán comenzarse 10% antes de la finalización de la misma (fecha estimada de finalización de las tareas involucradas).

Tabla 4.16 (Transcripta). Prueba de Concepto - Modelo de Ciclo de Vida

Reporte de Versionado (G.So.GeC.ReVe): durante el desarrollo del proyecto se realizaron siete modificaciones a productos internos del mismo, a partir del progreso de las actividades, las actividades de control y los pedidos de cambios. De las alteraciones realizadas, solo una ha procedido de la evaluación formal del líder del proyecto (ReCa.1). La tabla 4.35 ilustra los resultados registrados para la prueba de concepto.

Reporte de Versionado					
Responsable:	Rodriguez H.		**Fecha:**	13/06/2016	
ID#:	G.So.GeC.ReVe				
Fecha	**Elemento**	**Versión previa**	**Versión Actual**	**Cambio Asociado**	**Descripción**
20/04/16	Suposiciones del Proyecto	1.0	1.1	Incorporación de nuevos elementos	Nuevos conocimientos adquiridos a partir de la entrevista 2
06/05/16	Plan de Acción	1.0	1.1	Ajuste por reporte de estado	
03/06/16	Plan de Acción	1.1	1.2	Ajuste por reporte de estado y Reporte de Evaluación del Cambio (ReCa.1)	
03/06/16	Registro de Mediciones	1.0	1.1	Ajuste por reporte de estado	
03/06/16	Plan de Necesidad de Recursos	1.0	1.1	Ajuste de la necesidad de recursos causado por riesgo (risk.1)	Descripción del impacto en: Reporte de Evaluación del Cambio (ReCa.1)
15/06/16	Registro de Mediciones	1.1	1.2	Ajustes a cierre del proyecto	
15/06/16	Plan de Acción	1.2	1.3	Ajustes a cierre del proyecto	

Tabla 4.35. Prueba de Concepto - Reporte de Versionado

Informe del Estado de la Configuración (G.So.GeC.InEC): durante el desarrollo del proyecto, no han surgido cambios en sus objetivos, en la estrategia de ejecución del proyecto o iteraciones en el ciclo de vida, por lo cual se identifica una única versión, detallándose su estado interno (versiones vigentes de los productos internos). Las tablas 4.36.a y 4.36.b ilustran los resultados registrados para la prueba de concepto.

Informe del Estado de la Configuración				
Responsable:	Rodriguez H.	**Fecha:**	13/06/2016	
ID#:	G.So.GeC.InEC			
ID Actividad	**Actividad**	**Elemento**	**Versión del Proyecto**	
			V. 1.0 (Actual)	
G.In	**Iniciación**			
G.In.EIP	Exploración Inicial del Proyecto	Recursos Humanos Involucrados Riesgos del Proyecto Plan de Contingencias	1.0 1.0 1.0	
G.In.DeC	Definición de la Comunicación	Plan de Comunicación	1.0	
G.In.EvS	Evaluación de la Situación	Reporte de Evaluación de Herramientas Reporte de Evaluación de Viabilidad	1.0 1.0	
G.In.DCV	Definición del Ciclo de Vida	Modelo de Ciclo de Vida	1.0	

Tabla 4.36.a. Prueba de Concepto - Informe de Estado de la Configuración

G.Pl	**Planificación**		
G.Pl.PIM	Planificación de la Mediciones	Listado de Métricas Estimación del Proyecto	1.0 1.0
G.Pl.PIA	Planificación de las Actividades	Mapa de Actividades Plan de Acción	1.0 1.3
G.Pl.PIR	Planificación de los Recursos	Plan de Necesidad de Recursos	1.1
G.Pl.PRe	Planificación de las Responsabilidades	Matriz de Responsabilidades Propuesta del Proyecto	1.0 1.0
G.So	**Soporte**		
G.So.MeP	Mediciones del Proyecto	Registro de Mediciones	1.2
G.So.GeC	Gestión de la Configuración	Reporte de Versionado Informe del Estado de la Configuración	- -
G.Co	**Control**		
G.Co.GeD	Gestión del Desarrollo	Reporte de Estado	-
G.Co.CoA	Control de las Actividades	Registro de Riesgos Acontecidos	-
G.Co.Gca	Gestión del Cambio	Reporte de Evaluación del Cambio	-
G.Ci	**Cierre**		
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	Documento de Aceptación	-
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	Reporte de Cierre	-
D.EN	**Entendimiento del Negocio**		
D.EN.AnN	Análisis del Negocio	Fuentes de Información del Cliente Definiciones, Acrónimos y Abreviaciones Objetivos del Proyecto Criterios de Éxito del Proyecto Expectativas del Proyecto Suposiciones del Proyecto Restricciones del Proyecto	1.0 1.0 1.0 1.0 1.0 1.1 1.0
D.EN.CPN	Comprensión del Problema de Negocio	Problema de Negocio Criterios de Éxito del Problema de Negocio	1.0 1.0
D.ED	**Entendimiento de los Datos**		
D.ED.AnD	Análisis de los Datos	Diccionario de Fuente de Datos Campos Relacionados con el Problema de Negocio	1.0 1.0
D.ED.ExD	Exploración de los Datos	Reporte de Datos Explorados Fuente Integrada de datos	1.0 -
D.ED.EvD	Evaluación de los Datos	Reporte de la Calidad de los Datos	1.0
D.Mo	**Modelado**		
D.Mo.MoP	Modelado del problema	Diseño del Proceso de Explotación de Información	1.0
D.Mo.CoM	Configuración del Modelo	Selección de Algoritmos de Explotación de Información Selección de Variables del Modelo Estrategias de Evaluación de Modelos	1.0 1.0 1.0
D.PD	**Preparación de los Datos**		
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	Reporte de Generación de la Fuente Temporaria de datos Fuente Temporaria de Datos	1.0 -
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos	Reporte de Adecuación de la Fuente Temporaria de Datos	1.0
D.Im	**Implementación**		
D.Im.SeM	Selección del Modelo	Reporte de Estrategia de Parametrización del Modelo	1.0
D.Im.ExI	Explotación de Información	Reporte de Implementación del Modelo	1.0
D.EP	**Evaluación y Presentación**		
D.EP.EvR	Evaluación de los Resultados	Reporte de Evaluación de los Resultados	1.0
D.EP.PrR	Presentación de los Resultados	Reporte del Proyecto	1.0

Tabla 4.36.b. Prueba de Concepto - Informe de Estado de la Configuración

4.3.4. Fase: Control (G.Co)

En la fase Control se evalúa el desarrollo del proyecto, verificando que el mismo se efectúe de acuerdo a lo planificado y pactado con el cliente, y dejando registro formal de cualquier desvío, cambio o posible evento riesgoso que acontezca. Como resultado general de dicha fase se

formalizan los sucesos o imprevistos que ocurran durante el desarrollo del proyecto y sus correspondientes acciones derivadas.

Esta fase está conformada por tres actividades: Gestión del Desarrollo (sección 4.3.4.1), Control de las Actividades (sección 4.3.4.2) y Gestión del Cambio (sección 4.3.4.3). La figura 4.7, presenta una visión resumida de las actividades que integran la fase y sus elementos de entrada y salida (las imágenes de cada formalismo son representaciones miniatura de los mismos).

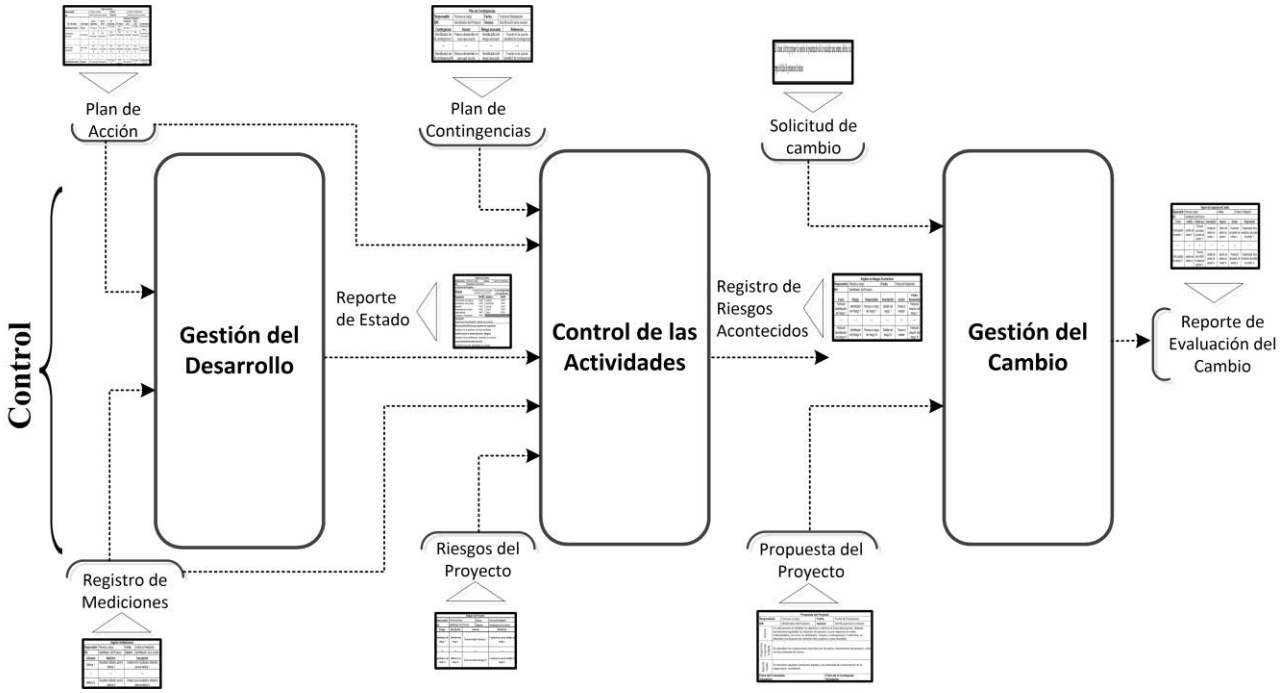


Figura 4.7. Fase Control

4.3.4.1. Actividad: Gestión del Desarrollo (G.Co.GeD)

En esta actividad se realiza el seguimiento del proyecto, dejando registro formal del progreso del mismo. El resultado de esta tarea contribuye en la toma de decisiones del líder del proyecto, con respecto al cumplimiento de lo planificado, pudiendo identificar la necesidad de reajustar las acciones previstas.

Información de Entrada

- Plan de Acción (G.Pl.PlA.PlAc)
- Registro de Mediciones (G.Co.MeP.ReMe)

Información de Salida

- Reporte de Estado (G.Co.GeD.ReEs)

4.3.4.1.1. Formalismo Identificado

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se propone el Reporte de Estado, el cual se presenta a continuación.

Reporte de Estado (G.Co.GeD.ReEs): basado en la propuestas de [Project Management Institute, Inc., 2013a; Verzuh, 2015] (sección 2.4.2.4, pág. 47), este formalismo registra formalmente la situación del proyecto hasta la fecha de reporte. Su estructura está conformada por dos secciones: *Evaluación del Programa*, donde se asienta la diferencia entre el tiempo planificado y el real, en porcentaje (PePR), para el proyecto general, sus subprocesos y fases (así como de los costos asociados, en caso que fuese necesario), y *Descripción del Periodo*, detallando situaciones identificadas como posibles riesgos que requieran seguimiento, cambios acontecidos (si hubiese) y los logros realizados durante el periodo. La tabla 4.37 ilustra la estructura del formalismo previamente descripto.

Reporte de Estado			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto		
Evaluación del Programa			
Global	Porcentaje diferencia entre el tiempo planificado y el real (PePR)		**(% por debajo/arriba de lo planificado)**
Desarrollo	**PePR**	**Gestión**	**PePR**
Entendimiento del Negocio	PePR	Iniciación	PePR
Entendimiento de los Datos	PePR	Planificación	PePR
Modelado	PePR	Soporte	PePR
Preparación de los Datos	PePR	Control	PePR
Implementación	PePR	Cierre	PePR
Evaluación y Presentación	PePR		
Descripción: Detalle de las fases/actividades cubiertas por el reporte			
Situaciones identificadas que requieren de seguimiento			
Descripción de las situaciones de riesgo identificadas			
Cambios durante el periodo (alcances, tiempos)			
Descripción de las modificaciones acontecidas en el periodo			
Logros principales durante el periodo			
Descripción de los hitos desarrollados en el periodo			

Tabla 4.37. Formalismo: Reporte de Estado

4.3.4.1.2. Técnica Identificada

Para el desarrollo de esta actividad se propone la técnica **"Seguimiento de Avance"**, mediante la cual se deja registro del progreso del proyecto con respecto a lo planificado. La ejecución de la actividad debe ser previamente planificada, dejando registro del periodo y alcance (fases/actividades) que el mismo cubre. A partir de ello, se analizan las expectativas de avance de las etapas del proyecto con respecto al avance real, para lo cual se utiliza el plan de acción como elemento de soporte, dejando registro de dicha información. Como resultado de esta actividad se verá actualizado el plan de acción y/o modificado en caso que se requiera ajustar el plan del proyecto, debido a un porcentaje elevado de desemejanza entre lo estimado y lo real. Finalmente, se resumen aquellas situaciones identificadas en el período de análisis que requieran ser controladas

durante el próximo ciclo, debido a que presenta una potencial situación de riesgo para el proyecto, detallándose las alteraciones en el proyecto con respecto a objetivos y/o recursos, y listando los hitos alcanzados de mayor trascendencia.

En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo para la prueba de concepto.

4.3.4.1.3. Ejecución de la Actividad en la Prueba de Concepto

Para aplicar la técnica Seguimiento de Avance en el proyecto seleccionado como prueba de concepto, se utilizan como insumos: el Plan de Acción (Tabla 4.24) y el Registro de Mediciones (Tabla 4.32), los cuales son transcriptos con el mismo número de tabla, para facilitar al lector en la comprensión de la implementación de la técnica.

Plan de Acción								
Responsable:		Rodriguez H.			**Fecha:**		15/06/16	
ID#:		G.PI.PIA.PIAc			**Versión:**		1.3	
ID Actividad	**Actividad**	**Inicio Estimado**	**Inicio Real**	**Fin Estimado**	**Fin Real**	**Esfuerzo Estimado (Hs)**	**Esfuerzo Real (Hs)**	**Comentarios**
G.In	**Iniciación**	**04/04/16**	**04/04/16**	**22/04/16**	**22/04/16**	**12**	**11**	
G.In.EIP	Exploración Inicial del Proyecto	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DeC	Definición de la Comunicación	04/04/16	04/04/16	18/04/16	18/04/16	2	2	
G.In.EvS	Evaluación de la Situación	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DCV	Definición del Ciclo de Vida	19/04/16	19/04/16	22/04/16	22/04/16	2	1	
G.PI	**Planificación**	**04/04/16**	**04/04/16**	**29/04/16**	**29/04/16**	**10**	**9**	
G.PI.PIM	Planificación de la Mediciones	04/04/16	04/04/16	29/04/16	29/04/16	2	1	
G.PI.PIA	Planificación de las Actividades	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.PI.PIR	Planificación de los Recursos	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.PI.PRe	Planificación de las Responsabilidades	04/04/16	04/04/16	29/04/16	29/04/16	4	4	
G.So	**Soporte**	**20/04/16**	**20/04/16**	**15/06/16**	**15/06/16**	**8**	**12**	
G.So.MeP	Mediciones del Proyecto	20/04/16	20/04/16	15/06/16	15/06/16	4	6	Se prevé el registro de las métricas de manera mensual
G.So.GeC	Gestión de la Configuración	20/04/16	20/04/16	15/06/16	15/06/16	4	6	

Tabla 4.24.a. (Transcripta). Prueba de Concepto - Plan de Acción (fin del proyecto)

G.Co	**Control**	**20/04/16**	**20/04/16**	**13/06/16**	**13/06/16**	**12**	**15**	
G.Co.GeD	Gestión del Desarrollo	20/04/16	20/04/16	13/06/16	13/06/16	4	6	Se prevé la aplicación del reporte de estado de manera mensual
G.Co.CoA	Control de las Actividades	20/04/16	20/04/16	13/06/16	13/06/16	6	8	
G.Co.Gca	Gestión del Cambio	20/04/16	20/04/16	13/06/16	13/06/16	2	1	
G.Ci	**Cierre**	**06/06/16**	**06/06/16**	**15/06/16**	**15/06/16**	**4**	**4**	
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	06/06/16	06/06/16	13/06/16	13/06/16	2	2	
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	14/06/16	14/06/16	15/06/16	15/06/16	2	2	
D.EN	**Entendimiento del Negocio**	**04/04/16**	**04/04/16**	**22/04/16**	**20/04/16**	**44**	**32**	
D.EN.AnN	Análisis del Negocio	04/04/16	04/04/16	20/04/16	20/04/16	28	22	
D.EN.CPN	Comprensión del Problema de Negocio	18/04/16	18/04/16	22/04/16	20/04/16	16	10	
D.ED	**Entendimiento de los Datos**	**25/04/16**	**25/04/16**	**06/05/16**	**06/05/16**	**56**	**52**	
D.ED.AnD	Análisis de los Datos	25/04/16	25/04/16	29/04/16	29/04/16	22	20	
D.ED.ExD	Exploración de los Datos	28/04/16	28/04/16	06/05/16	06/05/16	22	22	
D.ED.EvD	Evaluación de los Datos	02/05/16	02/05/16	06/05/16	06/05/16	12	10	
D.Mo	**Modelado**	**09/05/16**	**09/05/16**	**13/05/16**	**13/05/16**	**26**	**16**	
D.Mo.MoP	Modelado del problema	09/05/16	09/05/16	10/05/16	10/05/16	14	8	
D.Mo.CoM	Configuración del Modelo	11/05/16	11/05/16	13/05/16	13/05/16	12	8	
D.PD	**Preparación de los Datos**	**16/05/16**	**16/05/16**	**20/05/16**	**20/05/16**	**36**	**28**	
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	16/05/16	16/05/16	18/05/16	18/05/16	20	16	
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos	17/05/16	17/05/16	20/05/16	20/05/16	16	12	
D.Im	**Implementación**	**23/05/16**	**23/05/16**	**27/05/16**	**27/05/16**	**30**	**26**	
D.Im.SeM	Selección del Modelo	23/05/16	23/05/16	24/05/16	24/05/16	8	6	
D.Im.ExI	Explotación de Información	24/05/16	24/05/16	27/05/16	27/05/16	22	20	
D.EP	**Evaluación y Presentación**	**30/05/16**	**30/05/16**	**13/06/16**	**13/06/16**	**26**	**18**	
D.EP.EvR	Evaluación de los Resultados	30/05/16	30/05/16	02/06/16	02/06/16	10	6	
D.EP.PrR	Presentación de los Resultados	01/06/16	01/06/16	13/06/16	13/06/16	16	12	

Tabla 4.24.b. (Transcripta). Prueba de Concepto - Plan de Acción (fin del proyecto)

Registro de Mediciones			
Responsable:	Esposito E.	**Fecha:**	15/06/2016
ID#:	G.So.MeP.ReMe	**Versión:**	1.2

Indicador	Medición	Descripción
Tiempo total requerido para el desarrollo del proyecto	**DRPY** = 223	Tdesarrollo = 172 Tgestion = 51
Grado de Utilidad de Atributos	**GUA** = 6,71	NA = 392 NASE = 15 NAUD= 2 NO_UTILES = 275

Tabla 4.32 (Transcripta). Prueba de Concepto - Registro de Mediciones

Reporte de Estado (G.So.GeD.ReEs): de acuerdo al progreso del proyecto en los distintos hitos de control definidos en el plan de acción, se llevaron a cabo dos reportes de estado en las fechas 06/05/2016 y 03/06/2016 (tablas 4.38 y 4.39 respectivamente), en los cuales se registraron los desvíos temporales con respecto a lo planificado, junto con los logros y desafíos alcanzados durante dicho periodo.

Reporte de Estado			
Responsable:	Rodriguez H.	**Fecha:**	06/05/2016
ID#:	G.Co.GeD.ReEs.1		
Evaluación del Programa			
Global	**-14,75%**		**(% por debajo de lo planificado)**
Desarrollo	**-16%**	**Gestión**	**-9,09%**
Entendimiento del Negocio	-27,27%	Iniciación	-8,33%
Entendimiento de los Datos	-7,14%	Planificación	-10%
Modelado	-	Soporte	-
Preparación de los Datos	-	Control	-
Implementación	-	Cierre	-
Evaluación y Presentación	-		
Descripción: *Se evaluaron las actividades finalizadas hasta la fecha. (ver G.Pl.PIA.PIAc versión 1.1)*			
Situaciones identificadas que requieren de seguimiento - **Cambios durante el periodo (alcances, tiempos)** - **logros principales durante el periodo** Definición de los alcances del proyecto y los problemas de negocio			

Tabla 4.38. Prueba de Concepto - Reporte de Estado (G.Co.GeD.ReEs.1)

Reporte de Estado			
Responsable:	Rodriguez H.	**Fecha:**	03/06/2016
ID#:	G.Co.GeD.ReEs.2		
Evaluación del Programa			
Global	**-19,64%**		**(% por debajo de lo planificado)**
Desarrollo	**-20,79%**	**Gestión**	**-9,09%**
Entendimiento del Negocio	-27,27%	Iniciación	-8,33%
Entendimiento de los Datos	-7,14%	Planificación	-10%
Modelado	-38,46%	Soporte	-
Preparación de los Datos	-22,22%	Control	-
Implementación	-13,33%	Cierre	-
Evaluación y Presentación	-40%		
Descripción: *Se evaluaron las actividades finalizadas hasta la fecha. (ver G.Pl.PlA.PlAc versión 1.2)*			
Situaciones identificadas que requieren de seguimiento Se identifica un incremento en el tiempo de respuesta del cliente/experto **Cambios durante el periodo (alcances, tiempos)** - **logros principales durante el periodo** Generación y validación de los resultados vinculados con el problema de negocio			

Tabla 4.39. Prueba de Concepto - Reporte de Estado (G.Co.GeD.ReEs.2)

4.3.4.2. Actividad: Control de las Actividades (G.Co.CoA)

En esta actividad se evalúan las situaciones potencialmente peligrosas para el desarrollo del proyecto, realizando un seguimiento, control y registro de acontecimientos, así como de las acciones realizadas. El resultado de esta actividad contribuye en la calidad del proceso.

Información de Entrada

- Riesgos del Proyecto (G.In.EIP.RiPr)
- Plan de Contingencias (G.In.EIP.PCon)
- Plan de Acción (G.Pl.PlA.PlAc)
- Registro de Mediciones (G.So.MeP.ReMe)
- Reporte de Estado (G.Co.GeD.ReEs)

Información de Salida

- Registro de Riesgos Acontecidos (G.Co.CoA.ReRA)

4.3.4.2.1. Formalismos Identificados

Para la formalización de la información de salida esperada en la actividad, se propone el Registro de Riesgos Acontecidos, el cual se presenta a continuación.

Registro de Riesgos Acontecidos (G.Co.CoA.ReRA): se registran formalmente las situaciones riesgosas que ocurrieron en el proyecto, detallándose: la fecha en la cual se registra, el identificador del riesgo, la persona responsable de su resolución, una descripción del riesgo ocurrido, las acciones a realizar para solucionar la contingencia y la fecha en la cual la misma fue resuelta (siendo cada campo registrado en las columnas homónimas). La tabla 4.40 ilustra la estructura del formalismo previamente descripta.

Registro de Riesgos Acontecidos					
Responsable:	Persona a cargo		**Fecha:**	Fecha de Realización	
ID#:	Identificador del Producto				
Fecha	**Riesgo**	**Responsable**	**Descripción**	**Acción**	**Fecha Resolución**
Fecha de identificación del riesgo 1	Identificador del Riesgo 1	Persona a cargo del riesgo 1	Detalle del riesgo 1	Tareas a realizar	Fecha de solución del riesgo 1
…	…	…	…	…	…
Fecha de identificación del riesgo N	Identificador del Riesgo N	Persona a cargo del riesgo N	Detalle del riesgo N	Tareas a realizar	Fecha de solución del riesgo N

Tabla 4.40. Formalismo: Registro de Riesgos Acontecidos

4.3.4.2.2. Técnica Identificada

Para el desarrollo de esta actividad se propone la técnica **"Evaluación de Riesgos"**, mediante la cual se evalúan las posibles situaciones críticas del proyecto, realizando un seguimiento de aquellos indicadores que permitan identificar a tiempo el acontecimiento de las mismas y realizar acciones para corregir o reducir su impacto. Dicha evaluación se realiza teniendo en consideración los riesgos del proyecto (identificados en la fase de iniciación) y el plan de acción, que junto con los controles de estado y el registro de mediciones se identifican desvíos en lo planificado, permitiendo realizar acciones correctivas, o el acontecimiento de una situación de riesgo para el desarrollo predefinido del proyecto realizando las acciones previstas en caso de contingencia (plan de contingencias). Como resultado de la técnica se debe dejar registro de la situación acontecida, detallando la misma y las acciones a realizar, junto con la asignación de un responsable. Finalmente, se deberá registrar la fecha en la cual se concluyeron las acciones previstas alcanzando el efecto esperado.

En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo para la prueba de concepto.

4.3.4.2.3. Ejecución de la Actividad en la Prueba de Concepto

En esta sección, se presentan los resultados obtenidos de aplicar la técnica Evaluación de Riesgos, la cual utiliza como insumos los formalismos: Riesgos del Proyecto (Tabla 4.7), Plan de Contingencias (Tabla 4.8), Plan de Acción (Tabla 4.24), Registro de Mediciones (Tabla 4.32) y Reporte de Estado (Tabla 4.39). Dichos formalismos son transcriptos con el mismo número de tabla, para facilitar al lector en la comprensión de la aplicación de la técnica.

Riesgos del Proyecto			
Responsable:	Rodriguez H.	**Fecha:**	05/04/2016
ID#:	G.In.EIP.RiPr	**Versión:**	1.0
Riesgo	**Descripción**	**Alcance**	**Referencia**
risk.1	No contar con la presencia del experto en etapas críticas de necesidad de interacción con el mismo	proyecto	

Tabla 4.7 (Transcripta). Prueba de Concepto - Riesgos del Proyecto

Plan de Contingencias			
Responsable:	Rodriguez H.	**Fecha:**	05/04/2016
ID#:	G.In.EIP.PCon	**Versión:**	1.0
Contingencia	**Acción**	**Riesgo asociado**	**Referencia**
cont.1	Ajustes en los plazos del proyecto	(risk.1) No contar con la presencia del experto en etapas críticas de necesidad de interacción con el mismo	

Tabla 4.8 (Transcripta). Prueba de Concepto - Plan de Contingencias

Plan de Acción								
Responsable:		Rodriguez H.			**Fecha:**		15/06/16	
ID#:		G.Pl.PlA.PlAc			**Versión:**		1.3	
ID Actividad	**Actividad**	**Inicio Estimado**	**Inicio Real**	**Fin Estimado**	**Fin Real**	**Esfuerzo Estimado (Hs)**	**Esfuerzo Real (Hs)**	**Comentarios**
G.In	**Iniciación**	**04/04/16**	**04/04/16**	**22/04/16**	**22/04/16**	**12**	**11**	
G.In.EIP	Exploración Inicial del Proyecto	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DeC	Definición de la Comunicación	04/04/16	04/04/16	18/04/16	18/04/16	2	2	
G.In.EvS	Evaluación de la Situación	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DCV	Definición del Ciclo de Vida	19/04/16	19/04/16	22/04/16	22/04/16	2	1	
G.Pl	**Planificación**	**04/04/16**	**04/04/16**	**29/04/16**	**29/04/16**	**10**	**9**	
G.Pl.PIM	Planificación de la Mediciones	04/04/16	04/04/16	29/04/16	29/04/16	2	1	
G.Pl.PlA	Planificación de las Actividades	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.Pl.PlR	Planificación de los Recursos	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.Pl.PRe	Planificación de las Responsabilidades	04/04/16	04/04/16	29/04/16	29/04/16	4	4	

Tabla 4.24.a (Transcripta). Prueba de Concepto - Plan de Acción (fin del proyecto)

G.So	**Soporte**	**20/04/16**	**20/04/16**	**15/06/16**	**15/06/16**	**8**	**12**	
G.So.MeP	Mediciones del Proyecto	20/04/16	20/04/16	15/06/16	15/06/16	4	6	Se prevé el registro de las métricas de manera mensual
G.So.GeC	Gestión de la Configuración	20/04/16	20/04/16	15/06/16	15/06/16	4	6	
G.Co	**Control**	**20/04/16**	**20/04/16**	**13/06/16**	**13/06/16**	**12**	**15**	
G.Co.GeD	Gestión del Desarrollo	20/04/16	20/04/16	13/06/16	13/06/16	4	6	Se prevé la aplicación del reporte de estado de manera mensual
G.Co.CoA	Control de las Actividades	20/04/16	20/04/16	13/06/16	13/06/16	6	8	
G.Co.Gca	Gestión del Cambio	20/04/16	20/04/16	13/06/16	13/06/16	2	1	
G.Ci	**Cierre**	**06/06/16**	**06/06/16**	**15/06/16**	**15/06/16**	**4**	**4**	
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	06/06/16	06/06/16	13/06/16	13/06/16	2	2	
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	14/06/16	14/06/16	15/06/16	15/06/16	2	2	
D.EN	**Entendimiento del Negocio**	**04/04/16**	**04/04/16**	**22/04/16**	**20/04/16**	**44**	**32**	
D.EN.AnN	Análisis del Negocio	04/04/16	04/04/16	20/04/16	20/04/16	28	22	
D.EN.CPN	Comprensión del Problema de Negocio	18/04/16	18/04/16	22/04/16	20/04/16	16	10	
D.ED	**Entendimiento de los Datos**	**25/04/16**	**25/04/16**	**06/05/16**	**06/05/16**	**56**	**52**	
D.ED.AnD	Análisis de los Datos	25/04/16	25/04/16	29/04/16	29/04/16	22	20	
D.ED.ExD	Exploración de los Datos	28/04/16	28/04/16	06/05/16	06/05/16	22	22	
D.ED.EvD	Evaluación de los Datos	02/05/16	02/05/16	06/05/16	06/05/16	12	10	
D.Mo	**Modelado**	**09/05/16**	**09/05/16**	**13/05/16**	**13/05/16**	**26**	**16**	
D.Mo.MoP	Modelado del problema	09/05/16	09/05/16	10/05/16	10/05/16	14	8	
D.Mo.CoM	Configuración del Modelo	11/05/16	11/05/16	13/05/16	13/05/16	12	8	
D.PD	**Preparación de los Datos**	**16/05/16**	**16/05/16**	**20/05/16**	**20/05/16**	**36**	**28**	
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	16/05/16	16/05/16	18/05/16	18/05/16	20	16	
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos	17/05/16	17/05/16	20/05/16	20/05/16	16	12	
D.Im	**Implementación**	**23/05/16**	**23/05/16**	**27/05/16**	**27/05/16**	**30**	**26**	
D.Im.SeM	Selección del Modelo	23/05/16	23/05/16	24/05/16	24/05/16	8	6	
D.Im.ExI	Explotación de Información	24/05/16	24/05/16	27/05/16	27/05/16	22	20	
D.EP	**Evaluación y Presentación**	**30/05/16**	**30/05/16**	**13/06/16**	**13/06/16**	**26**	**18**	
D.EP.EvR	Evaluación de los Resultados	30/05/16	30/05/16	02/06/16	02/06/16	10	6	
D.EP.PrR	Presentación de los Resultados	01/06/16	01/06/16	13/06/16	13/06/16	16	12	

Tabla 4.24.b (Transcripta). Prueba de Concepto - Plan de Acción (fin del proyecto)

Registro de Mediciones			
Responsable:	Esposito E.	**Fecha:**	15/06/2016
ID#:	G.So.MeP.ReMe	**Versión:**	1.2

Indicador	**Medición**	**Descripción**
Tiempo total requerido para el desarrollo del proyecto	**DRPY** = 223	Tdesarrollo = 172 Tgestion = 51
Grado de Utilidad de Atributos	**GUA** = 6,71	NA = 392 NASE = 15 NAUD= 2 NO_UTILES = 275

Tabla 4.32 (Transcripta). Prueba de Concepto - Registro de Mediciones

Reporte de Estado			
Responsable:	Rodriguez H.	**Fecha:**	03/06/2016
ID#:	G.Co.GeD.ReEs.2		
Evaluación del Programa			
Global	**-19,64%**		**(% por debajo de lo planificado)**
Desarrollo	**-20,79%**	**Gestión**	**-9,09%**
Entendimiento del Negocio	-27,27%	Iniciación	-8,33%
Entendimiento de los Datos	-7,14%	Planificación	-10%
Modelado	-38,46%	Soporte	-
Preparación de los Datos	-22,22%	Control	-
Implementación	-13,33%	Cierre	-
Evaluación y Presentación	-40%		
Descripción: *Se evaluaron las actividades finalizadas hasta la fecha. (ver G.Pl.PIA.PIAc versión 1.2)*			
Situaciones identificadas que requieren de seguimiento Se identifica un incremento en el tiempo de respuesta del cliente/experto **Cambios durante el periodo (alcances, tiempos)** - **logros principales durante el periodo** Generación y validación de los resultados vinculados con el problema de negocio			

Tabla 4.39 (Transcripta). Prueba de Concepto - Reporte de Estado (G.Co.GeD.ReEs.2)

Registro de Riesgos Acontecidos (G.Co.CoA.ReRA): durante el desarrollo del proyecto seleccionado como prueba de concepto, aconteció en las etapas finales del mismo el riesgo previsto en el formalismo Riesgos del Proyecto (risk1), vinculado con la existencia de un único experto y las posibles demoras que se pudiesen ocasionar al no contar con la disponibilidad del mismo. Específicamente, durante la etapa de presentación de los resultados debiendo reajustar las fechas previstas para dicha actividad. En la tabla 4.41 se ilustra el riesgo registrado en el formalismo.

Registro de Riesgos Acontecidos					
Responsable:		Rodriguez H.	**Fecha:**		03/06/2016
ID#:		G.Co.CoA.ReRA			
Fecha	**Riesgo**	**Responsable**	**Descripción**	**Acción**	**Fecha Resolución**
03/06/16	risk.1	Rodriguez H.	Imposibilidad de realizar la presentación de los resultados en la fecha estipulada, por ausencia del cliente/experto	Posponer 1 semana las actividades pendientes	03/06/2016

Tabla 4.41. Prueba de Concepto – Registro de Riesgos Acontecidos

4.3.4.3. Actividad: Gestión del Cambio (G.Co.Gca)

Los proyectos de explotación de información, de manera similar a otros tipos de proyectos, presentan un entorno cambiante (por distintas razones como: el cliente desconoce los logros que se

pueden obtener mediante el uso de ingeniería de explotación de información, la aparición de información que genera nuevas inquietudes o necesidades, entre otros) en el cual las modificaciones al proyecto deben ser aceptadas, pero las mismas deben tener un coste para el cliente, no solo respecto a lo económico, sino en cuanto al trabajo asociado con el análisis de la oportunidad del mismo y su factibilidad, para evitar continuos cambios en el proyecto que impidan o dificulten el progreso del mismo.

En este contexto, en la actividad de Gestión del Cambio se realiza un proceso de evaluación formal de las peticiones de cambio, determinando como resultado la procedencia o no del mismo y sus efectos asociados.

Información de Entrada

- Solicitud de cambio (Externo)
- Propuesta del Proyecto (G.PI.PIR.PrPr)

Información de Salida

- Reporte de Evaluación del Cambio (G.Co.Gca.RECa)

4.3.4.3.1. Formalismos Identificados

La salida esperada para esta actividad se documenta mediante el formalismo Reporte de Evaluación del Cambio, el cual se presenta a continuación.

Reporte de Evaluación del Cambio (G.Co.Gca.RECa): se deja registro formal de las peticiones de cambios realizadas por los interesados, indicando la fecha, el pedido de cambio, la/s personas que realizan el pedido, una descripción detallada del pedido y el efecto que el mismo tiene con respecto al proyecto (asentando cada campo en su columna homónima). Además, se registra el resultado de la evaluación del pedido (Aprobado/Rechazado) y el responsable asignado. La tabla 4.42 ilustra la estructura del formalismo previamente descripto.

Reporte de Evaluación del Cambio						
Responsable:	Persona a cargo			**Fecha:**	Fecha de Realización	
ID#:	Identificador del Producto					
Fecha	**Cambio**	**Pedido por**	**Descripción**	**Impacto**	**Estado**	**Responsable**
Fecha pedido de cambio 1	pedido de cambio 1	Persona que realiza el pedido de cambio 1	Detalle del pedido de cambio 1	Efecto del pedido de cambio 1	Resolución del pedido de cambio 1	Responsable de la resolución del pedido de cambio 1
…	…	…	…	…	…	…
Fecha pedido de cambio N	pedido de cambio N	Persona que realiza el pedido de cambio N	Detalle del pedido de cambio N	Efecto del pedido de cambio N	Resolución del pedido de cambio N	Responsable de la resolución del pedido de cambio N

Tabla 4.42. Formalismo: Reporte de Evaluación del Cambio

4.3.4.3.2. Técnica Identificada

Para el desarrollo de esta actividad se propone la técnica **"Evaluación del Cambio"**, mediante la cual se evalúan las peticiones de cambios realizadas tanto por los interesados del proyecto de la organización cliente, así como por los miembros del equipo de explotación de información, con el objetivo de analizar la naturaleza de la petición y determinar el efecto que el mismo tiene con respecto a los objetivos del proyecto, el programa, las obligaciones acordadas y la posibilidad de introducir nuevos riesgos. Todo pedido de cambio debe ser respondido con la aprobación o rechazo, dejando previamente definido la persona responsable de realizar las evaluaciones (siendo necesario detallar el motivo del rechazo).

Los cambios aprobados pueden requerir de ajustes en la planificación del proyecto (plazos, recursos y responsables), así como disparar una serie de acciones que incluyan la necesidad de ajustar la configuración del proyecto y su comunicación.

El pedido de cambio es una propuesta formal para modificar cualquier producto interno del proyecto. Los motivos a partir de los cuales surgen las peticiones de cambios son múltiples: pueden ocasionarse debido a ajustes durante el desarrollo de una actividad, situaciones identificadas en controles, o peticiones formales de cambio en las necesidades del proyecto. Sin embargo, no todo pedido debe ser requerido formalmente (evaluado por un miembro externo a la actividad), dado que esto impactaría negativamente en la agilidad y el desarrollo continúo del proyecto, sino que se debe definir alguna línea base que permita determinar las mejoras en el proceso.

En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo para la prueba de concepto.

4.3.4.3.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica Evaluación del Cambio, la cual utiliza como insumos la solicitud de cambio (Fuente de Información 4.2) y la Propuesta del Proyecto (Tabla 4.30), la cual es transcripta con el mismo número de tabla que el asignado al momento de su presentación, para facilitar al lector en la comprensión de la técnica.

El cliente solicita posponer la reunión de presentación de los resultados una semana, debido a la imposibilidad de presenciar la misma.

Fuente de Información 4.2. Prueba de Concepto - Solicitud de Cambio

Propuesta del Proyecto			
Responsable:	Rodriguez H.	**Fecha:**	29/04/2016
ID#:	G.PI.PIR.PrPr	**Versión:**	1.0
Alcance	Se establece como objetivo del proyecto analizar la población de encuestados de Encuesta Nacional sobre Prevalencias de Consumo de Sustancias Psicoactivas del 2011, con el fin de comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos. Se espera definir un proceso que automatice o semi-automatice el análisis de los datos, reduciendo el costo y tiempo asociado con la generación de resultados y el correspondiente accionar a partir de los mismos, que brinde indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales serán validadas por el cliente. Específicamente, se requiere identificar las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo.		
Obligaciones y responsabilidades	La parte contratante se compromete a brindar disposición de todos los recursos requeridos en tiempo y forma, informando con una antelación no menor a 72hs en caso que estos no pudiesen ser entregados. Cualquier demora fuera del plazo estipulado, requerirá del ajuste de los plazos del proyecto. La contraparte se compromete a dar solución a las problemáticas requeridas por el cliente (ver sección Alcance del Proyecto), así como la veracidad de los resultados presentados. Asimismo, la organización se compromete a informar de forma mensual los avances del proyecto. Las partes acuerdan como fecha de finalización del proyecto el 06/06/2016.		
Firma del Contratante: Silva H.		**Firma de la Contraparte:** Rodriguez H.	
Aclaración: Silva H.		**Aclaración:** Rodriguez H.	

Tabla 4.30 (Transcripta). Prueba de Concepto - Propuesta del Proyecto

Reporte de Evaluación del Cambio (G.Co.Gca.RECa): se evalúa y aprueba la petición de cambio del cliente, generando como impacto el ajuste de las fechas estimadas para el desarrollo de las actividades pendientes en el proyecto. La tabla 4.43 describe la evaluación del pedido de cambio, el responsable y el impacto asociado.

Reporte de Evaluación del Cambio						
Responsable:	Esposito E.		**Fecha :**	03/06/2016		
ID#:	G.Co.Gca.ReCa					
Fecha	**Cambio**	**Pedido por**	**Descripción**	**Impacto**	**Estado**	**Responsable**
03/06/16	(ReCa.1) Ajuste de plazos de hitos del proyecto	Silva H. (rehi.3)	Modificación de la fecha de cumplimiento de los hitos debido a la imposibilidad del cliente de cumplimentar con los hitos pactados	Modificación en la fecha de las actividades pendientes en una semana. La fecha inicial de Formalización Interna del Cierre del Proyecto, se cambia al 14/06/16. La fecha final de las siguientes actividades se posponen al 13/06/16: Gestión del Desarrollo, Control de las Actividades, Gestión del Cambio, Formalización Externa del Cierre del Proyecto y Presentación de los Resultados, y la fecha final de las siguientes se pospone al 15/06/16: Mediciones del Proyecto, Gestión de la Configuración y Formalización Interna del Cierre del Proyecto. Se realizan ajustes en los recursos requeridos para dichas actividades.	Aprobado	Rodriguez H. (rehi.1)

Tabla 4.43. Prueba de Concepto - Reporte de Evaluación del Cambio

4.3.5. Fase: Cierre (G.Ci)

En la fase Cierre se deja registro formal de la finalización de las actividades asociadas al proyecto, evaluando los resultados obtenidos y el proceso tanto por los interesados de la organización contratante, así como de la organización propia.

En este contexto, se define a la fase como el compendio de dos actividades: Formalización Externa del Cierre del Proyecto (sección 4.3.5.1) y Formalización Interna del Cierre del Proyecto (sección 4.3.5.2). La figura 4.8, presenta resume las actividades que integran la fase y sus elementos de entrada y salida (las imágenes de cada formalismo son representaciones miniatura de los mismos)..

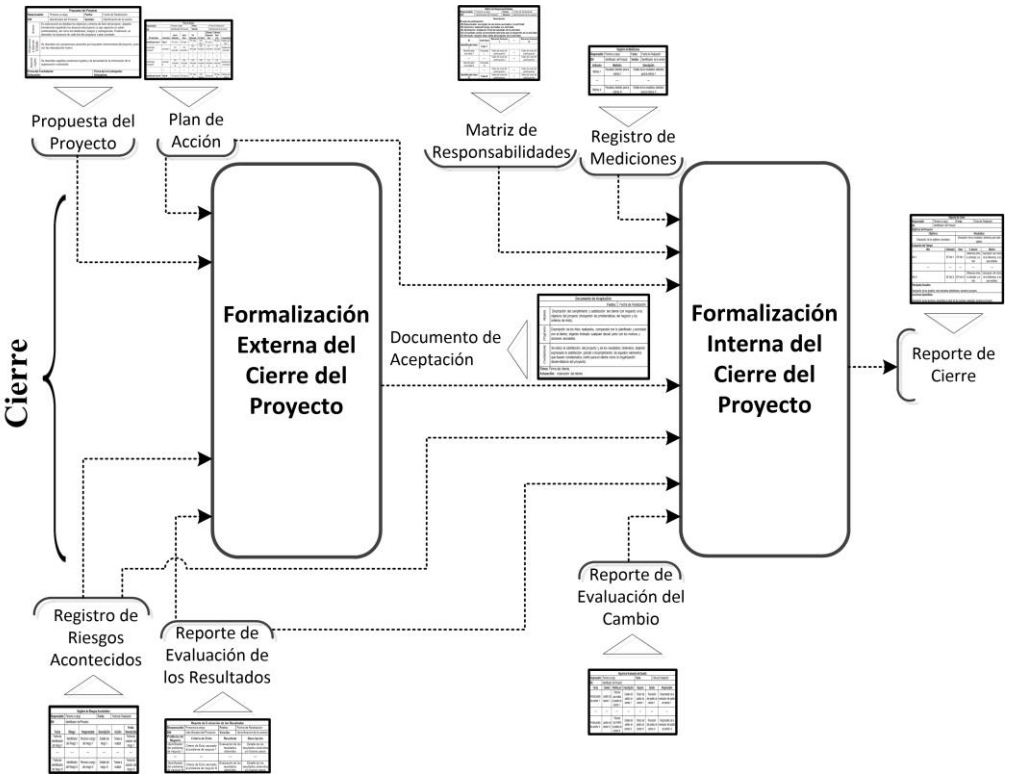


Figura 4.8. Fase Cierre

4.3.5.1. Actividad: Formalización Externa del Cierre del Proyecto (G.Ci.FEC)

En esta actividad se obtiene la conformidad del cliente, respecto a los compromisos asumidos en la propuesta del proyecto, dejando registro formal de la finalización del mismo.

Información de Entrada

- Reporte de Evaluación de los Resultados (D.EP.EvR.ReER)
- Registro de Riesgos Acontecidos (G.Co.CoA.ReRA)
- Plan de Acción (G.Pl.PlA.PlAc)
- Propuesta del Proyecto (G.Pl.PlR.PrPr)

Información de Salida

- Documento de Aceptación (G.Ci.FEC.DoAc)

4.3.5.1.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se propone el Documento de Aceptación, el cual se presenta a continuación.

Documento de Aceptación (G.Ci.FEC.DoAc): se resume los resultados derivados del desarrollo del proyecto, contemplando tres aspectos globales:

- *Alcance:* donde se describen las metas del proyecto acordadas (objetivos del proyecto, problemas de negocio y sus criterios de éxito), los logros alcanzados y la evaluación de los expertos respecto al interés de los resultados.
- *Programa*: detallando el progreso del proyecto y el cumplimiento de los hitos acordados, indicando desvíos acontecidos (si hubiese), las razones que derivaron en dicha situación y las acciones realizadas.
- *Conclusiones:* se resume las resoluciones del proyecto, expresando la satisfacción de las obligaciones acordadas entre las partes (pudiendo la misma ser parcial o no satisfecha).

En la parte final del formalismo se acepta el cumplimiento de los compromisos asumidos entre las partes. En la tabla 4.44 se ilustra la estructura del formalismo propuesto.

Documento de Aceptación	
	Fecha: Fecha de Realización
Alcance	Descripción del cumplimiento y satisfacción del cliente con respecto a los objetivos del proyecto (incluyendo las problemáticas del negocio y los criterios de éxito).
Programa	Descripción de los hitos realizados, comparado con lo planificado y acordado con el cliente, dejando indicado cualquier desvió junto con los motivos y acciones asociadas.
Conclusiones	Se indica la satisfacción del proyecto y de los resultados obtenidos, dejando expresada la satisfacción parcial o incumplimiento de aquellos elementos que fuesen considerados (tanto para el cliente como la organización desarrolladora del proyecto).
Firma: Firma del cliente **Aclaración:** Aclaración del cliente	

Tabla 4.44. Formalismo: Documento de Aceptación

4.3.5.1.2. Técnica Identificada

Para el desarrollo de esta actividad se propone la técnica **"Presentación de Conformidad"**, mediante la cual se deja registro formal de que las necesidades requeridas por el cliente (y acordadas en la propuesta del proyecto) han sido aprobadas por el requirente, dejando constancia de cualquier contravención acontecida durante el desarrollo del proyecto, junto con las decisiones y

acciones aplicadas. La descripción completa del proceso, se realiza en tres dimensiones: en la primera, se describen los objetivos del proyecto junto con el criterio de evaluación y los resultados establecidos de acuerdo a los criterios de éxito definidos, los alcances definidos del proyecto y los resultados obtenidos en la ejecución del mismo.

En la segunda dimensión, se describe el proceso realizado para alcanzar los objetivos descriptos previamente, evaluando el desempeño de los hitos realizados con respecto a los acordados entre las partes, informando (si fuese necesario) los motivos del desvío junto con las acciones asociadas a dicho evento. Finalmente, en la tercera dimensión se resume el resultado de los compromisos asumidos (indicando las acciones derivadas para aquellos que no fueron completamente satisfechos).

Como resultado de la aplicación de la técnica, se formaliza el cumplimiento de las necesidades del cliente, estableciendo la finalización exitosa del proyecto. En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo para la prueba de concepto.

4.3.5.1.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica Presentación de Conformidad, la cual utiliza como insumos los formalismos: Reporte de Evaluación de los Resultados (Tabla 4.91), Plan de Acción (Tabla 4.24), Propuesta del Proyecto (Tabla 4.30) y Registro de Riesgos Acontecidos (Tabla 4.41).

Los formalismos indicados como elementos de entrada, son transcriptos con el mismo número de tabla, para facilitar al lector en la comprensión de la aplicación de la técnica.

Reporte de Evaluación de los Resultados			
Responsable:	Rodriguez H.	**Fecha:**	02/06/2016
ID#:	D.EP.EvR.ReER	**Versión:**	1.0
Problema de Negocio	**Criterio de Éxito**	**Resultado**	**Descripción**
(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo	(cepn.1) Identificar aspectos que permitan comprender el comportamiento de grupos masivos de personas de acuerdo a la evaluación de Silva H. (rehi.3)	Satisfactorio	Las reglas identificadas permiten comprender los aspectos generales de la población estudiada. Se señala que los resultados obtenidos pueden estar dispersos por el consumo de algunos tipos de sustancias psicoactivas que se encuentran con menor presencia en la población encuestada, siendo de interés profundizar en el estudio del comportamiento de la población mediante un análisis geo-referencial.

Tabla 4.91 (Transcripta). Prueba de Concepto - Reporte de Evaluación de los Resultados

Plan de Acción								
Responsable:		Rodriguez H.			**Fecha:**		15/06/16	
ID#:		G.PI.PIA.PIAc			**Versión:**		1.3	
ID Actividad	**Actividad**	**Inicio Estimado**	**Inicio Real**	**Fin Estimado**	**Fin Real**	**Esfuerzo Estimado (Hs)**	**Esfuerzo Real (Hs)**	**Comentarios**
G.In	**Iniciación**	**04/04/16**	**04/04/16**	**22/04/16**	**22/04/16**	**12**	**11**	
G.In.EIP	Exploración Inicial del Proyecto	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DeC	Definición de la Comunicación	04/04/16	04/04/16	18/04/16	18/04/16	2	2	
G.In.EvS	Evaluación de la Situación	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DCV	Definición del Ciclo de Vida	19/04/16	19/04/16	22/04/16	22/04/16	2	1	
G.Pl	**Planificación**	**04/04/16**	**04/04/16**	**29/04/16**	**29/04/16**	**10**	**9**	
G.Pl.PlM	Planificación de la Mediciones	04/04/16	04/04/16	29/04/16	29/04/16	2	1	
G.Pl.PlA	Planificación de las Actividades	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.Pl.PlR	Planificación de los Recursos	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.Pl.PRe	Planificación de las Responsabilidades	04/04/16	04/04/16	29/04/16	29/04/16	4	4	
G.So	**Soporte**	**20/04/16**	**20/04/16**	**15/06/16**	**15/06/16**	**8**	**12**	
G.So.MeP	Mediciones del Proyecto	20/04/16	20/04/16	15/06/16	15/06/16	4	6	Se prevé el registro de las métricas de manera mensual
G.So.GeC	Gestión de la Configuración	20/04/16	20/04/16	15/06/16	15/06/16	4	6	
G.Co	**Control**	**20/04/16**	**20/04/16**	**13/06/16**	**13/06/16**	**12**	**15**	
G.Co.GeD	Gestión del Desarrollo	20/04/16	20/04/16	13/06/16	13/06/16	4	6	Se prevé la aplicación del reporte de estado de manera mensual
G.Co.CoA	Control de las Actividades	20/04/16	20/04/16	13/06/16	13/06/16	6	8	
G.Co.Gca	Gestión del Cambio	20/04/16	20/04/16	13/06/16	13/06/16	2	1	
G.Ci	**Cierre**	**06/06/16**	**06/06/16**	**15/06/16**	**15/06/16**	**4**	**4**	
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	06/06/16	06/06/16	13/06/16	13/06/16	2	2	
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	14/06/16	14/06/16	15/06/16	15/06/16	2	2	
D.EN	**Entendimiento del Negocio**	**04/04/16**	**04/04/16**	**22/04/16**	**20/04/16**	**44**	**32**	
D.EN.AnN	Análisis del Negocio	04/04/16	04/04/16	20/04/16	20/04/16	28	22	
D.EN.CPN	Comprensión del Problema de Negocio	18/04/16	18/04/16	22/04/16	20/04/16	16	10	
D.ED	**Entendimiento de los Datos**	**25/04/16**	**25/04/16**	**06/05/16**	**06/05/16**	**56**	**52**	
D.ED.AnD	Análisis de los Datos	25/04/16	25/04/16	29/04/16	29/04/16	22	20	
D.ED.ExD	Exploración de los Datos	28/04/16	28/04/16	06/05/16	06/05/16	22	22	
D.ED.EvD	Evaluación de los Datos	02/05/16	02/05/16	06/05/16	06/05/16	12	10	

Tabla 4.24.a (Transcripta). Prueba de Concepto - Plan de Acción (fin del proyecto)

D.Mo	**Modelado**	**09/05/16**	**09/05/16**	**13/05/16**	**13/05/16**	**26**	**16**	
D.Mo.MoP	Modelado del problema	09/05/16	09/05/16	10/05/16	10/05/16	14	8	
D.Mo.CoM	Configuración del Modelo	11/05/16	11/05/16	13/05/16	13/05/16	12	8	
D.PD	**Preparación de los Datos**	**16/05/16**	**16/05/16**	**20/05/16**	**20/05/16**	**36**	**28**	
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	16/05/16	16/05/16	18/05/16	18/05/16	20	16	
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos	17/05/16	17/05/16	20/05/16	20/05/16	16	12	
D.Im	**Implementación**	**23/05/16**	**23/05/16**	**27/05/16**	**27/05/16**	**30**	**26**	
D.Im.SeM	Selección del Modelo	23/05/16	23/05/16	24/05/16	24/05/16	8	6	
D.Im.ExI	Explotación de Información	24/05/16	24/05/16	27/05/16	27/05/16	22	20	
D.EP	**Evaluación y Presentación**	**30/05/16**	**30/05/16**	**13/06/16**	**13/06/16**	**26**	**18**	
D.EP.EvR	Evaluación de los Resultados	30/05/16	30/05/16	02/06/16	02/06/16	10	6	
D.EP.PrR	Presentación de los Resultados	01/06/16	01/06/16	13/06/16	13/06/16	16	12	

Tabla 4.24.b (Transcripta). Prueba de Concepto - Plan de Acción (fin del proyecto)

Propuesta del Proyecto			
Responsable:	Rodriguez H.	**Fecha:**	29/04/2016
ID#:	G.PI.PIR.PrPr	**Versión:**	1.0
Alcance	Se establece como objetivo del proyecto analizar la población de encuestados de Encuesta Nacional sobre Prevalencias de Consumo de Sustancias Psicoactivas del 2011, con el fin de comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos. Se espera definir un proceso que automatice o semi-automatice el análisis de los datos, reduciendo el costo y tiempo asociado con la generación de resultados y el correspondiente accionar a partir de los mismos, que brinde indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales serán validadas por el cliente. Específicamente, se requiere identificar las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo.		
Obligaciones y responsabilidades	La parte contratante se compromete a brindar disposición de todos los recursos requeridos en tiempo y forma, informando con una antelación no menor a 72hs en caso que estos no pudiesen ser entregados. Cualquier demora fuera del plazo estipulado, requerirá del ajuste de los plazos del proyecto. La contraparte se compromete a dar solución a las problemáticas requeridas por el cliente (ver sección Alcance del Proyecto), así como la veracidad de los resultados presentados. Asimismo, la organización se compromete a informar de forma mensual los avances del proyecto. Las partes acuerdan como fecha de finalización del proyecto el 06/06/2016.		
Firma del Contratante: Silva H.		**Firma de la Contraparte:** Rodriguez H.	
Aclaración: Silva H.		**Aclaración:** Rodriguez H.	

Tabla 4.30 (Transcripta). Prueba de Concepto - Propuesta del Proyecto

Registro de Riesgos Acontecidos					
Responsable:		Rodriguez H.	**Fecha:**		03/06/2016
ID#:		G.Co.CoA.ReRA			
Fecha	**Riesgo**	**Responsable**	**Descripción**	**Acción**	**Fecha Resolución**
03/06/16	risk.1	Rodriguez H.	Imposibilidad de realizar la presentación de los resultados en la fecha estipulada, por ausencia del cliente/experto	Posponer 1 semana las actividades pendientes	03/06/2016

Tabla 4.41 (Transcripta). Prueba de Concepto - Registro de Riesgos Acontecidos

Documento de Aceptación (G.Ci.FEC.DoAc): en el proyecto, se deriva a partir de los objetivos, las problemáticas identificadas y los criterios de éxito, los siguientes alcances del proyecto: "*Los objetivos definidos consisten en analizar la población de encuestados de Encuesta Nacional sobre Prevalencias de Consumo de Sustancias Psicoactivas del 2011, con el fin de identificar reglas que permitan comprender las condiciones (sociodemográficas, socioeconómicas, educativas y del entorno familiar social, considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo. Como resultado de la evaluación realizada por el experto Silva. H., se concluyó que: „Las reglas identificadas son de interés y permiten comprender los aspectos generales de la población estudiada.*"".

A partir de lo registrado en el plan de acción y el riesgo acontecido, se describe el programa llevado a cabo y su desvío con respecto a la planificación inicial: "*Se realizaron dos informes del progreso de las actividades de manera mensual (en las fechas 06/05/2016 y 03/06/2016) y se pautó como fecha de entrega del proyecto el día 06/06/16, la cual debió ser pospuesta al 13/06/2016 por pedido de Silva H."* y se describen como conclusiones de los resultados del proyecto: "*Las obligaciones acordadas en el contrato del proyecto fueron llevadas a cabo con un desvío de 7 días respecto al tiempo pactado a causa de la imposibilidad de llevar a cabo la reunión de presentación de los resultados pactada para el 06/06/2016, siendo realizada el 13/06/2016. Mediante la presente se deja de manifiesto que se ha cumplimentado exitosamente los requerimientos realizados, dando por finalizado el proyecto.*".

Finalmente, el cliente certifica el cumplimiento exitoso del proyecto. En la tabla 4.45 se ilustra el formalismo generado.

Documento de Aceptación	
	Fecha: 13/06/2016
Objetivos	Los objetivos definidos consisten en analizar la población de encuestados de Encuesta Nacional sobre Prevalencias de Consumo de Sustancias Psicoactivas del 2011, con el fin de identificar reglas que permitan comprender las condiciones (sociodemográficas, socioeconómicas, educativas y del entorno familiar social, considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo. Como resultado de la evaluación realizada por el experto Silva. H., se concluyó que: "Las reglas identificadas son de interés y permiten comprender los aspectos generales de la población estudiada."
Programa	Se realizaron dos informes del progreso de las actividades de manera mensual (en las fechas 06/05/2016 y 03/06/2016) y se pautó como fecha de entrega del proyecto el día 06/06/16, la cual debió ser pospuesta al 13/06/2016 por pedido de Silva H.
Conclusiones	Las obligaciones acordadas en el contrato del proyecto fueron llevadas a cabo con un desvío de 7 días respecto al tiempo pactado a causa de la imposibilidad de llevar a cabo la reunión de presentación de los resultados pactada para el 06/06/2016, siendo realizada el 13/06/2016. Mediante la presente se deja de manifiesto que se ha cumplimentado exitosamente los requerimientos realizados, dando por finalizado el proyecto.
Firma: Silva H. **Aclaración:** Silva H.	

Tabla 4.45. Prueba de Concepto – Documento de Aceptación

4.3.5.2. Actividad: Formalización Interna del Cierre del Proyecto (G.Ci.FIC)

En esta actividad se llevan a cabo las últimas tareas del proyecto, en el cual se evalúa la performance del equipo de trabajo, la propuesta, las acciones realizadas y el cumplimiento del plan de acción. Como resultado de esta actividad se resume el progreso del proyecto, dejando registro de aquellos aspectos que sean de valor para proyectos futuros.

Información de Entrada

- Plan de Acción (G.Pl.PlA.PlAc)
- Matriz de Responsabilidades (G.Pl.PlR.MaRe)
- Registro de Mediciones (G.So.MeP.ReMe)
- Registro de Riesgos Acontecidos (G.Co.CoA.ReRA)
- Reporte de Evaluación del Cambio (G.Co.Gca.RECa)
- Reporte de Evaluación de los Resultados (D.EP.EvR.ReER)
- Documento de Aceptación (G.Ci.FEC.DoAc)

Información de Salida

- Reporte de Cierre (G.Ci.FIC.ReCi)

4.3.5.2.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se propone el Reporte de Cierre, el cual se presenta a continuación.

Reporte de Cierre (G.Ci.FIC.ReCi): Se presenta un resumen formal del proyecto, basado en los propuestos en [Project Management Institute, Inc., 2013a; Verzuh, 2015] (sección 2.4.2.5, pág. 47), el cual se estructura en cinco secciones: en la primera de ellas, se registran los objetivos pactados con el cliente y los resultados obtenidos, en la segunda se listan el tiempo estimado para cada etapa del proyecto, junto con los valores reales, haciendo un análisis del porcentaje de diferencia entre ambos y registrando el motivo que derivó en la diferencia identificada. Los hitos a evaluar en la sección "Evaluación del tiempo" varían de acuerdo al modelo de ciclo de vida seleccionado y el flujo e iteración de las actividades, y los alcances de las mismas (es decir, la estrategia determinada en el plan de acción). Finalmente, se listan los principales desafíos y aprendizajes realizados durante el proyecto (en las secciones "Principales Desafíos" y "Lecciones Aprendidas"). La tabla 4.46 ilustra la estructura del formalismo previamente descripto.

Reporte de Cierre				
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización	
ID#:	Identificador del Producto			
Objetivos del Proyecto				
Objetivos		**Resultados**		
Descripción de los objetivos acordados		Descripción de los resultados obtenidos para cada objetivo		
Evaluación del Tiempo				
Hito	**Estimado**	**Real**	**% desvío**	**Motivo**
Hito 1	EE hito 1	ER hito 1	Diferencia entre lo estimado y lo real	Descripción del motivo de la diferencia, si es que existiese.
…	…	…	…	…
Hito N	EE hito N	ER hito N	Diferencia entre lo estimado y lo real	Descripción del motivo de la diferencia, si es que existiese.
Principales Desafíos Descripción de los desafíos más relevantes identificados durante el proyecto				
Lecciones Aprendidas Descripción de las lecciones aprendidas a partir de las acciones realizadas durante el proyecto				

Tabla 4.46. Formalismo: Reporte de Cierre

4.3.5.2.2. Técnica Identificada

Para el desarrollo de esta actividad se propone la técnica **"Evaluación del Proceso"**, mediante la cual se deja registro de la experiencia realizada por el equipo de trabajo, documentando los conocimientos relevantes del proceso, y registrando lecciones aprendidas en el proyecto para su posterior uso en la organización. Dicho conocimiento permite comprender los distintos aspectos del proceso que fueron realizados de manera correcta e incorrecta, identificando los posibles aspectos a mejorar. Adicionalmente, el registro de las lecciones aprendidas del proyecto (junto con la aplicación de una misma metodología o proceso), permite a la organización contar con el historial

de trabajos evitando cometer los mismos errores, así como proveer de referencias para futuros proyectos.

El proceso de evaluación, usualmente en responsabilidad del líder del proyecto, debe realizarse de manera conjunta con los miembros del equipo, consultando respecto de su experiencia en el proyecto, las dificultades y facilidades experimentadas, aspectos que pudieron hacerse mejor y aquellos que dieron soporte o facilitaron sus tareas. Además, en caso de ser posible, contar con la experiencia de aquellos clientes/expertos relevantes (es decir, que han formado parte activa del proceso o que sean usuarios directos de los resultados obtenidos), consultando respecto de su experiencia en el proyecto y su nivel de satisfacción los resultados.

El responsable de la actividad debe garantizar que las opiniones obtenidas cubran todos los aspectos del proyecto (incluyendo las visiones de los distintos roles) y posibilitando la libertad de expresar las distintas perspectivas de los miembros (evitando que se individualicen las críticas realizadas o que sean omitidas), permitiendo de esta forma sintetizar de la manera más parcial y representativa posible los resultados del desarrollo del proyecto.

En esta técnica se cubren los siguientes aspectos:

- Se evalúan los objetivos iniciales del proyecto y los resultados obtenidos (incluyendo limitaciones o dificultades en los mismos), identificando aquellos aspectos que pudieron ser mejorados desde el inicio del proyecto (si hubiese), así como decisiones correctamente tomadas. Dicha información se deriva del Documento de Aceptación, el Reporte de Evaluación de los Resultados y la interacción con los clientes/expertos.
- Se evalúan las estimaciones realizadas respecto a los resultados obtenidos, identificándose aquellas cuestiones asociadas con el desvío del programa. Esta información se obtiene y respalda de los formalismos: Plan de Acción, Registro de Mediciones, Reporte de Evaluación del Cambio, Registro de Riesgos Acontecidos y la interacción con los miembros del equipo.
- Finalmente, se resume aquellas dificultades enfrentadas durante el desarrollo del proyecto (como por ejemplo, riesgos acontecidos, cambios de las necesidades o cambios en la planificación del proyecto) y aquellas lecciones aprendidas (respecto al progreso del proyecto, los recursos utilizados y las decisiones realizadas), derivadas a partir del Plan de Acción, Reporte de Evaluación del Cambio, Registro de Riesgos Acontecidos y la interacción con los interesados (miembros del equipo y clientes/expertos).

La matriz de responsabilidades brinda una visión general de los miembros encargados y vinculados con cada actividad, identificando aquellas personas que serán afectadas a la tarea actual.

En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo para la prueba de concepto.

4.3.5.2.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica Evaluación del Proceso. Dicha técnica utiliza como insumos los formalismos: Plan de Acción (Tabla 4.24), Matriz de Responsabilidades (Tablas 4.29.a y 4.29.b), Registro de Mediciones (Tabla 4.32), Registro de Riesgos Acontecidos (Tabla 4.41), Reporte de Evaluación del Cambio (Tabla 4.43), Reporte de Evaluación de los Resultados (Tabla 4.91) y Documento de Aceptación (Tabla 4.45).

Los formalismos previamente indicados como elementos de entrada, son transcriptos con el mismo número de tabla, para facilitar al lector en la comprensión de la técnica.

Plan de Acción								
Responsable:		Rodriguez H.			**Fecha:**		15/06/16	
ID#:		G.PI.PIA.PIAc			**Versión:**		1.3	
ID Actividad	**Actividad**	**Inicio Estimado**	**Inicio Real**	**Fin Estimado**	**Fin Real**	**Esfuerzo Estimado (Hs)**	**Esfuerzo Real (Hs)**	**Comentarios**
G.In	**Iniciación**	**04/04/16**	**04/04/16**	**22/04/16**	**22/04/16**	**12**	**11**	
G.In.EIP	Exploración Inicial del Proyecto	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DeC	Definición de la Comunicación	04/04/16	04/04/16	18/04/16	18/04/16	2	2	
G.In.EvS	Evaluación de la Situación	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DCV	Definición del Ciclo de Vida	19/04/16	19/04/16	22/04/16	22/04/16	2	1	
G.Pl	**Planificación**	**04/04/16**	**04/04/16**	**29/04/16**	**29/04/16**	**10**	**9**	
G.Pl.PIM	Planificación de la Mediciones	04/04/16	04/04/16	29/04/16	29/04/16	2	1	
G.Pl.PIA	Planificación de las Actividades	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.Pl.PIR	Planificación de los Recursos	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.Pl.PRe	Planificación de las Responsabilidades	04/04/16	04/04/16	29/04/16	29/04/16	4	4	
G.So	**Soporte**	**20/04/16**	**20/04/16**	**15/06/16**	**15/06/16**	**8**	**12**	
G.So.MeP	Mediciones del Proyecto	20/04/16	20/04/16	15/06/16	15/06/16	4	6	Se prevé el registro de las métricas de manera mensual
G.So.GeC	Gestión de la Configuración	20/04/16	20/04/16	15/06/16	15/06/16	4	6	

Tabla 4.24.a (Transcripta). Prueba de Concepto - Plan de Acción (fin del proyecto)

G.Co	**Control**	**20/04/16**	**20/04/16**	**13/06/16**	**13/06/16**	**12**	**15**	
G.Co.GeD	Gestión del Desarrollo	20/04/16	20/04/16	13/06/16	13/06/16	4	6	Se prevé la aplicación del reporte de estado de manera mensual
G.Co.CoA	Control de las Actividades	20/04/16	20/04/16	13/06/16	13/06/16	6	8	
G.Co.Gca	Gestión del Cambio	20/04/16	20/04/16	13/06/16	13/06/16	2	1	
G.Ci	**Cierre**	**06/06/16**	**06/06/16**	**15/06/16**	**15/06/16**	**4**	**4**	
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	06/06/16	06/06/16	13/06/16	13/06/16	2	2	
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	14/06/16	14/06/16	15/06/16	15/06/16	2	2	
D.EN	**Entendimiento del Negocio**	**04/04/16**	**04/04/16**	**22/04/16**	**20/04/16**	**44**	**32**	
D.EN.AnN	Análisis del Negocio	04/04/16	04/04/16	20/04/16	20/04/16	28	22	
D.EN.CPN	Comprensión del Problema de Negocio	18/04/16	18/04/16	22/04/16	20/04/16	16	10	
D.ED	**Entendimiento de los Datos**	**25/04/16**	**25/04/16**	**06/05/16**	**06/05/16**	**56**	**52**	
D.ED.AnD	Análisis de los Datos	25/04/16	25/04/16	29/04/16	29/04/16	22	20	
D.ED.ExD	Exploración de los Datos	28/04/16	28/04/16	06/05/16	06/05/16	22	22	
D.ED.EvD	Evaluación de los Datos	02/05/16	02/05/16	06/05/16	06/05/16	12	10	
D.Mo	**Modelado**	**09/05/16**	**09/05/16**	**13/05/16**	**13/05/16**	**26**	**16**	
D.Mo.MoP	Modelado del problema	09/05/16	09/05/16	10/05/16	10/05/16	14	8	
D.Mo.CoM	Configuración del Modelo	11/05/16	11/05/16	13/05/16	13/05/16	12	8	
D.PD	**Preparación de los Datos**	**16/05/16**	**16/05/16**	**20/05/16**	**20/05/16**	**36**	**28**	
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	16/05/16	16/05/16	18/05/16	18/05/16	20	16	
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos	17/05/16	17/05/16	20/05/16	20/05/16	16	12	
D.Im	**Implementación**	**23/05/16**	**23/05/16**	**27/05/16**	**27/05/16**	**30**	**26**	
D.Im.SeM	Selección del Modelo	23/05/16	23/05/16	24/05/16	24/05/16	8	6	
D.Im.ExI	Explotación de Información	24/05/16	24/05/16	27/05/16	27/05/16	22	20	
D.EP	**Evaluación y Presentación**	**30/05/16**	**30/05/16**	**13/06/16**	**13/06/16**	**26**	**18**	
D.EP.EvR	Evaluación de los Resultados	30/05/16	30/05/16	02/06/16	02/06/16	10	6	
D.EP.PrR	Presentación de los Resultados	01/06/16	01/06/16	13/06/16	13/06/16	16	12	

Tabla 4.24.b (Transcripta). Prueba de Concepto - Plan de Acción (fin del proyecto)

Registro de Riesgos Acontecidos					
Responsable:		Rodriguez H.		**Fecha:**	03/06/2016
ID#:		G.Co.CoA.ReRA			
Fecha	**Riesgo**	**Responsable**	**Descripción**	**Acción**	**Fecha Resolución**
03/06/16	risk.1	Rodriguez H.	Imposibilidad de realizar la presentación de los resultados en la fecha estipulada, por ausencia del cliente/experto	Posponer 1 semana las actividades pendientes	03/06/2016

Tabla 4.41 (Transcripta). Prueba de Concepto – Registro de Riesgos Acontecidos

Matriz de Responsabilidades			
Responsable:	Rodriguez H.	**Fecha:**	25/04/2016
ID#:	G.PI.PIR.MaRe	**Versión:**	1.0

Descripción

Niveles de participación:
(R) Responsable: encargado de las tareas asociadas a la actividad.
(E) Ejecución: asignado tareas asociadas a la actividad.
(A) Aprobación: aceptación Final del resultado de la actividad.
(C) Consultado: posee conocimiento relevante para el desarrollo de la actividad.
(I) Informado: requiere estar alerta del progreso de la actividad.

ID Actividad	Actividad	Rodriguez H. (rehi.1)	Esposito E. (rehi.2)	Silva H. (rehi.3)
G.In	**Iniciación**			
G.In.EIP	Exploración Inicial del Proyecto	R	I	
G.In.DeC	Definición de la Comunicación	R	I	I
G.In.EvS	Evaluación de la Situación	R	C	I
G.In.DCV	Definición del Ciclo de Vida	R	I	
G.PI	**Planificación**			
G.PI.PIM	Planificación de la Mediciones	R	I	
G.PI.PIA	Planificación de las Actividades	R	I	
G.PI.PIR	Planificación de los Recursos	R	I	
G.PI.PRe	Planificación de las Responsabilidades	R	I	A
G.So	**Soporte**			
G.So.MeP	Mediciones del Proyecto	R	I	
G.So.GeC	Gestión de la Configuración	R	I	
G.Co	**Control**			
G.Co.GeD	Gestión del Desarrollo	R	I	I
G.Co.CoA	Control de las Actividades	R	I	I
G.Co.Gca	Gestión del Cambio	A	R	I
G.Ci	**Cierre**			
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	R		A
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	R	C	
D.EN	**Entendimiento del Negocio**			
D.EN.AnN	Análisis del Negocio	E	R	C
D.EN.CPN	Comprensión del Problema de Negocio	E	R	C
D.ED	**Entendimiento de los Datos**			
D.ED.AnD	Análisis de los Datos	E	R	C
D.ED.ExD	Exploración de los Datos	E	R	
D.ED.EvD	Evaluación de los Datos	E	R	C
D.Mo	**Modelado**			
D.Mo.MoP	Modelado del problema	E	R	I
D.Mo.CoM	Configuración del Modelo	C/A	R	I
D.PD	**Preparación de los Datos**			
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	I	R	
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos		R	
D.Im	**Implementación**			
D.Im.SeM	Selección del Modelo	C	R	
D.Im.ExI	Explotación de Información	C	R	I
D.EP	**Evaluación y Presentación**			
D.EP.EvR	Evaluación de los Resultados	R	E	C
D.EP.PrR	Presentación de los Resultados	R	E	I

Tabla 4.29 (Transcripta). Prueba de Concepto - Matriz de Responsabilidades

Registro de Mediciones			
Responsable:	Esposito E.	**Fecha:**	15/06/2016
ID#:	G.So.MeP.ReMe	**Versión:**	1.2

Indicador	Medición	Descripción
Tiempo total requerido para el desarrollo del proyecto	**DRPY** = 223	Tdesarrollo = 172 Tgestion = 51
Grado de Utilidad de Atributos	**GUA** = 6,71	NA = 392 NASE = 15 NAUD= 2 NO_UTILES = 275

Tabla 4.32 (Transcripta). Prueba de Concepto – Registro de Mediciones (fin del proyecto)

Reporte de Evaluación del Cambio						
Responsable:	Esposito E.		**Fecha :**	03/06/2016		
ID#:	G.Co.Gca.ReCa					
Fecha	**Cambio**	**Pedido por**	**Descripción**	**Impacto**	**Estado**	**Responsable**
03/06/16	(ReCa.1) Ajuste de plazos de hitos del proyecto	Silva H. (rehi.3)	Modificación de la fecha de cumplimiento de los hitos debido a la imposibilidad del cliente de cumplimentar con los hitos pactados	Modificación en la fecha de las actividades pendientes en una semana. La fecha inicial de Formalización Interna del Cierre del Proyecto, se cambia al 14/06/16. La fecha final de las siguientes actividades se posponen al 13/06/16: Gestión del Desarrollo, Control de las Actividades, Gestión del Cambio, Formalización Externa del Cierre del Proyecto y Presentación de los Resultados, y la fecha final de las siguientes se pospone al 15/06/16: Mediciones del Proyecto, Gestión de la Configuración y Formalización Interna del Cierre del Proyecto. Se realizan ajustes en los recursos requeridos para dichas actividades.	Aprobado	Rodriguez H. (rehi.1)

Tabla 4.43 (Transcripta). Prueba de Concepto – Reporte de Evaluación del Cambio

Documento de Aceptación	
	Fecha: 13/06/2016
Objetivos	Los objetivos definidos consisten en analizar la población de encuestados de Encuesta Nacional sobre Prevalencias de Consumo de Sustancias Psicoactivas del 2011, con el fin de identificar reglas que permitan comprender las condiciones (sociodemográficas, socioeconómicas, educativas y del entorno familiar social, considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo. Como resultado de la evaluación realizada por el experto Silva. H., se concluyó que: "Las reglas identificadas son de interés y permiten comprender los aspectos generales de la población estudiada."
Programa	Se realizaron dos informes del progreso de las actividades de manera mensual (en las fechas 06/05/2016 y 03/06/2016) y se pautó como fecha de entrega del proyecto el día 06/06/16, la cual debió ser pospuesta al 13/06/2016 por pedido de Silva H.
Conclusiones	Las obligaciones acordadas en el contrato del proyecto fueron llevadas a cabo con un desvío de 7 días respecto al tiempo pactado a causa de la imposibilidad de llevar a cabo la reunión de presentación de los resultados pactada para el 06/06/2016, siendo realizada el 13/06/2016. Mediante la presente se deja de manifiesto que se ha cumplimentado exitosamente los requerimientos realizados, dando por finalizado el proyecto.
Firma: Silva H. **Aclaración:** Silva H.	

Tabla 4.45 (Transcripta). Prueba de Concepto – Documento de Aceptación

Reporte de Cierre (G.Ci.FIC.ReCi): se describe el objetivo acordado y la evaluación realizada por el cliente respecto a los resultados obtenidos, y se evalúa el progreso del proyecto con respecto a los tiempos planificados al inicio del mismo, identificando desvíos de subestimación para el subproceso Gestión y sobreestimación en Desarrollo, identificando como único motivo para la fase Preparación de los Datos, el "Gran número de atributos con baja utilidad reflejada en la métrica GUA", explicando su sobreestimación. Adicionalmente, a partir de los resultados registrados en el desarrollo del proyecto, se determinan entre el equipo de trabajo los principales desafíos encarados en el proyecto y las lecciones aprendidas que el mismo dejó, identificándose dos desafíos (asociados con la estimación del esfuerzo y la validación del modelo) y una lección aprendida (vinculada con el trabajo diario del equipo). La tabla 4.47 sintetiza los resultados de interés para la organización como cierre del proyecto.

Reporte de Cierre				
Responsable:	Rodriguez H.	**Fecha:**	15/06/2016	
ID#:	G.Ci.FIC.ReCi			
Objetivos del Proyecto				
Objetivos		**Resultados**		
identificar reglas que permitan comprender las condiciones (sociodemográficas, socioeconómicas, educativas y del entorno familiar social, considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		Identificación de reglas de interés para el experto las cuales permiten comprender los aspectos generales de la población estudiada		
Evaluación del Tiempo (en HS.)				
Hito	**Estimado**	**Real**	**% desvío**	**Motivo**
Gestión	**46**	**51**	**10,87%**	
Iniciación	12	11	-8,33%	
Planificación	10	9	-10,00%	
Soporte	8	12	50,00%	
Control	12	15	25,00%	
Cierre	4	4	0,00%	
Desarrollo	**218**	**172**	**-21,10%**	
Entendimiento del Negocio	44	32	-27,27%	
Entendimiento de los Datos	56	52	-7,14%	
Modelado	26	16	-38,46%	
Preparación de los Datos	36	28	-22,22%	Gran número de atributos con baja utilidad reflejada en la métrica GUA
Implementación	30	26	-13,33%	
Evaluación y Presentación	26	18	-30,77%	
TOTAL	**264**	**223**	**-15,53%**	
Principales Desafíos - Estimación de los tiempos: el modelo utilizado para determinar la carga de trabajo infravaloró la necesidad para las actividades de gestión (en un 11% aproximadamente), mientras que el modelo del subproceso de desarrollo sobreestimó para este proyecto de tamaño pequeño en aproximadamente un 21% el esfuerzo requerido. - Imposibilidad de determinar junto con el cliente un criterio de éxito cuantitativamente verificable.				
Lecciones Aprendidas - Apoyar el desarrollo del proyecto en reuniones informales diarias evitando problemas por falta de comunicación entre los miembros del equipo				

Tabla 4.47. Prueba de Concepto – Reporte de Cierre

4.4. MoProPEI-D: Subproceso Desarrollo (D)

El subproceso de desarrollo comprende el conjunto de fases y actividades orientadas a la generación del producto resultante del proyecto, esto es, la identificación de patrones relevantes, novedosos y de calidad, así como su análisis y comprensión para la generación de piezas de conocimiento validables y de interés, que sean de valor para el proceso de toma de decisión.

La estructuración y ejecución del subproceso se define de acuerdo a la estrategia de implementación del proyecto, según las características del mismo (alcances, intereses, conocimiento y comprensión del dominio y del problema, entre otros). Es decir, la aplicación e iteración de las fases se ven definidas por el Ciclo de Vida seleccionado como mejor estrategia para el desarrollo del proyecto.

Como se mencionó previamente, la figura 4.1, presenta las relaciones y dependencias desde la perspectiva de las fases, abstrayéndose del elemento específicamente vinculado y las asociaciones internas entre las fases y actividades pertenecientes a un mismo subproceso. La figura 4.9, amplía dicho concepto ilustrando las relaciones entre las fases y las distintas actividades que componen al subproceso Desarrollo, haciendo uso del formalismo de representación propuesto en [Hossian, 2012; Rodriguez, 2015]. En el extremo superior de la figura, utilizando un elemento en forma de elipse dividido en dos partes, se presenta el nombre de la fase y la actividad (en las partes superior e inferior, respectivamente) vinculada del subproceso de Gestión. Es decir, la actividad Evaluación de la Situación, de la etapa Iniciación que pertenece al subproceso de Gestión. Las flechas representan dependencias entre las actividades (indicándose las entradas del lado izquierdo de las actividades y las salidas del lado derecho).

El subproceso Desarrollo, se encuentra conformado por seis fases: Entendimiento del Negocio (sección 4.4.1), Entendimiento de los Datos (sección 4.4.2), Modelado (sección 4.4.3), Preparación de los Datos (sección 4.4.4), Implementación (sección 4.4.5), y Evaluación y presentación (sección 4.4.6).

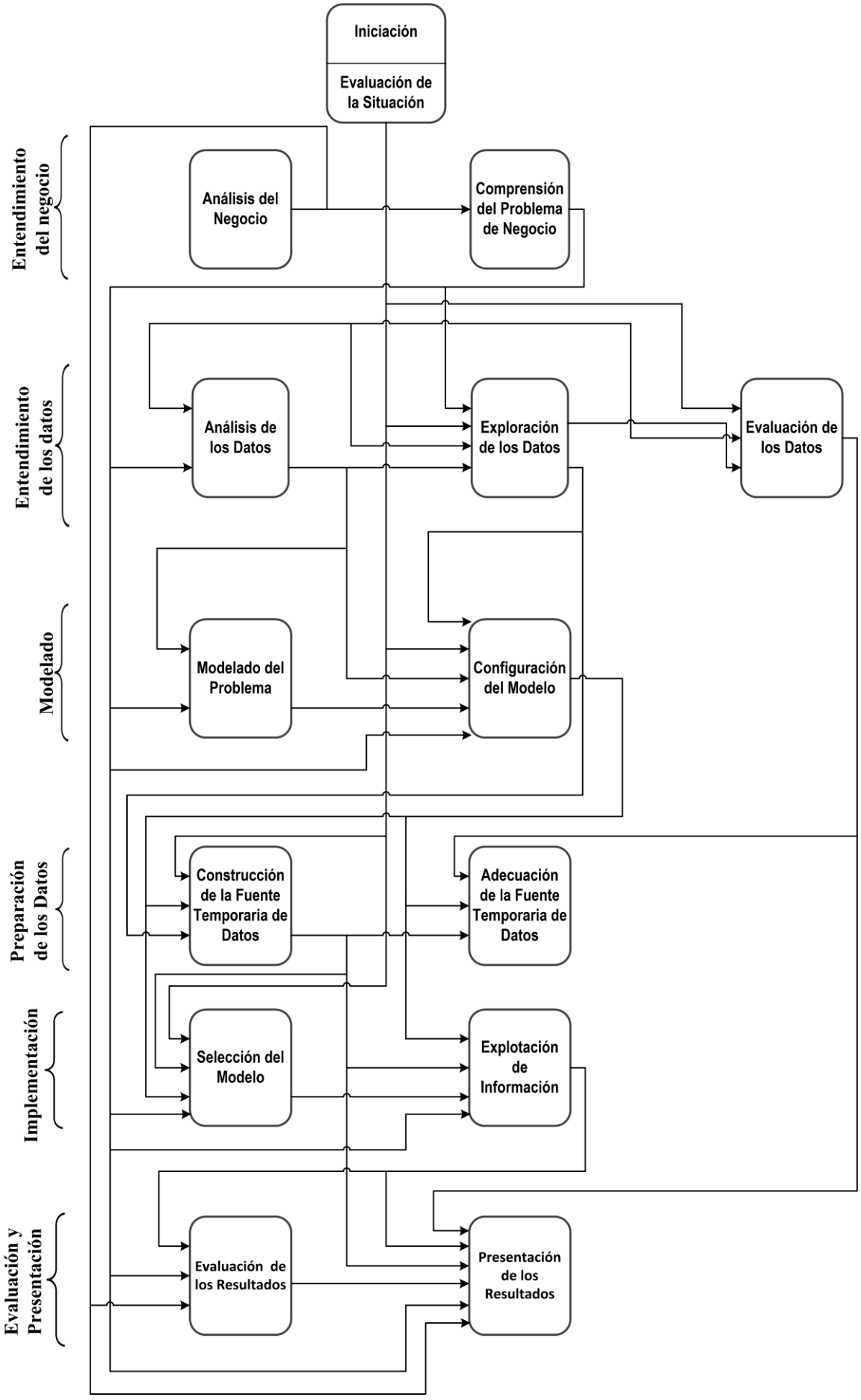


Figura 4.9. MoProPEI-D: Subproceso de Desarrollo

4.4.1. Fase: Entendimiento del Negocio (D.EN)

La fase Entendimiento del Negocio busca recolectar de las partes interesadas del proyecto (requirente, expertos involucrados en el proceso de toma de decisiones, expertos vinculados con el almacenamiento de los datos, etc.), así como en fuentes de información tanto internas como externas a la entidad requirente, aquellos conceptos relevantes que permitan comprender y establecer el marco del proyecto, los intereses de las partes involucradas y las características del mismo. Como resultado general de dicha fase se espera comprender en detalle las necesidades y las situaciones problemáticas (también referidas como pregunta-problema) a partir de las cuales se desea hacer uso de los datos disponibles para obtener piezas de información que favorezcan a dar respuesta a dichas problemáticas.

En este contexto, la fase de entendimiento del negocio se compone de dos actividades: Análisis del Negocio, donde se identifican las características generales del proyecto (sección 4.4.1.1) y Comprensión del Problema de Negocio, en la cual se establecen los problemas a resolver (sección 4.4.1.2). En la figura 4.10, se resume las dependencias (elementos de entrada y salida) de las actividades que conforman la fase actual.

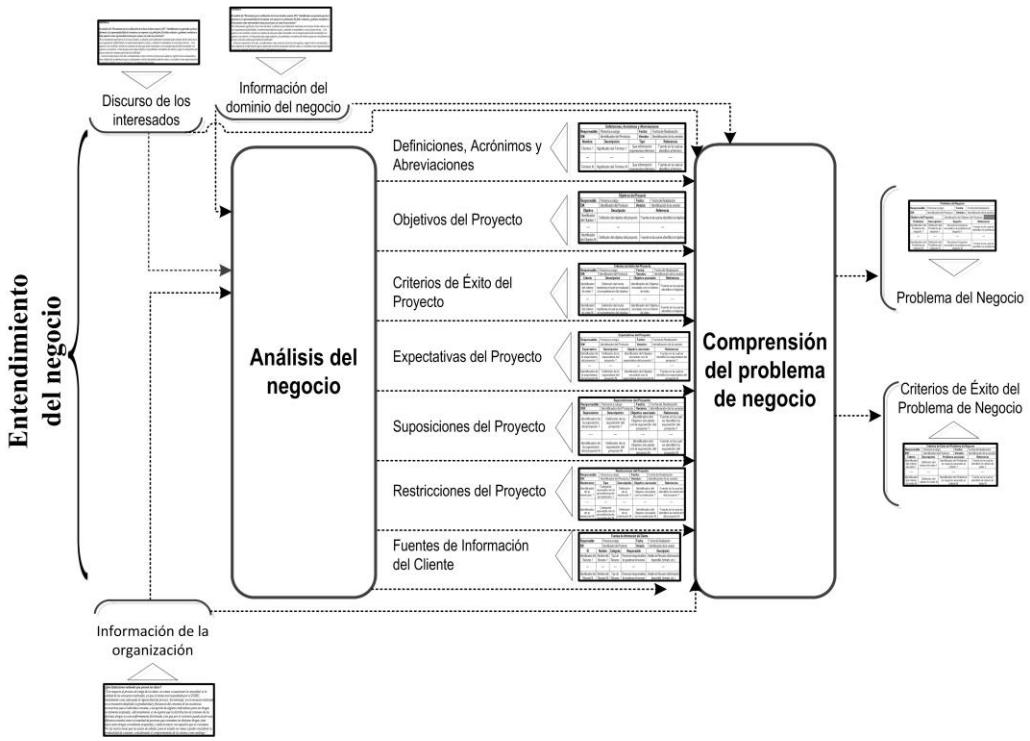


Figura 4.10. Fase: Entendimiento del Negocio

4.4.1.1. Actividad: Análisis del Negocio (D.EN.AnN)

El objetivo de esta actividad es identificar y comprender las metas del proyecto, en base a las necesidades del requirente y los interesados. Para ello, es necesario realizar un análisis detallado

acerca de las características del dominio del negocio, abarcando los distintos aspectos específicos del mismo (los cuales pueden ser poco conocido por los ingenieros de explotación de información) para garantizar una correcta comprensión de los requerimientos del cliente.

En la actividad actual se realizan interacciones con los interesados (cliente / expertos) con el objetivo de identificar y recabar toda información de interés para la comprensión y desarrollo del proyecto. Algunos de los aspectos relevantes a obtener son: las necesidades/objetivos del cliente, la identificación de recursos (humanos y materiales) que sean de interés para el proyecto, ya sea por su conocimiento con respecto a la lógica del negocio, así como su vinculación con los objetivos del proyecto, el contexto de la organización (principales competencias, comportamientos y terminologías específicas del dominio, entre otros). Los objetivos del proyecto definen la dirección del proceso, siendo su correcta definición fundamental para el desarrollo exitoso del mismo.

Información de Entrada

- Discursos de los interesados (externo)
- Información de la Organización (externo)
- Información del dominio del negocio (externo)

Información de Salida

- Fuentes de Información del Cliente (D.EN.ANN.FUIC)
- Definiciones, Acrónimos y Abreviaciones (D.EN.ANN.DEAA)
- Objetivos del Proyecto (D.EN.ANN.OBPR)
- Criterios de Éxito del Proyecto (D.EN.ANN.CREP)
- Expectativas del Proyecto (D.EN.ANN.EXPR)
- Suposiciones del Proyecto (D.EN.ANN.SUPR)
- Restricciones del Proyecto (D.EN.ANN.REPR)

4.4.1.1.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se hacen uso de los siguientes formalismos, propuestos en [Britos et al., 2008], y presentados en la sección 2.4.1.2 (pág. 35).

Fuentes de Información del Cliente (D.EN.ANN.FUIC): El objetivo del mismo es identificar aquellos recursos (de la organización o externos) que sean de interés para la obtención de conocimiento para el desarrollo del proyecto, el tipo de recurso que es (documento, almacén de datos, planilla o manual), el responsable de gestionar el mismo (si existiese) y una descripción general sobre su contenido. El formalismo propuesto se ilustra en la tabla 4.48.

Fuentes de Información del Cliente				
Responsable:	Persona a cargo		**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto		**Versión:**	Identificación de la versión
ID	**Nombre**	**Categoría**	**Responsable**	**Descripción**
Identificador del Recurso 1	Nombre del Recurso 1	Tipo de Recurso	Persona/s responsables de gestionar el recurso	Detalle del Recurso (información disponible, formato, etc.)
…	…	…	…	…
Identificador del Recurso N	Nombre del Recurso N	Tipo de Recurso	Persona/s responsables de gestionar el recurso	Detalle del Recurso (información disponible, formato, etc.)

Tabla 4.48. Formalismo: Fuentes de Información del Cliente.

Definiciones, Acrónimos y Abreviaciones (D.EN.ANN.DEAA): Se registran aquellas terminologías específicas del dominio que no sean familiares para el equipo de trabajo, favoreciendo la comprensión de los distintos aspectos del dominio y la ejecución del mismo, así como también la interacción con los expertos/clientes (educción de requerimientos, presentación de reportes, etc.). En este formalismo se registra una descripción del término, el tipo (definición, acrónimo o abreviación) y la referencia en la cual fue identificado. En la tabla 4.49 se presenta el formalismo propuesto en [Britos, et al., 2008].

Definiciones, Acrónimos y Abreviaciones			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Nombre	**Descripción**	**Tipo**	**Referencia**
Término 1	Significado del Término 1	Qué información representa el término	Fuente en la cual se identificó al término
…	…	…	…
Término N	Significado del Término N	Qué información representa el término	Fuente en la cual se identificó al término

Tabla 4.49. Formalismo: Definiciones, Acrónimos y Abreviaciones

Objetivos del Proyecto (D.EN.ANN.OBPR): En este apartado se registran las metas del proyecto desde la perspectiva del cliente, identificando cuales son los logros que el requirente pretende alcanzar. La tabla 4.50 (propuesta en [Britos, et al., 2008]) presenta el formalismo a utilizar en el cual se registra el identificador del objetivo del proyecto, la descripción del alcance del mismo y la referencia al elemento a partir del cual la meta fue definida.

Objetivos del Proyecto			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Objetivo	**Descripción**		**Referencia**
Identificador del Objetivo 1	Definición del objetivo del proyecto		Fuente en la cual se identificó el objetivo
…	…		…
Identificador del Objetivo N	Definición del objetivo del proyecto		Fuente en la cual se identificó el objetivo

Tabla 4.50. Formalismo: Objetivos del Proyecto

Criterios de Éxito del Proyecto (D.EN.ANN.CREP): Mediante el formalismo seleccionado se deja registro de la metodología a utilizar para evaluar la satisfacción de los objetivos definidos por el cliente, proveyendo un identificador para el criterio, detallándose cómo será dicha evaluación, el objetivo asociado y la referencia al elemento a partir del cual fue definido. En la tabla 4.51 (propuesta en [Britos, et al., 2008]) se presenta el formalismo a utilizar.

Criterios de Éxito del Proyecto			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Criterio	**Descripción**	**Objetivo asociado**	**Referencia**
Identificador del criterio de éxito 1	Definición del modo mediante el cual se evaluará el cumplimiento del objetivo	Identificador del Objetivo vinculado con el criterio de éxito	Fuente en la cual se identificó el objetivo
…	…	…	…
Identificador del criterio de éxito N	Definición del modo mediante el cual se evaluará el cumplimiento del objetivo	Identificador del Objetivo vinculado con el criterio de éxito	Fuente en la cual se identificó el objetivo

Tabla 4.51. Formalismo: Criterios de Éxito del Proyecto

Expectativas del Proyecto (D.EN.ANN.EXPR): Las expectativas del proyecto presentan una visión complementaria a la definición de los objetivos asociados, en la cual se vincula a los mismos con las pretensiones que tienen los clientes/expertos respecto al producto resultante. En este contexto, las expectativas del proyecto describen el uso que los interesados pretenden dar a los resultados provistos, brindando un identificador, junto con su descripción detallada, el objetivo vinculado y la referencia al elemento a partir del cual fue definida. En la tabla 4.52 (propuesta en [Britos, et al., 2008]) se presenta el formalismo previamente descripto.

Expectativas del Proyecto			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Expectativa	**Descripción**	**Objetivo asociado**	**Referencia**
Identificador de la expectativa del proyecto 1	Definición de la expectativa del proyecto 1	Identificador del Objetivo vinculado con la expectativa del proyecto 1	Fuente en la cual se identificó la expectativa del proyecto 1
…	…	…	…
Identificador de la expectativa del proyecto N	Definición de la expectativa del proyecto N	Identificador del Objetivo vinculado con la expectativa del proyecto N	Fuente en la cual se identificó la expectativa del proyecto N

Tabla 4.52. Formalismo: Expectativas del Proyecto

Suposiciones del Proyecto (D.EN.ANN.SUPR): Se identifican aquellas hipótesis o conjeturas asociadas con los objetivos de negocio del proyecto. Dichas hipótesis impactan de forma general sobre los problemas de negocio a desarrollar, los datos y los resultados derivados de los mismos. En la tabla 4.53 (propuesta en [Britos, et al., 2008]) se ilustra el formalismo donde se registran las suposiciones del proyecto junto con un identificador, el objetivo asociado con dicha suposición y la referencia al elemento a partir del cual fue definida.

Suposiciones del Proyecto			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Suposición	**Descripción**	**Objetivo asociado**	**Referencia**
Identificador de la suposición del proyecto 1	Definición de la suposición del proyecto 1	Identificador del Objetivo vinculado con la suposición del proyecto 1	Fuente en la cual se identificó la suposición del proyecto 1
…	…	…	…
Identificador de la suposición del proyecto N	Definición de la suposición del proyecto N	Identificador del Objetivo vinculado con la suposición del proyecto N	Fuente en la cual se identificó la suposición del proyecto N

Tabla 4.53. Formalismo: Suposiciones del Proyecto.

Restricciones del Proyecto (D.EN.ANN.REPR): Se identifican aquellos aspectos que presentan limitaciones para el cumplimiento de los objetivos del negocio, los cuales puedan demorar, afectar o imposibilitar el desarrollo de los mismos y la categoría a la cual dicha restricción está asociada (recurso humano, datos, cuestiones técnicas del proyecto o de la organización). Además, se identifica el objetivo vinculado y la referencia al elemento a partir del cual se define. En la tabla 4.54 (propuesta en [Britos, et al., 2008]) se presenta el formalismo descripto.

Restricciones del Proyecto				
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización	
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión	
Restricción	**Tipo**	**Descripción**	**Objetivo asociado**	**Referencia**
Identificador de la restricción 1	Categoría asociada con la procedencia de la restricción 1	Definición de la restricción 1	Identificador del Objetivo vinculado con la restricción 1	Fuente en la cual se identificó la restricción del proyecto 1
…	…	…	…	…
Identificador de la restricción N	Categoría asociada con la procedencia de la restricción N	Definición de la restricción N	Identificador del Objetivo vinculado con la restricción N	Fuente en la cual se identificó la restricción del proyecto N

Tabla 4.54. Formalismo: Restricciones del Proyecto.

4.4.1.1.2. Técnica A Utilizar

Para el desarrollo de esta actividad se utiliza la técnica **Definición de los objetivos del proyecto** que forma parte de la Metodología para la educción de requerimientos para proyectos de explotación de información propuesta en [Britos et al., 2008], la cual se describe en la sección 2.4.1.2 (pág. 35).

Esta técnica tiene como insumos la información recabada de las partes involucradas con el proyecto (Discursos de los interesados), y aquellos recursos pertenecientes a la organización contratante o públicos que brinden información de interés asociada con el dominio de la organización (Información de la Organización y del dominio del negocio respectivamente), y posee como productos de salida: la identificación de los recursos de información relevantes para el proyecto (Fuentes de Información del Cliente), descripción de conceptos específicos del dominio del negocio de interés para el proyecto (Definiciones, Acrónimos y Abreviaciones) y la descripción y caracterización de las necesidades del cliente (Objetivos del Proyecto, Criterios de Éxito del Proyecto, Expectativas del Proyecto, Suposiciones del Proyecto y Restricciones del Proyecto).

Para la aplicación de la técnica se utilizan los formalismos descriptos en la sección precitada, con el propósito de proveer al lector de claridad en los conceptos expuestos. En la posterior sección, se ilustra la aplicación de la técnica y los formalismos resultantes en el proyecto seleccionado como prueba de concepto (sección 4.2).

4.4.1.1.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica de definición de los objetivos del proyecto que forma parte de la Metodología para la educción de requerimientos para proyectos de explotación de información propuesta en [Britos et al., 2008]. Los mismos se definieron a partir del conocimiento extraído en distintas entrevistas con el cliente.

Para comprender cómo se obtuvieron los resultados, se listan los elementos de entrada para la actividad actual: se describen los resultados obtenidos en las entrevistas 1 y 2 (Fuente de Información 4.3), la información del dominio de negocio y de la Organización, página web INDEC y Documento para la utilización de la base de datos usuario 2011 (Fuente de Información 4.4 y 4.5, respectivamente).

Discursos de los interesados: Entrevistas con el cliente (Silva H.)

Entrevista 1:

"Nuestro interés en el proyecto reside en la posibilidad de identificar nuevas técnicas que nos permitan comprender la población encuestada, y posibilitar el análisis del comportamiento de la población en distintos periodos. La información a utilizar para el estudio se encuentra disponible en el sitio web del INDEC dentro de la categoría sociedad/salud. La información consiste en una encuesta nacional acerca de consumo de sustancias psicoactivas, de la cual se dispone en el sitio de los datos aplicados en dos años: 2008 y 2011, siendo de interés para el estudio la segunda encuesta."

¿Cuáles son sus expectativas con respecto a los alcances y resultados del proyecto?

"Con el uso de estas tecnologías pretendemos definir un proceso que nos permita automatizar o semi-automatizar el análisis de los datos, reduciendo el costo y tiempo asociado con la generación de resultados y el correspondiente accionar a partir de los mismos."

¿Conoce de la existencia algún modelo vigente?

"Nuestro interés es identificar a nivel global cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social de las personas que han consumido sustancias psicoactivas ilegales. En primera instancia nos interesa realizar un estudio básico del comportamiento general de la población, el cual nos permita entender y contrastar los resultados. No tenemos un marco de referencia para establecer un criterio de evaluación, ni uno en mente, pero deseamos obtener mediante este método aquellos aspectos que me permitan comprender el comportamiento de grupos masivos de personas, pudiendo derivar conclusiones a partir de los mismos y brindar indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población".

¿Qué limitaciones entiende que poseen los datos?

"Con respecto al proceso de carga de los datos, no vamos a cuestionar la veracidad, ni la calidad de las encuestas realizadas, ya que la misma está respaldada por el INDEC, entendiendo como adecuada la rigurosidad del proceso. Sin embargo, en la encuesta realizada no se encuentra detallado la gradualidad y frecuencia del consumo de las sustancias psicoactivas que el individuo consume, a excepción de algunos indicadores para las drogas socialmente aceptadas. Adicionalmente, es de esperar que la distribución de consumo de las distintas drogas no esté uniformemente distribuida, sino que por el contrario pueda existir una diferencia notable entre la cantidad de personas que consumen las distintas drogas (aún mayor para drogas socialmente aceptadas) y todavía mayor con aquellos que no consumen. Por las restricciones que les acabo de señalar, para el estudio no vamos a poder considerar la gradualidad de consumo, considerando el comportamiento de los mismos como análogo.".

Entrevista 2:

Del análisis del "Documento para la utilización de la base de datos usuario 2011" identificamos un apartado que hace referencia a la representatividad de la muestra con respecto a la población. En dicho contexto: ¿podemos considerar a dicha muestra como representativa tanto para el país, así como las provincias?

"En el documento explicativo de la base de datos, se detalla el procedimiento realizado para obtener dichos datos en el cual se garantiza la fiabilidad y la representatividad en el país, y también lo entendemos a nivel de provincias… Con respecto a las variables, existen un conjunto de ellas que están vinculadas con la autopercepción del encuestado con respecto a su entorno, si bien poseen una carga subjetiva, las podríamos considerar de interés ya que en cierta forma nos dan un visión del contexto percibido del individuo."

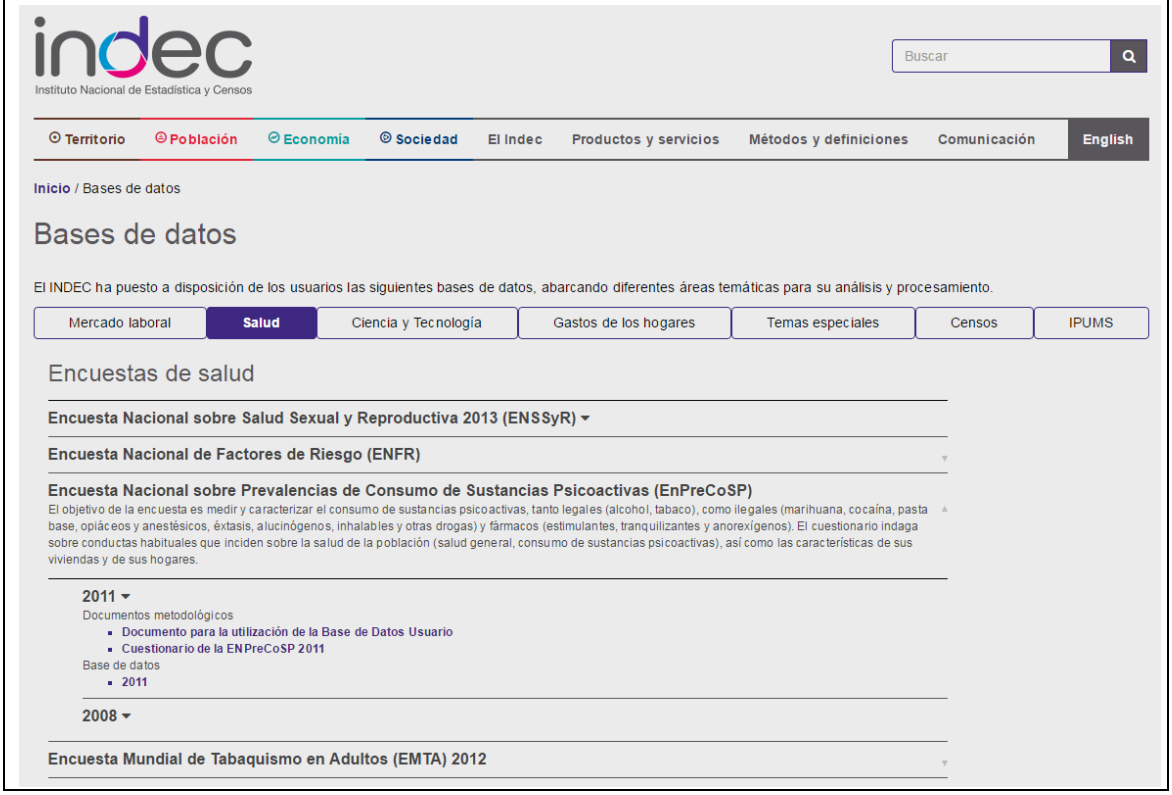
"…Como les mencioné el otro día, consideraremos como correcto el proceso de captura y registro de los encuestados y como señalan de la información que se desprende en dicho documento también vamos a considerar como representativas las muestras tomadas para las provincias y obviamente para el país.

Fuente de Información 4.3. Prueba de Concepto - Entrevistas 1 y 2

***Información del dominio del negocio**: Página web. INDEC*

De acuerdo a la información obtenida en la entrevista, se accede a la página web del Instituto Nacional de Estadística y Censos [Instituto Nacional de Estadística y Censos, 2016], en el cual se identifican los siguientes recursos:

- *Documento para la utilización de la Base de Datos Usuario*
- *Cuestionario de la ENPreCoSP 2011*
- *Base de datos 2011*



Fuente de Información 4.4. Prueba de Concepto - Información del Dominio del Negocio

***Información de la Organización:** Documento para la utilización de la base de datos usuario 2011*

"…Se entiende por sustancias psicoactivas a las drogas legales o sociales (tabaco, bebidas alcohólicas), ilegales (marihuana, cocaína, pasta base, éxtasis, opiáceos y anestésicos, crack, alucinógenos, inhalables y otras drogas) y fármacos (estimulantes, tranquilizantes, anorexígenos)."

*"**4.3. Control de estructuras de población***

Con el fin de evaluar la muestra obtenida, se comparó la estructura de edad de población expandida sin calibrar de la ENPreCoSP-2011 con las proyecciones de población elaboradas por el INDEC correspondientes al 30 de junio de 2011.

La comparación no arrojó diferencias por lo que se deduce que la muestra es representativa de la población, y al momento de la calibración ésta se realizó con respecto a la estructura por edad y sexo de la propia encuesta."

Fuente de Información 4.5. Prueba de Concepto - Información de la Organización

Fuentes de Información del Cliente (D.EN.ANN.FUIC): Mediante la técnica utilizada y el formalismo indicado, se registran las fuentes de información de interés para el desarrollo del proyecto, las cuales poseen conocimiento que puede ser de utilidad para comprender y/o definir los distintos aspectos a considerar durante el desarrollo del proyecto.

El primer paso consiste en identificar las fuentes de información pertenecientes a la organización o externas que han sido señaladas de interés para el desarrollo del proyecto, definiendo un identificador único para cada una de ellas (en las columnas nombre e ID respectivamente). Para la prueba de concepto presentada en la sección 4.2, de acuerdo a los resultados obtenidos presentados en las fuente de información 4.3 a 4.5, se identificaron 3 recursos de interés: a) Documento para la utilización de la Base de Datos Usuario 2011, b) Cuestionario de la ENPreCoSP 2011 y c) Base de datos 2011, a los cuales se les asignó los siguientes identificadores fuic.1, fuic.2 y fuic.3 respectivamente.

Luego, se procede a categorizar el documento de acuerdo al tipo de elemento que pertenece (en la columna "Categoría"). Esto es, si es un documento (de la organización o externo), un tipo de almacén de datos, una planilla de ejemplo, muestra o un manual operativo de tecnologías utilizadas. A continuación se presenta mediante par ordenado el recurso y la categoría asignada (ID; Categoría) durante la clasificación de los recursos previamente identificados: (fuic.1; Documento), (fuic.2; Planilla) y (fuic.3; Base de datos).

El siguiente paso es identificar (en caso que existiese) que miembro es el **responsable** de gestionar dicho elemento (en la columna homónima), el mismo debe estar identificado en el formalismo Recursos Humanos Involucrados (G.In.EvS.ReHI) o en caso contrario, deberá ser registrado en dicho formalismo. En la prueba de concepto, las fuentes de información identificadas son de carácter público, no existiendo un responsable disponible para realizar consultas al respecto.

Por último, se detalla en la columna "Descripción" aquella información complementaria del recurso, la cual facilite la comprensión y el trabajo con el mismo. Por ejemplo, el contenido que posee, el formato del recurso, su tecnología, el periodo de cubrimiento de la información, entre otros. A continuación, se indican las descripciones realizadas por el miembro del equipo encargado del análisis de las fuentes:

- **Documento para la utilización de la base de datos usuario 2011**: Describe distintas consideraciones acerca de la base de datos obtenida a partir de la encuesta ENPreCoSP 2011, indicando los métodos de recolección aplicados, los objetivos de la encuesta y la descripción de los datos.
- **Cuestionario ENPreCoSP 2011**: Ejemplo de cuestionario ENPreCoSP 2011.
- **Base ENPreCoSP 2011**: Almacén de registros de respuestas del cuestionario ENPreCoSP 2011 en formato txt (separado por el carácter "|") conformado por 34343 personas que respondieron 292 preguntas. El primer renglón contiene los nombres de los campos. Decimales separados por punto (.)

La tabla 4.55 identifica las fuentes de información de interés halladas y la descripción de los campos previamente señalados para el caso previamente expuesto.

Fuentes de Información del Cliente					
Responsable:		Esposito E.		**Fecha:**	05/04/2016
ID#:		D.EN.ANN.FUIC		**Versión:**	1.0
ID	**Nombre**	**Categoría**	**Responsable**	**Descripción**	
fuic.1	Documento para la utilización de la base de datos usuario 2011	Documento	-	Describe distintas consideraciones acerca de la base de datos obtenida a partir de la encuesta ENPreCoSP 2011, indicando los métodos de recolección aplicados, los objetivos de la encuesta y la descripción de los datos.	
fuic.2	Cuestionario ENPreCoSP 2011	Planilla	-	Ejemplo de cuestionario ENPreCoSP 2011	
fuic.3	Base ENPreCoSP 2011	Almacén de datos	-	Almacén de registros de respuestas del cuestionario ENPreCoSP 2011 en formato txt (separado por el carácter "|") conformado por 34343 personas que respondieron 292 preguntas. El primer renglón contiene los nombres de los campos. Decimales separados por punto (.)	

Tabla 4.55. Prueba de Concepto - Fuentes de Información del Cliente

Definiciones, Acrónimos y Abreviaciones (D.EN.ANN.DEAA): En esta etapa se propone el registro de aquellas terminologías específicas del dominio que no sean familiares para el equipo de trabajo. Por dicho motivo el resultado del mismo puede variar de acuerdo a la pericia de los miembros en el dominio del proyecto. En la prueba de concepto, el equipo de trabajo identifica el siguiente listado de términos y sus correspondientes definiciones (registradas en la primer y segunda columna del formalismo respectivamente):

- **ENPreCoSP**: Encuesta Nacional sobre Prevalencias de Consumo de Sustancias Psicoactivas
- **Sustancias psicoactivas**: drogas legales o sociales (tabaco, bebidas alcohólicas), ilegales (marihuana, cocaína, pasta base, éxtasis, opiáceos y anestésicos, crack, alucinógenos, inhalables y otras drogas) y fármacos (estimulantes, tranquilizantes, anorexígenos)
- **Sustancias psicoactivas ilegal:** ilegales (marihuana, cocaína, pasta base, éxtasis, opiáceos y anestésicos, crack, alucinógenos, inhalables y otras drogas) y fármacos (estimulantes, tranquilizantes, anorexígenos).
- **Drogas legales o sociales:** tabaco y bebidas alcohólicas.

Luego, se cataloga el término de acuerdo al tipo (Definición, Acrónimo o Abreviación) y por último se registra la referencia al momento (durante una entrevista, Brainstorming, etc.) o documento en el cual fue identificado (en las columnas respectivas). A continuación se presenta la descripción de dichos campos para los términos identificados en forma de terna (Nombre; Tipo; Referencia): (ENPreCoSP; Acrónimo; “Documento para la utilización de la base de datos usuario (fuic.1)”), (sustancias psicoactivas; Definición; “Documento para la utilización de la base de datos usuario

(fuic.1)"), (sustancias psicoactivas ilegal; Definición; "Documento para la utilización de la base de datos usuario (fuic.1)") y (drogas legales o sociales; Definición; "Documento para la utilización de la base de datos usuario (fuic.1)").

En la tabla 4.56 se ilustran los términos previamente descriptos registrados en el formalismo correspondiente.

Definiciones, Acrónimos y Abreviaciones			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.AnN.DeAA	**Versión:**	1.0
Nombre	**Descripción**	**Tipo**	**Referencia**
ENPreCoSP	Encuesta Nacional sobre Prevalencias de Consumo de Sustancias Psicoactivas	Acrónimo	Documento para la utilización de la base de datos usuario (fuic.1)
sustancias psicoactivas	drogas legales o sociales (tabaco, bebidas alcohólicas), ilegales (marihuana, cocaína, pasta base, éxtasis, opiáceos y anestésicos, crack, alucinógenos, inhalables y otras drogas) y fármacos (estimulantes, tranquilizantes, anorexígenos)	Definición	Documento para la utilización de la base de datos usuario (fuic.1)
sustancias psicoactivas ilegal	ilegales (marihuana, cocaína, pasta base, éxtasis, opiáceos y anestésicos, crack, alucinógenos, inhalables y otras drogas) y fármacos (estimulantes, tranquilizantes, anorexígenos)	Definición	Documento para la utilización de la base de datos usuario (fuic.1)
drogas legales o sociales	tabaco y bebidas alcohólicas	Definición	Documento para la utilización de la base de datos usuario (fuic.1)

Tabla 4.56. Prueba de Concepto - Definiciones, Acrónimos y Abreviaciones

Objetivos del Proyecto (D.EN.ANN.OBPR): A partir de la interacción con el cliente y los expertos, se identifican distintas necesidades o metas que los miembros de la organización poseen, entendiendo que la ingeniería de explotación de información es la disciplina que puedes brindarle acceso al cumplimiento de las mismas. En este contexto, haciendo uso del formalismo indicado, se registran aquellos objetivos que la organización contratante posee, los cual pueden estar conformados por varias problemáticas (las cuales serán registradas posteriormente en el formalismos "Problema del Negocio" en la sección 4.4.1.2).

Los objetivos del proyecto brindan un marco o enfoque general de las necesidades del cliente, permitiendo a los miembros del equipo comprender el motivo del proyecto, lo cual favorece a los ingenieros de explotación de información en el entendimiento de los alcances del proyecto y de los productos resultantes. Los objetivos identificados permiten al equipo de trabajo determinar el tipo de necesidad que el cliente posee, así como enmarcar la estrategia a aplicar para resolver los problemas de negocio (a identificar posteriormente en la fase Modelado, sección 4.4.3).

Del discurso del cliente obtenido durante la primera entrevista, se identifica de interés el siguiente párrafo: "*Nuestro interés en el proyecto reside en la posibilidad de identificar nuevas técnicas que nos permitan comprender la población encuestada, y posibilitar el análisis del comportamiento de la población en distintos periodos*" en el cual se señala el objetivo general del cliente. A partir de ello, se registra el mismo como "Obpr.1" en la columna objetivo, y con el propósito de mejorar la comprensión del mismo, se reescribe en la columna descripción el concepto identificado en el párrafo previamente citado, como "*Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos*" y por último, se registra en la columna referencia, que el mismo se obtuvo en la entrevista 1. En la tabla 4.57 se presenta el objetivo del proyecto identificado para la prueba de concepto.

Objetivos del Proyecto			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.ANN.OBPR	**Versión:**	1.0

Objetivo	**Descripción**	**Referencia**
obpr.1	Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1

Tabla 4.57. Prueba de Concepto - Objetivos del Proyecto

Criterios de Éxito del Proyecto (D.EN.ANN.CREP): A partir de los objetivos del proyecto, se procede a definir, de forma conjunta con los clientes/expertos vinculados con dicho propósito, los criterios que se tendrán en cuenta al momento de evaluar el cumplimiento satisfactorio de las metas definidas. El criterio de éxito debe ser lo más objetivo posible, es decir, definir con precisión la forma mediante la cual se medirá y evaluará el cumplimiento del mismo. En caso que este no pueda ser definido de dicha forma, deberá quedar registrado con el mayor nivel de detalle posible los lineamientos que se utilizarán para evaluar el resultado y (de ser necesario) quien determinará su cumplimiento.

A partir del extracto obtenido de las entrevistas realizadas al cliente, se identifica el siguiente párrafo: "*deseamos obtener mediante este método aquellos aspectos que permitan comprender el comportamiento de grupos masivos de personas, brindando indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población*", el cual se registra en la columna descripción, siendo el mismo adaptado a "*obtener piezas de conocimiento que favorezcan la comprensión del comportamiento de grupos masivos de personas, brindando indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales serán validadas por el cliente*" para favorecer la comprensión del criterio de éxito, asignándole el identificador "crexpr.1". El criterio de

éxito se encuentra asociado al único objetivo del proyecto (Obpr.1), obteniendo esta información de la entrevista 1 realizada al cliente (Silva H.). Dicha información se registra en las columnas objetivo asociado y referencia respectivamente.

En la tabla 4.58 se presenta el criterio de éxito definido para el objetivo del proyecto identificado para el caso presentado como prueba de concepto.

Criterios de Éxito del Proyecto			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.ANN.CREP	**Versión:**	1.0

Criterio	**Descripción**	**Objetivo asociado**	**Referencia**
crexpr.1	obtener piezas de conocimiento que favorezcan la comprensión del comportamiento de grupos masivos de personas, brindando indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales serán validadas por el cliente	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1

Tabla 4.58. Prueba de Concepto - Criterios de Éxito del Proyecto

Expectativas del Proyecto (D.EN.ANN.EXPR): Las expectativas del proyecto presentan una visión complementaria a la definición del objetivo asociado, en el cual se vinculan las pretensiones que tienen los clientes/expertos con respecto al producto resultante como respuesta a cada objetivo de proyecto identificado. La expectativa debe estar alineada con el objetivo del proyecto y sus criterios de éxito. Estas describen el uso que los interesados pretenden dar a los resultados provistos.

En primera instancia, se identifica del conocimiento educido durante las entrevistas, cúal es el interés que tiene el cliente con respecto a los resultados esperados para los objetivos del proyecto (asignando un identificador unívoco para cada uno). En la prueba de concepto, se identifican los párrafos: "*obtener mediante este método aquellos aspectos que me permitan comprender el comportamiento de grupos masivos de personas, pudiendo derivar conclusiones a partir de los mismos y brindar indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población*" y "*Con el uso de estas tecnologías pretendemos definir un proceso que nos permita automatizar o semi-automatizar el análisis de los datos, reduciendo el costo y tiempo asociado con la generación de resultados y el correspondiente accionar a partir de los mismos.*", a partir de los cuales se convirtió para simplificar la comprensión del mismo en "*Definir un proceso que automatice o semi-automatice el análisis de los datos, reduciendo el costo y tiempo asociado con la generación de resultados y el correspondiente accionar a partir de los mismos, que brinde indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales*

serán validadas por el cliente" asignando el identificador "expr.1". La expectativa se encuentra asociada al único objetivo del proyecto "Obpr.1", obteniendo esta información en la entrevista 1 realizada al único cliente. Dicha información se registra en las columnas objetivo asociado y referencia, respectivamente.

En la tabla 4.59 se presenta la expectativa asociada al objetivo del proyecto identificada para el caso presentado como prueba de concepto.

Expectativas del Proyecto			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.ANN.EXPR	**Versión:**	1.0

Expectativa	**Descripción**	**Objetivo asociado**	**Referencia**
expr.1	Definir un proceso que automatice o semi-automatice el análisis de los datos, reduciendo el costo y tiempo asociado con la generación de resultados y el correspondiente accionar a partir de los mismos, que brinde indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales serán validadas por el cliente	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1

Tabla 4.59. Prueba de Concepto - Expectativas del Proyecto

Suposiciones del Proyecto (D.EN.ANN.SUPR): De la interacción del ingeniero de explotación de información con los expertos vinculados con el objetivo del proyecto, se identifican aquellas hipótesis o conjeturas asociadas con los objetivos de negocio del proyecto. Dichas suposiciones establecen hechos considerados como ciertos para el desarrollo del proyecto (los cuales pueden o no ser verificables), que impactan sobre la forma en que se concebirán y se interpretarán los problemas de negocio a desarrollar, los datos y los resultados derivados de los mismos.

En la entrevista realizada para el proyecto señalado como prueba de concepto, se identificaron las siguientes suposiciones:

De la Entrevista 1 se identifica el siguiente párrafo:

- "*Con respecto al proceso de carga de los datos, no vamos a cuestionar la veracidad, ni la calidad de las encuestas realizadas, ya que la misma está respaldada por el INDEC, entendiendo como adecuada la rigurosidad del proceso*."
- "*…en la encuesta realizada no se encuentra detallado la gradualidad y frecuencia del consumo de las sustancias psicoactivas que el individuo consume… para el estudio no vamos a poder considerar la gradualidad de consumo, considerando el comportamiento de los mismos como análogo*"

Adicionalmente, en la entrevista 2, se identifican los siguientes párrafos:

- "*Con respecto a las variables, existen un conjunto de ellas que están vinculadas con la autopercepción del encuestado con respecto a su entorno, si bien poseen una carga subjetiva, las podríamos considerar de interés ya que en cierta forma nos dan un visión del contexto percibido del individuo*"
- "*…como señalan de la información que se desprende en dicho documento también vamos a considerar como representativas las muestras tomadas para las provincias y obviamente para el país."*

Para el registro de las suposiciones previamente indicadas, se asignó un identificador continuando a partir del valor previo, se describió dicha suposición adaptando los mismos para facilitar la comprensión de las suposiciones, y se las vinculó al objetivo y al elemento a partir del cual fue definida. A continuación se presenta en forma de cuaternas (Suposición; Descripción; Objetivo asociado; Referencia) los registros de los párrafos previamente listados:

- (supr.1; Los cuestionarios y la carga de la información se ha realizado de manera correcta; (obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos; Entrevista 1)
- (supr.2; Las conductas de consumo se considerarán como análogas sin importar la gradualidad del mismo; (obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos; Entrevista 1)
- (supr.3; Las variables vinculadas con autopercepción brindan información fiable respecto al entorno real del individuo; (obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos; Entrevista 2)
- (supr.4; El proceso de diseño de la muestra es representativo a nivel nacional y provincial; (obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos; Entrevista 2/fuic.1)

En la tabla 4.60 se ilustran las suposiciones del proyecto asociadas al objetivo de proyecto identificado para el caso presentado como prueba de concepto. Las suposiciones del proyecto identificadas fueron ampliadas a partir de la segunda entrevista, por lo cual el formalismo fue modificado incorporando las últimas tres suposiciones previamente mencionadas. Esto generó un cambio en el versionado del producto (versión 1.1) y la ejecución de los pasos asociados con dicho evento, los cuales se describen en la actividad Gestión de la Configuración (sección 4.3.3.2) perteneciente a la fase de soporte del subproceso de gestión.

Suposiciones del Proyecto			
Responsable:	Esposito E.	**Fecha:**	20/04/2016
ID#:	D.EN.ANN.SUPR	**Versión:**	1.1

Suposición	Descripción	Objetivo asociado	Referencia
supr.1	Los cuestionarios y la carga de la información se ha realizado de manera correcta	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1
supr.2	Las conductas de consumo se considerarán como análogas sin importar la gradualidad del mismo	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1
supr.3	Las variables vinculadas con autopercepción brindan información fiable respecto al entorno real del individuo	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 2
supr.4	El proceso de diseño de la muestra es representativo a nivel nacional y provincial	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 2 / fuic.1

Tabla 4.60. Prueba de Concepto - Suposiciones del Proyecto

Restricciones del Proyecto (D.EN.ANN.REPR): Se identifican aquellos aspectos que presentan limitaciones para el cumplimiento de los objetivos del negocio, los cuales puedan demorar, afectar o imposibilitar el desarrollo de los mismos. Las limitaciones pueden estar asociadas al recurso humano (conocimiento de las técnicas o tecnologías, disponibilidades), a los datos (posibilidad de acceso, calidad) o a cuestiones técnicas del proyecto (hardware o software) o de la organización (aspectos políticos o legales).

A partir del conocimiento extraído del experto, se identifican las siguientes restricciones:

- "*...es de esperar que la distribución de consumo de las distintas drogas no esté uniformemente distribuida, sino que por el contrario pueda existir una diferencia notable entre la cantidad de personas que consumen las distintas drogas (aún mayor para drogas socialmente aceptadas) y todavía mayor con aquellos que no consumen*"
- "*Por las restricciones que les acabo de señalar, para el estudio no vamos a poder considerar la gradualidad de consumo, considerando el comportamiento de los mismos como análogo.*"

Para el registro de las restricciones precitadas, se asignó un identificador unívoco a las mismas, se adaptó su descripción con el propósito de facilitar su comprensión, y se las vinculó al objetivo y al elemento a partir del cual fueron definidas. A continuación se presenta en forma de quinterna (Restricción; Tipo; Descripción; Objetivo asociado; Referencia) los registros de los párrafos previamente listados:

- (repr.1; datos; Se identifica un desbalance entre la cantidad de registros que han consumido distintas Sustancias Psicoactivas y quienes no han consumido; (obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos; Entrevista 1)

- (repr.2; datos; Se carece de información detallada respecto al grado o frecuencia de consumo de las sustancias psicoactivas; (obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos; Entrevista 1)

Se considera relevante destacar que la categoría definida para las restricciones previamente señaladas, se asignó el valor "datos" debido a la naturaleza a la cual las mismas están asociadas, es decir, ambas restricciones indican una carencia en la información disponible en las fuentes de información del cliente, la primera de ellas respecto a la cantidad de individuos con valor positivo en el consumo o no de sustancias psicoactivas y la segunda por la faltante de datos que hubiesen permitido ampliar el entendimiento y conocimiento de la población de estudio. En la tabla 4.61 se presenta las restricciones del proyecto previamente descriptas.

Restricciones del Proyecto				
Responsable:		Esposito E.	**Fecha:**	20/04/2016
ID#:		D.EN.ANN.REPR	**Versión:**	1.0
Restricción	**Tipo**	**Descripción**	**Objetivo asociado**	**Referencia**
repr.1	datos	Se identifica un desbalance entre la cantidad de registros que han consumido distintas Sustancias Psicoactivas y quienes no han consumido	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1
repr.2	datos	Se carece de información detallada respecto al grado o frecuencia de consumo de las sustancias psicoactivas	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1

Tabla 4.61. Prueba de Concepto - Restricciones del Proyecto

4.4.1.2. Actividad: Comprensión del Problema de Negocio (D.EN.CPN)

El objetivo de proyecto define cuál es la meta que el cliente desea alcanzar como resultado del proceso, mientras que el problema de negocio, presenta una visión detallada de preguntas-problema específicas que el cliente desea responder, las cuales permiten alcanzar los objetivos planteados. El problema de negocio, está fuertemente vinculado con los datos y las piezas de conocimiento que se desean extraer. Asimismo, es a partir del problema de negocio que se determina el tipo de modelo a utilizar.

En este contexto, durante esta actividad debe definirse con claridad no solo las problemáticas a responder, sino también qué tipos de variables se deberán considerar para la generación del modelo, sin preocuparse por definir en detalle las variables que específicamente se utilizarán, siendo dicha acción realizada en la actividad de análisis de los datos (sección 4.4.2.1) perteneciente a la fase de entendimiento de los datos del corriente subproceso.

En esta actividad, se interactúa con los clientes/expertos vinculados con el objetivo de proyecto, con el fin de profundizar en la comprensión de las problemáticas a abordar, identificando los problemas de negocio y sus características.

Información de Entrada

- Discursos de los interesados (externo)
- Información de la Organización (externo)
- Información del dominio del negocio (externo)
- Definiciones, Acrónimos y Abreviaciones (D.EN.ANN.DEAA)
- Objetivos del Proyecto (D.EN.ANN.OBPR)
- Criterios de Éxito del Proyecto (D.EN.ANN.CREP)
- Expectativas del Proyecto (D.EN.ANN.EXPR)
- Suposiciones del Proyecto (D.EN.ANN.SUPR)
- Restricciones del Proyecto (D.EN.ANN.REPR)

Información de Salida

- Problema del Negocio (D.EN.CPN.PRNE)
- Criterios de Éxito del Problema de Negocio (D.EN.CPN.CEPN)

4.4.1.2.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se hacen uso de los siguientes formalismos, propuestos en [Britos et al., 2008] e introducidos en la sección 2.4.1.2 (pág. 35).

Problema del Negocio (D.EN.CPN.PRNE): En este formalismo se deja registro de las necesidades de negocio vinculadas con los objetivos del proyecto. Estas representan preguntas-problemas que definen con claridad la problemática a abordar y las variables que serán consideradas.

En la fila "Objetivo del Proyecto", se identifica la meta global a la cual estarán asociados los problemas de negocios a registrar, asignando para cada uno, un identificador (en la columna "problema") y su definición en la columna "descripción". Adicionalmente, se listan los expertos o interesados vinculados con el problema y las fuentes en la cual se identifica dicha información. La tabla 4.62 (propuesta en [Britos, et al., 2008]) ilustra el formalismo descripto para registrar los problemas de negocio.

Criterios de Éxito del Problema de Negocio (D.EN.CPN.CEPN): Mediante este formalismo se deja registro de la metodología que se utilizará para evaluar el cumplimiento de los problemas de negocio. Para ello se registra un identificador asociado al criterio, se detalla cómo será dicha evaluación (en los campos "Criterio" y "descripción" respectivamente), se identifica el problema de negocio asociado y la referencia al elemento a partir del cual este fue definido. En la tabla 4.63 (propuesta en [Britos, et al., 2008]) se presenta el formalismo a utilizar.

Problema del Negocio			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Objetivo del Proyecto	Identificador del Objetivo del Proyecto		

Problema	Descripción	Experto	Referencia
Identificador del Problema de negocio 1	Definición del Problema de negocio 1	Recursos Humanos vinculados al problema de negocio 1	Fuente en la cual se identificó el problema
…	…	…	…
Identificador del Problema de negocio N	Definición del Problema de negocio N	Recursos Humanos vinculados al problema de negocio N	Fuente en la cual se identificó el problema

Tabla 4.62. Formalismo: Problema del Negocio.

Criterios de Éxito del Problema de Negocio			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión

Criterio	Descripción	Problema asociado	Referencia
Identificador del criterio de éxito 1	Definición del criterio de éxito 1	Identificador del Problema de negocio asociado al criterio 1	Fuente en la cual se identificó el criterio de éxito 1
…	…	…	…
Identificador del criterio de éxito N	Definición del criterio de éxito N	Identificador del Problema de negocio asociado al criterio N	Fuente en la cual se identificó el criterio de éxito N

Tabla 4.63. Formalismo: Criterios de Éxito del Problema de Negocio.

4.4.1.2.2. Técnicas Identificadas

Para el desarrollo de esta actividad se utiliza la técnica **Definición de los Problema de Negocio** que forma parte de la Metodología para la educción de requerimientos para proyectos de explotación de información propuesta en [Britos et al., 2008], la cual se introduce en la sección 2.4.1.2 (pág. 35).

Esta técnica tiene como insumos la información recabada de las partes involucradas con el proyecto (Discursos de los interesados), recursos pertenecientes a la organización contratante o públicos que brinden información asociada con el dominio de la organización que sean de interés para el desarrollo del mismo (Información de la Organización y del dominio del negocio respectivamente), las terminologías específicas del dominio (Definiciones, Acrónimos y Abreviaciones) y la descripción y caracterización de las necesidades del cliente (Objetivos del Proyecto, Expectativas del Proyecto, Suposiciones del Proyecto y Restricciones del Proyecto). Sus productos resultantes son la identificación de problemáticas específicas que guiarán al equipo en el proceso de extracción de patrones de conocimiento, las cuales permiten dar respuesta a uno o más aspectos comprendidos por el objetivo del proyecto (Problema del Negocio) y el mecanismo que se utilizará para evaluar el cumplimiento exitoso de los mismos (Criterios de Éxito del Problema de Negocio).

Para favorecer al lector en la comprensión de la aplicación de la técnica, se utilizarán los formalismos previamente descriptos en el proyecto seleccionado como prueba de concepto, brindando al lector una guía que lo ayude a entender los pasos que la misma involucra.

4.4.1.2.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica de definición del problema de negocio que forma parte de la Metodología para la educción de requerimientos para proyectos de explotación de información propuesta en [Britos et al., 2008]. Los mismos se obtuvieron a partir del conocimiento extraído en distintas entrevistas con el interesado y los formalismos derivados en la actividad previa.

Con el objetivo de facilitar al lector en la comprensión de la técnica y la identificación de los distintos conceptos, se listan los elementos de entrada utilizados por la técnica mencionada en el párrafo anterior: entrevista con el cliente (Fuente de Información 4.6), detallando el conocimiento obtenido en la tercera entrevista, Definiciones, Acrónimos y Abreviaciones (Tabla 4.56), Objetivos del Proyecto (Tabla 4.57), Criterios de Éxito del Proyecto (Tabla 4.58), Expectativas del Proyecto (Tabla 4.59), Suposiciones del Proyecto (Tabla 4.60) y Restricciones del Proyecto (Tabla 4.61).

Las tablas indicadas como elementos de entrada, son transcriptas con el mismo número de referencia, para facilitar al lector en la comprensión de la aplicación de la técnica.

Discursos de los interesados: Entrevistas con el cliente (Silva H.)

Entrevista 3:

Entendemos que el principal interés del proyecto es comprender los aspectos vinculados con el consumo de sustancias psicoactivas. ¿Puede identificar problemáticas de acuerdo a los datos disponibles que le permita alcanzar dicho objetivo?

"Así es, nuestro interés es evaluar las circunstancias que más favorecen en una persona a que consuma este tipo de sustancias, pero especialmente estamos interesados en aquellas ilegales… Específicamente, sería en primera instancia analizar uno de los objetivo definidos en la encuesta realizada por el INDEC: determinar las características sociodemográficas, socioeconómicas, educativas y del entorno familiar social de la población de 16 a 65 años de edad que consume sustancias psicoactivas. Pero no descartamos a futuro ampliar la lista de objetivos o la población de estudio."

¿Qué variable o conjunto de ellas consideraría de interés para evaluar la población?

"…Como mencione previamente aquellas vinculadas con los aspectos sociodemográficas, socioeconómicas, educativas y del entorno familiar social de la persona, pero siempre centrado en aquellas características específicas del encuestado. Hay un conjunto de aspectos vinculados con el jefe del hogar que no son de interés para el proyecto…"

¿Conoce indicadores o modelos actuales para analizar la problemática previamente mencionada? En caso que no: ¿Qué elemento considera factible para la evaluación de los resultados producidos?

"…desconozco la existencia de un mecanismo o modelo que permita analizar el problema, y dado las características del mismo, lo que me interesa es que los resultados obtenidos me permitan comprender el comportamiento de grupos masivos de personas. Particularmente, me parece que la mejor forma es que evalué los resultados obtenidos y juzgue la calidad de los mismos…"

Fuente de Información 4.6. Prueba de Concepto - Entrevista 3

Fuente de Información 4.7. Objetivos del Proyecto			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.AnN.ObPr	**Versión:**	1.0
Objetivo	**Descripción**		**Referencia**
obpr.1	Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos		Entrevista 1

Tabla 4.57 (Transcripta). Prueba de Concepto - Objetivos del Proyecto

Definiciones, Acrónimos y Abreviaciones			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.AnN.DeAA	**Versión:**	1.0

Nombre	Descripción	Tipo	Referencia
ENPreCoSP	Encuesta Nacional sobre Prevalencias de Consumo de Sustancias Psicoactivas	Acrónimo	Documento para la utilización de la base de datos usuario (fuic.1)
sustancias psicoactivas	drogas legales o sociales (tabaco, bebidas alcohólicas), ilegales (marihuana, cocaína, pasta base, éxtasis, opiáceos y anestésicos, crack, alucinógenos, inhalables y otras drogas) y fármacos (estimulantes, tranquilizantes, anorexígenos)	Definición	Documento para la utilización de la base de datos usuario (fuic.1)
sustancias psicoactivas ilegal	ilegales (marihuana, cocaína, pasta base, éxtasis, opiáceos y anestésicos, crack, alucinógenos, inhalables y otras drogas) y fármacos (estimulantes, tranquilizantes, anorexígenos)	Definición	Documento para la utilización de la base de datos usuario (fuic.1)
drogas legales o sociales	tabaco y bebidas alcohólicas	Definición	Documento para la utilización de la base de datos usuario (fuic.1)

Tabla 4.56 (Transcripta). Prueba de Concepto - Definiciones, Acrónimos y Abreviaciones

Criterios de Éxito del Proyecto			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.AnN.CrEP	**Versión:**	1.0

Criterio	Descripción	Objetivo asociado	Referencia
crexpr.1	obtener piezas de conocimiento que favorezcan la comprensión del comportamiento de grupos masivos de personas, brindando indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales serán validadas por el cliente	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1

Tabla 4.58 (Transcripta). Prueba de Concepto - Criterios de Éxito del Proyecto

Expectativas del Proyecto			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.AnN.ExPr	**Versión:**	1.0

Expectativa	Descripción	Objetivo asociado	Referencia
expr.1	Definir un proceso que automatice o semi-automatice el análisis de los datos, reduciendo el costo y tiempo asociado con la generación de resultados y el correspondiente accionar a partir de los mismos, que brinde indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales serán validadas por el cliente	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1

Tabla 4.59 (Transcripta). Prueba de Concepto - Expectativas del Proyecto

Suposiciones del Proyecto			
Responsable:	Esposito E.	**Fecha:**	20/04/2016
ID#:	D.EN.AnN.SuPr	**Versión:**	1.1

Suposición	Descripción	Objetivo asociado	Referencia
supr.1	Los cuestionarios y la carga de la información se ha realizado de manera correcta	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1
supr.2	Las conductas de consumo se considerarán como análogas sin importar la gradualidad del mismo	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1
supr.3	Las variables vinculadas con autopercepción brindan información fiable respecto al entorno real del individuo	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 2
supr.4	El proceso de diseño de la muestra es representativo a nivel nacional y provincial	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 2 / fuic.1

Tabla 4.60 (Transcripta). Prueba de Concepto - Suposiciones del Proyecto

Restricciones del Proyecto				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.AnN.RePr	**Versión:**	1.0	

Restricción	Tipo	Descripción	Objetivo asociado	Referencia
repr.1	datos	Se identifica un desbalance entre la cantidad de registros que han consumido distintas Sustancias Psicoactivas y quienes no han consumido	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1
repr.2	datos	Se carece de información detallada respecto al grado o frecuencia de consumo de las sustancias psicoactivas	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1

Tabla 4.61 (Transcripta). Prueba de Concepto - Restricciones del Proyecto

Problema del Negocio (D.EN.CPN.PRNE): a partir de la interacción con los expertos y de acuerdo a los objetivos del proyecto (definidos en la actividad previa), y en consideración de los aspectos vinculados con los objetivos del proyecto: expectativas, suposiciones y restricciones del proyecto, se identifican aquellas preguntas-problemas que se quieran responder de acuerdo a los objetivos finales. Señalando los datos (o tipos de datos) a considerar, así como los expertos o interesados vinculados con el problema.

El objetivo de negocio, permite identificar el alcance del problema de negocio (dado que este debe cubrir un subconjunto del elemento superior). Las expectativas, suposiciones y restricciones, permiten al equipo determinar el alcance o marco del problema de negocio, es decir, estos limitan

los distintos aspectos a cubrir por el mismo (por ejemplo: los datos a utilizar, o el resultado esperado).

Por cada objetivo del proyecto identificado, se profundiza en el interés del cliente por resolver distintas problemáticas que estén asociadas al mismo. Para la prueba de concepto, se identifica en la entrevista 3 el siguiente párrafo: "…*nuestro interés es evaluar las circunstancias que más favorecen en una persona a que consuma este tipo de sustancias, pero especialmente estamos interesados en aquellas ilegales… Específicamente, sería en primera instancia analizar uno de los objetivo definidos en la encuesta realizada por el INDEC: determinar las características sociodemográficas, socioeconómicas, educativas y del entorno familiar social de la población de 16 a 65 años de edad que consume sustancias psicoactivas*".

Luego, se procede a evaluar la necesidad identificada respecto a los descriptores del objetivo del proyecto en cuestión (expectativas, suposiciones y restricciones del proyecto) garantizando que la misma no se contrapongan. Las expectativas del proyecto brindan al problema de negocio la comprensión del tipo de necesidad y los resultados esperados. Del caso actual se desprende la necesidad de obtener patrones que puedan ser comprendidos por el cliente.

Las suposiciones y restricciones ayudan a establecer el marco del problema, debiendo contemplar los elementos registrados en los mismos para la definición del problema de negocio. En la prueba de concepto, las suposiciones "(supr.2) Las conductas de consumo se considerarán como análogas sin importar la gradualidad del mismo" y "(supr.3) Las variables vinculadas con autopercepción brindan información fiable respecto al entorno real del individuo", complementan la visión de los párrafos de la entrevista 3 previamente mencionados, incorporando información respecto al uso de variables para el problema.

En la tabla 4.64 se presenta el problema de negocio vinculado con el objetivo del proyecto identificado en la prueba de concepto.

Problema del Negocio				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.PRNE	**Versión:**	1.0	
Objetivo del Proyecto	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos			
Problema	**Descripción**		**Experto**	**Referencia**
prne.1	Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		(rehi.3) Silva H.	Entrevista 3

Tabla 4.64. Prueba de Concepto - Problema del Negocio

Criterios de Éxito del Problema de Negocio (D.EN.CPN.CEPN): De acuerdo a las características del problema de negocio y las necesidades del cliente, se definen los criterios que se tendrán en cuenta al momento de evaluar el cumplimiento satisfactorio del problema. En ocasiones, la presencia de modelos existentes puede ser un elemento a considerar al momento de definir los criterios de éxito. De forma análoga al criterio del objetivo del proyecto, el criterio de éxito del problema de negocio debe definir con precisión la forma mediante la cual se medirá y evaluará el cumplimiento del mismo. En caso que este no pueda ser definido de dicha forma, deberá quedar registrado con el mayor nivel de detalle posible los lineamientos que se utilizarán para evaluar el resultado y (de ser necesario) quien determinará su satisfacción. Debe controlarse que el criterio de éxito del problema de negocio, no se contraponga con el criterio de éxito del objetivo del proyecto asociado.

En la prueba de concepto se identifica en la entrevista 3, realizada al cliente Silva H. (rehi.3), el siguiente párrafo: *"...los resultados obtenidos me permitan comprender el comportamiento de grupos masivos de personas. Particularmente, me parece que la mejor forma es que evalué los resultados obtenidos y juzgue la calidad de los mismos.*". A partir de lo expresado por el cliente y en concordancia con el criterio de éxito del objetivo asociado, se define el siguiente criterio de éxito: "*Identificar aspectos que permitan comprender el comportamiento de grupos masivos de personas de acuerdo a la evaluación de Silva H. (rehi.3)*". En la tabla 4.65 se presenta el criterio de éxito definido para el problema de negocio identificado para el caso presentado como prueba de concepto.

Criterios de Éxito del Problema de Negocio				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.CEPN	**Versión:**	1.0	
Criterio	**Descripción**	**Problema asociado**		**Referencia**
cepn.1	Identificar aspectos que permitan comprender el comportamiento de grupos masivos de personas de acuerdo a la evaluación de Silva H. (rehi.3)	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		Entrevista 3

Tabla 4.65. Prueba de Concepto - Criterios de Éxito del Problema de Negocio

4.4.2. Fase: Entendimiento de los Datos (D.ED)

En la fase de entendimiento de los datos, se realiza un estudio exhaustivo de la información disponible tanto dentro como fuera de la organización de interés para el desarrollo del proyecto. Comprender el significado, las características del estado de una variable, así como el motivo por el

cual se considera la misma para realizar la búsqueda de patrones, es un paso relevante respecto a la calidad, validez y originalidad de los resultados obtenidos.

En este contexto, se realizan actividades de análisis de los datos (sección 4.4.2.1), donde se profundiza en la comprensión del significado de las variables disponibles y sus valores, exploración de los datos (sección 4.4.2.2), donde se describe en detalle las variables a considerar por el modelo, y evaluación de los datos (sección 4.4.2.3), donde se evalúan los distintos aspectos vinculados con la calidad de las variables seleccionadas. En la figura 4.11, se resume las actividades que conforman la fase, junto con sus elementos de entrada y salida (las imágenes de cada formalismo son representaciones miniatura de los mismos).

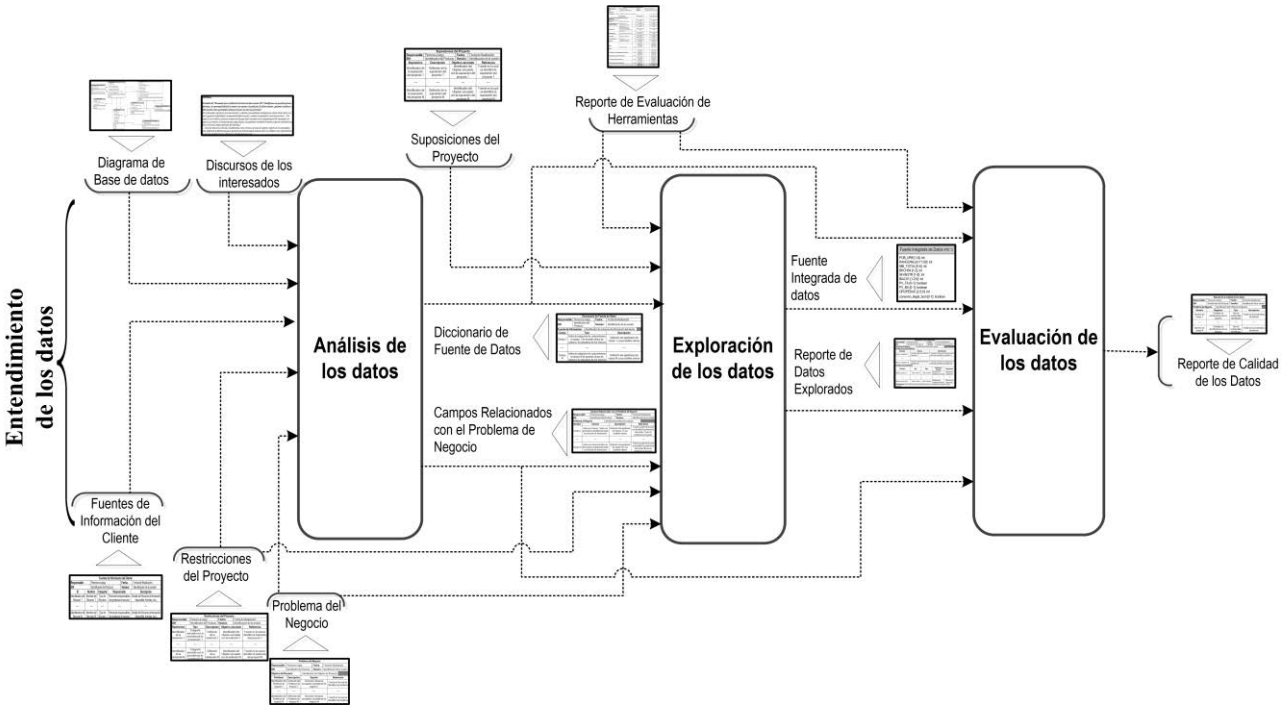


Figura 4.11. Fase: Entendimiento de los Datos

4.4.2.1. Actividad: Análisis de los Datos (D.ED.AnD)

Durante la actividad de análisis de los datos se evalúan las variables disponibles en las distintas fuentes de información, con el objetivo de comprender sus significados, valoraciones, así como cualquier otro aspecto relevante del proceso aplicado para el registro de dicha información (por ejemplo: valores por defecto del sistema, forma en la cual los datos son recolectados, etc.). Las fuentes de información a explorar pueden ser tanto internas como externas a la organización. Del estudio de los datos, se identifica de forma conjunta con los expertos las variables que el modelo tendrá en consideración.

Información de Entrada

- Discursos de los interesados (externo)
- Diagrama de Base de datos (externo)
- Fuentes de Información del Cliente (D.EN.AnN.FuIC)
- Restricciones del Proyecto (D.EN.AnN.RePr)
- Problema del Negocio (D.EN.CPN.PrNe)

Información de Salida

- Diccionario de Fuente de Datos (D.ED.AnD.DiFD)
- Campos Relacionados con el Problema de Negocio (D.ED.AnD.CRPN)

4.4.2.1.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se hace uso del formalismo campos relacionados con el problema de negocio (sección 2.4.1.2, pág. 35), propuesto en [Britos et al., 2008] y el Diccionario de Fuentes de Datos.

Diccionario de Fuentes de Datos (D.ED.AnD.DiFD): Se realiza una descripción detallada de los campos identificados en las fuentes de datos, garantizando la correcta comprensión del significado de las variables y de sus valores posibles. Mediante este formalismo se describen las variables existentes en cada una de las fuentes de información (identificadas en la parte superior), indicando el nombre de la variable, el tipo de valor que representa (cualitativo [ordinal/nominal] o cuantitativo [continuo/discreto]) y la descripción del significado de la variable y sus valores posibles en las columnas homónimas. La tabla 4.66 ilustra la estructura del formalismo previamente descripto.

Diccionario de Fuente de Datos			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Fuente de Información:	Identificador de la fuente de información del cliente		
Campo	**Tipo**		**Descripción**
Campo 1	Indica la categoría a la cual pertenece el campo 1 de acuerdo al tipo de valores y la naturaleza de los mismos		Definición del significado del campo 1 y sus posibles valores
…	…		…
Campo N	Indica la categoría a la cual pertenece el campo N de acuerdo al tipo de valores y la naturaleza de los mismos		Definición del significado del campo N y sus posibles valores

Tabla 4.66. Formalismo: Diccionario de Fuente de Datos

Campos Relacionados con el Problema de Negocio (D.ED.AnD.CRPN): Se deja registro formal del conjunto de variables a utilizar para cada problema de negocio identificado. En este formalismo se indica el nombre del campo (incluyendo el de la fuente de información en caso que fuese necesario para su identificación unívoca), si el campo es una variable compuesta la cual debe ser generada a partir de otros datos (indicando con una equis "X" en caso que la variable tuviese que

ser generada), la descripción del campo en el cual se incorpora la definición de la variable (ya sea la lógica a utilizar para generar el nuevo campo, así como el significado de una variable existente) y se señala la referencia a partir de la cual se definió el campo y/o su importancia para el problema de negocio. En la tabla 4.67 se ilustra el formalismo propuesto.

Campos Relacionados con el Problema de Negocio			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Problema de Negocio	Identificador problema de negocio		
Nombre	**Generar**	**Descripción**	**Referencia**
Campo 1	Indica si el campo 1 debe ser generado o actualmente existe en la fuente de información	Definición del significado del campo 1 y sus posibles valores	Fuente a partir de la cual se identificó la pertinencia del campo 1 para el problema de negocio
…	…	…	…
Campo N	Indica si el campo N debe ser generado o actualmente existe en la fuente de información	Definición del significado del campo N y sus posibles valores	Fuente a partir de la cual se identificó la pertinencia del campo N para el problema de negocio

Tabla 4.67. Formalismo: Campos Relacionados con el Problema de Negocio

4.4.2.1.2. Técnica Identificada

Para el desarrollo de esta actividad se utiliza la técnica **"Identificación de atributos relacionados con el Problema de Negocio"** definida en la Metodología para la educción de requerimientos para proyectos de explotación de información [Britos et al., 2008], la cual se describe en la sección 2.4.1.2 (pág. 35). Para alcanzar los objetivos completos de la actividad, se requiere incorporar a la técnica previamente mencionada el uso de la técnica Diccionario de Datos, la cual provee de una herramienta que permite ganar mayor conocimiento sobre los datos y su importancia en el dominio del negocio, y así obtener mejores resultados en el proceso de identificación de los atributos.

Los resultados generados a partir de esta actividad, utilizan como elemento de entrada la información recabada de entrevistar a las partes involucradas con el proyecto, en particular con los miembros vinculados con los problemas de negocio identificados y los expertos a cargo de la administración de las fuentes de información, el conocimiento vinculado con las fuentes de información identificadas en la fase de entendimiento del negocio (Fuentes de Información del Cliente), diagramas de las bases de datos (por ejemplo: Diagrama Entidad Relación, Modelo de datos orientado a objetos, etc.), así como información del proyecto que tenga impacto en los datos a utilizar (restricciones del proyecto y problema del negocio).

En el Diccionario de datos, se registran y describen los elementos que conforman la estructura de las fuentes de información identificadas para el proyecto, detallando el conjunto de posibles valores

para cada campo y su significado. Para ello, se analizan las fuentes de información disponibles (formalismo Fuentes de Información del Cliente), y se categoriza cada uno de los campos existentes de acuerdo al tipo de variable: si es cualitativa (expresa cualidades, características o modalidad), se analiza si dicha característica tiene algún criterio de orden clasificándose en ordinal o nominal, de acuerdo a si estas poseen o no ordenamiento alguno, respectivamente. Si es cuantitativa (expresa cantidades), se evalúa si presentan valores ininterrumpidos dentro de un intervalo (continuos) o no (discretos). Finalmente, se definen los campos, listando el rango de valores posibles y su interpretación (de ser necesario), así como cualquier otro aspecto relevante para comprender el significado de la variable y sus valores (por ejemplo: procedimiento de carga de los datos, valores por defecto de la herramienta, etc.).

En la técnica identificación de atributos relacionados con el problema de negocio, se realiza un proceso de identificación, evaluación y selección de las variables (o campos) que desde el punto de vista del negocio son significativos para el estudio de la problemática definida. A partir de la comprensión de los datos disponibles, se procede a evaluar las vinculaciones entre los datos, haciendo uso de diagramas de bases de datos (siendo recomendado generar el mismo, en caso que este no existiese), así como la utilidad de los mismos para los problemas de negocio identificados, teniendo en consideración la necesidad de generar atributos con mayor significado para el problema en cuestión, dejando registrado con precisión la lógica que se deberá utilizar para generar dicho atributo y sus valores, así como una descripción del mismo. La formulación de nuevas variables derivadas debe ser realizada de manera cuidadosa y justificada con la lógica del negocio, dado que de lo contrario, puede verse afectada la calidad de comprensión de los resultados [Siddiqi, 2012].

Se considera relevante destacar que la descripción de los campos a utilizar, puede verse restringida por el alcance del problema de negocio o las limitaciones del proyecto (por ejemplo: recortando el espacio de interés). Adicionalmente, del análisis de los datos y la interacción con los expertos, puede surgir la necesidad de generar nuevas variables (no existentes en las fuentes de información y por lo tanto en el diccionario de datos) que brinden mayor representatividad de la información que las mismas contienen.

En la siguiente sección se presenta la aplicación de la técnica junto con los formalismos, provistos en la sección previa, a la prueba de concepto.

4.4.2.1.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica Identificación de atributos relacionados con el Problema de Negocio que forma parte de la Metodología para la educción de

requerimientos para proyectos de explotación de información propuesta en [Britos et al., 2008], en conjunto con el diccionario de datos. En la prueba de conceptos, se utiliza como insumos la información externa provista por el experto: Discursos de los interesados (Fuente de Información 4.7), y los formalismos producidos en la fase de entendimiento del negocio: Fuentes de Información del Cliente (Tabla 4.55), Restricciones del Proyecto (Tabla 4.61) y Problema del Negocio (Tabla 4.64). Los formalismos indicados como elementos de entrada, son transcriptos con el mismo número de tabla, para facilitar al lector en la comprensión de la implementación de la técnica.

Discursos de los interesados: Entrevistas con el cliente (Silva H.)

Entrevista 4:

A partir de lo conversado en la entrevista previa y del análisis realizado de los datos, se determinó como variables relevantes las siguientes: la cantidad de personas en los agrupamientos urbanos, el rango de ingresos mensuales del hogar, el género de la persona, la edad, su nivel de instrucción académica, si posee personas en su entorno que consumen sustancias psicoactivas y el detalle de si consume cada una de las distintas sustancias psicoactivas. ¿Cree que alguna otra variable debería ser considerada?

"Prefiero que la edad del individuo sea utilizada con los rangos definidos por el instituto, siendo más representativo para los estudios realizados. Además, se debería incluir una variable adicional como indicador económico de la persona. El indicador de las necesidades básicas insatisfechas, nos permite profundizar en la situación económica del individuo junto con sus ingresos."

¿Desea que el análisis de consumo de sustancias psicoactivas sea realizado por tipo de droga?

"No, para un primer análisis prefiero evaluar a la población sin distinguir entre el tipo de sustancias que consumen, con excepción de aquellas que no son ilegales."

¿Es de interés realizar un análisis de las poblaciones por separado, ya sea por región u otro indicador geográfico?

"No, consideremos en primer lugar la población total (geográficamente hablando) y a partir de los resultados obtenidos, analizaremos la posibilidad de posteriormente ampliar el estudio."

¿Considera relevante la conformación de una variable que indique el interés o deseo de la persona de consumir este tipo de sustancias?

"Si bien acordamos utilizar las variables de autopercepción, para este trabajo preferimos no considerar esa información dado el nivel de subjetividad que posee y la ambigüedad en la información que brinda"

¿Qué variable considera como más representativa para evaluar el consumo de sustancias psicoactivas? Es decir, que periodo de prevalencia utilizaría para evaluar la población.

"Como les había mencionado, no discerniremos por la cantidad o frecuencia de consumo, dada la limitada información que se dispone al respecto. Ahora bien, respecto al periodo vamos a el de mayor cubrimiento."

¿Incluso para las sustancias socialmente aceptadas, utilizaremos la prevalencia de vida?

"Así es, para todas."

Fuente de Información 4.7. Prueba de Concepto - Entrevista 4

Fuentes de Información del Cliente					
Responsable:		Esposito E.		**Fecha:**	05/04/2016
ID#:		D.EN.ANN.FUIC		**Versión:**	1.0
ID	**Nombre**	**Categoría**	**Responsable**	**Descripción**	
fuic.1	Documento para la utilización de la base de datos usuario 2011	Documento	-	Describe distintas consideraciones acerca de la base de datos obtenida a partir de la encuesta ENPreCoSP 2011, indicando los métodos de recolección aplicados, los objetivos de la encuesta y la descripción de los datos.	
fuic.2	Cuestionario ENPreCoSP 2011	Planilla	-	Ejemplo de cuestionario ENPreCoSP 2011	
fuic.3	Base ENPreCoSP 2011	Almacén de datos	-	Almacén de registros de respuestas del cuestionario ENPreCoSP 2011 en formato txt (separado por el carácter "|") conformado por 34343 personas que respondieron 292 preguntas. El primer renglón contiene los nombres de los campos. Decimales separados por punto (.)	

Tabla 4.45 (Transcripta). Prueba de Concepto - Fuentes de Información del Cliente

Restricciones del Proyecto				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.ANN.REPR	**Versión:**	1.0	
Restricción	**Tipo**	**Descripción**	**Objetivo asociado**	**Referencia**
repr.1	datos	Se identifica un desbalance entre la cantidad de registros que han consumido distintas Sustancias Psicoactivas y quienes no han consumido	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1
repr.2	datos	Se carece de información detallada respecto al grado o frecuencia de consumo de las sustancias psicoactivas	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1

Tabla 4.61 (Transcripta). Prueba de Concepto - Restricciones del Proyecto

Problema del Negocio				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.PRNE	**Versión:**	1.0	
Objetivo del Proyecto	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos			
Problema	**Descripción**		**Experto**	**Referencia**
prne.1	Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		(rehi.3) Silva H.	Entrevista 3

Tabla 4.64 (Transcripta). Prueba de Concepto - Problema del Negocio

Diccionario de Fuente de Datos (D.ED.AnD.DiFD): A partir de las fuentes de información previamente identificadas (Fuentes de Información del Cliente), se procede a registrar en la columna “campo” el nombre de las variables que la componen. Luego se categorizan y definen cada

uno de los campos existentes de acuerdo al tipo de variable, listando el rango de valores posibles y su significado.

Para el caso de prueba de concepto escogido, en las tablas 4.68.a y 4.68.b se ilustra un subconjunto de los atributos disponibles, existiendo una descripción pública completa de los mismos disponible en [Instituto Nacional de Estadística y Censos, 2011].

Diccionario de Fuente de Datos				
Responsable:		Esposito E.	**Fecha:**	28/04/2016
ID#:		D.ED.AnD.DiFD	**Versión:**	1.0
Fuente de Información		(fuic.3) Base ENPreCoSP 2011		
Campo	**Tipo**	**Descripción**		
ABU_C	Nominal	**Abuso de cerveza** 1 Sí 2 No		
ABU_V	Nominal	**Abuso de vino** 1 Sí 2 No		
BHCH04	Nominal	**Sexo** 1 Varón 2 Mujer		
BHCH05	Discreta	**¿Cuál es su edad en años cumplidos?** (Edad en años cumplidos)		
BIAC01	Nominal	**¿Conoce personas cercanas a usted que en la actualidad consuman alguna sustancia como marihuana, cocaína, éxtasis, etc.?** 1 Sí 2 No 9 Ns/nc		
BIES04	Nominal	**¿Cuándo fue la primera vez que probó estimulantes sin indicación médica?** 1 Durante los últimos 30 días 2 Hace más de un mes, pero menos de un año 3 Hace más de un año 9 Ns/nc		
CAT_OCUP	Nominal	**Categoría ocupacional** 1 Patrón o empleador 2 Cuenta propia 3 Asalariado (excluye servicio doméstico) 4 Asalariado (sólo servicio doméstico) 5 Trabajador familiar sin pago		
CONDACT	Nominal	**Condición de actividad** 1 Ocupado 2 Desocupado 3 Inactivo		
GRUPEDAD	Ordinal	**Grupo de edad** 2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años 9 Ns/nc		
NBI_TOTAL	Ordinal	**INDICADORES DE NECESIDADES BÁSICAS INSATISFECHAS DE HOGAR: NBI Total** 0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI		
NIVINSTR	Ordinal	**Nivel de instrucción** 1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial		
POB_URB	Ordinal	**Agrupamiento de poblaciones urbanas** 1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes		

Tabla 4.68.a Prueba de Concepto - Diccionario de Fuente de Datos

PROV	Nominal	**Jurisdicción del país** 02 Ciudad Autónoma de Buenos Aires 06 Buenos Aires 10 Catamarca 14 Córdoba 18 Corrientes 22 Chaco 26 Chubut 30 Entre Ríos 34 Formosa 38 Jujuy 42 La Pampa 46 La Rioja 50 Mendoza 54 Misiones 58 Neuquén 62 Río Negro 66 Salta 70 San Juan 74 San Luis 78 Santa Cruz 82 Santa Fe 86 Santiago del Estero 94 Tierra del Fuego, Antártida Argentina e Islas del Atlántico Sur 90 Tucumán
…	…	…
PV_AL	Nominal	**Prevalencia de vida de consumo de alucinógenos** 1 Sí 2 No
PV_ANX	Nominal	**Prevalencia de vida de consumo de medicamentos para adelgazar (SIN INDICACIÓN MÉDICA)** 1 Sí 2 No
PV_BA	Nominal	**Prevalencia de vida de consumo de bebidas alcohólicas** 1 Sí 2 No
PV_CK	Nominal	**Prevalencia de vida de consumo de crack** 1 Sí 2 No
PV_CO	Nominal	**Prevalencia de vida de consumo de cocaína** 1 Sí 2 No
PV_ES	Nominal	**Prevalencia de vida de consumo de estimulantes** 1 Sí 2 No
PV_EX	Nominal	**Prevalencia de vida de consumo de éxtasis** 1 Sí 2 No
PV_IN	Nominal	**Prevalencia de vida de consumo de inhalables** 1 Sí 2 No
PV_MA	Nominal	**Prevalencia de vida de consumo de marihuana** 1 Sí 2 No
PV_OA	Nominal	**Prevalencia de vida de consumo de opiáceos y anestésicos** 1 Sí 2 No
PV_PB	Nominal	**Prevalencia de vida de consumo de pasta base** 1 Sí 2 No
PV_TA	Nominal	**Prevalencia de vida de consumo de tabaco** 1 Sí 2 No
PV_TR	Nominal	**Prevalencia de vida de consumo de tranquilizantes** 1 Sí 2 No
RANGOING	Ordinal	**Rango del Ingreso total mensual del hogar en pesos** 0 Sin ingresos 1 1 a 600 2 601 a 800 3 801 a 1.000 4 1.001 a 1.500 5 1.501 a 2.000 6 2.001 a 2.500 7 2.501 a 3.000 8 3.001 a 3.500 9 3.501 a 4.000
REGION	Nominal	**Región estadística** 1 Gran Buenos Aires (Ciudad Autónoma de Bs. As. y 24 Partidos del GBA) 2 Pampeana (Resto de Buenos Aires, Córdoba, La Pampa, Santa Fe y Entre Ríos) 3 Noroeste (Catamarca, Jujuy, La Rioja, Salta, Santiago del Estero y Tucumán) 4 Noreste (Corrientes, Chaco, Formosa y Misiones) 5 Cuyo (Mendoza, San Juan y San Luis) 6 Patagónica (Chubut, Neuquén, Río Negro, Santa Cruz y Tierra del Fuego)

Tabla 4.68.b Prueba de Concepto - Diccionario de Fuente de Datos

Campos Relacionados con el Problema de Negocio (D.ED.AnD.CRPN): A partir de la comprensión de los datos disponibles, se evalúa de forma conjunta con el experto las variables

relevantes para el problema de negocio identificado, así como aquellas que sea necesario construir a partir de otros campos (precisando el procedimiento para su generación).

Como resultado de este paso, se listan todos los atributos que serán considerados para incluir en el modelo (en la columna "nombre"), se identifica al atributo "consumo_ilegal_bool" como atributo a generar, brindando una definición del mismo e indicando las reglas para su construcción (en la columna "descripción") y por último, se definen las referencias a partir de la cuales se identifica el interés de cada variable para el problema de negocio (en la columna "Referencia").

Las tablas 4.69.a y 4.69.b ilustran la selección de campos relacionados con el problema de negocio identificado en la prueba de concepto.

Campos Relacionados con el Problema de Negocio			
Responsable:	Esposito E.	**Fecha:**	29/04/2016
ID#:	D.ED.AnD.CRPN	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		

Nombre	**Generar**	**Descripción**	**Referencia**
POB_URB		**Agrupamiento de poblaciones urbanas** 1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes	(fuic.3) Base ENPreCoSP 2011
RANGOING		**Rango del Ingreso total mensual del hogar en pesos** 0 Sin ingresos 1 1 a 600 2 601 a 800 3 801 a 1.000 4 1.001 a 1.500 5 1.501 a 2.000 6 2.001 a 2.500 7 2.501 a 3.000 8 3.001 a 3.500 9 3.501 a 4.000 10 4.001 a 4.500 11 4.501 a 5.500 12 5.001 a 6.000 13 6.001 a 7.000 14 7.001 a 8.000 15 8.001 a 10.000 16 10.001 a 15.000 17 15.001 y más 99 Ns/nc	(fuic.3) Base ENPreCoSP 2011
NBI_TOTAL		**INDICADORES DE NECESIDADES BÁSICAS INSATISFECHAS DE HOGAR: NBI Total** 0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
BHCH04		**Sexo** 1 Varón 2 Mujer	(fuic.3) Base ENPreCoSP 2011
GRUPEDAD		**Grupo de edad** 2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años 9 Ns/nc	(fuic.3) Base ENPreCoSP 2011 / entrevista 4

Tabla 4.69.a Prueba de Concepto - Campos Relacionados con el Problema de Negocio

NIVINSTR		**Nivel de instrucción** 1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial	(fuic.3) Base ENPreCoSP 2011
BIAC01		**¿Conoce personas cercanas a usted que en la actualidad consuman alguna sustancia como marihuana, cocaína, éxtasis, etc.?** 1 Sí 2 No 9 Ns/nc	(fuic.3) Base ENPreCoSP 2011
PV_TA		**Prevalencia de vida de consumo de tabaco** 1 Sí 2 No	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
PV_BA		**Prevalencia de vida de consumo de bebidas alcohólicas** 1 Sí 2 No	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
consumo_ilegal_bool	x	**Prevalencia de vida de consumo de alguna de las sustancias psicoactivas ilegales: si alguna de las variables (PV_ES, PV_MA, PV_CO, PV_PB, PV_EX, PV_IN, PV_ANX, PV_OA, PV_CK, PV_AL) es igual a 1 (SI) entonces 1 sino 0.** 1 Sí 0 No	entrevista 4

Tabla 4.69.b Prueba de Concepto - Campos Relacionados con el Problema de Negocio

4.4.2.2. Actividad: Exploración de los Datos (D.ED.ExD)

En esta actividad se analizan los valores de los campos identificados de interés para los distintos problemas de negocio, con el objetivo de comprender las características de la población o muestra de estudio, identificando relaciones iniciales entre las distintas variables estudiadas. El resultado de esta actividad contribuye en la comprensión de la calidad de los datos.

Información de Entrada

- Suposiciones del Proyecto (D.EN.AnN.SuPr)
- Restricciones del Proyecto (D.EN.AnN.RePr)
- Problema del Negocio (D.EN.CPN.PrNe)
- Diccionario de Fuente de Datos (D.ED.AnD.DiFD)
- Campos Relacionados con el Problema de Negocio (D.ED.AnD.CRPN)
- Reporte de Evaluación de Herramientas (G.In.EvS.EvHe)

Información de Salida

- Fuente Integrada de datos (D.ED.ExD.FuID)
- Reporte de Datos Explorados (D.ED.ExD.ReDE)

4.4.2.2.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se propone el Reporte de Datos Explorados, el cual se presenta a continuación.

Reporte de Datos Explorados (D.ED.ExD.ReDE): Se formaliza la descripción de la población o muestra disponible, centrado en las variables identificadas como relevantes para el problema de negocio. Para ello, se define el problema de negocio (en la fila homónima) y se realiza un análisis estadístico de los datos el cual varía de acuerdo a la característica de los mismos. Para las variables discretas se indica el nombre y la distribución de los distintos valores posibles en la fuente de datos. Para las variables continuas se indican los valores mínimos y máximos, y distintas medidas de tendencia central y dispersión. La tabla 4.70 ilustra la estructura del formalismo previamente descripto.

Reporte de Datos Explorados				
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización	
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión	
Problema de Negocio	Identificador problema de negocio			
ATRIBUTOS CUALITATIVOS				
Nombre	**Valores**	**Distribución**		
Atributo cualitativo 1	Posibles valores del atributo cualitativo 1	Cantidad y frecuencia de aparición de cada valor del atributo cualitativo 1		
…	…	…		
Atributo cualitativo N	Posibles valores del atributo cualitativo N	Cantidad y frecuencia de aparición de cada valor del atributo cualitativo N		
ATRIBUTOS CUANTITATIVOS				
Nombre	**Min**	**Max**	**Tendencia Central**	**Dispersión**
Atributo cuantitativo 1	Valor mínimo	Valor máximo	Medida/s de tendencia central	Medida/s de dispersión
…	…	…	…	…
Atributo cuantitativo N	Valor mínimo	Valor máximo	Medida/s de tendencia central	Medida/s de dispersión
Comentarios: Se indica si se anexan visualizaciones que amplíen la descripción de las variables				

Tabla 4.70. Formalismo: Reporte de Datos Explorados

4.4.2.2.2. Técnica Identificada

Para el desarrollo de esta actividad se propone la técnica de **"Exploración de los Datos"**, que a partir de la integración de los campos relevantes para el problema de negocio, se realiza una descripción detallada de la información disponible, la cual permita comprender las características de la población o muestra sobre la cual se va a obtener los patrones de conocimiento e identificar

relaciones iniciales entre las variables. Adicionalmente, es posible entender mediante esta técnica la complejidad del modelo requerido con respecto a los límites de decisión.

La aplicación de la técnica consiste en caracterizar las variables del problema de negocio, haciendo uso de distintas medidas de estadística descriptiva y de visualizaciones para comprender con el mayor detalle posible la información estudiada. Para ello, se debe tener en consideración la naturaleza de la información que cada variable y sus valores representan, es decir, el tipo de variable (cuantitativa o cualitativa), a partir de lo cual se implementan distintas herramientas de análisis.

Para aquellas variables cualitativas, se registran los valores presentes en la fuente de información y la frecuencia de apariciones de los mismos, indicando la cantidad y el porcentaje de representación. Para las variables cuantitativas, se registran los valores extremos presentes en la fuente de información (mínimo y máximo), y medidas estadísticas de tendencia central (media, mediana, moda, entre otras) y de dispersión (cuartiles, desvió estándar, entre otras). La mínima información reflejada es también conocida como el resumen de cinco datos, el cual brinda una descripción completa de la forma de distribución de los mismos [Han et al., 2011].

De forma complementaria, en caso que se considere de utilidad respecto a la información que aporten, se incorporan visualizaciones de una variable (histograma, gráfica de caja, violín, barras y poligonal), así como de más de una variable (gráfica o matriz de dispersión, mapa de calor o correlación y coordenadas paralelas), indicando su uso en la fila de “comentarios”.

La técnica utiliza como elementos de entrada: el problema de negocio, campos relacionados con el problema de negocio y el diccionario de fuentes de datos, a partir de las cuales se realiza el análisis y se identifican las variables a utilizar y su significado. Las suposiciones del proyecto, brindan una preconcepción de los datos, las cuales pueden ser confirmadas o rectificadas de acuerdo al análisis realizado. Mientras que la herramienta seleccionada limita los posibles instrumentos a utilizar.

En la siguiente sección se presenta la aplicación de la técnica junto con los formalismos, provistos en la sección previa, a la prueba de concepto.

4.4.2.2.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica Exploración de los Datos. Dicha técnica utiliza como insumos los formalismos: Suposiciones del Proyecto (Tabla 4.60), Restricciones del Proyecto (Tabla 4.61), Problema del Negocio (Tabla 4.64), Diccionario de Fuente

de Datos (Tablas 4.68.a y 4.68.b), Campos Relacionados con el Problema de Negocio (Tablas 4.69.a y 4.69.b) y Reporte de Evaluación de Herramientas (Tablas 4.13.a y 4.13.b).

Los formalismos identificados como elementos de entrada, son transcriptos con el mismo número de tabla, para facilitar al lector en la comprensión de la aplicación de la técnica.

Suposiciones del Proyecto			
Responsable:	Esposito E.	**Fecha:**	20/04/2016
ID#:	D.EN.ANN.SUPR	**Versión:**	1.1

Suposición	**Descripción**	**Objetivo asociado**	**Referencia**
supr.1	Los cuestionarios y la carga de la información se ha realizado de manera correcta	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1
supr.2	Las conductas de consumo se considerarán como análogas sin importar la gradualidad del mismo	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1
supr.3	Las variables vinculadas con autopercepción brindan información fiable respecto al entorno real del individuo	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 2
supr.4	El proceso de diseño de la muestra es representativo a nivel nacional y provincial	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 2 / fuic.1

Tabla 4.60 (Transcripta). Prueba de Concepto - Suposiciones del Proyecto

Restricciones del Proyecto			
Responsable:	Esposito E.	**Fecha:**	20/04/2016
ID#:	D.EN.ANN.REPR	**Versión:**	1.0

Restricción	**Tipo**	**Descripción**	**Objetivo asociado**	**Referencia**
repr.1	datos	Se identifica un desbalance entre la cantidad de registros que han consumido distintas Sustancias Psicoactivas y quienes no han consumido	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1
repr.2	datos	Se carece de información detallada respecto al grado o frecuencia de consumo de las sustancias psicoactivas	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1

Tabla 4.61 (Transcripta). Prueba de Concepto - Restricciones del Proyecto

Problema del Negocio				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.PRNE	**Versión:**	1.0	
Objetivo del Proyecto	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos			
Problema	**Descripción**		**Experto**	**Referencia**
prne.1	Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		(rehi.3) Silva H.	Entrevista 3

Tabla 4.64 (Transcripta). Prueba de Concepto - Problema del Negocio

Diccionario de Fuente de Datos				
Responsable:	Esposito E.		**Fecha:**	28/04/2016
ID#:	D.ED.AnD.DiFD		**Versión:**	1.0
Fuente de Información	(fuic.3) Base ENPreCoSP 2011			
Campo	**Tipo**	**Descripción**		
ABU_C	Nominal	**Abuso de cerveza** 1 Sí 2 No		
ABU_V	Nominal	**Abuso de vino** 1 Sí 2 No		
BHCH04	Nominal	**Sexo** 1 Varón 2 Mujer		
BHCH05	Discreta	**¿Cuál es su edad en años cumplidos?** (Edad en años cumplidos)		
BIAC01	Nominal	**¿Conoce personas cercanas a usted que en la actualidad consuman alguna sustancia como marihuana, cocaína, éxtasis, etc.?** 1 Sí 2 No 9 Ns/nc		
BIES04	Nominal	**¿Cuándo fue la primera vez que probó estimulantes sin indicación médica?** 1 Durante los últimos 30 días 2 Hace más de un mes, pero menos de un año 3 Hace más de un año 9 Ns/nc		
CAT_OCUP	Nominal	**Categoría ocupacional** 1 Patrón o empleador 2 Cuenta propia 3 Asalariado (excluye servicio doméstico) 4 Asalariado (sólo servicio doméstico) 5 Trabajador familiar sin pago		
CONDACT	Nominal	**Condición de actividad** 1 Ocupado 2 Desocupado 3 Inactivo		
GRUPEDAD	Ordinal	**Grupo de edad** 2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años 9 Ns/nc		
NBI_TOTAL	Ordinal	**INDICADORES DE NECESIDADES BÁSICAS INSATISFECHAS DE HOGAR: NBI Total** 0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI		
NIVINSTR	Ordinal	**Nivel de instrucción** 1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial		
POB_URB	Ordinal	**Agrupamiento de poblaciones urbanas** 1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes		

Tabla 4.68.a (Transcripta). Prueba de Concepto - Diccionario de Fuente de Datos

PROV	Nominal	**Jurisdicción del país** 02 Ciudad Autónoma de Buenos Aires 06 Buenos Aires 10 Catamarca 14 Córdoba 18 Corrientes 22 Chaco 26 Chubut 30 Entre Ríos 34 Formosa 38 Jujuy 42 La Pampa 46 La Rioja 50 Mendoza 54 Misiones 58 Neuquén 62 Río Negro 66 Salta 70 San Juan 74 San Luis 78 Santa Cruz 82 Santa Fe 86 Santiago del Estero 94 Tierra del Fuego, Antártida Argentina e Islas del Atlántico Sur 90 Tucumán
…	…	…
PV_AL	Nominal	**Prevalencia de vida de consumo de alucinógenos** 1 Sí 2 No
PV_ANX	Nominal	**Prevalencia de vida de consumo de medicamentos para adelgazar (SIN INDICACIÓN MÉDICA)** 1 Sí 2 No
PV_BA	Nominal	**Prevalencia de vida de consumo de bebidas alcohólicas** 1 Sí 2 No
PV_CK	Nominal	**Prevalencia de vida de consumo de crack** 1 Sí 2 No
PV_CO	Nominal	**Prevalencia de vida de consumo de cocaína** 1 Sí 2 No
PV_ES	Nominal	**Prevalencia de vida de consumo de estimulantes** 1 Sí 2 No
PV_EX	Nominal	**Prevalencia de vida de consumo de éxtasis** 1 Sí 2 No
PV_IN	Nominal	**Prevalencia de vida de consumo de inhalables** 1 Sí 2 No
PV_MA	Nominal	**Prevalencia de vida de consumo de marihuana** 1 Sí 2 No
PV_OA	Nominal	**Prevalencia de vida de consumo de opiáceos y anestésicos** 1 Sí 2 No
PV_PB	Nominal	**Prevalencia de vida de consumo de pasta base** 1 Sí 2 No
PV_TA	Nominal	**Prevalencia de vida de consumo de tabaco** 1 Sí 2 No
PV_TR	Nominal	**Prevalencia de vida de consumo de tranquilizantes** 1 Sí 2 No
RANGOING	Ordinal	**Rango del Ingreso total mensual del hogar en pesos** 0 Sin ingresos 1 1 a 600 2 601 a 800 3 801 a 1.000 4 1.001 a 1.500 5 1.501 a 2.000 6 2.001 a 2.500 7 2.501 a 3.000 8 3.001 a 3.500 9 3.501 a 4.000
REGION	Nominal	**Región estadística** 1 Gran Buenos Aires (Ciudad Autónoma de Bs. As. y 24 Partidos del GBA) 2 Pampeana (Resto de Buenos Aires, Córdoba, La Pampa, Santa Fe y Entre Ríos) 3 Noroeste (Catamarca, Jujuy, La Rioja, Salta, Santiago del Estero y Tucumán) 4 Noreste (Corrientes, Chaco, Formosa y Misiones) 5 Cuyo (Mendoza, San Juan y San Luis) 6 Patagónica (Chubut, Neuquén, Río Negro, Santa Cruz y Tierra del Fuego)

Tabla 4.68.b (Transcripta). Prueba de Concepto - Diccionario de Fuente de Datos

Campos Relacionados con el Problema de Negocio			
Responsable:	Esposito E.	**Fecha:**	29/04/2016
ID#:	D.ED.AnD.CRPN	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		

Nombre	Generar	Descripción	Referencia
POB_URB		**Agrupamiento de poblaciones urbanas** 1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes	(fuic.3) Base ENPreCoSP 2011
RANGOING		**Rango del Ingreso total mensual del hogar en pesos** 0 Sin ingresos 1 1 a 600 2 601 a 800 3 801 a 1.000 4 1.001 a 1.500 5 1.501 a 2.000 6 2.001 a 2.500 7 2.501 a 3.000 8 3.001 a 3.500 9 3.501 a 4.000 10 4.001 a 4.500 11 4.501 a 5.500 12 5.001 a 6.000 13 6.001 a 7.000 14 7.001 a 8.000 15 8.001 a 10.000 16 10.001 a 15.000 17 15.001 y más 99 Ns/nc	(fuic.3) Base ENPreCoSP 2011
NBI_TOTAL		**INDICADORES DE NECESIDADES BÁSICAS INSATISFECHAS DE HOGAR: NBI Total** 0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
BHCH04		**Sexo** 1 Varón 2 Mujer	(fuic.3) Base ENPreCoSP 2011
GRUPEDAD		**Grupo de edad** 2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años 9 Ns/nc	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
NIVINSTR		**Nivel de instrucción** 1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial	(fuic.3) Base ENPreCoSP 2011
BIAC01		**¿Conoce personas cercanas a usted que en la actualidad consuman alguna sustancia como marihuana, cocaína, éxtasis, etc.?** 1 Sí 2 No 9 Ns/nc	(fuic.3) Base ENPreCoSP 2011

Tabla 4.69.a (Transcripta). Prueba de Concepto - Campos Relacionados con el Problema de Negocio

PV_TA		**Prevalencia de vida de consumo de tabaco** 1 Sí 2 No	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
PV_BA		**Prevalencia de vida de consumo de bebidas alcohólicas** 1 Sí 2 No	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
consumo_ilegal_bool	x	**Prevalencia de vida de consumo de alguna de las sustancias psicoactivas ilegales: si alguna de las variables (PV_ES, PV_MA, PV_CO, PV_PB, PV_EX, PV_IN, PV_ANX, PV_OA, PV_CK, PV_AL) es igual a 1 (SI) entonces 1 sino 0.** 1 Sí 0 No	entrevista 4

Tabla 4.69.b (Transcripta). Prueba de Concepto - Campos Relacionados con el Problema de Negocio

Reporte de Evaluación de Herramientas					
Responsable:	Rodriguez H.			**Fecha:**	07/04/2016
ID#:	G.In.EvS.REHe			**Versión:**	1.0
Criterios: **Evaluación:** 1 = Malo, 2 = débil, 3 = Bueno, 4 = Excelente					1 = No, 4 = SI
Herramientas			**Tanagra V.1.4.50**	**Weka V.3.7.11**	**Orange V.2.7.8**
1. Funcional - Características Técnicas					
Soporte de Metodología / Ciclo de vida	Soporte del proceso	3	2	2	2
Compatibilidad con fuentes de datos	Base de datos	8	--	--	--
	Otras fuentes (word, excel, etc.)	8	3	2	3
Integración	Soporte de distintas técnicas asociadas al proceso de explotación de Información	5	4	4	4
Multilenguaje	Soporta distintas idiomas	2	1	1	1
Técnicas	Variedad de técnicas que provee	18	4	4	4
Reporte y visualización	Permite generar reportes y visualizaciones	12	2	2	2
Multiplataforma	Soporta múltiples plataformas	5	1	4	4
Instalación remota	La administración y mantenimiento son remotos	5	--	--	--
Usuarios Múltiples	Posee perfiles de usuarios	2	1	1	1
Seguridad	Provee seguridad de la información configurada por perfiles	2	1	1	1
Backup	Metodología de backup	2	1	1	1
Amigable	Interfaz de usuario	10	4	2	4
Configuraciones	Permite la configuración del perfil	8			
Documentación	Servicio de soporte y ayuda	5	4	1	3
Conexión	Soporta conexión por: Internet, FTP, ERPs.	2	1	1	1
Soporte de sistemas de mensaje	Soporta compartir información (por mail u otro medio)	3	1	1	1
Total			224	196	234
	Peso del Grupo	40%	89,6	78,4	93,6

Tabla 4.13.b (Transcripta). Prueba de Concepto - Reporte de Evaluación de Herramientas

2. Características del Proveedor					
Características del proveedor	Historia	30	3	3	1
Crecimiento	Perspectiva a futuro	10	2	3	2
Ubicación Geográfica	Oficinas	30	--	--	--
Implementación	Otras implementaciones de la misma herramienta	5	--	--	--
	Contacto con otros clientes	5	--	--	--
Confidencialidad	Confidencialidad de la información	20	--	--	--
Total			110	120	50
	Peso del Grupo	25%	27,5	30	12,5
3. Características del Servicio					
Garantía del producto	Duración y Alcance	30	--	--	--
Mejora	Brinda soporte a versiones previas	20	1	1	1
Licencia	Costo, alcances y soporte postventa	30	--	--	--
Soporte	Tiempo de respuesta y disponibilidad	20	--	--	--
Total			20	20	20
	Peso del Grupo	20%	4	4	4
4. Características Económicas					
Costo del software	Costo de la herramienta	30	--	--	--
Costo del Hardware	Necesidad de mejorar o comprar nuevo hardware compatible con la herramienta	20	--	--	--
Otros costos software	Costos adicionales al producto (backup, web servers, bases de datos, etc.)	20	--	--	--
Licencias	Política de licencia	10	--	--	--
Financiamiento	Existencia	10	--	--	--
Mejoras	Costo promedio de la mejora del producto	10	--	--	--
Total			0	0	0
	Peso del Grupo	-15%	0	0	0
Final					
1. Funcional - Características Técnicas		40%	89,6	78,4	93,6
2. Características del Proveedor		25%	27,5	30	12,5
3. Características del Servicio		20%	4	4	4
4. Características Económicas		-15%	0	0	0
TOTAL			**121,1**	**112,4**	**110,1**

Tabla 4.13.b (Transcripta). Prueba de Concepto - Reporte de Evaluación de Herramientas

Fuente Integrada de datos (D.ED.ExD.FuID): A partir del conjunto de campos identificados de interés para el problema de negocio, así como las suposiciones y restricciones vinculadas con el problema de negocio, se procede a integrar los campos a utilizar en una única fuente de información, la cual posee el conjunto de registros de interés con los campos sin alterar. Es una

práctica recomendada, utilizar o incorporar un campo identificador que permita mantener la trazabilidad de los registros con respecto a la fuente de almacenamiento original. En la prueba de concepto, si bien los datos se encontraban integrados en un único elemento, se genera la fuente integrada de datos, reduciendo el conjunto de variables a utilizar a 10 (aquellas indicadas en los campos relacionados al problema de negocio, generando la variable "consumo_ilegal_bool"). La figura 4.12 ilustra la estructura de la fuente de integrada de datos mediante su representación en un Diagrama Entidad-Relación.

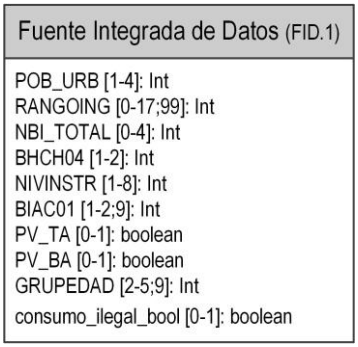


Figura 4.12. Prueba de Concepto – Fuente Integrada de Datos (Diagrama Entidad-Relación)

Reporte de Datos Explorados (D.ED.ExD.ReDE): A partir de la fuente integrada de datos, se describe la distribución de valores para cada atributo significativo para el problema de negocio, resaltando la cantidad de registros y la proporción que los mismos representan con respecto a la muestra. En la prueba de concepto, todos los atributos a utilizar son del tipo cualitativos, por lo que se registró la distribución de cada uno de sus valores. Las tabla 4.71.a y 4.71.b ilustran el resultado obtenido. Para simplificar el formalismo resultante, se omitió la sección vacía de atributos cuantitativos.

Reporte de Datos Explorados			
Responsable:	Esposito E.	**Fecha:**	06/05/2016
ID#:	D.ED.ExD.ReDE	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		
ATRIBUTOS CUALITATIVOS			
Nombre	**Valores**	**Distribución**	
POB_URB	1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes	1190 (3,47%) 3749 (10,92%) 14282 (41,59%) 15122 (44,03%)	

Tabla 4.71.a Prueba de Concepto - Reporte de Datos Explorados

RANGOING	0 Sin ingresos 1 1 a 600 2 601 a 800 3 801 a 1.000 4 1.001 a 1.500 5 1.501 a 2.000 6 2.001 a 2.500 7 2.501 a 3.000 8 3.001 a 3.500 9 3.501 a 4.000 10 4.001 a 4.500 11 4.501 a 5.500 12 5.001 a 6.000 13 6.001 a 7.000 14 7.001 a 8.000 15 8.001 a 10.000 16 10.001 a 15.000 17 15.001 y más 99 Ns/nc	138 (0,4%) 971 (2,83%) 1044 (3,04%) 1621 (4,72%) 3136 (9,13%) 3986 (11,61%) 2689 (7,83%) 3757 (10,94%) 1716 (5%) 2575 (7,5%) 985 (2,87%) 2120 (6,17%) 1748 (5,09%) 1246 (3,63%) 1071 (3,12%) 1348 (3,93%) 934 (2,72%) 464 (1,35%) 2794 (8,14%)
NBI_TOTAL	0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI	29813 (86,81%) 3670 (10,69%) 743 (2,16%) 113 (0,33%) 4 (0,01%)
BHCH04	1 Varón 2 Mujer	15787 (45,97%) 18556 (54,03%)
NIVINSTR	1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial	372 (1,08%) 2484 (7,23%) 6309 (18,37%) 7092 (20,65%) 8103 (23,59%) 4486 (13,06%) 5454 (15,88%) 43 (0,13%)
BIAC01	1 Sí 2 No 9 Ns/nc	7970 (23,21%) 26276 (76,51%) 97 (0,28%)
PV_TA	1 Sí 0 No	17636 (51,35%) 16707 (48,65%)
PV_BA	1 Sí 0 No	25709 (74,86%) 8634 (25,14%)
consumo_ilegal_bool	1 Sí 0 No	30850 (89,83%) 3493 (10,17%)
GRUPEDAD	2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años 9 Ns/nc	6052 (19,26%) 8130 (25,87%) 9420 (29,98%) 7824 (24,90%) 0 (0%)
Comentarios:		

Tabla 4.71.b Prueba de Concepto - Reporte de Datos Explorados

4.4.2.3. Actividad: Evaluación de los Datos (D.ED.EvD)

En esta actividad se analizan los campos de interés para los distintos problemas de negocio, identificando aquellas características que puedan afectar la calidad del modelo. Es decir, registrar

variables y/o valores que posean datos incorrectos, nulos (considerando si dicho valor posee o no significancia), atípicos (outliers) o valores que puedan ser considerados como anómalos.

Información de Entrada

- Diccionario de Fuente de Datos (D.ED.AnD.DiFD)
- Campos Relacionados con el Problema de Negocio (D.ED.AnD.CRPN)
- Reporte de Datos Explorados (D.ED.ExD.ReDE)
- Fuente Integrada de datos (D.ED.ExD.FuID)
- Reporte de Evaluación de Herramientas (G.In.EvS.EvHe)

Información de Salida

- Reporte de la Calidad de los Datos (D.ED.EvD.ReCD)

4.4.2.3.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se propone el Reporte de la Calidad de los Datos, el cual se presenta a continuación.

Reporte de la Calidad de los Datos (D.ED.EvD.ReCD): En este formalismo se identifican aquellos atributos y campos que presenten características anómalas, las cuales deban ser consideradas previamente a la generación del modelo, registrando el tipo de problemática identificada (nulos, outliers o inconsistencias), los registros afectados por la misma y su descripción en las columnas homónimas. La tabla 4.72 ilustra la estructura del formalismo previamente descripto.

Reporte de la Calidad de los Datos			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Problema de Negocio	Identificador del Problema de Negocio		
Nombre	**Registros**	**Tipo**	**Descripción**
Nombre del campo 1	Cantidad y/o identificadores de los registros	Tipo de problemática identificada	Detalle de la problemática identificada
…	…	…	…
Nombre del campo N	Cantidad y/o identificadores de los registros	Tipo de problemática identificada	Detalle de la problemática identificada

Tabla 4.72. Formalismo: Reporte de la Calidad de los Datos

4.4.2.3.2. Técnicas Identificadas

Para el desarrollo de esta actividad se propone la técnica **Exploración de la Calidad de los Datos**, que a partir de la integración de los campos relevantes para el problema de negocio, se realiza una

evaluación detallada de la información disponible, comprendiendo aquellas características de los datos que impliquen un problema o riesgo en la calidad de los mismos y por consiguiente afecten la calidad del modelo y los resultados derivados.

La aplicación de la técnica consiste en, a partir del conocimiento previamente obtenido del negocio y en particular de los datos, detectar mediante el uso de estadística descriptiva (distribución/cuartiles), visualizaciones (histogramas, diagrama de cajas, entre otros) y/o procedimientos de detección de anomalías en los datos (por ejemplo: Procedimientos de explotación de información para la identificación de datos faltantes con ruido e inconsistentes [Kuna, H. 2013]), los campos o registros anómalos (registrándose en la columna "nombre"), dejando constancia de la existencia de los mismos, detallando la cantidad de registros involucrados, el tipo de anomalía identificada y una descripción de la misma (así como las posibles acciones a tomar), registrándose en las columnas homónimas.

En esta etapa, no se realiza el tratamiento de las anomalías detectadas, dado que estas deben ser evaluadas de acuerdo al modelo a utilizar y el significado de las mismas (por ejemplo: en ocasiones la ausencia de valores puede significar información relevante que deba ser contemplada por el modelo). En este contexto, la información registrada impactará en la evaluación de la correspondencia y utilidad de los campos en el modelo a generar, así como de los campos y/o registros para la fase de preparación de los datos.

La técnica utiliza como elementos de entrada: el diccionario de fuente de datos y los campos relacionados con el problema de negocio, como herramientas para comprender los distintos aspectos de los datos, el reporte de datos explorados como elemento de soporte para la identificación de problemáticas iniciales en los datos (presencia de nulos y outliers) y la fuente integrada de datos, sobre la cual se profundizará en el análisis. El reporte de evaluación de herramientas brinda una visión de los instrumentos disponibles para la actividad. En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo para la prueba de concepto.

4.4.2.3.3. Ejecución de la Actividad en la Prueba de Concepto

En el presente apartado, se exhiben los resultados obtenidos de aplicar la técnica Exploración de la Calidad de los Datos en la prueba de concepto, para el cual se utilizan como elementos de entrada los formalismos: Diccionario de Fuente de Datos (Tablas 4.68.a y 4.68.b), Campos Relacionados con el Problema de Negocio (Tablas 4.69.a y 4.68.b), Reporte de Datos Explorados (Tablas 4.71.a y 4.71.b), Fuente Integrada de datos (Figura 4.12) y Reporte de Evaluación de Herramientas (Tablas

4.13.a y 4.13.b). Dichos formalismos, los cuales fueron generados en etapas previas del proceso, son transcriptos con el mismo número de referencia, para facilitar al lector en la comprensión de la aplicación de la técnica.

Diccionario de Fuente de Datos				
Responsable:		Esposito E.	**Fecha:**	28/04/2016
ID#:		D.ED.AnD.DiFD	**Versión:**	1.0
Fuente de Información		(fuic.3) Base ENPreCoSP 2011		
Campo	**Tipo**	**Descripción**		
ABU_C	Nominal	**Abuso de cerveza** 1 Sí 2 No		
ABU_V	Nominal	**Abuso de vino** 1 Sí 2 No		
BHCH04	Nominal	**Sexo** 1 Varón 2 Mujer		
BHCH05	Discreta	**¿Cuál es su edad en años cumplidos?** (Edad en años cumplidos)		
BIAC01	Nominal	**¿Conoce personas cercanas a usted que en la actualidad consuman alguna sustancia como marihuana, cocaína, éxtasis, etc.?** 1 Sí 2 No 9 Ns/nc		
BIES04	Nominal	**¿Cuándo fue la primera vez que probó estimulantes sin indicación médica?** 1 Durante los últimos 30 días 2 Hace más de un mes, pero menos de un año 3 Hace más de un año 9 Ns/nc		
CAT_OCUP	Nominal	**Categoría ocupacional** 1 Patrón o empleador 2 Cuenta propia 3 Asalariado (excluye servicio doméstico) 4 Asalariado (sólo servicio doméstico) 5 Trabajador familiar sin pago		
CONDACT	Nominal	**Condición de actividad** 1 Ocupado 2 Desocupado 3 Inactivo		
GRUPEDAD	Ordinal	**Grupo de edad** 2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años 9 Ns/nc		
NBI_TOTAL	Ordinal	**INDICADORES DE NECESIDADES BÁSICAS INSATISFECHAS DE HOGAR: NBI Total** 0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI		
NIVINSTR	Ordinal	**Nivel de instrucción** 1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial		
POB_URB	Ordinal	**Agrupamiento de poblaciones urbanas** 1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes		

Tabla 4.68.a (Transcripta). Prueba de Concepto - Diccionario de Fuente de Datos

PROV	Nominal	**Jurisdicción del país** 02 Ciudad Autónoma de Buenos Aires 06 Buenos Aires 10 Catamarca 14 Córdoba 18 Corrientes 22 Chaco 26 Chubut 30 Entre Ríos 34 Formosa 38 Jujuy 42 La Pampa 46 La Rioja 50 Mendoza 54 Misiones 58 Neuquén 62 Río Negro 66 Salta 70 San Juan 74 San Luis 78 Santa Cruz 82 Santa Fe 86 Santiago del Estero 94 Tierra del Fuego, Antártida Argentina e Islas del Atlántico Sur 90 Tucumán
…	…	…
PV_AL	Nominal	**Prevalencia de vida de consumo de alucinógenos** 1 Sí 2 No
PV_ANX	Nominal	**Prevalencia de vida de consumo de medicamentos para adelgazar (SIN INDICACIÓN MÉDICA)** 1 Sí 2 No
PV_BA	Nominal	**Prevalencia de vida de consumo de bebidas alcohólicas** 1 Sí 2 No
PV_CK	Nominal	**Prevalencia de vida de consumo de crack** 1 Sí 2 No
PV_CO	Nominal	**Prevalencia de vida de consumo de cocaína** 1 Sí 2 No
PV_ES	Nominal	**Prevalencia de vida de consumo de estimulantes** 1 Sí 2 No
PV_EX	Nominal	**Prevalencia de vida de consumo de éxtasis** 1 Sí 2 No
PV_IN	Nominal	**Prevalencia de vida de consumo de inhalables** 1 Sí 2 No
PV_MA	Nominal	**Prevalencia de vida de consumo de marihuana** 1 Sí 2 No
PV_OA	Nominal	**Prevalencia de vida de consumo de opiáceos y anestésicos** 1 Sí 2 No
PV_PB	Nominal	**Prevalencia de vida de consumo de pasta base** 1 Sí 2 No
PV_TA	Nominal	**Prevalencia de vida de consumo de tabaco** 1 Sí 2 No
PV_TR	Nominal	**Prevalencia de vida de consumo de tranquilizantes** 1 Sí 2 No
RANGOING	Ordinal	**Rango del Ingreso total mensual del hogar en pesos** 0 Sin ingresos 1 1 a 600 2 601 a 800 3 801 a 1.000 4 1.001 a 1.500 5 1.501 a 2.000 6 2.001 a 2.500 7 2.501 a 3.000 8 3.001 a 3.500 9 3.501 a 4.000
REGION	Nominal	**Región estadística** 1 Gran Buenos Aires (Ciudad Autónoma de Bs. As. y 24 Partidos del GBA) 2 Pampeana (Resto de Buenos Aires, Córdoba, La Pampa, Santa Fe y Entre Ríos) 3 Noroeste (Catamarca, Jujuy, La Rioja, Salta, Santiago del Estero y Tucumán) 4 Noreste (Corrientes, Chaco, Formosa y Misiones) 5 Cuyo (Mendoza, San Juan y San Luis) 6 Patagónica (Chubut, Neuquén, Río Negro, Santa Cruz y Tierra del Fuego)

Tabla 4.68.b (Transcripta). Prueba de Concepto - Diccionario de Fuente de Datos

Campos Relacionados con el Problema de Negocio			
Responsable:	Esposito E.	**Fecha:**	29/04/2016
ID#:	D.ED.AnD.CRPN	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		

Nombre	Generar	Descripción	Referencia
POB_URB		**Agrupamiento de poblaciones urbanas** 1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes	(fuic.3) Base ENPreCoSP 2011
RANGOING		**Rango del Ingreso total mensual del hogar en pesos** 0 Sin ingresos 1 1 a 600 2 601 a 800 3 801 a 1.000 4 1.001 a 1.500 5 1.501 a 2.000 6 2.001 a 2.500 7 2.501 a 3.000 8 3.001 a 3.500 9 3.501 a 4.000 10 4.001 a 4.500 11 4.501 a 5.500 12 5.001 a 6.000 13 6.001 a 7.000 14 7.001 a 8.000 15 8.001 a 10.000 16 10.001 a 15.000 17 15.001 y más 99 Ns/nc	(fuic.3) Base ENPreCoSP 2011
NBI_TOTAL		**INDICADORES DE NECESIDADES BÁSICAS INSATISFECHAS DE HOGAR: NBI Total** 0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
BHCH04		**Sexo** 1 Varón 2 Mujer	(fuic.3) Base ENPreCoSP 2011
GRUPEDAD		**Grupo de edad** 2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años 9 Ns/nc	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
NIVINSTR		**Nivel de instrucción** 1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial	(fuic.3) Base ENPreCoSP 2011
BIAC01		**¿Conoce personas cercanas a usted que en la actualidad consuman alguna sustancia como marihuana, cocaína, éxtasis, etc.?** 1 Sí 2 No 9 Ns/nc	(fuic.3) Base ENPreCoSP 2011

Tabla 4.69.a (Transcripta). Prueba de Concepto - Campos Relacionados con el Problema de Negocio

PV_TA		**Prevalencia de vida de consumo de tabaco** 1 Sí 2 No	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
PV_BA		**Prevalencia de vida de consumo de bebidas alcohólicas** 1 Sí 2 No	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
consumo_ilegal_bool	x	**Prevalencia de vida de consumo de alguna de las sustancias psicoactivas ilegales: si alguna de las variables (PV_ES, PV_MA, PV_CO, PV_PB, PV_EX, PV_IN, PV_ANX, PV_OA, PV_CK, PV_AL) es igual a 1 (SI) entonces 1 sino 0.** 1 Sí 0 No	entrevista 4

Tabla 4.69.b (Transcripta). Prueba de Concepto - Campos Relacionados con el Problema de Negocio

Reporte de Datos Explorados			
Responsable:	Esposito E.	**Fecha:**	06/05/2016
ID#:	D.ED.ExD.ReDE	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		
ATRIBUTOS CUALITATIVOS			
Nombre	**Valores**	**Distribución**	
POB_URB	1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes	1190 (3,47%) 3749 (10,92%) 14282 (41,59%) 15122 (44,03%)	
RANGOING	0 Sin ingresos 1 1 a 600 2 601 a 800 3 801 a 1.000 4 1.001 a 1.500 5 1.501 a 2.000 6 2.001 a 2.500 7 2.501 a 3.000 8 3.001 a 3.500 9 3.501 a 4.000 10 4.001 a 4.500 11 4.501 a 5.500 12 5.001 a 6.000 13 6.001 a 7.000 14 7.001 a 8.000 15 8.001 a 10.000 16 10.001 a 15.000 17 15.001 y más 99 Ns/nc	138 (0,4%) 971 (2,83%) 1044 (3,04%) 1621 (4,72%) 3136 (9,13%) 3986 (11,61%) 2689 (7,83%) 3757 (10,94%) 1716 (5%) 2575 (7,5%) 985 (2,87%) 2120 (6,17%) 1748 (5,09%) 1246 (3,63%) 1071 (3,12%) 1348 (3,93%) 934 (2,72%) 464 (1,35%) 2794 (8,14%)	
NBI_TOTAL	0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI	29813 (86,81%) 3670 (10,69%) 743 (2,16%) 113 (0,33%) 4 (0,01%)	

Tabla 4.71.a (Transcripta). Prueba de Concepto - Reporte de Datos Explorados

BHCH04	1 Varón 2 Mujer	15787 (45,97%) 18556 (54,03%)
NIVINSTR	1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial	372 (1,08%) 2484 (7,23%) 6309 (18,37%) 7092 (20,65%) 8103 (23,59%) 4486 (13,06%) 5454 (15,88%) 43 (0,13%)
BIAC01	1 Sí 2 No 9 Ns/nc	7970 (23,21%) 26276 (76,51%) 97 (0,28%)
PV_TA	1 Sí 0 No	17636 (51,35%) 16707 (48,65%)
PV_BA	1 Sí 0 No	25709 (74,86%) 8634 (25,14%)
consumo_ilegal_bool	1 Sí 0 No	30850 (89,83%) 3493 (10,17%)
GRUPEDAD	2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años 9 Ns/nc	6052 (19,26%) 8130 (25,87%) 9420 (29,98%) 7824 (24,90%) 0 (0%)
Comentarios:		

Tabla 4.71.b (Transcripta). Prueba de Concepto - Reporte de Datos Explorados

Reporte de Evaluación de Herramientas					
Responsable:	Rodriguez H.			**Fecha:**	07/04/2016
ID#:	G.In.EvS.REHe			**Versión:**	1.0
Criterios:					
Evaluación: 1 = Malo, 2 = débil, 3 = Bueno, 4 = Excelente					1 = No, 4 = SI
Herramientas			**Tanagra V.1.4.50**	**Weka V.3.7.11**	**Orange V.2.7.8**
1. Funcional - Características Técnicas					
Soporte de Metodología / Ciclo de vida	Soporte del proceso	3	2	2	2
Compatibilidad con fuentes de datos	Base de datos	8	--	--	--
	Otras fuentes (word, excel, etc.)	8	3	2	3
Integración	Soporte de distintas técnicas asociadas al proceso de explotación de Información	5	4	4	4
Multilenguaje	Soporta distintas idiomas	2	1	1	1
Técnicas	Variedad de técnicas que provee	18	4	4	4
Reporte y visualización	Permite generar reportes y visualizaciones	12	2	2	2
Multiplataforma	Soporta múltiples plataformas	5	1	4	4
Instalación remota	La administración y mantenimiento son remotos	5	--	--	--
Usuarios Múltiples	Posee perfiles de usuarios	2	1	1	1
Seguridad	Provee seguridad de la información configurada por perfiles	2	1	1	1

Tabla 4.13.a (Transcripta). Prueba de Concepto - Reporte de Evaluación de Herramientas

Backup	Metodología de backup	2	1	1	1
Amigable	Interfaz de usuario	10	4	2	4
Configuraciones	Permite la configuración del perfil	8			
Documentación	Servicio de soporte y ayuda	5	4	1	3
Conexión	Soporta conexión por: Internet, FTP, ERPs.	2	1	1	1
Soporte de sistemas de mensaje	Soporta compartir información (por mail u otro medio)	3	1	1	1
Total			224	196	234
	Peso del Grupo	40%	89,6	78,4	93,6
2. Características del Proveedor					
Características del proveedor	Historia	30	3	3	1
Crecimiento	Perspectiva a futuro	10	2	3	2
Ubicación Geográfica	Oficinas	30	--	--	--
Implementación	Otras implementaciones de la misma herramienta	5	--	--	--
	Contacto con otros clientes	5	--	--	--
Confidencialidad	Confidencialidad de la información	20	--	--	--
Total			110	120	50
	Peso del Grupo	25%	27,5	30	12,5
3. Características del Servicio					
Garantía del producto	Duración y Alcance	30	--	--	--
Mejora	Brinda soporte a versiones previas	20	1	1	1
Licencia	Costo, alcances y soporte postventa	30	--	--	--
Soporte	Tiempo de respuesta y disponibilidad	20	--	--	--
Total			20	20	20
	Peso del Grupo	20%	4	4	4
4. Características Económicas					
Costo del software	Costo de la herramienta	30	--	--	--
Costo del Hardware	Necesidad de mejorar o comprar nuevo hardware compatible con la herramienta	20	--	--	--
Otros costos software	Costos adicionales al producto (backup, web servers, bases de datos, etc.)	20	--	--	--
Licencias	Política de licencia	10	--	--	--
Financiamiento	Existencia	10	--	--	--
Mejoras	Costo promedio de la mejora del producto	10	--	--	--
Total			0	0	0
	Peso del Grupo	-15%	0	0	0
Final					
1. Funcional - Características Técnicas		40%	89,6	78,4	93,6
2. Características del Proveedor		25%	27,5	30	12,5
3. Características del Servicio		20%	4	4	4
4. Características Económicas		-15%	0	0	0
TOTAL			**121,1**	**112,4**	**110,1**

Tabla 4.13.b (Transcripta). Prueba de Concepto - Reporte de Evaluación de Herramientas

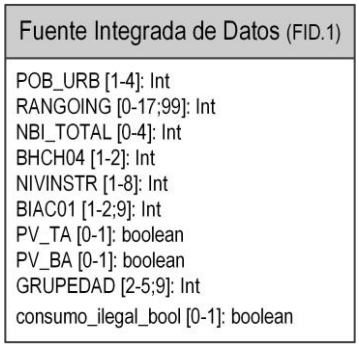


Figura 4.12 (Transcripta). Prueba de Concepto – Fuente Integrada de Datos

Reporte de la Calidad de los Datos (D.ED.EvD.ReCD): A partir del análisis realizado en la fuente integrada de datos y de la descripción de la misma, se identificaron 3 campos con registros anómalos: RANGOING, NIVINSTR y BIAC01. Se detecta, a partir del conocimiento del negocio, que la primer y tercer variable poseen 2794 y 97 registros respectivamente, los cuales no informan dicho campo (clasificados como *nulos*) y en la segunda variable, se detectan 43 registros con valor "8" que no solo se identifica como una minoría no representativa para la población de estudio, sino que además el valor clave no correspondía con la ordinalidad de la variable (clasificándolo como *outliers*). Los motivos previamente expuestos para cada variable fueron indicados en la columna "descripción". La tabla 4.73 ilustra el resultado obtenido del análisis de la calidad de los datos.

Reporte de la Calidad de los Datos			
Responsable:	Esposito E.	**Fecha:**	06/05/2016
ID#:	D.ED.EvD.ReCD	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		
Nombre	**Registros**	**Tipo**	**Descripción**
RANGOING	2794	nulos	Valor 99. Ingresos no informados
NIVINSTR	43	Outlier	Valor 8: Minoría no representativa para el problema de negocio
BIAC01	97	nulos	Valor 9. no informado

Tabla 4.73. Prueba de Concepto - Reporte de la Calidad de los Datos

4.4.3. Fase: Modelado (D.Mo)

En la fase de modelado se realizan las tareas de conversión del escenario del proyecto, desde la perspectiva del negocio, a un formalismo que permita estandarizar el conocimiento de interés para el desarrollo del proceso e identificar las técnicas que pueden ser utilizadas para dar solución a la problemática definida por el cliente. El objetivo principal de la fase es identificar y definir los modelos a utilizar para la extracción de conocimiento.

Esta fase es una aportación introducida a la estructura tradicional, con respecto a los procesos y metodologías existentes (descriptas en la sección 2.2, pág. 12, y analizadas en la sección 3.2.3, pág. 61). Su objetivo es comprender y definir la estrategia que se utilizará para extraer los patrones de conocimientos ocultos en los datos y la evaluación de los mismos en una etapa temprana del proceso, lo que permite identificar las características de los datos, con el propósito de reducir las iteraciones y el tiempo dedicado al tratamiento de los datos (señalada como una de las fases que mayor esfuerzo requiere [Kurgan y Musilek, 2006; Rodriguez et al., 2010]). Mediante la implementación de la fase actual, no solo se logra la comprensión de las necesidades de los datos previamente mencionada, sino que además es posible determinar la necesidad de incorporar nuevas técnicas o herramientas para alcanzar los objetivos establecidos, así como identificar riesgos en el cumplimiento de los mismos. Finalmente, la identificación y planificación del modelo y su evaluación nos facilita la estructuración requerida para llevar a cabo las tareas de preparación de los datos y de explotación de información.

En este contexto, la fase Modelado está conformada por 2 actividades: Modelado del Problema (sección 4.4.3.1), en la cual se traducen las necesidades del cliente desde la perspectiva del dominio del negocio a la perspectiva de explotación de información y se identifican a partir de ellas los posibles modelos de explotación de información a utilizar, y Configuración del Modelo (sección 4.4.3.2), donde se establecen las características del modelo y la forma en la cual serán evaluados. En la figura 4.13, se ilustran las dependencias (elementos de entrada y salida) de las actividades que conforman la fase actual (las imágenes de cada formalismo son representaciones miniatura de los mismos).

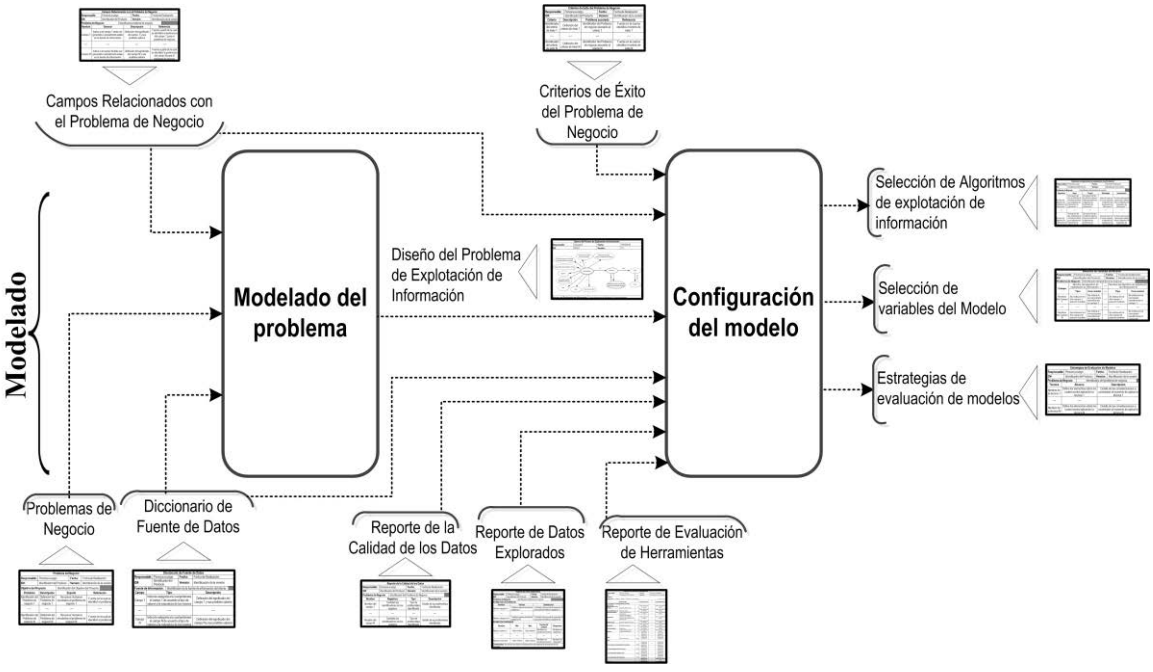


Figura 4.13. Fase: Modelado

4.4.3.1. Actividad: Modelado del Problema (D.Mo.MoP)

En esta actividad se realizan esfuerzos por traducir los requerimientos del cliente desde la perspectiva del dominio del negocio a la de explotación de información. Para ello, se utilizan formalismos de representación del conocimiento que permitan identificar cuáles son los tipos de técnicas de explotación de información (o procesos de explotación de información) adecuados para extraer el conocimiento requerido por la problemática planteada.

Información de Entrada

- Problema del Negocio (D.EN.CPN.PrNe)
- Diccionario de Fuente de Datos (D.ED.AnD.DiFD)
- Campos Relacionados con el Problema de Negocio (D.ED.AnD.CRPN)

Información de Salida

- Diseño del Proceso de Explotación de Información (D.Mo.MoP.DPEI)

4.4.3.1.1. Formalismos Identificados

Para la representación de las necesidades del cliente desde la perspectiva de explotación de información, se propone el formalismo **Diseño del Proceso de Explotación de Información**, el cual se presenta a continuación.

Diseño del Proceso de Explotación de Información (D.Mo.MoP.DPEI): Este formalismo utiliza el gráfico de representación del conocimiento para proyectos de explotación de información (Red Semántica del Problema de Explotación de Información) propuesto en [Martins, et. al, 2014], a partir del cual se deriva el proceso de explotación de información (sección 2.4.1.3, pág. 35) a utilizar. La tabla 4.74 ilustra la estructura del formalismo previamente descripto.

Diseño del Proceso de Explotación de Información			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Problema de Negocio:	Identificador del Problema de Negocio		
<Red Semántica del Problema de Explotación de Información>			
Proceso de Explotación de Información:	Nombre del proceso de explotación de información derivado		

Tabla 4.74. Formalismo: Diseño del Proceso de Explotación de Información

4.4.3.1.2. Técnicas Identificadas

Para el desarrollo de esta actividad se propone la técnica **Derivación del Proceso de Explotación de Información** (sección 2.4.1.4, pág. 39) propuesta en [Martins, et. al, 2014], que adicionalmente

hace uso como su nombre lo indica de las técnicas o procesos de explotación de información [Britos y García-Martínez, 2009; García-Martínez et al., 2013] (sección 2.4.1.3, pág. 35).

Esta técnica de representación visual hace uso de las herramientas de modelado del conocimiento: redes semánticas y marcos, para diseñar el dominio de negocio y el problema de negocio. Está conformada por 3 etapas: análisis del dominio del negocio, análisis del problema de explotación de información y determinación del proceso de explotación de información. Las primeras dos producen un formalismo visual de representación del dominio del negocio y del problema de explotación de información, mientras que la última brinda un conjunto de reglas que permiten identificar a partir de los formalismos obtenidos en las etapas previas el proceso de explotación de información, y por consiguiente los tipos de algoritmos de explotación de información y la estrategia de implementación a utilizar.

La técnica utiliza como elementos de entrada los formalismos: problema de negocio, campos relacionados con el problema de negocio y diccionario de fuentes de datos, mediante los cuales se representan y definen las problemáticas a abordar. En la siguiente sección se presenta la aplicación de la técnica en el proyecto seleccionado como prueba de concepto.

4.4.3.1.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica Derivación del Proceso de Explotación de Información en la prueba de concepto, el cual utiliza como elementos de entrada el Problema del Negocio (Tabla 4.64), el Diccionario de Fuente de Datos (Tablas 4.68.a y 4.68.b) y Campos Relacionados con el Problema de Negocio (Tablas 4.69.a y 4.69.b).

Los formalismos indicados como elementos de entrada, son transcriptos con el mismo número de referencia, para facilitar al lector en la comprensión de la aplicación de la técnica.

Problema del Negocio				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.PRNE	**Versión:**	1.0	
Objetivo del Proyecto	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos			
Problema	**Descripción**		**Experto**	**Referencia**
prne.1	Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		(rehi.3) Silva H.	Entrevista 3

Tabla 4.64 (Transcripta). Prueba de Concepto - Problema del Negocio

Diccionario de Fuente de Datos				
Responsable:		Esposito E.	**Fecha:**	28/04/2016
ID#:		D.ED.AnD.DiFD	**Versión:**	1.0
Fuente de Información		(fuic.3) Base ENPreCoSP 2011		
Campo	**Tipo**	**Descripción**		
ABU_C	Nominal	**Abuso de cerveza** 1 Sí 2 No		
ABU_V	Nominal	**Abuso de vino** 1 Sí 2 No		
BHCH04	Nominal	**Sexo** 1 Varón 2 Mujer		
BHCH05	Discreta	**¿Cuál es su edad en años cumplidos?** (Edad en años cumplidos)		
BIAC01	Nominal	**¿Conoce personas cercanas a usted que en la actualidad consuman alguna sustancia como marihuana, cocaína, éxtasis, etc.?** 1 Sí 2 No 9 Ns/nc		
BIES04	Nominal	**¿Cuándo fue la primera vez que probó estimulantes sin indicación médica?** 1 Durante los últimos 30 días 2 Hace más de un mes, pero menos de un año 3 Hace más de un año 9 Ns/nc		
CAT_OCUP	Nominal	**Categoría ocupacional** 1 Patrón o empleador 2 Cuenta propia 3 Asalariado (excluye servicio doméstico) 4 Asalariado (sólo servicio doméstico) 5 Trabajador familiar sin pago		
CONDACT	Nominal	**Condición de actividad** 1 Ocupado 2 Desocupado 3 Inactivo		
GRUPEDAD	Ordinal	**Grupo de edad** 2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años 9 Ns/nc		
NBI_TOTAL	Ordinal	**INDICADORES DE NECESIDADES BÁSICAS INSATISFECHAS DE HOGAR: NBI Total** 0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI		
NIVINSTR	Ordinal	**Nivel de instrucción** 1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial		
POB_URB	Ordinal	**Agrupamiento de poblaciones urbanas** 1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes		
PROV	Nominal	**Jurisdicción del país** 02 Ciudad Autónoma de Buenos Aires 06 Buenos Aires 10 Catamarca 14 Córdoba 18 Corrientes 22 Chaco 26 Chubut 30 Entre Ríos 34 Formosa 38 Jujuy 42 La Pampa 46 La Rioja 50 Mendoza 54 Misiones 58 Neuquén 62 Río Negro 66 Salta 70 San Juan 74 San Luis 78 Santa Cruz 82 Santa Fe 86 Santiago del Estero 94 Tierra del Fuego, Antártida Argentina e Islas del Atlántico Sur 90 Tucumán		

Tabla 4.68.a (Transcripta). Prueba de Concepto - Diccionario de Fuente de Datos

PV_AL	Nominal	**Prevalencia de vida de consumo de alucinógenos** 1 Sí 2 No
…	…	…
PV_ANX	Nominal	**Prevalencia de vida de consumo de medicamentos para adelgazar (SIN INDICACIÓN MÉDICA)** 1 Sí 2 No
PV_BA	Nominal	**Prevalencia de vida de consumo de bebidas alcohólicas** 1 Sí 2 No
PV_CK	Nominal	**Prevalencia de vida de consumo de crack** 1 Sí 2 No
PV_CO	Nominal	**Prevalencia de vida de consumo de cocaína** 1 Sí 2 No
PV_ES	Nominal	**Prevalencia de vida de consumo de estimulantes** 1 Sí 2 No
PV_EX	Nominal	**Prevalencia de vida de consumo de éxtasis** 1 Sí 2 No
PV_IN	Nominal	**Prevalencia de vida de consumo de inhalables** 1 Sí 2 No
PV_MA	Nominal	**Prevalencia de vida de consumo de marihuana** 1 Sí 2 No
PV_OA	Nominal	**Prevalencia de vida de consumo de opiáceos y anestésicos** 1 Sí 2 No
PV_PB	Nominal	**Prevalencia de vida de consumo de pasta base** 1 Sí 2 No
PV_TA	Nominal	**Prevalencia de vida de consumo de tabaco** 1 Sí 2 No
PV_TR	Nominal	**Prevalencia de vida de consumo de tranquilizantes** 1 Sí 2 No
RANGOING	Ordinal	**Rango del Ingreso total mensual del hogar en pesos** 0 Sin ingresos 5 1.501 a 2.000 1 1 a 600 6 2.001 a 2.500 2 601 a 800 7 2.501 a 3.000 3 801 a 1.000 8 3.001 a 3.500 4 1.001 a 1.500 9 3.501 a 4.000
REGION	Nominal	**Región estadística** 1 Gran Buenos Aires (Ciudad Autónoma de Bs. As. y 24 Partidos del GBA) 2 Pampeana (Resto de Buenos Aires, Córdoba, La Pampa, Santa Fe y Entre Ríos) 3 Noroeste (Catamarca, Jujuy, La Rioja, Salta, Santiago del Estero y Tucumán) 4 Noreste (Corrientes, Chaco, Formosa y Misiones) 5 Cuyo (Mendoza, San Juan y San Luis) 6 Patagónica (Chubut, Neuquén, Río Negro, Santa Cruz y Tierra del Fuego)

Tabla 4.68.b (Transcripta). Prueba de Concepto - Diccionario de Fuente de Datos

Campos Relacionados con el Problema de Negocio			
Responsable:	Esposito E.	**Fecha:**	29/04/2016
ID#:	D.ED.AnD.CRPN	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		

Nombre	**Generar**	**Descripción**	**Referencia**
POB_URB		**Agrupamiento de poblaciones urbanas** 1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes	(fuic.3) Base ENPreCoSP 2011

Tabla 4.69.a (Transcripta). Prueba de Concepto - Campos Relacionados con el Problema de Negocio

RANGOING		**Rango del Ingreso total mensual del hogar en pesos**	(fuic.3) Base ENPreCoSP 2011
		0 Sin ingresos 1 1 a 600 2 601 a 800 3 801 a 1.000 4 1.001 a 1.500 5 1.501 a 2.000 6 2.001 a 2.500 7 2.501 a 3.000 8 3.001 a 3.500 9 3.501 a 4.000 10 4.001 a 4.500 11 4.501 a 5.500 12 5.001 a 6.000 13 6.001 a 7.000 14 7.001 a 8.000 15 8.001 a 10.000 16 10.001 a 15.000 17 15.001 y más 99 Ns/nc	
NBI_TOTAL		**INDICADORES DE NECESIDADES BÁSICAS INSATISFECHAS DE HOGAR: NBI Total** 0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
BHCH04		**Sexo** 1 Varón 2 Mujer	(fuic.3) Base ENPreCoSP 2011
GRUPEDAD		**Grupo de edad** 2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años 9 Ns/nc	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
NIVINSTR		**Nivel de instrucción** 1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial	(fuic.3) Base ENPreCoSP 2011
BIAC01		**¿Conoce personas cercanas a usted que en la actualidad consuman alguna sustancia como marihuana, cocaína, éxtasis, etc.?** 1 Sí 2 No 9 Ns/nc	(fuic.3) Base ENPreCoSP 2011
PV_TA		**Prevalencia de vida de consumo de tabaco** 1 Sí 2 No	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
PV_BA		**Prevalencia de vida de consumo de bebidas alcohólicas** 1 Sí 2 No	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
consumo_ilegal_bool	x	**Prevalencia de vida de consumo de alguna de las sustancias psicoactivas ilegales: si alguna de las variables (PV_ES, PV_MA, PV_CO, PV_PB, PV_EX, PV_IN, PV_ANX, PV_OA, PV_CK, PV_AL) es igual a 1 (SI) entonces 1 sino 0.** 1 Sí 0 No	entrevista 4

Tabla 4.69.b (Transcripta). Prueba de Concepto - Campos Relacionados con el Problema de Negocio

Diseño del Proceso de Explotación de Información (D.Mo.MoP.DPEI): a partir del problema de negocio previamente identificado, se procede a representar los aspectos relevantes del mismo (conceptos, atributos, relaciones y valores), generando un conjunto de marcos que permiten sistematizar el proceso de representación del conocimiento desde el formato texto al formato gráfico basado en redes semánticas, identificando los distintos elementos de interés para determinar el proceso de explotación de información a utilizar. El proceso de explotación de información identifica el tipo de técnica o algoritmo de explotación de información y la estrategia (orden e interconexión) a utilizar para extraer las piezas de conocimiento que den soporte al problema de negocio identificado. En la prueba de concepto, se identifica a partir del problema de negocio, el concepto "persona", limitando su interés de estudio por la edad (entre 16 y 65 años). La acción de interés (representada con el nodo variable) es la de caracterizar a las persona, en relación a su prevalencia de vida en el consumo de sustancias psicoactivas ilegales. Identificando como características a considerar, aquellas definidas en el formalismo campos relacionados con el problema de negocio.

La tabla 4.75, ilustra la gráfica resultante a partir de la cual se identifica al **proceso de descubrimiento de reglas de comportamiento** el cual indica como estrategia de implementación el uso de algoritmos TDIDT.

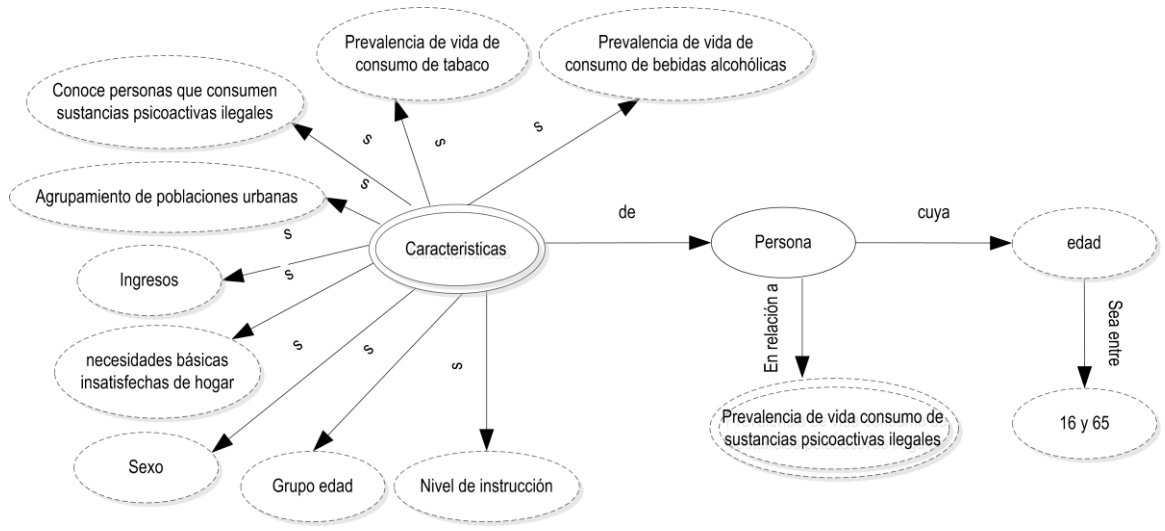
Diseño del Proceso de Explotación de Información			
Responsable:	Esposito E.	**Fecha:**	10/05/2016
ID#:	DPEI.1	**Versión:**	1.0
Problema de Negocio:	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		
			
Proceso de Explotación de Información:	Descubrimiento de Reglas de Comportamiento		

Tabla 4.75. Prueba de Concepto - Diseño del Proceso de Explotación de Información

4.4.3.2. Actividad: Configuración del Modelo (D.Mo.CoM)

En la actividad de configuración del modelo se definen los elementos que conforman la estrategia de implementación y evaluación de los distintos modelos para la extracción de patrones vinculados con el problema de negocio. Esto comprende seleccionar y definir el orden de aplicación de los algoritmos de explotación de información, seleccionar y determinar el formato requerido de las variables, y determinar el proceso a aplicar para la obtención y evaluación de los resultados obtenidos.

Dada la fuerte dependencia entre los datos y los posibles modelos a utilizar, esta actividad permite evaluar en una instancia temprana del proyecto, la viabilidad de las soluciones propuestas con respecto a las necesidades del cliente, los datos y las técnicas disponibles. Riesgos con respecto a la solución del problema pueden ser detectados en esta etapa.

Información de Entrada

- Criterios de Éxito del Problema de Negocio (D.EN.CPN.CEPN)
- Diccionario de Fuente de Datos (D.ED.AnD.DiFD)
- Campos Relacionados con el Problema de Negocio (D.ED.AnD.CRPN)
- Reporte de Datos Explorados (D.ED.ExD.ReDE)
- Reporte de la Calidad de los Datos (D.ED.EvD.ReCD)
- Diseño del Proceso de Explotación de Información (D.Mo.MoP.DPEI)
- Reporte de Evaluación de Herramientas (G.In.EvS.EvHe)

Información de Salida

- Selección de Algoritmos de Explotación de Información (D.Mo.CoM.SAEI)
- Selección de Variables del Modelo (D.Mo.CoM.SeVM)
- Estrategias de Evaluación de Modelos (D.Mo.CoM.EsEM)

4.4.3.2.1. Formalismos Identificados

Para registrar la información de salida esperada para el desarrollo de la actividad, se proponen los formalismos: Selección de Algoritmos de explotación de información, Selección de variables del Modelo y Estrategias de evaluación de modelos, los cuales se presentan a continuación.

Selección de Algoritmos de Explotación de Información (D.Mo.CoM.SAEI): Se formaliza el conjunto de algoritmos de explotación de información que se utilizarán para extraer las piezas de conocimiento requeridas de acuerdo al problema de negocio. En la columna “algoritmo”, se indica el nombre de la técnica a utilizar, definiendo las condiciones que posee el algoritmo con respecto a

los tipos de datos que soporta como elementos de entrada y (si posee) como atributo clase (registrados en las columnas “input”, y “target” respectivamente). En la columna “estrategia” se define la forma en la cual se aplicarán los algoritmos, indicando el orden de vinculación (si existiese), y en el campo “descripción”, se brinda información adicional sobre el mismo. La tabla 4.76 ilustra la estructura del formalismo previamente descripto.

Selección de Algoritmos de explotación de información				
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización	
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión	
Problema de Negocio	Identificador del problema de negocio			
Algoritmo	**Input**	**Target**	**Estrategia**	**Descripción**
Nombre del Algoritmo de Explotación de Información 1	Descripción del tipo de atributos de entrada permitidos por el algoritmo de Explotación de Información 1	Descripción del tipo de atributos clases o target permitidos por el algoritmo de Explotación de Información 1	Indica el orden en el cual se aplicará el algoritmo de Explotación de Información 1	Se indica información adicional de utilidad sobre el Algoritmo de Explotación de Información 1
…	…	…	…	…
Nombre del Algoritmo de Explotación de Información N	Descripción del tipo de atributos de entrada permitidos por el algoritmo de Explotación de Información N	Descripción del tipo de atributos clases o target permitidos por el algoritmo de Explotación de Información N	Indica el orden en el cual se aplicará el algoritmo de Explotación de Información N	Se indica información adicional de utilidad sobre el Algoritmo de Explotación de Información N

Tabla 4.76. Formalismo: Selección de Algoritmos de Explotación de Información

Selección de Variables del Modelo (D.Mo.CoM.SeVM): Se deja registro formal de los campos que se utilizan para el/los modelos (o algoritmos) seleccionados para el problema de negocio, indicando el nombre del atributo a utilizar (y en caso que fuese necesario la fuente de datos a la que pertenece) en la columna “campo”. Para cada algoritmo identificado, se procede a señalar si el atributo se utiliza (como input o target) o no es utilizado para dicha instancia (dejándolo vacío), y si es necesario aplicar alguna transformación, en las columnas “tipo” y “conversión”, respectivamente. En la tabla 4.77 se presenta la estructura del formalismo.

Estrategias de Evaluación de Modelos (D.Mo.CoM.EsEM): Se deja registro formal de las técnicas a utilizar para medir y valorar la mejora de los resultados obtenidos por los modelos. Se indica el nombre de la técnica a utilizar y se describe el marco de aplicación de los mismos (si su alcance es específico para algún algoritmo, o si es global, aplicando en los datos para todos los algoritmos) en las columnas “Técnica” y “Alcance”, respectivamente. Por último, en la columna “descripción” se indican aquellas características a tener en consideración al momento de aplicar la técnica. La tabla 4.78 ilustra la estructura del formalismo previamente descripto.

Selección de variables del Modelo					
Responsable:	Persona a cargo		**Fecha:**	Fecha de Realización	
ID#:	Identificador del Producto		**Versión:**	Identificación de la versión	
Problema de Negocio	Identificador del problema de negocio				
Campo	Nombre del algoritmo de explotación de información 1		…	Nombre del algoritmo de explotación de información N	
	Tipo	**Conversión**	…	**Tipo**	**Conversión**
Nombre del Campo 1	Se indica el rol del campo 1 para el modelo	Se indica si es necesario transformar el campo 1	…	Se indica el rol del campo 1 para el modelo	Se indica si es necesario transformar el campo 1
…	…	…	…	…	…
Nombre del Campo N	Se indica el rol del campo N para el modelo	Se indica si es necesario transformar el campo N	…	Se indica el rol del campo N para el modelo	Se indica si es necesario transformar el campo N

Tabla 4.77. Formalismo: Selección de Variables del Modelo

Estrategias de Evaluación de Modelos			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Problema de Negocio	Identificador del problema de negocio		
Técnica	**Alcance**	**Descripción**	
Nombre de la técnica 1	Define los elementos sobre los cuales tendrá aplicación la técnica 1	Detalle de las consideraciones a contemplar al momento de aplicar la técnica 1	
…	…	…	
Nombre de la técnica N	Define los elementos sobre los cuales tendrá aplicación la técnica N	Detalle de las consideraciones a contemplar al momento de aplicar la técnica N	

Tabla 4.78. Formalismo: Estrategias de Evaluación de Modelos

4.4.3.2.2. Técnicas Identificadas

Para el desarrollo de esta actividad se propone la técnica **"Determinación de la Configuración del Modelo"**, mediante la cual se define la estructura general del modelo cubriendo los aspectos vinculados con los elementos de entrada, los cuales tienen dependencia con las restricciones de los algoritmos de explotación de información y los pasos que se aplicarán para mejorar la calidad de los resultados y evaluar los mismos, en consideración con las necesidades del problema de negocio. La técnica está conformada por 3 etapas:

- En la primera de ellas se determina a partir de la representación del problema de negocio y la identificación del proceso de explotación de información (que identifica la familia o tipo de algoritmos que pueden utilizarse), cual/es de los algoritmos, teniendo en consideración las opciones provistas por las herramientas seleccionadas, van a utilizarse para generar la extracción de patrones de conocimiento. Para alcanzar este objetivo, se analiza las

descripciones de los datos, realizadas en la fase de entendimiento de los datos, su naturaleza (es decir, sin importar si la variable está representada por claves numéricas, qué información representa el significado de cada valor y que tipo de información es, cuantitativa o cualitativa, en el segundo de los casos, si la distancia entre puntos es de escala lineal o no, etc.), identificando de la familia de algoritmos cual/es pueden interpretar de la mejor forma el conocimiento oculto en los datos. Como resultado, se espera una vez identificado los algoritmos, describir las condiciones de entrada y target del mismo. Luego, se define la estrategia de aplicación la cual se obtiene a partir del proceso de explotación de información, es decir, si hay una vinculación entre algoritmos (dependencia de la salida de uno, como atributo entrada del otro) enumerando la misma de forma consecutiva. Por ejemplo, si se deben aplicar dos algoritmos **A** y **B**, siendo el segundo alimentado por el resultado del primero, en el campo "estrategia" se identifica *1* (uno) para **A** y *2* (dos) para **B**. Entendiéndose que todos los algoritmos con mismo número se aplican como opciones en el mismo nivel para identificar cuál es el que obtiene mejores resultados para el problema de negocio en cuestión. Finalmente, en el campo "descripción" se registran aquellos detalles a tener en consideración para la implementación del modelo (por ejemplo: necesidades específicas con los datos, como normalización, utilización de kernels o medidas de distancia específicas, entre otros).

- En el segundo paso, se identifica de acuerdo a los campos relacionados con el problema de negocio (evaluados con el experto de acuerdo a su significado con respecto a la problemática en cuestión), los reportes de los datos (exploración y calidad de los datos) y las necesidades específicas de los algoritmos definidos en el paso previo, cómo serán utilizados los datos (pudiendo algunos de ellos ser descartados), es decir, si serán elementos de entrada o clase (target) y si es necesario realizar algún tipo de conversión, detallando la forma de la misma (por ejemplo: convertir de discreto a continuo por valor, de forma binaria, o mediante una escala, convertir de continuo a discreto, mediante rangos definidos con el experto, por distancia o frecuencia, estandarización o normalización, etc.). Por simplicidad, se aconseja que en caso que distintos algoritmos utilicen la misma estructuración de los datos, se introduzcan de forma conjunta.
- En el tercer paso, se definen las técnicas a aplicar para el proceso de evaluación del modelo (teniendo en consideración los criterios de éxito definidos para los problemas de negocio), las cuales incluyen tanto la separación o filtrado de los registros para optimizar el modelo (por ejemplo: muestra o muestra estratificada), así como para garantizar la precisión y veracidad de las medidas de evaluación de los datos (o función de costo) y la correcta generalización de los mismos (independencia entre los datos de entrenamiento y los de

testeo), separando los registros en set de entrenamiento, testeo y validación o utilizando técnicas de evaluación de resultados, como por ejemplo: validación cruzada (cross-validation), bootstrap, entre otras. Definiéndose el alcance de aplicación de las técnicas (en la columna homónima), es decir, si se aplica a todos los algoritmos, o si es para evaluar un tipo de algoritmo específico, describiendo todos los detalles requeridos para la implementación de la estrategia de evaluación seleccionada (en la columna "descripción"), como por ejemplo: consideraciones respecto al porcentaje de datos para el muestreo, la variable sobre la cual se estratificará la población, el porcentaje de cada set de datos, los parámetros de los algoritmos de evaluación, etc.

En la siguiente sección se presenta la aplicación de la técnica y el registro de los formalismos para la prueba de concepto.

4.4.3.2.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica, a partir del cual se definen los formalismos: selección de algoritmos de explotación de información, selección de variables del modelo y estrategias de evaluación de modelos.

Para realizar dichos formalismos, se utilizan como insumos: Diseño del Proceso de Explotación de Información (Tabla 4.75), para definir los algoritmos a utilizar junto con Diccionario de Fuente de Datos (Tablas 4.68.a y 4.68.b), Campos Relacionados con el Problema de Negocio (Tablas 4.69.a y 4.69.b), Reporte de Datos Explorados (Tabla 4.71), Reporte de la Calidad de los Datos (Tabla 4.73) y el Reporte de Evaluación de Herramientas (Tablas 4.13.a y 4.13.b). Estos últimos, además fueron utilizados para definir las variables y el formato de las mismas. Por último, de los resultados desprendidos en los pasos anteriores, junto con los Criterios de Éxito del Problema de Negocio (Tabla 4.65), se definieron las estrategias de evaluación de los modelos. El Reporte de Evaluación de Herramientas limita el conjunto de instrumentos a utilizar y sus características. Los formalismos indicados como elementos de entrada, son transcriptos con el mismo número de referencia, para facilitar al lector en la comprensión de la aplicación de la técnica.

Criterios de Éxito del Problema de Negocio				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.CEPN	**Versión:**	1.0	
Criterio	**Descripción**	**Problema asociado**		**Referencia**
cepn.1	Identificar aspectos que permitan comprender el comportamiento de grupos masivos de personas de acuerdo a la evaluación de Silva H. (rehi.3)	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		Entrevista 3

Tabla 4.65 (Transcripta). Prueba de Concepto - Criterios de Éxito del Problema de Negocio

Reporte de la Calidad de los Datos			
Responsable:	Esposito E.	**Fecha:**	06/05/2016
ID#:	D.ED.EvD.ReCD	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		

Nombre	Registros	Tipo	Descripción
RANGOING	2794	nulos	Valor 99. Ingresos no informados
NIVINSTR	43	Outlier	Valor 8: Minoría no representativa para el problema de negocio
BIAC01	97	nulos	Valor 9. no informado

Tabla 4.73 (Transcripta). Prueba de Concepto - Reporte de la Calidad de los Datos

Diccionario de Fuente de Datos				
Responsable:		Esposito E.	**Fecha:**	28/04/2016
ID#:		D.ED.AnD.DiFD	**Versión:**	1.0
Fuente de Información		(fuic.3) Base ENPreCoSP 2011		
Campo	**Tipo**	**Descripción**		
ABU_C	Nominal	**Abuso de cerveza** 1 Sí 2 No		
ABU_V	Nominal	**Abuso de vino** 1 Sí 2 No		
BHCH04	Nominal	**Sexo** 1 Varón 2 Mujer		
BHCH05	Discreta	**¿Cuál es su edad en años cumplidos?** (Edad en años cumplidos)		
BIAC01	Nominal	**¿Conoce personas cercanas a usted que en la actualidad consuman alguna sustancia como marihuana, cocaína, éxtasis, etc.?** 1 Sí 2 No 9 Ns/nc		
BIES04	Nominal	**¿Cuándo fue la primera vez que probó estimulantes sin indicación médica?** 1 Durante los últimos 30 días 2 Hace más de un mes, pero menos de un año 3 Hace más de un año 9 Ns/nc		
CAT_OCUP	Nominal	**Categoría ocupacional** 1 Patrón o empleador 2 Cuenta propia 3 Asalariado (excluye servicio doméstico) 4 Asalariado (sólo servicio doméstico) 5 Trabajador familiar sin pago		
CONDACT	Nominal	**Condición de actividad** 1 Ocupado 2 Desocupado 3 Inactivo		
GRUPEDAD	Ordinal	**Grupo de edad** 2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años 9 Ns/nc		
NBI_TOTAL	Ordinal	**INDICADORES DE NECESIDADES BÁSICAS INSATISFECHAS DE HOGAR: NBI Total** 0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI		
NIVINSTR	Ordinal	**Nivel de instrucción** 1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial		

Tabla 4.68.a (Transcripta). Prueba de Concepto - Diccionario de Fuente de Datos

POB_URB	Ordinal	**Agrupamiento de poblaciones urbanas** 1 Más de 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 4 De 5.000 a 100.000 habitantes
PROV	Nominal	**Jurisdicción del país** 02 Ciudad Autónoma de Buenos Aires 50 Mendoza 06 Buenos Aires 54 Misiones 10 Catamarca 58 Neuquén 14 Córdoba 62 Río Negro 18 Corrientes 66 Salta 22 Chaco 70 San Juan 26 Chubut 74 San Luis 30 Entre Ríos 78 Santa Cruz 34 Formosa 82 Santa Fe 38 Jujuy 86 Santiago del Estero 42 La Pampa 94 Tierra del Fuego, Antártida Argentina 46 La Rioja e Islas del Atlántico Sur 90 Tucumán
…	…	…
PV_AL	Nominal	**Prevalencia de vida de consumo de alucinógenos** 1 Sí 2 No
PV_ANX	Nominal	**Prevalencia de vida de consumo de medicamentos para adelgazar (SIN INDICACIÓN MÉDICA)** 1 Sí 2 No
PV_BA	Nominal	**Prevalencia de vida de consumo de bebidas alcohólicas** 1 Sí 2 No
PV_CK	Nominal	**Prevalencia de vida de consumo de crack** 1 Sí 2 No
PV_CO	Nominal	**Prevalencia de vida de consumo de cocaína** 1 Sí 2 No
PV_ES	Nominal	**Prevalencia de vida de consumo de estimulantes** 1 Sí 2 No
PV_EX	Nominal	**Prevalencia de vida de consumo de éxtasis** 1 Sí 2 No
PV_IN	Nominal	**Prevalencia de vida de consumo de inhalables** 1 Sí 2 No
PV_MA	Nominal	**Prevalencia de vida de consumo de marihuana** 1 Sí 2 No
PV_OA	Nominal	**Prevalencia de vida de consumo de opiáceos y anestésicos** 1 Sí 2 No
PV_PB	Nominal	**Prevalencia de vida de consumo de pasta base** 1 Sí 2 No
PV_TA	Nominal	**Prevalencia de vida de consumo de tabaco** 1 Sí 2 No
PV_TR	Nominal	**Prevalencia de vida de consumo de tranquilizantes** 1 Sí 2 No
RANGOING	Ordinal	**Rango del Ingreso total mensual del hogar en pesos** 0 Sin ingresos 5 1.501 a 2.000 1 1 a 600 6 2.001 a 2.500 2 601 a 800 7 2.501 a 3.000 3 801 a 1.000 8 3.001 a 3.500 4 1.001 a 1.500 9 3.501 a 4.000
REGION	Nominal	**Región estadística** 1 Gran Buenos Aires (Ciudad Autónoma de Bs. As. y 24 Partidos del GBA) 2 Pampeana (Resto de Buenos Aires, Córdoba, La Pampa, Santa Fe y Entre Ríos) 3 Noroeste (Catamarca, Jujuy, La Rioja, Salta, Santiago del Estero y Tucumán) 4 Noreste (Corrientes, Chaco, Formosa y Misiones) 5 Cuyo (Mendoza, San Juan y San Luis) 6 Patagónica (Chubut, Neuquén, Río Negro, Santa Cruz y Tierra del Fuego)

Tabla 4.68.b (Transcripta). Prueba de Concepto - Diccionario de Fuente de Datos

Campos Relacionados con el Problema de Negocio			
Responsable:	Esposito E.	**Fecha:**	29/04/2016
ID#:	D.ED.AnD.CRPN	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		

Nombre	Generar	Descripción	Referencia
POB_URB		**Agrupamiento de poblaciones urbanas** 1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes	(fuic.3) Base ENPreCoSP 2011
RANGOING		**Rango del Ingreso total mensual del hogar en pesos** 0 Sin ingresos 1 1 a 600 2 601 a 800 3 801 a 1.000 4 1.001 a 1.500 5 1.501 a 2.000 6 2.001 a 2.500 7 2.501 a 3.000 8 3.001 a 3.500 9 3.501 a 4.000 10 4.001 a 4.500 11 4.501 a 5.500 12 5.001 a 6.000 13 6.001 a 7.000 14 7.001 a 8.000 15 8.001 a 10.000 16 10.001 a 15.000 17 15.001 y más 99 Ns/nc	(fuic.3) Base ENPreCoSP 2011
NBI_TOTAL		**INDICADORES DE NECESIDADES BÁSICAS INSATISFECHAS DE HOGAR: NBI Total** 0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
BHCH04		**Sexo** 1 Varón 2 Mujer	(fuic.3) Base ENPreCoSP 2011
GRUPEDAD		**Grupo de edad** 2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años 9 Ns/nc	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
NIVINSTR		**Nivel de instrucción** 1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial	(fuic.3) Base ENPreCoSP 2011
BIAC01		**¿Conoce personas cercanas a usted que en la actualidad consuman alguna sustancia como marihuana, cocaína, éxtasis, etc.?** 1 Sí 2 No 9 Ns/nc	(fuic.3) Base ENPreCoSP 2011
PV_TA		**Prevalencia de vida de consumo de tabaco** **1 Sí** **2 No**	(fuic.3) Base ENPreCoSP 2011 / entrevista 4

Tabla 4.69.a (Transcripta). Prueba de Concepto - Campos Relacionados con el Problema de Negocio

PV_BA		**Prevalencia de vida de consumo de bebidas alcohólicas** 1 Sí 2 No	(fuic.3) Base ENPreCoSP 2011 / entrevista 4
consumo_ilegal_bool	x	**Prevalencia de vida de consumo de alguna de las sustancias psicoactivas ilegales: si alguna de las variables (PV_ES, PV_MA, PV_CO, PV_PB, PV_EX, PV_IN, PV_ANX, PV_OA, PV_CK, PV_AL) es igual a 1 (SI) entonces 1 sino 0.** 1 Sí 0 No	entrevista 4

Tabla 4.69.b (Transcripta). Prueba de Concepto - Campos Relacionados con el Problema de Negocio

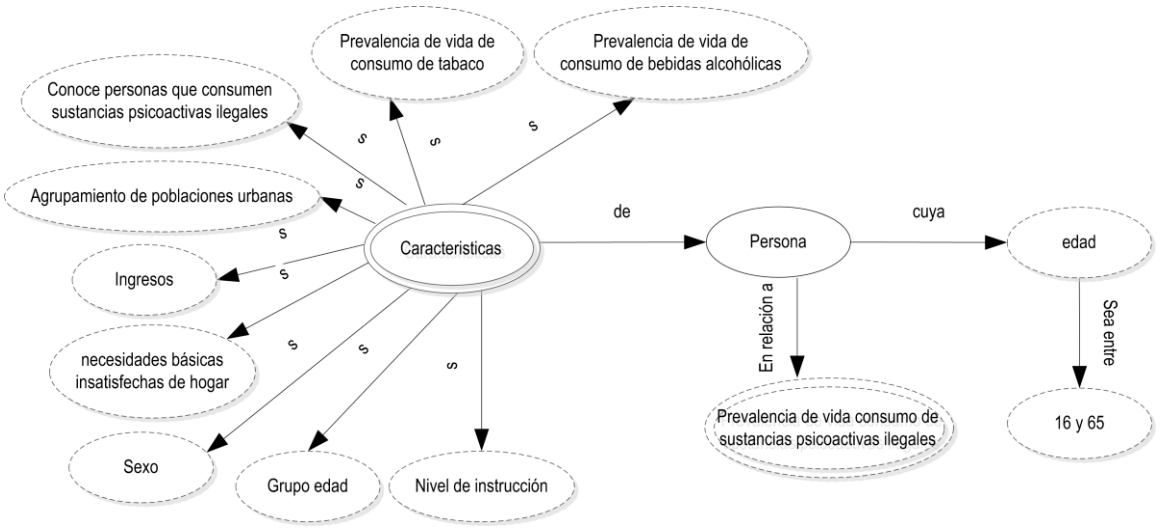
Reporte de Datos Explorados			
Responsable:	Esposito E.	**Fecha:**	06/05/2016
ID#:	D.ED.ExD.ReDE	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		
ATRIBUTOS CUALITATIVOS			
Nombre	**Valores**	**Distribución**	
POB_URB	1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes	1190 (3,47%) 3749 (10,92%) 14282 (41,59%) 15122 (44,03%)	
RANGOING	0 Sin ingresos 1 1 a 600 2 601 a 800 3 801 a 1.000 4 1.001 a 1.500 5 1.501 a 2.000 6 2.001 a 2.500 7 2.501 a 3.000 8 3.001 a 3.500 9 3.501 a 4.000 10 4.001 a 4.500 11 4.501 a 5.500 12 5.001 a 6.000 13 6.001 a 7.000 14 7.001 a 8.000 15 8.001 a 10.000 16 10.001 a 15.000 17 15.001 y más 99 Ns/nc	138 (0,4%) 971 (2,83%) 1044 (3,04%) 1621 (4,72%) 3136 (9,13%) 3986 (11,61%) 2689 (7,83%) 3757 (10,94%) 1716 (5%) 2575 (7,5%) 985 (2,87%) 2120 (6,17%) 1748 (5,09%) 1246 (3,63%) 1071 (3,12%) 1348 (3,93%) 934 (2,72%) 464 (1,35%) 2794 (8,14%)	
NBI_TOTAL	0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI	29813 (86,81%) 3670 (10,69%) 743 (2,16%) 113 (0,33%) 4 (0,01%)	
BHCH04	1 Varón 2 Mujer	15787 (45,97%) 18556 (54,03%)	
NIVINSTR	1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial	372 (1,08%) 2484 (7,23%) 6309 (18,37%) 7092 (20,65%) 8103 (23,59%) 4486 (13,06%) 5454 (15,88%) 43 (0,13%)	
BIAC01	1 Sí 2 No 9 Ns/nc	7970 (23,21%) 26276 (76,51%) 97 (0,28%)	

Tabla 4.71.a (Transcripta). Prueba de Concepto - Reporte de Datos Explorados

PV_TA	1 Sí 0 No	17636 (51,35%) 16707 (48,65%)
PV_BA	1 Sí 0 No	25709 (74,86%) 8634 (25,14%)
consumo_ilegal_bool	1 Sí 0 No	30850 (89,83%) 3493 (10,17%)
GRUPEDAD	2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años 9 Ns/nc	6052 (19,26%) 8130 (25,87%) 9420 (29,98%) 7824 (24,90%) 0 (0%)
Comentarios:		

Tabla 4.71.b (Transcripta). Prueba de Concepto - Reporte de Datos Explorados

Diseño del Proceso de Explotación de Información			
Responsable:	Esposito E.	**Fecha:**	10/05/2016
ID#:	DPEI.1	**Versión:**	1.0
Problema de Negocio:	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		



Proceso de Explotación de Información:	Descubrimiento de Reglas de Comportamiento

Tabla 4.75 (Transcripta). Prueba de Concepto - Diseño del Proceso de Explotación de Información

Reporte de Evaluación de Herramientas					
Responsable:	Rodriguez H.			**Fecha:**	07/04/2016
ID#:	G.In.EvS.REHe			**Versión:**	1.0
Criterios: **Evaluación:** 1 = Malo, 2 = débil, 3 = Bueno, 4 = Excelente					1 = No, 4 = SI
Herramientas			**Tanagra V.1.4.50**	**Weka V.3.7.11**	**Orange V.2.7.8**
1. Funcional - Características Técnicas					
Soporte de Metodología / Ciclo de vida	Soporte del proceso	3	2	2	2
Compatibilidad con fuentes de datos	Base de datos	8	--	--	--
	Otras fuentes (word, excel, etc.)	8	3	2	3

Tabla 4.13.a (Transcripta). Prueba de Concepto - Reporte de Evaluación de Herramientas

Integración	Soporte de distintas técnicas asociadas al proceso de explotación de Información	5	4	4	4
Multilenguaje	Soporta distintas idiomas	2	1	1	1
Técnicas	Variedad de técnicas que provee	18	4	4	4
Reporte y visualización	Permite generar reportes y visualizaciones	12	2	2	2
Multiplataforma	Soporta múltiples plataformas	5	1	4	4
Instalación remota	La administración y mantenimiento son remotos	5	--	--	--
Usuarios Múltiples	Posee perfiles de usuarios	2	1	1	1
Seguridad	Provee seguridad de la información configurada por perfiles	2	1	1	1
Backup	Metodología de backup	2	1	1	1
Amigable	Interfaz de usuario	10	4	2	4
Configuraciones	Permite la configuración del perfil	8			
Documentación	Servicio de soporte y ayuda	5	4	1	3
Conexión	Soporta conexión por: Internet, FTP, ERPs.	2	1	1	1
Soporte de sistemas de mensaje	Soporta compartir información (por mail u otro medio)	3	1	1	1
Total			224	196	234
	Peso del Grupo	40%	89,6	78,4	93,6
2. Características del Proveedor					
Características del proveedor	Historia	30	3	3	1
Crecimiento	Perspectiva a futuro	10	2	3	2
Ubicación Geográfica	Oficinas	30	--	--	--
Implementación	Otras implementaciones de la misma herramienta	5	--	--	--
	Contacto con otros clientes	5	--	--	--
Confidencialidad	Confidencialidad de la información	20	--	--	--
Total			110	120	50
	Peso del Grupo	25%	27,5	30	12,5
3. Características del Servicio					
Garantía del producto	Duración y Alcance	30	--	--	--
Mejora	Brinda soporte a versiones previas	20	1	1	1
Licencia	Costo, alcances y soporte postventa	30	--	--	--
Soporte	Tiempo de respuesta y disponibilidad	20	--	--	--
Total			20	20	20
	Peso del Grupo	20%	4	4	4
4. Características Económicas					
Costo del software	Costo de la herramienta	30	--	--	--
Costo del Hardware	Necesidad de mejorar o comprar nuevo hardware compatible con la herramienta	20	--	--	--

Tabla 4.13.b (Transcripta). Prueba de Concepto - Reporte de Evaluación de Herramientas

Otros costos software	Costos adicionales al producto (backup, web servers, bases de datos, etc.)	20	--	--	--
Licencias	Política de licencia	10	--	--	--
Financiamiento	Existencia	10	--	--	--
Mejoras	Costo promedio de la mejora del producto	10	--	--	--
Total			0	0	0
	Peso del Grupo	-15%	0	0	0
Final					
1. Funcional - Características Técnicas		40%	89,6	78,4	93,6
2. Características del Proveedor		25%	27,5	30	12,5
3. Características del Servicio		20%	4	4	4
4. Características Económicas		-15%	0	0	0
TOTAL			**121,1**	**112,4**	**110,1**

Tabla 4.13.c (Transcripta). Prueba de Concepto - Reporte de Evaluación de Herramientas

Selección de Algoritmos de Explotación de Información (D.Mo.CoM.SAEI): a partir del proceso de explotación de información definido en la actividad previa, se seleccionan los algoritmos de árboles de decisión (pertenecientes a la familia TDIDT, de sus siglas en inglés Top Down Induction Decision Trees). En particular, en base a la herramienta utilizada, se identifican dos algoritmos: C4.5 e ID3 (registrándose en la columna homónima). Se determinan las restricciones de implementación de los algoritmos en la herramienta para la entrada: atributos discretos y/o continuos, y para la variable clase atributos discretos (registrando dicha información en las columnas Input y target respectivamente). Dado que ambos algoritmos serán utilizados como alternativa para evaluar cual obtiene mejores resultados, a ambos se le asigna el valor uno (1) como "estrategia". Por último, en la "descripción" se agrega como comentario que ambos cumplen con la restricción establecida por el proceso de descubrimiento de reglas de comportamiento, de ser algoritmos pertenecientes a la familia TDIDT. La tabla 4.79 ilustra la estructura del formalismo obtenido.

Selección de Algoritmos de explotación de información			
Responsable:	Esposito E.	**Fecha:**	11/05/2016
ID#:	D.Mo.CoM.SAEI	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		

Algoritmo	**Input**	**Target**	**Estrategia**	**Descripción**
C4.5	Discretos/Continuos	Discretos	1	familia TDIDT
ID3	Discretos/Continuos	Discretos	1	familia TDIDT

Tabla 4.79. Prueba de Concepto - Selección de Algoritmos de Explotación de Información

Selección de Variables del Modelo (D.Mo.CoM.SeVM): A partir del análisis realizado de los datos (exploración y calidad), se mantienen todos los campos identificados, utilizando a todos los atributos como "tipo" entrada, con excepción del atributo generado consumo_ilegal_bool, el cual se utiliza como variable target. Adicionalmente, como los valores se encuentran indicados con un valor numérico clave (y conociendo que la herramienta seleccionada categoriza a dichos datos como continuos, debido a que tienen asignados un valor numérico de referencia) se define la necesidad de convertir a discretos, entendiéndose sin aclaración que se utilizará la transformación lineal de 1 a 1, dado que las variables son de naturaleza discreta. Por último, se considera relevante destacar que debido a que ambos algoritmos presentan las mismas restricciones en los tipos de datos que utilizan, para simplificar el formalismo, se incluyen ambos en la misma columna. En la tabla 4.80 se puede observar el resultado obtenido de aplicar el segundo paso de la técnica previamente descripto.

Selección de variables del Modelo			
Responsable:	Esposito E.	**Fecha:**	13/05/2016
ID#:	D.Mo.CoM.SeVM	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		

Campo	**C4.5/ID3**	
	Tipo	**Conversión**
POB_URB	Input	Discretizar
RANGOING	Input	Discretizar
NBI_TOTAL	Input	Discretizar
BHCH04	Input	Discretizar
NIVINSTR	Input	Discretizar
BIAC01	Input	Discretizar
consumo_ilegal_bool	Target	Discretizar
GRUPEDAD	Input	Discretizar
PV_TA	Input	Discretizar
PV_BA	Input	Discretizar

Tabla 4.80. Prueba de Concepto - Selección de Variables del Modelo

Estrategias de Evaluación de Modelos (D.Mo.CoM.EsEM): a partir de las restricciones del proyecto, los criterios de éxito del problema de negocio y la exploración de los datos, se identifica la necesidad de aplicar dos técnicas para la evaluación de los datos. La primera de ellas se identifica a partir de la comprobación del desbalance entre las personas que consumen y no sustancias psicoactivas ilegales ("repr.1 - Se identifica un desbalance entre la cantidad de registros que han consumido distintas Sustancias Psicoactivas y quienes no han consumido"), decidiendo realizar un muestreo estratificado dado que el interés del problema de negocio se centra en comprender la conducta de las personas que consumen, logrando de este modo normalizar la penalización del error para ambas clases (positiva y negativa). La aplicación de la misma será a nivel global (columna

"alcance", dado que aplicará en los datos para todos los algoritmos (C4.5 e ID3). Por último, en la "descripción" se detalla el procedimiento a tener en cuenta al momento de aplicar la técnica: balancear la distribución de la variable target al doble de la población de consumidores de sustancias psicoactivas ilegales.

La segunda técnica, se define para comprobar la generalización del patrón obtenido y a su vez la independencia del conocimiento entre la fuente de entrenamiento y la de testeo, y para evitar reducir más el set de datos a utilizar para entrenar el modelo. Se identifica la "técnica" cross-validation para los algoritmos C4.5 e ID3 (registrados en el campo "alcance") y con la configuración: 10 fold Cross-Validation con 10 repeticiones (registrado en la "descripción"). La configuración elegida, se debe a que estudios han demostrado que presenta el balance adecuado entre precisión del resultado y costo de cómputo, y en forma conjunta con estratificación se obtiene un mejor resultado con respecto al sesgo y varianza de los resultados [Kohavi, R., 1995].

En la tabla 4.81 se puede observar el resultado obtenido de aplicar el tercer paso de la técnica previamente descripto.

Estrategias de evaluación de modelos			
Responsable:	Esposito E.	**Fecha:**	13/05/2016
ID#:	D.Mo.CoM.EsEM	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		
Técnica	**Alcance**	**Descripción**	
Muestreo Estratificado	global	balancear la distribución de la variable target al doble de la población de consumidores de sustancias psicoactivas ilegales	
Cross-Validation	C4.5/ID3	10 fold Cross-Validation con 10 repeticiones	

Tabla 4.81. Prueba de Concepto - Estrategias de Evaluación de Modelos

4.4.4. Fase: Preparación de los Datos (D.PD)

En esta fase se realizan transformaciones a los datos preparándolos para su correcta aplicación en los algoritmos de explotación de información. Esto incluye el formateo de los datos de acuerdo a las restricciones de los modelos a utilizar, la limpieza de los mismos (de acuerdo a las problemas de calidad identificados), la conformación de las fuentes de datos a utilizar para las distintas instancias de aplicación y evaluación del modelo y la validación de la composición de la misma por parte de los expertos del negocio. Esta etapa es crítica con respecto a la calidad y el éxito de los resultados obtenidos, contribuyendo fuertemente en el éxito del proyecto [Ye, 2003].

En este contexto, la fase Preparación de los Datos está conformada por 2 actividades: Construcción de la Fuente Temporaria de Datos (sección 4.4.4.1), donde se preparan y describen las distintas fuentes de datos a utilizar para la extracción del conocimiento y la selección del modelo, y Adecuación de la Fuente Temporaria de Datos (sección 4.4.4.2), en la cual se realizan las tareas de limpieza y formateo de los datos. En la figura 4.14, se presentan las actividades que conforman la fase, junto con sus elementos de entrada y salida (las imágenes de cada formalismo son representaciones miniatura de los mismos).

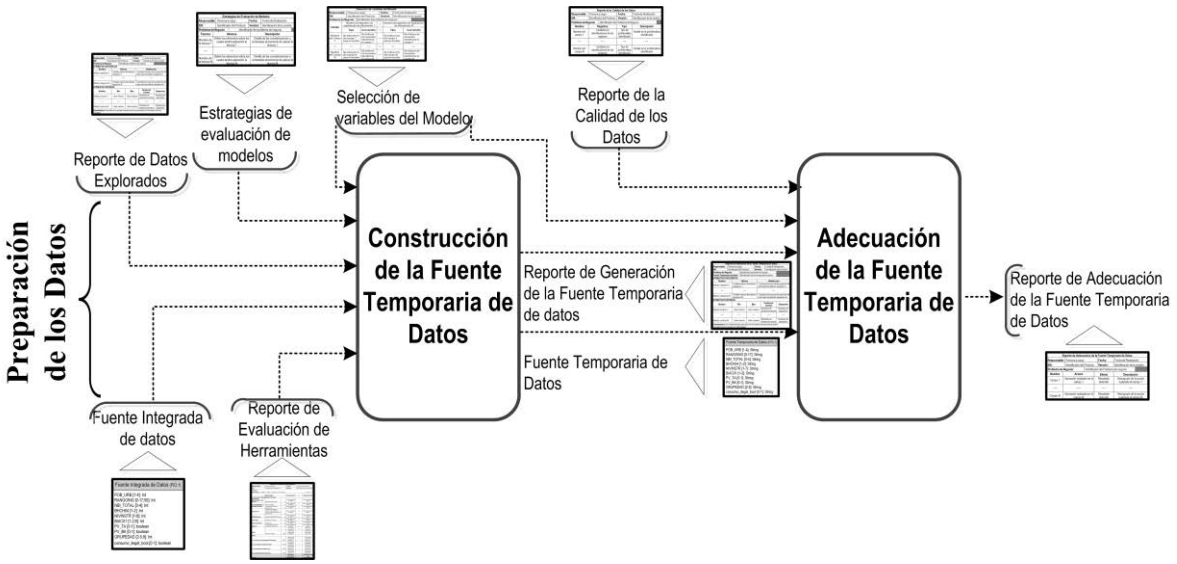


Figura 4.14. Fase: Preparación de los Datos

4.4.4.1. Actividad: Construcción de la Fuente Temporaria de Datos (D.PD.CFT)

En esta actividad se realizan las tareas finales para la generación de las fuentes de datos requeridas para las distintas etapas de implementación del modelo (entrenamiento, validación y testeo). Las fuentes generadas se definen como fuente temporaria de datos, debido a que dicha fuente de almacenamiento es distinta a aquella utilizada en producción y la misma solo será de utilidad para la formación del modelo, la extracción del conocimiento y la evaluación del mismo.

Información de Entrada

- Reporte de Datos Explorados (D.ED.ExD.ReDE)
- Estrategias de Evaluación de Modelos (D.Mo.CoM.EsEM)
- Fuente Integrada de datos (D.ED.ExD.FuID)
- Reporte de Evaluación de Herramientas (G.In.EvS.EvHe)
- Selección de variables del Modelo (D.Mo.CoM.SeVM)

Información de Salida

- Fuente Temporaria de Datos (D.PD.CFT.FuTD)
- Reporte de Generación de la Fuente Temporaria de datos (D.PD.CFT.RGFT)

4.4.4.1.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se propone el Reporte de Generación de la Fuente Temporaria de datos, el cual se presenta a continuación.

Reporte de Generación de la Fuente Temporaria de datos (D.PD.CFT.RGFT): Se formaliza la descripción de las fuentes temporaria de datos desarrollada para el problema de negocio en cuestión, brindando una descripción estadística de las variables que la conforman, distinguidas por el tipo de dato (cualitativo o cuantitativo). Para las variables cualitativas se indica el nombre y la distribución de los distintos valores posibles en la fuente de datos. Para las variables cuantitativas se indica los valores mínimos y máximos, y distintas medidas de tendencia central y dispersión. Finalmente, en los comentarios se podrán agregar indicaciones relevantes acerca de la fuente temporaria de datos, así como visualizaciones. La tabla 4.82 ilustra la estructura del formalismo previamente descripto.

Reporte de Generación de la Fuente Temporaria de datos				
Responsable:	Persona a cargo		**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto		**Versión:**	Identificación de la versión
Problema de Negocio	Identificador problema de negocio			
Fuente Temporaria de datos	Identificador fuente temporaria de datos			
ATRIBUTOS CUALITATIVOS				
Nombre	**Valores**		**Distribución**	
Atributo cualitativo 1	Posibles valores del atributo cualitativo 1		Cantidad y frecuencia de aparición de cada valor del atributo cualitativo 1	
…	…		…	
Atributo cualitativo N	Posibles valores del atributo cualitativo N		Cantidad y frecuencia de aparición de cada valor del atributo cualitativo N	
ATRIBUTOS CUANTITATIVO				
Nombre	**Min**	**Max**	**Tendencia Central**	**Dispersión**
Atributo cuantitativo 1	Valor mínimo	Valor máximo	Medida/s de tendencia central	Medida/s de dispersión
…	…	…	…	…
Atributo cuantitativo N	Valor mínimo	Valor máximo	Medida/s de tendencia central	Medida/s de dispersión
Comentarios: Se complementa la descripción de la fuente temporaria de datos (visualizaciones pueden ser adjuntadas).				

Tabla 4.82. Formalismo: Reporte de Generación de la Fuente Temporaria de datos

4.4.4.1.2. Técnica Identificada

Para el desarrollo de esta actividad se propone la técnica **"Generación de la Fuente Temporaria de Datos"**, mediante la cual se producen las fuentes de datos a utilizar para la implementación de los modelos, realizando un análisis descriptivo similar al realizado por la técnica Exploración de los Datos (sección 4.4.2.2), proveyendo una descripción detallada de los campos disponibles. Esta

descripción permite evaluar la correcta implementación de las tareas de preparación de las fuentes, así como dejar registro de los pasos realizados posibilitando su replicación.

Los elementos de entrada que se utilizan son: el reporte de datos explorados, la fuente integrada de datos y la selección de variables del modelo, que brindan información respecto al total de información disponible, el formato requerido, y las técnicas de separación o muestreo de los registros a aplicar de acuerdo a lo definido en el formalismo estrategias de evaluación de modelos. Finalmente, las herramientas seleccionadas nos permiten comprender las suposiciones que las mismas hacen respecto de los datos (por ejemplo el orden de ubicación de la columna clase), determinando la forma de estructurar la fuente temporaria de datos. Los objetivos de la técnica son generar las fuentes temporarias de datos y dejar registro de los parámetros a partir de los cuales se realizaron las tareas de muestreo (si estas fuesen realizadas).

Finalmente, a partir de las fuentes generadas se procede a describir las variables que integran el modelo, haciendo uso de distintas medidas de estadística descriptiva y de visualizaciones para comprender con el mayor detalle posible la información estudiada. Para ello, se debe tener en consideración la naturaleza de la información que cada variable y sus valores representa, es decir, si la variable es cualitativa o cuantitativa, a partir de lo cual se implementarán distintas herramientas de análisis. Para aquellas variables cualitativas, se registran los valores presentes en la fuente de información y la frecuencia de apariciones de los mismos, indicando la cantidad y el porcentaje de representación. Para las variables cuantitativas, se registran los valores extremos presentes en la fuente de información (mínimo y máximo), y medidas estadísticas de tendencia central (media, mediana, entre otras) y de dispersión (cuartiles, desvió estándar, entre otras).

De forma complementaria, en caso que se considere de utilidad respecto a la información que aporten, se adjuntan visualizaciones de una variable (histograma, gráfica de caja, violín, barras y poligonal), así como de más de una variable (gráfica o matriz de dispersión, mapa de calor o correlación y coordenadas paralelas), indicando su uso en la fila de “comentarios”. En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo para la prueba de concepto.

4.4.4.1.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica Generación de la Fuente Temporaria de Datos, la cual utiliza como insumos los formalismos: Reporte de Datos Explorados (Tabla 4.71), Fuente Integrada de datos (Figura 4.12), Estrategias de evaluación de modelos (Tabla 4.81), Reporte de Evaluación de Herramientas (Tabla 4.13) y Selección de variables del Modelo

(Tabla 4.80), los cuales son transcriptos con el mismo número de referencia, para facilitar al lector en la comprensión de la aplicación de la técnica.

Reporte de Datos Explorados			
Responsable:	Esposito E.	**Fecha:**	06/05/2016
ID#:	D.ED.ExD.ReDE	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		
ATRIBUTOS CUALITATIVOS			
Nombre	**Valores**	**Distribución**	
POB_URB	1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes	1190 (3,47%) 3749 (10,92%) 14282 (41,59%) 15122 (44,03%)	
RANGOING	0 Sin ingresos 1 1 a 600 2 601 a 800 3 801 a 1.000 4 1.001 a 1.500 5 1.501 a 2.000 6 2.001 a 2.500 7 2.501 a 3.000 8 3.001 a 3.500 9 3.501 a 4.000 10 4.001 a 4.500 11 4.501 a 5.500 12 5.001 a 6.000 13 6.001 a 7.000 14 7.001 a 8.000 15 8.001 a 10.000 16 10.001 a 15.000 17 15.001 y más 99 Ns/nc	138 (0,4%) 971 (2,83%) 1044 (3,04%) 1621 (4,72%) 3136 (9,13%) 3986 (11,61%) 2689 (7,83%) 3757 (10,94%) 1716 (5%) 2575 (7,5%) 985 (2,87%) 2120 (6,17%) 1748 (5,09%) 1246 (3,63%) 1071 (3,12%) 1348 (3,93%) 934 (2,72%) 464 (1,35%) 2794 (8,14%)	
NBI_TOTAL	0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI	29813 (86,81%) 3670 (10,69%) 743 (2,16%) 113 (0,33%) 4 (0,01%)	
BHCH04	1 Varón 2 Mujer	15787 (45,97%) 18556 (54,03%)	
NIVINSTR	1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial	372 (1,08%) 2484 (7,23%) 6309 (18,37%) 7092 (20,65%) 8103 (23,59%) 4486 (13,06%) 5454 (15,88%) 43 (0,13%)	
BIAC01	1 Sí 2 No 9 Ns/nc	7970 (23,21%) 26276 (76,51%) 97 (0,28%)	
PV_TA	1 Sí 0 No	17636 (51,35%) 16707 (48,65%)	
PV_BA	1 Sí 0 No	25709 (74,86%) 8634 (25,14%)	
consumo_ilegal_bool	1 Sí 0 No	30850 (89,83%) 3493 (10,17%)	
GRUPEDAD	2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años 9 Ns/nc	6052 (19,26%) 8130 (25,87%) 9420 (29,98%) 7824 (24,90%) 0 (0%)	
Comentarios:			

Tabla 4.71 (Transcripta). Prueba de Concepto - Reporte de Datos Explorados

Figura 4.12 (Transcripta). Prueba de Concepto – Fuente Integrada de Datos (Diagrama Entidad-Relación)

Estrategias de evaluación de modelos			
Responsable:	Esposito E.	**Fecha:**	13/05/2016
ID#:	D.Mo.CoM.EsEM	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		
Técnica	**Alcance**	**Descripción**	
Muestreo Estratificado	global	balancear la distribución de la variable target al doble de la población de consumidores de sustancias psicoactivas ilegales	
Cross-Validation	C4.5/ID3	10 fold Cross-Validation con 10 repeticiones	

Tabla 4.81 (Transcripta). Prueba de Concepto - Estrategias de evaluación de modelos

Reporte de Evaluación de Herramientas					
Responsable:	Rodriguez H.			**Fecha:**	07/04/2016
ID#:	G.In.EvS.REHe			**Versión:**	1.0
Criterios: **Evaluación:** 1 = Malo, 2 = débil, 3 = Bueno, 4 = Excelente					1 = No, 4 = SI
Herramientas			**Tanagra V.1.4.50**	**Weka V.3.7.11**	**Orange V.2.7.8**
1. Funcional - Características Técnicas					
Soporte de Metodología / Ciclo de vida	Soporte del proceso	3	2	2	2
Compatibilidad con fuentes de datos	Base de datos	8	--	--	--
	Otras fuentes (word, excel, etc.)	8	3	2	3
Integración	Soporte de distintas técnicas asociadas al proceso de explotación de Información	5	4	4	4
Multilenguaje	Soporta distintas idiomas	2	1	1	1
Técnicas	Variedad de técnicas que provee	18	4	4	4
Reporte y visualización	Permite generar reportes y visualizaciones	12	2	2	2
Multiplataforma	Soporta múltiples plataformas	5	1	4	4
Instalación remota	La administración y mantenimiento son remotos	5	--	--	--
Usuarios Múltiples	Posee perfiles de usuarios	2	1	1	1
Seguridad	Provee seguridad de la información configurada por perfiles	2	1	1	1

Tabla 4.13.a (Transcripta). Prueba de Concepto - Reporte de Evaluación de Herramientas

Backup	Metodología de backup	2	1	1	1
Amigable	Interfaz de usuario	10	4	2	4
Configuraciones	Permite la configuración del perfil	8			
Documentación	Servicio de soporte y ayuda	5	4	1	3
Conexión	Soporta conexión por: Internet, FTP, ERPs.	2	1	1	1
Soporte de sistemas de mensaje	Soporta compartir información (por mail u otro medio)	3	1	1	1
Total			224	196	234
	Peso del Grupo	40%	89,6	78,4	93,6
2. Características del Proveedor					
Características del proveedor	Historia	30	3	3	1
Crecimiento	Perspectiva a futuro	10	2	3	2
Ubicación Geográfica	Oficinas	30	--	--	--
Implementación	Otras implementaciones de la misma herramienta	5	--	--	--
	Contacto con otros clientes	5	--	--	--
Confidencialidad	Confidencialidad de la información	20	--	--	--
Total			110	120	50
	Peso del Grupo	25%	27,5	30	12,5
3. Características del Servicio					
Garantía del producto	Duración y Alcance	30	--	--	--
Mejora	Brinda soporte a versiones previas	20	1	1	1
Licencia	Costo, alcances y soporte postventa	30	--	--	--
Soporte	Tiempo de respuesta y disponibilidad	20	--	--	--
Total			20	20	20
	Peso del Grupo	20%	4	4	4
4. Características Económicas					
Costo del software	Costo de la herramienta	30	--	--	--
Costo del Hardware	Necesidad de mejorar o comprar nuevo hardware compatible con la herramienta	20	--	--	--
Otros costos software	Costos adicionales al producto (backup, web servers, bases de datos, etc.)	20	--	--	--
Licencias	Política de licencia	10	--	--	--
Financiamiento	Existencia	10	--	--	--
Mejoras	Costo promedio de la mejora del producto	10	--	--	--
Total			0	0	0
	Peso del Grupo	-15%	0	0	0
Final					
1. Funcional - Características Técnicas		40%	89,6	78,4	93,6
2. Características del Proveedor		25%	27,5	30	12,5
3. Características del Servicio		20%	4	4	4
4. Características Económicas		-15%	0	0	0
TOTAL			**121,1**	**112,4**	**110,1**

Tabla 4.13.b (Transcripta). Prueba de Concepto - Reporte de Evaluación de Herramientas

Selección de variables del Modelo			
Responsable:	Esposito E.	**Fecha:**	13/05/2016
ID#:	D.Mo.CoM.SeVM	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		

Campo	C4.5/ID3	
	Tipo	**Conversión**
POB_URB	Input	Discretizar
RANGOING	Input	Discretizar
NBI_TOTAL	Input	Discretizar
BHCH04	Input	Discretizar
NIVINSTR	Input	Discretizar
BIAC01	Input	Discretizar
consumo_ilegal_bool	Target	Discretizar
GRUPEDAD	Input	Discretizar
PV_TA	Input	Discretizar
PV_BA	Input	Discretizar

Tabla 4.80 (Transcripta). Prueba de Concepto - Selección de Variables del Modelo

Fuente Temporaria de datos (D.PD.CFT.FuTD): Se produce una única fuente temporaria de datos (figura 4.15, representada mediante una entidad del formalismo DER), la cual surge de realizar el muestreo estratificado respecto a la variable "consumo_ilegal_bool" de la Fuente Integrada de Datos, conformada por 6450 registros.

Figura 4.15. Prueba de Concepto – Fuente Temporaria de Datos (Diagrama Entidad-Relación)

Reporte de Generación de la Fuente Temporaria de datos (D.PD.CFT.RGFT): A partir de la fuente integrada de datos, se aplica con la herramienta seleccionada la técnica de muestreo estratificado (identificada en Estrategias de evaluación de modelos), indicando como monto total de la muestra el doble de la cantidad de consumidores de sustancias psicoactivas ilegales, utilizando como número de semilla el valor 1802. Adicionalmente, se describe la distribución de valores para cada atributo que formará parte del modelo, resaltando la cantidad de registros y la proporción que los mismos representan con respecto muestra. En la prueba de concepto, todos los atributos a utilizar son del tipo cualitativos, por lo que se registró la distribución de cada uno de sus valores.

Las tablas 4.83.a y 4.83.b ilustran el resultado obtenido. Para simplificar el formalismo resultante, se omitió la sección vacía de atributos cuantitativos.

Reporte de Generación de la Fuente Temporaria de datos			
Responsable:	Esposito E.	**Fecha:**	18/05/16
ID#:	D.PD.CFT.RGFT	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		
Fuente Temporaria de datos	(FTD.1) Fuente Temporaria de Datos		

ATRIBUTOS CUALITATIVOS		
Nombre	**Valores**	**Distribuciones**
POB_URB	1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes	396 (6,14%) 817 (12,67%) 2605 (40,39%) 2632 (40,81%)
RANGOING	0 Sin ingresos 1 1 a 600 2 601 a 800 3 801 a 1.000 4 1.001 a 1.500 5 1.501 a 2.000 6 2.001 a 2.500 7 2.501 a 3.000 8 3.001 a 3.500 9 3.501 a 4.000 10 4.001 a 4.500 11 4.501 a 5.500 12 5.001 a 6.000 13 6.001 a 7.000 14 7.001 a 8.000 15 8.001 a 10.000 16 10.001 a 15.000 17 15.001 y más 99 Ns/nc	19 (0,29%) 191 (2,96%) 177 (2,74%) 272 (4,22%) 565 (8,76%) 742 (11,5%) 545 (8,45%) 697 (10,81%) 379 (5,88%) 571 (8,85%) 207 (3,21%) 486 (7,53%) 378 (5,86%) 272 (4,22%) 266 (4,12%) 310 (4,81%) 241 (3,74%) 132 (2,05%) 0 (0%)
NBI_TOTAL	0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI	5644 (87,5%) 621 (9,63%) 156 (2,42%) 29 (0,45%) 0 (0%)
BHCH04	1 Varón 2 Mujer	3183 (49,35%) 3267 (50,65%)
NIVINSTR	1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial	54 (0,84%) 385 (5,97%) 969 (15,02%) 1352 (20,96%) 1509 (23,4%) 1072 (16,62%) 1109 (17,19%) 0 (0%)
BIAC01	1 Sí 2 No 9 Ns/nc	2646 (41,02%) 3804 (58,98%) 0 (0%)
PV_TA	1 Sí 0 No	4114 (63,78%) 2336 (36,22%)
PV_BA	1 Sí 0 No	5359 (83,09%) 1091 (16,91%)

Tabla 4.83.a Prueba de Concepto - Reporte de Generación de la Fuente Temporaria de datos

consumo_ilegal_bool	1 Sí 0 No	3245 (50,31%) 3205 (49,69%)
GRUPEDAD	2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años	1392 (21,58%) 1971 (30,56%) 1258 (19,50%) 1829 (28,36%)
Comentarios: número de semilla: 1802, atributo de estratificación: consumo_ilegal_bool, definición del tamaño de la muestra por valor absoluto: 6450		

Tabla 4.83.b Prueba de Concepto - Reporte de Generación de la Fuente Temporaria de datos

4.4.4.2. Actividad: Adecuación de la Fuente Temporaria de Datos (D.PD.AFT)

En esta actividad se analizan las características de los campos seleccionados para los distintos problemas de negocio, con el objetivo de identificar y realizar actividades de conversión y ajuste de los registros, preparando los datos para la adecuada extracción de patrones de conocimiento.

Información de Entrada

- Reporte de la Calidad de los Datos (D.ED.EvD.ReCD)
- Selección de variables del Modelo (D.Mo.CoM.SeVM)
- Fuente Temporaria de Datos (D.PD.CFT.FuTD)
- Reporte de Generación de la Fuente Temporaria de datos (D.PD.CFT.RGFT)

Información de Salida

- Reporte de Adecuación de la Fuente Temporaria de Datos (D.PD.AFT.RAFT)

4.4.4.2.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se propone el Reporte de Adecuación de la Fuente Temporaria de Datos, el cual se presenta a continuación.

Reporte de Adecuación de la Fuente Temporaria de Datos (D.PD.AFT.RAFT): Se formalizan los ajustes realizados a la fuente de información que integra todas las variables de interés para el problema de negocio, de acuerdo a las restricciones de los algoritmos de explotación de información a utilizar. Se registra el nombre de los campos y la acción que se aplicará (en las columnas homónimas), se registra el resultado producido de aplicar la acción (en la columna "Efecto") y se detallan las operaciones realizadas en la acción (en la columna "descripción"), en caso que fuese necesario. La tabla 4.84 ilustra la estructura del formalismo previamente descripto.

Reporte de Adecuación de la Fuente Temporaria de Datos			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Problema de Negocio:	Identificador del Problema de negocio		
Nombre	**Acción**	**Efecto**	**Descripción**
Campo 1	Operación realizada en el campo 1	Resultado obtenido	Descripción de la acción realizada al campo 1
…	…	…	…
Campo N	Operación realizada en el campo N	Resultado obtenido	Descripción de la acción realizada al campo N

Tabla 4.84. Formalismo: Reporte de Adecuación de la Fuente Temporaria de Datos

4.4.4.2.2. Técnica Identificada

Para el desarrollo de esta actividad se propone la técnica **"Adecuación de los Datos"**, en la cual se llevan a cabo las tareas de formateo, estandarización/normalización, reducción y limpieza de los campos, con el objetivo de mejorar la calidad de los mismos y por consiguiente, de los resultados del modelo. Las acciones a llevar a cabo surgen del análisis de la calidad de los datos (Reporte de la Calidad de los Datos) y del formato requerido de acuerdo a los procesos de explotación de información identificados (Selección de variables del Modelo).

A partir de las evaluaciones de la calidad de los datos, se identifican los campos que requieren ser ajustados, analizando la mejor solución para las problemáticas señaladas. Estas pueden ser la imputación de un valor estimado, la eliminación del registro o la eliminación del campo, en caso que esté presente la mayoría de sus valores incorrectos. De forma complementaria, se detallan las acciones a realizar con el propósito de mejorar la calidad de los datos o adecuar el formato de los campos de acuerdo a las necesidades del negocio y de los algoritmos a utilizar (identificadas en la Selección de Variables del Modelo), optimizando de esta forma los resultados de los algoritmos de explotación de información, así como la compatibilidad de los datos con los mismos. Una vez definida las acciones a ejecutar, estas se aplican a la fuente temporaria de datos, modificando la composición de la misma, registrando el impacto realizado, esto incluye asentar la cantidad de registros afectados, y opcionalmente, su identificador.

Se considera relevante destacar que las acciones de reducción o estandarización de los datos, deben ser aplicadas a todas las fuentes a utilizar (entrenamiento, validación y testeo) para el problema. En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo en la prueba de concepto.

4.4.4.2.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica Adecuación de los Datos, la cual utiliza como insumos los formalismos: Reporte de la Calidad de los Datos (Tabla 4.73), Selección de variables del Modelo (Tabla 4.80), Fuente Temporaria de Datos (Figura 4.15) y Reporte de Generación de la Fuente Temporaria de Datos (Tabla 4.83.a y 4.83.b). Estos elementos de entrada, son transcriptos con el mismo número de referencia, para facilitar al lector en la comprensión de la aplicación de la técnica.

Reporte de la Calidad de los Datos			
Responsable:	Esposito E.	**Fecha:**	06/05/2016
ID#:	D.ED.EvD.ReCD	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		

Nombre	**Registros**	**Tipo**	**Descripción**
RANGOING	2794	nulos	Valor 99. Ingresos no informados
NIVINSTR	43	Outlier	Valor 8: Minoría no representativa para el problema de negocio
BIAC01	97	nulos	Valor 9. no informado

Tabla 4.73 (Transcripta). Prueba de Concepto - Reporte de la Calidad de los Datos

Selección de variables del Modelo			
Responsable:	Esposito E.	**Fecha:**	13/05/2016
ID#:	D.Mo.CoM.SeVM	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		

Campo	**C4.5/ID3**	
	Tipo	**Conversión**
POB_URB	Input	Discretizar
RANGOING	Input	Discretizar
NBI_TOTAL	Input	Discretizar
BHCH04	Input	Discretizar
NIVINSTR	Input	Discretizar
BIAC01	Input	Discretizar
consumo_ilegal_bool	Target	Discretizar
GRUPEDAD	Input	Discretizar
PV_TA	Input	Discretizar
PV_BA	Input	Discretizar

Tabla 4.80 (Transcripta). Prueba de Concepto - Selección de Variables del Modelo

Figura 4.15 (Transcripta). Prueba de Concepto - Fuente Temporaria de datos (Diagrama Entidad-Relación)

Reporte de Generación de la Fuente Temporaria de datos			
Responsable:	Esposito E.	**Fecha:**	18/05/16
ID#:	D.PD.CFT.RGFT	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		
Fuente Temporaria de datos	(FTD.1) Fuente Temporaria de Datos		

ATRIBUTOS CUALITATIVOS		
Nombre	**Valores**	**Distribuciones**
POB_URB	1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes	396 (6,14%) 817 (12,67%) 2605 (40,39%) 2632 (40,81%)
RANGOING	0 Sin ingresos 1 1 a 600 2 601 a 800 3 801 a 1.000 4 1.001 a 1.500 5 1.501 a 2.000 6 2.001 a 2.500 7 2.501 a 3.000 8 3.001 a 3.500 9 3.501 a 4.000 10 4.001 a 4.500 11 4.501 a 5.500 12 5.001 a 6.000 13 6.001 a 7.000 14 7.001 a 8.000 15 8.001 a 10.000 16 10.001 a 15.000 17 15.001 y más 99 Ns/nc	19 (0,29%) 191 (2,96%) 177 (2,74%) 272 (4,22%) 565 (8,76%) 742 (11,5%) 545 (8,45%) 697 (10,81%) 379 (5,88%) 571 (8,85%) 207 (3,21%) 486 (7,53%) 378 (5,86%) 272 (4,22%) 266 (4,12%) 310 (4,81%) 241 (3,74%) 132 (2,05%) 0 (0%)
NBI_TOTAL	0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI	5644 (87,5%) 621 (9,63%) 156 (2,42%) 29 (0,45%) 0 (0%)
BHCH04	1 Varón 2 Mujer	3183 (49,35%) 3267 (50,65%)

Tabla 4.83.a (Transcripta). Prueba de Concepto - Reporte de Generación de la Fuente Temporaria de Datos

NIVINSTR	1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial	54 (0,84%) 385 (5,97%) 969 (15,02%) 1352 (20,96%) 1509 (23,4%) 1072 (16,62%) 1109 (17,19%) 0 (0%)
BIAC01	1 Sí 2 No 9 Ns/nc	2646 (41,02%) 3804 (58,98%) 0 (0%)
PV_TA	1 Sí 0 No	4114 (63,78%) 2336 (36,22%)
PV_BA	1 Sí 0 No	5359 (83,09%) 1091 (16,91%)
consumo_ilegal_bool	1 Sí 0 No	3245 (50,31%) 3205 (49,69%)
GRUPEDAD	2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años	1392 (21,58%) 1971 (30,56%) 1258 (19,50%) 1829 (28,36%)
Comentarios: número de semilla: 1802, atributo de estratificación: consumo_ilegal_bool, definición del tamaño de la muestra por valor absoluto: 6450		

Tabla 4.83.b (Transcripta). Prueba de Concepto - Reporte de Generación de la Fuente Temporaria de Datos

Reporte de Adecuación de la Fuente Temporaria de Datos (D.PD.AFT.RAFT): A partir del formalismo selección de variables del modelo, se identifica la necesidad de convertir a discretas las diez variables seleccionadas, mediante una conversión lineal (un valor numérico equivale a un valor discreto). Finalmente, se registra el conjunto de valores que fueron identificados para cada campo, destacando que para grupo edad no se encontraba presencia de un valor en la fuente de información, sirviendo el reporte de datos explorados como respaldo para corroborar la correcta ejecución de las conversiones. Adicionalmente, en el reporte de la calidad de los datos se identifican que tres campos presentan valores nulos o outliers, decidiendo remover los registros con dichos valores dado la relevancia de los campos, indicando la cantidad de registros que se vieron afectados, pudiendo corroborar la correcta ejecución en el reporte de datos explorados. Las tablas 4.85.a y 4.85.b ilustran el resultado previamente descripto en el formalismo propuesto.

Reporte de Adecuación de la Fuente Temporaria de Datos			
Responsable:	Esposito E.	**Fecha:**	20/05/16
ID#:	D.PD.AFT.RAFT	**Versión:**	1.0
Problema de Negocio:	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		

Nombre	**Acción**	**Efecto**	**Descripción**
POB_URB	Conversión a discreto	4 valores identificados	Cada valor numérico corresponde a un valor discreto
RANGOING	Conversión a discreto	19 valores identificados	Cada valor numérico corresponde a un valor discreto

Tabla 4.85.a Prueba de Concepto - Reporte de Adecuación de la Fuente Temporaria de Datos

NBI_TOTAL	Conversión a discreto	5 valores identificados	Cada valor numérico corresponde a un valor discreto
BHCH04	Conversión a discreto	2 valores identificados	Cada valor numérico corresponde a un valor discreto
NIVINSTR	Conversión a discreto	8 valores identificados	Cada valor numérico corresponde a un valor discreto
BIAC01	Conversión a discreto	3 valores identificados	Cada valor numérico corresponde a un valor discreto
consumo_ilegal_bool	Conversión a discreto	2 valores identificados	Cada valor numérico corresponde a un valor discreto
PV_BA	Conversión a discreto	2 valores identificados	Cada valor numérico corresponde a un valor discreto
PV_TA	Conversión a discreto	2 valores identificados	Cada valor numérico corresponde a un valor discreto
GRUPEDAD	Conversión a discreto	4 valores identificados. El valor Ns/nc no se encontraba en los registros	Cada valor numérico corresponde a un valor discreto
RANGOING	remover registros con ingreso no indicado (99)	97 registros eliminados	
NIVINSTR	remover registros con nivel de estudios Educación especial (8)	42 registros eliminados	No se poseen suficientes datos para considerar dicho valor
BIAC01	remover registros con valor no indicado (9)	2778 registros eliminados	

Tabla 4.85.b Prueba de Concepto - Reporte de Adecuación de la Fuente Temporaria de Datos

4.4.5. Fase: Implementación (D.Im)

En la quinta fase del subproceso de desarrollo, se llevan a cabo las tareas de definición del modelo y la extracción de los patrones de conocimiento de la fuente de información identificadas. Está conformada por dos actividades: Selección del Modelo (sección 4.4.5.1), donde se define la estrategia a utilizar para identificar la mejor configuración del modelo, y explotación de información (sección 4.4.5.2), donde se realiza la extracción y descripción de los patrones de conocimiento ocultos en los datos. La figura 4.16, presenta una visión resumida de las actividades que integran la fase y sus elementos de entrada y salida (las imágenes de cada formalismo son representaciones miniatura de los mismos).

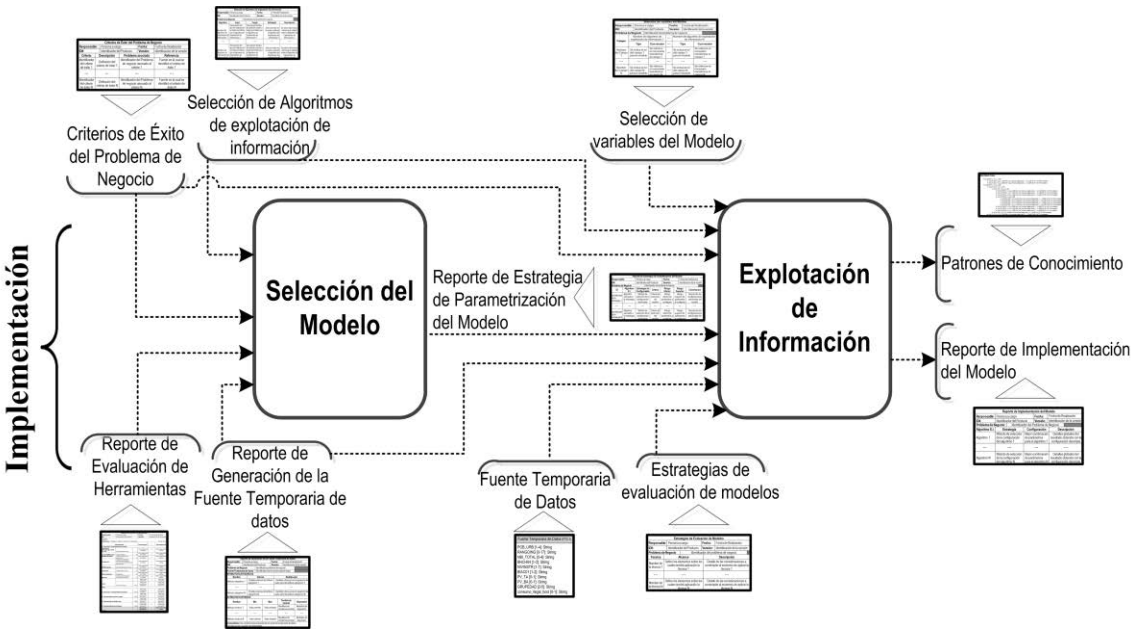


Figura 4.16. Fase: Implementación

4.4.5.1. Actividad: Selección del Modelo (D.Im.SeM)

En esta actividad se definen el criterio y la forma mediante la cual se determinará cuál de los posibles algoritmos o combinación de algoritmos logran capturar (o generalizar) con mayor precisión los patrones ocultos en los datos. Las comparaciones y evaluaciones a realizar, no solo contemplan distintos algoritmos, sino también sus distintos parámetros de configuración. Mediante esta actividad se pretende responder a las preguntas de: ¿Cómo se elige el mejor algoritmo para el problema de negocio? Y ¿Cómo se definen los parámetros de configuración del algoritmo? [Shalev-Shwartz & Ben-David, 2014]. Como resultado de la misma debe formalizarse el método a utilizar para analizar los algoritmos y el criterio para su selección.

Información de Entrada

- Selección de Algoritmos de explotación de información (D.Mo.CoM.SAEI)
- Reporte de Generación de la Fuente Temporaria de datos (D.PD.CFT.RGFT)
- Criterios de Éxito del Problema de Negocio (D.EN.CPN.CEPN)
- Reporte de Evaluación de Herramientas (G.In.EvS.EvHe)

Información de Salida

- Reporte de Estrategia de Parametrización del Modelo (D.Im.SeM.REPM)

4.4.5.1.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se propone el Reporte de Estrategia de Parametrización del Modelo, el cual se describe a continuación.

Reporte de Estrategia de Parametrización del Modelo (D.Im.SeM.REPM): Se formalizan las estrategias a utilizar para seleccionar el mejor modelo. Para ello, se identifica el método a utilizar para evaluar la mejor configuración del modelo (en la columna "estrategia de configuración"), indicando el algoritmo de explotación de información (E.I) asociado y un identificador único para el mismo (en las columnas homónimas). Se registran el criterio a utilizar para determinar la mejor configuración y el rango de valores a evaluar (inferior y superior). Finalmente, se incorporan otros aspectos que deban ser considerados al momento de implementar las estrategias seleccionadas. La tabla 4.86 ilustra la estructura del formalismo previamente descripto.

Reporte de Estrategia de Parametrización del Modelo						
Responsable:		Persona a cargo		**Fecha:**	Fecha de Realización	
ID#:		Identificador del Producto		**Versión:**	Identificación de la versión	
Problema de Negocio		Identificador del problema de negocio				
ID	**Algoritmo E.I.**	**Estrategia de Configuración**	**Criterio**	**Rango Inferior**	**Rango Superior**	**Comentarios**
Identificador de la estrategia 1	Algoritmo asociado a la estrategia 1	Método de selección de la configuración del modelo	Criterio de selección del modelo	Rango inferior de parámetros a configurar	Rango superior de parámetros a configurar	Descripción de configuraciones adicionales del modelo
…	…	…	…	…	…	…
Identificador de la estrategia N	Algoritmo asociado a la estrategia N	Método de selección de la configuración del modelo	Criterio de selección del modelo	Rango inferior de parámetros a configurar	Rango superior de parámetros a configurar	Descripción de configuraciones adicionales del modelo

Tabla 4.86. Formalismo: Reporte de Estrategia de Parametrización del Modelo

4.4.5.1.2. Técnica Identificada

Para el desarrollo de esta actividad se propone la técnica **"Selección de la Estrategia de Hiperparametrización"**, mediante la cual se establecen los aspectos a considerar para la evaluación y selección del modelo. Para ello, se debe tener en consideración el tipo de problema de explotación de información a resolver, los criterios de éxito definidos para el mismo y las características de la muestra sobre la cual se extraerán los patrones. Adicionalmente, se deberá tener en consideración las posibilidades y restricciones que las herramientas seleccionadas brindan con respecto a los algoritmos de explotación de información seleccionados, las técnicas de hiperparametrización, de medición y evaluación de los resultados.

El primer paso consiste en evaluar el método o proceso de explotación de información a aplicar, el cual puede identificarse de acuerdo a la estrategia de implementación de algoritmos (formalismo de Selección de Algoritmos de Explotación de Información), comprendiendo el objetivo final del proceso de extracción del conocimiento. Una vez identificado los algoritmos que intervienen en la resolución del problema y las características de la implementación de los mismos en la herramienta

seleccionada, se determina que parámetros se pueden configurar, cuáles serán predefinidos por el ingeniero de explotación de información y cuáles serán determinados a partir de alguna estrategia (si se considerase oportuno), y a partir de ello, evaluar cuál es la mejor estrategia para hallar la configuración óptima, teniendo en cuenta la complejidad y el costo de las mismas.

Según los criterios de éxito del problema de negocio y las características de la muestra a utilizar, se define el mejor criterio para evaluar la calidad del modelo y por consiguiente, de los resultados obtenidos. Este varía de acuerdo al tipo de algoritmo que se debe utilizar (clasificación, agrupamiento, etc.) y a si la distribución de la muestra se encuentra o no equilibrada respecto a la variable clase (o target). Por ejemplo: sí los algoritmos a utilizar son de clasificación se pueden utilizar medidas como precisión, tasa de error, exactitud, recall, etc., en cambio sí es de agrupamiento, se pueden utilizar métricas de similitud inter e intra-cluster, como por ejemplo: coeficiente de silueta.

Adicionalmente, es posible utilizar distintas visualizaciones como herramientas que faciliten la comprensión y selección del modelo, cuando el mismo no puede ser automatizado (como por ejemplo, para agrupamiento: el uso de gráfico de polígono considerando la consistencia de los clusters identificados de acuerdo a distintas medidas, conocido como elbow method, o para clasificación: graficando las curvas de aprendizaje en los distintos estadios como indicador de precisión, sesgo y varianza del modelo, curva ROC (del inglés Receiver Operating Characteristic, o Característica Operativa del Receptor) como indicador de la performance del modelo, entre otros). Como resultado de esta actividad, debe definirse el método y los criterios a utilizar para seleccionar los modelos.

En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo para la prueba de concepto.

4.4.5.1.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica Selección de la Estrategia de Hiperparametrización, la cual utiliza como insumos los formalismos: Selección de Algoritmos de explotación de información (Tabla 4.79), el Reporte de Generación de la Fuente Temporaria de datos (Tablas 4.83.a y 4.83.b), los Criterios de Éxito del Problema de Negocio (Tabla 4.65) y la Reporte de Evaluación de Herramientas (Tablas 4.13.a y 4.13.b).

Los formalismos indicados como elementos de entrada, son transcriptos con el mismo número de tabla, para facilitar al lector en la comprensión de la aplicación de la técnica.

Selección de Algoritmos de explotación de información				
Responsable:	Esposito E.	**Fecha:**	11/05/2016	
ID#:	D.Mo.CoM.SAEI	**Versión:**	1.0	
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo			
Algoritmo	**Input**	**Target**	**Estrategia**	**Descripción**
C4.5	Discretos/Continuos	Discretos	1	familia TDIDT
ID3	Discretos/Continuos	Discretos	1	familia TDIDT

Tabla 4.79 (Transcripta). Prueba de Concepto - Selección de Algoritmos de explotación de información

Reporte de Generación de la Fuente Temporaria de datos			
Responsable:	Esposito E.	**Fecha:**	18/05/16
ID#:	D.PD.CFT.RGFT	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		
Fuente Temporaria de datos	(FTD.1) Fuente Temporaria de datos		
ATRIBUTOS CUALITATIVOS			
Nombre	**Valores**	**Distribuciones**	
POB_URB	1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes	396 (6,14%) 817 (12,67%) 2605 (40,39%) 2632 (40,81%)	
RANGOING	0 Sin ingresos 1 1 a 600 2 601 a 800 3 801 a 1.000 4 1.001 a 1.500 5 1.501 a 2.000 6 2.001 a 2.500 7 2.501 a 3.000 8 3.001 a 3.500 9 3.501 a 4.000 10 4.001 a 4.500 11 4.501 a 5.500 12 5.001 a 6.000 13 6.001 a 7.000 14 7.001 a 8.000 15 8.001 a 10.000 16 10.001 a 15.000 17 15.001 y más 99 Ns/nc	19 (0,29%) 191 (2,96%) 177 (2,74%) 272 (4,22%) 565 (8,76%) 742 (11,5%) 545 (8,45%) 697 (10,81%) 379 (5,88%) 571 (8,85%) 207 (3,21%) 486 (7,53%) 378 (5,86%) 272 (4,22%) 266 (4,12%) 310 (4,81%) 241 (3,74%) 132 (2,05%) 0 (0%)	
NBI_TOTAL	0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI	5644 (87,5%) 621 (9,63%) 156 (2,42%) 29 (0,45%) 0 (0%)	

Tabla 4.83.a (Transcripta). Prueba de Concepto - Reporte de Generación de la Fuente Temporaria de datos

BHCH04	1 Varón 2 Mujer	3183 (49,35%) 3267 (50,65%)
NIVINSTR	1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial	54 (0,84%) 385 (5,97%) 969 (15,02%) 1352 (20,96%) 1509 (23,4%) 1072 (16,62%) 1109 (17,19%) 0 (0%)
BIAC01	1 Sí 2 No 9 Ns/nc	2646 (41,02%) 3804 (58,98%) 0 (0%)
PV_TA	1 Sí 0 No	4114 (63,78%) 2336 (36,22%)
PV_BA	1 Sí 0 No	5359 (83,09%) 1091 (16,91%)
consumo_ilegal_bool	1 Sí 0 No	3245 (50,31%) 3205 (49,69%)
GRUPEDAD	2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años	1392 (21,58%) 1971 (30,56%) 1258 (19,50%) 1829 (28,36%)
Comentarios: número de semilla: 1802, atributo de estratificación: consumo_ilegal_bool, definición del tamaño de la muestra por valor absoluto: 6450		

Tabla 4.83.b (Transcripta). Prueba de Concepto - Reporte de Generación de la Fuente Temporaria de datos

Reporte de Evaluación de Herramientas					
Responsable:	Rodriguez H.			**Fecha:**	07/04/2016
ID#:	G.In.EvS.REHe			**Versión:**	1.0
Criterios: **Evaluación:** 1 = Malo, 2 = débil, 3 = Bueno, 4 = Excelente					1 = No, 4 = SI
Herramientas			**Tanagra V.1.4.50**	**Weka V.3.7.11**	**Orange V.2.7.8**
1. Funcional - Características Técnicas					
Soporte de Metodología / Ciclo de vida	Soporte del proceso	3	2	2	2
Compatibilidad con fuentes de datos	Base de datos	8	--	--	--
	Otras fuentes (word, excel, etc.)	8	3	2	3
Integración	Soporte de distintas técnicas asociadas al proceso de explotación de Información	5	4	4	4
Multilenguaje	Soporta distintas idiomas	2	1	1	1
Técnicas	Variedad de técnicas que provee	18	4	4	4
Reporte y visualización	Permite generar reportes y visualizaciones	12	2	2	2
Multiplataforma	Soporta múltiples plataformas	5	1	4	4
Instalación remota	La administración y mantenimiento son remotos	5	--	--	--
Usuarios Múltiples	Posee perfiles de usuarios	2	1	1	1
Seguridad	Provee seguridad de la información configurada por perfiles	2	1	1	1

Tabla 4.13.a (Transcripta). Prueba de Concepto - Reporte de Evaluación de Herramientas

Backup	Metodología de backup	2	1	1	1
Amigable	Interfaz de usuario	10	4	2	4
Configuraciones	Permite la configuración del perfil	8			
Documentación	Servicio de soporte y ayuda	5	4	1	3
Conexión	Soporta conexión por: Internet, FTP, ERPs.	2	1	1	1
Soporte de sistemas de mensaje	Soporta compartir información (por mail u otro medio)	3	1	1	1
Total			224	196	234
	Peso del Grupo	40%	89,6	78,4	93,6
2. Características del Proveedor					
Características del proveedor	Historia	30	3	3	1
Crecimiento	Perspectiva a futuro	10	2	3	2
Ubicación Geográfica	Oficinas	30	--	--	--
Implementación	Otras implementaciones de la misma herramienta	5	--	--	--
	Contacto con otros clientes	5	--	--	--
Confidencialidad	Confidencialidad de la información	20	--	--	--
Total			110	120	50
	Peso del Grupo	25%	27,5	30	12,5
3. Características del Servicio					
Garantía del producto	Duración y Alcance	30	--	--	--
Mejora	Brinda soporte a versiones previas	20	1	1	1
Licencia	Costo, alcances y soporte postventa	30	--	--	--
Soporte	Tiempo de respuesta y disponibilidad	20	--	--	--
Total			20	20	20
	Peso del Grupo	20%	4	4	4
4. Características Económicas					
Costo del software	Costo de la herramienta	30	--	--	--
Costo del Hardware	Necesidad de mejorar o comprar nuevo hardware compatible con la herramienta	20	--	--	--
Otros costos software	Costos adicionales al producto (backup, web servers, bases de datos, etc.)	20	--	--	--
Licencias	Política de licencia	10	--	--	--
Financiamiento	Existencia	10	--	--	--
Mejoras	Costo promedio de la mejora del producto	10	--	--	--
Total			0	0	0
	Peso del Grupo	-15%	0	0	0
Final					
1. Funcional - Características Técnicas		40%	89,6	78,4	93,6
2. Características del Proveedor		25%	27,5	30	12,5
3. Características del Servicio		20%	4	4	4
4. Características Económicas		-15%	0	0	0
TOTAL			**121,1**	**112,4**	**110,1**

Tabla 4.13.b (Transcripta). Prueba de Concepto - Reporte de Evaluación de Herramientas

Criterios de Éxito del Problema de Negocio				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.CEPN	**Versión:**	1.0	
Criterio	**Descripción**	**Problema asociado**		**Referencia**
cepn.1	Identificar aspectos que permitan comprender el comportamiento de grupos masivos de personas de acuerdo a la evaluación de Silva H. (rehi.3)	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		Entrevista 3

Tabla 4.65 (Transcripta). Prueba de Concepto - Criterios de Éxito del Problema de Negocio

Reporte de Estrategia de Parametrización del Modelo (D.Im.SeM.REPM): De acuerdo al tipo de problema a resolver (descubrimiento de reglas de comportamiento), los algoritmos a utilizar (clasificación TDIDT) y la herramienta seleccionada, se definen los siguientes parámetros y rangos de valores posibles para su optimización, los cuales serán definidos mediante la estrategia de búsqueda aleatoria (o Random Search):

- Algoritmo C4.5: Mínima cantidad de hojas por rama (Min size of leaves) seleccionando como rango de valores posibles de 100 a 200. Adicionalmente, se define el nivel de confidencia (Confidence Level) de manera constante en 0.25.
- Algoritmo ID3: el tamaño mínimo de hojas para dividir una rama (min size for Split) entre 50 y 150, y la Mínima cantidad de hojas por rama (Min size of leaves) seleccionando como rango de valores posibles de 100 a 200. Además, se definen los parámetros máxima profundidad del árbol (max depth of the tree) y mínima ganancia de entropía para su división (Min entropy gain for splitting) como constantes con valor 10 y 0.03 respectivamente.

Adicionalmente, de acuerdo a las características de la fuente temporaria de datos (la cual fue estratificada), se decide utilizar como criterio de evaluación de performance la tasa de error. En la tabla 4.87 se formaliza la información previamente descripta.

Reporte de Estrategia de Parametrización del Modelo						
Responsable:	Esposito E.		**Fecha:**	24/05/2016		
ID#:	D.Im.SeM.REPM		**Versión:**	1.0		
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo					
ID	**Algoritmo E.I.**	**Estrategia de Configuración**	**Criterio**	**Rango Inferior**	**Rango Superior**	**Comentarios**
repm.1	C4.5	RadomSearch	Tasa de Error	Min size of leaves: 100	Min size of leaves: 200	Confidence Level:0,25
repm.2	ID3	RadomSearch	Tasa de Error	min size for split: 50 min size of leaves: 50	min size for split: 150 min size of leaves: 200	max depth of the tree: 10 Min entropy gain for splitting: 0.03

Tabla 4.87. Prueba de Concepto - Reporte de Estrategia de Parametrización del Modelo

4.4.5.2. Actividad: Explotación de Información (D.Im.ExI)

En esta actividad se aplican los algoritmos de explotación de información (o minería de datos), con el objetivo de extraer los patrones de conocimientos ocultos en las fuentes de información, dejando constancia de los resultados obtenidos para poder reproducir y comparar los mismos. El resultado de esta actividad debe dar respuesta al problema de negocio definido y contribuir con respecto al objetivo general del proyecto.

Información de Entrada

- Criterios de Éxito del Problema de Negocio (D.EN.CPN.CEPN)
- Selección de Algoritmos de explotación de información (D.Mo.CoM.SAEI)
- Selección de variables del Modelo (D.Mo.CoM.SeVM)
- Estrategias de evaluación de modelos (D.Mo.CoM.EsEM)
- Reporte de Estrategia de Parametrización del Modelo (D.Im.SeM.REPM)
- Fuente Temporaria de Datos (D.PD.CFT.FuTD)
- Reporte de Generación de la Fuente Temporaria de datos (D.PD.CFT.RGFT)

Información de Salida

- Reporte de Implementación del Modelo (D.Im.ExI.ReIM)
- Patrones de Conocimiento (D.Im.ExI.PaCo)

4.4.5.2.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se propone el Reporte de Implementación del Modelo, el cual se presenta a continuación.

Reporte de Implementación del Modelo (D.Im.ExI.ReIM): Se deja registro formal de los parámetros de la mejor configuración obtenida para los algoritmos a analizar, de acuerdo a la estrategia de configuración seleccionada (asentando dicha información en las columnas homónimas). Adicionalmente, se detalla las características generales de los patrones obtenidos por el modelo en la columna “descripción”. La tabla 4.88 ilustra la estructura del formalismo previamente descripto.

Reporte de Implementación del Modelo			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Problema de Negocio:	Identificador del Problema de Negocio		
Algoritmo E.I.	**Estrategia**	**Configuración**	**Descripción**
Algoritmo 1	Método de selección de la configuración del algoritmo 1	Mejor combinación de parámetros para el algoritmo 1	Detalles globales del resultado obtenido con la configuración descripta
…	…	…	…
Algoritmo N	Método de selección de la configuración del algoritmo N	Mejor combinación de parámetros para el algoritmo N	Detalles globales del resultado obtenido con la configuración descripta

Tabla 4.88. Formalismo: Reporte de Implementación del Modelo

4.4.5.2.2. Técnica Identificada

Para el desarrollo de esta actividad se aplica la técnica **"Extracción de Conocimiento"**, en la cual se obtienen los patrones de conocimientos ocultos en la información disponible, identificados por la mejor configuración de los modelos seleccionados.

En esta etapa, se describen los resultados obtenidos con respecto a la calidad (asociado con las métricas para evaluar resultados), siendo la primera instancia en la cual puede identificarse si los resultados obtenidos no satisfacen los criterios de éxito, y en dicho caso determinar cuáles son los posibles pasos a seguir. Se debe tener en consideración que en esta etapa, se realiza una evaluación inicial vinculada con las características cuantitativas del modelo, en la próxima fase se realizará una validación de los resultados obtenidos.

Una vez aplicadas las técnicas de configuración y evaluación del modelo, se detalla para cada algoritmo y estrategia utilizada, cuál fue el mejor resultado obtenido y las características del mismo (configuración de parámetros, valores de las métricas de interés, de ser posible la complejidad del patrón, entre otros.), con el objetivo de permitir la reproducción y comparación de resultados.

En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo para la prueba de concepto.

4.4.5.2.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de implementar la técnica Extracción de Conocimiento, la cual utiliza como insumos los formalismos: Criterios de Éxito del Problema de Negocio (Tabla 4.65), Selección de Algoritmos de explotación de información (Tabla 4.79), Selección de variables del Modelo (Tabla 4.80), Estrategias de evaluación de modelos (Tabla 4.81),

Fuente Temporaria de Datos (Figura 4.15), Reporte de Generación de la Fuente Temporaria de datos (Tabla 4.83) y Reporte de Estrategia de Parametrización del Modelo (Tabla 4.87), los cuales son transcriptos con el mismo número de referencia, para facilitar al lector en la comprensión de la aplicación de la técnica.

Criterios de Éxito del Problema de Negocio					
Responsable:	Esposito E.		**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.CEPN		**Versión:**	1.0	
Criterio	**Descripción**	**Problema asociado**			**Referencia**
cepn.1	Identificar aspectos que permitan comprender el comportamiento de grupos masivos de personas de acuerdo a la evaluación de Silva H. (rehi.3)	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo			Entrevista 3

Tabla 4.65 (Transcripta). Prueba de Concepto - Criterios de Éxito del Problema de Negocio

Selección de Algoritmos de explotación de información				
Responsable:	Esposito E.	**Fecha:**	11/05/2016	
ID#:	D.Mo.CoM.SAEI	**Versión:**	1.0	
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo			
Algoritmo	**Input**	**Target**	**Estrategia**	**Descripción**
C4.5	Discretos/Continuos	Discretos	1	familia TDIDT
ID3	Discretos/Continuos	Discretos	1	familia TDIDT

Tabla 4.79 (Transcripta). Prueba de Concepto - Selección de Algoritmos de explotación de información

Selección de variables del Modelo			
Responsable:	Esposito E.	**Fecha:**	13/05/2016
ID#:	D.Mo.CoM.SeVM	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		
Campo	**C4.5/ID3**		
	Tipo	**Conversión**	
POB_URB	Input	Discretizar	
RANGOING	Input	Discretizar	
NBI_TOTAL	Input	Discretizar	
BHCH04	Input	Discretizar	
NIVINSTR	Input	Discretizar	
BIAC01	Input	Discretizar	
consumo_ilegal_bool	Target	Discretizar	
GRUPEDAD	Input	Discretizar	
PV_TA	Input	Discretizar	
PV_BA	Input	Discretizar	

Tabla 4.80 (Transcripta). Prueba de Concepto - Selección de variables del Modelo

Estrategias de evaluación de modelos			
Responsable:	Esposito E.	**Fecha:**	13/05/2016
ID#:	D.Mo.CoM.EsEM	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		
Técnica	**Alcance**	**Descripción**	
Muestreo Estratificado	global	balancear la distribución de la variable target al doble de la población de consumidores de sustancias psicoactivas ilegales	
Cross-Validation	C4.5/ID3	10 fold Cross-Validation con 10 repeticiones	

Tabla 4.81 (Transcripta). Prueba de Concepto - Estrategias de Evaluación de Modelos

Reporte de Estrategia de Parametrización del Modelo						
Responsable:		Esposito E.		**Fecha:**	24/05/2016	
ID#:		D.Im.SeM.REPM		**Versión:**	1.0	
Problema de Negocio		(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo				
ID	**Algoritmo E.I.**	**Estrategia de Configuración**	**Criterio**	**Rango Inferior**	**Rango Superior**	**Comentarios**
repm.1	C4.5	RadomSearch	Tasa de Error	Min size of leaves: 100	Min size of leaves: 200	Confidence Level:0,25
repm.2	ID3	RadomSearch	Tasa de Error	min size for split: 50 min size of leaves: 50	min size for split: 150 min size of leaves: 200	max depth of the tree: 10 Min entropy gain for splitting: 0.03

Tabla 4.87 (Transcripta). Prueba de Concepto - Reporte de Estrategia de Parametrización del Modelo

Figura 4.15 (Transcripta). Prueba de Concepto - Fuente Temporaria de datos (Diagrama Entidad-Relación)

Reporte de Generación de la Fuente Temporaria de datos			
Responsable:	Esposito E.	**Fecha:**	18/05/16
ID#:	D.PD.CFT.RGFT	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		
Fuente Temporaria de datos	(FTD.1) Fuente Temporaria de Datos		

ATRIBUTOS CUALITATIVOS		
Nombre	**Valores**	**Distribuciones**
POB_URB	1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes	396 (6,14%) 817 (12,67%) 2605 (40,39%) 2632 (40,81%)
RANGOING	0 Sin ingresos 1 1 a 600 2 601 a 800 3 801 a 1.000 4 1.001 a 1.500 5 1.501 a 2.000 6 2.001 a 2.500 7 2.501 a 3.000 8 3.001 a 3.500 9 3.501 a 4.000 10 4.001 a 4.500 11 4.501 a 5.500 12 5.001 a 6.000 13 6.001 a 7.000 14 7.001 a 8.000 15 8.001 a 10.000 16 10.001 a 15.000 17 15.001 y más 99 Ns/nc	19 (0,29%) 191 (2,96%) 177 (2,74%) 272 (4,22%) 565 (8,76%) 742 (11,5%) 545 (8,45%) 697 (10,81%) 379 (5,88%) 571 (8,85%) 207 (3,21%) 486 (7,53%) 378 (5,86%) 272 (4,22%) 266 (4,12%) 310 (4,81%) 241 (3,74%) 132 (2,05%) 0 (0%)
NBI_TOTAL	0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI	5644 (87,5%) 621 (9,63%) 156 (2,42%) 29 (0,45%) 0 (0%)
BHCH04	1 Varón 2 Mujer	3183 (49,35%) 3267 (50,65%)
NIVINSTR	1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial	54 (0,84%) 385 (5,97%) 969 (15,02%) 1352 (20,96%) 1509 (23,4%) 1072 (16,62%) 1109 (17,19%) 0 (0%)
BIAC01	1 Sí 2 No 9 Ns/nc	2646 (41,02%) 3804 (58,98%) 0 (0%)
PV_TA	1 Sí 0 No	4114 (63,78%) 2336 (36,22%)
PV_BA	1 Sí 0 No	5359 (83,09%) 1091 (16,91%)
consumo_ilegal_bool	1 Sí 0 No	3245 (50,31%) 3205 (49,69%)
GRUPEDAD	2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años	1392 (21,58%) 1971 (30,56%) 1258 (19,50%) 1829 (28,36%)
Comentarios: número de semilla: 1802, atributo de estratificación: consumo_ilegal_bool, definición del tamaño de la muestra por valor absoluto: 6450		

Tabla 4.83 (Transcripta). Prueba de Concepto - Reporte de Generación de la Fuente Temporaria de Datos

Reporte de Implementación del Modelo (D.Im.ExI.ReIM): A partir de la implementación de los algoritmos de explotación de información seleccionados (C4.5 e ID3) en la Fuente Temporaria de Datos (FTD.1), mediante la aplicación de las técnicas de búsqueda aleatoria (Radom Search) y la implementación de 10 fold Cross-Validation con 10 repeticiones, se obtuvieron como mejores configuraciones para cada algoritmo (Min size of leaves: 150; Confidence Level: 0,25) y (min size for split: 50; min size of leaves: 50; max depth of the tree: 10; Min entropy gain for splitting: 0.03) respectivamente, y se obtuvieron como patrón de conocimiento, árboles de decisiones con las siguientes características:

- La estructura del árbol obtenido para el algoritmo C4.5, está conformada por 20 nodos, de los cuales 14 son nodos hojas, y presenta una tasa de error (promedio de las repeticiones de cross-validation) de 0.2565.
- La estructura del árbol obtenido para el algoritmo ID3, está conformada por 34 nodos, de los cuales 28 son nodos hojas, y presenta una tasa de error (promedio de las repeticiones de cross-validation) de 0.2697.

Debido a la imposibilidad de definir un criterio cuantitativo para la verificación de la calidad del modelo, se procederá en la próxima fase a evaluar con el experto de forma cualitativa los resultados obtenidos. La tabla 4.89 ilustra el resultado obtenido de la implementación de los procesos de explotación de información.

Reporte de Implementación del Modelo			
Responsable:	Esposito E.	**Fecha:**	27/05/2016
ID#:	D.Im.ExI.ReIM	**Versión:**	1.0
Problema de Negocio:	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		

Algoritmo E.I.	**Estrategia**	**Configuración**	**Descripción**
C4.5	(repm.1) RadomSearch	Min size of leaves: 150 Confidence Level: 0,25	Número de nodos: 20 Número de Hojas: 14 Tasa de error (promedio): 0.2558
ID3	(repm.2) RadomSearch	min size for split: 50 min size of leaves: 50 max depth of the tree: 10 Min entropy gain for splitting: 0.03	Número de nodos: 34 Número de Hojas: 28 Tasa de error (promedio): 0.2697

Tabla 4.89. Prueba de Concepto - Reporte de Implementación del Modelo

Patrones de Conocimiento (D.Im.ExI.PaCo): Se detallan los patrones obtenidos por el modelo que obtuvo mejores resultados (C4.5), presentando las reglas de comportamiento generadas a partir la muestra estudiada. La figura 4.17 ilustra los resultados obtenidos, los cuales conforman 8 reglas que se describen a continuación:

SI Conoce personas cercanas que en la actualidad consuman alguna sustancia
Y posee Prevalencia de vida de consumo de bebidas alcohólicas
ENTONCES presenta prevalencia de vida de consumo de sustancias psicoactivas ilegales (79,91% de 2459)

SI Conoce personas cercanas que en la actualidad consuman alguna sustancia
Y NO posee Prevalencia de vida de consumo de bebidas alcohólicas
ENTONCES NO presenta prevalencia de vida de consumo de sustancias psicoactivas ilegales (67.38% de 187)

SI NO conoce personas cercanas que en la actualidad consuman alguna sustancia
Y posee Prevalencia de vida de consumo de tabaco
Y posee Prevalencia de vida de consumo de bebidas alcohólicas
Y s u nivel de instrucción es Terciario o universitario incompleto o superior
ENTONCES presenta prevalencia de vida de consumo de sustancias psicoactivas ilegales (58,89% de 579)

SI NO conoce personas cercanas que en la actualidad consuman alguna sustancia
Y posee Prevalencia de vida de consumo de tabaco
Y posee Prevalencia de vida de consumo de bebidas alcohólicas
Y su nivel de instrucción es Secundario completo
Y reside en un agrupamiento de poblaciones urbanas de 500.001 habitantes o más
ENTONCES presenta prevalencia de vida de consumo de sustancias psicoactivas ilegales (64,38% de 73)

SI NO conoce personas cercanas que en la actualidad consuman alguna sustancia
Y posee Prevalencia de vida de consumo de tabaco
Y posee Prevalencia de vida de consumo de bebidas alcohólicas
Y su nivel de instrucción es Secundario completo
Y reside en un agrupamiento de poblaciones urbanas de 500.000 habitantes o menos
ENTONCES NO presenta prevalencia de vida de consumo de sustancias psicoactivas ilegales (58,57% de 391)

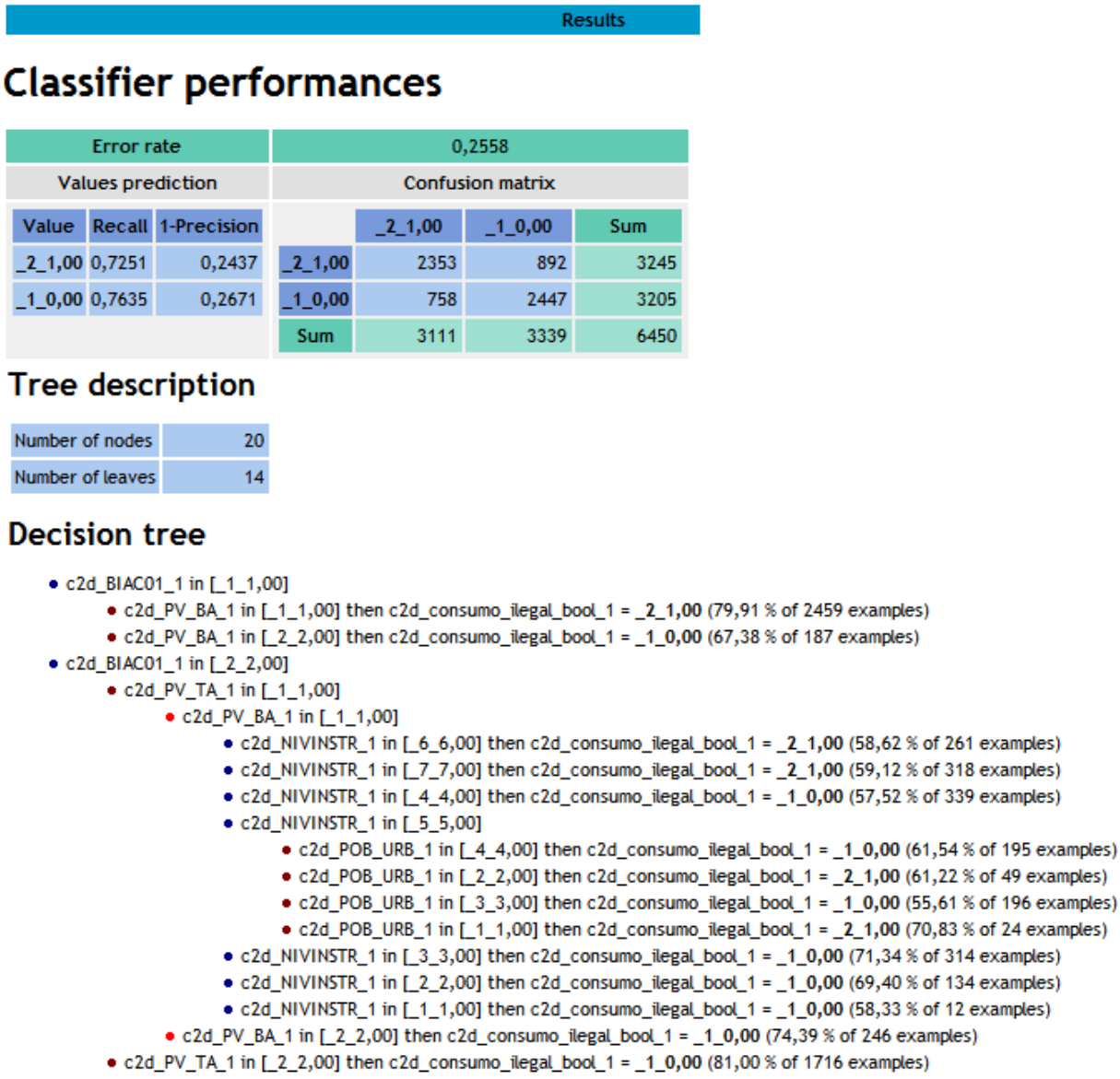
SI NO conoce personas cercanas que en la actualidad consuman alguna sustancia
Y posee Prevalencia de vida de consumo de tabaco
Y posee Prevalencia de vida de consumo de bebidas alcohólicas
Y su nivel de instrucción es Secundario incompleto o inferior
ENTONCES NO presenta prevalencia de vida de consumo de sustancias psicoactivas ilegales (64,96% de 799)

SI NO conoce personas cercanas que en la actualidad consuman alguna sustancia
Y posee Prevalencia de vida de consumo de tabaco
Y NO posee Prevalencia de vida de consumo de bebidas alcohólicas
ENTONCES NO presenta prevalencia de vida de consumo de sustancias psicoactivas ilegales (74.39% de 246)

SI NO conoce personas cercanas que en la actualidad consuman alguna sustancia
Y **NO** posee Prevalencia de vida de consumo de tabaco
ENTONCES NO presenta prevalencia de vida de consumo de sustancias psicoactivas ilegales (81% de 1716)

4.4.6. Fase: Evaluación y Presentación (D.EP)

La fase Evaluación y Presentación tiene como objetivo realizar las validaciones finales del producto (identificando nuevas necesidades en caso que existiesen) y presentar los resultados obtenidos a los interesados, garantizando una adecuada comprensión del procedimiento y las conclusiones derivadas para el proceso de toma de decisión.



Results

Classifier performances

Error rate	0,2558

Values prediction

Value	Recall	1-Precision
_2_1,00	0,7251	0,2437
_1_0,00	0,7635	0,2671

Confusion matrix

	_2_1,00	_1_0,00	Sum
_2_1,00	2353	892	3245
_1_0,00	758	2447	3205
Sum	3111	3339	6450

Tree description

Number of nodes	20
Number of leaves	14

Decision tree

- c2d_BIAC01_1 in [_1_1,00]
  - c2d_PV_BA_1 in [_1_1,00] then c2d_consumo_ilegal_bool_1 = **_2_1,00** (79,91 % of 2459 examples)
  - c2d_PV_BA_1 in [_2_2,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (67,38 % of 187 examples)
- c2d_BIAC01_1 in [_2_2,00]
  - c2d_PV_TA_1 in [_1_1,00]
    - c2d_PV_BA_1 in [_1_1,00]
      - c2d_NIVINSTR_1 in [_6_6,00] then c2d_consumo_ilegal_bool_1 = **_2_1,00** (58,62 % of 261 examples)
      - c2d_NIVINSTR_1 in [_7_7,00] then c2d_consumo_ilegal_bool_1 = **_2_1,00** (59,12 % of 318 examples)
      - c2d_NIVINSTR_1 in [_4_4,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (57,52 % of 339 examples)
      - c2d_NIVINSTR_1 in [_5_5,00]
        - c2d_POB_URB_1 in [_4_4,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (61,54 % of 195 examples)
        - c2d_POB_URB_1 in [_2_2,00] then c2d_consumo_ilegal_bool_1 = **_2_1,00** (61,22 % of 49 examples)
        - c2d_POB_URB_1 in [_3_3,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (55,61 % of 196 examples)
        - c2d_POB_URB_1 in [_1_1,00] then c2d_consumo_ilegal_bool_1 = **_2_1,00** (70,83 % of 24 examples)
      - c2d_NIVINSTR_1 in [_3_3,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (71,34 % of 314 examples)
      - c2d_NIVINSTR_1 in [_2_2,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (69,40 % of 134 examples)
      - c2d_NIVINSTR_1 in [_1_1,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (58,33 % of 12 examples)
    - c2d_PV_BA_1 in [_2_2,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (74,39 % of 246 examples)
  - c2d_PV_TA_1 in [_2_2,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (81,00 % of 1716 examples)

Figura 4.17. Prueba de Concepto - Patrones de Conocimiento

Esta fase se encuentra conformada por dos actividades: Evaluación de los Resultados (sección 4.4.6.1), donde se analiza la validez y utilidad de los patrones hallados, y Presentación de los Resultados (sección 4.4.6.2), garantizando la adecuada transmisión del conocimiento extraído para su utilización. La figura 4.18, presenta una visión resumida de las actividades que integran la fase y sus elementos de entrada y salida (las imágenes de cada formalismo son representaciones miniatura de los mismos).

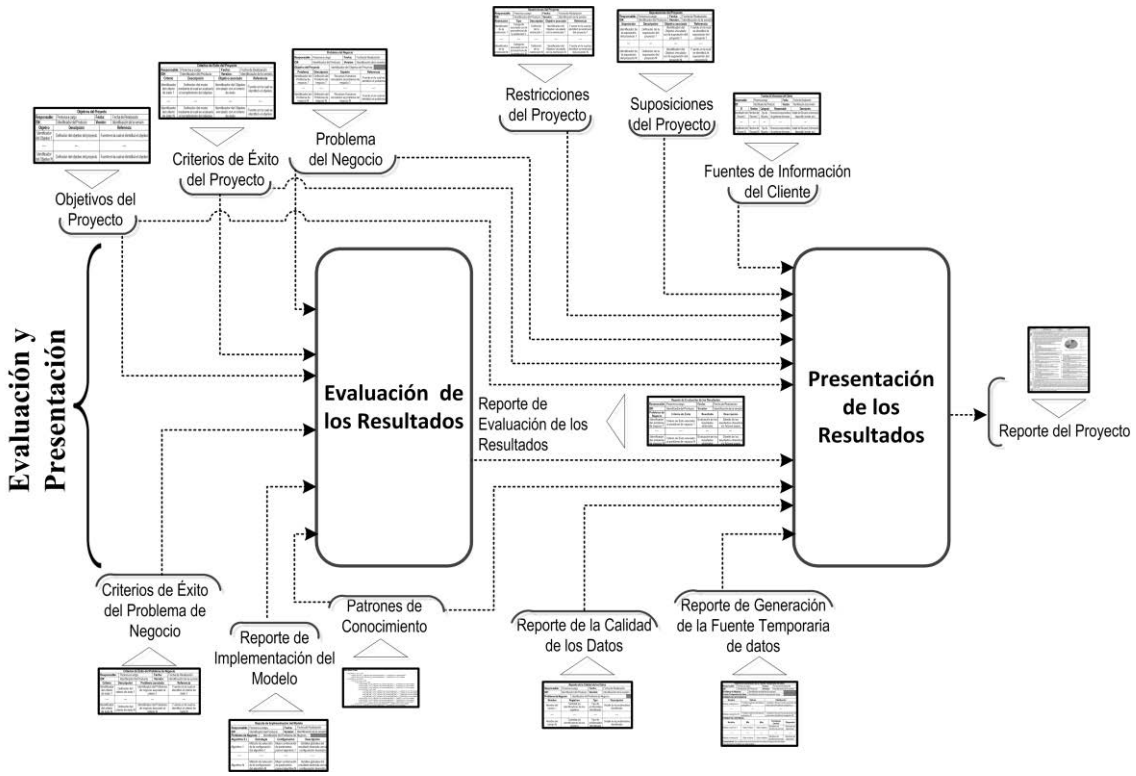


Figura 4.18. Fase: Evaluación y Presentación

4.4.6.1. Actividad: Evaluación de los Resultados (D.EP.EvR)

En esta actividad se evalúa la validez de los patrones de conocimiento obtenidos para el dominio de negocio y en particular para las problemática de negocio en cuestión. Como resultado de esta etapa debe corroborarse la satisfacción de las problemáticas del cliente o identificar la necesidad de refinar o resolver nuevas.

Información de Entrada

- Objetivos del Proyecto (D.EN.AnN.ObPr)
- Criterios de Éxito del Proyecto (D.EN.AnN.CrEP)
- Problema del Negocio (D.EN.CPN.PrNe)
- Criterios de Éxito del Problema de Negocio (D.EN.CPN.CEPN)
- Reporte de Implementación del Modelo (D.Im.ExI.ReIM)
- Patrones de Conocimiento (D.Im.ExI.PaCo)

Información de Salida

- Reporte de Evaluación de los Resultados (D.EP.EvR.ReER)

4.4.6.1.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se propone el Reporte de Evaluación de los Resultados, el cual se presenta a continuación.

Reporte de Evaluación de los Resultados (D.EP.EvR.ReER): Se formaliza la validación de los resultados, con respecto a su representatividad, validez y novedad para los expertos del negocio. Para ello, se indican el problema de negocio y los criterios de éxito asociado al mismo (en las columnas homónimas), se describe la valoración de los resultados obtenidos con respecto al cumplimiento o no de las necesidades y criterios del problema de negocio (en la columna "resultado"), y se detalla la interpretación de los resultados obtenidos y las futuras tareas a realizar (en caso que se identifiquen) en la columna "Descripción". La tabla 4.90 ilustra la estructura del formalismo previamente descripto.

Reporte de Evaluación de los Resultados			
Responsable:	Persona a cargo	**Fecha:**	Fecha de Realización
ID#:	Identificador del Producto	**Versión:**	Identificación de la versión
Problema de Negocio	**Criterio de Éxito**	**Resultado**	**Descripción**
Identificador del problema de negocio 1	Criterio de Éxito asociado al problema de negocio 1	Evaluación de los resultados obtenidos	Detalle de los resultados obtenidos y/o futuros pasos
…	…	…	…
Identificador del problema de negocio N	Criterio de Éxito asociado al problema de negocio N	Evaluación de los resultados obtenidos	Detalle de los resultados obtenidos y/o futuros pasos

Tabla 4.90. Formalismo: Reporte de Evaluación de los Resultados

4.4.6.1.2. Técnica Identificada

Para el desarrollo de esta actividad se propone la técnica **"Validación del Conocimiento"**, evaluando a partir de los problemas de negocio, los objetivos del proyecto y los criterios de éxitos respectivos, junto con la opinión de los expertos, la calidad y validez de los resultados obtenidos, así como la necesidad de refinar el procedimiento aplicado o la búsqueda de nuevos patrones que respondan a nuevas necesidades del cliente. En esta actividad, puede requerirse el desarrollo de procedimientos que permitan visualizar a los expertos del negocio (los cuales pueden no tener conocimientos sobre ingeniería de explotación de información) la validez de los resultados obtenidos. Como resultado de su implementación, deben quedar formalizados los siguientes pasos a realizar en el proyecto, es decir, si se cumplimentó con todos sus requerimientos (siendo el resultado *satisfactorio*), si es necesario ampliar los estudios realizados (siendo el resultado *ampliatorio*), reconsiderando nuevas hipótesis o modificando las configuraciones de los modelos o si los mismos no fueron cumplimentados (siendo el resultado *insatisfactorio*).

En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo para la prueba de concepto.

4.4.6.1.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica Validación del Conocimiento, la cual utiliza como insumos los formalismos: objetivos del proyecto (Tabla 4.57), criterios de éxito del proyecto (Tabla 4.58), problema del negocio (Tabla 4.64), criterios de éxito del problema de negocio (Tabla 4.65), Reporte de Implementación del Modelo (Tabla 4.89) y patrones de conocimiento (Figura 4.17), los cuales son transcriptos con el mismo número de referencia, para facilitar al lector en la comprensión de la aplicación de la técnica.

Objetivos del Proyecto			
Responsable:	Esposito E.	**Fecha:**	15/04/2016
ID#:	D.EN.AnN.ObPr	**Versión:**	1.0
Objetivo	**Descripción**		**Referencia**
obpr.1	Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos		Entrevista 1

Tabla 4.57 (Transcripta). Prueba de Concepto - Objetivos del Proyecto

Criterios de Éxito del Proyecto				
Responsable:	Esposito E.	**Fecha:**	15/04/2016	
ID#:	D.EN.AnN.CrEP	**Versión:**	1.0	
Criterio	**Descripción**		**Objetivo asociado**	**Referencia**
crexpr.1	obtener piezas de conocimiento que favorezcan la comprensión del comportamiento de grupos masivos de personas, brindando indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales serán validadas por el cliente		(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1

Tabla 4.58 (Transcripta). Prueba de Concepto - Criterios de Éxito del Proyecto

Problema del Negocio				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.PRNE	**Versión:**	1.0	
Objetivo del Proyecto	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos			
Problema	**Descripción**		**Experto**	**Referencia**
prne.1	Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		(rehi.3) Silva H.	Entrevista 3

Tabla 4.64 (Transcripta). Prueba de Concepto - Problema del Negocio

Reporte de Implementación del Modelo			
Responsable:	Esposito E.	**Fecha:**	27/05/2016
ID#:	D.Im.Exl.ReIM	**Versión:**	1.0
Problema de Negocio:	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		

Algoritmo E.I.	Estrategia	Configuración	Descripción
C4.5	(repm.1) RadomSearch	Min size of leaves: 150 Confidence Level: 0,25	Número de nodos: 20 Número de Hojas: 14 Tasa de error (promedio): 0.2558
ID3	(repm.2) RadomSearch	min size for split: 50 min size of leaves: 50 max depth of the tree: 10 Min entropy gain for splitting: 0.03	Número de nodos: 34 Número de Hojas: 28 Tasa de error (promedio): 0.2697

Tabla 4.89 (Transcripta). Prueba de Concepto - Reporte de Implementación del Modelo

Results

Classifier performances

Error rate	0,2558

Values prediction

Value	Recall	1-Precision
_2_1,00	0,7251	0,2437
_1_0,00	0,7635	0,2671

Confusion matrix

	_2_1,00	_1_0,00	Sum
_2_1,00	2353	892	3245
_1_0,00	758	2447	3205
Sum	3111	3339	6450

Tree description

Number of nodes	20
Number of leaves	14

Decision tree

- c2d_BIAC01_1 in [_1_1,00]
  - c2d_PV_BA_1 in [_1_1,00] then c2d_consumo_ilegal_bool_1 = **_2_1,00** (79,91 % of 2459 examples)
  - c2d_PV_BA_1 in [_2_2,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (67,38 % of 187 examples)
- c2d_BIAC01_1 in [_2_2,00]
  - c2d_PV_TA_1 in [_1_1,00]
    - c2d_PV_BA_1 in [_1_1,00]
      - c2d_NIVINSTR_1 in [_6_6,00] then c2d_consumo_ilegal_bool_1 = **_2_1,00** (58,62 % of 261 examples)
      - c2d_NIVINSTR_1 in [_7_7,00] then c2d_consumo_ilegal_bool_1 = **_2_1,00** (59,12 % of 318 examples)
      - c2d_NIVINSTR_1 in [_4_4,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (57,52 % of 339 examples)
      - c2d_NIVINSTR_1 in [_5_5,00]
        - c2d_POB_URB_1 in [_4_4,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (61,54 % of 195 examples)
        - c2d_POB_URB_1 in [_2_2,00] then c2d_consumo_ilegal_bool_1 = **_2_1,00** (61,22 % of 49 examples)
        - c2d_POB_URB_1 in [_3_3,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (55,61 % of 196 examples)
        - c2d_POB_URB_1 in [_1_1,00] then c2d_consumo_ilegal_bool_1 = **_2_1,00** (70,83 % of 24 examples)
      - c2d_NIVINSTR_1 in [_3_3,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (71,34 % of 314 examples)
      - c2d_NIVINSTR_1 in [_2_2,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (69,40 % of 134 examples)
      - c2d_NIVINSTR_1 in [_1_1,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (58,33 % of 12 examples)
    - c2d_PV_BA_1 in [_2_2,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (74,39 % of 246 examples)
  - c2d_PV_TA_1 in [_2_2,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (81,00 % of 1716 examples)

Figura 4.17 (Transcripta). Prueba de Concepto - Patrones de Conocimiento

Criterios de Éxito del Problema de Negocio				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.CEPN	**Versión:**	1.0	
Criterio	**Descripción**	**Problema asociado**		**Referencia**
cepn.1	Identificar aspectos que permitan comprender el comportamiento de grupos masivos de personas de acuerdo a la evaluación de Silva H. (rehi.3)	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		Entrevista 3

Tabla 4.65 (Transcripta). Prueba de Concepto - Criterios de Éxito del Problema de Negocio

Reporte de Evaluación de los Resultados (D.EP.EvR.ReER): De acuerdo a los patrones de conocimiento identificados, se evaluó de forma conjunta con el experto del problema de negocio (Silva H.), la validez e interés de los resultados, determinando que los mismos satisficieron las necesidades cubiertas por la pregunta-problema y los criterios de éxito vinculados. En adición, el experto señaló: "Las reglas identificadas permiten comprender los aspectos generales de la población estudiada. Se señala que los resultados obtenidos pueden estar dispersos por el consumo de algunos tipos de sustancias psicoactivas que se encuentran con menor presencia en la población encuestada, siendo de interés profundizar en el estudio del comportamiento de la población mediante un análisis geo-referencial.". Si bien fueron identificadas nuevas necesidades de investigación, se acordó que las mismas serán desarrolladas en un proyecto futuro. La tabla 4.91 ilustra el resultado obtenido.

Reporte de Evaluación de los Resultados			
Responsable:	Rodriguez H.	**Fecha:**	02/06/2016
ID#:	D.EP.EvR.ReER	**Versión:**	1.0
Problema de Negocio	**Criterio de Éxito**	**Resultado**	**Descripción**
(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo	(cepn.1) Identificar aspectos que permitan comprender el comportamiento de grupos masivos de personas de acuerdo a la evaluación de Silva H. (rehi.3)	Satisfactorio	Las reglas identificadas permiten comprender los aspectos generales de la población estudiada. Se señala que los resultados obtenidos pueden estar dispersos por el consumo de algunos tipos de sustancias psicoactivas que se encuentran con menor presencia en la población encuestada, siendo de interés profundizar en el estudio del comportamiento de la población mediante un análisis geo-referencial.

Tabla 4.91. Prueba de Concepto - Reporte de Evaluación de los Resultados

4.4.6.2. Actividad: Presentación de los Resultados (D.EP.PrR)

En esta actividad se llevan a cabo las tareas finales del proyecto, con respecto a la documentación los resultados obtenidos y la presentación de los mismos a los interesados. El objetivo de esta

actividad es la correcta transmisión de los patrones obtenidos y el conocimiento extraído para dar soporte al proceso decisorio del cliente.

Información de Entrada

- Fuentes de Información del Cliente (D.EN.AnN.FuIC)
- Objetivos del Proyecto (D.EN.AnN.ObPr)
- Criterios de Éxito del Proyecto (D.EN.AnN.CrEP)
- Suposiciones del Proyecto (D.EN.AnN.SuPr)
- Restricciones del Proyecto (D.EN.AnN.RePr)
- Problema del Negocio (D.EN.CPN.PrNe)
- Patrones de Conocimiento (D.Im.ExI.PaCo)
- Reporte de Generación de la Fuente Temporaria de datos (D.PD.CFT.RGFT)
- Reporte de la Calidad de los Datos (D.ED.EvD.ReCD)
- Reporte de Evaluación de los Resultados (D.EP.EvR.ReER)

Información de Salida

- Reporte del Proyecto (D.EP.PrR.RepP)

4.4.6.2.1. Formalismos Identificados

Para la formalización de la información de salida esperada para el desarrollo de la actividad, se propone el Reporte del Proyecto, el cual se presenta a continuación.

Reporte del Proyecto (D.EP.PrR.RepP): Se formaliza el conocimiento extraído a partir de un reporte enfocado en brindar a los expertos del dominio del negocio, de la información de soporte para el proceso decisorio. Para ello, se identifican cinco secciones generales que cubren los siguientes temas principales del proyecto: descripción del dominio y de las problemáticas definidas, de la información disponible y utilizada, de los resultados obtenidos, interpretación de los resultados y recomendaciones. En adición, dichas secciones pueden incluir visualizaciones para favorecer la comprensión de los resultados presentados. La tabla 4.92 ilustra la estructura del formalismo previamente descripto.

Reporte del Proyecto			
Responsable:	Rodriguez H.	**Fecha:**	10/06/2016
ID#:	D.EP.PrR.RepP	**Versión:**	1.0
DESCRIPCIÓN DEL PROBLEMA	Introducción del dominio del negocio y de los objetivos y problemáticas a abordar en el proyecto		
DESCRIPCIÓN DE LOS DATOS	Descripción de la información disponible y utilizada para el desarrollo del proyecto, así como cualquier acción relevante realizada en los datos para la interpretación de los resultados		
RESULTADOS DE EXPLOTACIÓN DE INFORMACIÓN	Descripción de los patrones de conocimientos obtenidos mediante la aplicación de las técnicas de ingeniería de explotación de información		
EVALUACIÓN DE LOS RESULTADOS	Descripción de los resultados obtenidos desde la perspectiva del negocio, indicando el significado de los mismos, así como su validez y precisión de los mismos.		
DIFICULTADES Y RECOMEN-DACIONES	Sugerencias de mejoras en el proceso de recolección de los datos, así como posibles objetivos o preguntas problemas de interés adicionales para resolver mediante ingeniería de explotación de información. Así como la descripción de dificultades identificadas durante el proceso.		

Tabla 4.92. Formalismo: Reporte del Proyecto

4.4.6.2.2. Técnica Identificada

Para el desarrollo de esta actividad se propone la técnica **"Síntesis del Proyecto"**, en la cual se analiza y resume la información más relevante del proceso de extracción de conocimiento, orientado a dar al experto del contexto y el procedimiento realizado para obtener las piezas de conocimiento a utilizar para dar soporte al proceso de toma de decisiones. En este contexto, la redacción del reporte debe estar pensada y ajustada de acuerdo a necesidades y características de los expertos del dominio del negocio (los cuales pueden no tener conocimientos técnicos sobre explotación de información), para que estos comprendan y utilicen las piezas de conocimientos extraídas, los resultados obtenidos en las distintas etapas del proyecto, así como cualquier elemento que pueda producirse para facilitar la interpretación de los mismos (por ejemplo: visualizaciones, casos de prueba, etc.).

Para la generación del informe se señalan cinco secciones generales, que cubren los aspectos más relevantes del proyecto (desde la perspectiva del dominio de negocio):

- **Descripción del problema**: se introduce las características del dominio y de las necesidades del cliente, detallando aquellos problemas de negocio que se abarcaron en el proyecto. Las limitaciones y asunciones que se hayan realizado deberán ser incorporadas.

- **Descripción de los Datos**: se describen los datos utilizados, indicando en caso de considerarse necesario las tareas de recolección e integración realizadas, las tareas de limpieza, formateo e imputación de valores (en caso que corresponda), detallando la composición de la fuente de información utilizada para el análisis y la derivación de los resultados. Se recomienda la incorporación de visualizaciones para facilitar la lectura y comprensión de la información presentada.
- **Resultados de Explotación de Información**: se detallan aquellas piezas de conocimiento que sean consideradas relevantes, novedosas y de interés para los objetivos del proyecto. En esta sección se presentan los patrones obtenidos sin adentrarse en los detalles de la interpretación o evaluación de la calidad de los mismos, los cuales serán tratados en la siguiente sección.
- **Evaluación de los Resultados**: se presenta la interpretación de los resultados, así como los indicadores de calidad que se hayan evaluado durante el proyecto. Se describe las estimaciones de los resultados tanto desde la perspectiva cuantitativa (métricas de evaluación de performance del modelo), así como cualitativa (validación de los resultados por los expertos). Se recomienda la utilización de visualizaciones para facilitar la lectura y comprensión de la información presentada.
- **Dificultades y Recomendaciones**: Se describen aquellas dificultades que se hayan identificado en el desarrollo del proceso, así como la incorporación de recomendaciones de mejoras (como por ejemplo problemáticas identificadas en el almacenamiento de los datos del negocio por falta de consideración o inconvenientes en el proceso y/o las herramientas de carga de datos) y posibles futuros pasos a realizar.

En la siguiente sección se presenta la aplicación de la técnica y el registro del formalismo para la prueba de concepto.

4.4.6.2.3. Ejecución de la Actividad en la Prueba de Concepto

A continuación se presentan los resultados obtenidos de aplicar la técnica Síntesis del Proyecto, la cual utiliza como insumos los formalismos: Fuentes de Información del Cliente (Tabla 4.55), Objetivos del Proyecto (Tabla 4.57), Criterios de Éxito del Proyecto (Tabla 4.58), Suposiciones del Proyecto (Tabla 4.60), Restricciones del Proyecto (Tabla 4.61), Problema del Negocio (Tabla 4.64), Patrones de Conocimiento (Figura 4.17), Reporte de Generación de la Fuente Temporaria de datos (Tablas 4.83.a y 4.83.b), Reporte de la Calidad de los Datos (Tabla 4.73) y Reporte de Evaluación de los Resultados (Tabla 4.91), los cuales son transcriptos con el mismo número de referencia, para facilitar al lector en la comprensión de la aplicación de la técnica.

Fuentes de Información del Cliente					
Responsable:		Esposito E.		**Fecha:**	05/04/2016
ID#:		D.EN.ANN.FUIC		**Versión:**	1.0
ID	**Nombre**	**Categoría**	**Responsable**	**Descripción**	
fuic.1	Documento para la utilización de la base de datos usuario 2011	Documento	-	Describe distintas consideraciones acerca de la base de datos obtenida a partir de la encuesta ENPreCoSP 2011, indicando los métodos de recolección aplicados, los objetivos de la encuesta y la descripción de los datos.	
fuic.2	Cuestionario ENPreCoSP 2011	Planilla	-	Ejemplo de cuestionario ENPreCoSP 2011	
fuic.3	Base ENPreCoSP 2011	Almacén de datos	-	Almacén de registros de respuestas del cuestionario ENPreCoSP 2011 en formato txt (separado por el carácter "|") conformado por 34343 personas que respondieron 292 preguntas. El primer renglón contiene los nombres de los campos. Decimales separados por punto (.)	

Tabla 4.45 (Transcripta). Prueba de Concepto - Fuentes de Información del Cliente

Objetivos del Proyecto				
Responsable:	Esposito E.	**Fecha:**	15/04/2016	
ID#:	D.EN.ANN.OBPR	**Versión:**	1.0	
Objetivo	**Descripción**			**Referencia**
obpr.1	Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos			Entrevista 1

Tabla 4.57 (Transcripta). Prueba de Concepto - Objetivos del Proyecto

Criterios de Éxito del Proyecto				
Responsable:	Esposito E.	**Fecha:**	15/04/2016	
ID#:	D.EN.ANN.CREP	**Versión:**	1.0	
Criterio	**Descripción**	**Objetivo asociado**		**Referencia**
crexpr.1	obtener piezas de conocimiento que favorezcan la comprensión del comportamiento de grupos masivos de personas, brindando indicadores para la implementación de políticas públicas más eficaces, orientadas a mejorar las condiciones de salud de la población las cuales serán validadas por el cliente	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos		Entrevista 1

Tabla 4.58 (Transcripta). Prueba de Concepto - Criterios de Éxito del Proyecto

Restricciones del Proyecto					
Responsable:		Esposito E.	**Fecha:**	20/04/2016	
ID#:		D.EN.AnN.RePr	**Versión:**	1.0	
Restricción	**Tipo**	**Descripción**	**Objetivo asociado**		**Referencia**
repr.1	datos	Se identifica un desbalance entre la cantidad de registros que han consumido distintas Sustancias Psicoactivas y quienes no han consumido	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos		Entrevista 1
repr.2	datos	Se carece de información detallada respecto al grado o frecuencia de consumo de las sustancias psicoactivas	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos		Entrevista 1

Tabla 4.61 (Transcripta). Prueba de Concepto - Restricciones del Proyecto

Suposiciones del Proyecto			
Responsable:	Esposito E.	**Fecha:**	20/04/2016
ID#:	D.EN.AnN.SuPr	**Versión:**	1.1

Suposición	Descripción	Objetivo asociado	Referencia
supr.1	Los cuestionarios y la carga de la información se ha realizado de manera correcta	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1
supr.2	Las conductas de consumo se considerarán como análogas sin importar la gradualidad del mismo	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 1
supr.3	Las variables vinculadas con autopercepción brindan información fiable respecto al entorno real del individuo	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 2
supr.4	El proceso de diseño de la muestra es representativo a nivel nacional y provincial	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos	Entrevista 2 / fuic.1

Tabla 4.60 (Transcripta). Prueba de Concepto - Suposiciones del Proyecto

Results

Classifier performances

Error rate	0,2558
Values prediction	Confusion matrix

Value	Recall	1-Precision
_2_1,00	0,7251	0,2437
_1_0,00	0,7635	0,2671

	_2_1,00	_1_0,00	Sum
_2_1,00	2353	892	3245
_1_0,00	758	2447	3205
Sum	3111	3339	6450

Tree description

Number of nodes	20
Number of leaves	14

Decision tree

- c2d_BIAC01_1 in [_1_1,00]
  - c2d_PV_BA_1 in [_1_1,00] then c2d_consumo_ilegal_bool_1 = **_2_1,00** (79,91 % of 2459 examples)
  - c2d_PV_BA_1 in [_2_2,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (67,38 % of 187 examples)
- c2d_BIAC01_1 in [_2_2,00]
  - c2d_PV_TA_1 in [_1_1,00]
    - c2d_PV_BA_1 in [_1_1,00]
      - c2d_NIVINSTR_1 in [_6_6,00] then c2d_consumo_ilegal_bool_1 = **_2_1,00** (58,62 % of 261 examples)
      - c2d_NIVINSTR_1 in [_7_7,00] then c2d_consumo_ilegal_bool_1 = **_2_1,00** (59,12 % of 318 examples)
      - c2d_NIVINSTR_1 in [_4_4,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (57,52 % of 339 examples)
      - c2d_NIVINSTR_1 in [_5_5,00]
        - c2d_POB_URB_1 in [_4_4,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (61,54 % of 195 examples)
        - c2d_POB_URB_1 in [_2_2,00] then c2d_consumo_ilegal_bool_1 = **_2_1,00** (61,22 % of 49 examples)
        - c2d_POB_URB_1 in [_3_3,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (55,61 % of 196 examples)
        - c2d_POB_URB_1 in [_1_1,00] then c2d_consumo_ilegal_bool_1 = **_2_1,00** (70,83 % of 24 examples)
      - c2d_NIVINSTR_1 in [_3_3,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (71,34 % of 314 examples)
      - c2d_NIVINSTR_1 in [_2_2,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (69,40 % of 134 examples)
      - c2d_NIVINSTR_1 in [_1_1,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (58,33 % of 12 examples)
    - c2d_PV_BA_1 in [_2_2,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (74,39 % of 246 examples)
  - c2d_PV_TA_1 in [_2_2,00] then c2d_consumo_ilegal_bool_1 = **_1_0,00** (81,00 % of 1716 examples)

Figura 4.17 (Transcripta). Prueba de Concepto - Patrones de Conocimiento

Problema del Negocio				
Responsable:	Esposito E.	**Fecha:**	20/04/2016	
ID#:	D.EN.CPN.PRNE	**Versión:**	1.0	
Objetivo del Proyecto	(obpr.1) Comprender el comportamiento de la población encuestada, siendo posible la comparación del mismo con otros periodos			
Problema	**Descripción**		**Experto**	**Referencia**
prne.1	Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		(rehi.3) Silva H.	Entrevista 3

Tabla 4.64 (Transcripta). Prueba de Concepto - Problema del Negocio

Reporte de Generación de la Fuente Temporaria de datos			
Responsable:	Esposito E.	**Fecha:**	18/05/16
ID#:	D.PD.CFT.RGFT	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		
Fuente Temporaria de datos	(FTD.1) Fuente Temporaria de datos		
ATRIBUTOS CUALITATIVOS			
Nombre	**Valores**	**Distribuciones**	
POB_URB	1 Más de 1.500.000 habitantes 2 De 500.001 a 1.500.000 habitantes 3 De 100.001 a 500.000 habitantes 4 De 5.000 a 100.000 habitantes	396 (6,14%) 817 (12,67%) 2605 (40,39%) 2632 (40,81%)	
RANGOING	0 Sin ingresos 1 1 a 600 2 601 a 800 3 801 a 1.000 4 1.001 a 1.500 5 1.501 a 2.000 6 2.001 a 2.500 7 2.501 a 3.000 8 3.001 a 3.500 9 3.501 a 4.000 10 4.001 a 4.500 11 4.501 a 5.500 12 5.001 a 6.000 13 6.001 a 7.000 14 7.001 a 8.000 15 8.001 a 10.000 16 10.001 a 15.000 17 15.001 y más 99 Ns/nc	19 (0,29%) 191 (2,96%) 177 (2,74%) 272 (4,22%) 565 (8,76%) 742 (11,5%) 545 (8,45%) 697 (10,81%) 379 (5,88%) 571 (8,85%) 207 (3,21%) 486 (7,53%) 378 (5,86%) 272 (4,22%) 266 (4,12%) 310 (4,81%) 241 (3,74%) 132 (2,05%) 0 (0%)	
NBI_TOTAL	0 Ningún indicador de NBI 1 Al menos un indicador de NBI 2 Al menos dos indicadores de NBI 3 Al menos tres indicadores de NBI 4 Al menos cuatro indicadores de NBI	5644 (87,5%) 621 (9,63%) 156 (2,42%) 29 (0,45%) 0 (0%)	
BHCH04	1 Varón 2 Mujer	3183 (49,35%) 3267 (50,65%)	

Tabla 4.83.a (Transcripta). Prueba de Concepto - Reporte de Generación de la Fuente Temporaria de datos

NIVINSTR	1 Sin instrucción 2 Primario incompleto 3 Primario completo 4 Secundario incompleto 5 Secundario completo 6 Terciario o universitario incompleto 7 Terciario o universitario completo y más 8 Educación especial	54 (0,84%) 385 (5,97%) 969 (15,02%) 1352 (20,96%) 1509 (23,4%) 1072 (16,62%) 1109 (17,19%) 0 (0%)
BIAC01	1 Sí 2 No 9 Ns/nc	2646 (41,02%) 3804 (58,98%) 0 (0%)
PV_TA	1 Sí 0 No	4114 (63,78%) 2336 (36,22%)
PV_BA	1 Sí 0 No	5359 (83,09%) 1091 (16,91%)
consumo_ilegal_bool	1 Sí 0 No	3245 (50,31%) 3205 (49,69%)
GRUPEDAD	2 16 a 24 años 3 25 a 34 años 4 35 a 49 años 5 50 a 65 años	1392 (21,58%) 1971 (30,56%) 1258 (19,50%) 1829 (28,36%)
Comentarios: número de semilla: 1802, atributo de estratificación: consumo_ilegal_bool, definición del tamaño de la muestra por valor absoluto: 6450		

Tabla 4.83.b (Transcripta). Prueba de Concepto - Reporte de Generación de la Fuente Temporaria de datos

Reporte de la Calidad de los Datos			
Responsable:	Esposito E.	**Fecha:**	06/05/2016
ID#:	D.ED.EvD.ReCD	**Versión:**	1.0
Problema de Negocio	(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo		

Nombre	**Registros**	**Tipo**	**Descripción**
RANGOING	2794	nulos	Valor 99. Ingresos no informados
NIVINSTR	43	Outlier	Valor 8: Minoría no representativa para el problema de negocio
BIAC01	97	nulos	Valor 9. no informado

Tabla 4.73 (Transcripta). Prueba de Concepto - Reporte de la Calidad de los Datos

Reporte de Evaluación de los Resultados			
Responsable:	Rodriguez H.	**Fecha:**	02/06/2016
ID#:	D.EP.EvR.ReER	**Versión:**	1.0
Problema de Negocio	**Criterio de Éxito**	**Resultado**	**Descripción**
(prne.1) Cuáles son las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social (considerando la autopercepción) de las personas de 16 a 65 años que han consumido sustancias psicoactivas ilegales, sin contemplar las graduaciones del consumo	(cepn.1) Identificar aspectos que permitan comprender el comportamiento de grupos masivos de personas de acuerdo a la evaluación de Silva H. (rehi.3)	Satisfactorio	Las reglas identificadas permiten comprender los aspectos generales de la población estudiada. Se señala que los resultados obtenidos pueden estar dispersos por el consumo de algunos tipos de sustancias psicoactivas que se encuentran con menor presencia en la población encuestada, siendo de interés profundizar en el estudio del comportamiento de la población mediante un análisis geo-referencial.

Tabla 4.91 (Transcripta). Prueba de Concepto - Reporte de Evaluación de los Resultados

Reporte del Proyecto (D.EP.PrR.RepP): En primera instancia a partir de los reportes generados en la fase de entendimiento del negocio (aquellos vinculados con los objetivos del proyecto y los problema de negocio), se completó la sección de descripción del problema, en la cual se resume los objetivos generales e individuales del proyecto, alcances y recursos utilizados. Luego, a partir de los formalismos de identificación y descripción de las fuentes de datos, se describe en la segunda sección, las fuentes de datos utilizadas, el proceso de preparación de los datos realizados y los atributos que se utilizaron en el modelo. En tercera instancia se evalúan los patrones de conocimiento obtenidos y se presentan los resultados obtenidos de mayor relevancia. Posteriormente, haciendo uso de los criterios de éxito, las evaluaciones del experto y de los resultados de las métricas asociadas a la evaluación de los algoritmos, se interpretan los resultados obtenidos en la sección "Evaluación de los Resultados". Finalmente, se identifica a partir de la evaluación de los resultados obtenidos, recomendaciones sobre futuros aspectos a evaluar en nuevos proyectos. Las tablas 4.93.a y 4.93.b ilustran el resultado obtenido para la prueba de concepto.

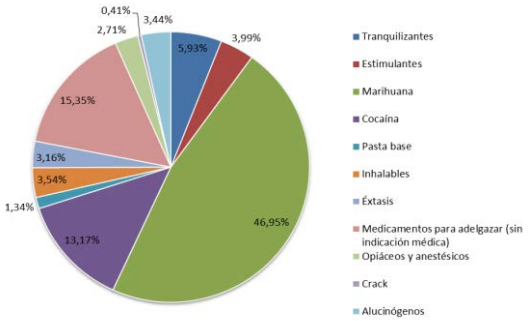
<table>
<tr><th colspan="4">Reporte del Proyecto</th></tr>
<tr><td>Responsable:</td><td>Rodriguez H.</td><td>Fecha:</td><td>10/06/2016</td></tr>
<tr><td>ID#:</td><td>D.EP.PrR.RepP</td><td>Versión:</td><td>1.0</td></tr>
<tr><td>DESCRIPCIÓN DEL PROBLEMA</td><td colspan="3">El Ministerio de Salud de la Nación, a través de la Dirección de Salud Mental y Adicciones, en conjunto con el Instituto Nacional de Estadística y Censos y con la colaboración de las Direcciones Provinciales de Estadística, llevó a cabo la Encuesta Nacional sobre Prevalencias de Consumo de Sustancias Psicoactivas 2011 (ENPreCoSP-2011).
En este contexto, se requirió analizar características sociodemográficas, socioeconómicas, educativas y del entorno familiar social de la población de 16 a 65 años de edad que consume sustancias psicoactivas ilegales, comprendiendo el comportamiento de la población encuestada y permitiendo evaluar el mismo con respecto a otros periodos. Identificando como objetivo del estudio: comprender las condiciones sociodemográficas, socioeconómicas, educativas y del entorno familiar social de las personas que han consumido sustancias psicoactivas ilegales. Quedando excluidas de la variable de estudio aquellas drogas socialmente aceptadas (cigarrillo y alcohol).</td></tr>
<tr><td>DESCRIPCIÓN DE LOS DATOS</td><td colspan="3">El análisis inicial fue realizado sobre 31426 individuos (población total encuestada exceptuando personas que no hayan indicado su ingreso o nivel de estudio, así como una minoría con educación especial), aplicando un muestreo estratificado sobre el consumo de sustancias psicoactivas ilegales. Para ello se tuvieron en cuenta las siguientes variables:
▪ Agrupamiento de poblaciones urbanas
▪ Rango del ingreso total mensual del hogar en pesos
▪ Indicadores de necesidades básicas insatisfechas de hogar: NBI total
▪ Sexo
▪ Grupo de edad
▪ Nivel de instrucción
▪ ¿Conoce personas cercanas a usted que en la actualidad consuman alguna sustancia como marihuana, cocaína, éxtasis, etc.?
▪ Prevalencia de vida de consumo de tabaco
▪ Prevalencia de vida de consumo de bebidas alcohólicas
▪ Prevalencia de vida de consumo de alguna de las sustancias psicoactivas ilegales
▪ La distribución de casos que presentan prevalencia de vida en consumo de sustancias psicoactivas ilegales puede verse en la figura 1.
 

Figura 1. Distribución de consumo de drogas ilegales</td></tr>
</table>

Tabla 4.93.a Prueba de Concepto - Reporte del Proyecto

RESULTADOS DE EXPLOTACIÓN DE INFORMACIÓN	**SI** Conoce personas cercanas que en la actualidad consuman alguna sustancia **Y** posee Prevalencia de vida de consumo de bebidas alcohólicas **ENTONCES** presenta prevalencia de vida de consumo de sustancias psicoactivas ilegales (79,91% de 2459) **SI** Conoce personas cercanas que en la actualidad consuman alguna sustancia **Y** NO posee Prevalencia de vida de consumo de bebidas alcohólicas **ENTONCES** NO presenta prevalencia de vida de consumo de sustancias psicoactivas ilegales (67.38% de 187) **SI** NO conoce personas cercanas que en la actualidad consuman alguna sustancia **Y** posee Prevalencia de vida de consumo de tabaco **Y** posee Prevalencia de vida de consumo de bebidas alcohólicas **Y** s u nivel de instrucción es Terciario o universitario incompleto o superior **ENTONCES** presenta prevalencia de vida de consumo de sustancias psicoactivas ilegales (58,89% de 579) **SI** NO conoce personas cercanas que en la actualidad consuman alguna sustancia **Y** posee Prevalencia de vida de consumo de tabaco **Y** posee Prevalencia de vida de consumo de bebidas alcohólicas **Y** su nivel de instrucción es Secundario completo Y reside en un agrupamiento de poblaciones urbanas de 500.001 habitantes o más **ENTONCES** presenta prevalencia de vida de consumo de sustancias psicoactivas ilegales (64,38% de 73) **SI** NO conoce personas cercanas que en la actualidad consuman alguna sustancia **Y** posee Prevalencia de vida de consumo de tabaco **Y** posee Prevalencia de vida de consumo de bebidas alcohólicas **Y** su nivel de instrucción es Secundario completo **Y** reside en un agrupamiento de poblaciones urbanas de 500.000 habitantes o menos **ENTONCES** NO presenta prevalencia de vida de consumo de sustancias psicoactivas ilegales (58,57% de 391) **SI** NO conoce personas cercanas que en la actualidad consuman alguna sustancia **Y** posee Prevalencia de vida de consumo de tabaco **Y** posee Prevalencia de vida de consumo de bebidas alcohólicas **Y** su nivel de instrucción es Secundario incompleto o inferior **ENTONCES** NO presenta prevalencia de vida de consumo de sustancias psicoactivas ilegales (64,96% de 799) **SI** NO conoce personas cercanas que en la actualidad consuman alguna sustancia **Y** posee Prevalencia de vida de consumo de tabaco **Y** NO posee Prevalencia de vida de consumo de bebidas alcohólicas **ENTONCES** NO presenta prevalencia de vida de consumo de sustancias psicoactivas ilegales (74.39% de 246) **SI** NO conoce personas cercanas que en la actualidad consuman alguna sustancia **Y** **NO** posee Prevalencia de vida de consumo de tabaco **ENTONCES** NO presenta prevalencia de vida de consumo de sustancias psicoactivas ilegales (81% de 1716)
EVALUACIÓN DE LOS RESULTADOS	Como conclusiones del análisis realizado, se destaca el impacto del entorno, así como el consumo de las drogas legales o socialmente aceptadas en el consumo de sustancias psicoactivas, caracterizando el comportamiento de una población mayoritaria que consume este tipo de sustancias de aproximadamente el 75%. En aquellos casos en el cual el individuo no posee en su entorno personas con acceso a las sustancias psicoactivas ilegales, aparecen nuevas características que describen el comportamiento de dicha población de forma conjunta al consumo de las drogas legales. Estos son el nivel de instrucción y número de habitantes. Identificando el consumo en la población cuyos valores son los más elevados para ambas variables. Es importante tener en consideración al momento de interpretar los resultados la distribución de los casos de consumo de distintas sustancias, viéndose los resultados ponderados por aquellas drogas con mayor cantidad de consumidores (especialmente marihuana con una presencia casi del 50%), donde a interpretación del experto dicho factor se ve reflejado particularmente en las reglas donde la persona que presenta prevalencia de vida en el consumo de sustancias posee estudios universitario o superior, así como reside en lugares con los más alto números de habitantes en la escala presente.
DIFICULTADES Y RECOMENDACIONES	Se indica como posibles trabajos a realizar, el análisis y comparación del cambio en el comportamiento de las poblaciones en el transcurso del tiempo, así como el análisis geo referenciado de los individuos.

Tabla 4.93.b Prueba de Concepto - Reporte del Proyecto

En síntesis, en este capítulo se introdujo el Modelo de Proceso para Proyectos de Explotación de Información (MoProPEI), cuya estructura completa (subprocesos, fases, actividades, objetivos, técnicas, elementos de entrada y salida) se resume en las tablas 4.94 y 4.95. De forma conjunta, se presentó el primero de los casos de validación (perteneciente al área de la salud), realizándose un análisis respecto a los patrones de consumo de sustancias psicoactivas a partir de la información disponible en la bases de datos del INDEC [Instituto Nacional de Estadística y Censos, 2016]. Como resultado del proyecto, se obtuvieron patrones de interés que ayudaron a comprender y describir las características predominantes del consumo de psicoactivas de la muestra de estudio.

FASE	ACTIVI-DAD	ENTRADA	TÉCNICA	SALIDA	OBJETIVO
Iniciación	Exploración Inicial del Proyecto	• Fuentes de Información del Cliente • Objetivos del Proyecto • Criterios de Éxito del Proyecto • Expectativas del Proyecto • Restricciones del Proyecto • Problema del Negocio • Criterios de Éxito del Problema de Negocio	Caracterización del desarrollo del proyecto perteneciente a la Metodología para la educción de requerimientos para proyectos de explotación de información [Britos et al., 2008]	• Recursos Humanos Involucrados • Riesgos del Proyecto • Plan de Contingencias	Se identifican los miembros de interés para el proyecto y las posibles situaciones de riesgo
	Definición de la Comunicación	• Recursos Humanos Involucrados	Definición de la Comunicación [Verzuh, 2015]	• Plan de Comunicación	Se definen las necesidades y canales de comunicación durante el desarrollo del proyecto
	Evaluación de la Situación	• Objetivos del Proyecto • Problema del Negocio • Fuentes de Información del Cliente • Recursos Humanos Involucrados • Expectativas del Proyecto • Suposiciones del Proyecto	Metodología para la selección de Herramientas de Explotación de Información [Britos et al., 2006] y Modelo de Evaluación de Viabilidad para Proyectos de Explotación de Información [Pytel et al., 2015]	• Reporte de Evaluación de Herramientas • Reporte de Evaluación de Viabilidad	Se analizan las herramientas de utilidad para el desarrollo del proyecto, determinando la viabilidad del mismo
	Definición del Ciclo de Vida	• Objetivos del Proyecto • Expectativas del Proyecto • Problema del Negocio • Riesgos del Proyecto • Recursos Humanos Involucrados	Selección del Ciclo de Vida	• Modelo de Ciclo de Vida	Se establece de acuerdo a las características del proyecto, el flujo mediante el cual se llevarán a cabo las tareas de desarrollo
Planificación	Planificación de la Mediciones	• Objetivos del Proyecto • Problema del Negocio • Fuentes de Información del Cliente • Recursos Humanos Involucrados • Reporte de Evaluación de Herramientas	Métricas para Proyectos de Explotación de Información [Basso et al., 2013] y Modelo de Estimación para Proyectos de Explotación de Información [Pytel et al., 2015]	• Listado de Métricas • Estimación del Proyecto	Se definen las mediciones que se llevarán a cabo durante el proyecto, realizando una estimación inicial del esfuerzo requerido
	Planificación de las Actividades	• Modelo de Ciclo de Vida • Estimación del Proyecto • Objetivos del Proyecto • Problema del Negocio • Fuentes de Información del Cliente	Definición del Programa del Proyecto	• Mapa de Actividades • Plan de Acción	Se definen las tareas a realizar y sus alcances en el transcurso del tiempo para el desarrollo del proyecto
	Planificación de los Recursos	• Recursos Humanos Involucrados • Reporte de Evaluación de Herramientas • Plan de Acción • Problema del Negocio • Fuentes de Información del Cliente	Planificación de los Recursos Necesarios	• Plan de Necesidad de Recursos	Se prevén los recursos (humanos y materiales) necesarios para el desarrollo de las actividades en el tiempo
	Planificación de las Responsabilidades	• Recursos Humanos Involucrados • Plan de Comunicación • Plan de Acción • Plan de Necesidad de Recursos • Objetivos del Proyecto • Criterios de Éxito del Proyecto • Expectativas del Proyecto • Restricciones del Proyecto • Problema del Negocio • Criterios de Éxito del Problema de Negocio • Riesgos del Proyecto • Plan de Contingencias	Designación de Responsabilidades [Project Management Institute, Inc., 2013a]	• Matriz de Responsabilidades • Propuesta del Proyecto	Se deja registro formal de las responsabilidades y obligaciones de las partes involucradas
Soporte	Mediciones del Proyecto	• Listado de Métricas • Plan de Acción	Cálculo de Métricas	• Registro de Mediciones	Se calculan las métricas durante el desarrollo del proyecto
	Gestión de la Configuración	• Reglas de Versionado (Externo) • Reporte de Evaluación del Cambio • Modelo de Ciclo de Vida	Configuración del versionado	• Reporte de Versionado • Informe del Estado de la Configuración	Se mantiene registro de la evolución de los productos, garantizando su trazabilidad

Tabla 4.94.a. MoProPEI: Estructura subproceso Gestión

FASE	ACTIVI-DAD	ENTRADA	TÉCNICA	SALIDA	OBJETIVO
Control	Gestión del Desarrollo	• Plan de Acción • Registro de Mediciones	Seguimiento de Avance	• Reporte de Estado	Se evalúa el desarrollo del proyecto, verificando que el mismo se efectúe de acuerdo a lo planificado y pactado con el cliente, y dejando registro formal de cualquier desvió, cambio o posible evento riesgoso que aconteciese
	Control de las Actividades	• Riesgos del Proyecto • Plan de Contingencias • Plan de Acción • Registro de Mediciones • Reporte de Estado	Evaluación de Riesgos	• Registro de Riesgos Acontecidos	Se evalúan las situaciones potencialmente peligrosas para el desarrollo del proyecto, realizando un seguimiento, control y registro de acontecimientos, así como de las acciones realizadas
	Gestión del Cambio	• Solicitud de cambio (Externo) • Propuesta del Proyecto	Evaluación del Cambio	• Reporte de Evaluación del Cambio	Se realiza un proceso de evaluación formal de las peticiones de cambio, determinando como resultado la procedencia o no del mismo y sus efectos asociados
Cierre	Formalización Externa del Cierre del	• Reporte de Evaluación de los Resultados • Registro de Riesgos Acontecidos • Plan de Acción • Propuesta del Proyecto	Presentación de Conformidad	• Documento de Aceptación	Se obtiene la conformidad del cliente, respecto a los compromisos asumidos en la propuesta del proyecto, dejando registro formal de la finalización del mismo
	Formalización Interna del Cierre del Proyecto	• Plan de Acción • Matriz de Responsabilidades • Registro de Mediciones • Registro de Riesgos Acontecidos • Reporte de Evaluación del Cambio • Reporte de Evaluación de los Resultados • Documento de Aceptación	Evaluación del Proceso	• Reporte de Cierre	Se evalúa el desarrollo del proyecto, dejando registro de aquellos aspectos que sean de valor para proyectos futuros

Tabla 4.94.b. MoProPEI: Estructura subproceso Gestión

FASE	ACTIV-IDAD	ENTRADA	TÉCNICA	SALIDA	OBJETIVO
Entendimiento del Negocio	Análisis del Negocio	• Discursos de los interesados (externo) • Información de la Organización (externo) • Información del dominio del negocio (externo)	Definición de los objetivos del proyecto que forma parte de la Metodología para la educción de requerimientos para proyectos de explotación de información [Britos et al., 2008]	• Fuentes de Información del Cliente • Definiciones, Acrónimos y Abreviaciones • Objetivos del Proyecto • Criterios de Éxito del Proyecto • Expectativas del Proyecto • Suposiciones del Proyecto • Restricciones del Proyecto	Identificar y comprender las metas del proyecto, en base a las necesidades del requirente y los interesados
	Comprensión del Problema de Negocio	• Discursos de los interesados (externo) • Información de la Organización (externo) • Información del dominio del negocio (externo) • Definiciones, Acrónimos y Abreviaciones • Objetivos del Proyecto • Criterios de Éxito del Proyecto • Expectativas del Proyecto • Suposiciones del Proyecto • Restricciones del Proyecto	Definición de los Problema de Negocio que forma parte de la Metodología para la educción de requerimientos para proyectos de explotación de información [Britos et al., 2008]	• Problema del Negocio • Criterios de Éxito del Problema de Negocio	Se definen las problemáticas específicas del negocio a analizar

Tabla 4.95.a. MoProPEI: Estructura subproceso Desarrollo

FASE	ACTIV-IDAD	ENTRADA	TÉCNICA	SALIDA	OBJETIVO
Entendimiento de los Datos	Análisis de los Datos	• Discursos de los interesados (externo) • Diagrama de Base de datos (externo) • Fuentes de Información del Cliente • Restricciones del Proyecto • Problema del Negocio	Identificación de atributos relacionados con el Problema de Negocio definida en la Metodología para la educción de requerimientos para proyectos de explotación de información [Britos et al., 2008] y Diccionario de Datos	• Diccionario de Fuente de Datos • Campos Relacionados con el Problema de Negocio	Se evalúan las variables disponibles en las distintas fuentes de información
	Exploración de los Datos	• Suposiciones del Proyecto • Problema del Negocio • Diccionario de Fuente de Datos • Campos Relacionados con el Problema de Negocio • Reporte de Evaluación de Herramientas	Exploración de los Datos	• Fuente Integrada de datos • Reporte de Datos Explorados	Se analizan los valores de los campos identificados de interés para los distintos problemas de negocio, con el objetivo de comprender las características de la población o muestra de estudio, identificando relaciones iniciales entre las distintas variables estudiadas
	Evaluación de los Datos	• Diccionario de Fuente de Datos • Campos Relacionados con el Problema de Negocio • Reporte de Datos Explorados • Fuente Integrada de datos • Reporte de Evaluación de Herramientas	Exploración de la Calidad de los Datos	• Reporte de la Calidad de los Datos	Se analizan los campos identificados de interés para los distintos problemas de negocio, identificando aquellas características que puedan afectar la calidad del modelo
Modelado	Modelado del Problema	• Problema del Negocio • Diccionario de Fuente de Datos • Campos Relacionados con el Problema de Negocio	Derivación del Proceso de Explotación de Información	• Diseño del Proceso de Explotación de Información	Se realiza un modelado de representación de los problemas de negocio, identificando los métodos de explotación de información a utilizar
	Configuración del Modelo	• Criterios de Éxito del Problema de Negocio • Diccionario de Fuente de Datos • Campos Relacionados con el Problema de Negocio • Reporte de Datos Explorados • Reporte de la Calidad de los Datos • Diseño del Proceso de Explotación de Información • Reporte de Evaluación de Herramientas	Determinación de la Configuración del Modelo	• Selección de Algoritmos de Explotación de Información • Selección de Variables del Modelo • Estrategias de Evaluación de Modelos	Se definen los elementos que conforman la estrategia de implementación y evaluación de los distintos modelos para la extracción de patrones vinculados con el problema de negocio
Preparación de los Datos	Construcción de la Fuente Temporaria de Datos	• Reporte de Datos Explorados • Estrategias de Evaluación de Modelos • Fuente Integrada de datos • Reporte de Evaluación de Herramientas • Selección de variables del Modelo	Generación de la Fuente Temporaria de Datos	• Fuente Temporaria de Datos • Reporte de Generación de la Fuente Temporaria de datos	Se realizan las tareas finales para la generación de las fuentes de datos requeridas para las distintas etapas de implementación del modelo
	Adecuación de la Fuente Temporaria de Datos	• Reporte de la Calidad de los Datos • Selección de variables del Modelo • Fuente Temporaria de Datos • Reporte de Generación de la Fuente Temporaria de datos	Adecuación de los Datos	• Reporte de Adecuación de la Fuente Temporaria de Datos	Se transforman los datos de acuerdo a las necesidades del modelo

Tabla 4.95.b. MoProPEI: Estructura subproceso Desarrollo

Implementación	Selección del Modelo	• Selección de Algoritmos de explotación de información • Reporte de Generación de la Fuente Temporaria de datos • Criterios de Éxito del Problema de Negocio • Reporte de Evaluación de Herramientas	Selección de la Estrategia de Hiperparametrización	• Reporte de Estrategia de Parametrización del Modelo	Se define el proceso mediante el cual se evalúa la calidad de los modelos y el criterio de selección del mismo
	Explotación de Información	• Criterios de Éxito del Problema de Negocio • Selección de Algoritmos de explotación de información • Selección de variables del Modelo • Estrategias de evaluación de modelos • Reporte de Estrategia de Parametrización del Modelo • Fuente Temporaria de Datos • Reporte de Generación de la Fuente Temporaria de datos	Extracción de Conocimiento	• Reporte de Implementación del Modelo • Patrones de Conocimiento	Se aplican los algoritmos de explotación de información, documentando los resultados obtenidos
Evaluación y Presentación	Evaluación de los Resultados	• Objetivos del Proyecto • Criterios de Éxito del Proyecto • Problema del Negocio • Criterios de Éxito del Problema de Negocio • Reporte de Implementación del Modelo • Patrones de Conocimiento	Validación del Conocimiento	• Reporte de Evaluación de los Resultados	Se evalúa la validez de los patrones de conocimiento para los problemas de negocio
	Presentación de los Resultados	• Fuentes de Información del Cliente • Objetivos del Proyecto • Criterios de Éxito del Proyecto • Suposiciones del Proyecto • Restricciones del Proyecto • Problema del Negocio • Patrones de Conocimiento • Reporte de Generación de la Fuente Temporaria de datos • Reporte de la Calidad de los Datos • Reporte de Evaluación de los Resultados	Síntesis del Proyecto	• Reporte del Proyecto	Se garantiza la correcta transferencia del conocimiento extraído para dar soporte al proceso de toma de decisiones

Tabla 4.95.c. MoProPEI: Estructura subproceso Desarrollo

5. VALIDACIÓN

En [Hevner et al., 2004] se definen a los artefactos de tecnologías de información como construcciones (vocabulario y símbolos), modelos (abstracciones y representaciones), métodos (algoritmos y prácticas) e instancias (sistemas implementados y prototipos). Para este tipo de artefactos, los autores proponen 5 tipos de métodos para su evaluación:

- **Observacional**
  - *Caso de estudio*: se analiza el artefacto en profundidad en el ambiente de negocio.
  - *Estudio de campo*: se estudia el uso del artefacto en múltiples proyectos.
- **Analítico**
  - *Análisis estático*: examina la estructura del artefacto desde sus características estáticas.
  - *Análisis de la arquitectura*: estudia si el artefacto se ajusta a las características técnicas de la arquitectura de ingeniería de software.
  - *Optimización*: demuestra propiedades óptimas inherentes al artefacto.
  - *Análisis dinámico*: estudia las características dinámicas del artefacto en uso.
- **Experimental**
  - *Experimento controlado*: estudia las características del artefacto en un ambiente controlado.
  - *Simulación*: ejecuta el artefacto con datos artificiales.
- **Testeo**
  - *Funcional*: ejecuta las interfaces del artefacto con el objetivo de descubrir fallas e identificar defectos.
  - *Estructural*: se realizan pruebas por cubrimiento de algunas métricas acorde a la implementación del artefacto.
- **Descriptivo**
  - *Argumento informado*: se utiliza la información del marco teórico de la disciplina para generar un argumento convincente de las utilidades del artefacto.
  - *Escenarios*: se construyen escenarios detallados en torno al artefacto para demostrar su utilidad.

Para el método descriptivo, los autores señalan que sólo debe utilizarse en caso que no pueda aplicarse alguno de los métodos restantes.

A partir de lo previamente expuesto, se seleccionan como viables de acuerdo a las características del artefacto a construir, los métodos de validación: Observacional, Analítico y Experimental.

En las primeras dos secciones se pone en análisis el uso del artefacto en nuevos dominios de negocio (método observacional), los cuales poseen diferentes características para la aplicación del modelo de proceso propuesto. En la sección 5.1, se realiza un proyecto orientado a identificar el comportamiento de los usuarios que visitan una página web mediante el estudio de sus registros de navegación. En la sección 5.2, se presenta el segundo proyecto perteneciente al ámbito de la educación superior, en el cual se realiza un estudio con el objetivo de comprender el rendimiento académico de los estudiantes en contextos de masividad. En la sección 5.3 se examinan las características estáticas de la propuesta (método analítico) aplicando el marco comparativo de metodologías para proyectos de explotación de información [Moine, 2013]. El análisis se aplicará entre el modelo de proceso propuesto (MoProPEI) y los modelos de procesos / metodologías más relevantes (descriptas en los capítulos 2 y 3). Finalmente, en la sección 5.4, se presentan los resultados de aplicar un experimento controlado (método experimental), en el cual se implementa el trabajo realizado por el único experimento replicable identificado en la disciplina, comparando la propuesta superadora (IKDDM) con la realizada en el presente trabajo de investigación.

5.1. CASO DE VALIDACIÓN: WEB LOG

El primer caso de validación consiste en el análisis del sitio web perteneciente al Campus Virtual de la Universidad Nacional de Lanús. Esta área se configura como un entorno de enseñanza y de aprendizaje en el que se produce un encuentro académico entre profesores y estudiantes independientemente de la situación real de tiempo y espacio. Esta particularidad abre nuevas oportunidades de desarrollo curricular a través de la utilización de una plataforma educativa que se actualiza periódicamente.

El proyecto surge a partir del interés en aplicar nuevas tecnologías con la meta de realizar acciones que mejoren la experiencia del usuario. Mediante este, se provee de recursos, herramientas y servicios educativos a la comunidad, pensados principalmente para aquellos docentes y estudiantes pertenecientes a la comunidad educativa.

El problema general a abordar, consiste en la identificación de patrones en la navegación e intereses de los usuarios, con el objetivo de optimizar su experiencia, facilitando su uso y mejorando la disposición de los contenidos. Para alcanzar dicho objetivo se dispone de los registros (logs) de los usuarios que accedieron a la página web, junto con información del dispositivo y navegador

utilizado, el recurso accedido, etc., correspondiente a los últimos dos años. A continuación se provee el listado de variables disponibles:

- **IP**: dirección IP del usuario,
- **Referer**: página desde donde accede el usuario a la primera página del sitio web (ruta absoluta),
- **Searchterms**: términos de búsqueda utilizados para acceder al sitio web (si proviene de un buscador),
- **Resource**: nombre de la página accedida (ruta relativa),
- **Plugins**: programas adicionales instalados al navegador (por ejemplo: flash, acrobat, etc),
- **Visit_id**: identificador unívoco de sesión de usuarios durante una visita,
- **Server_latency**: tiempo requerido para recibir y procesar el pedido,
- **Page_performance**: tiempo requerido en cargarse la página al usuario,
- **Browser**: tipo de navegador que usa (por ejemplo: chrome, safari, mozilla, etc.),
- **Browser_version**: versión del navegador,
- **Browser_type**: identifica el tipo de dispositivo que utiliza (0 = pc; 2 = celular),
- **Platform**: sistema operativo del usuario (por ejemplo: win7, win10, android, etc.),
- **Language**: configuración del lenguaje del navegador (por ejemplo: es-es, es-ar, es-41, etc.),
- **User_agent**: agente de usuario (aplicación que utiliza como cliente),
- **Resolution**: resolución de la pantalla del cliente,
- **Screen_width**: ancho de la pantalla del cliente,
- **Screen_height**: alto de la pantalla del cliente,
- **Content_type**: tipo de acción que realizó el usuario.
- **Category**: categoría a la que pertenece el recurso (por ejemplo: Búsquedas = 1, Novedades =30, etc.),
- **Autor**: usuario que creó el recurso,
- **Content_id**: identificador único del recurso (página web)
- **Dt**: codificación de la fecha en formato Unix.

En las siguientes secciones, se presentan los elementos obtenidos y desarrollados a lo largo del proyecto, a partir de MoProPEI. Para facilitar la comprensión del proyecto, los resultados obtenidos se presentan acorde a la estructura del proceso propuesto, y no de manera temporal. Las actividades y tareas realizadas junto con los resultados obtenidos al aplicar las técnicas pertinentes a los subprocesos de gestión y desarrollo se detallan en las secciones 5.1.1 y 5.1.2 respectivamente.

5.1.1. MoProPEI-G: Subproceso Gestión (G)

El subproceso de gestión se implementa de manera transversal al subproceso de desarrollo. Se encuentra conformado por cinco fases: Iniciación (sección 5.1.1.1), Planificación (sección 5.1.1.2), Soporte (sección 5.1.1.3), Control (sección 5.1.1.4) y Cierre (sección 5.1.1.5).

5.1.1.1. Fase: Iniciación (G.IN)

Esta fase se encuentra integrada por cuatro actividades: Exploración Inicial del Proyecto (sección 5.1.1.1.1): donde se identifican los miembros de interés para el proyecto y las posibles situaciones de riesgo durante el desarrollo del mismo, Definición de la Comunicación (sección 5.1.1.1.2): se prevén las necesidades y canales de comunicación durante el desarrollo del proyecto, Evaluación de la Situación (sección 5.1.1.1.3): se analizan las herramientas de utilidad para el desarrollo del proyecto, determinando la viabilidad del mismo y Definición del Ciclo de Vida (sección 5.1.1.1.4): donde se establece de acuerdo a las características del proyecto, el flujo mediante el cual se llevarán a cabo las tareas de desarrollo.

Se destaca que las actividades que integran la fase, son realizadas de manera paralela a la fase Entendimiento del negocio del subproceso Desarrollo (sección 5.1.2.1), utilizando resultados de la fase como elementos de entrada. Dichas dependencias son señaladas al inicio de cada actividad, indicando los formalismos que utilizan.

5.1.1.1.1. Actividad: Exploración Inicial del Proyecto (G.In.EIP)

Mediante esta actividad se identifican y describen las personas involucradas en el proyecto, los posibles riesgos y las acciones en caso de contingencia.

A continuación se presentan los resultados obtenidos de aplicar la técnica caracterización del desarrollo del proyecto (sección 4.3.1.1.2, pág. 79) perteneciente a la metodología para la educción de requerimientos para proyectos de explotación de información, la cual utiliza como insumos los formalismos: Fuentes de Información del Cliente (Tabla 5.20), Objetivos del Proyecto (Tabla 5.22), Criterios de Éxito del Proyecto (Tabla 5.23), Expectativas del Proyecto (Tabla 5.24), Restricciones del Proyecto (Tabla 5.26), Problema del Negocio (Tabla 5.27) y Criterios de Éxito del Problema de Negocio (Tabla 5.28). Si bien estos formalismos son presentados de forma posterior en el documento (sección 5.1.2.1), fueron desarrollados en paralelo durante sus etapas iniciales.

Recursos Humanos Involucrados (G.In.EIP.ReHI): se deja registro de la información de los miembros de la organización que desarrolla el proyecto, así como de la organización cliente que realizó el contacto con el equipo de trabajo, siendo el único individuo partícipe del proyecto. Por cuestiones de privacidad, no se presenta la información de contacto de las personas.

En la tabla 5.1, se muestran los tres involucrados en el proyecto, teniendo en consideración las salvedades previamente mencionadas.

Riesgos del Proyecto (G.In.EIP.RiPr): no se identificaron riesgos para el proyecto.

Plan de Contingencias (G.In.EIP.Pcon): no se identificaron riesgos para el proyecto.

Recursos Humanos Involucrados					
Responsable:	Sebastian M.			**Fecha:**	06/02/2017
ID#:	G.In.EIP.ReHI			**Versión:**	1.0
ID	**Nombre**	**Rol/Posición**	**Pertenece a**	**Descripción**	**Información de contacto**
rehi.1	Sebastian M.	Líder de Proyecto	recurso interno	Persona encargada de la dirección del proyecto	Skype: XXXX
rehi.2	Santiago B.	Ingeniero de Explotación de información	recurso interno	Ingeniero de Explotación de información	Skype: XXXXX
rehi.3	Dario R.	Cliente	Organización Contratante	Responsable del sitio web	Correo: XXXX@gmail.com

Tabla 5.1. Caso de Validación: Web Log - Recursos Humanos Involucrados

5.1.1.1.2. Actividad: Definición de la Comunicación (G.In.DeC)

En esta actividad, se establecen estrategias formales de comunicación a partir de la necesidad e intereses de las partes involucradas en el proyecto. A continuación se presentan los resultados obtenidos de aplicar la técnica Definición de la Comunicación (sección 4.3.1.2.2, pág. 85), la cual utiliza como insumo los Recursos Humanos Involucrados (Tabla 5.1).

Plan de Comunicación (G.In.DeC.PCom): en el formalismo de entrada, se identificaron tres recursos asociados al proyecto, dos miembros del equipo de trabajo y el cliente, previendo tres tipos de comunicaciones: de comprensión del proyecto, de reporte de avances y de estado interno del proyecto.

El primero, tiene como objetivo mantener un continuo vínculo con el experto e interesado del negocio, analizando los alcances y restricciones del proyecto, participando la totalidad de los miembros. Dicha comunicación se realiza de manera semanal durante el periodo planificado de entendimiento del negocio.

El segundo tipo de comunicación, es para mantener al cliente informado durante el desarrollo del proyecto, pudiendo identificar posibles cambios de intereses o nuevas necesidades. La información se brinda de manera bimensual y participan los tres involucrados.

Finalmente, el último tipo de comunicación tiene como objetivo mantener al equipo de trabajo informado respecto al estado del proyecto y los problemáticas que pudiesen ocurrir durante el desarrollo del mismo. Participan los miembros del equipo de trabajo y serán realizadas de manera bisemanal.

Los modos mediante los cuales las comunicaciones serán realizadas son: presencial para las comunicaciones con el cliente, y videollamada, mediante la herramienta Skype, para las reuniones bisemanales de seguimiento del trabajo. El responsable de las mismas es líder del proyecto. En la tabla 5.2 se deja registro de la información previamente descripta.

Plan de Comunicación				
Responsable:	Sebastian M.	**Fecha:**		08/02/2017
ID#:	G.In.DeC.PCom	**Versión:**		1.0
Interesados	**Información**	**Frecuencia**	**Medio**	**Responsable**
(rehi.1) Sebastian M. (rehi.2) Santiago B. (rehi.3) Dario R.	Comprensión del Proyecto	semanal durante el periodo de entendimiento del negocio	Presencial	(rehi.1) Sebastian M.
(rehi.1) Sebastian M. (rehi.2) Santiago B. (rehi.3) Dario R.	Avances del Proyecto	bimensual	Presencial	(rehi.1) Sebastian M.
(rehi.1) Sebastian M. (rehi.2) Santiago B.	Estado del Proyecto	bisemanal	Skype	(rehi.1) Sebastian M.

Tabla 5.2. Caso de Validación: Web Log - Plan de Comunicación

5.1.1.1.3. Actividad: Evaluación de la Situación (G.In.EvS)

En esta actividad se analiza la posibilidad de éxito del proyecto, teniendo en consideración los objetivos y las soluciones que brindan las distintas herramientas de explotación de información existentes. Los objetivos de la actividad son: seleccionar las herramientas a utilizar y determinar la viabilidad del proyecto.

A continuación se presentan los resultados obtenidos de aplicar la metodología para la selección de herramientas de explotación de información y el modelo de evaluación de viabilidad para proyectos de explotación de información (sección 4.3.1.3.2, pág. 90), utilizando como insumos los formalismos: Objetivos del Proyecto (Tabla 5.22), Problema del Negocio (Tabla 5.27), Fuentes de Información del Cliente (Tabla 5.20), Recursos Humanos Involucrados (Tabla 5.1), Expectativas del Proyecto (Tabla 5.24) y Suposiciones del Proyecto (Tabla 5.25).

Reporte de Evaluación de Herramientas (G.In.EvS.REHe): para el desarrollo del proyecto, se identifican como posibles herramientas para utilizar, de acuerdo a las necesidades del cliente y la experiencia de los miembros del proyecto: RapidMiner Versión 7.4, Weka Versión 3.7.11 y Orange Versión 2.7.8. A partir de ello, se califica cada una de las características de acuerdo a la escala de valores entre 1 y 4 de acuerdo al tipo de pregunta, señalando con "--" aquellos aspectos que no han sido evaluados. Como resultado, se identifica a la herramienta RapidMiner como la más adecuada para el proyecto. En la tabla 5.3 se ilustra las valoraciones realizadas y los resultados obtenidos para cada una de las herramientas en las cuatro características generales.

Reporte de Evaluación de Herramientas					
Responsable:	Sebastian M.			**Fecha:**	20/02/2017
ID#:	G.In.EvS.REHe			**Versión:**	1.0
Criterios:					
Evaluación: 1 = Malo, 2 = débil, 3 = Bueno, 4 = Excelente					1 = No, 4 = SI
Herramientas		**Peso**	**RapidMiner V.7.4**	**Weka V.3.7.11**	**Orange V.2.7.8**
1. Funcional - Características Técnicas					
Soporte de Metodología / Ciclo de vida	Soporte del proceso	3	3	2	2
Compatibilidad con fuentes de datos	Base de datos	8	3	2	1
	Otras fuentes (word, excel, etc.)	8	4	2	3
Integración	Soporte de distintas técnicas asociadas al proceso de explotación de Información	5	4	4	4
Multilenguaje	Soporta distintas idiomas	2	1	1	1
Técnicas	Variedad de técnicas que provee	18	4	4	4
Reporte y visualización	Permite generar reportes y visualizaciones	12	2	2	2
Multiplataforma	Soporta múltiples plataformas	5	4	4	4
Instalación remota	La administración y mantenimiento son remotos	5	4	1	1
Usuarios Múltiples	Posee perfiles de usuarios	2	1	1	1
Seguridad	Provee seguridad de la información configurada por perfiles	2	1	1	1
Backup	Metodología de backup	2	1	1	1
Amigable	Interfaz de usuario	10	4	2	4
Configuraciones	Permite la configuración del perfil	8	--	--	--
Documentación	Servicio de soporte y ayuda	5	4	1	3
Conexión	Soporta conexión por: Internet, FTP, ERPs.	2	1	1	1
Soporte de sistemas de mensaje	Soporta compartir información (por mail u otro medio)	3	3	1	1
Total			300	217	247
	Peso del Grupo	40%	120	86,8	98,8
2. Características del Proveedor					
Características del proveedor	Historia	30	3	3	1
Crecimiento	Perspectiva a futuro	10	3	3	2
Ubicación Geográfica	Oficinas	30	--	--	--
Implementación	Otras implementaciones de la misma herramienta	5	--	--	--
	Contacto con otros clientes	5	--	--	--
Confidencialidad	Confidencialidad de la información	20	--	--	--
Total			120	120	50
	Peso del Grupo	25%	30	30	12,5

Tabla 5.3.a. Caso de Validación: Web Log - Reporte de Evaluación de herramientas

3. Características del Servicio					
Garantía del producto	Duración y Alcance	30	--	--	--
Mejora	Brinda soporte a versiones previas	20	--	--	--
Licencia	Costo, alcances y soporte postventa	30	--	--	--
Soporte	Tiempo de respuesta y disponibilidad	20	--	--	--
Total			0	0	0
	Peso del Grupo	20%	0	0	0
4. Características Económicas					
Costo del software	Costo de la herramienta	30	--	--	--
Costo del Hardware	Necesidad de mejorar o comprar nuevo hardware compatible con la herramienta	20	--	--	--
Otros costos software	Costos adicionales al producto (backup, web servers, bases de datos, etc.)	20	--	--	--
Licencias	Política de licencia	10	--	--	--
Financiamiento	Existencia	10	--	--	--
Mejoras	Costo promedio de la mejora del producto	10	--	--	--
Total			0	0	0
	Peso del Grupo	-15%	0	0	0
Final					
1. Funcional - Características Técnicas		40%	120	86,8	98,8
2. Características del Proveedor		25%	30	30	12,5
3. Características del Servicio		20%	0	0	0
4. Características Económicas		-15%	0	0	0
TOTAL			**150**	**116,8**	**111,3**

Tabla 5.3.b. Caso de Validación: Web Log - Reporte de Evaluación de herramientas

Reporte de Evaluación de Viabilidad (G.In.EvS.REVi): a partir de la información recabada sobre la fuente de datos, el problema de negocio y los miembros que forman parte del mismo, se valúan las trece características a considerar, determinando las valoraciones por dimensión y global. De los resultados obtenidos, se verifica en primera instancia que las valoraciones individuales de cada característica son superiores al umbral, y que las valoraciones de las dimensiones y global son mayores a 5, por lo cual se determina al proyecto como viable. En la tabla 5.4 se ilustra las valoraciones realizadas y los resultados obtenidos para cada dimensión del proyecto.

5.1.1.1.4. Actividad: Definición del Ciclo de Vida (G.In.DCV)

En la actividad actual, se analizan las características del proyecto con el objetivo de definir la estrategia de implementación más adecuada para el desarrollo del mismo. Del resultado de esta actividad se establece la estructura y el flujo de ejecución de las fases en el caso de validación.

Reporte de Evaluación de Viabilidad												
Responsable:	Sebastian M.							**Fecha:**			21/02/2017	
ID#:	G.In.EvS.REVi							**Versión:**			1.0	
Datos						**Problema de Negocio**			**Proyecto**		**Equipo de Trabajo**	
P1	**P2**	**A1**	**A2**	**A3**	**E1**	**P3**	**A4**	**A5**	**E2**	**E3**	**P4**	**E4**
todo	mucho	mucho	regular	mucho	mucho	regular	mucho	regular	todo	mucho	mucho	mucho
Umbral												
poco	poco	poco	poco	poco	nada	poco	poco	poco	nada	nada	poco	nada

Dimensiones		***Viabilidad global***	***Resultado***
Plausibilidad	7.035	***6.91***	***Viable***
Adecuación	6.169		
Éxito	7.73		

Tabla 5.4. Caso de Validación: Web Log - Evaluación de Viabilidad

A continuación se presentan los resultados obtenidos de aplicar la técnica Selección del Ciclo de Vida (sección 4.3.1.4.2, pág. 97). Esta utiliza como insumos: Objetivos del Proyecto (Tabla 5.22), Expectativas del Proyecto (Tabla 5.24), Problema del Negocio (Tabla 5.27), Riesgos del Proyecto (no aplicable en este proyecto) y Recursos Humanos Involucrados (Tabla 5.1).

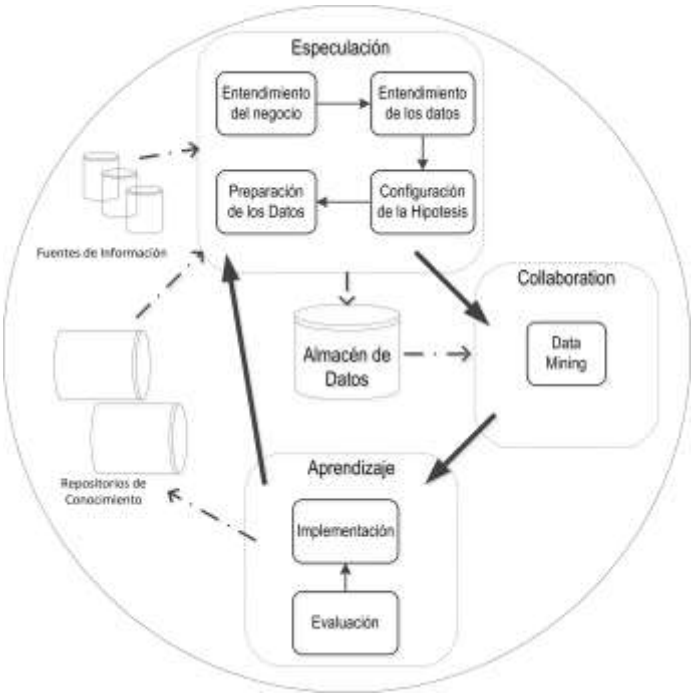
Modelo de Ciclo de Vida (G.In.DCV.MoCV): debido a que el cliente no tiene bien definidos los alcances del proyecto y la experiencia del equipo en proyectos con objetivos similares, se determina utilizar el ciclo de vida ASD-BI (sección 2.2.3, pág. 46), haciendo uso de sus características iterativas para facilitar a comprender al cliente las problemáticas de interés a resolver. Se prevé la realización de dos iteraciones siendo necesaria la finalización integral de la primera para comenzar con la segunda de ellas. La tabla 5.5 ilustra el formalismo generado, presentando la visualización del modelo de ciclo de vida y su criterio de transición.

5.1.1.2. Fase: Planificación (G.Pl)

En la fase Planificación se define el curso de acciones requeridas para alcanzar los objetivos del proyecto. Se encuentra conformada por cuatro actividades: Planificación de la Mediciones (sección 5.1.1.2.1), Planificación de las Actividades (sección 5.1.1.2.2), Planificación de los Recursos (sección 5.1.1.2.3) y Planificación de las Responsabilidades (sección 5.1.1.2.4).

Se destaca que las actividades que integran la fase, son realizadas de manera paralela a la fase Entendimiento del negocio (sección 5.1.2.1) del subproceso Desarrollo, utilizando resultados de dicha fase como elementos de entrada. Dichas dependencias son señaladas al inicio de cada actividad, indicando los formalismos que utilizan.

Modelo de Ciclo de Vida			
Responsable:	Sebastian M.	**Fecha:**	22/02/2017
ID#:	G.In.DCV.MoCV	**Versión:**	1.0



Criterios de transición:
Finalización completa de la iteración

Tabla 5.5. Caso de Validación: Web Log - Modelo de Ciclo de Vida

5.1.1.2.1. Actividad: Planificación de la Mediciones (G.Pl.PlM)

En esta actividad se realiza una estimación inicial del tiempo requerido para el desarrollo del programa del proyecto y se definen las mediciones que se llevarán a cabo durante el transcurso del mismo.

A continuación se presentan los resultados obtenidos de aplicar las técnicas Métricas para Proyectos de Explotación de Información y Modelo de Estimación para Proyectos de Explotación de Información (sección 4.3.2.1.2, pág. 102), las cuales utilizan como insumo los siguientes formalismos: Objetivos del Proyecto (Tabla 5.22), Problema del Negocio (Tabla 5.27), Fuentes de Información del Cliente (Tabla 5.20), Recursos Humanos Involucrados (Tabla 5.1) y Reporte de Evaluación de Herramientas (Tabla 5.3).

Listado de Métricas (G.Pl.PlA.LiMe): para este proyecto se propone el uso de tres métricas, dos enfocadas en el proyecto y una al modelo. La tabla 5.6 se presenta cada una de ellas detallando su forma de cálculo.

Listado de Métricas				
Responsable:	Sebastian M.		**Fecha:**	16/02/2017
ID#:	G.Pl.PIA.LiMe		**Versión:**	1.0
Nombre	**Tipo**	**Fórmula**	**Comentarios**	
Tiempo total requerido para el desarrollo del proyecto	Proyecto	DRPY = $\sum$ trA trA = tiempo requerido por actividad	Sumatoria de los tiempos requeridos para cada actividad del proyecto	
Tiempo medio requerido para el desarrollo de un problema de explotación de información	Proyecto	($\sum$ trA.SD) / NPEI trA.SD = tiempo requerido por actividad del subproceso de desarrollo NPEI = cantidad de problemas de explotación de información	Solo se considera el tiempo de las actividades pertenecientes al subproceso de desarrollo	
Número medio de atributos significativos por modelo	Modelo	($\sum$ AtS.M) / NMOD AtS.M = cantidad de atributos significativos de un modelo NMOD = cantidad de modelos	atributos significativos: aquellos que participan en la definición del patrón identificado	

Tabla 5.6. Caso de Validación: Web Log - Listado de Métricas

Estimación del Proyecto (G.Pl.PlA.EsPr): de acuerdo a las características del proyecto, se evalúan y fijan los factores de acuerdo a las escalas establecidas en la técnica. A partir de los formalismos objetivos del proyecto y problema de negocio, se identifica que las necesidades del cliente están asociadas con: "conocer los atributos que poseen mayor incidencia sobre la identificación de una clase desconocida a priori" (asignando el valor "5" al factor OBTY). A partir del formalismo fuente de información del cliente, se identifica que se posee un repositorio de datos con tecnología compatible para su integración (asignando el valor "1" al factor AREP), la cantidad de tuplas existentes es aproximadamente de 150000 (asignando el valor "5" al factor QTUM), no existen tablas auxiliares (asignando el valor "1" al factor QTUA) y se posee un registro no muy detallado de las variables del sistema, pero se encuentra a disposición el experto para explicar el significado de los datos disponibles (asignando el valor "3" al factor KLDS). De acuerdo a los interesados del proyecto registrados en Recursos Humanos Involucrados (Tabla 5.1), se identifica las variables LECO con el valor "1" (dado que se posee completo respaldo de los interesados) y KEXT, con el valor "3" (debido a que el equipo trabajó en otro tipo de organizaciones, con datos y objetivos similares). Finalmente, se identifica que la herramienta RapidMiner, seleccionada a partir del Reporte de Evaluación de Herramientas (Tabla 5.3), posee amplias funciones para las etapas de formateo, integración, descripción de los datos y de implementación y optimización de modelos (asignando al factor TOOL, el valor "1").

A partir de los valores definidos, se obtiene como esfuerzo total del subproceso desarrollo: 2.96 meses/hombre y de acuerdo al porcentaje estimado para tareas de gestión (15%), se determina como esfuerzo para dicha fase: 0.44 meses/hombre, siendo el esfuerzo total del proyecto igual a: 3.40 meses/hombre. La tabla 5.7 ilustra las valoraciones realizadas y el resultado obtenido.

Estimación del Proyecto										
Responsable:	Sebastian M.					**Fecha:**		24/02/2017		
ID#:	G.Pl.PlM.EsPr					**Versión:**		1.0		
Esfuerzo										
OBTY	**LECO**	**AREP**	**QTUM**	**QTUA**	**KLDS**	**KEXT**	**TOOL**	**Total Desarrollo**	**Total Gestión**	**Total**
5	1	1	5	1	3	3	1	*2,96*	*0,44*	**3,40**

Tabla 5.7. Caso de Validación: Web Log - Estimación del Proyecto

5.1.1.2.2. Actividad: Planificación de las Actividades (G.Pl.PlA)

En esta actividad se prevén las acciones a realizar durante el transcurso del proyecto y sus alcances, definiendo la ejecución de las actividades en el tiempo. Como resultado de esta tarea, se genera el programa de actividades del proyecto.

A continuación se presentan los resultados obtenidos de aplicar la técnica Definición del Programa del Proyecto (sección 4.3.2.2.2, pág. 109), la cual utiliza como insumos los formalismos: Modelo de Ciclo de Vida (Tabla 5.5), Estimación del Proyecto (Tabla 5.7), Objetivos del Proyecto (Tabla 5.22), Problema del Negocio (Tabla 5.27) y Fuentes de Información del Cliente (Tabla 5.20).

Mapa de Actividades (G.Pl.PlA.MaAc): A partir del modelo de ciclo de vida seleccionado, se determinan las etapas durante las cuales se desarrollarán las distintas actividades del proceso. Se prevé la realización de dos iteraciones, con el objetivo de profundizar en las problemáticas a abordar, a partir del conocimiento extraído en la primera parte del proyecto. La tabla 5.8, presenta la distribución de las actividades según la estructura del modelo de ciclo de vida seleccionado.

Plan de Acción (G.Pl.PlA.PlAc): a partir de la estimación de tiempos y la selección de las actividades a realizar en cada etapa del modelo de ciclo de vida (mapa de actividades, Tabla 5.8), se asigna la duración y rango de fechas de ejecución a cada una de las actividades usando de base las mediciones de esfuerzo requeridas para proyectos de explotación de información [Rodríguez et al., 2010], teniendo en consideración que para el proyecto se utiliza una base de jornada diaria de cuatro horas.

El plan de acción se mantiene actualizado durante el desarrollo del proyecto, registrando en los hitos de control y reporte de estado, los avances del proyecto (tiempos y fechas reales). El resultado obtenido al final del proyecto, se presenta en la tabla 5.9. En la sección A.2.1, se exhiben las dos versiones previas del formalismo: 1.0 y 1.1 (Tablas A.6 y A.7 respectivamente), y el diagrama Gantt (Figura A.2).

Mapa de Actividades			
Responsable:	Sebastian M.	**Fecha:**	23/02/2017
ID#:	G.Pl.PlA.MaAc	**Versión:**	1.0

ID	Fase/Actividad	Iteración 1							Iteración 2						
		E.N.	E.D.	C.H.	P.D.	D.M.	E.	I.	E.N.	E.D.	C.H.	P.D.	D.M.	E.	I.
G.In	**Iniciación**														
G.In.EIP	Exploración Inicial del Proyecto	x													
G.In.DeC	Definición de la Comunicación	x							x						
G.In.EvS	Evaluación de la Situación	x							x						
G.In.DCV	Definición del Ciclo de Vida	x													
G.Pl	**Planificación**														
G.Pl.PIM	Planificación de la Mediciones	x	x												
G.Pl.PIA	Planificación de las Actividades	x	x						x	x					
G.Pl.PIR	Planificación de los Recursos	x	x						x	x					
G.Pl.PRe	Planificación de las Responsabilidades	x	x						x	x					
G.So	**Soporte**														
G.So.MeP	Mediciones del Proyecto	x	x	x	x	x	x	x	x	x	x	x	x	x	x
G.So.GeC	Gestión de la Configuración	x	x	x	x	x	x	x	x	x	x	x	x	x	x
G.Co	**Control**														
G.Co.GeD	Gestión del Desarrollo	x	x	x	x	x	x	x	x	x	x	x	x	x	x
G.Co.CoA	Control de las Actividades	x	x	x	x	x	x	x	x	x	x	x	x	x	x
G.Co.Gca	Gestión del Cambio	x	x	x	x	x	x	x	x	x	x	x	x	x	x
G.Ci	**Cierre**														
G.Ci.FEC	Formalización Externa del Cierre del Proyecto														x
G.Ci.FIC	Formalización Interna del Cierre del Proyecto														x
D.EN	**Entendimiento del Negocio**														
D.EN.AnN	Análisis del Negocio	x							x						
D.EN.CPN	Comprensión del Problema de Negocio	x							x						
D.ED	**Entendimiento de los Datos**														
D.ED.AnD	Análisis de los Datos		x							x					
D.ED.ExD	Exploración de los Datos		x							x					
D.ED.EvD	Evaluación de los Datos		x							x					
D.Mo	**Modelado**														
D.Mo.MoP	Modelado del problema			x							x				
D.Mo.CoM	Configuración del Modelo			x							x				
D.PD	**Preparación de los Datos**														
D.PD.AFI	Construcción de la Fuente Temporaria de Datos				x							x			
D.PD.CFT	Adecuación de la Fuente Temporaria de Datos				x							x			
D.Im	**Implementación**														
D.Im.SeM	Selección del Modelo					x							x		
D.Im.ExI	Explotación de Información					x							x		
D.EP	**Evaluación y Presentación**														
D.EP.EvR	Evaluación de los Resultados							x							x
D.EP.PrR	Presentación de los Resultados														x

Tabla 5.8. Caso de Validación: Web Log - Mapa de Actividades

Plan de Acción								
Responsable:		Sebastian M.			**Fecha:**	11/05/17		
ID#:		G.PI.PIA.PIAc			**Versión:**	1.2		
ID Actividad	**Actividad**	**Inicio Estimado**	**Inicio Real**	**Fin Estimado**	**Fin Real**	**Esfuerzo Estimado (Hs)**	**Esfuerzo Real (Hs)**	**Comentarios**
G.In	**Iniciación**	**06/02/17**	**06/02/17**	**07/04/17**	**07/04/17**	**16**	**18**	
G.In.EIP	Exploración Inicial del Proyecto	06/02/17	06/02/17	17/02/17	17/02/17	4	4	
G.In.DeC	Definición de la Comunicación	06/02/17	06/02/17	07/04/17	07/04/17	4	3	
G.In.EvS	Evaluación de la Situación	06/02/17	06/02/17	07/04/17	07/04/17	6	7	
G.In.DCV	Definición del Ciclo de Vida	20/02/17	20/02/17	22/02/17	22/02/17	2	4	
G.PI	**Planificación**	**06/02/17**	**06/02/17**	**14/04/17**	**14/04/17**	**16**	**25**	
G.PI.PIM	Planificación de la Mediciones	06/02/17	06/02/17	24/02/17	24/02/17	2	4	
G.PI.PIA	Planificación de las Actividades	06/02/17	06/02/17	14/04/17	14/04/17	4	8	
G.PI.PIR	Planificación de los Recursos	06/02/17	06/02/17	14/04/17	14/04/17	4	3	
G.PI.PRe	Planificación de las Responsabilidades	06/02/17	06/02/17	14/04/17	14/04/17	6	10	
G.So	**Soporte**	**06/02/17**	**06/02/17**	**11/05/17**	**11/05/17**	**18**	**19**	
G.So.MeP	Mediciones del Proyecto	06/02/17	06/02/17	11/05/17	11/05/17	8	9	Se prevé el registro de mediciones al fin de cada iteración
G.So.GeC	Gestión de la Configuración	06/02/17	06/02/17	11/05/17	11/05/17	10	10	
G.Co	**Control**	**06/02/17**	**06/02/17**	**11/05/17**	**11/05/17**	**22**	**20**	
G.Co.GeD	Gestión del Desarrollo	06/02/17	06/02/17	11/05/17	11/05/17	8	9	Se prevé la aplicación del reporte de estado al fin de cada iteración
G.Co.CoA	Control de las Actividades	06/02/17	06/02/17	11/05/17	11/05/17	10	9	
G.Co.Gca	Gestión del Cambio	13/02/17	13/02/17	11/05/17	11/05/17	4	2	
G.Ci	**Cierre**	**09/05/17**	**09/05/17**	**11/05/17**	**11/05/17**	**4**	**6**	
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	09/05/17	09/05/17	10/05/17	10/05/17	2	2	
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	09/05/17	09/05/17	11/05/17	11/05/17	2	4	
Iteración 1		**06/02/17**	**06/02/17**	**27/03/17**	**27/03/17**	**270**	**251**	
D.EN	**Entendimiento del Negocio**	**06/02/17**	**06/02/17**	**24/02/17**	**20/02/17**	**50**	**50**	
D.EN.AnN	Análisis del Negocio	06/02/17	06/02/17	20/02/17	20/02/17	30	30	
D.EN.CPN	Comprensión del Problema de Negocio	14/02/17	14/02/17	24/02/17	20/02/17	20	20	
D.ED	**Entendimiento de los Datos**	**22/02/17**	**22/02/17**	**06/03/17**	**06/03/17**	**68**	**70**	
D.ED.AnD	Análisis de los Datos	22/02/17	22/02/17	01/03/17	28/02/17	30	30	
D.ED.ExD	Exploración de los Datos	27/02/17	27/02/17	03/03/17	03/03/17	30	32	
D.ED.EvD	Evaluación de los Datos	01/03/17	01/03/17	06/03/17	06/03/17	8	8	
D.Mo	**Modelado**	**06/03/17**	**06/03/17**	**15/03/17**	**15/03/17**	**42**	**38**	
D.Mo.MoP	Modelado del problema	06/03/17	06/03/17	13/03/17	13/03/17	24	20	
D.Mo.CoM	Configuración del Modelo	10/03/17	10/03/17	15/03/17	15/03/17	18	18	
D.PD	**Preparación de los Datos**	**13/03/17**	**13/03/17**	**20/03/17**	**17/03/17**	**40**	**28**	
D.PD.AFI	Construcción de la Fuente Temporaria de Datos	13/03/17	13/03/17	17/03/17	17/03/17	30	20	
D.PD.CFT	Adecuación de la Fuente Temporaria de Datos	15/03/17	15/03/17	20/03/17	20/03/17	10	8	

Tabla 5.9.a. Caso de Validación: Web Log - Plan de Acción (fin del proyecto)

D.Im	**Implementación**	**17/03/17**	**17/03/17**	**27/03/17**	**27/03/17**	**56**	**53**	
D.Im.SeM	Selección del Modelo	17/03/17	17/03/17	21/03/17	21/03/17	16	8	
D.Im.ExI	Explotación de Información	21/03/17	21/03/17	27/03/17	27/03/17	40	45	
D.EP	**Evaluación y Presentación**	**24/03/17**	**24/03/17**	**27/03/17**	**27/03/17**	**14**	**12**	
D.EP.EvR	Evaluación de los Resultados	24/03/17	24/03/17	27/03/17	27/03/17	14	12	
D.EP.PrR	Presentación de los Resultados	-	-	-	-	0	0	
Iteración 2		**28/03/17**	**28/03/17**	**09/05/17**	**09/05/17**	**184**	**147**	
D.EN	**Entendimiento del Negocio**	**28/03/17**	**28/03/17**	**07/04/17**	**03/04/17**	**26**	**20**	
D.EN.AnN	Análisis del Negocio	28/03/17	28/03/17	03/04/17	30/03/17	10	6	
D.EN.CPN	Comprensión del Problema de Negocio	03/04/17	31/03/17	07/04/17	03/04/17	16	14	
D.ED	**Entendimiento de los Datos**	**03/04/17**	**03/04/17**	**14/04/17**	**13/04/17**	**40**	**39**	
D.ED.AnD	Análisis de los Datos	03/04/17	03/04/17	10/04/17	05/04/17	16	12	
D.ED.ExD	Exploración de los Datos	03/04/17	03/04/17	12/04/17	13/04/17	20	26	
D.ED.EvD	Evaluación de los Datos	12/04/17	12/04/17	14/04/17	13/04/17	4	1	
D.Mo	**Modelado**	**14/04/17**	**14/04/17**	**20/04/17**	**19/04/17**	**30**	**22**	
D.Mo.MoP	Modelado del problema	14/04/17	14/04/17	18/04/17	16/04/17	18	10	
D.Mo.CoM	Configuración del Modelo	17/04/17	17/04/17	20/04/17	19/04/17	12	12	
D.PD	**Preparación de los Datos**	**18/04/17**	**18/04/17**	**24/04/17**	**24/04/17**	**26**	**14**	
D.PD.AFI	Construcción de la Fuente Temporaria de Datos	18/04/17	18/04/17	21/04/17	20/04/17	18	12	
D.PD.CFT	Adecuación de la Fuente Temporaria de Datos	21/04/17	21/04/17	24/04/17	24/04/17	8	2	
D.Im	**Implementación**	**24/04/17**	**24/04/17**	**02/05/17**	**28/04/17**	**28**	**20**	
D.Im.SeM	Selección del Modelo	24/04/17	24/04/17	26/04/17	25/04/17	8	4	
D.Im.ExI	Explotación de Información	26/04/17	25/04/17	02/05/17	28/04/17	20	16	
D.EP	**Evaluación y Presentación**	**02/05/17**	**02/05/17**	**09/05/17**	**09/05/17**	**34**	**32**	
D.EP.EvR	Evaluación de los Resultados	02/05/17	02/05/17	05/05/17	05/05/17	14	12	
D.EP.PrR	Presentación de los Resultados	05/05/17	05/05/17	09/05/17	09/05/17	20	20	

Tabla 5.9.b. Caso de Validación: Web Log - Plan de Acción (fin del proyecto)

5.1.1.2.3. Actividad: Planificación de los Recursos (G.Pl.PlR)

En esta actividad se prevén los recursos (tanto humanos como materiales) que se necesitarán para el desarrollo de las actividades en el tiempo.

A continuación se presentan los resultados obtenidos de aplicar la técnica Planificación de los Recursos Necesarios (sección 4.3.2.3.2, pág. 116), la cual utiliza como insumos los formalismos: Recursos Humanos Involucrados (Tabla 5.1), Reporte de Evaluación de Herramientas (Tabla 5.3), Plan de Acción (Tabla 5.9), Problema del Negocio (Tabla 5.27) y Fuentes de Información del Cliente (Tabla 5.20).

Plan de Necesidad de Recursos (G.Pl.PlR.PlNR): para este proyecto, se contará con dos recursos humanos: un líder del proyecto y un ingeniero de explotación de información, los cuales serán requeridos durante todo el proyecto. Adicionalmente, se detalla la necesidad de dos computadoras para cada uno de los miembros con las siguientes características: SO Linux o Windows 7 (en

adelante), RAM 4GB o más y 10GB o más espacio en disco, de acuerdo con las necesidades requeridas por la herramienta seleccionada. En la tabla 5.10 se registra la información previamente descripta en el formalismo.

Plan de Necesidad de Recursos					
Responsable:	Sebastian M.			**Fecha:**	24/02/2017
ID#:	G.Pl.PIR.PINR			**Versión:**	1.0
Recursos Humanos					
ID	**Recurso**	**Cantidad**	**F. Inicio**	**F. Fin**	**Descripción**
rhr.1	Líder de Proyecto	1	06/02/17	11/05/17	
rhr.2	Ingeniero de Explotación de Información	1	06/02/17	11/05/17	
Recursos Materiales					
ID	**Recurso**	**Cantidad**	**F. Inicio**	**F. Fin**	**Descripción**
rmr.1	Computadora Personal	2	06/02/17	11/05/17	SO Linux o Windows 7 (en adelante) RAM 4GB o más 10GB o más espacio en disco

Tabla 5.10. Caso de Validación: Web Log - Plan de Necesidad de Recursos

5.1.1.2.4. Actividad: Planificación de las Responsabilidades (G.Pl.PRe)

En esta actividad se definen las responsabilidades y obligaciones de las partes involucradas en el proyecto, dejando formalizado quien es el encargado de realizar cada tarea y los compromisos asumidos por cada una de las partes intervinientes en el acuerdo.

A continuación se presentan los resultados obtenidos de aplicar la técnica Designación de Responsabilidades (sección 4.3.2.4.2, pág. 123), la cual utiliza como insumos los formalismos: Recursos Humanos Involucrados (Tabla 5.1), Plan de Comunicación (Tabla 5.2), Plan de Acción (Tabla 5.9), Plan de Necesidad de Recursos (Tabla 5.10), Objetivos del Proyecto (Tabla 5.22), Criterios de Éxito del Proyecto (Tabla 5.23), Expectativas del Proyecto (Tabla 5.24), Restricciones del Proyecto (Tabla 5.26), Problema del Negocio (Tabla 5.27), Criterios de Éxito del Problema de Negocio (Tabla 5.28), Riesgos del Proyecto (no aplicable en este proyecto) y Plan de Contingencias (no aplicable en este proyecto).

Matriz de Responsabilidades (G.Pl.PlR.MaRe): en el formalismo Recursos Humanos Involucrados, se identifican tres interesados (dos miembros del equipo y un cliente/experto) introduciendo a cada uno de ellos en una columna, y asignando el nivel de participación en cada una de las actividades de acuerdo al interés de información y el conocimiento de los mismos. La tabla 5.11 ilustra el nivel de participación de cada miembro en las actividades del proyecto.

Matriz de Responsabilidades			
Responsable:	Sebastian M.	**Fecha:**	27/02/2017
ID#:	G.Pl.PlR.MaRe	**Versión:**	1.0

Descripción

Niveles de participación:
(R) Responsable: encargado de las tareas asociadas a la actividad.
(E) Ejecución: asignado tareas asociadas a la actividad.
(A) Aprobación: aceptación Final del resultado de la actividad.
(C) Consultado: posee conocimiento relevante para el desarrollo de la actividad.
(I) Informado: requiere estar alerta del progreso de la actividad.

ID Actividad	Actividad	Sebastian M. (rehi.1)	Santiago B. (rehi.2)	Dario R. (rehi.3)
G.In	**Iniciación**			
G.In.EIP	Exploración Inicial del Proyecto	R	I	
G.In.DeC	Definición de la Comunicación	R	I	I
G.In.EvS	Evaluación de la Situación	R	C	I
G.In.DCV	Definición del Ciclo de Vida	R	I	
G.Pl	**Planificación**			
G.Pl.PlM	Planificación de la Mediciones	R	I	
G.Pl.PlA	Planificación de las Actividades	R	I	I
G.Pl.PlR	Planificación de los Recursos	R	I	
G.Pl.PRe	Planificación de las Responsabilidades	R	I	A
G.So	**Soporte**			
G.So.MeP	Mediciones del Proyecto	R	E	
G.So.GeC	Gestión de la Configuración	R	E	
G.Co	**Control**			
G.Co.GeD	Gestión del Desarrollo	R	I	I
G.Co.CoA	Control de las Actividades	R	I	I
G.Co.Gca	Gestión del Cambio	A	R	I
G.Ci	**Cierre**			
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	R		A
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	R	C	
Iteración 1				
D.EN	**Entendimiento del Negocio**			
D.EN.AnN	Análisis del Negocio	R	E	C
D.EN.CPN	Comprensión del Problema de Negocio	R	E	C
D.ED	**Entendimiento de los Datos**			
D.ED.AnD	Análisis de los Datos	I	R	C
D.ED.ExD	Exploración de los Datos	I	R	
D.ED.EvD	Evaluación de los Datos	I	R	C
D.Mo	**Modelado**			
D.Mo.MoP	Modelado del problema	I	R	
D.Mo.CoM	Configuración del Modelo	C/A	R	
D.PD	**Preparación de los Datos**			
D.PD.AFI	Construcción de la Fuente Temporaria de Datos	I	R	
D.PD.CFT	Adecuación de la Fuente Temporaria de Datos	I	R	
D.Im	**Implementación**			
D.Im.SeM	Selección del Modelo	R	I	
D.Im.ExI	Explotación de Información	C/A	R	I
D.EP	**Evaluación y Presentación**			
D.EP.EvR	Evaluación de los Resultados	R	E	C
D.EP.PrR	Presentación de los Resultados	-	-	-

Tabla 5.11.a. Caso de Validación: Web Log - Matriz de Responsabilidades

Iteración 2				
D.EN	**Entendimiento del Negocio**			
D.EN.AnN	Análisis del Negocio	R	E	C
D.EN.CPN	Comprensión del Problema de Negocio	R	E	C
D.ED	**Entendimiento de los Datos**			
D.ED.AnD	Análisis de los Datos	I	R	C
D.ED.ExD	Exploración de los Datos	I	R	
D.ED.EvD	Evaluación de los Datos	I	R	C
D.Mo	**Modelado**			
D.Mo.MoP	Modelado del problema	I	R	
D.Mo.CoM	Configuración del Modelo	C/A	R	
D.PD	**Preparación de los Datos**			
D.PD.AFI	Construcción de la Fuente Temporaria de Datos	I	R	
D.PD.CFT	Adecuación de la Fuente Temporaria de Datos	I	R	
D.Im	**Implementación**			
D.Im.SeM	Selección del Modelo	R	I	
D.Im.ExI	Explotación de Información	R	E	I
D.EP	**Evaluación y Presentación**			
D.EP.EvR	Evaluación de los Resultados	R	E	C
D.EP.PrR	Presentación de los Resultados	R	E	I

Tabla 5.11.b. Caso de Validación: Web Log - Matriz de Responsabilidades

Propuesta del Proyecto (G.Pl.PlR.PrPr): a partir del Objetivo del Proyecto, el Problema del Negocio y sus Criterios de Éxito asociados, se describe como alcance del proyecto: "*Mejorar la experiencia del usuario mediante la identificación de patrones en la navegación del sitio, facilitando el acceso de los recursos, de acuerdo al perfil de interés de los usuarios.*"

En la sección de Obligaciones y Responsabilidades, se determinan los compromisos asumidos por las partes intervinientes, los cuales se derivan de los formalismos Matriz de Responsabilidades, Plan de Comunicación y el Plan de Acción: "*La parte contratante se compromete a brindar disposición de todos los recursos requeridos en tiempo y forma, informando con una antelación no menor a 48hs en caso que estos no pudiesen ser entregados. Cualquier demora fuera del plazo estipulado, requerirá del ajuste de los plazos del proyecto. La contraparte se compromete a dar solución a las problemáticas requeridas por el cliente (ver sección Alcance), así como la veracidad de los resultados presentados. Asimismo, la organización se compromete a informar de forma bimensual los avances del proyecto. Las partes acuerdan como fecha de finalización del proyecto el 10/05/2017.*"

La sección de aspectos legales fue omitida dado que no existe ninguna consideración de dicho tipo. La tabla 5.12 ilustra la información previamente mencionada registrada en el formalismo.

Propuesta del Proyecto			
Responsable:	Sebastian M.	**Fecha:**	01/03/2017
ID#:	G.Pl.PlR.PrPr	**Versión:**	1.0
Alcance	Mejorar la experiencia del usuario mediante la identificación de patrones en la navegación del sitio, facilitando el acceso de los recursos, de acuerdo al perfil de interés de los usuarios.		
Obligaciones y responsabilidades	La parte contratante se compromete a brindar disposición de todos los recursos requeridos en tiempo y forma, informando con una antelación no menor a 48hs en caso que estos no pudiesen ser entregados. Cualquier demora fuera del plazo estipulado, requerirá del ajuste de los plazos del proyecto. La contraparte se compromete a dar solución a las problemáticas requeridas por el cliente (ver sección Alcance), así como la veracidad de los resultados presentados. Asimismo, la organización se compromete a informar de forma bimensual los avances del proyecto. Las partes acuerdan como fecha de finalización del proyecto el 10/05/2017.		
Firma del Contratante: Dario R. **Aclaración:** Dario R.		**Firma de la Contraparte:** Sebastian M. **Aclaración:** Sebastian M.	

Tabla 5.12. Caso de Validación: Web Log - propuesta del Proyecto

5.1.1.3. Fase: Soporte (G.So)

La fase Soporte se encuentra conformada por dos actividades: Mediciones del Proyecto (sección 5.1.1.3.1) y Gestión de la Configuración (sección 5.1.1.3.2).

5.1.1.3.1. Actividad: Mediciones del Proyecto (G.So.MeP)

En esta actividad se calculan las métricas durante el desarrollo del proyecto, dejando registro formal del progreso de los indicadores. El resultado de esta actividad contribuye en la toma de decisiones del líder del proyecto, así como en la evaluación de la calidad del proceso y/o del modelo.

A continuación se presentan los resultados obtenidos de aplicar la técnica Cálculo de Métricas (sección 4.3.3.1.2, pág. 133), que utiliza como insumos el Listado de Métricas (Tabla 5.6) y Plan de Acción (Tabla 5.9).

Registro de Mediciones (G.So.MeP.ReMe): como se definió en el plan de acción, el registro de mediciones se realiza al finalizar cada iteración. El resultado de la primera de ellas, se exhibe en la Tabla A.8 (sección A.2.1). El resultado obtenido de cada uno de los indicadores, una vez finalizada la segunda iteración (y por consiguiente en el fin del proyecto) se ilustra en la tabla 5.13.

Registro de Mediciones			
Responsable:	Sebastian M.	**Fecha:**	11/05/2017
ID#:	G.So.MeP.ReMe	**Versión:**	1.1

Indicador	Medición	Descripción
Tiempo total requerido para el desarrollo del proyecto	**DRPY** = 486hs	Tdesarrollo = 398 Tgestion = 88
Tiempo medio requerido para el desarrollo de un problema de explotación de información	**DRPEI** = 132.6hs	Tdesarrollo(iter1) = 251 Tdesarrollo(iter2) = 147 NPEI = 3
Número medio de atributos significativos por modelo	4,33	AtS.M(prne.1) = 4 AtS.M(prne.2) = 4 AtS.M(prne.4) = 5 NMOD = 3

Tabla 5.13. Caso de Validación: Web Log - Registro de Mediciones (fin del proyecto)

5.1.1.3.2. Actividad: Gestión de la Configuración (G.So.GeC)

En esta actividad se realizan las tareas vinculadas con la trazabilidad de los productos generados durante el desarrollo del proyecto, garantizando que en todo momento los miembros del equipo estén informados de las versiones actuales de los resultados producidos en cada fase.

A continuación se presentan los resultados obtenidos de aplicar la técnica Configuración del versionado (sección 4.3.3.2.2, pág. 137). Dicha técnica utiliza como insumos las Reglas de Versionado (Fuente de Información 5.1), existentes en la organización, Reporte de Evaluación del Cambio (no aplicable en este proyecto) y Modelo de Ciclo de Vida (Tabla 5.5).

Se utilizan dos dígitos para reflejar el progreso de los productos a lo largo del proyecto X.Y: el primero de ellos (X) indica la versión mayor del documento, incrementándose de a uno cada vez que se modifican o eliminan elementos del producto (generando incompatibilidad con otros productos o versiones anteriores). En caso que el producto se encuentre en un estadio temprano, el cual no puede ser utilizado para su uso como entrada en otras tareas, este dígito debe ser indicado con cero. El segundo (Y), describe la versión menor del documento, la cual se incrementa en uno reflejando incorporaciones o alteraciones que no modifiquen la funcionalidad en el producto. Cuando se modifique la versión superior, este valor será restituido al valor cero.
Para el registro del estado del proyecto, se utiliza la misma lógica (X.Y). El primer elemento (X) indica alteraciones en las necesidades o estrategias de ejecución del proyecto, mientras que el segundo (Y) representa iteraciones en el ciclo de vida.

Fuente de Información 5.1. Caso de Validación: Web Log - Reglas de Versionado

Reporte de Versionado (G.So.GeC.ReVe): durante el desarrollo del proyecto se realizaron siete ajustes a sus productos internos, a partir del progreso de las actividades y las tareas de control. La tabla 5.14 ilustra los resultados registrados para el caso de validación.

Reporte de Versionado					
Responsable:	Sebastian M.		**Fecha:**	11/05/2017	
ID#:	G.So.GeC.ReVe				
Fecha	**Elemento**	**Versión previa**	**Versión Actual**	**Cambio Asociado**	**Descripción**
20/02/2017	Restricciones del Proyecto	1.0	1.1	incorporación de restricciones	Se incorporación de restricciones (repr.2 y repr.3) a partir de la ampliación del conocimiento obtenida de la entrevista 3
28/03/2017	Plan de Acción	1.0	1.1	Actualización del plan del proyecto fin de la iteración 1	Se incorporaron las fechas y esfuerzos reales
29/03/2017	Problema del Negocio	1.0	1.1	Validación de los resultados obtenidos para los problemas prne.1 y prne.2	Incorporación de problemas de negocio prne.3 y prne.4
31/03/2017	Criterios de Éxito del Problema de Negocio	1.0	1.1	Validación de los resultados obtenidos para los problemas prne.1 y prne.2	Incorporación de los criterios de éxito cepn.3 y cepn.4
05/05/2017	Reporte de Evaluación de los Resultados	1.0	1.1	Validación del problema de negocio prne.4 (iteración 2)	Incorporación de los resultados de la validación del problema de negocio prne.4
11/05/2017	Registro de Mediciones	1.0	1.1	Actualización de las mediciones al cierre del proyecto	Actualización de las métricas
11/05/2017	Plan de Acción	1.1	1.2	Actualización del plan del proyecto fin del proyecto	Se incorporaron y ajustan las fechas y esfuerzos reales

Tabla 5.14. Caso de Validación: Web Log - Reporte de Versionado

Informe del Estado de la Configuración (G.So.GeC.InEC): en las tablas 5.15.a y 5.15.b, se observan los estados de configuración de las versiones internas de los productos del proyecto, dejando registro sus alteraciones, permitiendo a los miembros del equipo estar notificados de los avances del mismo. La versión 1.0 está asociada con el desarrollo de los productos durante la primera iteración del proyecto, mientras que la versión 1.1 está asociada con la versión final del proyecto (finalización de la segunda iteración y cierre del mismo).

Informe del Estado de la Configuración					
Responsable:	Sebastian M.		**Fecha:**	11/05/2017	
ID#:	G.So.GeC.InEC				
ID Actividad	**Actividad**	**Elemento**		**Versión del Proyecto**	
				V. 1.0	**V. 1.1**
G.In	**Iniciación**				
G.In.EIP	Exploración Inicial del Proyecto	Recursos Humanos Involucrados Riesgos del Proyecto Plan de Contingencias		1.0 - -	1.0 - -
G.In.DeC	Definición de la Comunicación	Plan de Comunicación		1.0	1.0
G.In.EvS	Evaluación de la Situación	Reporte de Evaluación de Herramientas Reporte de Evaluación de Viabilidad		1.0 1.0	1.0 1.0
G.In.DCV	Definición del Ciclo de Vida	Modelo de Ciclo de Vida		1.0	1.0

Tabla 5.15.a. Caso de Validación: Web Log - Informe de Estado de la Configuración

G.Pl	**Planificación**			
G.Pl.PIM	Planificación de la Mediciones	Listado de Métricas Estimación del Proyecto	1.0 1.0	1.0 1.0
G.Pl.PIA	Planificación de las Actividades	Mapa de Actividades Plan de Acción	1.0 1.1	1.0 1.2
G.Pl.PIR	Planificación de los Recursos	Plan de Necesidad de Recursos	1.0	1.0
G.Pl.PRe	Planificación de las Responsabilidades	Matriz de Responsabilidades Propuesta del Proyecto	1.0 1.0	1.0 1.0
G.So	**Soporte**			
G.So.MeP	Mediciones del Proyecto	Registro de Mediciones	1.0	1.1
G.So.GeC	Gestión de la Configuración	Reporte de Versionado Informe del Estado de la Configuración	- -	- -
G.Co	**Control**			
G.Co.GeD	Gestión del Desarrollo	Reporte de Estado	-	-
G.Co.CoA	Control de las Actividades	Registro de Riesgos Acontecidos	-	-
G.Co.Gca	Gestión del Cambio	Reporte de Evaluación del Cambio	-	-
G.Ci	**Cierre**			
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	Documento de Aceptación	-	-
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	Reporte de Cierre	-	-
D.EN	**Entendimiento del Negocio**			
D.EN.AnN	Análisis del Negocio	Fuentes de Información del Cliente Definiciones, Acrónimos y Abreviaciones Objetivos del Proyecto Criterios de Éxito del Proyecto Expectativas del Proyecto Suposiciones del Proyecto Restricciones del Proyecto	1.0 1.0 1.0 1.0 1.0 1.0 1.1	1.0 1.0 1.0 1.0 1.0 1.0 1.1
D.EN.CPN	Comprensión del Problema de Negocio	Problema de Negocio Criterios de Éxito del Problema de Negocio	1.0 1.0	1.1 1.1
D.ED	**Entendimiento de los Datos**			
D.ED.AnD	Análisis de los Datos	Diccionario de Fuente de Datos Campos Relacionados con el Problema de Negocio	1.0 1.0	1.0 1.0
D.ED.ExD	Exploración de los Datos	Reporte de Datos Explorados Fuente Integrada de datos	1.0 -	1.0 -
D.ED.EvD	Evaluación de los Datos	Reporte de la Calidad de los Datos	1.0	1.0
D.Mo	**Modelado**			
D.Mo.MoP	Modelado del problema	Diseño del Proceso de Explotación de Información	1.0	1.0
D.Mo.CoM	Configuración del Modelo	Selección de Algoritmos de Explotación de Información Selección de Variables del Modelo Estrategias de Evaluación de Modelos	1.0 1.0 1.0	1.0 1.0 1.0
D.PD	**Preparación de los Datos**			
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	Reporte de Generación de la Fuente Temporaria de datos Fuente Temporaria de Datos	1.0 -	1.0 -
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos	Reporte de Adecuación de la Fuente Temporaria de Datos	1.0	1.0
D.Im	**Implementación**			
D.Im.SeM	Selección del Modelo	Reporte de Estrategia de Parametrización del Modelo	1.0	1.0
D.Im.ExI	Explotación de Información	Reporte de Implementación del Modelo	1.0	1.0
D.EP	**Evaluación y Presentación**			
D.EP.EvR	Evaluación de los Resultados	Reporte de Evaluación de los Resultados	1.0	1.0
D.EP.PrR	Presentación de los Resultados	Reporte del Proyecto	1.0	1.0

Tabla 5.15.b. Caso de Validación: Web Log - Informe de Estado de la Configuración

5.1.1.4. Fase: Control (G.Co)

Esta fase está conformada por tres actividades: Gestión del Desarrollo (sección 5.1.1.4.1), Control de las Actividades (sección 5.1.1.4.2) y Gestión del Cambio (sección 5.1.1.4.3).

5.1.1.4.1. Actividad: Gestión del Desarrollo (G.Co.GeD)

En esta actividad se realiza el seguimiento del proyecto, dejando registro formal del progreso del mismo. El resultado de esta actividad contribuye en la toma de decisiones del líder del proyecto, con respecto al cumplimiento de lo planificado, pudiendo identificarse la necesidad de reajustar las acciones programadas.

A continuación se presentan los resultados obtenidos de aplicar la técnica Seguimiento de Avance (sección 4.3.4.1.2, pág. 143) en el proyecto, utilizando como insumos: el Plan de Acción (Tabla 5.9) y el Registro de Mediciones (Tabla 5.13).

Reporte de Estado (G.So.GeD.ReEs): a partir de lo expuesto en el plan de acción, al finalizar cada iteración del proyecto, se lleva a cabo el control del progreso del proyecto obteniendo como resultado el reporte de estado del mismo hasta la fecha de ejecución. El día 28/03/2017 se realizó el control de avance del proyecto, cuyo resultado se expone en la tabla 5.16. El segundo reporte de estado, se realizó al finalizar el proyecto (11/05/2017) registrando su resultado en la tabla 5.17. Se destaca que fueron registradas las mediciones de aquellas fases que al momento del reporte habían finalizado.

5.1.1.4.2. Actividad: Control de las Actividades (G.Co.CoA)

En esta actividad se evalúan las situaciones potencialmente peligrosas para el desarrollo del proyecto, realizando un seguimiento, control y registro de acontecimientos, así como de las acciones realizadas. El resultado de esta actividad contribuye en la calidad del proceso.

Registro de Riesgos Acontecidos (G.Co.CoA.ReRA): durante el desarrollo del caso de validación seleccionado no han acontecido riesgos.

Reporte de Estado			
Responsable:	Sebastian M.	**Fecha:**	28/03/2017
ID#:	G.Co.GeD.ReEs.1		
Evaluación del Programa			
Global	**-7.04%**		**(% por debajo de lo planificado)**
Desarrollo (iter. 1)	**-7.04%**	**Gestión**	**-**
Entendimiento del Negocio	0.0%	Iniciación	-
Entendimiento de los Datos	2.94%	Planificación	-
Modelado	-9.52%	Soporte	-
Preparación de los Datos	-30.0%	Control	-
Implementación	-5.36%	Cierre	-
Evaluación y Presentación	-14.29%		
Descripción: *Se evaluaron las actividades finalizadas hasta la fecha. (ver G.Pl.PIA.PIAc versión 1.1)*			
Situaciones identificadas que requieren de seguimiento - **Cambios durante el periodo (alcances, tiempos)** - **logros principales durante el periodo** - Se identificaron tres perfiles de usuarios - Se definieron dos nuevas problemáticas asociadas al objetivo (obpr.1)			

Tabla 5.16. Caso de Validación: Web Log - Reporte de Estado (G.Co.GeD.ReEs.1)

Reporte de Estado			
Responsable:	Sebastian M.	**Fecha:**	11/05/2017
ID#:	G.Co.GeD.ReEs.2		
Evaluación del Programa			
Global	**-8,30%**		**(% por debajo de lo planificado)**
Desarrollo (iter. 2)	**-20,11%**	**Gestión**	**15,79%**
Entendimiento del Negocio	-23,08%	Iniciación	12,50%
Entendimiento de los Datos	-2,50%	Planificación	56,25%
Modelado	-26,67%	Soporte	5,56%
Preparación de los Datos	-46,15%	Control	-9,09%
Implementación	-28,57%	Cierre	50,00%
Evaluación y Presentación	-5,88%		
Descripción:			
Situaciones identificadas que requieren de seguimiento - **Cambios durante el periodo (alcances, tiempos)** Se descartó una de las problemáticas a partir de la exploración de los datos **logros principales durante el periodo** Definición de los alcances del proyecto y los problemas de negocio			

Tabla 5.17. Caso de Validación: Web Log - Reporte de Estado (G.Co.GeD.ReEs.2)

5.1.1.4.3. Actividad: Gestión del Cambio (G.Co.Gca)

En la actividad de Gestión del Cambio se realiza un proceso de evaluación formal de las peticiones de modificación del proyecto, determinando como resultado la procedencia o no de la misma y sus efectos asociados.

Reporte de Evaluación del Cambio (G.Co.Gca.RECa): durante el desarrollo del proyecto, no se han realizado peticiones de cambios por parte del cliente, sin embargo, a partir de las suposiciones refutadas, se realizó una revaluación de la planificación del proyecto, concluyéndose que no era necesario ajustar la misma.

5.1.1.5. Fase: Cierre (G.Ci)

La fase Cierre está compuesta por dos actividades: Formalización Externa del Cierre del Proyecto (sección 5.1.1.5.1) y Formalización Interna del Cierre del Proyecto (sección 5.1.1.5.2).

5.1.1.5.1. Actividad: Formalización Externa del Cierre del Proyecto (G.Ci.FEC)

En esta actividad se obtiene la conformidad del cliente, respecto a los compromisos asumidos en la propuesta del proyecto, dejando registro formal de la finalización del mismo.

A continuación se presentan los resultados obtenidos de aplicar la técnica Presentación de Conformidad (sección 4.3.5.1.2, pág. 156), la cual utiliza como insumos los formalismos: Reporte de Evaluación de los Resultados (Tabla 5.59), Registro de Riesgos Acontecidos (no aplicable en este proyecto), Plan de Acción (Tabla 5.9) y Propuesta del Proyecto (Tabla 5.12).

Documento de Aceptación (G.Ci.FEC.DoAc): a partir de la propuesta del proyecto y el reporte de evaluación de los resultados, se definen los objetivos abordados en el proyecto. La sección programa, se determina a partir del plan de acción y las obligaciones y responsabilidades acordadas en la propuesta del proyecto. Finalmente, las conclusiones se derivan principalmente del reporte de evaluación de los resultados y de la evaluación del cumplimiento de las obligaciones acordadas, confirmándose el cumplimiento exitoso del proyecto. En la tabla 5.18 se ilustra el formalismo resultante.

Documento de Aceptación	
	Fecha: 10/05/2017
Objetivos	El alcance del proyecto consiste en mejorar la experiencia del usuario mediante la identificación de patrones frecuentes (15-20% de soporte) en la navegación del sitio, facilitando el acceso de los recursos, de acuerdo al perfil de interés de los usuarios.
Programa	Se realizaron informes del progreso de las actividades de manera bimensual y se finalizó el proyecto el día 10/05/2017, cumplimentando con las obligaciones acordadas.
Conclusiones	Como resultado de la evaluación se concluyó que los patrones y perfiles identificados proveen aportes de interés, para profundizar en la comprensión del comportamiento de los usuarios del sitio web, mejorando el acceso y presentación de los recursos. El proyecto fue finalizado acorde a lo estipulado, en tiempo y forma. Mediante la presente se deja de manifiesto que se ha cumplimentado exitosamente los requerimientos realizados, dando por finalizado el proyecto.
Firma: Dario R. **Aclaración:** Dario R.	

Tabla 5.18. Caso de Validación: Web Log - Documento de Aceptación

5.1.1.5.2. Actividad: Formalización Interna del Cierre del Proyecto (G.Ci.FIC)

En esta actividad se llevan a cabo las últimas tareas del proyecto, en el cual se evalúa la performance del equipo de trabajo, la propuesta, las acciones realizadas y el cumplimiento del plan de acción. Como resultado de esta actividad se resume el progreso del proceso, dejando registro de aquellos aspectos que sean de valor para futuros proyecto.

A continuación se presentan los resultados obtenidos de aplicar la técnica Evaluación del Proceso (sección 4.3.5.2.2, pág. 162), la cual utiliza como insumos los formalismos: Plan de Acción (Tabla 5.9), Matriz de Responsabilidades (Tabla 5.11), Registro de Mediciones (Tabla 5.13), Registro de Riesgos Acontecidos (no aplicable en este proyecto), Reporte de Evaluación del Cambio (no aplicable en este proyecto), Reporte de Evaluación de los Resultados (Tabla 5.59) y Documento de Aceptación (Tabla 5.18).

Reporte de Cierre (G.Ci.FIC.ReCi): se describe el objetivo acordado y la validación realizada por el cliente de los resultados obtenidos, y se evalúa el progreso del proyecto con respecto a los tiempos planificados, identificando desvíos de subestimación para el subproceso Gestión y sobreestimación en Desarrollo, asociado con la desestimación de la suposición (supr.1). De la evaluación de los resultados del proyecto, se identifica como principal desafío: "*Estimación de los tiempos: el modelo utilizado para determinar la carga de trabajo sobreestimó en aproximadamente un 12% el esfuerzo requerido para el proceso de desarrollo, mientras que el criterio utilizado para el proceso de gestión, subestimo los tiempos requeridos para sus actividades casi en un 16%.",* mientras que no se identifican lecciones aprendidas (sección omitida en el reporte de cierre). La tabla 5.19 sintetiza aspectos de interés para la organización resultantes del cierre del proyecto.

Reporte de Cierre			
Responsable:	Sebastian M.	**Fecha:**	11/05/2017
ID#:	G.Ci.FIC.ReCi		

Objetivos del Proyecto

Objetivos	Resultados
Optimizar la experiencia del usuario a partir del análisis de sus acciones en el sitio web, facilitando su uso y mejorando la disposición de los contenidos de interés	Se identifican tres perfiles de usuarios de acuerdo a los recursos que acceden durante su navegación. Se identificaron rutas frecuentes de navegación globales y específicas por perfiles, las cuales permiten comprender y mejorar la experiencia de los usuarios a partir de la disposición de recursos asociados con las características en común con otros usuarios.

Evaluación del Tiempo (en HS.)

Hito	Estimado	Real	% desvío	Motivo
Gestión	**76**	**88**	**15,79%**	
Iniciación	16	18	12,50%	
Planificación	16	25	56,25%	
Soporte	18	19	5,56%	
Control	22	20	-9,09%	
Cierre	4	6	50,00%	
Desarrollo	**530**	**486**	**-12,33%**	
Entendimiento del Negocio	76	70	-7,89%	
Entendimiento de los Datos	108	109	0,93%	
Modelado	72	60	-16,67%	Descarte de la problemáticas de negocio (prne.3)
Preparación de los Datos	66	42	-36,36%	Descarte de la problemáticas de negocio (prne.3)
Implementación	84	73	-13,10%	Descarte de la problemáticas de negocio (prne.3)
Evaluación y Presentación	48	44	-8,33%	Descarte de la problemáticas de negocio (prne.3)
TOTAL	**530**	**486**	**-8,30%**	

Principales Desafíos

- Estimación de los tiempos: el modelo utilizado para determinar la carga de trabajo sobreestimó en aproximadamente un 12% el esfuerzo requerido para el proceso de desarrollo, mientras que el criterio utilizado para el proceso de gestión, subestimo los tiempos requeridos para sus actividades casi en un 16%.

Tabla 5.19. Caso de Validación: Web Log - Reporte de Cierre

5.1.2. MoProPEI-D: Subproceso Desarrollo (D)

El subproceso Desarrollo, se encuentra conformado por seis fases: Entendimiento del Negocio (sección 5.1.2.1), Entendimiento de los Datos (sección 5.1.2.2), Modelado (sección 5.1.2.3), Preparación de los Datos (sección 5.1.2.4), Implementación (sección 5.1.2.5), y Evaluación y presentación (sección 5.1.2.6).

5.1.2.1. Fase: Entendimiento del Negocio (D.EN)

La fase de entendimiento del negocio se compone de dos actividades: Análisis del Negocio (sección 5.1.2.1.1), donde se identifican las características generales del proyecto y Comprensión del Problema de Negocio (sección 5.1.2.1.2), en la cual se establecen los problemas a resolver.

5.1.2.1.1. Actividad: Análisis del Negocio (D.EN.AnN)

El objetivo de esta actividad es identificar y comprender las metas generales del proyecto, en base a las necesidades del requirente y los interesados. A continuación se presentan los resultados obtenidos de aplicar la técnica de definición de los objetivos del proyecto (sección 4.4.1.1.2, pág. 175). Los mismos se obtuvieron a partir del conocimiento extraído en las distintas entrevistas con el cliente (Fuente de Información 5.2.a y 5.2.b) y del análisis del sitio web.

Discursos de los interesados: Reuniones con el cliente (Dario R.)

Entrevista 1:

El Campus Virtual de la UNLa se configura como un entorno de enseñanza y de aprendizaje en el que se produce un encuentro académico entre profesores y estudiantes independientemente de la situación real de tiempo y espacio. Esta particularidad abre nuevas oportunidades de desarrollo curricular a través de la utilización de una plataforma educativa que se actualiza periódicamente.
El proyecto surge a partir del interés en aplicar nuevas tecnologías al sitio web (http://campus.unla.edu.ar) de la institución con el objetivo de realizar acciones que mejoren la experiencia del usuario. Mediante este, se provee de recursos, herramientas y servicios educativos a la comunidad, pensados principalmente para aquellos docentes y estudiantes pertenecientes a la comunidad educativa.
El sitio web, está conformado principalmente por cuatro secciones generales: centro de ayuda, en la cual se guía a los usuarios sobre las aulas virtuales y las distintas herramientas disponibles, Docente, se orienta al docente respecto a la virtualización y el enriquecimiento de las prácticas didácticas, Capacitación, se provee asesoramiento a los usuarios, y Novedades, se brindan noticias relacionadas con distintas propuestas y colaboraciones de la comunidad, como participaciones en congresos, capacitaciones, proyectos, etc.

¿Cuáles son sus intereses y expectativas con respecto a la mejora de la experiencia del usuario?
Se desea estudiar el uso que realizan los usuarios a la página web esperando obtener un mayor conocimiento de las características de navegación, con el objetivo de mejorar la estructuración de los contenidos, facilitando y mejorando la experiencia del usuario.

¿Cuáles son las condiciones a partir de las cuales consideraría exitoso el resultado del proyecto?
No hay un criterio que tenga en consideración, sino el poder identificar intereses comunes entre los usuarios que nos permitan identificar su posible comportamiento en la página.

¿Qué información se disponible para el desarrollo del proyecto?
Si bien el Campus Virtual, dispone principalmente de dos sitios: las aulas virtuales y el sitio web, en este proyecto nos interesa estudiar el comportamiento de los usuarios en el sitio web, para el cual se dispone de los registros de navegación (logs) de los últimos dos años.

¿Qué limitaciones entiende que poseen los datos?
El sitio web, se encuentra desarrollado sobre la plataforma wordpress, utilizando un plugin que deja registro, como previamente comente, de los accesos de los usuarios, junto con otra información del dispositivo utilizado y demás.

Entrevista 2:

¿Identifica alguna problemática de interés respecto a la mejora de la experiencia del usuario?
No en particular, sin embargo, nos gustaría comprender mejor las necesidades que cada usuario posee, pudiendo brindarle contenido de manera más eficiente.

Fuente de Información 5.2.a. Caso de Validación: Web Log - Entrevistas 1 y 2

¿Poseen un sistema de registro de usuarios?
No, los datos de navegación son anónimos, pero se dispone de campos para identificar un mismo usuario.

¿Desde qué fecha se encuentra la última versión del sitio web en vigencia?
La versión actual fue puesta en producción los primeros días de febrero del año 2016. Sin embargo, la estructura global de los contenidos se mantiene estable desde los últimos tres años.

Fuente de Información 5.2.b. Caso de Validación: Web Log - Entrevistas 1 y 2

Fuentes de Información del Cliente (D.EN.ANN.FUIC): a partir de la interacción con el cliente, se identifica como fuente de información los registros de logs de usuarios en wordpress, del tipo almacén de datos, siendo el único miembro externo del proyecto el responsable. La tabla 5.20 ilustra el resultado obtenido.

Fuentes de Información del Cliente					
Responsable:		Sebastian M.		**Fecha:**	08/02/2017
ID#:		D.EN.ANN.FUIC		**Versión:**	1.0
ID	**Nombre**	**Categoría**	**Responsable**	**Descripción**	
fuic.1	Wordpress Logs	Almacén de datos	(rehi.3) Dario R.	Registro de los usuarios que accedieron a la página web, junto con otra información del dispositivo utilizado y demás, correspondiente a los últimos dos años.	

Tabla 5.20. Caso de Validación: Web Log - Fuentes de Información del Cliente

Definiciones, Acrónimos y Abreviaciones (D.EN.ANN.DEAA): En esta etapa se propone el registro de aquellas terminologías específicas del dominio que no sean familiares para el equipo de trabajo, identificándose a partir del siguiente párrafo: *"...centro de ayuda, en la cual se guía a los usuarios sobre las aulas virtuales y las distintas herramientas disponibles, Docente, se orienta al docente respecto a la virtualización y el enriquecimiento de las prácticas didácticas, Capacitación, se provee asesoramiento a los usuarios, y Novedades, se brindan noticias relacionadas con distintas propuestas y colaboraciones de la comunidad, como participaciones en congresos, capacitaciones, proyectos, etc…*" cuatro términos de interés, que representan los contenidos de las secciones principales del dominio. En la tabla 5.21 se ilustran los términos identificados y su descripción.

Objetivos del Proyecto (D.EN.ANN.OBPR): A partir de la interacción con el cliente, se identifica de interés el siguiente párrafo: "…*estudiar el uso que realizan los usuarios a la página web esperando obtener un mayor conocimiento de las características de navegación, con el objetivo de mejorar la estructuración de los contenidos, facilitando y mejorando la experiencia del usuario...*" en el cual se señala el objetivo general del cliente. Se registra como "Obpr.1" en la columna objetivo, reescribiéndose (en la columna descripción) con el propósito de mejorar la comprensión del mismo: "*Optimizar la experiencia del usuario a partir del análisis de sus acciones en el sitio*

web, facilitando su uso y mejorando la disposición de los contenidos de interés". Finalmente, se registra en la columna referencia, que el mismo se obtuvo en la entrevista 1. En la tabla 5.22 se presenta el formalismo resultante.

Definiciones, Acrónimos y Abreviaciones			
Responsable:	Sebastian M.	**Fecha:**	08/02/2017
ID#:	D.EN.AnN.DeAA	**Versión:**	1.0
Nombre	**Descripción**	**Tipo**	**Referencia**
Centro de ayuda	Sección que guía a los usuarios sobre las aulas virtuales y las distintas herramientas disponibles	Definición	Entrevista 1
Docente	Sección donde se orienta al docente respecto a la virtualización y el enriquecimiento de las prácticas didácticas	Definición	Entrevista 1
Capacitación	Sección en la cual se provee asesoramiento a los usuarios	Definición	Entrevista 1
Novedades	Sección donde se brindan noticias relacionadas con distintas propuestas y colaboraciones de la comunidad, como participaciones en congresos, capacitaciones, proyectos, etc.	Definición	Entrevista 1

Tabla 5.21. Caso de Validación: Web Log - Definiciones, Acrónimos y Abreviaciones

Objetivos del Proyecto			
Responsable:	Sebastian M.	**Fecha:**	08/02/2017
ID#:	D.EN.ANN.OBPR	**Versión:**	1.0
Objetivo	**Descripción**		**Referencia**
obpr.1	Optimizar la experiencia del usuario a partir del análisis de sus acciones en el sitio web, facilitando su uso y mejorando la disposición de los contenidos de interés		Entrevista 1

Tabla 5.22. Caso de Validación: Web Log - Objetivos del Proyecto

Criterios de Éxito del Proyecto (D.EN.ANN.CREP): A partir del objetivo previamente identificado y de los extractos obtenidos de la primera entrevista con el cliente: "*No hay un criterio que tenga en consideración, sino el poder identificar intereses comunes entre los usuarios que nos permitan identificar su posible comportamiento en la página…*" se reescribe el criterio identificado para facilitar su comprensión a "*Identificar relaciones entre la navegación de los usuarios los cuales favorezcan la comprensión de su comportamiento en la página web*", asignándole el identificador "crexpr.1". El criterio de éxito se encuentra asociado al único objetivo del proyecto (Obpr.1), obteniendo esta información de la entrevista 1 realizada al cliente. Dicha información se registra en las columnas objetivo asociado y referencia respectivamente. En la tabla 5.23 se presenta el resultado obtenido.

Criterios de Éxito del Proyecto			
Responsable:	Sebastian M.	**Fecha:**	08/02/2017
ID#:	D.EN.ANN.CREP	**Versión:**	1.0

Criterio	Descripción	Objetivo asociado	Referencia
crexpr.1	Identificar relaciones entre la navegación de los usuarios los cuales favorezcan la comprensión de su comportamiento en la página web	(obpr.1) Optimizar la experiencia del usuario a partir del análisis de sus acciones en el sitio web, facilitando su uso y mejorando la disposición de los contenidos de interés	Entrevista 1

Tabla 5.23. Caso de Validación: Web Log - Criterios de Éxito del Proyecto

Expectativas del Proyecto (D.EN.ANN.EXPR): Las expectativas del proyecto presentan una visión complementaria a la definición del objetivo asociado, en el cual se vinculan las pretensiones que tienen los clientes/expertos con respecto al producto resultante como respuesta a cada objetivo de proyecto identificado. En el caso de validación se identifica como expectativa asociadas al objetivo del proyecto: *"...mejorar la estructuración de los contenidos, facilitando y mejorando la experiencia del usuario."*. En la tabla 5.24 se presentan las expectativas asociadas al objetivo del proyecto.

Expectativas del Proyecto			
Responsable:	Sebastian M.	**Fecha:**	08/02/2017
ID#:	D.EN.ANN.EXPR	**Versión:**	1.0

Expectativa	Descripción	Objetivo asociado	Referencia
expr.1	Renovar la estructuración de los contenidos, facilitando y mejorando la experiencia del usuario	(obpr.1) Optimizar la experiencia del usuario a partir del análisis de sus acciones en el sitio web, facilitando su uso y mejorando la disposición de los contenidos de interés	Entrevista 1

Tabla 5.24. Caso de Validación: Web Log - Expectativas del Proyecto

Suposiciones del Proyecto (D.EN.ANN.SUPR): a partir de las nuevas necesidades identificadas en la segunda iteración del proyecto, se identifica en la entrevista 5 (fuente de información 5.5) como suposición, que: "el comportamiento de navegación de los usuarios varían según el tipo de dispositivo que utiliza para acceder el sitio". La tabla 5.25 ilustra el formalismo resultante.

Suposiciones del Proyecto			
Responsable:	Sebastian M.	**Fecha:**	30/03/2017
ID#:	D.EN.AnN.SuPr	**Versión:**	1.0

Suposición	Descripción	Objetivo asociado	Referencia
supr.1	El comportamiento de navegación de los usuarios varían según el tipo de dispositivo que utiliza para acceder el sitio	(obpr.1) Optimizar la experiencia del usuario a partir del análisis de sus acciones en el sitio web, facilitando su uso y mejorando la disposición de los contenidos de interés	Entrevista 5

Tabla 5.25. Caso de Validación: Web Log - Suposiciones del Proyecto

Restricciones del Proyecto (D.EN.ANN.REPR): se identifican aquellos aspectos que presentan limitaciones para el cumplimiento de los objetivos del negocio, los cuales puedan demorar, afectar o imposibilitar el desarrollo de los mismos. Las limitaciones pueden estar asociadas al recurso humano (conocimiento de las técnicas o tecnologías, disponibilidades), a los datos (posibilidad de acceso, calidad) o a cuestiones técnicas del proyecto (hardware o software) u organización (aspectos políticos o legales). A partir de la segunda entrevista, se identifica el siguiente párrafo: *"...La versión actual fue puesta en producción los primeros días de febrero del año 2016. Sin embargo, la estructura global de los contenidos se mantiene estable desde los últimos tres años."*, a partir del cual se define la restricción en los registros que son de interés para el objetivo identificado (obpr.1). A partir de la información relevada en la entrevista 3 (fuente de información 5.3), se identifican dos nuevas restricciones también asociadas con los datos. En la tabla 5.26, se presentan las restricciones del proyecto (reescritas para mejorar su comprensibilidad). En la sección A.2.2, se presenta la primera versión (Tabla A.9).

Restricciones del Proyecto				
Responsable:		Sebastian M.	**Fecha:**	20/02/2017
ID#:		D.EN.ANN.REPR	**Versión:**	1.1
Restricción	**Tipo**	**Descripción**	**Objetivo asociado**	**Referencia**
repr.1	datos	La versión actual fue puesta en producción los primeros días de febrero del año 2016	(obpr.1) Optimizar la experiencia del usuario a partir del análisis de sus acciones en el sitio web, facilitando su uso y mejorando la disposición de los contenidos de interés	Entrevista 2
repr.2	datos	Serán consideradas aquellas navegaciones en las cuales el usuario haya realizado más de una petición (accedido a más de un recurso)	(obpr.1) Optimizar la experiencia del usuario a partir del análisis de sus acciones en el sitio web, facilitando su uso y mejorando la disposición de los contenidos de interés	Entrevista 3
repr.3	datos	Serán recursos considerados aquellos cuyo número de accesos sea mayor al 1% del total de peticiones realizadas.	(obpr.1) Optimizar la experiencia del usuario a partir del análisis de sus acciones en el sitio web, facilitando su uso y mejorando la disposición de los contenidos de interés	Entrevista 3

Tabla 5.26. Caso de Validación: Web Log - Restricciones del Proyecto (versión final)

5.1.2.1.2. Actividad: Comprensión del Problema de Negocio (D.EN.CPN)

En esta actividad, se presenta una visión detallada de preguntas-problema específicas que el cliente desea responder, las cuales permiten alcanzar los objetivos generales establecidos.

A continuación se presentan los resultados obtenidos de aplicar la técnica de definición del problema de negocio (sección 4.4.1.2.2, pág. 189), los cuales se obtuvieron a partir de la interacción con el interesado (Fuente de Información 5.3) y los formalismos derivados de la actividad previa.

Discursos de los interesados: Reunión con el cliente (Dario R.)

Entrevista 3 (Minuta):

La reunión tuvo como objetivo, colaborar con el cliente para definir las necesidades o problemáticas de interés vinculadas con el objetivo del negocio definido. A partir de ello, se definen los siguientes problemas de interés:

a) Detallar las posibles rutas más frecuentes de navegación que realizan los usuarios.
b) Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación

A continuación se listan los criterios de éxito definidos para cada uno de los problemas previamente mencionados:

a) Las rutas de navegación frecuentes sean representativas de al menos un 15% del total de usuarios,
b) La caracterización de los perfiles tenga una tasa de error inferior al 20%.

Para los problemas previamente mencionados, serán consideradas aquellas navegaciones en las cuales el usuario haya realizado más de una petición (accedido a más de un recurso), dado a que el cliente especifica que un gran número de usuarios accede mediante el sitio web, a las aulas virtuales. En adición, para el análisis se tendrán en consideración aquellos recursos cuyo número total de accesos sea superior al 1% del total de peticiones realizadas.

Fuente de Información 5.3. Caso de Validación: Web Log – Entrevista 3

Problema del Negocio (D.EN.CPN.PRNE): En la primera iteración del proyecto, se realizó una entrevista con el cliente (fuente de información 5.3), a partir de la cual se definieron dos problemas de negocio (prne.1 y prne.2). Como resultado de la validación del cliente (fuente de información 5.5), se identificaron dos nuevos problemas de interés asociados al objetivo "obpr.1" (prne.3 y prne.4). En la tabla 5.27, se detallan todos los problemas de negocio identificados durante el proyecto. En la tabla A.10, se ilustra la primera versión del formalismo.

Problema del Negocio			
Responsable:	Sebastian M.	**Fecha:**	31/03/2017
ID#:	D.EN.CPN.PRNE	**Versión:**	1.1
Objetivo del Proyecto	(obpr.1) Optimizar la experiencia del usuario a partir del análisis de sus acciones en el sitio web, facilitando su uso y mejorando la disposición de los contenidos de interés		

Problema	**Descripción**	**Experto**	**Referencia**
prne.1	Detallar las posibles rutas más frecuentes de navegación que realizan los usuarios	(rehi.3) Dario R.	Entrevista 3
prne.2	Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación	(rehi.3) Dario R.	Entrevista 3
prne.3	Determinar las rutas más frecuentes de navegación según el tipo de dispositivo utilizado para navegar	(rehi.3) Dario R.	Entrevista 5
prne.4	Determinar las rutas más frecuentes de navegación de cada perfil de usuario	(rehi.3) Dario R.	Entrevista 5

Tabla 5.27. Caso de Validación: Web Log - Problema del Negocio (versión final)

Criterios de Éxito del Problema de Negocio (D.EN.CPN.CEPN): a partir de la reunión con el cliente (Fuente de Información 5.3), se acordaron los criterios de éxito para los problemas definidos en la misma. Como resultado de la validación del cliente (fuente de información 5.5), se identificaron dos nuevos problemas de interés y sus respectivos criterios de éxito (cepn.3 y cepn.4). La tabla 5.28 describe todos los criterios de éxito identificados para cada problema del negocio. En la tabla A.11, se ilustra la primera versión del formalismo.

Criterios de Éxito del Problema de Negocio			
Responsable:	Sebastian M.	**Fecha:**	03/04/2017
ID#:	D.EN.CPN.CEPN	**Versión:**	1.1

Criterio	Descripción	Problema asociado	Referencia
cepn.1	Las rutas de navegación frecuentes sean representativas de al menos un 15% del total de usuarios	(prne.1) Detallar las posibles rutas más frecuentes de navegación que realizan los usuarios	Entrevista 3
cepn.2	La caracterización de los perfiles tenga una tasa de error inferior al 20%.	(prne.2) Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación	Entrevista 3
cepn.3	Las rutas de navegación frecuentes sean representativas de al menos un 15% del total de usuarios	(prne.3) Determinar las rutas más frecuentes de navegación según el tipo de dispositivo utilizado para navegar	Entrevista 5
cepn.4	Las rutas de navegación frecuentes sean representativas de al menos un 20% del total de usuarios	(prne.4) Determinar las rutas más frecuentes de navegación de cada perfil de usuario	Entrevista 5

Tabla 5.28. Caso de Validación: Web Log - Criterios de Éxito del Problema de Negocio (versión final)

5.1.2.2. Fase: Entendimiento de los Datos (D.ED)

La fase de entendimiento de los datos, está conformada por tres actividades: análisis de los datos (sección 5.1.2.2.1), donde se profundiza en la comprensión del significado de las variables disponibles y sus valores, exploración de los datos (sección 5.1.2.2.2), donde se describe en detalle las variables a considerar por el modelo, y evaluación de los datos (sección 5.1.2.2.3), donde se evalúan los distintos aspectos vinculados con la calidad de las variables seleccionadas.

5.1.2.2.1. Actividad: Análisis de los Datos (D.ED.AnD)

Durante la actividad de análisis de los datos se evalúan las variables disponibles en las distintas fuentes de información, con el objetivo de comprender sus significados, valoraciones, así como cualquier otro aspecto relevante del proceso aplicado para el registro de dicha información (por ejemplo: valores por defecto del sistema, forma en la cual los datos son recolectados, etc.).

A continuación se presentan los resultados obtenidos de aplicar la técnica Identificación de atributos relacionados con el Problema de Negocio (sección 4.4.2.1.2, pág. 197), utilizando como insumos la información externa provista por los interesados: Discursos de los interesados (Fuente de

Información 5.4), y los formalismos producidos en la fase de entendimiento del negocio: Fuentes de Información del Cliente (Tabla 5.20), Restricciones del Proyecto (Tabla 5.26) y Problema del Negocio (Tabla 5.27).

Discursos de los interesados: Reuniones con el cliente (Dario R.)

Entrevista 4 (Minuta):

A partir del acceso a los datos almacenados, se identificaron las siguientes variables disponibles en la fuente de información:

- **IP**: dirección IP del usuario,
- **Referer**: página desde donde accede el usuario a la primera página del sitio web (ruta absoluta),
- **Searchterms**: términos de búsqueda utilizados para acceder al sitio web (si proviene de un buscador),
- **Resource**: nombre de la página accedida (ruta relativa),
- **Plugins**: programas adicionales instalados al navegador (por ejemplo: flash, acrobat, etc),
- **Visit_id**: identificador unívoco de sesión de usuarios durante una visita,
- **Server_latency**: tiempo requerido para recibir y procesar el pedido,
- **Page_performance**: tiempo requerido en cargarse la página al usuario,
- **Browser**: tipo de navegador que usa (por ejemplo: chrome, safari, mozilla, etc.),
- **Browser_version**: versión del navegador,
- **Browser_type**: identifica el tipo de dispositivo que utiliza (0 = pc; 2 = celular),
- **Platform**: sistema operativo del usuario (por ejemplo: win7, win10, android, etc.),
- **Language**: configuración del lenguaje del navegador (por ejemplo: es-es, es-ar, es-41, etc.),
- **User_agent**: agente de usuario (aplicación que utiliza como cliente),
- **Resolution**: resolución de la pantalla del cliente,
- **Screen_width**: ancho de la pantalla del cliente,
- **Screen_height**: alto de la pantalla del cliente,
- **Content_type**: tipo de acción que realizó el usuario.
- **Category**: categoría a la que pertenece el recurso (por ejemplo: Busquedas = 1, Novedades =30, etc.),
- **Autor**: usuario que creó el recurso,
- **Content_id**: identificador único del recurso (página web)
- **Dt**: codificación de la fecha en formato Unix.

Junto con el cliente, se definieron el conjunto de variables de interés para la resolución de las problemáticas identificadas. A continuación se listan:

- ***IP, visit_id** y **dt**: como variables que permiten identificar la navegación de cada usuario.*
- **Resource y content_id**: identificar los distintos tipos de recursos. La variable content_id, facilita la distinción de aquellos recursos, como por ejemplo la página de búsqueda, en los cuales existen múltiples rutas para una misma página.

Adicionalmente, se acordó la generación de las siguientes variables, detallando el procedimiento para su construcción:

- **Fecha**: convirtiendo la codificación almacenada en el campo “dt” a formato “dd/mm/aa hh:mm”.
- **nVisitas**: cantidad de recursos únicos que un usuario accede en una visita al sitio.
- Columna binaria por cada recurso, indicando con uno si el usuario accedió al mismo.

Para ambos problemas, se utilizarán los campos binarios que identifican el acceso a los distintos recursos y la cantidad de recursos accedidos en una visita al sitio.

Fuente de Información 5.4. Caso de Validación: Web Log - Entrevista 4

Diccionario de Fuente de Datos (D.ED.AnD.DiFD): a partir de las fuentes de información previamente identificadas (Fuentes de Información del Cliente), se procede a registrar en la columna "campo" el nombre de las variables que la componen, categorizando cada uno de los campos existentes de acuerdo al tipo de variable e incorporando su significado. La tabla 5.29 ilustra los atributos disponibles.

Diccionario de Fuente de Datos			
Responsable:	Santiago B.	**Fecha:**	24/02/2017
ID#:	D.ED.AnD.DiFD	**Versión:**	1.0
Fuente de Información	(fuic.1) Wordpress Logs		
Campo	**Tipo**	**Descripción**	
Autor	Nominal	usuario que creó el recurso	
Browser	Nominal	Tipo de navegador que usa (por ejemplo: chrome, safari, mozilla, etc.)	
Browser_type	Nominal	identifica el tipo de dispositivo que utiliza (0 = pc; 2 = celular)	
Browser_version	Nominal	versión del navegador	
Category	Nominal	Categoría a la que pertenece el recurso (por ejemplo: Búsquedas = 1, Novedades =30, etc.)	
Content_id	Nominal	identificador único del recurso (página web)	
Content_type	Nominal	tipo de acción que realizó el usuario	
Dt	Continuo	Codificación de la fecha en formato Unix.	
IP	Nominal	dirección IP del usuario	
Language	Nominal	Configuración del lenguaje del navegador (por ejemplo: es-es, es-ar, es-41, etc.)	
Page_performance	Nominal	tiempo requerido en cargarse la página al usuario	
Platform	Nominal	Sistema operativo del usuario (por ejemplo: win7, win10, android, etc.)	
Plugins	Nominal	programas adicionales instalados al navegador (por ejemplo: flash, acrobat, etc)	
Referer	Nominal	página desde donde accede el usuario a la primera página del sitio web (ruta absoluta)	
Resolution	Nominal	resolución de la pantalla del cliente	
Resource	Nominal	nombre de la página accedida (ruta relativa)	
Screen_height	Continuo	alto de la pantalla del cliente	
Screen_width	Continuo	ancho de la pantalla del cliente	
Searchterms	Nominal	términos de búsqueda utilizados para acceder al sitio web (si proviene de un buscador)	
Server_latency	Continuo	tiempo requerido para recibir y procesar el pedido	
User_agent	Nominal	agente de usuario (aplicación que utiliza como cliente)	
Visit_id	Nominal	identificador unívoco de sesión de usuarios durante una visita	

Tabla 5.29. Caso de Validación: Web Log - Diccionario de Fuente de Datos

Campos Relacionados con el Problema de Negocio (D.ED.AnD.CRPN): a partir del análisis de la fuente de datos disponible (Tabla 5.30), se evalúa de forma conjunta con el experto el conjunto de variables relevantes para el problema de negocio identificado, así como aquellos que son necesario construir a partir de otros valores (Fuente de Información 5.4), precisando el procedimiento para su generación. Los campos de interés se ven limitados por las Restricciones del Proyecto (Tabla 5.26). La tabla 5.30 ilustra la selección de campos relacionados con los problemas de negocio *prne.1* y *prne.2*. En la segunda iteración del proyecto, se identifican los problemas de

negocio *prne.3* y *prne.4* reflejándose los campos de interés en las tablas 5.31 (a y b), y 5.32 respectivamente.

Campos Relacionados con el Problema de Negocio			
Responsable:	Santiago B.	**Fecha:**	28/02/2017
ID#:	D.ED.AnD.CRPN.1	**Versión:**	1.0
Problema de Negocio	(prne.1) Detallar las posibles rutas más frecuentes de navegación que realizan los usuarios. (prne.2) Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación.		

Nombre	**Gene -rar**	**Descripción**	**Referencia**
Visit_id		identificador unívoco de sesión de usuarios durante una visita	Entrevista 4 / fuic.1
nVisitas	**x**	Indica la cantidad de recursos únicos que un usuario accede en una visita al sitio	Entrevista 4
/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/usuarios/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/contacto/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/quienes-somos/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/docentes/experiencias-didacticas/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/capacitacion/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/aulas-extendidas-o-ampliadas-como-y-para-que-usarlas/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/ciudadania-digital-2/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/docentes/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/la-importancia-de-los-foros-virtuales-en-los-procesos-educativos/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/oferta-academica/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/videos-tutoriales/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/novedades/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/docentes/recursos/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/tutorias/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/proyectos/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/ciudadania-digital-en-vivo/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/las-prácticas-de-ensenanza-y-aprendizaje-en-cuestion/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4

Tabla 5.30. Caso de Validación: Web Log - Campos Relacionados con los Problemas de Negocio (prne.1 y prne.2)

Campos Relacionados con el Problema de Negocio			
Responsable:	Santiago B.	**Fecha:**	05/04/2017
ID#:	D.ED.AnD.CRPN.2	**Versión:**	1.0
Problema de Negocio	(prne.3) Determinar las rutas más frecuentes de navegación según el tipo de dispositivo utilizado para navegar		

Nombre	**Gene -rar**	**Descripción**	**Referencia**
Visit_id		identificador unívoco de sesión de usuarios durante una visita	Entrevista 4 / fuic.1
Browser_type		identifica el tipo de dispositivo que utiliza (0 = pc; 2 = celular)	Entrevista 5
nVisitas	**x**	Indica la cantidad de recursos únicos que un usuario accede en una visita al sitio	Entrevista 4
/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/usuarios/	**x**	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4

Tabla 5.31.a. Caso de Validación: Web Log - Campos Relacionados con los Problemas de Negocio (prne.3)

/contacto/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/quienes-somos/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/docentes/experiencias-didacticas/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/capacitacion/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/aulas-extendidas-o-ampliadas-como-y-para-que-usarlas/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/ciudadania-digital-2/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/docentes/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/la-importancia-de-los-foros-virtuales-en-los-procesos-educativos/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/oferta-academica/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/videos-tutoriales/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/novedades/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/docentes/recursos/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/tutorias/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/proyectos/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/ciudadania-digital-en-vivo/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/las-prácticas-de-ensenanza-y-aprendizaje-en-cuestion/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4

Tabla 5.31.b. Caso de Validación: Web Log - Campos Relacionados con los Problemas de Negocio (prne.3)

Campos Relacionados con el Problema de Negocio			
Responsable:	Santiago B.	**Fecha:**	05/04/2017
ID#:	D.ED.AnD.CRPN.3	**Versión:**	1.0
Problema de Negocio	(prne.4) Determinar las rutas más frecuentes de navegación de cada perfil de usuario		

Nombre	**Gene-rar**	**Descripción**	**Referencia**
Visit_id		identificador unívoco de sesión de usuarios durante una visita	Entrevista 4 / fuic.1
/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/usuarios/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/contacto/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/quienes-somos/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/docentes/experiencias-didacticas/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/capacitacion/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/aulas-extendidas-o-ampliadas-como-y-para-que-usarlas/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/ciudadania-digital-2/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/docentes/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/la-importancia-de-los-foros-virtuales-en-los-procesos-educativos/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/oferta-academica/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/videos-tutoriales/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/novedades/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/docentes/recursos/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/tutorias/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/proyectos/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/ciudadania-digital-en-vivo/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
/las-prácticas-de-ensenanza-y-aprendizaje-en-cuestion/	x	Variable binaria que indica si el usuario accedió al recurso en la visita	Entrevista 4
Perfiles	x	Tipología de usuario de acuerdo a sus similitudes en el acceso de recursos del sitio web (generado a partir del modelo de agrupamiento)	Entrevista 5

Tabla 5.32. Caso de Validación: Web Log - Campos Relacionados con los Problemas de Negocio (prne.4)

5.1.2.2.2. Actividad: Exploración de los Datos (D.ED.ExD)

En esta actividad se analizan los valores de los campos identificados de interés para los distintos problemas de negocio, con el objetivo de comprender las características de la población o muestra de estudio, reflejando las relaciones iniciales entre las distintas variables estudiadas.

A continuación se presentan los resultados obtenidos de aplicar la técnica Exploración de los Datos (sección 4.4.2.2.2, pág. 205), la cual utiliza como insumos los formalismos: Suposiciones del Proyecto (Tabla 5.25), Restricciones del Proyecto (Tabla 5.26), Problema del Negocio (Tabla 5.27), Diccionario de Fuente de Datos (Tabla 5.29), Campos Relacionados con el Problema de Negocio (Tabla 5.30) y Reporte de Evaluación de Herramientas (Tabla 5.3).

Fuente Integrada de datos (D.ED.ExD.FuID): a partir del conjunto de campos identificados de interés para el problema de negocio, se procede a integrar los mismos en una única fuente de información, la cual posee 134207 registros. Luego de realizar el filtrado acorde a las restricciones en los datos señaladas, la fuente queda conformada por 8811 registros y 20 campos. La figura 5.1 ilustra la estructura de la fuente de integrada de datos (mediante su representación en un Diagrama Entidad-Relación) a utilizar para las problemáticas prne.1 y *prne.2*. Las figuras 5.2 y 5.3 ilustran la fuente integrada de datos correspondiente a los problemas de negocio prne.3 y *prne.4* desarrollados en la segunda iteración del proyecto.

Figura 5.1. Caso de Validación: Web Log - Fuente Integrada de Datos correspondiente a prne.1 y prne.2

Figura 5.2. Caso de Validación: Web Log - Fuente Integrada de Datos correspondiente a prne.3

Figura 5.3. Caso de Validación: Web Log - Fuente Integrada de Datos correspondiente a prne.4

Reporte de Datos Explorados (D.ED.ExD.ReDE): a partir de las fuentes integradas de datos, se describe la distribución de valores para cada atributo relevante para cada problema de negocio, detallando la cantidad de registros y la proporción que los mismos representan con respecto a la muestra. La tabla 5.33 ilustra la descripción de los atributos correspondientes a la fuente FuID.1. En la segunda iteración del proyecto, se definen las fuentes FuID.2 y FuID.3, las cuales se describen en las tablas 5.34 y 5.36 (a y b) respectivamente. A partir de la suposición asociada con la variación del comportamiento de los usuarios según el dispositivo utilizado para navegar el sitio (supr.1), se amplía el análisis de los datos de la fuente FuID.2, realizándose un diagrama de violín respecto a la cantidad de recursos visitados por un usuario en una navegación según el tipo de dispositivo utilizado (figura 5.4) y una descripción de los recursos accedidos por cada tipo de dispositivo (tabla

5.35). A partir del análisis realizado, no se desprende evidencia que soporte la suposición evaluada, determinándose la no pertinencia del estudio del comportamiento de los usuarios según el dispositivo utilizado (prne.3).

Reporte de Datos Explorados			
Responsable:	Santiago B.	**Fecha:**	03/03/2017
ID#:	D.ED.ExD.ReDE.FuID.1	**Versión:**	1.0
Problema de Negocio	(prne.1) Detallar las posibles rutas más frecuentes de navegación que realizan los usuarios. (prne.2) Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación.		

ATRIBUTOS CUALITATIVOS		
Nombre	**Valores**	**Distribución**
/	0 1	1368 (15,53%) 7443 (84,47%)
/usuarios/	0 1	6544 (74,27%) 2267 (25,73%)
/contacto/	0 1	8128 (92,25%) 683 (7,75%)
/quienes-somos/	0 1	8320 (94,43%) 491 (5,57%)
/docentes/experiencias-didacticas/	0 1	8399 (95,32%) 412 (4,68%)
/capacitacion/	0 1	6444 (73,14%) 2367 (26,86%)
/aulas-extendidas-o-ampliadas-como-y-para-que-usarlas/	0 1	8729 (99,07%) 82 (0,93%)
/ciudadania-digital-2/	0 1	8375 (95,05%) 436 (4,95%)
/docentes/	0 1	7533 (85,50%) 1278 (14,50%)
/la-importancia-de-los-foros-virtuales-en-los-procesos-educativos/	0 1	8677 (98,48%) 134 (1,52%)
/oferta-academica/	0 1	5534 (62,81%) 3277 (37,19%)
/videos-tutoriales/	0 1	8174 (92,77%) 637 (7,23%)
/novedades/	0 1	7796 (88,48%) 1015 (11,52%)
/docentes/recursos/	0 1	7966 (90,41%) 845 (9,59%)
/tutorias/	0 1	8459 (96,00%) 352 (4,00%)
/proyectos/	0 1	8459 (96,00%) 352 (4,00%)
/ciudadania-digital-en-vivo/	0 1	8578 (97,36%) 233 (2,64%)
/las-prácticas-de-ensenanza-y-aprendizaje-en-cuestion/	0 1	8772 (99,56%) 39 (0,44%)

ATRIBUTOS CUANTITATIVOS				
Nombre	**Min**	**Max**	**Tendencia Central**	**Dispersión**
nVisitas	2	16	Media=2.53 Moda=2	[2-3] = 87,8% [4-5] = 10.01% [6-7] = 1.78% [8-9] = 0.34% [10;16] = 0.07%
Comentarios:				

Tabla 5.33. Caso de Validación: Web Log - Reporte de Datos Explorados (FuID.1)

Reporte de Datos Explorados			
Responsable:	Santiago B.	**Fecha:**	12/04/2017
ID#:	D.ED.ExD.ReDE.FuID.2	**Versión:**	1.0
Problema de Negocio	(prne.3) Determinar las rutas más frecuentes de navegación según el tipo de dispositivo utilizado para navegar		

ATRIBUTOS CUALITATIVOS		
Nombre	**Valores**	**Distribución**
/	0 1	1367 (15,53%) 7438 (84,47%)
/usuarios/	0 1	6544 (74,32%) 2261 (25,68%)
/contacto/	0 1	8127 (92,30%) 678 (7,70%)
/quienes-somos/	0 1	8319 (94,48%) 486 (5,52%)
/docentes/experiencias-didacticas/	0 1	8399 (95,39%) 406 (4,61%)
/capacitacion/	0 1	6442 (73,16%) 2363 (26,84%)
/aulas-extendidas-o-ampliadas-como-y-para-que-usarlas/	0 1	8724 (99,08%) 81 (0,92%)
/ciudadania-digital-2/	0 1	8373 (95,09%) 432 (4,91%)
/docentes/	0 1	7532 (85,54%) 1273 (14,46%)
/la-importancia-de-los-foros-virtuales-en-los-procesos-educativos/	0 1	8672 (98,49%) 133 (1,51%)
/oferta-academica/	0 1	5534 (62,85%) 3271 (37,15%)
/videos-tutoriales/	0 1	8172 (92,81%) 633 (7,19%)
/novedades/	0 1	7794 (88,52%) 1011 (11,48%)
/docentes/recursos/	0 1	7964 (90,45%) 841 (9,55%)
/tutorias/	0 1	8459 (96,07%) 346 (3,93%)
/proyectos/	0 1	8457 (96,05%) 348 (3,95%)
/ciudadania-digital-en-vivo/	0 1	8573 (97,37%) 232 (2,63%)
/las-prácticas-de-ensenanza-y-aprendizaje-en-cuestion/	0 1	8767 (99,57%) 38 (0,43%)
browser_type	0 2	6744 (76,59%) 2061 (23,41%)

ATRIBUTOS CUANTITATIVOS

Nombre	Min	Max	Tendencia Central	Dispersión
nVisitas	2	9	Media=2.53 Moda=2	[2-3] = 87,86% [4-5] = 10.01% [6-7] = 1.79% [8-9] = 0.34%

Comentarios:
Se adjunta el análisis de acceso de recursos por dispositivo (figuras 5.4) y tabla (5.35)

Tabla 5.34. Caso de Validación: Web Log - Reporte de Datos Explorados (FuID.2)

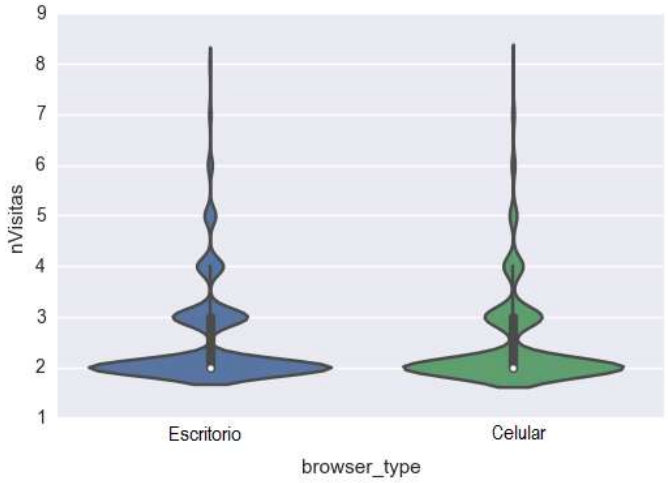


Figura 5.4. Caso de Validación: Web Log – Gráfica de cantidad de recursos visitados según dispositivo (FuID.2)

Recursos	Escritorio	Celular
/	0,83 (0,37)	0,88 (0,32)
/aulas-extendidas-o-ampliadas-como-y-para-que-usarlas/	0,01 (0,10)	0,01 (0,09)
/capacitacion/	0,28 (0,45)	0,22 (0,41)
/ciudadania-digital-2/	0,05 (0,22)	0,05 (0,22)
/ciudadania-digital-en-vivo/	0,03 (0,17)	0,02 (0,13)
/contacto/	0,08 (0,27)	0,07 (0,26)
/docentes/	0,16 (0,36)	0,11 (0,31)
/docentes/experiencias-didacticas/	0,05 (0,23)	0,02 (0,15)
/docentes/recursos/	0,10 (0,30)	0,08 (0,27)
/la-importancia-de-los-foros-virtuales-en-los-procesos-educativos/	0,02 (0,13)	0,01 (0,11)
/las-prácticas-de-ensenanza-y-aprendizaje-en-cuestion/	0,01 (0,07)	0,00 (0,04)
/novedades/	0,10 (0,29)	0,18 (0,38)
/oferta-academica/	0,40 (0,49)	0,39 (0,45)
/proyectos/	0,04 (0,20)	0,03 (0,17)
/quienes-somos/	0,06 (0,25)	0,03 (0,17)
/tutorias/	0,04 (0,20)	0,03 (0,17)
/usuarios/	0,22 (0,41)	0,28 (0,49)
/videos-tutoriales/	0,08 (0,27)	0,05 (0,22)

Tabla 5.35. Caso de Validación: Web Log – Descripción de recursos visitados según dispositivo (FuID.2)

Reporte de Datos Explorados			
Responsable:	Santiago B.	**Fecha:**	13/04/2017
ID#:	D.ED.ExD.ReDE.FuID.3	**Versión:**	1.0
Problema de Negocio	(prne.4) Determinar las rutas más frecuentes de navegación de cada perfil de usuario		
ATRIBUTOS CUALITATIVOS			
Nombre	**Valores**	**Distribución**	
/	0 1	1367 (15,53%) 7438 (84,47%)	
/usuarios/	0 1	6544 (74,32%) 2261 (25,68%)	
/contacto/	0 1	8127 (92,30%) 678 (7,70%)	

Tabla 5.36.a. Caso de Validación: Web Log - Reporte de Datos Explorados (FuID.3)

/quienes-somos/	0 1	8319 (94,48%) 486 (5,52%)
/docentes/experiencias-didacticas/	0 1	8399 (95,39%) 406 (4,61%)
/capacitacion/	0 1	6442 (73,16%) 2363 (26,84%)
/aulas-extendidas-o-ampliadas-como-y-para-que-usarlas/	0 1	8724 (99,08%) 81 (0,92%)
/ciudadania-digital-2/	0 1	8373 (95,09%) 432 (4,91%)
/docentes/	0 1	7532 (85,54%) 1273 (14,46%)
/la-importancia-de-los-foros-virtuales-en-los-procesos-educativos/	0 1	8672 (98,49%) 133 (1,51%)
/oferta-academica/	0 1	5534 (62,85%) 3271 (37,15%)
/videos-tutoriales/	0 1	8172 (92,81%) 633 (7,19%)
/novedades/	0 1	7794 (88,52%) 1011 (11,48%)
/docentes/recursos/	0 1	7964 (90,45%) 841 (9,55%)
/tutorias/	0 1	8459 (96,07%) 346 (3,93%)
/proyectos/	0 1	8457 (96,05%) 348 (3,95%)
/ciudadania-digital-en-vivo/	0 1	8573 (97,37%) 232 (2,63%)
/las-prácticas-de-ensenanza-y-aprendizaje-en-cuestion/	0 1	8767 (99,57%) 38 (0,43%)
Perfiles	0 1 2	3448 (39,16%) 864 (9,81%) 4493 (51,03%)
Comentarios:		

Tabla 5.36.b. Caso de Validación: Web Log - Reporte de Datos Explorados (FuID.3)

5.1.2.2.3. Actividad: Evaluación de los Datos (D.ED.EvD)

En esta actividad se analizan los campos de interés para los distintos problemas de negocio, identificando aquellas características que puedan afectar la calidad del modelo.

A continuación, se exhiben los resultados obtenidos de aplicar la técnica Exploración de la Calidad de los Datos (sección 4.4.2.3.2, pág. 215) en el caso de validación, para el cual se utilizan como elementos de entrada los formalismos: Diccionario de Fuente de Datos (Tabla 5.29), Campos Relacionados con el Problema de Negocio (Tabla 5.30), Reporte de Datos Explorados (Tablas 5.33, 5.34 y 5.36), Fuente Integrada de datos (Figuras 5.1 - 5.3) y Reporte de Evaluación de Herramientas (Tabla 5.3).

Reporte de la Calidad de los Datos (D.ED.EvD.ReCD): a partir del análisis realizado en la fuente integrada de datos (FuID.1) y su descripción, se identifica en el campo "nVisitas", registros cuya cantidad de recursos accedidos es mayor a diez, entendiéndose como valores atípicos (los cuales tienen una representatividad menor al 0.05% de los registros). Dicho rango de valores se encuentra presente en seis registros. La tabla 5.37 ilustra el resultado obtenido del análisis de la calidad de los datos. Durante la segunda iteración del proyecto, no se identificaron problemas de calidad en los datos.

Reporte de la Calidad de los Datos			
Responsable:	Santiago B.	**Fecha:**	06/03/17
ID#:	D.ED.EvD.ReCD.1	**Versión:**	1.0
Problema de Negocio	(prne.1) Detallar las posibles rutas más frecuentes de navegación que realizan los usuarios. (prne.2) Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación.		

Nombre	**Registros**	**Tipo**	**Descripción**
nVisitas	6	outlier	Valores mayores o iguales a 10

Tabla 5.37. Caso de Validación: Web Log - Reporte de la Calidad de los Datos

5.1.2.3. Fase: Modelado (D.Mo)

La fase Modelado está conformada por 2 actividades: Modelado del Problema (sección 5.1.2.3.1), en la cual se traduce las necesidades del cliente desde la perspectiva del dominio del negocio a la perspectiva de explotación de información y se identifica a partir de ella los posibles modelos de explotación de información a utilizar y Configuración del Modelo (sección 5.1.2.3.2), donde se establecen las características del modelo a utilizar.

5.1.2.3.1. Actividad: Modelado del Problema (D.Mo.MoP)

En esta actividad se traducen los requerimientos del cliente desde la perspectiva del dominio del negocio a explotación de información.

A continuación se presentan los resultados obtenidos de aplicar la técnica Derivación del Proceso de Explotación de Información (sección 4.4.3.1.2, pág. 225), la cual utiliza como elementos de entrada el Problema del Negocio (Tabla 5.27), el Diccionario de Fuente de Datos (Tabla 5.29) y Campos Relacionados con el Problema de Negocio (Tabla 5.30).

Diseño del Proceso de Explotación de Información (D.Mo.MoP.DPEI): para cada problema de negocio, se identifican sus aspectos relevantes (conceptos, atributos, relaciones y valores), generando un conjunto de marcos que permiten sistematizar el proceso de representación del conocimiento desde formato texto al formato gráfico basado en redes semánticas, identificando los

distintos elementos de interés para determinar el proceso de explotación de información a utilizar. Para el primer problema de negocio (prne.1), se identifica el concepto "usuario" y siendo de interés la identificación de rutas más frecuentes (representado con el nodo variable), de acuerdo a los recursos accedidos. La tabla 5.38, ilustra la gráfica resultante a partir de la cual se identifica al **proceso de Ponderación de Interdependencia de Atributos** el cual indica como estrategia de implementación el uso de algoritmos de frecuencia de ocurrencia (por ejemplo: redes bayesianas). Para el segundo problema de negocio (prne.2), se identifica el concepto navegación y como acciones a realizar la categorización y determinación de perfiles de usuario (representados como nodos variables). La tabla 5.39, ilustra la gráfica resultante a partir de la cual se identifica al **proceso de Descubrimiento de Reglas de Pertenencia a Grupos**, utilizando una combinación de algoritmos de tipo clustering y árboles de decisión.

En la segunda iteración del proyecto, de acuerdo a los resultados identificados en la etapa de exploración de los datos (sección 5.1.2.2), se aplica únicamente la técnica de derivación del proceso de explotación de información al problema de negocio prne.4, observándose su resultado en la tabla 5.40 (la utilización del **proceso de Ponderación de Reglas de Pertenencia a Grupos con Grupos definidos**).

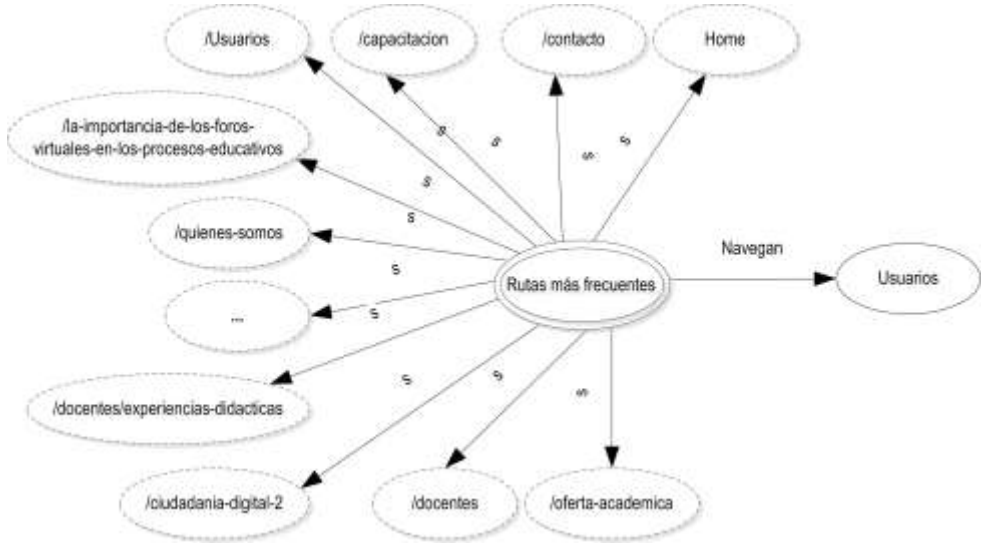
Diseño del Proceso de Explotación de Información			
Responsable:	Santiago B.	**Fecha:**	10/03/2017
ID#:	D.Mo.MoP.DPEI.1	**Versión:**	1.0
Problema de Negocio:	(prne.1) Detallar las posibles rutas más frecuentes de navegación que realizan los usuarios		
			
Proceso de Explotación de Información:	Ponderación de Interdependencia de Atributos		

Tabla 5.38. Caso de Validación: Web Log - Diseño del Proceso de Explotación de Información (prne.1)

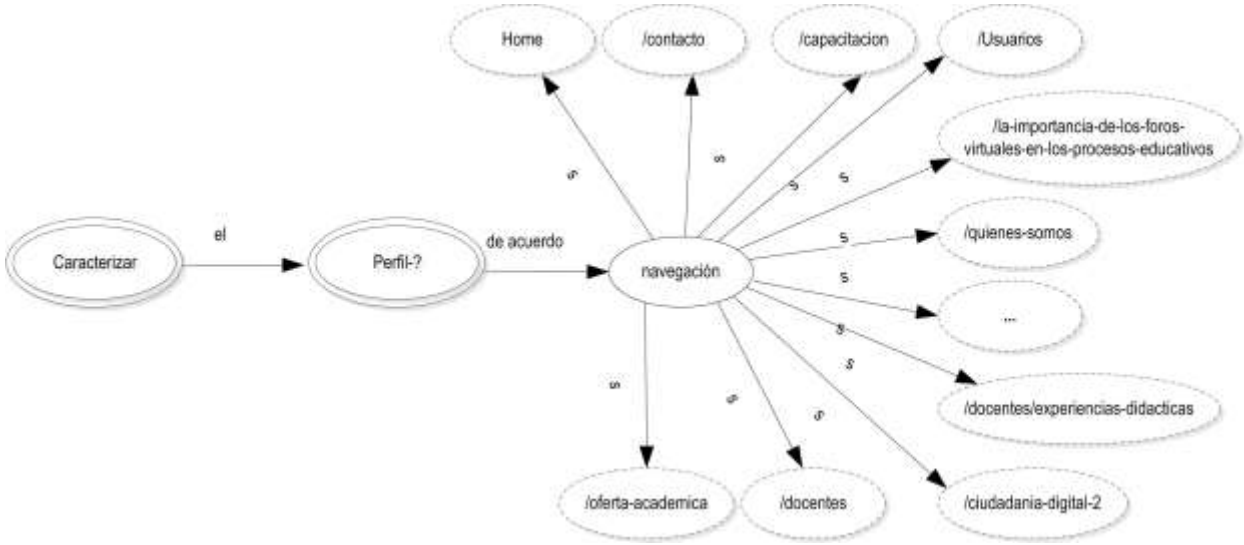
Diseño del Proceso de Explotación de Información			
Responsable:	Santiago B.	**Fecha:**	13/03/2017
ID#:	D.Mo.MoP.DPEI.2	**Versión:**	1.0
Problema de Negocio:	(prne.2) Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación		
			
Proceso de Explotación de Información:	Descubrimiento de reglas de pertenencia a grupos		

Tabla 5.39. Caso de Validación: Web Log - Diseño del Proceso de Explotación de Información (prne.2)

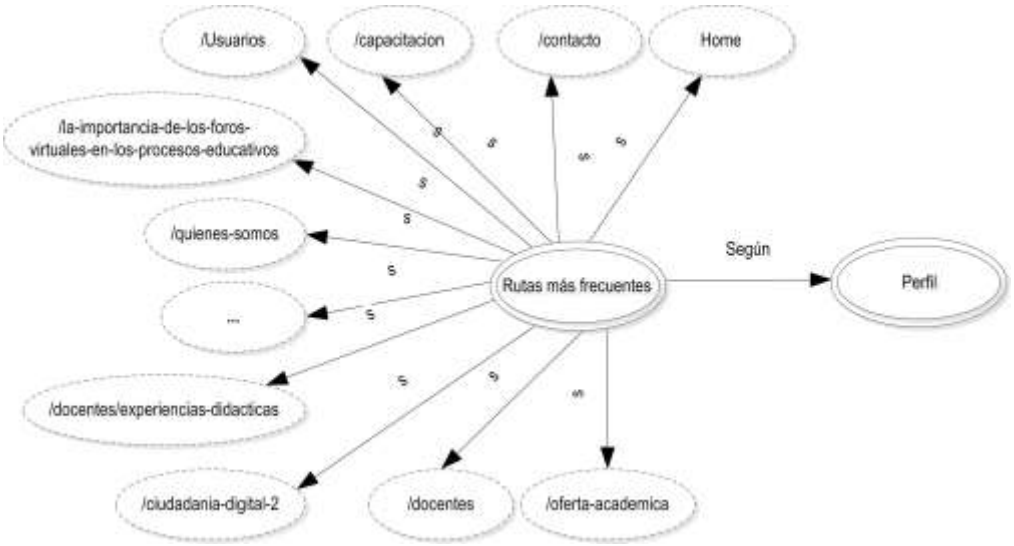
Diseño del Proceso de Explotación de Información			
Responsable:	Santiago B.	**Fecha:**	16/04/2017
ID#:	D.Mo.MoP.DPEI.3	**Versión:**	1.0
Problema de Negocio:	(prne.4) Determinar las rutas más frecuentes de navegación de cada perfil de usuario		
			
Proceso de Explotación de Información:	Proceso de Ponderación de Reglas de Pertenencia a Grupos (Grupos definidos)		

Tabla 5.40. Caso de Validación: Web Log - Diseño del Proceso de Explotación de Información (prne.2)

5.1.2.3.2. Actividad: Configuración del Modelo (D.Mo.CoM)

En esta actividad se definen los elementos que conforman la estrategia de implementación y evaluación de los distintos modelos para la extracción de patrones vinculados con el problema de negocio.

A continuación se presentan los resultados obtenidos de aplicar la técnica Determinación de la Configuración del Modelo (sección 4.4.3.2.2, pág. 233), la cual utiliza como insumos los formalismos: Diseño del Proceso de Explotación de Información (Tablas 5.38 - 5.40), para definir los algoritmos a utilizar junto con Diccionario de Fuente de Datos (Tabla 5.29), Campos Relacionados con el Problema de Negocio (Tabla 5.30), Reporte de Datos Explorados (Tablas 5.33, 5.34 y 5.36), Reporte de la Calidad de los Datos (Tabla 5.37) y Criterios de Éxito del Problema de Negocio (Tabla 5.28).

Selección de Algoritmos de Explotación de Información (D.Mo.CoM.SAEI): a partir de los procesos de explotación de información identificados en la primera iteración de la actividad previa, se utiliza el algoritmo de frecuencia de ocurrencia de ítems (FP-Growth) para la problemática "prne.1" y los algoritmos K-Medoids y Árboles de decisiones para la problemática "prne.2". Las tablas 5.41 y 5.42 ilustran las estructuras de los formalismos obtenidos para cada problemática, respectivamente. En la segunda iteración se define el uso del algoritmo de frecuencia de ocurrencia de ítems (FP-Growth) para la problemática "prne.4" (tabla 5.43).

Selección de Algoritmos de explotación de información				
Responsable:	Sebastian M.		**Fecha:**	13/03/2017
ID#:	D.Mo.CoM.SAEI.1		**Versión:**	1.0
Problema de Negocio	(prne.1) Detallar las posibles rutas más frecuentes de navegación que realizan los usuarios			
Algoritmo	**Input**	**Target**	**Estrategia**	**Descripción**
FP-Growth	Booleano		1	Identificación de ítems frecuentes

Tabla 5.41. Caso de Validación: Web Log - Selección de Algoritmos de Explotación de Información (prne.1)

Selección de Algoritmos de explotación de información				
Responsable:	Sebastian M.		**Fecha:**	13/03/2017
ID#:	D.Mo.CoM.SAEI.2		**Versión:**	1.0
Problema de Negocio	(prne.2) Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación			
Algoritmo	**Input**	**Target**	**Estrategia**	**Descripción**
K-Medoids	Booleano/Continuo		1	
K-Means	Booleano/Continuo		1	
DecisionTree	Discretos/Continuos	Discretos	2	Versión genérica de árboles de decisión

Tabla 5.42. Caso de Validación: Web Log - Selección de Algoritmos de Explotación de Información (prne.2)

Selección de Algoritmos de explotación de información				
Responsable:	Sebastian M.	**Fecha:**	17/04/2017	
ID#:	D.Mo.CoM.SAEI.3	**Versión:**	1.0	
Problema de Negocio	(prne.4) Determinar las rutas más frecuentes de navegación de cada perfil de usuario			
Algoritmo	**Input**	**Target**	**Estrategia**	**Descripción**
FP-Growth	Booleano		1	Identificación de ítems frecuentes

Tabla 5.43. Caso de Validación: Web Log - Selección de Algoritmos de Explotación de Información (prne.4)

Selección de Variables del Modelo (D.Mo.CoM.SeVM): para la conformación del modelo, y a partir del análisis realizado de los datos (exploración y calidad), se define la utilización de todas las variables asociadas con el acceso de recursos de la visita (para ambos problemas de negocios), identificándose como entrada (Input) en la columna tipo. Para estos algoritmos, no se requiere convertir las variables. En las tablas 5.44 y 5.45 se ilustran los resultados obtenidos. En la tabla 5.46, se definen las variables a utilizar para el problema de negocio "prne.4" (perteneciente a la segunda iteración del proceso).

Selección de variables del Modelo			
Responsable:	Sebastian M.	**Fecha:**	15/03/2017
ID#:	D.Mo.CoM.SeVM.1	**Versión:**	1.0
Problema de Negocio	(prne.1) Detallar las posibles rutas más frecuentes de navegación que realizan los usuarios		
Campo	**FP-Growth**		
	Tipo	**Conversión**	
Visit_id			
nVisitas			
/	Input		
/usuarios/	Input		
/contacto/	Input		
/quienes-somos/	Input		
/docentes/experiencias-didacticas/	Input		
/capacitacion/	Input		
/aulas-extendidas-o-ampliadas-como-y-para-que-usarlas/	Input		
/ciudadania-digital-2/	Input		
/docentes/	Input		
/la-importancia-de-los-foros-virtuales-en-los-procesos-educativos/	Input		
/oferta-academica/	Input		
/videos-tutoriales/	Input		
/novedades/	Input		
/docentes/recursos/	Input		
/tutorias/	Input		
/proyectos/	Input		
/ciudadania-digital-en-vivo/	Input		
/las-prácticas-de-ensenanza-y-aprendizaje-en-cuestion/	Input		

Tabla 5.44. Caso de Validación: Web Log - Selección de Variables del Modelo (prne.1)

Selección de variables del Modelo			
Responsable:	Sebastian M.	**Fecha:**	15/03/2017
ID#:	D.Mo.CoM.SeVM.2	**Versión:**	1.0
Problema de Negocio	(prne.2) Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación		

Campo	K-medoids/k-means		DecisionTree	
	Tipo	Conversión	Tipo	Conversión
Visit_id				
nVisitas				
/	Input		Input	
/usuarios/	Input		Input	
/contacto/	Input		Input	
/quienes-somos/	Input		Input	
/docentes/experiencias-didacticas/	Input		Input	
/capacitacion/	Input		Input	
/aulas-extendidas-o-ampliadas-como-y-para-que-usarlas/	Input		Input	
/ciudadania-digital-2/	Input		Input	
/docentes/	Input		Input	
/la-importancia-de-los-foros-virtuales-en-los-procesos-educativos/	Input		Input	
/oferta-academica/	Input		Input	
/videos-tutoriales/	Input		Input	
/novedades/	Input		Input	
/docentes/recursos/	Input		Input	
/tutorias/	Input		Input	
/proyectos/	Input		Input	
/ciudadania-digital-en-vivo/	Input		Input	
/las-prácticas-de-ensenanza-y-aprendizaje-en-cuestion/	Input		Input	

Tabla 5.45. Caso de Validación: Web Log - Selección de Variables del Modelo (prne.2)

Selección de variables del Modelo			
Responsable:	Sebastian M.	**Fecha:**	18/04/2017
ID#:	D.Mo.CoM.SeVM.3	**Versión:**	1.0
Problema de Negocio	(prne.4) Determinar las rutas más frecuentes de navegación de cada perfil de usuario		

Campo	FP-Growth	
	Tipo	Conversión
Visit_id		
/	Input	
/usuarios/	Input	
/contacto/	Input	
/quienes-somos/	Input	
/docentes/experiencias-didacticas/	Input	
/capacitacion/	Input	
/aulas-extendidas-o-ampliadas-como-y-para-que-usarlas/	Input	
/ciudadania-digital-2/	Input	
/docentes/	Input	
/la-importancia-de-los-foros-virtuales-en-los-procesos-educativos/	Input	
/oferta-academica/	Input	
/videos-tutoriales/	Input	
/novedades/	Input	
/docentes/recursos/	Input	
/tutorias/	Input	
/proyectos/	Input	
/ciudadania-digital-en-vivo/	Input	
/las-prácticas-de-ensenanza-y-aprendizaje-en-cuestion/	Input	
Perfiles	Target	

Tabla 5.46. Caso de Validación: Web Log - Selección de Variables del Modelo (prne.1)

Estrategias de Evaluación de Modelos (D.Mo.CoM.EsEM): a partir de la cantidad de registros disponibles y el tiempo de cómputo requerido para la ejecución de los procesos se opta por utilizar para la evaluación del modelo asociado al segundo problema (prne.2) la técnica de separación del set de datos en dos particiones (entrenamiento y testeo) las cuales representan el 70% y 30% respectivamente. Para los modelos restantes (por las características del algoritmo a utilizar) no se requiere una estrategia de evaluación. En la tabla 5.47 se registran los criterios definidos.

Estrategias de evaluación de modelos			
Responsable:	Sebastian M.	**Fecha:**	15/03/2017
ID#:	D.Mo.CoM.EsEM	**Versión:**	1.0
Problema de Negocio	(prne.2) Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación		
Técnica	**Alcance**	**Descripción**	
Split validation	DecisionTree	Separación 70-30 (entrenamiento-test)	

Tabla 5.47. Caso de Validación: Web Log - Estrategias de Evaluación de Modelos

5.1.2.4. Fase: Preparación de los Datos (D.PD)

La fase Preparación de los Datos está conformada por 2 actividades: Construcción de la Fuente Temporaria de Datos (sección 5.1.2.4.1), donde se preparan y describen las distintas fuentes de datos a utilizar para la extracción del conocimiento y la selección del modelo, y Adecuación de la Fuente Temporaria de Datos (sección 5.1.2.4.2), en la cual se realizan las tareas de limpieza y formateo de los datos.

5.1.2.4.1. Actividad: Construcción de la Fuente Temporaria de Datos (D.PD.CFT)

En esta actividad se realizan las tareas finales para la generación de las fuentes de datos requeridas para las distintas etapas de implementación del modelo (entrenamiento, validación y testeo). Las fuentes generadas se definen como fuente temporaria de datos, debido a que dicha fuente de almacenamiento es distinta a aquella utilizada en producción y la misma solo será de utilidad para la formación del modelo, la extracción del conocimiento y la evaluación del mismo.

A continuación se presentan los resultados obtenidos de aplicar la técnica Generación de la Fuente Temporaria de Datos (sección 4.4.4.1.2, pág. 246), la cual utiliza como insumos los siguientes formalismos: Reporte de Datos Explorados (Tablas 5.33, 5.34 y 5.36), Fuente Integrada de datos (Figuras 5.1 - 5.3), Estrategias de evaluación de modelos (Tabla 5.47), Reporte de Evaluación de Herramientas (Tabla 5.3) y Selección de variables del Modelo (Tablas 5.44 - 5.46).

Fuente Temporaria de datos (D.PD.CFT.FuTD): la estructura de la fuente temporaria de datos (FuTD.1) a utilizar para los problemas de negocio "prne.1" y "prne.2" se ilustra en la figura 5.5. En la segunda iteración del proyecto, se construye la FuTD.2 para la problemática "prne.4" (figura 5.6).

Figura 5.5. Caso de Validación: Web Log - Fuente Temporaria de Datos FuTD.1 (Diagrama Entidad-Relación)

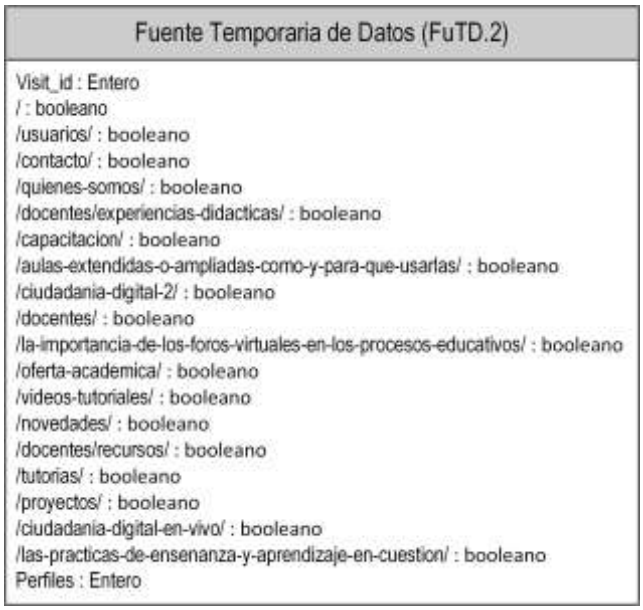


Figura 5.6. Caso de Validación: Web Log - Fuente Temporaria de Datos FuTD.2 (Diagrama Entidad-Relación)

Reporte de Generación de la Fuente Temporaria de datos (D.PD.CFT.RGFT): si bien la FuTD inicial para los problemas de negocio "prne.1" y "prne.2" es idéntica (tabla 5.48), para el entrenamiento del modelo correspondiente al segundo problema, se realiza una muestra del 70% de

la fuente integrada de datos, la cual se describe en la tabla 5.49. En la segunda iteración del proyecto, se construye la FuTD para la problemática "prne.4" (tabla 5.50).

Se destaca que la sección de atributos cuantitativos fue omitida por simplicidad, debido a la inexistencia de atributos de dicho tipo.

Reporte de Generación de la Fuente Temporaria de datos			
Responsable:	Santiago B.	**Fecha:**	16/03/2017
ID#:	D.PD.CFT.RGFT.1	**Versión:**	1.0
Problema de Negocio	(prne.1) Detallar las posibles rutas más frecuentes de navegación que realizan los usuarios.		
ATRIBUTOS CUALITATIVOS			
Nombre		**Valores**	**Distribución**
/		0 1	1367 (15,5%) 7438 (84,5%)
/usuarios/		0 1	6544 (74,3%) 2261 (25,7%)
/contacto/		0 1	8127 (92,3%) 678 (7,7%)
/quienes-somos/		0 1	8319 (94,5%) 486 (5,5%)
/docentes/experiencias-didacticas/		0 1	8399 (95,4%) 406 (4,6%)
/capacitacion/		0 1	6442 (73,2%) 2363 (26,8%)
/aulas-extendidas-o-ampliadas-como-y-para-que-usarlas/		0 1	8724 (99,1%) 81 (0,9%)
/ciudadania-digital-2/		0 1	8373 (95,1%) 432 (4,9%)
/docentes/		0 1	7532 (85,5%) 1273 (14,5%)
/la-importancia-de-los-foros-virtuales-en-los-procesos-educativos/		0 1	8672 (98,5%) 133 (1,5%)
/oferta-academica/		0 1	5534 (62,9%) 3271 (37,1%)
/videos-tutoriales/		0 1	8172 (92,8%) 633 (7,2%)
/novedades/		0 1	7794 (88,5%) 1011 (11,5%)
/docentes/recursos/		0 1	7964 (90,4%) 841 (9,6%)
/tutorias/		0 1	8459 (96,1%) 346 (3,9%)
/proyectos/		0 1	8457 (96,0%) 348 (4,0%)
/ciudadania-digital-en-vivo/		0 1	8573 (97,4%) 232 (2,6%)
/las-prácticas-de-ensenanza-y-aprendizaje-en-cuestion/		0 1	8767 (99,6%) 38 (0,4%)
Comentarios:			

Tabla 5.48. Caso de Validación: Web Log - Reporte de Generación de la Fuente Temporaria de datos asociadas al problema de negocio (prne.1)

Reporte de Generación de la Fuente Temporaria de datos			
Responsable:	Santiago B.	**Fecha:**	17/03/2017
ID#:	D.PD.CFT.RGFT.2	**Versión:**	1.0
Problema de Negocio	(prne.2) Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación.		
ATRIBUTOS CUALITATIVOS			
Nombre	**Valores**	**Distribución**	
/	0 1	952 (15,4%) 5212 (84,6%)	
/usuarios/	0 1	4621 (75%) 1543 (25%)	
/contacto/	0 1	5712 (92,7%) 452 (7,3%)	
/quienes-somos/	0 1	5822 (94,5%) 342 (5.5 %)	
/docentes/experiencias-didacticas/	0 1	5877 (95,3%) 287 (4,7%)	
/capacitacion/	0 1	4530 (73,5%) 1634 (26,5%)	
/aulas-extendidas-o-ampliadas-como-y-para-que-usarlas/	0 1	6109 (99,1%) 55 (0,9%)	
/ciudadania-digital-2/	0 1	5859 (95,1%) 305 (4,9%)	
/docentes/	0 1	5246 (85,1%) 918 (14,9%)	
/la-importancia-de-los-foros-virtuales-en-los-procesos-educativos/	0 1	6084 (98,7%) 80 (1,3%)	
/oferta-academica/	0 1	3842 (62,3%) 2322 (37,7%)	
/videos-tutoriales/	0 1	5734 (93%) 430 (7%)	
/novedades/	0 1	5454 (88,5%) 710 (11,5%)	
/docentes/recursos/	0 1	5574 (90,4%) 590 (9,6%)	
/tutorias/	0 1	5918 (96%) 246 (4%)	
/proyectos/	0 1	5932 (96,2%) 232 (3,8%)	
/ciudadania-digital-en-vivo/	0 1	5994 (97,2%) 170 (2,8%)	
/las-prácticas-de-ensenanza-y-aprendizaje-en-cuestion/	0 1	6140 (99,6%) 22 (0,4%)	
Comentarios: Aplicable al entrenamiento del algoritmo árbol de decisión (DecisionTree). Local random seed: 1992			

Tabla 5.49. Caso de Validación: Web Log - Reporte de Generación de la Fuente Temporaria de datos asociada al problema de negocio (prne.2)

Reporte de Generación de la Fuente Temporaria de datos			
Responsable:	Santiago B.	**Fecha:**	20/04/2017
ID#:	D.PD.CFT.RGFT.3	**Versión:**	1.0
Problema de Negocio	(prne.4) Determinar las rutas más frecuentes de navegación de cada perfil de usuario		
ATRIBUTOS CUALITATIVOS			
Nombre	**Valores**	**Distribución**	
/	0 1	1367 (15,53%) 7438 (84,47%)	
/usuarios/	0 1	6544 (74,32%) 2261 (25,68%)	
/contacto/	0 1	8127 (92,30%) 678 (7,70%)	
/quienes-somos/	0 1	8319 (94,48%) 486 (5,52%)	
/docentes/experiencias-didacticas/	0 1	8399 (95,39%) 406 (4,61%)	
/capacitacion/	0 1	6442 (73,16%) 2363 (26,84%)	
/aulas-extendidas-o-ampliadas-como-y-para-que-usarlas/	0 1	8724 (99,08%) 81 (0,92%)	
/ciudadania-digital-2/	0 1	8373 (95,09%) 432 (4,91%)	
/docentes/	0 1	7532 (85,54%) 1273 (14,46%)	
/la-importancia-de-los-foros-virtuales-en-los-procesos-educativos/	0 1	8672 (98,49%) 133 (1,51%)	
/oferta-academica/	0 1	5534 (62,85%) 3271 (37,15%)	
/videos-tutoriales/	0 1	8172 (92,81%) 633 (7,19%)	
/novedades/	0 1	7794 (88,52%) 1011 (11,48%)	
/docentes/recursos/	0 1	7964 (90,45%) 841 (9,55%)	
/tutorias/	0 1	8459 (96,07%) 346 (3,93%)	
/proyectos/	0 1	8457 (96,05%) 348 (3,95%)	
/ciudadania-digital-en-vivo/	0 1	8573 (97,37%) 232 (2,63%)	
/las-prácticas-de-ensenanza-y-aprendizaje-en-cuestion/	0 1	8767 (99,57%) 38 (0,43%)	
Perfiles	0 1 2	3448 (39,16%) 864 (9,81%) 4493 (51,03%)	
Comentarios:			

Tabla 5.50. Caso de Validación: Web Log - Reporte de Generación de la Fuente Temporaria de datos asociadas al problema de negocio (prne.4)

5.1.2.4.2. Actividad: Adecuación de la Fuente Temporaria de Datos (D.PD.AFT)

En esta actividad se analizan las características de los campos seleccionados para los distintos problemas de negocio, con el objetivo de identificar y realizar actividades de conversión y ajuste de los registros, preparando los datos para la adecuada extracción de patrones de conocimiento.

A continuación se presentan los resultados obtenidos de aplicar la técnica Adecuación de los Datos (sección 4.4.4.2.2, pág. 254), la cual utiliza como insumos los formalismos: Reporte de la Calidad de los Datos (Tabla 5.37), Selección de variables del Modelo (Tablas 5.44 - 5.46), Fuente Temporaria de Datos (Tabla 5.5 y 5.6) y Reporte de Generación de la Fuente Temporaria de Datos (Tablas 5.48 - 5.50).

Reporte de Adecuación de la Fuente Temporaria de Datos (D.PD.AFT.RAFT): para los primeros dos problemas de negocio (prne.1 y prne.2), se removieron en su construcción aquellos registros que habían navegado un número elevado de recursos, al considerarse como poco frecuente (posiblemente siendo bots). Para la segunda iteración del proyecto, no fueron requeridas tareas de adecuación de los datos. La tabla 5.51 ilustra los resultados.

Reporte de Adecuación de la Fuente Temporaria de Datos			
Responsable:	Santiago B	**Fecha:**	17/03/2017
ID#:	D.PD.AFT.RAFT	**Versión:**	1.0
Problema de Negocio:	(prne.1) Detallar las posibles rutas más frecuentes de navegación que realizan los usuarios. (prne.2) Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación.		

Nombre	**Acción**	**Efecto**	**Descripción**
nVisitas	Remover outliers	6 registros eliminados	Valores mayores o iguales a 10

Tabla 5.51. Caso de Validación: Web Log - Reporte de Adecuación de la Fuente Temporaria de Datos

5.1.2.5. Fase: Implementación (D.Im)

La fase de implementación, está conformada por las actividades: Selección del Modelo (sección 5.1.2.5.1), donde se define la estrategia a utilizar para identificar la mejor configuración del modelo, y explotación de información (sección 5.1.2.5.2), donde se realiza la extracción y descripción de los patrones de conocimiento ocultos en los datos.

5.1.2.5.1. Actividad: Selección del Modelo (D.Im.SeM)

En esta actividad se define el criterio y la forma mediante la cual se determina cuál de los posibles algoritmos o combinación de algoritmos logra capturar con mayor precisión los patrones ocultos en los datos.

A continuación se presentan los resultados obtenidos de aplicar la técnica Selección de la Estrategia de Hiperparametrización (sección 4.4.5.1.2, pág. 260). La misma utiliza como insumos los formalismos: Selección de Algoritmos de explotación de información (Tablas 5.41-5.43), el Reporte

de Generación de la Fuente Temporaria de Datos (Tablas 5.48-5.50), los Criterios de Éxito del Problema de Negocio (Tabla 5.28) y la Reporte de Evaluación de Herramientas (Tabla 5.3).

Reporte de Estrategia de Parametrización del Modelo (D.Im.SeM.REPM): de acuerdo al tipo de problema a resolver y los algoritmos asociados, se requiere para el segundo problema de negocio (prne.2) identificar la configuración óptima de los parámetros para los algoritmos de clutering (k-means y k-medoids) y de clasificación (árboles de decisión). La tabla 5.52 detalla la estrategia de configuración a utilizar, los parámetros a optimizar (indicando el rango de valores a evaluar) y aquellos predefinidos.

Reporte de Estrategia de Parametrización del Modelo						
Responsable:	Sebastian M.		**Fecha:**	21/03/2017		
ID#:	D.Im.SeM.REPM		**Versión:**	1.0		
Problema de Negocio	(prne.2) Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación					
ID	**Algoritmo E.I.**	**Estrategia de Configuración**	**Criterio**	**Rango Inferior**	**Rango Superior**	**Comentarios**
repm.1	K-Medoids / K-Means	Grid Search	La suma de la diferencia media cuadrada intra grupos (WSS)	K: 2	K: 6	Nominal measure: JaccardSimilarity
repm.2	DecisionTree	Random Search	Tasa de Error	maximal_depth: 10	maximal_depth: 100	Confidence: 0.25 Minimal gain: 0.1 Steps: 5

Tabla 5.52. Caso de Validación: Web Log - Reporte de Estrategia de Parametrización del Modelo

5.1.2.5.2. Actividad: Explotación de Información (D.Im.ExI)

En esta actividad se aplican los algoritmos de explotación de información (o minería de datos), con el objetivo de extraer los patrones de conocimientos ocultos en las fuentes de información, dejando constancia de los resultados obtenidos para poder reproducir y comparar los mismos.

A continuación se presentan los resultados obtenidos de aplicar la técnica Extracción de Conocimiento (sección 4.4.5.2.2, pág. 267), la cual utiliza como insumos los formalismos: Criterios de Éxito del Problema de Negocio (Tabla 5.28), Selección de Algoritmos de explotación de información (Tablas 5.41 - 5.43), Selección de variables del Modelo (Tablas 5.44 - 5.46), Estrategias de evaluación de modelos (Tabla 5.47), Fuente Temporaria de Datos (Tabla 5.5 y 5.6), Reporte de Generación de la Fuente Temporaria de Datos (Tablas 5.48 - 5.50) y Reporte de Estrategia de Parametrización del Modelo (Tabla 5.52).

Reporte de Implementación del Modelo (D.Im.ExI.ReIM): la tabla 5.53 ilustra la configuración utilizada para la problemática “prne.1”. Para el problema de negocio “prne.2”, se obtuvo como

mejor configuración: la combinación de algoritmos K-Medoids (k=3) y DecisionTree (maximal_depth=15). En la tabla 5.54 se formalizan los resultados obtenidos. La tabla 5.55 describe la configuración utilizada, en la segunda iteración del proyecto, para dar respuesta a la problemática "prne.4".

Reporte de Implementación del Modelo			
Responsable:	Santiago B.	**Fecha:**	27/03/2017
ID#:	D.Im.ExI.ReIM.1	**Versión:**	1.0
Problema de Negocio:	(prne.1) Detallar las posibles rutas más frecuentes de navegación que realizan los usuarios		

Algoritmo E.I.	**Estrategia**	**Configuración**	**Descripción**
FP-Growth	-	Min support: 0.15	

Tabla 5.53. Caso de Validación: Web Log - Reporte de Implementación del Modelo (prne.1)

Reporte de Implementación del Modelo			
Responsable:	Santiago B.	**Fecha:**	27/03/2017
ID#:	D.Im.ExI.ReIM.2	**Versión:**	1.0
Problema de Negocio:	(prne.2) Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación		

Algoritmo E.I.	**Estrategia**	**Configuración**	**Descripción**
K-Medoids	(repm.1) Grid Search	K: 3	WSS: 0.720
K-Means	(repm.1) Grid Search	K: 3	WSS: 0.781
K-Medoids-DecisionTree	(repm.2) Random Search	maximal_depth: 15	Tasa de error: 0.00
K-Means-DecisionTree	(repm.2) Random Search	maximal_depth: 15	Tasa de error: 0.05

Tabla 5.54. Caso de Validación: Web Log - Reporte de Implementación del Modelo (prne.2)

Reporte de Implementación del Modelo			
Responsable:	Santiago B.	**Fecha:**	28/04/2017
ID#:	D.Im.ExI.ReIM.3	**Versión:**	1.0
Problema de Negocio:	(prne.4) Determinar las rutas más frecuentes de navegación de cada perfil de usuario		

Algoritmo E.I.	**Estrategia**	**Configuración**	**Descripción**
FP-Growth	-	Min support: 0.20	

Tabla 5.55. Caso de Validación: Web Log - Reporte de Implementación del Modelo (prne.4)

Patrones de Conocimiento (D.Im.ExI.PaCo): para el primer problema de negocio (prne.1), se identifican tres rutas frecuentes de navegación con un soporte superior al 15%: desde la página de

inicio hacia las secciones generales Centro de Ayuda (/usuarios) y Capacitación, y la sección Oferta Académica. La tabla 5.56 detalla dicha información. Se destaca que fueron desestimados aquellos ítems frecuentes únicos. Para el segundo problema de negocio (prne.2) **s**e detallan los patrones obtenidos por el modelo que obtuvo mejores resultados (K- Medoids - DecisionTree), presentando las reglas de comportamiento obtenidas para cada uno de los tres grupos identificados y su matriz de confusión (figuras 5.7.a y 5.7.b). A partir de los patrones identificados y su análisis (tablas 5.57.a y 5.57.b), se identifican los siguientes perfiles:

- Cluster 0: conformado principalmente por usuarios que no acceden a las páginas "/oferta-académica", "/capacitación" y /novedades". Se destaca la presencia de visitantes que acceden al centro de ayuda ("/usuarios"), aunque se observa la conformación de otros grupos de menor envergadura.
- Cluster 1: integrado por usuarios que visitan la página en busca de novedades ("/novedades").
- Cluster 2: compuesto por usuarios que acceden en busca de información sobre los distintos cursos disponibles (/capacitación y /oferta-académica).

En la segunda iteración del proyecto, se obtuvieron los patrones detallados en la tabla 5.58, identificándose una ruta frecuente para el perfil Cluster 0: desde la página inicial al centro de ayuda ("/usuarios"); dos rutas frecuentes para Cluster 1: a) desde la página de inicio a novedades, y b) desde la página de inicio a centro de ayuda y luego novedades; y tres rutas frecuentes para el Cluster 2: a) desde la página de inicio a ofertas académicas, b) desde la página de inicio a capacitación y c) desde la página de ofertas académicas a capacitación.

Soporte	Ítem 1	Ítem 2
0.279	/	/oferta-académica/
0.206	/	/capacitación/
0.234	/	/usuarios/

Tabla 5.56. Caso de Validación: Web Log – Ítems frecuentes (prne.1)

Tasa de error: 0.00	real Cluster 0	real Cluster 1	real Cluster 2	precisión clase
pred. Cluster 0	1046	0	0	100%
pred. Cluster 1	0	1339	0	100%
pred. Cluster 2	0	0	256	100%
recall clase	100%	100%	100%	

Figura 5.7.a. Caso de Validación: Web Log - Reglas de Comportamiento y Matriz de Confusión (prne.2)

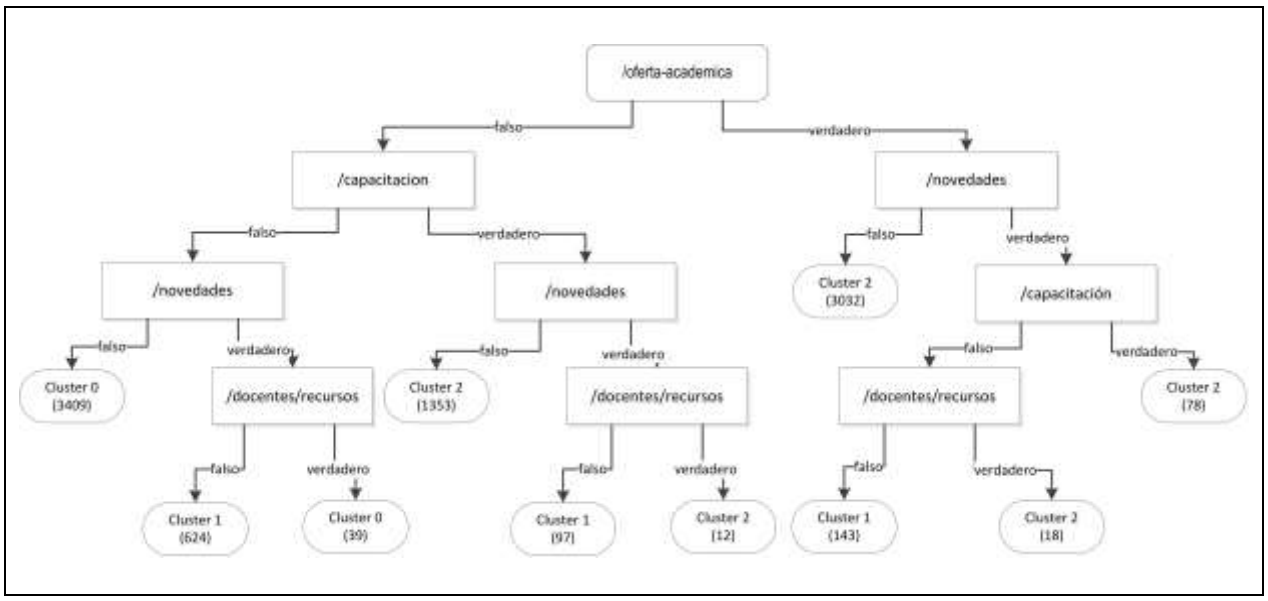


Figura 5.7.b. Caso de Validación: Web Log - Reglas de Comportamiento y Matriz de Confusión (prne.2)

Campo	Cluster 0 [Recall] Precisión	Cluster 1 [Recall] Precisión	Cluster 2 [Recall] Precisión	Global
/=false	[27,1 %] 10,7 %	[70,2 %] 21,3 %	[2,8 %] 4,4 %	15,5%
/=true	[41,4 %] 89,3 %	[47,5 %] 78,7 %	[11,1 %] 95,6 %	84,5%
/usuarios=false	[28,9 %] 54,9 %	[60,9 %] 88,7 %	[10,2 %] 77,2 %	74,3%
/usuarios=true	[68,8 %] 45,1 %	[22,5 %] 11,3 %	[8,7 %] 22,8 %	25,7%
aulas-extendidas-o-ampliadas-como-y-para-que-usarlas=false	[38,9 %] 98,4 %	[51,2 %] 99,4 %	[9,9 %] 100,0 %	99,1%
aulas-extendidas-o-ampliadas-como-y-para-que-usarlas=true	[69,1 %] 1,6 %	[30,9 %] 0,6 %	[0,0 %] 0,0 %	0,9%
capacitacion=false	[53,5 %] 100,0 %	[34,6 %] 49,6 %	[11,9 %] 88,8 %	73,2%
capacitacion=true	[0,0 %] 0,0 %	[95,9 %] 50,4 %	[4,1 %] 11,2 %	26,8%
ciudadania-digital-2=false	[36,8 %] 89,4 %	[53,1 %] 99,0 %	[10,1 %] 97,6 %	95,1%
ciudadania-digital-2=true	[85,0 %] 10,6 %	[10,2 %] 1,0 %	[4,9 %] 2,4 %	4,9%
ciudadania-digital-en-vivo=false	[37,7 %] 93,9 %	[52,3 %] 99,7 %	[10,0 %] 99,2 %	97,4%
ciudadania-digital-en-vivo=true	[91,4 %] 6,1 %	[5,6 %] 0,3 %	[3,0 %] 0,8 %	2,6%
contacto=false	[38,7 %] 91,3 %	[51,2 %] 92,6 %	[10,0 %] 94,4 %	92,3%
contacto=true	[44,1 %] 8,7 %	[48,8 %] 7,4 %	[7,1 %] 5,6 %	7,7%
docentes/experiencias-didacticas=false	[38,5 %] 93,7 %	[51,5 %] 96,2 %	[10,1 %] 98,0 %	95,4%
docentes/experiencias-didacticas=true	[53,7 %] 6,3 %	[42,1 %] 3,8 %	[4,2 %] 2,0 %	4,6%
docentes/recursos=false	[37,4 %] 86,4 %	[51,7 %] 91,7 %	[10,8 %] 100,0 %	90,4%
docentes/recursos=true	[55,6 %] 13,6 %	[44,4 %] 8,3 %	[0,0 %] 0,0 %	9,6%
docentes=false	[36,8 %] 80,4 %	[52,9 %] 88,6 %	[10,3 %] 90,2 %	85,5%
docentes=true	[53,1 %] 19,6 %	[40,2 %] 11,4 %	[6,7 %] 9,8 %	14,5%
la-importancia-de-los-foros-virtuales-en-los-procesos-educativos=false	[38,4 %] 96,7 %	[51,6 %] 99,6 %	[9,9 %] 99,8 %	98,5%
la-importancia-de-los-foros-virtuales-en-los-procesos-educativos=true	[85,7 %] 3,3 %	[12,8 %] 0,4 %	[1,5 %] 0,2 %	1,5%
las-prácticas-de-enseñanza-y-aprendizaje-en-cuestion=false	[39,0 %] 99,2 %	[51,1 %] 99,8 %	[9,9 %] 100,0 %	99,6%
las-prácticas-de-enseñanza-y-aprendizaje-en-cuestion=true	[71,1 %] 0,8 %	[28,9 %] 0,2 %	[0,0 %] 0,0 %	0,4%
novedades=false	[43,7 %] 98,9 %	[56,3 %] 97,6 %	[0,0 %] 0,0 %	88,5%
novedades=true	[3,9 %] 1,1 %	[10,7 %] 2,4 %	[85,5 %] 100,0 %	11,5%
oferta-academica=false	[62,3 %] 100,0 %	[24,7 %] 30,4 %	[13,0 %] 83,4 %	62,9%

Tabla 5.57.a. Caso de Validación: Web Log – Descripción por Clusters (prne.2)

oferta-academica=true	[0,0 %] 0,0 %	[95,6 %] 69,6 %	[4,4 %] 16,6 %	37,1%
proyectos=false	[39,2 %] 96,0 %	[51,0 %] 95,9 %	[9,9 %] 96,6 %	96,0%
proyectos=true	[39,4 %] 4,0 %	[52,3 %] 4,1 %	[8,3 %] 3,4 %	4,0%
quienes-somos=false	[39,1 %] 94,2 %	[50,8 %] 94,1 %	[10,1 %] 97,3 %	94,5%
quienes-somos=true	[40,9 %] 5,8 %	[54,3 %] 5,9 %	[4,7 %] 2,7 %	5,5%
tutorias=false	[39,7 %] 97,4 %	[50,2 %] 94,6 %	[10,1 %] 98,5 %	96,1%
tutorias=true	[26,0 %] 2,6 %	[70,2 %] 5,4 %	[3,8 %] 1,5 %	3,9%
videos-tutoriales=false	[37,4 %] 88,7 %	[52,5 %] 95,4 %	[10,1 %] 95,5 %	92,8%
videos-tutoriales=true	[61,5 %] 11,3 %	[32,4 %] 4,6 %	[6,2 %] 4,5 %	7,2%
Total Registros	**3448**	**864**	**4493**	

Tabla 5.57.b. Caso de Validación: Web Log – Descripción por Clusters (prne.2)

Cluster 0		
support	**ítem 1**	**ítem 2**
0,415	/	/usuarios/

Cluster 1			
support	**ítem 1**	**ítem 2**	**ítem 3**
0,956	/	/novedades/	
0,220	/	/usuarios/	/novedades/

Cluster 2		
support	**ítem 1**	**ítem 2**
0,518	/	/oferta-academica/
0,383	/	/capacitacion/
0,201	/oferta-academica/	/capacitacion/

Tabla 5.58. Caso de Validación: Web Log – Ítems frecuentes por perfil (prne.4)

5.1.2.6. Fase: Evaluación y Presentación (D.EP)

La fase Evaluación y Presentación se encuentra conformada por dos actividades: Evaluación de los Resultados (sección 5.1.2.6.1), donde se analiza la validez y utilidad de los patrones hallados, y Presentación de los Resultados (sección 5.1.2.6.2), garantizando la adecuada transmisión del conocimiento extraído para su utilización.

5.1.2.6.1. Actividad: Evaluación de los Resultados (D.EP.EvR)

En esta actividad se evalúa la validez de los patrones de conocimiento obtenidos para el dominio de negocio y en particular para las problemática de negocio en cuestión.

A continuación se presentan los resultados obtenidos de aplicar la técnica Validación del Conocimiento (sección 4.4.6.1.2, pág. 275), la cual utiliza como insumos los formalismos: Objetivos del Proyecto (Tabla 5.22), Criterios de Éxito del Proyecto (Tabla 5.23), Problema del Negocio (Tabla 5.27), Criterios de Éxito del Problema de Negocio (Tabla 5.28), Reporte de Implementación del Modelo (Tablas 5.53–5.55) y patrones de conocimiento (Tablas 5.56-5.58 y Figura 5.7).

Reporte de Evaluación de los Resultados (D.EP.EvR.ReER): de acuerdo a los patrones de conocimiento identificados, se evaluó de forma conjunta con el experto del problema de negocio

(Dario R.) la validez e interés de los resultados. En la fuente de información 5.5, se presentan los resultados obtenidos de la interacción con el cliente durante la primera iteración del proyecto. Se destaca la definición de nuevas necesidades ampliatorias de los patrones obtenidos para las primeras dos problemáticas de negocio (prne.1 y prne.2), identificándose dos nuevos problemas de negocio.

En la segunda iteración, como resultado del análisis exploratorio de los datos, se descarta la suposición asociada con el tercer problema definido (prne.3). Finalmente, se validan los resultados obtenidos para la problemática "prne.4". En la fuente de información 5.6, se presentan los resultados obtenidos de la interacción con el cliente durante la segunda parte del proyecto. La tabla 5.59, ilustra los resultados de la validación de los problemas de negocios (en su versión final). La versión previa se encuentra registrada en la sección A.2.2 (tabla A.12).

Discursos de los interesados: Reuniones con el cliente (Dario R.)

Entrevista 5 (Minuta):

Los resultados obtenidos respecto a las páginas de navegación más frecuentes de los usuarios eran esperables, dado que son tres de las cuatro secciones generales del sitio web, y en especial aquellas destinadas directamente a proveer recursos y herramientas que faciliten a los docentes principalmente (pero también a los estudiantes) en el uso del aula virtual. Nos interesaría ampliar el estudio, identificando si hay algún otro tipo relación oculta en la navegación de los usuarios, por ejemplo, si varía el acceso a los recursos entre un usuario que accede a nuestra versión Mobile con respecto a la versión de escritorio.
Con respecto a los perfiles de navegación, se consideran de gran relevancia los tipos de usuarios identificados. Entendemos que el primer grupo está conformado por visitantes ocasionales en busca de algún tipo de recurso que les permita solucionar una problemática específica con respecto al uso del aula virtual. Usualmente, son usuarios sin mucho conocimiento sobre herramientas informáticas. El segundo y tercer grupo, están asociados a usuarios con mayor experticia en el uso del aula virtual, en busca de herramientas o cursos para perfeccionar algún conocimiento específico (en el segundo caso) o con el objetivo de mantenerse informado en los cambios y eventos que realiza el área (para el primer caso).
A partir de la confirmación del interés de los grupos identificados, se acuerda con el cliente profundizar con la comprensión de las rutas frecuentes de navegación específicas de cada perfil de usuarios.
Se acuerda para las nuevas problemáticas identificadas, que el criterio de aceptación de los patrones de comportamiento de los usuarios debe ser superior al 15% para la primera problemática y 20% para la última, debido a que se espera que los perfiles posean un mayor nivel de coincidencia en los recursos de interés.

Fuente de Información 5.5. Caso de Validación: Web Log – Entrevista 5

Discursos de los interesados: Reuniones con el cliente (Dario R.)

Entrevista 6 (Minuta):

A partir del estudio realizado respecto a la variación de la visita de recursos entre un usuario que accede al sitio web en la versión Mobile y la versión de escritorio, se determina que el comportamiento es semejante, identificándose como posible explicación que la versión para celulares no es una aplicación especializada para dichos dispositivos, sino que es la optimización de la página de escritorio para las pantallas de menor tamaño (es decir, bajo un diseño responsivo). Esto implica que el usuario puede realizar las mismas secuencias de navegación sin importar el dispositivo utilizado, siendo un objetivo de interés para la organización, lograr que el usuario pueda alcanzar la misma experiencia sin importar el medio utilizado.
Con respecto al estudio individualizado de los usuarios por perfiles, las rutas de navegación identificadas son de interés, en especial las cuatro rutas que amplían los conocimientos previamente presentados. Si bien para la pieza de conocimiento obtenida para el perfil 0, es parte de las obtenidas en la población en general, esta descripción detallada ayuda a especificar el comportamiento y permitir ajustar la oferta de recursos a los usuarios a partir de los sitios que visitan y sus posibles intereses.

Fuente de Información 5.6. Caso de Validación: Web Log – Entrevista 6

Reporte de Evaluación de los Resultados			
Responsable:	Sebastian M.	**Fecha:**	05/05/2017
ID#:	D.EP.EvR.ReER	**Versión:**	1.1
Problema de Negocio	**Criterio de Éxito**	**Resultado**	**Descripción**
(prne.1) Detallar las posibles rutas más frecuentes de navegación que realizan los usuarios.	(cepn.1) Las rutas de navegación frecuentes sean representativas de al menos un 15% del total de usuarios	Ampliatorio	Se confirman los resultados, identificándose la necesidad de estudiar la variación de la navegación de los usuarios según el medio que utilicen.
(prne.2) Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación.	(cepn.2) La caracterización de los perfiles tenga una tasa de error inferior al 20%.	Ampliatorio	El primer grupo está conformado por visitantes ocasionales en busca de algún tipo de recurso que les permita solucionar una problemática específica con respecto al uso del aula virtual. Usualmente, son usuarios sin mucho conocimiento sobre herramientas informáticas. El segundo y tercer grupo, están asociados a usuarios con mayor experticia en el uso del aula virtual, en busca de herramientas o cursos para perfeccionar algún conocimiento específico (en el segundo caso) o con el objetivo de mantenerse informado en los cambios y eventos que realiza el área (para el primer caso). Se acuerda profundizar con la comprensión de las rutas frecuentes de navegación específicas de cada perfil de usuarios.
(prne.4) Determinar las rutas más frecuentes de navegación de cada perfil de usuario	(cepn.4) Las rutas de navegación frecuentes sean representativas de al menos un 20% del total de usuarios	Satisfactorio	A partir del análisis especializado en los perfiles de usuarios, se identifican seis rutas frecuentes de navegación, ampliando los resultados obtenidos en el problema de negocio prne.1.

Tabla 5.59. Caso de Validación: Web Log - Reporte de Evaluación de los Resultados (versión final)

5.1.2.6.2. Actividad: Presentación de los Resultados (D.EP.PrR)

En esta actividad se llevan a cabo las tareas finales del proyecto, con respecto a la documentación de los resultados obtenidos y la presentación de los mismos a los interesados. El objetivo de esta actividad es la correcta transmisión de los patrones obtenidos y el conocimiento extraído para dar soporte al proceso decisorio del cliente.

A continuación se presentan los resultados obtenidos de aplicar la técnica Síntesis del Proyecto (sección 4.4.6.2.2, pág. 280), la cual utiliza como insumos los formalismos: Fuentes de Información del Cliente (Tabla 5.20), Objetivos del Proyecto (Tabla 5.22), Criterios de Éxito del Proyecto (Tabla 5.23), Suposiciones del Proyecto (Tabla 5.25), Restricciones del Proyecto (Tabla 5.26), Problema del Negocio (Tabla 5.27), Patrones de Conocimiento (Tablas 5.56 - 5.58 y Figura 5.7), Reporte de Generación de la Fuente Temporaria de Datos (Tablas 5.48 - 5.50), Reporte de la Calidad de los Datos (Tabla 5.37) y Reporte de Evaluación de los Resultados (Tabla 5.59)

Reporte del Proyecto (D.EP.PrR.RepP): en la tabla 5.60 se ilustra el resumen del proyecto, presentado al cliente, describiendo las necesidades y problemáticas identificadas, junto con sus criterios de éxito, se detallan los datos disponibles, el proceso implementado junto con los

resultados obtenidos y su interpretación/evaluación. Finalmente, se presentan recomendaciones sobre futuros aspectos a evaluar.

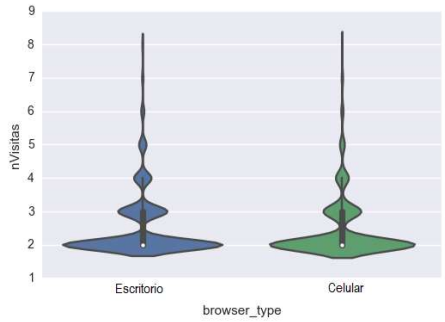
<table>
<tr><th colspan="4">Reporte del Proyecto</th></tr>
<tr><td>Responsable:</td><td>Sebastian M.</td><td>Fecha:</td><td>07/05/2017</td></tr>
<tr><td>ID#:</td><td>D.EP.PrR.RepP</td><td>Versión:</td><td>1.0</td></tr>
<tr><td>DESCRIPCIÓN DEL PROBLEMA</td><td colspan="3">El propósito de este proyecto es brindar piezas de conocimiento respecto a la navegación del usuario la cual permita optimizar la experiencia del mismo, facilitando su uso y mejorando la disposición de los contenidos de interés. Para ello, se dispuso de los registros de navegación del sitio (logs) de los usuarios correspondientes a los dos últimos años (febrero del 2016, cuando entra en vigencia la nueva estructura del sitio web).
El estudio se centró en comprender las características de navegación más frecuentes entre los usuarios, es decir, aquellos conjuntos de recursos que se visitan con mayor asiduidad. En este contexto, se estudió la población general y luego se realizaron tareas de segmentación de los usuarios por similitud en los recursos accedidos.
Se espera como resultado del proyecto, ampliar el conocimiento de los usuarios con el objetivo de renovar la estructuración de los contenidos, facilitando y mejorando la experiencia del usuario.</td></tr>
<tr><td>DESCRIPCIÓN DE LOS DATOS</td><td colspan="3">Para el estudio, se dispuso de 134207 registros correspondientes al registro de los recursos accedidos (e información de los dispositivos) por los usuarios en los últimos 2 años (desde inicios de febrero del 2016 hasta el inicio del proyecto).
En consideración con objetivo del proyecto, se utilizaron aquellos usuarios que realizaron más de una petición (accedieron a más de un recurso) durante su navegación en el sitio. Adicionalmente, dado la amplia cantidad de recursos disponibles, se decidió realizar el análisis sobre aquellos cuyo número de accesos sea mayor al 1% del total de peticiones realizadas. Para el estudio, se utilizó la información correspondiente al tipo de dispositivo y los recursos que accedió el usuario durante su navegación.
Como resultado de la adaptación de la base de datos, de acuerdo a las necesidades y restricciones del proyecto (mencionadas en el párrafo anterior), se obtuvieron 8811 registros de navegaciones por usuario con 20 campos. Finalmente, 6 registros fueron eliminados debido al alto número de recursos accedidos (valores infrecuentes asociados con el posible recorrido automatizado del sitio por “robots”), resultando en un total de 8805 registros.</td></tr>
<tr><td>RESULTADOS DE EXPLOTACIÓN DE INFORMACIÓN</td><td colspan="3">A partir del estudio de la muestra general, se identifican tres rutas frecuentes de navegación con un soporte superior al 15% de los registros: desde la página de inicio hacia las secciones Centro de Ayuda (/usuarios), Capacitación, y Oferta Académica, esto es, el acceso a tres de las cuatro secciones generales del sitio web. A partir de estos resultados, surge el interés de analizar si dicho comportamiento se variaba de acuerdo al tipo de dispositivo utilizado. Suposición que fue desechada luego de analizar los datos de cada uno de estos grupos. La figura 1 refleja la simetría entre el número de recursos accedidos por navegación de los usuarios por tipo de dispositivo.
 

Figura 1. Cantidad de recursos visitados según tipo de dispositivo
A partir del estudio de segmentación de usuarios de acuerdo a su similitud en el interés de los recursos accedidos, se identificaron 3 tipos:
• Perfil 1: conformado principalmente por usuarios que no acceden a las páginas “/oferta-académica”, “/capacitación” y /novedades”. Se destaca la presencia de visitantes que acceden al centro de ayuda (“/usuarios”), aunque se observa la conformación de otros grupos de menor envergadura. Su patrón de navegación frecuente es: acceder desde la página inicial al centro de ayuda (“/usuarios”)
• Perfil 2: integrado por usuarios que visitan la página en busca de novedades (“/novedades”). Se identifican dos patrones de navegación frecuentes: a) desde la página de inicio a novedades, y b) desde la página de inicio a centro de ayuda y luego novedades
• Perfil 3: compuesto por usuarios que acceden en busca de información sobre los distintos cursos disponibles (/capacitación) y (“/oferta-académica”). Se identifican tres patrones de navegación frecuentes: a) desde la página de inicio a ofertas académicas, b) desde la página de inicio a capacitación y c) desde la página de ofertas académicas a capacitación.</td></tr>
<tr><td>EVALUACIÓN DE LOS RESULTADOS</td><td colspan="3">A partir del análisis de segmentación de usuarios, se identifican tres perfiles: el primero (perfil 1) está conformado por visitantes ocasionales en busca de algún tipo de recurso que les permita solucionar una problemática específica con respecto al uso del aula virtual. Usualmente, son usuarios sin mucho conocimiento sobre herramientas informáticas. El segundo y tercer grupo (perfil 2 y 3 respectivamente), están asociados a usuarios con mayor experticia en el uso del aula virtual, en busca de herramientas o cursos para perfeccionar algún conocimiento específico (en el segundo caso) o con el objetivo de mantenerse informado en los cambios y eventos que realiza el área (para el primer caso).
Adicionalmente, se identifican seis rutas frecuentes de navegación, que junto con la segmentación de los usuarios por sus intereses, permiten comprender y mejorar la experiencia de los usuarios a partir de la disposición de recursos asociados con las características en común con otros usuarios.</td></tr>
<tr><td>DIFICULTADES Y RECOMENDACIONES</td><td colspan="3">El estudio realizado puede ser ampliado mediante la implementación de técnicas específicas de explotación de información en grafos.</td></tr>
</table>

Tabla 5.60. Caso de Validación: Web Log - Reporte del Proyecto

5.2. CASO DE VALIDACIÓN: EDUCACIÓN SUPERIOR

El segundo caso de validación es un proyecto proveniente del sector educativo, en el cual se desea obtener información sobre el comportamiento de los estudiantes de educación superior, proveyendo asistencia en cuestiones de políticas públicas universitarias vinculadas con la mejora de la calidad educativa. Específicamente, el caso se enmarca en un proyecto de investigación realizado por docentes de la Facultad de Ciencias Exactas, Físicas y Naturales (FCEFyN) en la Universidad Nacional de Córdoba y se focaliza en materias de educación superior en contextos de masividad.

El objetivo central del trabajo consiste en "discurrir entre distintos planos del problema del estudiante como actor protagónico del espacio natural histórico de transmisión de saberes hegemónicos de generación en generación, para permitir que emerjan las dimensiones hacia su comprensión y la de los procesos que lo constituyen en un problema público" [Diaz y Garcia-Martinez, 2016].

La problemática abordada consiste en identificar y caracterizar distintos perfiles de estudiantes de acuerdo a sus aspectos académicos, socioeconómicos y geográficos, esperando lograr un mejor conocimiento de las características del estudiante, protagonista principal de este escenario, que permita proporcionar contribuciones novedosas y valiosas que favorezcan la toma de decisiones en aspectos vinculados con la gestión de la Educación Superior en contextos de masividad.

Para alcanzar dicho objetivo se dispone de la fuente de información SIU-Guaraní, enfocando el análisis en aquellos estudiantes que cursaron la materia informática en los años 2012 y 2013 (contándose con más de 1500). Para cada estudiante, se dispone de información de su carrera, procedencia, aspectos personales y de su entorno, económicos, los cuales son actualizados a partir de las entrevistas que se realizan cada año, dejando registro de la última fecha de relevamiento. A continuación se provee el listado de variables disponibles:

- **Fecha de Relevamiento**: fecha de última actualización del registro
- **Carrera id**: identificador de la carrera del estudiante
- **Carrera**: nombre de la carrera
- **Año de ingreso**: año en el cuál él estudiante ingresó a la carrera
- **Localidad procedencia**: localidad que procede el estudiante
- **País procedencia**.
- **Provincia procedencia**.
- **Departamento procedencia**.
- **Fecha de nacimiento**.

- **Estado civil**.
- **Cantidad de hijos**.
- **Vive con**: descripción de las personas con las cuales habita el estudiante
- **Beca**: posee beca
- **Fuente de la Beca**: de donde proviene la beca
- **Costea sus estudios**: qué medios utiliza para solventar los costos de su estudio (múltiples valores).
- **Obra Social**: origen de la obra social (si posee).
- **Trabajo**: descripción de su trabajo o la intención de obtención.
- **Horas semanales de trabajo**.
- **Padre últimos estudios**: últimos estudios escolares realizados por el padre.
- **Padre trabajo**: descripción de la carga horaria del trabajo o la intención de obtención.
- **Madre últimos estudios**: últimos estudios escolares realizados por la madre.
- **Madre trabajo**: descripción de la carga horaria del trabajo o la intención de obtención.
- **Cantidad de materias cursa primer semestre**: monto total de materias que realizó el primer semestre.
- **Fecha aprobó materia**: fecha de aprobación de la materia de interés.
- **Nota aprobó materia**: calificación obtenida en la materia de interés.
- **Cantidad de materias aprobadas**.
- **Promedio con aplazo**: promedio de calificaciones obtenidas incluyendo las desaprobaciones.
- **Promedio sin aplazo**: promedio de calificaciones obtenidas excluyendo las desaprobaciones.

En las siguientes secciones, se presentan los elementos obtenidos y desarrollados a lo largo del proyecto, presentados acorde a MoProPEI. Para facilitar la comprensión del proceso, los resultados obtenidos se presentan según la estructura de la propuesta, y no de manera temporal. Las actividades y tareas realizadas junto con los resultados obtenidos al aplicar las técnicas pertinentes a los subprocesos de gestión y desarrollo se detallan en las secciones 5.2.1 y 5.2.2 respectivamente.

5.2.1. MoProPEI-G: Subproceso Gestión (G)

El subproceso de gestión se implementa de manera transversal al subproceso de desarrollo. Está conformado por cinco fases: Iniciación (sección 5.2.1.1), Planificación (sección 5.2.1.2), Soporte (sección 5.2.1.3), Control (sección 5.2.1.4) y Cierre (sección 5.2.1.5).

5.2.1.1. Fase: Iniciación (G.IN)

Esta fase se encuentra integrada por cuatro actividades: Exploración Inicial del Proyecto (sección 5.2.1.1.1): donde se identifican los miembros de interés para el proyecto y las posibles situaciones de riesgo durante el desarrollo del mismo, Definición de la Comunicación (sección 5.2.1.1.2): se prevén las necesidades y canales de comunicación durante el desarrollo del proyecto, Evaluación de la Situación (sección 5.2.1.1.3): se analizan las herramientas de utilidad para el desarrollo del proyecto, determinando la viabilidad del mismo y Definición del Ciclo de Vida (sección 5.2.1.1.4): donde se establece de acuerdo a las características del proyecto, el flujo mediante el cual se llevarán a cabo las tareas de desarrollo.

Se destaca que las actividades que integran la fase, son realizadas de manera paralela a la fase Entendimiento del negocio del subproceso Desarrollo (sección 5.2.2.1), utilizando resultados de dicha fase como elementos de entrada. Las dependencias son señaladas al inicio de cada actividad, indicando los formalismos que utilizan.

5.2.1.1.1. Actividad: Exploración Inicial del Proyecto (G.In.EIP)

Mediante esta actividad se identifican y describen las personas involucradas en el proyecto, los posibles riesgos y las acciones en caso de contingencia.

A continuación se presentan los resultados obtenidos de aplicar la técnica caracterización del desarrollo del proyecto (sección 4.3.1.1.2, pág. 79) perteneciente a la metodología para la educción de requerimientos para proyectos de explotación de información, la cual utiliza como insumos los formalismos: Fuentes de Información del Cliente (Tabla 5.82), Objetivos del Proyecto (Tabla 5.84), Criterios de Éxito del Proyecto (Tabla 5.85), Expectativas del Proyecto (Tabla 5.86), Restricciones del Proyecto (Tabla 5.87), Problema del Negocio (Tabla 5.88) y Criterios de Éxito del Problema de Negocio (Tabla 5.89). Estos formalismos, si bien son presentados de forma posterior en el documento (sección 5.2.2.1), fueron desarrollados en paralelo durante sus etapas iniciales.

Recursos Humanos Involucrados (G.In.EIP.ReHI): se registra la información de los miembros involucrados en el proyecto (de la organización que desarrolla el proyecto y de la organización cliente). Por cuestiones de privacidad, no se presenta la información de contacto de las personas.

En la tabla 5.61, se muestran los cinco involucrados en el proyecto registrados, teniendo en consideración las salvedades previamente mencionada.

Riesgos del Proyecto (G.In.EIP.RiPr): a partir de la restricción identificada durante la interacción con el cliente (Fuente de Información 5.8) *"(repr.1) El acceso de los datos debe ser evaluado por la entidad encargada de garantizar que se cumplan las políticas de aseguramiento de la privacidad de los datos personales de los estudiantes",* se identifica como evento crítico para el desarrollo del plan del proyecto, la posibilidad de demorarse en el desarrollo a causa de demoras en la entrega de los datos. El alcance del riesgo es transversal al desarrollo del proyecto (dado que de ocurrir impactaría en el flujo global del desarrollo del proyecto) y se asigna el identificador "risk.1". En la tabla 5.62, se ilustra la información previamente descripta registrada en el formalismo.

Recursos Humanos Involucrados					
Responsable:	Ramón G.			**Fecha:**	28/07/2015
ID#:	G.In.EIP.ReHI			**Versión:**	1.0
ID	**Nombre**	**Rol/Posición**	**Pertenece a**	**Descripción**	**Información de contacto**
rehi.1	Ramón G.	Líder de Proyecto	recurso interno	Persona encargada de la dirección del proyecto	Skype: XXXX
rehi.2	Ezequiel B.	Ingeniero de Explotación de información	recurso interno	Ingeniero de Explotación de información Senior	Skype: XXXXX
rehi.3	Diego J.	Ingeniero de Explotación de información	recurso interno	Ingeniero de Explotación de información Junior	Skype: XXXXX
rehi.4	Laura D.	Cliente	Organización Contratante	Cliente Experto en el área	Correo: xxxx@gmail.com Skype: XXXXX
rehi.5	Jorge P.	Interesado (Personal de Informática)	Organización Contratante	Responsable de las fuentes de información	Correo: xxxx@gmail.com

Tabla 5.61. Caso de Validación: Educación Superior - Recursos Humanos Involucrados

Riesgos del Proyecto			
Responsable:	Ramón G.	**Fecha:**	12/08/2015
ID#:	G.In.EIP.RiPr	**Versión:**	1.0
Riesgo	**Descripción**	**Alcance**	**Referencia**
risk.1	Demora en acceso a los datos	proyecto	Restricciones del Proyecto (repr.1)

Tabla 5.62. Caso de Validación: Educación Superior - Riesgos del Proyecto

Plan de Contingencias (G.In.EIP.Pcon): a partir del riesgo "(risk.1) Demora en acceso a los datos" previamente identificado, se define como única acción "ajustar los plazos del proyecto", asignándose como identificador "cont.1". En la tabla 5.63, se ilustra la información previamente descripta registrada en el formalismo.

Plan de Contingencias			
Responsable:	Ramón G.	**Fecha:**	12/08/2015
ID#:	G.In.EIP.PCon	**Versión:**	1.0

Contingencia	Acción	Riesgo asociado	Referencia
cont.1	Ajustes en los plazos del proyecto	(risk.1) Demora en acceso a los datos	

Tabla 5.63. Caso de Validación: Educación Superior - Plan de Contingencias

5.2.1.1.2. Actividad: Definición de la Comunicación (G.In.DeC)

En esta actividad, se establecen estrategias formales de comunicación a partir de la necesidad e intereses de las partes involucradas en el proyecto. A continuación se presentan los resultados obtenidos de aplicar la técnica Definición de la Comunicación (sección 4.3.1.2.2, pág. 85), la cual utiliza como insumo los Recursos Humanos Involucrados (Tabla 5.61).

Plan de Comunicación (G.In.DeC.PCom): en el formalismo de entrada, se identificaron cinco recursos humanos asociados al proyecto, tres miembros del equipo de trabajo y dos expertos/clientes, previendo tres tipos de comunicaciones: de comprensión del proyecto, de reporte de avances y de estado interno del proyecto.

La primera de ellas, tiene como objetivo mantener un continuo vínculo con el experto e interesado del negocio, analizando los alcances y restricciones del proyecto, formando parte tres miembros del proyecto (Ramón G., Ezequiel B. y Laura D.). Dicha comunicación se realiza de manera semanal durante el periodo planificado de entendimiento del negocio. El segundo tipo de comunicación, destinado a mantener al cliente informado durante el desarrollo del proyecto, pudiendo identificar posibles cambios en los intereses o nuevas necesidades. La información se brinda de manera mensual y participan los Ingenieros de Explotación de información y el cliente. Finalmente, el último tipo de comunicación tiene como objetivo mantener al equipo de trabajo informado respecto al estado del proyecto y los problemáticas que pudiesen ocurrir durante el desarrollo del mismo. Participan los miembros del equipo de trabajo y serán realizadas de manera semanal.

El modo en el cual las comunicaciones serán realizadas es por videollamada, mediante la herramienta Skype, siendo el responsable de las mismas el líder del proyecto (en la primera y la última) y el Ingeniero de Explotación de información sénior en la segunda. La tabla 5.64 ilustra el resultado obtenido de aplicar la técnica al caso de validación.

Plan de Comunicación				
Responsable:	Ramón G.	**Fecha:**		28/07/2015
ID#:	G.In.DeC.PCom	**Versión:**		1.0
Interesados	**Información**	**Frecuencia**	**Medio**	**Responsable**
(rehi.1) Ramón G. (rehi.2) Ezequiel B. (rehi.4) Laura D.	Comprensión del Proyecto	semanal durante el periodo de entendimiento del negocio	Skype	(rehi.1) Ramón G.
(rehi.2) Ezequiel B. (rehi.3) Diego J. (rehi.4) Laura D.	Avances del Proyecto	mensual	Skype	(rehi.2) Ezequiel B.
(rehi.1) Ramón G. (rehi.2) Ezequiel B. (rehi.3) Diego J.	Estado del Proyecto	semanal	Skype	(rehi.1) Ramón G.

Tabla 5.64. Caso de Validación: Educación Superior - Plan de Comunicación

5.2.1.1.3. Actividad: Evaluación de la Situación (G.In.EvS)

En esta actividad se analiza la posibilidad de éxito del proyecto, teniendo en consideración los objetivos y las posibles soluciones que brindan las distintas herramientas de explotación de información existentes. Los objetivos de la actividad son: seleccionar las herramientas a utilizar y determinar la viabilidad del proyecto.

A continuación se presentan los resultados obtenidos de aplicar la metodología para la selección de herramientas de explotación de información y el modelo de evaluación de viabilidad para proyectos de explotación de información (sección 4.3.1.3.2, pág. 90), la cual utiliza como insumos los formalismos: Objetivos del Proyecto (Tabla 5.84), Problema del Negocio (Tabla 5.88), Fuentes de Información del Cliente (Tabla 5.82), Recursos Humanos Involucrados (Tabla 5.61), Expectativas del Proyecto (Tabla 5.86) y Suposiciones del Proyecto (no aplicable en este proyecto).

Reporte de Evaluación de Herramientas (G.In.EvS.REHe): para el desarrollo del proyecto, se identifican como posibles herramientas para utilizar de acuerdo a las necesidades del cliente y la experiencia de los miembros del proyecto: Tanagra Versión 1.4.50, Weka Versión 3.7.11 y Orange Versión 2.7.8. A partir de ello, se califica cada una de las características de acuerdo a la escala de valores entre 1 y 4 de acuerdo al tipo de pregunta, señalando con "--" aquellos aspectos que no han sido evaluados. Como resultado se selecciona a la herramienta tanagra como la más adecuada para el proyecto. En la tabla 5.65 se ilustra las valoraciones realizadas y los resultados obtenidos para cada una de las herramientas evaluadas.

Reporte de Evaluación de Herramientas					
Responsable:	Ezequiel B.			**Fecha:**	25/08/2015
ID#:	G.In.EvS.REHe			**Versión:**	1.0
Criterios:					
Evaluación: 1 = Malo, 2 = débil, 3 = Bueno, 4 = Excelente					1 = No, 4 = SI
Herramientas		**Peso**	**Tanagra V.1.4.50**	**Weka V.3.7.11**	**Orange V.2.7.8**
1. Funcional - Características Técnicas					
Soporte de Metodología / Ciclo de vida	Soporte del proceso	3	2	2	2
Compatibilidad con fuentes de datos	Base de datos	8	--	--	--
	Otras fuentes (word, excel, etc.)	8	3	2	3
Integración	Soporte de distintas técnicas asociadas al proceso de explotación de Información	5	4	4	4
Multilenguaje	Soporta distintas idiomas	2	1	1	1
Técnicas	Variedad de técnicas que provee	18	4	4	4
Reporte y visualización	Permite generar reportes y visualizaciones	12	2	2	2
Multiplataforma	Soporta múltiples plataformas	5	1	4	4
Instalación remota	La administración y mantenimiento son remotos	5	--	--	--
Usuarios Múltiples	Posee perfiles de usuarios	2	1	1	1
Seguridad	Provee seguridad de la información configurada por perfiles	2	1	1	1
Backup	Metodología de backup	2	1	1	1
Amigable	Interfaz de usuario	10	4	2	4
Configuraciones	Permite la configuración del perfil	8			
Documentación	Servicio de soporte y ayuda	5	4	1	3
Conexión	Soporta conexión por: Internet, FTP, ERPs.	2	1	1	1
Soporte de sistemas de mensaje	Soporta compartir información (por mail u otro medio)	3	1	1	1
Total			224	196	234
	Peso del Grupo	40%	89,6	78,4	93,6
2. Características del Proveedor					
Características del proveedor	Historia	30	3	3	1
Crecimiento	Perspectiva a futuro	10	2	3	2
Ubicación Geográfica	Oficinas	30	--	--	--
Implementación	Otras implementaciones de la misma herramienta	5	--	--	--
	Contacto con otros clientes	5	--	--	--
Confidencialidad	Confidencialidad de la información	20	--	--	--
Total			110	120	50
	Peso del Grupo	25%	27,5	30	12,5
3. Características del Servicio					
Garantía del producto	Duración y Alcance	30	--	--	--
Mejora	Brinda soporte a versiones previas	20	1	1	1
Licencia	Costo, alcances y soporte postventa	30	--	--	--
Soporte	Tiempo de respuesta y disponibilidad	20	--	--	--
Total			20	20	20
	Peso del Grupo	20%	4	4	4

Tabla 5.65.a. Caso de Validación: Educación Superior - Reporte de Evaluación de herramientas

4. Características Económicas					
Costo del software	Costo de la herramienta	30	--	--	--
Costo del Hardware	Necesidad de mejorar o comprar nuevo hardware compatible con la herramienta	20	--	--	--
Otros costos software	Costos adicionales al producto (backup, web servers, bases de datos, etc.)	20	--	--	--
Licencias	Política de licencia	10	--	--	--
Financiamiento	Existencia	10	--	--	--
Mejoras	Costo promedio de la mejora del producto	10	--	--	--
Total			0	0	0
	Peso del Grupo	-15%	0	0	0
Final					
1. Funcional - Características Técnicas		40%	89,6	78,4	93,6
2. Características del Proveedor		25%	27,5	30	12,5
3. Características del Servicio		20%	4	4	4
4. Características Económicas		-15%	0	0	0
TOTAL			**121,1**	**112,4**	**110,1**

Tabla 5.65.b. Caso de Validación: Educación Superior - Reporte de Evaluación de herramientas

Reporte de Evaluación de Viabilidad (G.In.EvS.REVi): a partir de la información recabada sobre la fuente de datos, el problema de negocio y los miembros que forman parte del mismo, se determinan las trece características a considerar, definiendo las valoraciones por dimensión y global. De los resultados obtenidos, se verifica en primera instancia que las valoraciones individuales de cada característica son superiores al umbral, y que las valoraciones de las dimensiones y global son mayores a 5, por lo cual se determina al proyecto como viable, identificándose un alto nivel de soporte en los aspectos de plausibilidad (disposición de los elementos necesarios para el desarrollo exitoso del proyecto) y éxito (el apoyo de los interesados y la disposición para el desarrollo del proyecto). En la tabla 5.66 se ilustra las valoraciones realizadas y los resultados obtenidos para cada dimensión del proyecto.

Reporte de Evaluación de Viabilidad												
Responsable:	Ramón G.							**Fecha:**			27/08/2015	
ID#:	G.In.EvS.REVi							**Versión:**			1.0	
Datos						**Problema de Negocio**			**Proyecto**		**Equipo de Trabajo**	
P1	**P2**	**A1**	**A2**	**A3**	**E1**	**P3**	**A4**	**A5**	**E2**	**E3**	**P4**	**E4**
Todo	Todo	Todo	Mucho	Mucho	Todo	Mucho	Todo	Mucho	Todo	Mucho	Mucho	Regular
Umbral												
poco	poco	poco	poco	poco	nada	poco	poco	poco	nada	nada	poco	nada
Dimensiones						***Viabilidad global***					***Resultado***	
Plausibilidad					8,24	***7,1***					***Viable***	
Adecuación					5,66							
Éxito					7,49							

Tabla 5.66. Caso de Validación: Educación Superior - Evaluación de Viabilidad

5.2.1.1.4. Actividad: Definición del Ciclo de Vida (G.In.DCV)

Es en la actividad actual, se analizan las características del proyecto con el objetivo de definir la estrategia de implementación más adecuada para el desarrollo del mismo. Del resultado de esta actividad se establece la estructura y el flujo de ejecución de las fases en el caso de validación.

A continuación se presentan los resultados obtenidos de aplicar la técnica Selección del Ciclo de Vida (sección 4.3.1.4.2, pág. 97), la cual utiliza como insumos Objetivos del Proyecto (Tabla 5.84), Expectativas del Proyecto (Tabla 5.86), Problema del Negocio (Tabla 5.88), Riesgos del Proyecto (Tabla 5.62) y Recursos Humanos Involucrados (Tabla 5.61).

Modelo de Ciclo de Vida (G.In.DCV.MoCV): debido a que el cliente tiene bien definidas las necesidades y problemáticas del proyecto, y la experiencia del equipo en proyectos similares, se determina utilizar el ciclo de vida DMLC, basado en el definido en CRISP-DM (sección 2.4.1.8, pág. 43), destacándose además que el equipo conoce ampliamente dicha dinámica de trabajo. Se establece como criterio de transición, la finalización del 80% de las actividades de la fase. La tabla 5.67 ilustra el formalismo generado, presentando la visualización del modelo de ciclo de vida y su criterio de transición.

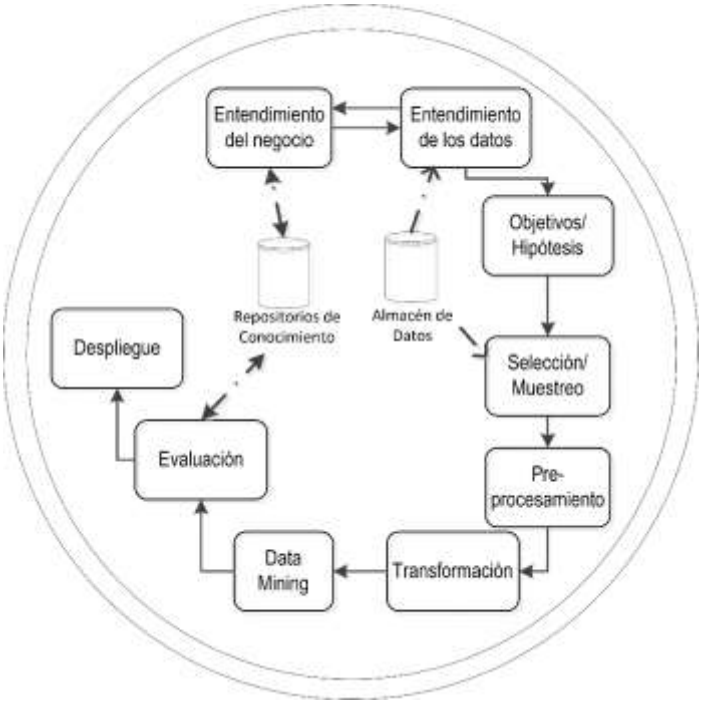
Modelo de Ciclo de Vida			
Responsable:	Ramón G.	**Fecha:**	28/08/2015
ID#:	G.In.DCV.MoCV	**Versión:**	1.0
			
Criterios de transición: Las fases posteriores podrán iniciarse una vez alcanzado el 80% de las tareas asociadas a la fase vigente			

Tabla 5.67. Caso de Validación: Educación Superior - Modelo de Ciclo de Vida

5.2.1.2. Fase: Planificación (G.Pl)

En la fase Planificación se define el curso de acciones requeridos para alcanzar los objetivos del proyecto. Se encuentra conformada por cuatro actividades: Planificación de la Mediciones (sección 5.2.1.2.1), Planificación de las Actividades (sección 5.2.1.2.2), Planificación de los Recursos (sección 5.2.1.2.3) y Planificación de las Responsabilidades (sección 5.2.1.2.4).

Se destaca que las actividades que integran la fase son realizadas de manera paralela a la fase Entendimiento del negocio del subproceso Desarrollo, utilizando resultados de dicha fase como elementos de entrada. Dichas dependencias son señaladas al inicio de cada actividad, indicando los formalismos que utilizan.

5.2.1.2.1. Actividad: Planificación de la Mediciones (G.Pl.PlM)

En esta actividad se realiza una estimación inicial del tiempo requerido para el desarrollo del programa del proyecto y se definen las mediciones que se llevarán a cabo durante el transcurso del mismo.

A continuación se presentan los resultados obtenidos de aplicar las técnicas Métricas para Proyectos de Explotación de Información y Modelo de Estimación para Proyectos de Explotación de Información (sección 4.3.2.1.2, pág. 102), las cuales utilizan como insumo los siguientes formalismos: Objetivos del Proyecto (Tabla 5.84), Problema del Negocio (Tabla 5.88), Fuentes de Información del Cliente (Tabla 5.82), Recursos Humanos Involucrados (Tabla 5.61) y Reporte de Evaluación de Herramientas (Tabla 5.65).

Listado de Métricas (G.Pl.PlA.LiMe): para este proyecto se propone el uso de cuatro métricas, dos enfocadas en el proyecto, una al modelo y la última a los datos. La tabla 5.68 se presenta cada una de ellas detallando su forma de cálculo.

Estimación del Proyecto (G.Pl.PlA.EsPr): de acuerdo a las características del proyecto, se evalúan y fijan los factores de acuerdo a las escalas establecidas en la técnica. A partir de los formalismos objetivos del proyecto y problema de negocio, se identifica que las necesidades del cliente están asociadas con “conocer los atributos que caracterizan a grupos sin poseer una clasificación conocida previamente”, asignando el valor “3” al factor OBTY. A partir del formalismo fuente de información del cliente, se identifica que se poseen dos repositorios de datos con tecnología compatible para su integración (asignando el valor “2” al factor AREP), la cantidad de tuplas existentes es aproximadamente de 1500 (asignando el valor “2” al factor QTUM), no

Listado de Métricas				
Responsable:	Ramón G.		**Fecha:**	31/08/2015
ID#:	G.PI.PIA.LiMe		**Versión:**	1.0
Nombre	**Tipo**	**Fórmula**	**Comentarios**	
Tiempo total requerido para el desarrollo del proyecto	Proyecto	DRPY = $\sum$ trA trA = tiempo requerido por actividad	Sumatoria de los tiempos requeridos para cada actividad del proyecto	
Tiempo medio requerido para el desarrollo de un problema de explotación de información	Proyecto	($\sum$ trA.SD) / NPEI trA.SD = tiempo requerido por actividad del subproceso de desarrollo NPEI = cantidad de problemas de explotación de información	Solo se considera el tiempo de las actividades pertenecientes al subproceso de desarrollo	
Número total de atributos que no son de utilidad en las tablas	Datos	$\sum$ aNu aNu = cantidad de atributos que no son de utilidad		
Número medio de atributos significativos por modelo	Modelo	($\sum$ AtS.M) / NMOD AtS.M = cantidad de atributos significativos de un modelo NMOD = cantidad de modelos		

Tabla 5.68. Caso de Validación: Educación Superior - Listado de Métricas

existen tablas auxiliares (asignando el valor "1" al factor QTUA) y se posee un registro detallado de cada columna de la fuente de datos, sus significados y un ejemplar del elemento utilizado para recabar la información (asignando el valor "1" al factor KLDS). De acuerdo a los interesados del proyecto registrados en Recursos Humanos Involucrados (Tabla 5.61), se identifican las variables LECO y KEXT, con el valor "1". Finalmente, del Reporte de Evaluación de Herramientas se determina las funcionalidades que la herramienta seleccionada brinda, asignando al factor TOOL, el valor "3".

A partir de los valores asignados, se obtiene como esfuerzo total del subproceso desarrollo: 2.08 meses/hombre y de acuerdo al porcentaje estimado para tareas de gestión (15%), se determina como esfuerzo para dicha instancia: 0.31 meses/hombre, siendo el esfuerzo total del proyecto igual a: 2.39 meses/hombre. La tabla 5.69 ilustra las valoraciones realizadas y el resultado obtenido.

Estimación del Proyecto										
Responsable:	Ramón G.				**Fecha:**			03/09/2015		
ID#:	G.PI.PIM.EsPr				**Versión:**			1.0		
Esfuerzo										
OBTY	**LECO**	**AREP**	**QTUM**	**QTUA**	**KLDS**	**KEXT**	**TOOL**	**Total Desarrollo**	**Total Gestión**	**Total**
3	1	2	2	1	1	2	3	2,08	0,31	**2,39**

Tabla 5.69. Caso de Validación: Educación Superior - Estimación del Proyecto

5.2.1.2.2. Actividad: Planificación de las Actividades (G.Pl.PlA)

En esta actividad se prevén las acciones a realizar durante el transcurso del proyecto y sus alcances, definiendo la ejecución de las actividades en el transcurso del tiempo. Como resultado de esta tarea, se genera el programa de actividades del proyecto.

A continuación se presentan los resultados obtenidos de aplicar la técnica Definición del Programa del Proyecto (sección 4.3.2.2.2, pág. 109), la cual utiliza como insumos los formalismos: Modelo de Ciclo de Vida (Tabla 5.67), Estimación del Proyecto (Tabla 5.69), Objetivos del Proyecto (Tabla 5.84), Problema del Negocio (Tabla 5.88) y Fuentes de Información del Cliente (Tabla 5.82)

Mapa de Actividades (G.Pl.PlA.MaAc): A partir del modelo de ciclo de vida seleccionado, se determina las etapas durante las cuales se desarrollarán las distintas actividades del proceso propuesto. La ejecución de las mismas es dependiente a las necesidades del proyecto. La tabla 5.70, presenta la distribución de las actividades a través del modelo de ciclo de vida, como puede observarse las actividades pertenecientes a la fase de soporte y control, se realizan de manera transversal al desarrollo del proyecto.

Mapa de Actividades										
Responsable:		Ramón G.				**Fecha:**		01/09/2015		
ID#:		G.Pl.PlA.MaAc				**Versión:**		1.0		
ID	**Fase/Actividad**	**E.N.**	**E.D.**	**H.**	**S.**	**Pre.P.**	**T.**	**D.M.**	**E.**	**D.**
G.In	**Iniciación**									
G.In.EIP	Exploración Inicial del Proyecto	x								
G.In.DeC	Definición de la Comunicación	x								
G.In.EvS	Evaluación de la Situación	x								
G.In.DCV	Definición del Ciclo de Vida	x								
G.Pl	**Planificación**									
G.Pl.PIM	Planificación de la Mediciones	x	x							
G.Pl.PlA	Planificación de las Actividades	x	x							
G.Pl.PlR	Planificación de los Recursos	x	x							
G.Pl.PRe	Planificación de las Responsabilidades	x	x							
G.So	**Soporte**									
G.So.MeP	Mediciones del Proyecto	x	x	x	x	x	x	x	x	x
G.So.GeC	Gestión de la Configuración	x	x	x	x	x	x	x	x	x
G.Co	**Control**									
G.Co.GeD	Gestión del Desarrollo	x	x	x	x	x	x	x	x	x
G.Co.CoA	Control de las Actividades	x	x	x	x	x	x	x	x	x
G.Co.Gca	Gestión del Cambio	x	x	x	x	x	x	x	x	x

Tabla 5.70.a. Caso de Validación: Educación Superior - Mapa de Actividades

ID	Fase/Actividad	E.N.	E.D.	H.	S.	Pre.P.	T.	D.M.	E.	D.
G.Ci	**Cierre**									
G.Ci.FEC	Formalización Externa del Cierre del Proyecto									x
G.Ci.FIC	Formalización Interna del Cierre del Proyecto									x
D.EN	**Entendimiento del Negocio**									
D.EN.AnN	Análisis del Negocio	x								
D.EN.CPN	Comprensión del Problema de Negocio	x								
D.ED	**Entendimiento de los Datos**									
D.ED.AnD	Análisis de los Datos		x							
D.ED.ExD	Exploración de los Datos		x							
D.ED.EvD	Evaluación de los Datos		x							
D.Mo	**Modelado**									
D.Mo.MoP	Modelado del problema			x						
D.Mo.CoM	Configuración del Modelo			x						
D.PD	**Preparación de los Datos**									
D.PD.CFT	Construcción de la Fuente Temporaria de Datos				x	x				
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos					x	x			
D.Im	**Implementación**									
D.Im.SeM	Selección del Modelo							x		
D.Im.ExI	Explotación de Información							x		
D.EP	**Evaluación y Presentación**									
D.EP.EvR	Evaluación de los Resultados								x	x
D.EP.PrR	Presentación de los Resultados									x

Tabla 5.70.b. Caso de Validación: Educación Superior - Mapa de Actividades

Plan de Acción (G.Pl.PlA.PlAc): a partir de la estimación de tiempos y la selección de las actividades a realizar en cada etapa del modelo de ciclo de vida (mapa de actividades), se asigna la duración y rango de fechas de ejecución de cada una de las actividades usando de base las mediciones de esfuerzo requeridas para proyectos de explotación de información [Rodríguez et al., 2010], teniendo en consideración que para el proyecto se utiliza una base de jornada diaria de cuatro horas.

El plan de acción se mantiene actualizado durante el desarrollo del proyecto, registrándose en los hitos de control y reporte de estado, los avances del proyecto (tiempos y fechas reales). El resultado obtenido al final del proyecto, se presenta en las tablas 5.71.a y 5.71.b. En la sección A.3.1, se presentan las dos versiones previas del formalismo: 1.0 y 1.1 (Tablas A.13 y A.14 respectivamente) y diagrama de Gantt actualizado al cierre del proyecto (Figura A.3).

Plan de Acción								
Responsable:		Ramón G.			**Fecha:**		03/09/2015	
ID#:		G.Pl.PIA.PlAc			**Versión:**		1.2	
ID Actividad	**Actividad**	**Inicio Estimado**	**Inicio Real**	**Fin Estimado**	**Fin Real**	**Esfuerzo Estimado (Hs)**	**Esfuerzo Real (Hs)**	**Comentarios**
G.In	**Iniciación**	**27/07/15**	**27/07/15**	**28/08/15**	**28/08/15**	**16**	**16**	
G.In.EIP	Exploración Inicial del Proyecto	27/07/15	27/07/15	28/08/15	28/08/15	6	5	
G.In.DeC	Definición de la Comunicación	27/07/15	27/07/15	28/08/15	28/08/15	3	3	
G.In.EvS	Evaluación de la Situación	27/07/15	27/07/15	28/08/15	28/08/15	4	5	
G.In.DCV	Definición del Ciclo de Vida	28/08/15	28/08/15	28/08/15	28/08/15	3	3	
G.Pl	**Planificación**	**31/08/15**	**31/08/15**	**08/09/15**	**08/09/15**	**18**	**20**	
G.Pl.PIM	Planificación de la Mediciones	31/08/15	31/08/15	03/09/15	03/09/15	4	4	
G.Pl.PIA	Planificación de las Actividades	31/08/15	31/08/15	03/09/15	03/09/15	5	6	
G.Pl.PIR	Planificación de los Recursos	03/09/15	03/09/15	04/09/15	04/09/15	4	4	
G.Pl.PRe	Planificación de las Responsabilidades	04/09/15	04/09/15	08/09/15	08/09/15	5	6	
G.So	**Soporte**	**27/07/15**	**27/07/15**	**17/11/15**	**17/11/15**	**8**	**8**	
G.So.MeP	Mediciones del Proyecto	27/07/15	27/07/15	17/11/15	17/11/15	4	4	Se prevé el registro de las métricas a mitad (09/09/15) y final del tiempo estimado para el proyecto
G.So.GeC	Gestión de la Configuración	27/07/15	27/07/15	17/11/15	17/11/15	4	4	
G.Co	**Control**	**27/07/15**	**27/07/15**	**20/11/15**	**20/11/15**	**14**	**6**	
G.Co.GeD	Gestión del Desarrollo	27/07/15	27/07/15	20/11/15	20/11/15	6	4	Se prevé la aplicación del reporte de estado a mitad (09/09/15) y final del tiempo estimado para el proyecto
G.Co.CoA	Control de las Actividades	27/07/15	27/07/15	20/11/15	20/11/15	6	2	
G.Co.Gca	Gestión del Cambio	27/07/15	27/07/15	20/11/15	20/11/15	2	0	
G.Ci	Cierre	18/11/15	18/11/15	20/11/15	20/11/15	12	12	
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	18/11/15	18/11/15	19/11/15	19/11/15	4	4	
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	18/11/15	18/11/15	20/11/15	20/11/15	8	8	
D.EN	Entendimiento del Negocio	27/07/15	27/07/15	31/08/15	31/08/15	52	40	
D.EN.AnN	Análisis del Negocio	27/07/15	27/07/15	28/08/15	28/08/15	36	34	
D.EN.CPN	Comprensión del Problema de Negocio	18/08/15	18/08/15	31/08/15	31/08/15	16	6	
D.ED	Entendimiento de los Datos	28/08/15	28/08/15	21/09/15	21/09/15	66	52	
D.ED.AnD	Análisis de los Datos	28/08/15	28/08/15	11/09/15	11/09/15	30	28	
D.ED.ExD	Exploración de los Datos	07/09/15	07/09/15	14/09/15	14/09/15	20	14	
D.ED.EvD	Evaluación de los Datos	14/09/15	14/09/15	21/09/15	21/09/15	16	10	
D.Mo	Modelado	22/09/15	22/09/15	29/09/15	29/09/15	24	20	
D.Mo.MoP	Modelado del problema	22/09/15	22/09/15	25/09/15	25/09/15	16	12	
D.Mo.CoM	Configuración del Modelo	28/09/15	28/09/15	29/09/15	29/09/15	8	8	

Tabla 5.71.a. Caso de Validación: Educación Superior - Plan de Acción (fin del proyecto)

D.PD	**Preparación de los Datos**	**01/10/15**	**01/10/15**	**19/10/15**	**19/10/15**	**34**	**26**	
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	01/10/15	01/10/15	12/10/15	12/10/15	20	16	
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos	12/10/15	12/10/15	19/10/15	19/10/15	14	10	
D.Im	**Implementación**	**19/10/15**	**19/10/15**	**30/10/15**	**30/10/15**	**40**	**34**	
D.Im.SeM	Selección del Modelo	19/10/15	19/10/15	23/10/15	23/10/15	10	8	
D.Im.ExI	Explotación de Información	26/10/15	26/10/15	30/10/15	30/10/15	30	26	
D.EP	**Evaluación y Presentación**	**02/11/15**	**02/11/15**	**16/11/15**	**16/11/15**	**36**	**28**	
D.EP.EvR	Evaluación de los Resultados	02/11/15	02/11/15	06/11/15	06/11/15	14	10	
D.EP.PrR	Presentación de los Resultados	09/11/15	09/11/15	16/11/15	16/11/15	22	18	

Tabla 5.71.b. Caso de Validación: Educación Superior - Plan de Acción (fin del proyecto)

5.2.1.2.3. Actividad: Planificación de los Recursos (G.Pl.PlR)

En esta actividad se prevén los recursos (tanto humanos como materiales) que se necesitan para el desarrollo de las actividades en el tiempo. A continuación se presentan los resultados obtenidos de aplicar la técnica Planificación de los Recursos Necesarios (sección 4.3.2.3.2, pág. 116), la cual utiliza como insumos los formalismos: Recursos Humanos Involucrados (Tabla 5.61), Reporte de Evaluación de Herramientas (Tabla 5.65), Plan de Acción (Tabla 5.71), Problema del Negocio (Tabla 5.88) y Fuentes de Información del Cliente (Tabla 5.82).

Plan de Necesidad de Recursos (G.Pl.PlR.PlNR): para este proyecto, se planifica la necesidad de tres personas durante todo el proyecto: un líder del proyecto y dos ingenieros de explotación de información (IEI, Sénior y Junior). Adicionalmente, se detalla la necesidad de tres computadoras para cada uno de los miembros con las siguientes características: SO Windows 7 (en adelante), RAM 4GB o más y 10GB o más espacio en disco, de acuerdo con las necesidades requeridas por la herramienta seleccionada. En la tabla 5.72 se registra la información previamente descripta en el formalismo.

Plan de Necesidad de Recursos					
Responsable:	Ramón G.			**Fecha:**	04/09/2015
ID#:	G.Pl.PIR.PINR			**Versión:**	1.0
Recursos Humanos					
ID	**Recurso**	**Cantidad**	**F. Inicio**	**F. Fin**	**Descripción**
rhr.1	Líder de Proyecto	1	27/07/15	20/11/15	
rhr.2	IEI Senior	1	27/07/15	20/11/15	
rhr.3	IEI Junior	1	27/07/15	20/11/15	
Recursos Materiales					
ID	**Recurso**	**Cantidad**	**F. Inicio**	**F. Fin**	**Descripción**
rmr.1	Computadora Personal	3	27/07/15	20/11/15	SO windows (7 en adelante) RAM 4 gb o más 10GB o más espacio en disco

Tabla 5.72. Caso de Validación: Educación Superior - Plan de Necesidad de Recursos

5.2.1.2.4. Actividad: Planificación de las Responsabilidades (G.Pl.PRe)

En esta actividad se definen las responsabilidades y obligaciones de las partes involucradas en el proyecto, dejando formalizado quién es el encargado de realizar cada tarea y los compromisos asumidos por cada una de las partes intervinientes en el acuerdo.

A continuación se presentan los resultados obtenidos de aplicar la técnica Designación de Responsabilidades (sección 4.3.2.4.2, pág. 123), la cual utiliza como insumos los formalismos: Recursos Humanos Involucrados (Tabla 5.61), Plan de Comunicación (Tabla 5.64), Plan de Acción (Tabla 5.71), Plan de Necesidad de Recursos (Tabla 5.72), Objetivos del Proyecto (Tabla 5.84), Criterios de Éxito del Proyecto (Tabla 5.85), Expectativas del Proyecto (Tabla 5.86), Restricciones del Proyecto (Tabla 5.87), Problema del Negocio (Tabla 5.88), Criterios de Éxito del Problema de Negocio (Tabla 5.89), Riesgos del Proyecto (Tabla 5.62) y Plan de Contingencias (Tabla 5.63).

Matriz de Responsabilidades (G.Pl.PlR.MaRe): en el formalismo Recursos Humanos Involucrados, se identifican cinco interesados (tres miembros del equipo y dos cliente/experto) introduciendo a cada uno de ellos en una columna, y asignando el nivel de participación en cada una de las actividades de acuerdo al interés de información y el conocimiento de los mismos. La tabla 5.73 ilustra el nivel de participación de cada miembro en las actividades del proyecto.

Propuesta del Proyecto (G.Pl.PlR.PrPr): a partir del Objetivo del Proyecto, el Problema del Negocio y los Criterios de Éxito asociados, se describe como alcance del proyecto: "*La organización desarrolladora se compromete a cumplimentar el proceso de extracción de conocimiento a partir de los datos disponibles, con el objetivo de generar patrones de conocimiento que permitan identificar y caracterizar distintos perfiles de estudiantes, permitiendo proporcionar contribuciones novedosas y valiosas que favorezcan la toma de decisiones en aspectos vinculados con la gestión de la Educación Superior en contextos de masividad.*"

En la sección de Obligaciones y Responsabilidades, se determinan los compromisos asumidos por las partes intervinientes, los cuales se derivan de los formalismos Matriz de Responsabilidades, Plan de Comunicación y el Plan de Acción: "*La parte contratante se compromete a brindar disposición de todos los recursos requeridos en tiempo y forma, informando con una antelación no menor a 72hs en caso que estos no pudiesen ser entregados. Cualquier demora fuera del plazo estipulado, requerirá del ajuste de los plazos del proyecto.*

Matriz de Responsabilidades			
Responsable:	Ramón G.	**Fecha:**	06/09/2015
ID#:	G.Pl.PIR.MaRe	**Versión:**	1.0

Descripción

Niveles de participación:
(R) Responsable: encargado de las tareas asociadas a la actividad.
(E) Ejecución: asignado tareas asociadas a la actividad.
(A) Aprobación: aceptación Final del resultado de la actividad.
(C) Consultado: posee conocimiento relevante para el desarrollo de la actividad.
(I) Informado: requiere estar alerta del progreso de la actividad.

ID Actividad	Actividad	Ramón G. (rehi.1)	Ezequiel B. (rehi.2)	Diego J. (rehi.3)	Laura D. (rehi.4)	Jorge P. (rehi.5)
G.In	**Iniciación**					
G.In.EIP	Exploración Inicial del Proyecto	R	I	I		
G.In.DeC	Definición de la Comunicación	R	I	I	I	
G.In.EvS	Evaluación de la Situación	A	R	I		
G.In.DCV	Definición del Ciclo de Vida	R	I	I		
G.Pl	**Planificación**					
G.Pl.PIM	Planificación de la Mediciones	R	I	I		
G.Pl.PIA	Planificación de las Actividades	R	I			
G.Pl.PIR	Planificación de los Recursos	R	I			
G.Pl.PRe	Planificación de las Responsabilidades	R	I	I	A	
G.So	**Soporte**					
G.So.MeP	Mediciones del Proyecto	I	R	I		
G.So.GeC	Gestión de la Configuración	I	R	I		
G.Co	**Control**					
G.Co.GeD	Gestión del Desarrollo	R	I	I	I	
G.Co.CoA	Control de las Actividades	R	I	I	I	
G.Co.Gca	Gestión del Cambio	A	R	I	I	
G.Ci	**Cierre**					
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	R			A	
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	R	C	C		
D.EN	**Entendimiento del Negocio**					
D.EN.AnN	Análisis del Negocio	E	R	E	C	
D.EN.CPN	Comprensión del Problema de Negocio	E	R	E	C	
D.ED	**Entendimiento de los Datos**					
D.ED.AnD	Análisis de los Datos	I	E	R	C	C
D.ED.ExD	Exploración de los Datos	I	E	R		
D.ED.EvD	Evaluación de los Datos	I	E	R		
D.Mo	**Modelado**					
D.Mo.MoP	Modelado del problema	I	R	E		
D.Mo.CoM	Configuración del Modelo	C	R			
D.PD	**Preparación de los Datos**					
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	I	I	R		
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos		I	R		
D.Im	**Implementación**					
D.Im.SeM	Selección del Modelo	C	R	E		
D.Im.ExI	Explotación de Información	C	R	E	I	
D.EP	**Evaluación y Presentación**					
D.EP.EvR	Evaluación de los Resultados	I	R	E	C	
D.EP.PrR	Presentación de los Resultados	E	R	E	I	

Tabla 5.73. Caso de Validación: Educación Superior - Matriz de Responsabilidades

La contraparte se compromete a dar solución a las problemáticas requeridas por el cliente (ver sección Alcance), así como la veracidad de los resultados presentados. Asimismo, la organización se compromete a informar de forma mensual los avances del proyecto.

Las partes acuerdan como fecha de finalización del proyecto el 20/11/2015."

En la sección de aspectos legales, se describe la restricción del proyecto identificada (repr.1), siendo la petición de datos evaluada por la comisión universitaria. Por último, se registra la conformidad de las partes interesadas. La tabla 5.74 ilustra la información previamente mencionada registrada en el formalismo.

Propuesta del Proyecto			
Responsable:	Ramón G.	**Fecha:**	08/09/2015
ID#:	G.Pl.PlR.PrPr	**Versión:**	1.0
Alcance	La organización desarrolladora se compromete a cumplimentar el proceso de extracción de conocimiento a partir de los datos disponibles, con el objetivo de generar patrones de conocimiento que permitan identificar y caracterizar distintos perfiles de estudiantes, permitiendo proporcionar contribuciones novedosas y valiosas que favorezcan la toma de decisiones en aspectos vinculados con la gestión de la Educación Superior en contextos de masividad.		
Obligaciones y responsabilidades	La parte contratante se compromete a brindar disposición de todos los recursos requeridos en tiempo y forma, informando con una antelación no menor a 72hs en caso que estos no pudiesen ser entregados. Cualquier demora fuera del plazo estipulado, requerirá del ajuste de los plazos del proyecto. La contraparte se compromete a dar solución a las problemáticas requeridas por el cliente (ver sección Alcance), así como la veracidad de los resultados presentados. Asimismo, la organización se compromete a informar de forma mensual los avances del proyecto. Las partes acuerdan como fecha de finalización del proyecto el 20/11/2015.		
Aspectos Legales	El acceso a los datos será provisto mediante petición previa, la cual será evaluada por la comisión universitaria con el fin de garantizar la privacidad de los datos personales de los estudiantes.		
Firma del Contratante: Laura D. **Aclaración:** Laura D.		**Firma de la Contraparte:** Ramón G. **Aclaración:** Ramón G.	

Tabla 5.74. Caso de Validación: Educación Superior - propuesta del Proyecto

5.2.1.3. Fase: Soporte (G.So)

La fase Soporte se encuentra conformada por dos actividades: Mediciones del Proyecto (sección 5.2.1.3.1) y Gestión de la Configuración (sección 5.2.1.3.1).

5.2.1.3.1. Actividad: Mediciones del Proyecto (G.So.MeP)

En esta actividad se calculan las métricas durante el desarrollo del proyecto, dejando registro formal del progreso de los indicadores. El resultado de esta actividad contribuye en la toma de decisiones del líder del proyecto, así como en la evaluación de la calidad del proceso y/o del modelo.

A continuación se presentan los resultados obtenidos de aplicar la técnica Cálculo de Métricas (sección 4.3.3.1.2, pág. 133), que utiliza como insumos el Listado de Métricas (Tabla 5.68) y Plan de Acción (Tabla 5.71).

Registro de Mediciones (G.So.MeP.ReMe): se presentan en la tabla 5.75 los resultados registrados en la última versión del producto interno previamente descripto, existiendo una versión previa, la cuales se ilustra en la Tabla A.15. En dicho formalismo, se registraron los valores a la fecha de creación de las métricas seleccionadas para el proyecto.

Registro de Mediciones			
Responsable:	Ezequiel B.	**Fecha:**	20/11/2015
ID#:	G.So.MeP.ReMe	**Versión:**	1.1
Indicador	**Medición**	**Descripción**	
Tiempo total requerido para el desarrollo del proyecto	**DRPY** = 262hs	Tdesarrollo = 200 Tgestion = 62	
Tiempo medio requerido para el desarrollo de un problema de explotación de información	**DRPEI** = 200hs		
Número total de atributos que no son de utilidad en las tablas	13		
Número medio de atributos significativos por modelo	4		

Tabla 5.75. Caso de Validación: Educación Superior - Registro de Mediciones (fin del proyecto)

5.2.1.3.2. Actividad: Gestión de la Configuración (G.So.GeC)

En esta actividad se realizan las tareas vinculadas con la trazabilidad de los productos generados durante el desarrollo del proyecto, garantizando que en todo momento los miembros del equipo estén informados de las versiones actuales de los resultados producidos en cada fase.

A continuación se presentan los resultados obtenidos de aplicar la técnica Configuración del versionado (sección 4.3.3.2.2, pág. 137). Dicha técnica utiliza como insumos las Reglas de Versionado (Fuente de Información 5.7), existentes en la organización, Reporte de Evaluación del Cambio (no aplicable en este proyecto) y Modelo de Ciclo de Vida (Tabla 5.67).

Se utilizan dos dígitos para reflejar el progreso de los productos a lo largo del proyecto X.Y: el primero de ellos (X) indica la versión mayor del documento, incrementándose de a uno cada vez que se modifican o eliminan elementos del producto (generando incompatibilidad con otros productos o versiones anteriores). En caso que el producto se encuentre en un estadio temprano, el cual no puede ser utilizado para su uso como entrada en otras tareas, este dígito debe ser indicado con cero. El segundo (Y), describe la versión menor del documento, la cual se incrementa en uno reflejando incorporaciones o alteraciones que no modifiquen la funcionalidad en el producto. Cuando se modifique la versión superior, este valor será restituido al valor cero.

Para el registro del estado del proyecto, se utiliza la misma lógica (X.Y). El primer elemento (X) indica alteraciones en las necesidades o estrategias de ejecución del proyecto, mientras que el segundo (Y) representa iteraciones en el ciclo de vida.

Fuente de Información 5.7. Caso de Validación: Educación Superior - Reglas de Versionado

Reporte de Versionado (G.So.GeC.ReVe): durante el desarrollo del proyecto se realizaron cuatro ajustes en sus productos internos, a partir del progreso de las actividades, y las tareas de control. La tabla 5.76 ilustra los resultados registrados para el caso de validación.

Reporte de Versionado					
Responsable:	Ezequiel B.		**Fecha:**	20/11/2015	
ID#:	G.So.GeC.ReVe				
Fecha	**Elemento**	**Versión previa**	**Versión Actual**	**Cambio Asociado**	**Descripción**
28/08/15	Fuentes de Información del Cliente	1.0	1.1	Actualización en la descripción	Registro afectado (fuic.1)
09/09/15	Plan de Acción	1.0	1.1	Actualización de mediciones al fin del proyecto	Cambio asociado con G.Co.GeD.ReEs.1
20/11/15	Plan de Acción	1.1	1.2	Actualización de mediciones al fin del proyecto	Cambio asociado con G.Co.GeD.ReEs.2
20/11/15	Registro de Mediciones	1.0	1.1	Actualización de mediciones al fin del proyecto	Cambio asociado con G.Co.GeD.ReEs.2

Tabla 5.76. Caso de Validación: Educación Superior - Reporte de Versionado

Informe del Estado de la Configuración (G.So.GeC.InEC): durante el desarrollo del proyecto, no han surgido cambios en sus objetivos, en la estrategia de ejecución o iteraciones en el ciclo de vida, por lo cual se identifica una única versión del proyecto junto con el estado interno de sus elementos (las versiones vigentes de los productos internos). Las tablas 5.77.a y 5.77.b ilustran los resultados registrados para el caso de validación.

Informe del Estado de la Configuración			
Responsable:	Ezequiel B.	**Fecha:**	17/11/2015
ID#:	G.So.GeC.InEC		

ID Actividad	**Actividad**	**Elemento**	**Versión del Proyecto V. 1.0 (Actual)**
G.In	**Iniciación**		
G.In.EIP	Exploración Inicial del Proyecto	Recursos Humanos Involucrados Riesgos del Proyecto Plan de Contingencias	1.0 1.0 1.0
G.In.DeC	Definición de la Comunicación	Plan de Comunicación	1.0
G.In.EvS	Evaluación de la Situación	Reporte de Evaluación de Herramientas Reporte de Evaluación de Viabilidad	1.0 1.0
G.In.DCV	Definición del Ciclo de Vida	Modelo de Ciclo de Vida	1.0
G.Pl	**Planificación**		
G.Pl.PIM	Planificación de la Mediciones	Listado de Métricas Estimación del Proyecto	1.0 1.0
G.Pl.PIA	Planificación de las Actividades	Mapa de Actividades Plan de Acción	1.0 1.2
G.Pl.PIR	Planificación de los Recursos	Plan de Necesidad de Recursos	1.0

Tabla 5.77.a. Caso de Validación: Educación Superior - Informe de Estado de la Configuración

G.Pl.PRe	Planificación de las Responsabilidades	Matriz de Responsabilidades Propuesta del Proyecto	1.0 1.0
G.So	**Soporte**		
G.So.MeP	Mediciones del Proyecto	Registro de Mediciones	1.1
G.So.GeC	Gestión de la Configuración	Reporte de Versionado Informe del Estado de la Configuración	- -
G.Co	**Control**		
G.Co.GeD	Gestión del Desarrollo	Reporte de Estado	-
G.Co.CoA	Control de las Actividades	Registro de Riesgos Acontecidos	-
G.Co.Gca	Gestión del Cambio	Reporte de Evaluación del Cambio	-
G.Ci	**Cierre**		
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	Documento de Aceptación	-
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	Reporte de Cierre	-
D.EN	**Entendimiento del Negocio**		
D.EN.AnN	Análisis del Negocio	Fuentes de Información del Cliente Definiciones, Acrónimos y Abreviaciones Objetivos del Proyecto Criterios de Éxito del Proyecto Expectativas del Proyecto Suposiciones del Proyecto Restricciones del Proyecto	1.1 1.0 1.0 1.0 1.0 1.0 1.0
D.EN.CPN	Comprensión del Problema de Negocio	Problema de Negocio Criterios de Éxito del Problema de Negocio	1.0 1.0
D.ED	**Entendimiento de los Datos**		
D.ED.AnD	Análisis de los Datos	Diccionario de Fuente de Datos Campos Relacionados con el Problema de Negocio	1.0 1.0
D.ED.ExD	Exploración de los Datos	Reporte de Datos Explorados Fuente Integrada de datos	1.0 -
D.ED.EvD	Evaluación de los Datos	Reporte de la Calidad de los Datos	1.0
D.Mo	**Modelado**		
D.Mo.MoP	Modelado del problema	Diseño del Proceso de Explotación de Información	1.0
D.Mo.CoM	Configuración del Modelo	Selección de Algoritmos de Explotación de Información Selección de Variables del Modelo Estrategias de Evaluación de Modelos	1.0 1.0 1.0
D.PD	**Preparación de los Datos**		
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	Reporte de Generación de la Fuente Temporaria de datos Fuente Temporaria de Datos	1.0 -
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos	Reporte de Adecuación de la Fuente Temporaria de Datos	1.0
D.Im	**Implementación**		
D.Im.SeM	Selección del Modelo	Reporte de Estrategia de Parametrización del Modelo	1.0
D.Im.ExI	Explotación de Información	Reporte de Implementación del Modelo	1.0
D.EP	**Evaluación y Presentación**		
D.EP.EvR	Evaluación de los Resultados	Reporte de Evaluación de los Resultados	1.0
D.EP.PrR	Presentación de los Resultados	Reporte del Proyecto	1.0

Tabla 5.77.b. Caso de Validación: Educación Superior - Informe de Estado de la Configuración

5.2.1.4. Fase: Control (G.Co)

Esta fase está conformada por tres actividades: Gestión del Desarrollo (sección 5.2.1.4.1), Control de las Actividades (sección 5.2.1.4.2) y Gestión del Cambio (sección 5.2.4.1.3).

5.2.1.4.1. Actividad: Gestión del Desarrollo (G.Co.GeD)

En esta actividad se realiza el seguimiento del proyecto, dejando registro formal del progreso del mismo. El resultado de esta actividad contribuye en la toma de decisiones del líder del proyecto, con respecto al cumplimiento de lo planificado, pudiendo identificarse la necesidad de reajustar las acciones definidas.

A continuación se presentan los resultados obtenidos de aplicar la técnica Seguimiento de Avance (sección 4.3.4.1.2, pág. 143) en el proyecto, utilizando como insumos: el Plan de Acción (Tabla 5.71) y el Registro de Mediciones (Tabla 5.75).

Reporte de Estado (G.So.GeD.ReEs): de acuerdo al progreso del proyecto en los distintos hitos de control definidos en el plan de acción, se realizó el primer reporte de estado en la fecha 09/09/2015 y en segundo correspondiente a la finalización del proyecto (tablas 5.78 y 5.79, respectivamente), registrándose los desvíos temporales con respecto a lo planificado, junto con los logros y desafíos alcanzados durante dichos periodos.

Reporte de Estado			
Responsable:	Ramón G.	**Fecha:**	09/09/15
ID#:	G.Co.GeD.ReEs.1		
Evaluación del Programa			
Global	**-13.64%**		**(% por debajo de lo planificado)**
Desarrollo	**-23.08%**	**Gestión**	**11.11%**
Entendimiento del Negocio	-23.08%	Iniciación	0%
Entendimiento de los Datos	-	Planificación	11.11%
Modelado	-	Soporte	-
Preparación de los Datos	-	Control	-
Implementación	-	Cierre	-
Evaluación y Presentación	-		
Descripción: *Se evaluaron las actividades finalizadas hasta la fecha. (ver G.Pl.PIA.PIAc versión 1.1)*			
Situaciones identificadas que requieren de seguimiento - **Cambios durante el periodo (alcances, tiempos)** - **logros principales durante el periodo** Definición de los alcances del proyecto y los problemas de negocio			

Tabla 5.78. Caso de Validación: Educación Superior - Reporte de Estado (G.Co.GeD.ReEs.1)

Reporte de Estado			
Responsable:	Ramón G.	**Fecha:**	20/11/15
ID#:	G.Co.GeD.ReEs.2		
Evaluación del Programa			
Global	**-18.13%**		**(% por debajo de lo planificado)**
Desarrollo	**-20,63%**	**Gestión**	**-8.82%**
Entendimiento del Negocio	-23,08%	Iniciación	0%
Entendimiento de los Datos	-21,21%	Planificación	11.11%
Modelado	-16,67%	Soporte	0%
Preparación de los Datos	-23,53%	Control	-57,14%
Implementación	-15,00%	Cierre	0,00%
Evaluación y Presentación	-22,22%		
Descripción: Al cierre del proyecto			
Situaciones identificadas que requieren de seguimiento - **Cambios durante el periodo (alcances, tiempos)** - **logros principales durante el periodo** Se validan los patrones asociados a las necesidades del cliente			

Tabla 5.79. Caso de Validación: Educación Superior - Reporte de Estado (G.Co.GeD.ReEs.2)

5.2.1.4.2. Actividad: Control de las Actividades (G.Co.CoA)

En esta actividad se evalúan las situaciones potencialmente peligrosas para el desarrollo del proyecto, realizando un seguimiento, control y registro de acontecimientos, así como de las acciones realizadas. El resultado de esta actividad contribuye en la calidad del proceso.

Registro de Riesgos Acontecidos (G.Co.CoA.ReRA): durante el desarrollo del caso de validación seleccionado no han acontecido riesgos.

5.2.1.4.3. Actividad: Gestión del Cambio (G.Co.Gca)

En la actividad de Gestión del Cambio se realiza un proceso de evaluación formal de las peticiones de modificación de los distintos aspectos del proyecto, determinando como resultado la procedencia o no del mismo y sus efectos asociados.

Reporte de Evaluación del Cambio (G.Co.Gca.RECa): durante el desarrollo del caso de validación, no se han realizado peticiones de cambios.

5.2.1.5. Fase: Cierre (G.Ci)

La fase Cierre está compuesta por dos actividades: Formalización Externa del Cierre del Proyecto (sección 5.2.1.5.1) y Formalización Interna del Cierre del Proyecto (sección 5.2.1.5.2).

5.2.1.5.1. Actividad: Formalización Externa del Cierre del Proyecto (G.Ci.FEC)

En esta actividad se obtiene la conformidad del cliente, respecto a los compromisos asumidos en la propuesta del proyecto, dejando registro formal de la finalización del mismo.

A continuación se presentan los resultados obtenidos de aplicar la técnica Presentación de Conformidad (sección 4.3.5.1.2, pág. 156), la cual utiliza como insumos los formalismos: Reporte de Evaluación de los Resultados (Tabla 5.102), Registro de Riesgos Acontecidos (no aplicable en este proyecto), Plan de Acción (Tabla 5.71) y Propuesta del Proyecto (Tabla 5.74).

Documento de Aceptación (G.Ci.FEC.DoAc): en el proyecto, se deriva a partir de los objetivos, las problemáticas identificadas y los criterios de éxito, los siguientes alcances del proyecto: "*Los objetivos definidos consisten en generar patrones de conocimiento que permitan identificar y caracterizar distintos perfiles de estudiantes, permitiendo proporcionar contribuciones novedosas y valiosas que favorezcan la toma de decisiones en aspectos vinculados con la gestión de la Educación Superior en contextos de masividad.*".

A partir de lo registrado en el plan de acción y la Propuesta del Proyecto, se describen las conclusiones sobre el desarrollo de lo planificado: "*Se realizaron dos informes del progreso de las actividades (en las fechas 09/09/2015 y 17/11/2015) cumplimentando con las obligaciones pactadas (fecha de finalización del proyecto el 20/11/2015).”* y del Reporte de Evaluación de los Resultados, se describen las conclusiones del proyecto: "*Como resultado de la evaluación realizada por el experto Laura D., se concluyó que los perfiles identificados proveen aportes novedosos y de interés, para confirmar y generar nuevas políticas. Considerándose los aspectos de procedencia de poco interés. El proyecto fue finalizado acorde a lo estipulado, en tiempo y forma. Mediante la presente se deja de manifiesto que se ha cumplimentado exitosamente los requerimientos realizados, dando por finalizado el proyecto.*". Por último, el cliente certifica el cumplimiento exitoso del proyecto. En la tabla 5.80 se ilustra el formalismo generado.

Documento de Aceptación	
	Fecha: 19/11/2015
Objetivos	Los objetivos definidos consisten en generar patrones de conocimiento que permitan identificar y caracterizar distintos perfiles de estudiantes, permitiendo proporcionar contribuciones novedosas y valiosas que favorezcan la toma de decisiones en aspectos vinculados con la gestión de la Educación Superior en contextos de masividad.
Programa	Se realizaron dos informes del progreso de las actividades (en las fechas 09/09/2015 y 17/11/2015) cumplimentando con las obligaciones pactadas (fecha de finalización del proyecto el 20/11/2015).
Conclusiones	Como resultado de la evaluación realizada por el experto Laura D., se concluyó que los perfiles identificados proveen aportes novedosos y de interés, para confirmar y generar nuevas políticas. Considerándose los aspectos de procedencia de poco interés. El proyecto fue finalizado acorde a lo estipulado, en tiempo y forma. Mediante la presente se deja de manifiesto que se ha cumplimentado exitosamente los requerimientos realizados, dando por finalizado el proyecto.
Firma: Laura D. **Aclaración:** Laura D.	

Tabla 5.80. Caso de Validación: Educación Superior - Documento de Aceptación

5.2.1.5.2. Actividad: Formalización Interna del Cierre del Proyecto (G.Ci.FIC)

En esta actividad se llevan a cabo las últimas tareas del proyecto, evaluándose el desempeño del equipo de trabajo, la propuesta, las acciones realizadas y el cumplimiento del plan de acción. Como resultado de esta actividad se resume el progreso del proyecto, dejando registro de aquellos aspectos que sean de valor para futuros desarrollos.

A continuación se presentan los resultados obtenidos de aplicar la técnica Evaluación del Proceso (sección 4.3.5.2.2, pág. 162), la cual utiliza como insumos los formalismos: Plan de Acción (Tabla 5.71), Matriz de Responsabilidades (Tabla 5.73), Registro de Mediciones (Tabla 5.75), Registro de Riesgos Acontecidos (no aplicable en este proyecto), Reporte de Evaluación del Cambio (no aplicable en este proyecto), Reporte de Evaluación de los Resultados (Tabla 5.102) y Documento de Aceptación (Tabla 5.80).

Reporte de Cierre (G.Ci.FIC.ReCi): se describe el objetivo acordado y la validación realizada por el cliente de los resultados obtenidos, y se evalúa el progreso del proyecto con respecto a los tiempos planificados al inicio del mismo, identificando desvíos de sobreestimación para el subproceso Gestión y sobreestimación en Desarrollo, parcialmente debido a la Imposibilidad de utilizar la fuente de información (fuic.2). Adicionalmente, a partir de los resultados registrados en el desarrollo del proyecto, se determina entre el equipo de trabajo los desafíos principales abordados en el proyecto (asociados con la sobreestimación y la subjetividad en la evaluación del modelo) y como lección aprendida: “Incrementar la frecuencia de los reportes de estado, para favorecer el

seguimiento del proyecto y el apoyo de los interesados.". La tabla 5.81 sintetizan los resultados de interés para la organización como cierre del proyecto.

Reporte de Cierre				
Responsable:	Ramón G.	**Fecha:**	20/11/2015	
ID#:	G.Ci.FIC.ReCi			
Objetivos del Proyecto				
Objetivos		**Resultados**		
Identificar a partir del progreso académico perfiles de estudiantes y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos		Se identifican cinco perfiles de estudiantes, siendo significativos cuatro de ellos, proveyendo aportes novedosos y de interés, para confirmar y generar nuevas políticas. Considerándose los aspectos de procedencia de poco interés.		
Evaluación del Tiempo (en HS.)				
Hito	**Estimado**	**Real**	**% desvío**	**Motivo**
Gestión	**68**	**62**	**-8,82%**	
Iniciación	16	16	0,00%	
Planificación	18	20	11,11%	
Soporte	8	8	0,00%	
Control	14	6	-57,14%	
Cierre	12	12	0,00%	
Desarrollo	**252**	**200**	**-20,63%**	
Entendimiento del Negocio	52	40	-23,08%	
Entendimiento de los Datos	66	52	-21,21%	Imposibilidad de utilizar la fuente de información (fuic.2)
Modelado	24	20	-16,67%	
Preparación de los Datos	34	26	-23,53%	Imposibilidad de utilizar la fuente de información (fuic.2)
Implementación	40	34	-15,00%	
Evaluación y Presentación	36	28	-22,22%	
TOTAL	**320**	**262**	**-18,13%**	
Principales Desafíos **-** Estimación de los tiempos: el modelo utilizado para determinar la carga de trabajo sobreestimó para este proyecto de tamaño pequeño en aproximadamente un 18% el esfuerzo requerido. **-** Imposibilidad de determinar junto con el cliente un criterio de éxito cuantitativamente verificable.				
Lecciones Aprendidas **-** Incrementar la frecuencia de los reportes de estado, para favorecer el seguimiento del proyecto y el apoyo de los interesados.				

Tabla 5.81. Caso de Validación: Educación Superior - Reporte de Cierre

5.2.2. MoProPEI-D: Subproceso Desarrollo (D)

El subproceso Desarrollo, se encuentra conformado por seis fases: Entendimiento del Negocio (sección 5.2.2.1), Entendimiento de los Datos (sección 5.2.2.2), Modelado (sección 5.2.2.3), Preparación de los Datos (sección 5.2.2.4), Implementación (sección 5.2.2.5), y Evaluación y presentación (sección 5.2.2.6).

5.2.2.1. Fase: Entendimiento del Negocio (D.EN)

La fase de entendimiento del negocio se compone de dos actividades: Análisis del Negocio (sección 5.2.2.1.1), donde se identifican las características generales del proyecto y Comprensión del Problema de Negocio (sección 5.2.2.1.2), en la cual se establecen los problemas a resolver.

5.2.2.1.1. Actividad: Análisis del Negocio (D.EN.AnN)

El objetivo de esta actividad es identificar y comprender las metas del proyecto, en base a las necesidades del requirente y los interesados.

A continuación se presentan los resultados obtenidos de aplicar la técnica de definición de los objetivos del proyecto (sección 4.4.1.1.2, pág. 175). Los mismos se obtuvieron a partir del conocimiento extraído en distintas entrevistas con el cliente (Fuentes de Información 5.8) y la información del dominio de negocio (Fuente de Información 5.9).

Fuentes de Información del Cliente (D.EN.ANN.FUIC): a partir de la interacción con el cliente, se identifican dos fuentes de información, del tipo almacén o base de datos, describiendo sus contenidos y responsables. La tabla 5.82 ilustra el resultado obtenido, la cual fue actualizada a partir de la interacción con el responsable, incrementando el detalle de la descripción de la fuente SIU_GUARANÍ, detallando el formato en el cual se proveerá dicha fuente. En Tabla A.16, puede observarse la primera versión del formalismo.

Discursos de los interesados: Entrevistas con el cliente (Laura D.)

Entrevista 1:

El propósito de esta investigación es contribuir a facilitar la apropiación del conocimiento en Educación Superior en contextos de masividad. Proveer de información para un adecuado diseño de las políticas públicas en Educación Superior, despierta el interés para contribuir con una mejor apropiación del conocimiento por parte de la sociedad. En esta dirección, una dimensión relevante es la asociada a las características del estudiante, principal actor de este complejo escenario. El objetivo principal del proyecto consiste en discurrir entre distintos planos del problema del estudiante como actor protagónico del espacio natural histórico de transmisión de saberes hegemónicos de generación en generación, para permitir que emerjan las dimensiones hacia su comprensión y la de los procesos que lo constituyen en un problema público.

En este marco, surge el interés de comprender cuáles son los aspectos de vulnerabilidad en el proceso de formación del estudiante, cuáles son las prácticas que favorecen este proceso, cuáles son las instituciones que normativizan estas acciones, cuáles son las tensiones en juego alrededor de los conocimientos hegemónicos en el contexto de sociedad en red. Emergen dos cuestiones esenciales: ¿cómo comprende este estudiante su compromiso social y político en un contexto de masividad? Y ¿cuáles prácticas e instituciones dan cuenta en sus procesos de formación? De esta interacción es posible pensar el surgimiento del problema, ya su definición, su constitución como público y los mecanismos institucionales que lo facilitan. Lo político se concibe como un proceso que emerge de los conflictos que surgen de las interpretaciones que los actores hacen como consecuencia de sus prácticas y cómo se institucionalizan.

Fuente de Información 5.8.a. Caso de Validación: Educación Superior - Entrevistas 1 y 2

Actualmente la Secretaría de Políticas Universitarias orienta importantes acciones materializadas en Programas de becas a estudiantes como las Becas TIC de finalización de carrera, adquisición de recursos físicos como el Programa de Mejora a la Enseñanza de Grado (PAMEG) y otros más integrales que se dirigen a la carrera en su totalidad como el Programa de Mejora a las carreras de Informática (PROMINF). En este contexto, el proyecto se enfoca en el análisis de los estudiantes de las carreras de Ingeniería de la Universidad Nacional de Córdoba.

¿Cuáles son las problemáticas a analizar?
La detección temprana de factores significativos que faciliten la mejora de los procesos de aprendizaje es un tema de relevancia en la Educación Superior en contextos de masividad, en tanto contribuye al establecimiento de políticas que deriven en la mejora de la calidad de los profesionales egresados. El esfuerzo se orienta a comprender al estudiante de Educación Superior y su compromiso social y político como protagonista del espacio natural de transmisión de los más altos niveles de conocimiento de generación en generación; inmerso además en una realidad contextual de vertiginosos cambios tecnológicos que impactan y redefinen formas de vida, estructuras sociales, culturales y políticas, promoviendo nuevas formas institucionales.

¿Cuáles son sus expectativas con respecto a los alcances y resultados del proyecto?
Se espera lograr un mejor conocimiento de las características del estudiante, protagonista principal de este escenario, permitiendo proporcionar contribuciones novedosas y valiosas que favorezcan la toma de decisiones en aspectos vinculados con la gestión de la Educación Superior en contextos de masividad. Se espera que esta contribución resulte novedosa, tanto en las interpretaciones sobre la información del estudiante que emerjan, así como para la construcción de un diseño metodológico aplicable en las muy diversas prácticas de este escenario.
Además, se espera que, de los resultados del procesamiento de la información del estudiante, surjan hallazgos para enriquecer a esos procesos decisionales con nuevas interpretaciones que impacten en una mejor apropiación del conocimiento.

¿Qué información se disponible para el desarrollo del proyecto?
La Universidad Nacional de Córdoba dispone de 2 sistemas fundamentalmente en los cuales se resguarda la información del estudiante. El Sistema de Información Universitaria SIU_GUARANI, de gestión académica, el cual contiene información académica y socioeconómica de los estudiantes, y las aulas virtuales desarrolladas sobre la plataforma MOODLE las cuales contienen información de las instancias de acreditación. La primera de ellas, dispone de más de 1500 registros de estudiantes que cursaron la materia informática en los años 2012 y 2013 que contienen variada información de los estudiantes tanto del tipo académico, como socio-económico y geográficas, relevada hasta julio de 2014.

¿Cuál es el procedimiento para el acceso a los datos?
No se podrá acceder de forma directa a las fuentes de información, pero el personal de informática (Jorge P.) podrá brindarles acceso al modelado y estructura de las fuentes de datos para poder realizar las peticiones de conjuntos de datos requeridos. Dichas peticiones serán evaluadas por la entidad universitaria con el fin de garantizar la privacidad de los datos personales de los estudiantes.

¿Qué limitaciones entiende que poseen los datos?
Las variables asociadas con las calificaciones de los estudiantes (como por ejemplo: al promedio de calificaciones con y sin aplazos) no deben ser tomadas en cuenta como variables representativas del rendimiento académico, en razón del sesgo proveniente de las subjetividades de los evaluadores al generar esas calificaciones, y de las diversas normativas vigentes en las distintas unidades académicas. En su lugar, se propone la generación de variables representativas a aspectos relativos al desempeño académico y al cumplimiento del plan de carrera; siendo éstas más permeables al momento de realizar comparaciones o generar estándares.

Entrevista 2:

A partir de la necesidad de identificar conocimiento que contribuya en el aprendizaje en contextos de masividad: ¿Qué problemáticas considera relevantes para abordar dicho objetivo?
Es interés comprender e identificar, a partir de variables que midan el progreso académico de los estudiantes con respecto al plan de estudios, distintos perfiles y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos, para así poder tomar decisiones tempranas de apoyo de acuerdo a sus necesidades.

Fuente de Información 5.8.b. Caso de Validación: Educación Superior - Entrevistas 1 y 2

¿Cuáles son las características consideradas que determinan los contextos masivos de interés para la investigación?
Se entiende por contextos de masividad, aquel contexto áulico real y presencial en el cual la comunicación personal entre el docente y el estudiante se ve impedida en razón del número de participantes, para el proyecto tomaremos a la asignatura Informática como eje del análisis.

Fuente de Información 5.8.c. Caso de Validación: Educación Superior - Entrevistas 1 y 2

***Información del dominio del negocio**: Página web Moodle DER*
En http://www.examulator.com/er/2.9/index.html puede observarse la representación de la estructura de la base de datos (Diagrama Entidad Relación) y sus esquemas.

Fuente de Información 5.9. Caso de Validación: Educación Superior - Información del Dominio del Negocio

Fuentes de Información del Cliente					
Responsable:		Ezequiel B.		**Fecha:**	28/08/2015
ID#:		D.EN.ANN.FUIC		**Versión:**	1.1
ID	**Nombre**	**Categoría**	**Responsable**	**Descripción**	
fuic.1	SIU_GUARANI	Almacén de datos	(rehi.5) Jorge P.	Sistema de gestión académica que posee información de los estudiantes del tipo académico, como socio-económico y geográficas, relevados hasta julio de 2014. Se dispone de más de 1500 registros de estudiantes que cursaron la materia informática en los años 2012 y 2013. Los cuales se entregan en formato csv.	
fuic.2	Moodle	Almacén de datos	(rehi.5) Jorge P.	Se registra la información del alumno respecto de la cursada de una materia específica	

Tabla 5.82. Caso de Validación: Educación Superior - Fuentes de Información del Cliente

Definiciones, Acrónimos y Abreviaciones (D.EN.ANN.DEAA): En esta etapa se propone el registro de aquellas terminologías específicas del dominio, que el equipo de trabajo determina de interés para precisar su significado:

- **Contextos de masividad**: contexto áulico real y presencial en el cual la comunicación personal entre el docente y el estudiante se ve impedida en razón del número de participantes.
- **Rendimiento académico:** desempeño del estudiante de acuerdo a lo esperado según su plan de estudios.

En la tabla 5.83 se ilustran los términos previamente descriptos registrados en el formalismo correspondiente.

Objetivos del Proyecto (D.EN.ANN.OBPR): A partir de la interacción con el cliente y los expertos, se identifica de interés el siguiente párrafo: "*El propósito de esta investigación es contribuir a facilitar la apropiación del conocimiento en Educación Superior en contextos de masividad. Proveer de información para un adecuado diseño de las políticas públicas en*

Definiciones, Acrónimos y Abreviaciones			
Responsable:	Ezequiel B.	**Fecha:**	21/08/2015
ID#:	D.EN.AnN.DeAA	**Versión:**	1.0

Nombre	Descripción	Tipo	Referencia
Contextos de masividad	contexto áulico real y presencial en el cual la comunicación personal entre el docente y el estudiante se ve impedida en razón del número de participantes	Definición	Entrevista 2
Rendimiento académico	desempeño del estudiante de acuerdo a lo esperado según su plan de estudios	Definición	Entrevista 1

Tabla 5.83. Caso de Validación: Educación Superior - Definiciones, Acrónimos y Abreviaciones

Educación Superior, despierta el interés para contribuir con una mejor apropiación del conocimiento por parte de la sociedad.", en el cual se señala el objetivo general del cliente. A partir de ello, se registra el mismo como "Obpr.1" en la columna objetivo, y con el propósito de mejorar la comprensión del mismo, se reescribe en la columna descripción el concepto identificado en el párrafo previamente citado, como "*Contribuir con conocimiento que favorezca al diseño de políticas públicas en Educación Superior, facilitando la apropiación del conocimiento en contextos de masividad*" y por último, se registra en la columna referencia, que el mismo se obtuvo en la entrevista 1. En la tabla 5.84 se presenta el formalismo resultante.

Objetivos del Proyecto			
Responsable:	Ezequiel B.	**Fecha:**	24/08/2015
ID#:	D.EN.ANN.OBPR	**Versión:**	1.0

Objetivo	Descripción	Referencia
obpr.1	Contribuir con conocimiento que favorezca al diseño de políticas públicas en Educación Superior, facilitando la apropiación del conocimiento en contextos de masividad	Entrevista 1

Tabla 5.84. Caso de Validación: Educación Superior - Objetivos del Proyecto

Criterios de Éxito del Proyecto (D.EN.ANN.CREP): A partir del objetivo previamente identificado y de los extractos obtenidos de la primera entrevista con el cliente: "*La detección temprana de factores significativos que faciliten la mejora de los procesos de aprendizaje es un tema de relevancia en la Educación Superior en contextos de masividad, en tanto contribuye al establecimiento de políticas que deriven en la mejora de la calidad de los profesionales egresados…*" y "*Se espera lograr un mejor conocimiento de las características del estudiante, protagonista principal de este escenario, permitiendo proporcionar contribuciones novedosas y valiosas que favorezcan la toma de decisiones en aspectos vinculados con la gestión de la Educación Superior en contextos de masividad*", se registra en la columna descripción, el siguiente párrafo: "*Identificar factores significativos en estadios tempranos del estudiante, los cuales faciliten la mejora de los procesos de aprendizaje, los cuales proporcionen contribuciones*

novedosas y valiosas que favorezcan a la toma de decisiones en aspectos vinculados con la gestión de la Educación Superior en contextos de masividad", asignándole el identificador "crexpr.1". El criterio de éxito se encuentra asociado al único objetivo del proyecto (Obpr.1), obteniendo esta información de la entrevista 1 realizada al cliente. Dicha información se registra en las columnas objetivo asociado y referencia respectivamente. En la tabla 5.85 se presenta el criterio de éxito definido previamente descripto.

Criterios de Éxito del Proyecto			
Responsable:	Ezequiel B.	**Fecha:**	24/08/2015
ID#:	D.EN.ANN.CREP	**Versión:**	1.0

Criterio	**Descripción**	**Objetivo asociado**	**Referencia**
crexpr.1	Identificar factores significativos en estadios tempranos del estudiante, los cuales faciliten la mejora de los procesos de aprendizaje, los cuales proporcionen contribuciones novedosas y valiosas que favorezcan a la toma de decisiones en aspectos vinculados con la gestión de la Educación Superior en contextos de masividad	(obpr.1) Contribuir con conocimiento que favorezca al diseño de políticas públicas en Educación Superior, facilitando la apropiación del conocimiento en contextos de masividad	Entrevista 1

Tabla 5.85. Caso de Validación: Educación Superior - Criterios de Éxito del Proyecto

Expectativas del Proyecto (D.EN.ANN.EXPR): Las expectativas del proyecto presentan una visión complementaria a la definición del objetivo asociado, en el cual se vinculan las pretensiones que tienen los clientes/expertos con respecto al producto resultante como respuesta a cada objetivo de proyecto identificado. En el caso de validación se identifican dos expectativas asociadas con el siguiente párrafo, proveniente de la entrevista 1: *"...Se espera que esta contribución resulte novedosa, tanto en las interpretaciones sobre la información del estudiante que emerjan, así como para la construcción de un diseño metodológico aplicable en las muy diversas prácticas de este escenario. Además, se espera que, de los resultados del procesamiento de la información del estudiante, surjan hallazgos para enriquecer a esos procesos decisionales con nuevas interpretaciones que impacten en una mejor apropiación del conocimiento."*. En la tabla 5.86 se presentan las expectativas asociadas al objetivo del proyecto.

Suposiciones del Proyecto (D.EN.ANN.SUPR): Durante la interacción del ingeniero de explotación de información con los interesados, no se identifican hipótesis o conjeturas asociadas al objetivo de negocio del proyecto.

Expectativas del Proyecto			
Responsable:	Ezequiel B.	**Fecha:**	26/08/2015
ID#:	D.EN.ANN.EXPR	**Versión:**	1.0

Expectativa	Descripción	Objetivo asociado	Referencia
expr.1	Provean contribuciones novedosas, tanto en las interpretaciones sobre la información del estudiante que emerjan, así como para la construcción de un diseño metodológico aplicable en las muy diversas prácticas de este escenario	(obpr.1) Contribuir con conocimiento que favorezca al diseño de políticas públicas en Educación Superior, facilitando la apropiación del conocimiento en contextos de masividad	Entrevista 1
expr.2	Enriquecer los procesos decisionales con nuevas interpretaciones que impacten en una mejor apropiación del conocimiento	(obpr.1) Contribuir con conocimiento que favorezca al diseño de políticas públicas en Educación Superior, facilitando la apropiación del conocimiento en contextos de masividad	Entrevista 1

Tabla 5.86. Caso de Validación: Educación Superior - Expectativas del Proyecto

Restricciones del Proyecto (D.EN.ANN.REPR): se identifican aquellos aspectos que presentan limitaciones para el cumplimiento de los objetivos del negocio, los cuales puedan demorar, afectar o imposibilitar el desarrollo de los mismos. Las limitaciones pueden estar asociadas al recurso humano (conocimiento de las técnicas o tecnologías, disponibilidades), a los datos (posibilidad de acceso, calidad) o a cuestiones técnicas del proyecto (hardware o software) u organización (aspectos políticos o legales). En la primera entrevista con el cliente, se identifican dos párrafos a partir de los cuales se definen dos restricciones asociadas a los datos: *"…las peticiones de conjuntos de datos requeridos. Dichas peticiones serán evaluadas por la entidad universitaria con el fin de garantizar la privacidad de los datos personales de los estudiantes."* y *"…las calificaciones de los estudiantes (como por ejemplo: al promedio de calificaciones con y sin aplazos) no deben ser tomadas en cuenta como variables representativas del rendimiento académico, en razón del sesgo proveniente de las subjetividades de los evaluadores al generar esas calificaciones, y de las diversas normativas vigentes en las distintas unidades académicas…".* En la tabla 5.87, se presentan las restricciones del proyecto previamente mencionadas (reescritas para mejorar su comprensibilidad).

Restricciones del Proyecto				
Responsable:		Ezequiel B.	**Fecha:**	28/08/2015
ID#:		D.EN.ANN.REPR	**Versión:**	1.0
Restricción	**Tipo**	**Descripción**	**Objetivo asociado**	**Referencia**
repr.1	datos	El acceso de los datos debe ser evaluado por la entidad encargada de garantizar que se cumplan las políticas de aseguramiento de la privacidad de los datos personales de los estudiantes	(obpr.1) Contribuir con conocimiento que favorezca al diseño de políticas públicas en Educación Superior, facilitando la apropiación del conocimiento en contextos de masividad	Entrevista 1
repr.2	datos	las calificaciones de los estudiantes (como por ejemplo: al promedio de calificaciones con y sin aplazos) no deben ser tomadas en cuenta como variables representativas del rendimiento académico, en razón del sesgo proveniente de las subjetividades de los evaluadores al generar esas calificaciones, y de las diversas normativas vigentes en las distintas unidades académicas	(obpr.1) Contribuir con conocimiento que favorezca al diseño de políticas públicas en Educación Superior, facilitando la apropiación del conocimiento en contextos de masividad	Entrevista 1

Tabla 5.87. Caso de Validación: Educación Superior - Restricciones del Proyecto

5.2.2.1.2. Actividad: Comprensión del Problema de Negocio (D.EN.CPN)

En esta actividad, se presenta una visión detallada de preguntas-problema específicas que el cliente desea responder, las cuales permiten alcanzar los objetivos planteados.

A continuación se introducen los resultados obtenidos de aplicar la técnica de definición del problema de negocio (sección 4.4.1.2.2, pág. 189). Los mismos se obtuvieron a partir del conocimiento extraído en distintas entrevistas con el interesado (Fuente de Información 5.8) y los formalismos derivados de la actividad previa.

Problema del Negocio (D.EN.CPN.PRNE): en la entrevista 2, se identifica el siguiente párrafo: "…*Es interés comprender e identificar, a partir de variables que midan el progreso académico de los estudiantes con respecto al plan de estudios, distintos perfiles y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos, para así poder tomar decisiones tempranas de apoyo de acuerdo a sus necesidades.*", a partir del cual se define el único problema de negocio asociado al proyecto: "*Identificar a partir del progreso académico perfiles de estudiantes y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos*", detallado en la tabla 5.88.

Problema del Negocio				
Responsable:	Ezequiel B.	**Fecha:**	31/08/2015	
ID#:	D.EN.CPN.PRNE	**Versión:**	1.0	
Objetivo del Proyecto	(obpr.1) Contribuir con conocimiento que favorezca al diseño de políticas públicas en Educación Superior, facilitando la apropiación del conocimiento en contextos de masividad			
Problema	**Descripción**		**Experto**	**Referencia**
prne.1	Identificar a partir del progreso académico perfiles de estudiantes y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos		(rehi.4) Laura D.	Entrevista 2

Tabla 5.88. Caso de Validación: Educación Superior - Problema del Negocio

Criterios de Éxito del Problema de Negocio (D.EN.CPN.CEPN): a partir de la interacción con el experto, no se pudo definir un criterio cuantitativo de éxito del problema de negocio, determinando que él mismo será el encargado de evaluar las características que permitan comprender el comportamiento en contexto de masividad. En la tabla 5.89 se presenta el criterio de éxito definido para el problema de negocio identificado en el caso de validación.

Criterios de Éxito del Problema de Negocio				
Responsable:	Ezequiel B.	**Fecha:**	31/08/2015	
ID#:	D.EN.CPN.CEPN	**Versión:**	1.0	
Criterio	**Descripción**	**Problema asociado**		**Referencia**
cepn.1	Identificar características que permitan comprender el comportamiento en contexto de masividad, los cuales serán evaluados por Laura D. (rehi.4)	(prne.1) Identificar a partir del progreso académico perfiles de estudiantes y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos		

Tabla 5.89. Caso de Validación: Educación Superior - Criterios de Éxito del Problema de Negocio

5.2.2.2. Fase: Entendimiento de los Datos (D.ED)

La fase de entendimiento de los datos, está conformada por tres actividades: análisis de los datos (sección 5.2.2.2.1), donde se profundiza en la comprensión del significado de las variables disponibles y sus valores, exploración de los datos (sección 5.2.2.2.2), donde se describe en detalle las variables a considerar por el modelo, y evaluación de los datos (sección 5.2.2.2.3), donde se evalúan los distintos aspectos vinculados con la calidad de las variables seleccionadas.

5.2.2.2.1. Actividad: Análisis de los Datos (D.ED.AnD)

Durante la actividad de análisis de los datos se evalúan las variables disponibles en las distintas fuentes de información, con el objetivo de comprender sus significados, valoraciones, así como cualquier otro aspecto relevante (por ejemplo: valores por defecto del sistema, forma en la cual los datos son recolectados, etc.).

A continuación se presentan los resultados obtenidos de aplicar la técnica Identificación de atributos relacionados con el Problema de Negocio (sección 4.4.2.1.2, pág. 197), utilizando como insumos la información externa provista por los interesados: Discursos de los interesados (Fuente de Información 5.10.a y 5.10.b), y los formalismos producidos en la fase de entendimiento del negocio: Fuentes de Información del Cliente (Tabla 5.82), Restricciones del Proyecto (Tabla 5.87) y Problema del Negocio (Tabla 5.88).

Discursos de los interesados: Entrevista con el personal de Informática (Jorge P.)

Entrevista 3:

¿Cuáles son los procedimientos para requerir acceso a las fuentes de información (Moodle y SIU-Guaraní?
Deberá el profesor enviar un pedido formal en el cual solicite los campos requeridos y el periodo de estudiantes. Aunque, únicamente se podrá brindar la información del sistema guaraní. Luego que el pedido fuese aprobado, se entrega un archivo csv con los resultados de la petición realizada.

¿Qué información se disponible acerca de los estudiantes en el SIU-Guaraní?
La base de datos posee información de la carrera, procedencia, aspectos personales y de su entorno, económicos. Adicionalmente, la fuente se actualiza a partir de las entrevistas que se realizan cada año, dejando registro de la última fecha de relevamiento.
A continuación se provee el listado de variables y su significado provisto por el equipo de informática.
• Fecha de Relevamiento: fecha de última actualización del registro
• Carrera id: identificador de la carrera del estudiante
• Carrera: nombre de la carrera
• Año de ingreso: año en el que él estudiante ingreso a la carrera
• Localidad procedencia: localidad que procede el estudiante
• País procedencia.
• Provincia procedencia.
• Departamento procedencia.
• Fecha de nacimiento.
• Estado civil.
• Cantidad de hijos.
• Vive con: descripción de las personas con las cuales habita el estudiante
• Beca: posee beca
• Fuente de la Beca: de donde proviene la beca
• Costea sus estudios: qué medios utiliza para solventar los costos de su estudio (múltiples valores).
• Obra Social: origen de la obra social (si posee).
• Trabajo: descripción de su trabajo o la intención de obtención.
• Horas semanales de trabajo.
• Padre últimos estudios: últimos estudios escolares realizados por el padre.
• Padre trabajo: descripción de la carga horaria del trabajo o la intención de obtención.
Madre últimos estudios: últimos estudios escolares realizados por la madre.
• Madre trabajo: descripción de la carga horaria del trabajo o la intención de obtención.
• Cantidad de materias cursa primer semestre: monto total de materias que realizó el primer semestre.
• Fecha aprobó materia: fecha de aprobación de la materia de interés.
• Nota aprobó materia: calificación obtenida en la materia de interés.
• Cantidad de materias aprobadas.
• Promedio con aplazo: promedio de calificaciones obtenidas incluyendo las desaprobaciones.
• Promedio sin aplazo: promedio de calificaciones obtenidas excluyendo las desaprobaciones.

Fuente de Información 5.10.a. Caso de Validación: Educación Superior - Entrevista 3 y 4

Reunión con el cliente (Laura D.)

Entrevista 4:

A partir de la interacción con el cliente, se definieron las siguientes variables de interés para el análisis de los problemas de negocio:

- *La fuente de ingresos del alumno: de su propio trabajo, de su familia y/o de beca. Las primeras dos obtienen a partir de la variable Costea sus estudios, con los valores "con su trabajo" y "Con el aporte de familiares" respectivamente. La última, de la variable homónima en la base de datos. Se destaca que el valor uno (1) indica la presencia de dicha fuente de ingreso.*
- *Los últimos estudios alcanzados por su padre y madre. Agrupados por: si realizó hasta el secundario sin finalizar (1), universitario sin finalizar (2), título superior o universitario finalizado (3) y al menos realizó cursos de posgrado (4).*
- *El género y la edad del estudiante. La edad calculada al momento de la encuesta, se calcula a partir de la diferencia entre las variables "F. Relevamiento" y "Fecha Nacimiento". Si la fecha de nacimiento es nula, se asigna el valor uno negativo (-1).*
- *La ubicación de procedencia (generando tres variables booleanas, si es argentino, si es de la Provincia de Córdoba, y si es de Córdoba capital).*
- *Si el alumno aprobó Informática durante la cursada. Indica si el estudiante aprobó la materia hasta un año luego de haber realizado la cursada (1), o no (0). Se utiliza la variable "Fecha Aprob Informatica" y el cuatrimestre de cursada según el plan de estudios.*
- *Si el alumno realizó la cursada de Informática acorde a lo establecido en el plan de estudios. Indica si el estudiante posee hasta tres materias menos que las correspondientes para su cohorte según el plan de estudios. El siguiente algoritmo establece la lógica del progreso del estudiante:*

  Si la cantidad de años en la carrera es mayor a 5 entonces 0,
  Sino:
  Si la cantidad de años en la carrera es menor a 5:
  *Si (cantidad de años en la carrera * 9 – 3 > Cant. Materias Aprob.) entonces 3,*
  Sino 4
  *Si ((cantidad de años en la carrera – 1) * 9 +7 > Cant. Materias Aprob.) entonces 1*
  Sino 2
- *Dos variables que determinan el rendimiento del alumno en su primer año de ingreso y su desempeño en el total de años cursados respecto al plan de estudios.*

La primera indica el tiempo que demora en cursar informática, según el plan de estudios: Si la cursa el mismo año (0), Si la cursa hasta 2 años antes (1), Sino (2).
La segunda, provee una escala que indica el progreso del alumno en el primer semestre de carrera: si realiza 5 materias (3), si realiza 3 o 4 materias (2), si realiza 2 o 3 materias (1), sino 0.
Para el análisis centrado en la provincia de córdoba, se desestimará la variable del país de procedencia.

Fuente de Información 5.10.b. Caso de Validación: Educación Superior - Entrevista 3 y 4

Diccionario de Fuente de Datos (D.ED.AnD.DiFD): a partir de las fuentes de información previamente identificadas (Fuentes de Información del Cliente, tabla 5.10), se procede a registrar en la columna "campo" el nombre de las variables que la componen, categorizando cada uno de los campos existentes de acuerdo al tipo de variable y definiendo su significado. La tabla 5.90 ilustra los atributos disponibles.

Campos Relacionados con el Problema de Negocio (D.ED.AnD.CRPN): a partir de la comprensión de los datos disponibles, se evalúa de forma conjunta con el experto el conjunto de variables relevantes para el problema de negocio identificado, así como aquellos que son necesario construir a partir de otros valores (precisando el procedimiento para su generación).

Diccionario de Fuente de Datos				
Responsable:		Diego J.	**Fecha:**	01/09/2015
ID#:		D.ED.AnD.DiFD	**Versión:**	1.0
Fuente de Información		(fuic.1) SIU_GUARANI		
Campo	**Tipo**	**Descripción**		
año Ingreso	Discreto	Año en el que el estudiante realizó el ingreso a la carrera		
Beca	Nominal	Posee beca		
Cant. Mat. Cursa 1er Sem Año Ingreso	Discreto	Cuantas materias curso el estudiante el primer semestre al ingresar a la universidad		
Cant. Materias Aprob.	Discreto	Cuantas materias tiene aprobadas en la universidad		
Cantidad Hijos	Discreto	Cantidad de hijos		
Carrera	Nominal	Código Único identificador para la carrera		
Carrera Nombre	Nominal	Nombre de la carrera		
Costea sus estudios	Nominal	Mediante qué medios costea sus estudios		
curso info	Discreto	Año en el que el estudiante cursó informática		
Dpto.	Nominal	Departamento de origen		
Estado Civil	Nominal	Estado civil		
F. Relevamiento	Discreto	Fecha en la cual se relevaron los datos		
Fecha Aprob Informatica	Discreto	Cuando aprobó informática		
Fecha Nacimiento	Discreto	Fecha en la que nació		
fuente de la Beca	Nominal	Cuál es la fuente de la cual proviene la beca		
Localidad	Nominal	Localidad de origen		
Madre Trabajo	Nominal	Tipo de trabajo de la madre		
Madre Ultimos estudios	Nominal	Últimos estudios alcanzados por la madre		
Madre Vive	Nominal	La madre vive		
Nota Aprob Informatica	Discreto	Calificación obtenida al aprobar informática (valores enteros de 0 a 10)		
Padre Trabajo	Nominal	Tipo de trabajo del padre		
Padre Ultimos estudios	Nominal	Últimos estudios alcanzados por el padre		
Padre Vive	Nominal	El padre vive		
Pais	Nominal	País de origen		
Promedio c/aplz	Continuo	Promedio de las calificaciones de las materias considerando los aplazos (valores de 0 a 10)		
Promedio s/aplz	Continuo	Promedio de las calificaciones de las materias sin considerar los aplazos (valores de 0 a 10)		
Prov.	Nominal	Provincia de origen		
Sexo	Nominal	Género		
Vive con	Nominal	Grupo de personas con las que vive		

Tabla 5.90. Caso de Validación: Educación Superior - Diccionario de Fuente de Datos

La tabla 5.91 ilustra la selección de campos relacionados con el problema de negocio identificado.

5.2.2.2.2. Actividad: Exploración de los Datos (D.ED.ExD)

En esta actividad se analizan los valores de los campos de interés para los distintos problemas de negocio, con el objetivo de comprender las características de la población o muestra de estudio, identificando relaciones iniciales entre las distintas variables estudiadas.

Campos Relacionados con el Problema de Negocio			
Responsable:	Diego J.	**Fecha:**	11/09/2015
ID#:	D.ED.AnD.CRPN	**Versión:**	1.0
Problema de Negocio	(prne.1) Identificar a partir del progreso académico perfiles de estudiantes y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos		

Nombre	Generar	Descripción	Referencia
trabaja	x	Si en la variable “Costea sus estudios” aparece el valor “con su trabajo” entonces se asigna el valor uno (1), sino cero (0)	Entrevista 4
familia	x	Si “Costea sus estudios” tiene como valor “Con el aporte de familiares” entonces se asigna el valor uno (1), sino cero (0)	Entrevista 4
beca		Posee beca	Entrevista 4 / fuic.1
Aprobo Inf en Cursada	x	Indica si el estudiante aprobó la materia hasta un año luego de haber realizado la cursada (1), o no (0). Se utiliza la variable “Fecha Aprob Informatica”	Entrevista 4
CumplePlan	x	Indica si el estudiante posee hasta tres materias menos que las correspondientes para su cohorte según el plan de estudios. Si la cantidad de años en la carrera es mayor a 5 entonces: 0, Sino: Si la cantidad de años en la carrera es menor a 5: Si (cantidad de años en la carrera * 9 – 3 > Cant. Materias Aprob.) entonces: 3, Sino: 4 Si ((cantidad de años en la carrera – 1) * 9 +7 > Cant. Materias Aprob.) entonces: 1 Sino: 2	Entrevista 4
Demora en cursarla	x	indica el tiempo que demora en cursar informática, según el plan de estudios: Si la cursa el mismo año => 0, Si la cursa hasta 2 años antes => 1, Sino 2.	Entrevista 4
Sexo		Género	Entrevista 4 / fuic.1
RitmoInicial	x	Escala que indica el progreso del alumno en el primer semestre de carrera Si realiza 5 materias (incluyendo informática) => 3, Si realiza 3 o 4 materias (incluyendo informática) => 2, Si realiza 2 o 3 materias (incluyendo informática) => 1, Sino 0.	Entrevista 4
Argentino	x	Si “País” es “Argentino” => 1, Sino 0	Entrevista 4
Cordoba	x	Si “Prov.” es “Córdoba” => 1, Sino 0	Entrevista 4
Capital	x	Si “Dpto.” es Córdoba “Capital” => 1, Sino 0	Entrevista 4
edad	x	Edad al momento de ser encuestado. Se calcula a partir de la diferencia entre las variables “F. Relevamiento” y “Fecha Nacimiento”. Si la fecha de nacimiento es nula, se asigna el valor uno negativo (-1).	Entrevista 4
Padre Ultimos estudios Grupos	x	Si realizó hasta secundario incompleto => 1, Si realizó hasta universitario incompleto => 2, Si posee título superior o universitario completo => 3, Si realizó posgrados (finalizados o cursados) => 4.	Entrevista 4
Madre Ultimos estudios grupos	x	Si realizó hasta secundario incompleto => 1, Si realizó hasta universitario incompleto => 2, Si posee título superior o universitario completo => 3, Si realizó posgrados (finalizados o cursados) => 4.	Entrevista 4

Tabla 5.91. Caso de Validación: Educación Superior - Campos Relacionados con el Problema de Negocio

A continuación se presentan los resultados obtenidos de aplicar la técnica Exploración de los Datos (sección 4.4.2.2.2, pág. 205), la cual utiliza como insumos los formalismos: Suposiciones del Proyecto (no aplicable en este proyecto), Problema del Negocio (Tabla 5.88), Diccionario de Fuente de Datos (Tabla 5.90), Campos Relacionados con el Problema de Negocio (Tabla 5.91) y Reporte de Evaluación de Herramientas (Tabla 5.65).

Fuente Integrada de datos (D.ED.ExD.FuID): a partir del conjunto de campos identificados de interés para el problema de negocio, se procede a integrarlos en una única fuente de información, la cual posee 1533 registros. La figura 5.8 ilustra la estructura de la fuente de integrada de datos mediante su representación en un Diagrama Entidad-Relación.

Figura 5.8. Caso de Validación: Educación Superior - Fuente Integrada de Datos (Diagrama Entidad-Relación)

Reporte de Datos Explorados (D.ED.ExD.ReDE): a partir de la fuente integrada de datos, se describe la distribución de valores para cada atributo significativo para el problema de negocio, resaltando la cantidad de registros y la proporción que los mismos representan con respecto a la muestra. La tabla 5.92 ilustra el resultado obtenido.

5.2.2.2.3. Actividad: Evaluación de los Datos (D.ED.EvD)

En esta actividad se analizan los campos de interés para los distintos problemas de negocio, identificando aquellas características que puedan afectar la calidad del modelo.

A continuación, se exhiben los resultados obtenidos de aplicar la técnica Exploración de la Calidad de los Datos (sección 4.4.2.3.2, pág. 215) en el caso de validación, para el cual se utilizan como elementos de entrada los formalismos: Diccionario de Fuente de Datos (Tabla 5.90), Campos Relacionados con el Problema de Negocio (Tabla 5.91), Reporte de Datos Explorados (Tabla 5.92), Fuente Integrada de datos (Figura 5.8) y Reporte de Evaluación de Herramientas (Tabla 5.65).

Reporte de Datos Explorados			
Responsable:	Diego J.	**Fecha:**	14/09/2015
ID#:	D.ED.ExD.ReDE	**Versión:**	1.0
Problema de Negocio	(prne.1) Identificar a partir del progreso académico perfiles de estudiantes y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos		

ATRIBUTOS CUALITATIVOS		
Nombre	**Valores**	**Distribución**
trabaja	0 1	1317 (85,91%) 216 (14,09%)
familia	0 1	121 (7,89%) 1412 (92,11%)
beca	0 1	1426 (93,02%) 107 (6,98%)
Aprobo Inf en Cursada	0 1	1053 (68,69%) 480 (31,31%)
CumplePlan	0 1 2 3 4	72 (4,70%) 52 (3,39%) 1 (0,07%) 887 (57,86%) 521 (33,99%)
Demora en cursarla	0 1 2	1199 (78,21%) 90 (5,87%) 244 (15,92%)
Sexo	M F	1130 (73,71%) 403 (26,29%)
RitmoInicial	0 1 2 3	334 (21,79%) 22 (1,44%) 985 (64,25%) 192 (12,52%)
Argentino	0 1	26 (1,70%) 1507 (98,30%)
Cordoba	0 1	629 (41,03%) 904 (58,97%)
Capital	0 1	1085 (70,78%) 448 (29,22%)
Padre Ultimos estudios Grupos	0 1 2 3 4	46 (3,00%) 317 (20,68%) 557 (36,33%) 510 (33,27%) 103 (6,72%)
Madre Ultimos estudios grupos	0 1 2 3 4	14 (0,91%) 245 (15,98%) 458 (29,88%) 708 (46,18%) 108 (7,05%)

ATRIBUTOS CUANTITATIVOS				
Nombre	**Min**	**Max**	**Tendencia Central**	**Dispersión**
Edad	-1	64	Media = 21.56	Des. Est. = 3.98
Comentarios:				

Tabla 5.92. Caso de Validación: Educación Superior - Reporte de Datos Explorados

Reporte de la Calidad de los Datos (D.ED.EvD.ReCD): a partir del análisis realizado en la fuente integrada de datos y de la descripción de la misma, se identifica al campo edad con registros anómalos (valor uno negativo). Dicho valor se encuentra presente en ocho registros y representa el valor nulo. La tabla 5.93 ilustra el resultado obtenido del análisis de la calidad de los datos.

Reporte de la Calidad de los Datos			
Responsable:	Diego J.	**Fecha:**	21/09/2015
ID#:	D.ED.EvD.ReCD	**Versión:**	1.0
Problema de Negocio	(prne.1) Identificar a partir del progreso académico perfiles de estudiantes y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos		
Nombre	**Registros**	**Tipo**	**Descripción**
Edad	8	nulos	Valores nulos asignados como con el valor "-1"

Tabla 5.93. Caso de Validación: Educación Superior - Reporte de la Calidad de los Datos

5.2.2.3. Fase: Modelado (D.Mo)

La fase Modelado está conformada por 2 actividades: Modelado del Problema (sección 5.2.2.3.1), en la cual se traduce las necesidades del cliente desde la perspectiva del dominio del negocio a la perspectiva de explotación de información y se identifica a partir de ella los posibles modelos de explotación de información a utilizar y Configuración del Modelo (sección 5.2.2.3.2), donde se establecen las características del modelo y su evaluación a utilizar.

5.2.2.3.1. Actividad: Modelado del Problema (D.Mo.MoP)

En esta actividad se traducen los requerimientos del cliente desde la perspectiva del dominio del negocio a explotación de información. A continuación se presentan los resultados obtenidos de aplicar la técnica Derivación del Proceso de Explotación de Información (sección 4.4.3.1.2, pág. 225), la cual utiliza como elementos de entrada el Problema del Negocio (Tabla 5.88), el Diccionario de Fuente de Datos (Tabla 5.90) y Campos Relacionados con el Problema de Negocio (Tabla 5.91).

Diseño del Proceso de Explotación de Información (D.Mo.MoP.DPEI): a partir del problema de negocio previamente identificado, se procede a representar los aspectos relevantes del mismo (conceptos, atributos, relaciones y valores), generando un conjunto de marcos que permiten sistematizar el proceso de representación del conocimiento desde formato texto al formato gráfico basado en redes semánticas, identificando los distintos elementos de interés para determinar el proceso de explotación de información a utilizar. En el caso de validación, se identifica a partir del problema de negocio, el concepto "estudiante", siendo de interés la identificación de características que definan sus similitudes (representada con los nodos variables), haciendo uso de las variables identificadas en el formalismo Campos Relacionados con el Problema de Negocio. La tabla 5.94, ilustra la gráfica resultante a partir de la cual se identifica al **proceso de Descubrimiento de reglas de pertenencia a grupos** el cual indica como estrategia de implementación el uso de algoritmos de agrupamiento (o clustering) junto con algoritmos de la familia TDIDT.

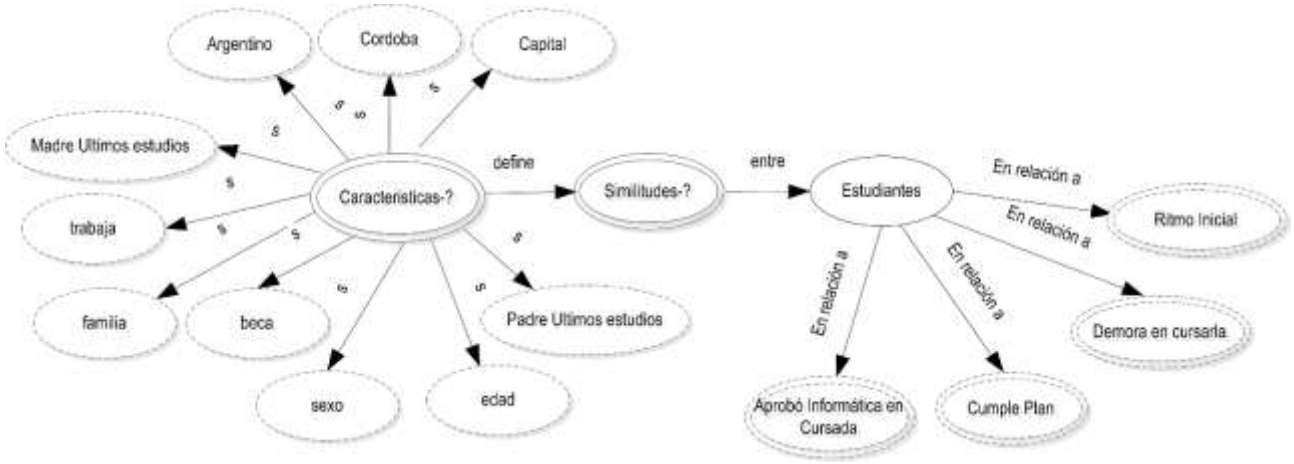
Diseño del Proceso de Explotación de Información			
Responsable:	Ezequiel B.	**Fecha:**	25/09/2015
ID#:	DPEI.1	**Versión:**	1.0
Problema de Negocio:	(prne.1) Identificar a partir del progreso académico perfiles de estudiantes y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos		
			
Proceso de Explotación de Información:	Descubrimiento de reglas de pertenencia a grupos		

Tabla 5.94. Caso de Validación: Educación Superior - Diseño del Proceso de Explotación de Información

5.2.2.3.2. Actividad: Configuración del Modelo (D.Mo.CoM)

En la actividad de configuración del modelo se definen los elementos que conforman la estrategia de implementación y evaluación de los distintos modelos para la extracción de patrones vinculados con el problema de negocio.

A continuación se presentan los resultados obtenidos de aplicar la técnica Determinación de la Configuración del Modelo (sección 4.4.3.2.2, pág. 233), la cual utiliza como insumos los formalismos: Diseño del Proceso de Explotación de Información (Tabla 5.94), para definir los algoritmos a utilizar junto con Diccionario de Fuente de Datos (Tabla 5.90), Campos Relacionados con el Problema de Negocio (Tabla 5.91), Reporte de Datos Explorados (Tabla 5.92), Reporte de la Calidad de los Datos (Tabla 5.93) y Criterios de Éxito del Problema de Negocio (Tabla 5.89).

Selección de Algoritmos de Explotación de Información (D.Mo.CoM.SAEI): a partir del proceso de explotación de información definido en la actividad previa, se seleccionan los algoritmos de agrupamiento SOM, HAC y K-Means, y el algoritmo de árboles de decisión C4.5 (perteneciente a la familia TDIDT, de sus siglas en inglés Top Down Induction Decision Trees), siendo ID3 otra alternativa disponible en la herramienta, la cual no será utilizada. Se identifican las restricciones de implementación de cada uno de los algoritmos (tipo de atributos de entrada y target) y se define la estrategia de implementación (primero los algoritmos de clustering y luego se combina el resultado

con el algoritmo C4.5 para cada uno de ellos). La tabla 5.95 ilustra la estructura del formalismo obtenido.

Selección de Algoritmos de explotación de información				
Responsable:	Ezequiel B.	**Fecha:**	28/09/2015	
ID#:	D.Mo.CoM.SAEI	**Versión:**	1.0	
Problema de Negocio	(prne.1) Identificar a partir del progreso académico perfiles de estudiantes y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos			
Algoritmo	**Input**	**Target**	**Estrategia**	**Descripción**
C4.5	Discretos/Continuos	Discretos	2	familia TDIDT
ID3	Discretos/Continuos	Discretos		familia TDIDT
HAC	Continuos	-	1	
SOM	Continuos	-	1	
K-Means	Continuos	-	1	

Tabla 5.95. Caso de Validación: Educación Superior - Selección de Algoritmos de Explotación de Información

Selección de Variables del Modelo (D.Mo.CoM.SeVM): para la conformación del modelo, y a partir del análisis realizado de los datos (exploración y calidad), se utilizan todos los campos identificados para el problema, seleccionando los campos académicos como entrada (Input) para los algoritmos de agrupamiento, y todos las variables para el algoritmo de clasificación (siendo el atributo target o clase, el generado por los algoritmos de agrupamiento). Adicionalmente, se señala la necesidad de normalizar las variables utilizadas para identificar grupos. En la tabla 5.96 se puede observar el resultado obtenido.

Selección de variables del Modelo				
Responsable:	Ezequiel B.	**Fecha:**	29/09/2015	
ID#:	D.Mo.CoM.SeVM	**Versión:**	1.0	
Problema de Negocio	(prne.1) Identificar a partir del progreso académico perfiles de estudiantes y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos			
Campo	**HAC/SOM/k-means**		**C4.5**	
	Tipo	**Conversión**	**Tipo**	**Conversión**
trabaja			Input	
familia			Input	
beca			Input	
Padre Ultimos estudios Grupos			Input	
Madre Ultimos estudios grupos			Input	
Aprobo Inf en Cursada	Input		Input	
CumplePlan	Input	Normalizar	Input	
Demora en cursarla	Input	Normalizar	Input	
Sexo			Input	
RitmoInicial	Input	Normalizar	Input	
Argentino			Input	
Cordoba			Input	
Capital			Input	
edad			Input	

Tabla 5.96. Caso de Validación: Educación Superior - Selección de Variables del Modelo

Estrategias de Evaluación de Modelos (D.Mo.CoM.EsEM): se define el uso del método cross-validation para comprobar la generalización del patrón obtenido y a su vez la independencia del conocimiento entre la fuente de entrenamiento y la de testeo, evitando reducir el set de datos a utilizar para entrenar el modelo. Dicho método se emplea con la configuración: 10 fold Cross-Validation con 10 repeticiones (registrado en la "descripción"). La configuración elegida, se debe a que estudios han demostrado que presenta el balance adecuado entre precisión del resultado y costo de cómputo [Kohavi, R., 1995]. En la tabla 5.97 se presenta la estrategia para evaluar los modelos seleccionados.

Estrategias de evaluación de modelos			
Responsable:	Ezequiel B.	**Fecha:**	29/09/2015
ID#:	D.Mo.CoM.EsEM	**Versión:**	1.0
Problema de Negocio	(prne.1) Identificar a partir del progreso académico perfiles de estudiantes y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos		
Técnica	**Alcance**	**Descripción**	
Cross-Validation	C4.5	10 fold Cross-Validation con 10 repeticiones	

Tabla 5.97. Caso de Validación: Educación Superior - Estrategias de Evaluación de Modelos

5.2.2.4. Fase: Preparación de los Datos (D.PD)

La fase Preparación de los Datos está conformada por 2 actividades: Construcción de la Fuente Temporaria de Datos (sección 5.2.2.4.1), donde se preparan y describen las distintas fuentes de datos a utilizar para la extracción del conocimiento y la selección del modelo, y Adecuación de la Fuente Temporaria de Datos (sección 5.2.2.4.2), en la cual se realizan las tareas de limpieza y formateo de los datos.

5.2.2.4.1. Actividad: Construcción de la Fuente Temporaria de Datos (D.PD.CFT)

En esta actividad se realizan las tareas finales para la generación de las fuentes de datos requeridas para las distintas etapas de implementación del modelo (entrenamiento, validación y testeo). Las fuentes generadas se definen como fuente temporaria de datos, debido a que dicha fuente de almacenamiento es distinta a aquella utilizada en producción y la misma solo será de utilidad para la formación del modelo, la extracción del conocimiento y la evaluación del mismo.

A continuación se presentan los resultados obtenidos de aplicar la técnica Generación de la Fuente Temporaria de Datos (sección 4.4.4.1.2, pág. 246), la cual utiliza como insumos los siguientes formalismos: Reporte de Datos Explorados (Tabla 5.92), Fuente Integrada de datos (Figura 5.8),

Estrategias de evaluación de modelos (Tabla 5.97), Reporte de Evaluación de Herramientas (Tabla 5.65) y Selección de variables del Modelo (Tabla 5.96).

Fuente Temporaria de datos (D.PD.CFT.FuTD): se utiliza una única fuente temporaria de datos (figura 5.9, representada mediante una entidad del formalismo DER), la cual mantiene la misma estructura que la fuente integrada de datos.

Figura 5.9. Caso de Validación: Educación Superior - Fuente Temporaria de Datos (Diagrama Entidad-Relación)

Reporte de Generación de la Fuente Temporaria de datos (D.PD.CFT.RGFT): se describe la fuente temporaria de datos, el cual está conformado por 1525 registros (ajustado a partir de los ocho registros con edad nula eliminados). La tabla 5.98 ilustra el resultado obtenido.

5.2.2.4.2. Actividad: Adecuación de la Fuente Temporaria de Datos (D.PD.AFT)

En esta actividad se analizan las características de los campos seleccionados para los distintos problemas de negocio, con el objetivo de identificar y realizar actividades de conversión y ajuste de los registros, preparando los datos para la adecuada extracción de patrones de conocimiento. A continuación se presentan los resultados obtenidos de aplicar la técnica Adecuación de los Datos (sección 4.4.4.2.2, pág. 254), la cual utiliza como insumos los formalismos: Reporte de la Calidad de los Datos (Tabla 5.93), Selección de variables del Modelo (Tabla 5.96), Fuente Temporaria de Datos (Figura 5.9) y Reporte de Generación de la Fuente Temporaria de Datos (Tabla 5.98).

Reporte de Adecuación de la Fuente Temporaria de Datos (D.PD.AFT.RAFT): a partir del formalismo selección de variables del modelo, se define eliminar aquellos atributos con edad nula (uno negativo), viéndose afectados ocho registros. Adicionalmente, se normalizan por varianza las variables CumplePlan, Demora en cursarla y RitmoInicial, para la implementación de los algoritmos de agrupamiento, viéndose afectados la totalidad de los registros. La tabla 5.99 ilustra el resultado previamente descripto.

Reporte de Generación de la Fuente Temporaria de datos			
Responsable:	Diego J.	**Fecha:**	12/10/2015
ID#:	D.PD.CFT.RGFT	**Versión:**	1.0
Problema de Negocio	(prne.1) Identificar a partir del progreso académico perfiles de estudiantes y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos		

ATRIBUTOS CUALITATIVOS		
Nombre	**Valores**	**Distribución**
trabaja	0 1	1311 (85,97%) 214 (14,03%)
familia	0 1	120 (7,87%) 1405 (92,13%)
beca	0 1	1418 (92,98%) 107 (7,02%)
Aprobo Inf en Cursada	0 1	1045 (68,52%) 480 (31,48%)
CumplePlan	0 1 2 3 4	64 (4,20%) 52 (3,41%) 1 (0,07%) 887 (58,16%) 521 (34,16%)
Demora en cursarla	0 1 2	1199 (78,62%) 90 (5,90%) 236 (15,48%)
Sexo	M F	1123 (73,64%) 402 (26,36%)
RitmoInicial	0 1 2 3	326 (21,38%) 22 (1,44%) 985 (64,59%) 192 (12,59%)
Argentino	0 1	26 (1,70%) 1499 (98,30%)
Cordoba	0 1	625 (40,98%) 900 (59,02%)
Capital	0 1	1080 (70,82%) 445 (29,18%)
Padre Ultimos estudios Grupos	0 1 2 3 4	46 (3,02%) 315 (20,66%) 554 (36,33%) 508 (33,31%) 102 (6,69%)
Madre Ultimos estudios grupos	0 1 2 3 4	14 (0,92%) 244 (16,00%) 455 (29,84%) 704 (46,16%) 108 (7,08%)

ATRIBUTOS CUANTITATIVOS				
Nombre	**Min**	**Max**	**Tendencia Central**	**Dispersión**
Edad	19	64	Media = 21.68	Des. Est. = 3.63

Comentarios:

Tabla 5.98. Caso de Validación: Educación Superior - Reporte de Generación de la Fuente Temporaria de datos

Reporte de Adecuación de la Fuente Temporaria de Datos			
Responsable:	Diego J.	**Fecha:**	19/10/2015
ID#:	D.PD.AFT.RAFT	**Versión:**	1.0
Problema de Negocio:	(prne.1) Identificar a partir del progreso académico perfiles de estudiantes y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos		

Nombre	Acción	Efecto	Descripción
Edad	remover registros con edad faltante (-1)	8 registros eliminados	
CumplePlan	Normalizar valores	1525 registros afectados	Normalización de valores por varianza
Demora en cursarla	Normalizar valores	1525 registros afectados	Normalización de valores por varianza
RitmoInicial	Normalizar valores	1525 registros afectados	Normalización de valores por varianza

Tabla 5.99. Caso de Validación: Educación Superior - Reporte de Adecuación de la Fuente Temporaria de Datos

5.2.2.5. Fase: Implementación (D.Im)

La fase de implementación, está conformada por las actividades: Selección del Modelo (sección 5.2.2.5.1), donde se define la estrategia a utilizar para identificar la mejor configuración del modelo, y explotación de información (sección 5.2.2.5.2), donde se realiza la extracción y descripción de los patrones de conocimiento ocultos en los datos.

5.2.2.5.1. Actividad: Selección del Modelo (D.Im.SeM)

En esta actividad se definen el criterio y la forma mediante la cual se determina, cuál de las posibles combinaciones de algoritmos logran capturar con mayor precisión los patrones ocultos en los datos. A continuación se presentan los resultados obtenidos de aplicar la técnica Selección de la Estrategia de Hiperparametrización (sección 4.4.5.1.2, pág. 260), la cual utiliza como insumos los formalismos: Selección de Algoritmos de explotación de información (Tabla 5.95), el Reporte de Generación de la Fuente Temporaria de Datos (Tabla 5.98), los Criterios de Éxito del Problema de Negocio (Tabla 5.89) y la Reporte de Evaluación de Herramientas (Tabla 5.65).

Reporte de Estrategia de Parametrización del Modelo (D.Im.SeM.REPM): de acuerdo al tipo de problema a resolver (Descubrimiento de reglas de pertenencia a grupos), los algoritmos a utilizar (agrupamiento y clasificación TDIDT) y la herramienta seleccionada, se definen los parámetros y los rangos de valores posibles para su optimización. Para el algoritmo de clasificación, se optimizan los parámetros: Mínima cantidad de hojas por rama (Min size of leaves) seleccionando como rango de valores posibles de 5 a 100 y el nivel de confidencia (Confidence Level) de 0.20 a 0.60. Utiliza como criterio de evaluación la tasa de error y Random Search como estrategia de configuración. Para los algoritmos de agrupamiento, se define como única variable a optimizar la cantidad de

clusters, utilizando la estrategia de configuración Grid Search (con excepción de la implementación de HAC que tiene automatizado dicho aspecto). En la tabla 5.100 se detalla la información previamente mencionada.

Reporte de Estrategia de Parametrización del Modelo						
Responsable:		Ezequiel B.		**Fecha:**	23/10/2015	
ID#:		D.Im.SeM.REPM		**Versión:**	1.0	
Problema de Negocio	(prne.1) Identificar a partir del progreso académico perfiles de estudiantes y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos					
ID	**Algoritmo E.I.**	**Estrategia de Configuración**	**Criterio**	**Rango Inferior**	**Rango Superior**	**Comentarios**
repm.1	C4.5	Random Search	Tasa de Error	Min size of leaves: 5 Confidence Level: 0,20	Min size of leaves: 100 Confidence Level: 0,60	Se buscará el mejor valor de acuerdo al criterio entre el rango definido por cada parámetro
repm.2	SOM	Grid Search	interpretabilidad	Row size: 2 Col size: 1	Row size: 4 Col size: 4	Distance normalization: Variance Learning rate: 0,05 Seed random generator: Standard
repm.3	K-Means	Grid Search	La suma de la diferencia media cuadrada intra grupos (WSS)	Clusters: 2	Clusters: 6	Max Iteration: 10 Trials: 5 Distance normalization: Variance Average computation: McQueen Seed random generator: Standard
repm.4	HAC	Automática	La suma de la diferencia media cuadrada entre grupos (BSS)	Clusters: 2	Clusters: 6	Distance normalization: Variance

Tabla 5.100. Caso de Validación: Educación Superior - Reporte de Estrategia de Parametrización del Modelo

5.2.2.5.2. Actividad: Explotación de Información (D.Im.ExI)

En esta actividad se aplican los algoritmos de explotación de información (o minería de datos), con el objetivo de extraer los patrones de conocimientos ocultos en las fuentes de información, dejando constancia de los resultados obtenidos para poder reproducir y comparar los mismos.

A continuación se presentan los resultados obtenidos de aplicar la técnica Extracción de Conocimiento (sección 4.4.5.2.2, pág. 267), la cual utiliza como insumos los formalismos: Criterios de Éxito del Problema de Negocio (Tabla 5.89), Selección de Algoritmos de explotación de información (Tabla 5.95), Selección de variables del Modelo (Tabla 5.96), Estrategias de evaluación de modelos (Tabla 5.97), Fuente Temporaria de Datos (Figura 5.9), Reporte de Generación de la Fuente Temporaria de Datos (Tabla 5.98) y Reporte de Estrategia de Parametrización del Modelo (Tabla 5.100).

Reporte de Implementación del Modelo (D.Im.ExI.ReIM): a partir de la estrategia de optimización definida en la actividad previa, se obtuvo como mejor configuración el modelo: HAC-C4.5 con la configuración: HAC (Clusters: 5) y C4.5 (Min size of leaves: 5 y Confidence Level: 0,25), generando un árbol con nueve nodos y cinco hojas, cuya tasa de error promedio es del 0.2%.

La tabla 5.101 ilustra el resultado obtenido de la implementación de los procesos de explotación de información.

Reporte de Implementación del Modelo			
Responsable:	Ezequiel B.	**Fecha:**	30/10/2015
ID#:	D.Im.ExI.ReIM	**Versión:**	1.0
Problema de Negocio:	(prne.1) Identificar a partir del progreso académico perfiles de estudiantes y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos		

Algoritmo E.I.	**Estrategia**	**Configuración**	**Descripción**
SOM	(repm.2) Grid Search	Row size: 3 Col size: 2	
K-Means	(repm.3) Grid Search	Clusters: 5	WSS = 12318,14
HAC	(repm.4) Automática	Clusters: 5	% BSS = 0.4077
SOM-C4.5	(repm.1) Random Search	Min size of leaves: 10 Confidence Level: 0,30	Número de nodos: 15 Número de Hojas: 8 Tasa de error (promedio): 0.0275
K-Means-C4.5	(repm.1) Random Search	Min size of leaves: 10 Confidence Level: 0,25	Número de nodos: 25 Número de Hojas: 13 Tasa de error (promedio): 0.0222
HAC-C4.5	(repm.1) Random Search	Min size of leaves: 5 Confidence Level: 0,25	Número de nodos: 9 Número de Hojas: 5 Tasa de error (promedio): 0.0020

Tabla 5.101. Caso de Validación: Educación Superior - Reporte de Implementación del Modelo

Patrones de Conocimiento (D.Im.ExI.PaCo): se detallan los patrones obtenidos por el modelo que obtuvo mejores resultados (HAC-C4.5), presentando las reglas de comportamiento obtenidas para cada uno de los cinco grupos identificados (figura 5.10).

A partir de las reglas obtenidas y del análisis estadístico de los grupos, se derivan las siguientes conclusiones: Los resultados de la aplicación de HAC clasifican al universo de estudiantes en cinco conjuntos. Las principales variables que participan en la caracterización de los grupos son: el origen de sus ingresos, el tiempo de demora en cursar la materia informática y la nacionalidad.

- H1 (102 individuos): Son argentinos, no tienen beca y trabajan. Su ingreso proviene, en su mayoría de su propio trabajo, y comenzaron con un bajo ritmo inicial. El máximo porcentaje es de Córdoba Capital. El nivel de estudios de ambos padres es el más bajo entre los estudiantes de nacionalidad argentina. Evidencian un nivel muy bajo de aprobación de Informática durante la cursada.
- H2 (107 individuos): Conformado por argentinos, beneficiarios de becas. Poseen el máximo porcentaje de aprobados en cursada y de cumplir actualmente el plan de carrera. Los niveles de estudios alcanzados por sus padres son acordes a la media del universo (MdU).

- H3 (1020 individuos): Argentinos que no trabajan y no tienen beca, comenzaron con el más alto ritmo su carrera, sin embargo, el grupo anterior posee mejores porcentajes de aprobación y de mantener el plan de carrera. En su casi totalidad, viven con su familia. Ambos padres tienen el nivel más alto de estudios.
- H4 (26 individuos): Conformado por los estudiantes extranjeros. Presentan el padre con el nivel de estudios más alto y son el segundo grupo con mayor demora en cursar la materia (después de H5). Sin embargo, es relevante destacar el reducido número de representatividad del mismo.
- H5 (270 individuos): Son argentinos, no trabajan ni tienen beca, comenzaron con el más bajo ritmo. Los más demorados en su plan de carrera, ninguno aprobó Informática en cursada y son los que más demoraron en cursarla con respecto a su año de ingreso. Su fuente de ingreso son mayoritariamente los aportes de su familia.

Classifier performances

Error rate	0,0020
Values prediction	Confusion matrix

Value	Recall	1-Precision
c_hac_1	0,9808	0,0000
c_hac_2	1,0000	0,0280
c_hac_3	1,0000	0,0000
c_hac_4	1,0000	0,0000
c_hac_5	0,9963	0,0000

	c_hac_1	c_hac_2	c_hac_3	c_hac_4	c_hac_5	Sum
c_hac_1	102	2	0	0	0	104
c_hac_2	0	104	0	0	0	104
c_hac_3	0	0	1020	0	0	1020
c_hac_4	0	0	0	26	0	26
c_hac_5	0	1	0	0	270	271
Sum	102	107	1020	26	270	1525

Tree description

Number of nodes	9
Number of leaves	5

Decision tree

- Argentino_bin in [SI]
  - beca < 0,5000
    - familia < 0,5000 then Cluster_HAC_1 = c_hac_1 (100,00 % of 102 examples)
    - familia >= 0,5000
      - Demora en cursarla < 0,5000 then Cluster_HAC_1 = c_hac_3 (100,00 % of 1020 examples)
      - Demora en cursarla >= 0,5000 then Cluster_HAC_1 = c_hac_5 (100,00 % of 270 examples)
  - beca >= 0,5000 then Cluster_HAC_1 = c_hac_2 (97,20 % of 107 examples)
- Argentino_bin in [NO] then Cluster_HAC_1 = c_hac_4 (100,00 % of 26 examples)

Figura 5.10. Caso de Validación: Educación Superior - Patrones de Conocimiento

5.2.2.6. Fase: Evaluación y Presentación (D.EP)

La fase Evaluación y Presentación se encuentra conformada por dos actividades: Evaluación de los Resultados (sección 5.2.2.6.1), donde se analiza la validez y utilidad de los patrones hallados, y Presentación de los Resultados (sección 5.2.2.6.2), garantizando la adecuada transmisión del conocimiento extraído para su utilización.

5.2.2.6.1. Actividad: Evaluación de los Resultados (D.EP.EvR)

En esta actividad se evalúa la validez de los patrones de conocimiento obtenidos para el dominio de negocio y en particular para las problemática de negocio en cuestión. A continuación se presentan los resultados obtenidos de aplicar la técnica Validación del Conocimiento (sección 4.4.6.1.2, pág. 275), la cual utiliza como insumos los formalismos: Objetivos del Proyecto (Tabla 5.84), Criterios de Éxito del Proyecto (Tabla 5.85), Problema del Negocio (Tabla 5.88), Criterios de Éxito del Problema de Negocio (Tabla 5.89), Reporte de Implementación del Modelo (Tabla 5.101) y patrones de conocimiento (Figura 5.10).

Reporte de Evaluación de los Resultados (D.EP.EvR.ReER): de acuerdo a los patrones de conocimiento identificados, se evaluó de forma conjunta con el experto del problema de negocio (Laura D.) la validez e interés de los resultados, determinando que los mismos satisficieron las necesidades cubiertas por la pregunta-problema y los criterios de éxito vinculados. En adición, el experto señaló: "*Se identifican cinco perfiles de estudiantes, siendo significativos cuatro de ellos, brindando aportes novedosos y de interés, para confirmar y generar nuevas políticas. Considerándose los aspectos de procedencia de poco interés.*". La tabla 5.102 ilustra el resultado obtenido.

Reporte de Evaluación de los Resultados			
Responsable:	Ezequiel B.	**Fecha:**	06/11/2015
ID#:	D.EP.EvR.ReER	**Versión:**	1.0
Problema de Negocio	**Criterio de Éxito**	**Resultado**	**Descripción**
(prne.1) Identificar a partir del progreso académico perfiles de estudiantes y caracterizarlos de acuerdo a sus aspectos socioeconómicos y geográficos	(cepn.1) Identificar características que permitan comprender el comportamiento en contexto de masividad, los cuales serán evaluados por Laura D. (rehi.4)	Satisfactorio	Se identifican cinco perfiles de estudiantes, siendo significativos cuatro de ellos, brindando aportes novedosos y de interés, para confirmar y generar nuevas políticas. Considerándose los aspectos de procedencia de poco interés.

Tabla 5.102. Caso de Validación: Educación Superior - Reporte de Evaluación de los Resultados

5.2.2.6.2. Actividad: Presentación de los Resultados (D.EP.PrR)

En esta actividad se llevan a cabo las tareas finales del proyecto, con respecto a la documentación de los resultados obtenidos y la presentación de los mismos a los interesados. El objetivo de esta actividad es la correcta transmisión de los patrones obtenidos y el conocimiento extraído para dar soporte al proceso de toma de decisión del cliente.

A continuación se presentan los resultados obtenidos de aplicar la técnica Síntesis del Proyecto (sección 4.4.6.2.2, pág. 280), la cual utiliza como insumos los formalismos: Fuentes de Información del Cliente (Tabla 5.82), Objetivos del Proyecto (Tabla 5.84), Criterios de Éxito del Proyecto (Tabla 5.85), Suposiciones del Proyecto (no aplicable en este proyecto), Restricciones del Proyecto (Tabla 5.87), Problema del Negocio (Tabla 5.88), Patrones de Conocimiento (Figura 5.10), Reporte de Generación de la Fuente Temporaria de Datos (Tabla 5.98), Reporte de la Calidad de los Datos (Tabla 5.93) y Reporte de Evaluación de los Resultados (Tabla 5.102)

Reporte del Proyecto (D.EP.PrR.RepP): en la tabla 5.103 se presenta un resumen del proyecto, describiendo las necesidades y problemáticas identificadas, junto con sus criterios de éxito. Se detallan los datos disponibles, el proceso implementado junto con los resultados obtenidos y su interpretación/evaluación. Finalmente, se presentan recomendaciones sobre futuros aspectos a evaluar.

Reporte del Proyecto			
Responsable:	Ezequiel B.	**Fecha:**	12/11/2015
ID#:	D.EP.PrR.RepP	**Versión:**	1.0
DESCRIPCIÓN DEL PROBLEMA	El propósito de esta investigación es contribuir a facilitar la apropiación del conocimiento en Educación Superior en contextos de masividad. Proveer de información para un adecuado diseño de las políticas públicas en Educación Superior, despierta el interés para contribuir con una mejor apropiación del conocimiento por parte de la sociedad. En este contexto, se requirió analizar estudiar el comportamiento de aquellos estudiantes de la Universidad Nacional de Córdoba centrándose en aquellos estudiantes que cursaron la materia informática en los años 2012 y 2013, con el objetivo de contribuir con conocimientos que favorezcan al diseño de políticas públicas en Educación Superior, facilitando la apropiación del conocimiento en contextos de masividad. A partir de dicha meta, se caracteriza el comportamiento de distintos tipos de estudiantes a sus aspectos académicos, socioeconómicos y geográficos, los cuales permitan comprender el comportamiento de los alumnos en contexto de masividad, proveyendo de contribuciones novedosas, tanto en las interpretaciones sobre la información del estudiante que emerjan, así como para la construcción de un diseño metodológico aplicable en las muy diversas prácticas de este escenario.		
DESCRIPCIÓN DE LOS DATOS	El análisis inicial fue realizado sobre 1533 de estudiantes que cursaron la materia informática en los años 2012 y 2013. Debido a anomalías en los datos (fecha de nacimiento), se redujo la muestra a 1525. Las variables consideradas para el análisis son: ▪ Origen de los ingresos del estudiante (Beca, Familia y Trabajo Propio) ▪ Género. ▪ Variables geográficas de procedencia (si es Argentino, de la provincia de Córdoba y de su capital) ▪ Edad, ▪ Nivel de instrucción de los padres ▪ Variables académicas: - Aprueba la materia informática durante el mismo año que la cursa, - Tiempo que demora en cursar la materia informática, - Gradualidad de cumplimiento del plan, y - Ritmo de materias que cursa durante el primer semestre. Se decide no considerar las calificaciones de los estudiantes (promedio de calificaciones con y sin aplazos) debido al sesgo proveniente de las subjetividades de los evaluadores al generar esas calificaciones, y de las diversas normativas vigentes en las distintas unidades académicas		

Tabla 5.103.a. Caso de Validación: Educación Superior - Reporte del Proyecto

RESULTADOS DE EXPLOTACIÓN DE INFORMACIÓN	A partir de las reglas obtenidas y del análisis estadístico de los grupos, se derivan las siguientes conclusiones: Los resultados de la aplicación de los algoritmos HAC y C4.5 clasifican al universo de estudiantes en cinco conjuntos. Las principales variables que participan en la caracterización de los grupos son: el origen de sus ingresos, el tiempo de demora en cursar la materia informática y la nacionalidad. • H1 (102 individuos): Son argentinos, no tienen beca y trabajan. Su ingreso proviene, en su mayoría de su propio trabajo, y comenzaron con un bajo ritmo inicial. El máximo porcentaje es de Córdoba Capital. El nivel de estudios de ambos padres es el más bajo entre los estudiantes de nacionalidad argentina. Evidencian un nivel muy bajo de aprobación de Informática durante la cursada. • H2 (107 individuos): Conformado por argentinos, beneficiarios de becas. Poseen el máximo porcentaje de aprobados en cursada y de cumplir actualmente el plan de carrera. Los niveles de estudios alcanzados por sus padres son acordes a la media del universo (MdU). • H3 (1020 individuos): Argentinos, que no trabajan y no tienen beca, comenzaron con el más alto ritmo su carrera, sin embargo el grupo anterior mejores porcentajes de aprobación y de mantener el plan de carrera. En su casi totalidad, viven con su familia. Ambos padres tienen el nivel más alto de estudios. • H4 (26 individuos): Conformado por los estudiantes extranjeros. Presentan el padre con el nivel de estudios más alto y son el segundo grupo con mayor demora en cursar la materia (después de H5). Sin embargo, es relevante destacar el reducido número de representatividad del mismo. • H5 (270 individuos): Son argentinos, no trabajan ni tienen beca, comenzaron con el más bajo ritmo. Los más demorados en su plan de carrera, ninguno aprobó Informática en cursada y son los que más demoraron en cursarla con respecto a su año de ingreso. Su fuente de ingreso son mayoritariamente los aportes de su familia.
EVALUACIÓN DE LOS RESULTADOS	Como conclusiones del análisis realizado, se destaca al instrumento de becas como una medida favorable para el desempeño académico, mientras que aquellos estudiantes que trabajan (en líneas generales) presentan un menor progreso. De aquellos estudiantes cuyo aporte principal proviene de la familia, se pueden identificar dos grandes grupos, observándose que el rendimiento inicial del mismo tiene fuerte relación con su cumplimiento del plan de estudios. De los cinco perfiles de estudiantes hallados, se identifican como significativos, novedosos y de interés para confirmar y generar nuevas políticas cuatro de ellos. Considerándose los aspectos de procedencia poco relevantes.
DIFICULTA-DES Y RECOMEN-DACIONES	Se recomienda ampliar el estudio realizado, mediante el registro de variables de residencia que permitan evaluar de manera geo referenciada la distancia y/o el tiempo de transporte requerido hasta la universidad.

Tabla 5.103.b. Caso de Validación: Educación Superior - Reporte del Proyecto

5.3. ANÁLISIS COMPARATIVO DE LA PROPUESTA

En esta sección se evalúa la propuesta (MoProPEI) con respecto a los abordajes metodológicos seleccionados (sección 3.2.1, pág. 57), haciendo uso del marco comparativo de metodologías para proyectos de explotación de información (o en palabras del autor: minería de datos) definido en [Moine, 2013]. Dicho marco, está conformado por cuatro aspectos (figura 5.11 [Moine, 2013]): *Nivel de detalle en las actividades de cada fase*, se evalúa si las propuestas guían al usuario en su desarrollo; *escenarios de aplicación*, la capacidad de adaptarse a distintas necesidades del proyecto; *actividades específicas que componen cada fase*, evaluando el cubrimiento de tareas requeridas en todo proyecto de explotación de información; y *actividades destinadas a la dirección de proyectos*, analizando la incorporación de aquellas actividades necesarias para la gestión de un proyecto. Para cada uno de ellos, se consideran una serie de características, las cuales permiten cuantificar el cubrimiento parcial de la propuesta en dichos aspectos. El marco cubre un total de 52 características a evaluar, las cuales se responden de manera afirmativa o negativa, determinándose como superior a la propuesta que totalice la mayor cantidad de afirmaciones.

En las tablas 5.104 - 5.107, se presentan las valuaciones realizadas para cada característica correspondiente a los cuatro aspectos previamente descriptos, considerándose el valor “SI” a aquellas propuestas que especifican técnicas o lineamientos para las actividades requeridas. La tabla 5.108, presenta las valoraciones obtenidas por cada aspecto y totales (la sumatoria de todos los aspectos). La figura 5.12 facilita la visualización de los resultados obtenidos.

A partir de los resultados previamente mencionados, se observa que el modelo de proceso propuesto en esta tesis, satisface un 87% de las características, siendo un 16% superior a la segunda metodología evaluada. Si bien la propuesta se encuentra con el mayor cubrimiento en 3 de los 4 aspectos (uno de ellos compartido), la principal diferencia se encuentra en los aspectos orientados a las actividades para la dirección del proyecto en el cual casi duplica el cubrimiento de las propuestas predecesoras (totalizando 15 puntos con respecto a los 8 que obtienen las dos siguientes propuestas CRISP-DM y Catalyst).

Figura 5.11. Marco Comparativo metodología de explotación de información.

Evaluación del nivel de detalle en las actividades de cada fase										
CARACTERÍSTICAS	**CRISP-DM**	**Catalyst**	**KDD**	**SEMMA**	**IKDDM**	**MPIMD**	**ASD-BI**	**FMDS**	**TDSP**	**MoProPEI**
¿Se definen actividades específicas para cada fase del proceso?	SI	SI	NO	NO	SI	SI	SI	NO	SI	SI
¿Se explicitan los pasos a seguir para llevar a cabo cada actividad?	SI	SI	NO	NO	SI	NO	NO	NO	SI	SI
¿Se definen las entradas de cada actividad?	NO	NO	NO	NO	SI	NO	NO	NO	NO	SI
¿Se definen las salidas de cada actividad?	SI	SI	SI	SI	SI	NO	NO	NO	SI	SI
¿Se provee una guía de buenas prácticas para cada una de las actividades específicas?	SI	SI	NO	NO	SI	NO	NO	NO	SI	SI
Detalle en las actividades - Sub total (5)	**4**	**4**	**1**	**1**	**5**	**1**	**1**	**0**	**4**	**5**

Tabla 5.104. Marco Comparativo – Nivel de detalle en la descripción de las actividades

Evaluación de los escenarios de aplicación										
CARACTERÍSTICAS	**CRISP-DM**	**Catalyst**	**KDD**	**SEMMA**	**IKDDM**	**MPIMD**	**ASD-BI**	**FMDS**	**TDSP**	**MoProPEI**
¿Se especifican actividades para la definición y el análisis del problema u oportunidad con el cual colaborará la minería de datos?	SI	SI	SI	NO	SI	SI	SI	SI	SI	SI
¿Se consideran puntos de partida alternativos donde el usuario no refiere un problema sino que sólo desea explorar sus datos?	NO	SI	NO	NO	NO	NO	NO	NO	NO	NO
¿La metodología es independiente del dominio de aplicación?	SI	SI	SI	SI	SI	SI	SI	SI	SI	SI
¿La metodología es aplicable a proyectos de diferente tamaño?	SI	SI	SI	SI	SI	SI	SI	SI	SI	SI
Escenarios de aplicación - Sub total (4)	**3**	**4**	**3**	**2**	**3**	**3**	**3**	**3**	**3**	**3**

Tabla 5.105. Marco Comparativo – Escenarios de aplicación

Evaluación de las actividades específicas en cada fase										
CARACTERÍSTICAS	**CRISP-DM**	**Catalyst**	**KDD**	**SEMMA**	**IKDDM**	**MPIMD**	**ASD-BI**	**FMDS**	**TDSP**	**MoProPEI**
Análisis del problema										
¿Se propone una evaluación general de la organización?	SI	SI	NO	NO	SI	NO	SI	NO	NO	SI
¿Se identifica al personal involucrado en el proyecto?	SI	SI	NO	NO	SI	NO	SI	NO	NO	SI
¿Se define el problema u oportunidad de negocio?	SI	SI	NO	NO	SI	SI	SI	SI	SI	SI
¿Se propone una evaluación de las fuentes de datos?	NO	SI	SI	SI	SI	NO	SI	SI	SI	SI
¿Se analizan todas las soluciones posibles al problema?	NO	SI	NO	NO	NO	NO	NO	NO	NO	NO
¿Se especifican los objetivos del proyecto?	SI	SI	NO	NO	SI	SI	SI	SI	SI	SI
¿Se define un criterio de éxito para el proyecto?	SI	NO	NO	NO	SI	SI	SI	SI	SI	SI
¿Se realiza una evaluación general de las técnicas de minería que podrían utilizarse?	SI	NO	SI	SI	SI	NO	SI	SI	SI	SI
¿Se especifica de qué forma el usuario utilizará el nuevo conocimiento?	NO	SI	NO	NO	NO	NO	NO	SI	NO	NO
Análisis del problema - Sub total (9)	**6**	**7**	**2**	**2**	**7**	**3**	**7**	**6**	**5**	**7**
Selección y preparación de los datos										
¿Se propone un análisis exploratorio inicial de los datos?	SI	SI	SI	SI	SI	SI	SI	SI	SI	SI
¿Se sugieren actividades para la limpieza de los datos?	SI	SI	SI	SI	SI	NO	NO	NO	SI	SI
¿Se contemplan actividades para la transformación de variables y la creación de atributos derivados?	SI	SI	SI	SI	Si	NO	SI	SI	SI	SI
¿Se realiza un análisis descriptivo final sobre los datos depurados?	NO	NO	NO	NO	NO	NO	NO	NO	NO	SI
¿Se verifica con el usuario la completitud del conjunto de datos final?	NO	SI	NO	NO	NO	NO	NO	NO	NO	SI
Selección y preparación de los datos - Sub total (5)	**3**	**4**	**3**	**3**	**3**	**1**	**2**	**2**	**3**	**5**
Modelado										
¿Se efectúa una selección de las técnicas que se utilizarán?	SI	SI	SI	SI	SI	SI	SI	SI	SI	SI
¿Se planifica la forma en la que se evaluarán los resultados?	SI	NO	NO	NO	SI	SI	NO	SI	SI	SI
¿Se efectúa una evaluación inicial de los modelos obtenidos?	SI	SI	SI	SI	SI	NO	SI	SI	SI	SI
¿Se proveen directivas para el caso donde se dificulta el descubrimiento de patrones?	NO	SI	NO	NO	NO	NO	NO	NO	NO	NO
Modelado - Sub total (4)	**3**	**3**	**2**	**2**	**3**	**2**	**2**	**3**	**3**	**3**
Evaluación										
¿Se interpretan los modelos en función de los objetivos organizacionales?	SI	SI	SI	SI	SI	SI	SI	SI	SI	SI
¿Se comparan y ponderan los modelos obtenidos?	SI	SI	SI	SI	SI	SI	SI	SI	SI	SI
¿Se propone una revisión general del proceso?	SI	SI	NO	NO	SI	SI	SI	SI	NO	SI
¿Se proveen directivas para el caso donde ninguno de los modelos obtenidos resulta viable?	SI	SI	SI	SI	SI	NO	NO	SI	SI	SI
Evaluación - Sub total (4)	**4**	**4**	**3**	**3**	**4**	**3**	**3**	**4**	**3**	**4**

Tabla 5.106.a. Marco Comparativo – Actividades específicas de cada fase

CARACTERÍSTICAS	CRISP-DM	Catalyst	KDD	SEMMA	IKDDM	MPIMD	ASD-BI	FMDS	TDSP	MoProPEI
Implementación										
¿Se planifica la implementación del nuevo conocimiento?	SI	SI	SI	NO	NO	SI	SI	SI	SI	SI
¿Se propone la creación de un programa de mantenimiento?	SI	SI	NO	NO	SI	NO	NO	NO	NO	NO
¿Se entrega al usuario un resumen del proyecto?	SI	SI	SI	NO	SI	NO	SI	NO	SI	SI
¿Se documenta la experiencia adquirida por el equipo de trabajo?	SI	NO	NO	NO	SI	SI	SI	NO	NO	SI
Implementación - Sub total (4)	**4**	**3**	**2**	**0**	**3**	**2**	**3**	**1**	**2**	**3**
Evaluación de las actividades específicas en cada fase - Sub Total (26)	**20**	**21**	**12**	**10**	**20**	**11**	**17**	**16**	**16**	**22**

Tabla 5.106.b. Marco Comparativo – Actividades específicas de cada fase

Evaluación de las actividades para la dirección del proyecto										
CARACTERÍSTICAS	**CRISP-DM**	**Catalyst**	**KDD**	**SEMMA**	**IKDDM**	**MPIMD**	**ASD-BI**	**FMDS**	**TDSP**	**MoProPEI**
Gestión del alcance										
¿Se propone la selección de los entregables que se generarán durante el proyecto?	SI	SI	NO	NO	NO	NO	NO	NO	NO	SI
¿Se especifican actividades de control del alcance?	NO	NO	NO	NO	NO	SI	NO	NO	NO	SI
Gestión del alcance - Sub total (2)	**1**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**2**
Gestión del tiempo										
¿Se realiza una definición y secuenciación de las actividades que se ejecutarán durante el proyecto?	SI	SI	NO	NO	NO	SI	NO	NO	SI	SI
¿Se realiza una estimación de la duración de cada actividad?	SI	SI	NO	NO	NO	SI	NO	NO	NO	SI
¿Se construye un cronograma para el proyecto?	SI	SI	NO	NO	NO	NO	NO	NO	SI	SI
¿Existen actividades de control del cronograma?	NO	NO	NO	NO	NO	NO	NO	NO	NO	SI
Gestión del tiempo - Sub total (4)	**3**	**3**	**0**	**0**	**0**	**2**	**0**	**0**	**2**	**4**
Gestión del costo										
¿Se efectúa una estimación de los recursos afectados por cada actividad?	SI	SI	NO	NO	NO	SI	NO	NO	NO	SI
¿Se realiza una estimación de los costos del proyecto?	NO	NO	NO	NO	NO	NO	NO	NO	NO	SI
¿Se construye un presupuesto de costos?	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
¿Existen actividades de control del presupuesto a medida que avanza el proyecto?	NO	NO	NO	NO	NO	NO	NO	NO	NO	SI
Gestión del costo - Sub total (4)	**1**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**3**

Tabla 5.107.a. Marco Comparativo – Actividades de dirección del proyecto

CARACTERÍSTICAS	CRISP-DM	Catalyst	KDD	SEMMA	IKDDM	MPIMD	ASD-BI	FMDS	TDSP	MoProPEI
Gestión del equipo de trabajo										
¿Se efectúa una planificación de los recursos humanos?	SI	SI	NO	NO	NO	NO	NO	NO	SI	SI
¿Se proponen actividades para motivar la interacción entre los miembros del Equipo?	NO	NO	NO	NO	NO	NO	NO	NO	NO	SI
¿Se efectúa un seguimiento del rendimiento de los recursos humanos?	NO	NO	NO	NO	NO	NO	NO	NO	NO	SI
Gestión del equipo de trabajo - Sub total (3)	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**3**
Gestión del riesgo										
¿Se efectúa una identificación de los riesgos del proyecto?	SI	SI	NO	NO	NO	SI	NO	NO	NO	SI
¿Se realiza una cuantificación de los riesgos?	NO	SI	NO	NO	NO	NO	NO	NO	NO	NO
¿Se planifican acciones de respuesta ante cada riesgo?	SI	NO	NO	NO	NO	SI	NO	NO	NO	SI
¿Existen actividades de supervisión y control de los riesgos?	NO	NO	NO	NO	NO	NO	NO	NO	NO	SI
Gestión del riesgo - Sub total (4)	**2**	**2**	**0**	**0**	**0**	**2**	**0**	**0**	**0**	**3**
Evaluación de las actividades para la dirección del proyecto - Sub Total (17)	**8**	**8**	**0**	**0**	**0**	**6**	**0**	**0**	**3**	**15**

Tabla 5.107.b. Marco Comparativo – Actividades de dirección del proyecto

ASPECTOS	CRISP-DM	Catalyst	KDD	SEMMA	IKDDM	MPIMD	ASD-BI	FMDS	TDSP	MoProPEI
Nivel de detalle en las actividades de cada fase (5)	4	4	1	1	5	1	1	0	4	5
Escenarios de aplicación (4)	3	4	3	2	3	3	3	3	3	3
Actividades específicas en cada fase (26)	20	21	12	10	20	11	17	16	16	22
Actividades para la dirección del proyecto (17)	8	8	0	0	0	6	0	0	3	15
TOTAL (52)	**35**	**37**	**16**	**13**	**28**	**21**	**21**	**19**	**26**	**45**

Tabla 5.108. Marco Comparativo – Evaluación general

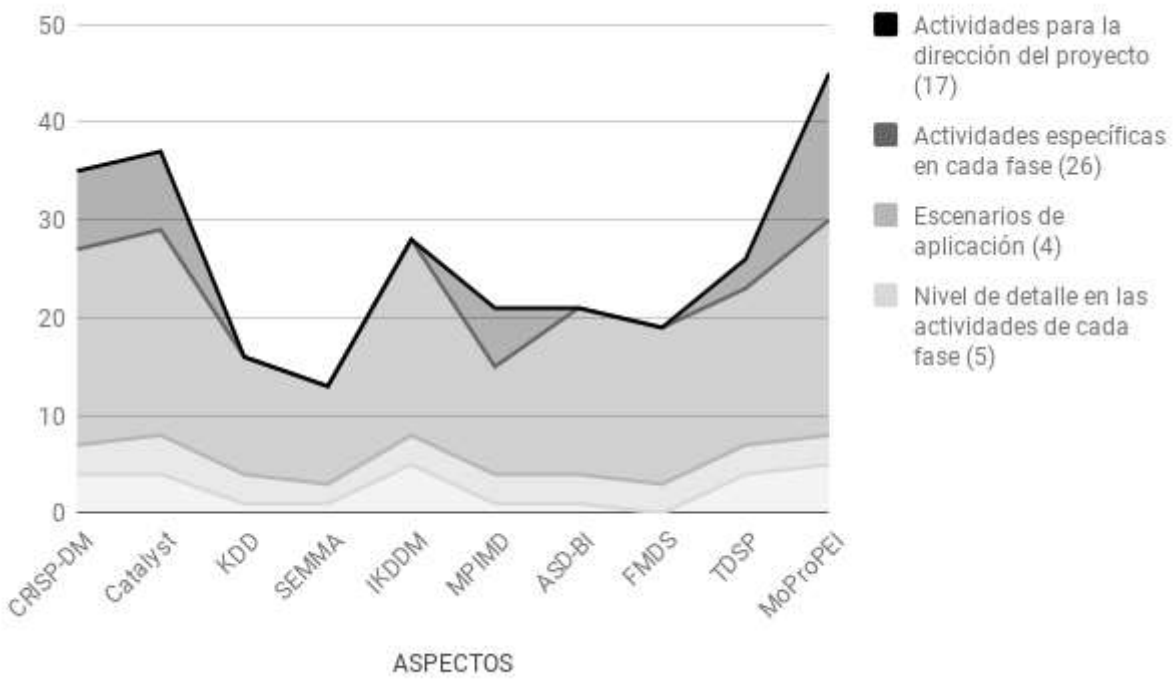


Figura 5.12. Marco Comparativo – Evaluación general (por características acumulativo)

5.4. ANÁLISIS EXPERIMENTAL

Del análisis sistemático de la literatura (Anexo B, pág. 471) se identificaron 3 trabajos previos en los cuales se realizan experimentos para evaluar el rendimiento de las propuestas de modelos de procesos / metodologías existentes:

- Sharma, S. (2008). An integrated knowledge discovery and data mining process model (Doctoral dissertation).
- Sharma, S., Osei-Bryson, K. M., & Kasper, G. M. (2012). Evaluation of an integrated Knowledge Discovery and Data Mining process model. Expert Systems with Applications, 39(13), 11335-11348.
- Saltz, J., & Crowston, K. (2017). Comparing data science project management methodologies via a controlled experiment. In Proceedings of the 50th Hawaii International Conference on System Sciences.

Los primeros dos trabajos representan el mismo proceso experimental en el cual se evalúa a la metodología IKDDM (propuesta por los autores del experimento) con el modelo de proceso más utilizado (CRISP-DM). En dicho trabajo se realizan encuestas a una muestra de 42 personas, aleatoriamente asignadas en 2 grupos (cada uno correspondiente a una propuesta). Por cada

participante, se recolectó el género, role, años de experiencia en minería de datos, y el tiempo de inicio y fin del test.

En primer lugar se le asignó a cada persona el documento resumido de la propuesta correspondiente al grupo al cual fue asignado. Luego un examen teórico sobre contenidos de minería de datos fue realizado. Por último, cada miembro del experimento realizó una encuesta de autopercepción sobre la propuesta asignada. Dicho test está orientado a medir la facilidad, utilidad, calidad y grado de satisfacción del usuario de acuerdo al modelo definido en [Maes y Poels, 2006]. La tabla 5.109 muestra las preguntas realizadas (traducidas al castellano). Se utilizó una escala de likert de 7 ítems para valorar cada criterio.

PEOU1	Me resultó simple entender lo que el modelo KDDM intentaba modelar.	**PSQ2**	El modelo KDDM es una representación realista de este tipo de procesos.
PEOU2	El uso del modelo KDDM fue en ocasiones frustrante.	**PSQ3**	El modelo KDDM contiene elementos contradictorios.
PEOU3	En general, el modelo KDDM fue fácil de utilizar.	**PSQ4**	Todos los elementos en el modelo KDDM son relevantes para la representación de este tipo de procesos.
PEOU4	Entender cómo leer el modelo KDDM me resultó fácil.	**PSQ5**	El modelo KDDM provee una completa representación de este tipo de procesos.
PU1	En general, creo que el modelo KDDM es una mejora a la descripción textual de las tareas involucradas en este tipo de procesos.	**US1**	El modelo KDDM satisfizo adecuadamente las necesidades de información que se me pidió que diera soporte.
PU2	En general, el modelo KDDM fue útil para entender el proceso modelado.	**US2**	El modelo KDDM no me proveyó de manera eficiente la información que necesitaba.
PU3	En general, creo que el modelo KDDM mejoró mi rendimiento.	**US3**	El modelo KDDM me proveyó de manera eficaz la información que necesitaba.
PSQ1	El modelo KDDM representa correctamente las tareas a realizar en este tipo de proyectos.	**US4**	En general, estoy satisfecho con la información que me proveyó el modelo KDDM.

Facilidad de uso percibida (PEOU)
Utilidad percibida (PU)

Calidad semántica (PSQ)
Satisfacción del usuario (US)

Tabla 5.109. Listado de sentencias aplicadas en el experimento [Sharma, 2008].

La validez del procedimiento fue evaluada a partir de un estudio piloto. La similitud de los grupos aleatorios generados fue evaluada a partir de la comparación de medias mediante la prueba T para muestras independientes.

Para validar los resultados del experimento, se utilizó el test de Mann-Whitney, obteniendo como conclusión: "El resultados del test Mann-Whitney con respecto a la calidad global de los modelos de procesos indica que existe un nivel significante de diferencia entre CRISP-DM y IKDDM. Los resultados del test indican claramente que el modelo IKDDM supera al modelo CRISP-DM por un margen ampliamente significante ($p<0.001$) … Los resultados del test de Mann-Whitney para cada una de las cuatro variables de percepción indican que el grupo IKDDM y el CRISP-DM difieren significativamente en la percepción de facilidad de uso, utilidad, calidad semántica y nivel de satisfacción del usuario de los modelos aplicados para la ejecución de las tareas de minería de datos." [Sharma, 2008].

El artículo realizado por Saltz et al. [2017] investiga el impacto del uso de distintos modelos de procesos para el desarrollo de proyectos de explotación de información. Un total de 85 estudiantes graduados participaron (donde más del 75% tenía experiencia previa en tecnologías de la información). Los estudiantes trabajaron en grupos durante un curso semestral en el desarrollo de un proyecto, utilizando uno de los siguientes cuatro modelo de procesos: CRISP-DM, Ágil Scrum, Ágil Kanban y sin metodología (línea base).

Los resultados del experimento fueron obtenidos a partir de las siguientes dos formas:

- La evaluación de 2 expertos con un rango de valores entre 1 y 10 (tabla 5.110).
- Un cuestionario con respuestas en escala de likert de 5 niveles (tabla 5.111).

La prueba utilizada para medir la significancia de los resultados es análisis de varianza (ANOVA).

Sección	Resultado promedio
Ágil Scrum	6.5
Ágil Kanban	7.8
CRISP-DM	8.4
Línea Base (sin modelo de proceso)	7

Tabla 5.110. Resultados del experimento [Saltz et al., 2017] - evaluación de los expertos

Como resultado de la pregunta de investigación: ¿Es un modelo de proceso mejor que los otros?, el equipo llegó a la siguientes conclusiones: "En nuestro experimento, hubo dos metodologías superiores a las otras (Agile Kanban y CRISP-DM), siendo Kanban levemente superior, debido a su mayor enfoque en el trabajo en equipo. Quizá de una forma sorprendente, la metodología Agile Scrum fue peor que la línea base".

Sentencia	Sección	Valor Promedio (Escala de Likert de 5 valores)
De ser posible, desearía trabajar con el mismo equipo en futuro proyectos	Ágil Scrum	3.4
	Ágil Kanban	4.2
	CRISP-DM	4.3
	Línea Base	3.8
Estoy muy satisfecho (con respecto a trabajar en el proyecto)	Ágil Scrum	3.9
	Ágil Kanban	4.3
	CRISP-DM	4.4
	Línea Base	4.4
El método de gestión del proyecto utilizado fue similar al utilizado en proyectos previos	Ágil Scrum	3.3
	Ágil Kanban	2.5
	CRISP-DM	3.3
	Línea Base	3.6
Me resultó complicado utilizar el método de gestión de proyectos con mi equipo	Ágil Scrum	2.6
	Ágil Kanban	3.1
	CRISP-DM	3
	Línea Base	2.6

Tabla 5.111. Resultados Experimento [Saltz et al., 2017] - Percepción de los estudiantes

Del análisis de los experimentos previamente descritos, se observa que el procedimiento más reciente realizado en [Saltz et al., 2017] no provee una descripción adecuada de los pasos la cual permita replicar, ni ampliar el experimento. No se dispone del cuestionario realizado, ni se detallan los criterios y herramientas de valoración de los resultados. En adición, los elementos comparados son disímiles. El modelo de proceso CRISP-DM es una guía diseñada para el desarrollo de proyectos de explotación de información principalmente enfocada en el proceso orientado al producto, mientras que los métodos ágiles Scrum y Kanban definen una estrategia de desarrollo de productos software tradicionales, sin indicarse procedimientos específicos para llevar a cabo proyectos de explotación de información.

Por otro lado, en el experimento realizado en [Sharma, 2008] se describen y presentan en detalles los pasos y técnicas utilizadas para la obtención de los resultados, siendo posible su replicación. En

adición, se analiza el modelo de proceso CRISP-DM (la propuesta más utilizada) con la metodología propuesta por el autor (IKDDM), derivándose como superadora la segunda.

5.4.1. Diseño del experimento

A partir de lo expuesto en la sección anterior, se determina replicar el diseño del experimento realizado en [Sharma, 2008] con el objetivo de comparar la metodología identificada como superadora (IKDDM) con la metodología propuesta en este trabajo (MoProPEI).

El experimento está conformado por las siguientes etapas:

1. Definición de la hipótesis,
2. Relevamiento de los datos,
3. Generación de las variables,
4. Determinación del test estadístico,
5. Implementación del test, y
6. Análisis de los resultados.

Se utilizará la herramienta IBM SPSS para la implementación del experimento.

5.4.1.1. Definición de la hipótesis

A partir de los objetivos del trabajo y las preguntas de investigación definidas en el capítulo 3, se deriva como objetivo de este trabajo de investigación la definición de un modelo de proceso para proyectos de ingeniería de explotación de información integral, el cual facilite la implementación efectiva y eficiente de este tipo de proyectos con respecto a las propuestas existentes.

Acorde a lo expresado en [Hevner et al., 2004], y lo realizado en [Sharma, 2008], lo expuesto en el párrafo anterior implica la creación de un artefacto que resuelve un problema previamente resuelto pero de manera más efectiva y eficiente.

La evaluación del artefacto se realiza a partir de cuatro aspectos utilizados en el experimento realizado en [Sharma, 2008]: facilidad de uso, utilidad percibida, calidad semántica y satisfacción del usuario.

En síntesis, se plantea como hipótesis alternativa:

> h1 = La metodología propuesta mejora la eficiencia de los usuarios para el desarrollo de proyectos.

Siendo la hipótesis nula:

h0 = La percepción de eficiencia no depende de la metodología.

5.4.1.2. Relevamiento de los datos

El relevamiento de los datos se realiza acorde a lo definido en [Sharma, 2008]. Se utilizan la misma cantidad de casos muestrales (42) divididos en 2 grupos (asociados a las metodologías en evaluación) de manera aleatoria simple. La encuesta realizada por cada participante es idéntica que la utilizada en la propuesta base, respondida a partir del estudio de la metodología asignada y su implementación en un proyecto de explotación de información. En dicha encuesta, se recaudaron los siguientes datos del encuestado y del proceso:

- Años de experiencia en la disciplina.
- Años de experiencia en la industria trabajando en la disciplina.
- El tiempo requerido para terminar la encuesta (calculado a partir de la hora de inicio y fin de la misma).

Se decidió, a diferencia que en el experimento base, no registrar el género del encuestado, debido a que se entiende que dicha variable no presenta razón alguna para suponer su interés en el estudio. En adición, y en base a lo expuesto en [Saltz et al., 2017] las personas encuestadas pertenecen al área de las ciencias de la computación, teniendo todos al menos un título en dicha disciplina.

Además, los participantes respondieron 16 ítems asociadas a 4 categorías (facilidad de uso percibida, utilidad percibida, calidad semántica y satisfacción del usuario) con 7 posibles opciones en escala de Likert (desde totalmente de acuerdo hasta totalmente en desacuerdo).

A continuación se listan las 16 declaraciones realizadas agrupadas por categoría. La sigla KDDM se utiliza para hacer referencia a cualquiera de las metodologías según corresponda. Entre paréntesis se muestra el identificador de la pregunta.

Facilidad de uso percibida: mide el grado en el cual una persona considera que el modelo conceptual utilizado para resolver un problema no requerirá un esfuerzo mental extra [Maes y Poels, 2006].

- Me resultó simple entender lo que el modelo KDDM intentaba modelar. (p1)
- El uso del modelo KDDM fue en ocasiones frustrante. (p3)
- En general, el modelo KDDM fue fácil de utilizar. (p5)
- Entender cómo leer el modelo KDDM me resultó fácil. (p7)

Utilidad percibida: mide el grado en el cual una persona considera que utilizar un modelo en particular a mejorado su performance [Maes y Poels, 2006].

- En general, creo que el modelo KDDM es una mejora a la descripción textual de las tareas involucradas en este tipo de procesos. (p2)
- En general, el modelo KDDM fue útil para entender el proceso modelado. (p4)
- En general, creo que el modelo KDDM mejoró mi rendimiento. (p6)

Calidad semántica: mide el grado de correspondencia entre la información provista por el modelo y el dominio del proyecto en el cual se aplica [Poels, et al., 2005; Maes y Poels, 2006].

- El modelo KDDM representa correctamente las tareas a realizar en este tipo de proyectos. (p8)
- El modelo KDDM es una representación realista de este tipo de procesos. (p10)
- El modelo KDDM contiene elementos contradictorios. (p12)
- Todos los elementos en el modelo KDDM son relevantes para la representación de este tipo de procesos. (p14)
- El modelo KDDM provee una completa representación de este tipo de procesos. (p16)

Satisfacción del usuario: mide cuán satisfecho se encuentra el usuario con el modelo de proceso con respecto a su propósito [Maes y Poels, 2006].

- El modelo KDDM satisfizo adecuadamente las necesidades de información que se me pidió que diera soporte. (p9)
- El modelo KDDM no me proveyó de manera eficiente la información que necesitaba. (p11)
- El modelo KDDM me proveyó de manera eficaz la información que necesitaba. (p13)
- En general, estoy satisfecho con la información que me proveyó el modelo KDDM. (p15)

Cabe aclarar que cada formulario permite identificar unívocamente la metodología asociada al encuestado. En la sección C.1 (pág. 483) del Anexo C, se ilustra un modelo de las encuestas realizadas. Los datos recolectados se detallan en la tabla C.1 (sección C.2, pág. 488 del Anexo C).

5.4.1.3. Generación de las variables

A partir de los datos recolectados mediante los cuestionarios, se generaron 5 nuevas variables. Una variable por cada categoría: facilidad de uso percibida (PEOU), utilidad percibida (PU), satisfacción del usuario (US), and calidad semántica percibida (PSQ) y una variable totalizadora denominada “surveyscore”.

El proceso para calcular los valores totalizadores por categorías consiste en:

1. Sumar el valor numérico asociado con cada valor de la escala de likert: 1 = totalmente en desacuerdo, 2 = en desacuerdo, 3 = levemente en desacuerdo, 4 = indefinido, 5 = levemente de acuerdo, 6 = de acuerdo, 7 = totalmente de acuerdo.
2. Para aquellas sentencias negativas (p3, p12 y p13), invertir la escala de valores numéricos, es decir, el valor 7 se asocia con el valor totalmente en desacuerdo y el 1 con el valor totalmente de acuerdo. Esta acción fue realizada para dar el impacto correcto de la percepción del encuestado en la totalización de la categoría.

La variable "surveyscore" se obtiene a partir de la sumatoria de los totales de cada categoría.
La tabla C.2 (pág. 489) en el Anexo C resume los resultados obtenidos de aplicar los procesos de generación de datos previamente descritos a la información cruda obtenida.

5.4.1.4. Determinación del test estadístico

La definición del test a utilizar depende del objetivo del experimento, las muestras y los tipos de datos disponibles. A partir del objetivo, se identifica que las variables de interés son 2: los grupos (variable discreta) y el resultado acumulado de la percepción de los encuestados, calculado en la variable "surveyscore" (variable ordinal). Del análisis de las muestras a utilizar, se observa que las mismas son independientes (es decir, no se encuentran emparejadas) y que se posee una muestra relativamente pequeña. Por último, a partir de los datos no puede asumirse que su distribución en la población es normal. Además, a partir del tipo de dato de la variable en estudio, se entiende que la mediana es la medida más representativa de la tendencia central.

En base al análisis previamente expuesto, y en concordancia con lo realizado en [Sharma, 2008], se determina utilizar el test no paramétrico de Mann-Whitney para muestras independientes. El nivel de significancia utilizado es del 5%.

5.4.1.5. Implementación de los test

Para la aplicación del test no paramétrico, en primer lugar se debe verificar que se cumplan las suposiciones que el mismo posee sobre los datos (sección 5.4.1.5.1). Luego, se detallan los resultados obtenidos de implementar el test estadístico seleccionado para evaluar la hipótesis de trabajo (sección 5.4.1.5.2). Por último, se aplica el test no paramétrico de Mann-Whitney con el objetivo de evaluar las diferencias entre los grupos con respecto a las cuatro características que componen a la variable en estudio (sección 5.4.1.5.3).

5.4.1.5.1. Verificación de suposiciones del Test No paramétrico de Mann-Whitney

Para verificar la validez de los resultados provistos por el test no paramétrico seleccionado, es necesario evaluar la suposición se igualdad de dispersión entre los grupos. Para ello, se aplica el test de Levene, de igual manera que en el experimento realizado en [Sharma, 2008], sin embargo, las variables a utilizar (como fue previamente indicado en la definición del experimento) son las siguientes: años de experiencia en la disciplina (exp), años de experiencias en la industria (exp_ind) y tiempo requerido para completar la encuesta (Time Taken). La tabla 5.112 presenta un resumen estadístico. La tabla 5.113 muestra los resultados de aplicar el test estadístico.

El test de Levene plantea como hipótesis que las varianzas en los grupos son distintas (p valor < 0.05), por lo tanto la hipótesis nula es que las varianzas en los grupos son iguales (p valor > 0.05). Si se obtiene un resultado significativo, se debe concluir que es poco probable que la diferencia de varianzas obtenidas sea resultado del proceso de selección aleatoria de las muestras de la población. Debiéndose rechazar la hipótesis nula y concluir que hay diferencia entre varianzas en la población.

De los resultados obtenidos en la tabla 5. 113 se observa que el test de Levene para las 3 variables analizadas (exp, exp_ind y Time Taken) no es significativo, con p valores iguales a 0.876, 0.741 y 0.496 respectivamente, y por lo tanto, podemos concluir que no existen pruebas suficientes para rechazar la hipótesis nula de igualdad de varianzas, verificándose la igualdad de varianzas en las muestras.

Estadísticas de grupo

	grupo	N	Media	Desv. Desviación	Desv. Error promedio
exp	ikddm	21	2,19	1,470	,321
	mopropei	21	2,14	1,459	,318
exp_ind	ikddm	21	,67	1,111	,242
	mopropei	21	,71	1,146	,250
TimeTaken	ikddm	21	21,95	6,352	1,386
	mopropei	21	21,71	5,596	1,221

Tabla 5.112. Descripción estadísticas de las variables experiencia y tiempo de respuesta

Prueba de muestras independientes

		Prueba de Levene de igualdad de varianzas		prueba t para la igualdad de medias						
									95% de intervalo de confianza de la diferencia	
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	Inferior	Superior
exp	Se asumen varianzas iguales	,025	,876	,105	40	,917	,048	,452	-,866	,961
	No se asumen varianzas iguales			,105	39,998	,917	,048	,452	-,866	,961
exp_ind	Se asumen varianzas iguales	,111	,741	-,137	40	,892	-,048	,348	-,752	,656
	No se asumen varianzas iguales			-,137	39,960	,892	-,048	,348	-,752	,656
TimeTaken	Se asumen varianzas iguales	,473	,496	,129	40	,898	,238	1,847	-3,495	3,972
	No se asumen varianzas iguales			,129	39,374	,898	,238	1,847	-3,497	3,973

Tabla 5.113. Test de Levene (igualdad de varianzas)

Por último, se analiza la igualdad de medias entre los grupos con el objetivo de identificar diferencias entre los mismos que puedan influir en la interpretación final de la hipótesis. Para ello se aplica la prueba T de igualdad de medias para muestras independientes. La tabla 5.113 resume el resultado de significancia bilateral para cada una de las variables analizadas (exp, exp_ind y Time Taken). Los p valores obtenidos son: 0.917, 0.892 y 9.896 respectivamente, concluyéndose que no se identifican diferencias significativas en los grupos analizados.

5.4.1.5.2. Evaluación de la hipótesis – Variable Principal

Una vez verificadas las suposiciones, sin encontrarse pruebas suficientes en contra de las condiciones bases para la aplicación del test de Mann-Whitney, se procede a aplicar dicho método para evaluar la hipótesis de investigación. Dicha prueba evalúa la diferencia entre los grupos a partir del rango de orden cada uno de los individuos. La tabla 5.114 ilustra los valores de rango medio y totales por grupo, el valor experimental (U) y la valor de aproximación por la normal (Z).

Finalmente, la tabla 5.115 muestra el p valor de significancia bilateral obtenido por el test (0.021), el cual es inferior al valor de significancia establecido (0.05) por lo cual se puede afirmar que existe una diferencia significativa entre los 2 grupos, es decir, que existe una relación entre la percepción de eficiencia de los usuarios y las metodologías utilizadas.

A partir de la tabla 5.114, se observa que el valor promedio de rangos para el grupo asociado con la metodología IKDDM es ocho puntos inferior al grupo asociado con la metodología MoProPEI y la diferencia en la suma de rangos es de más de 180 (a favor de la segunda propuesta). A partir de dichos valores puede identificarse que la diferencia en la percepción de eficiencia de los usuarios es favorable para el modelo propuesto en este trabajo de investigación.

Prueba de Mann-Whitney

Rangos

	grupo	N	Rango promedio	Suma de rangos
surveyscore	ikddm	21	17,14	360,00
	mopropei	21	25,86	543,00
	Total	42		

Estadísticos de prueba[a]

	surveyscore
U de Mann-Whitney	129,000
W de Wilcoxon	360,000
Z	-2,305
Sig. asintótica(bilateral)	,021

a. Variable de agrupación: grupo

Tabla 5.114. Detalles estadísticos test de Mann-Whitney - surveyscore

Resumen de prueba de hipótesis

	Hipótesis nula	Prueba	Sig.	Decisión
1	La distribución de surveyscore es la misma entre las categorías de grupo.	Prueba U de Mann-Whitney para muestras independientes	,021	Rechazar la hipótesis nula.

Se muestran significaciones asintóticas. El nivel de significación es de ,05.

Tabla 5.115. Resultado test de Mann-Whitney - surveyscore

En la próxima sección se presenta un análisis detallado del impacto de las características que integran la evaluación de calidad de las metodologías (facilidad de uso percibida, utilidad percibida, satisfacción del usuario, and calidad semántica percibida) con respecto a los resultados derivados en el presente estudio.

5.4.1.5.3. Evaluación de la hipótesis - Características

En esta sección tiene como objetivo ampliar los resultados obtenidos en la evaluación de la hipótesis de trabajo (sección 5.4.1.5.2), en la cual se encontraron diferencias significativas a favor de la propuesta realizada en esta tesis doctoral. Para ello, se propone implementar el test no paramétrico de Mann-Whitney (replicando el análisis realizado en [Sharma, 2008]) para analizar la existencia de diferencias significativas para las características que conforman a la variable principal de estudio (surveyscore): facilidad de uso percibida (PEOU), utilidad percibida (PU), satisfacción del usuario (US), y calidad semántica percibida (PSQ).

La tabla 5.116 ilustra los valores de rango medio y totales por grupo, el valor experimental (U) y la valor de aproximación por la normal (Z) para cada categoría de percepción. La tabla 5.117 muestra el p valor de significancia bilateral obtenido por el test por cada variable de análisis. Si el p valor obtenido es inferior al valor de significancia establecido (0.05) se puede afirmar que existe una diferencia significativa entre los 2 grupos, es decir, que existe una relación entre la categoría de percepción de eficiencia de los usuarios y las metodologías utilizadas.

Rangos

	grupo	N	Rango promedio	Suma de rangos
PEOU	ikddm	21	17,74	372,50
	mopropei	21	25,26	530,50
	Total	42		
US	ikddm	21	18,71	393,00
	mopropei	21	24,29	510,00
	Total	42		
PU	ikddm	21	18,21	382,50
	mopropei	21	24,79	520,50
	Total	42		
PSQ	ikddm	21	15,81	332,00
	mopropei	21	27,19	571,00
	Total	42		

Estadísticos de prueba[a]

	PEOU	US	PU	PSQ
U de Mann-Whitney	141,500	162,000	151,500	101,000
W de Wilcoxon	372,500	393,000	382,500	332,000
Z	-2,000	-1,481	-1,745	-3,022
Sig. asintótica(bilateral)	,046	,139	,081	,003

a. Variable de agrupación: grupo

Tabla 5.116. Detalles estadísticos test de Mann-Whitney - Componentes

A partir de la tabla 5.116, se observa que el valor promedio y suma total de rangos para cada una de las características es superior para la propuesta MoProPEI. A continuación se lista la diferencia entre las propuestas para cada una de las características, indicándose como par ordenado (valor promedio de rangos, suma total de rangos):

- PEOU = (7.82, 158)
- US = (5.58, 117)
- PU = (6.58, 138)
- PSQ = (11.38, 239)

Resumen de prueba de hipótesis

	Hipótesis nula	Prueba	Sig.	Decisión
1	La distribución de PEOU es la misma entre las categorías de grupo.	Prueba U de Mann-Whitney para muestras independientes	,046	Rechazar la hipótesis nula.
2	La distribución de US es la misma entre las categorías de grupo.	Prueba U de Mann-Whitney para muestras independientes	,139	Retener la hipótesis nula.
3	La distribución de PU es la misma entre las categorías de grupo.	Prueba U de Mann-Whitney para muestras independientes	,081	Retener la hipótesis nula.
4	La distribución de PSQ es la misma entre las categorías de grupo.	Prueba U de Mann-Whitney para muestras independientes	,003	Rechazar la hipótesis nula.

Se muestran significaciones asintóticas. El nivel de significación es de ,05.

Tabla 5.117. Resultado test de Mann-Whitney - Componentes

Para determinar la significancia de las diferencias identificas, debe observarse el p valor obtenido de aplicar el test de Mann-Whitney (tabla 5.117). Como resultado del mismo, se obtiene una diferencia significativa para los componentes PSQ (calidad semántica) y PEOU (facilidad de uso

percibida) con valores de significancia bilateral de 0.003 y 0.043 respectivamente, para los cuales se determina rechazar la hipótesis nula de igualdad de distribución entre los grupos. Sin embargo, para los componentes US (Satisfacción del usuario) y PU (Utilidad percibida), no se encuentran pruebas suficientes para rechazar la hipótesis nula (con p valores de 0,081 y 0.139) debiéndose retener la condición de similitud entre los grupos.

5.4.1.6. Análisis de los resultados

A partir del análisis experimental realizado, se derivan las siguientes conclusiones:

- Se identifica una diferencia significativa con respecto a la calidad del modelo de proceso propuesto, con respecto a la mejora de la eficiencia de las personas en el desarrollo de proyectos de explotación de información. Esta afirmación se concluye a partir de los resultados obtenidos (p valor = 0.021) de aplicar el test de Mann-Whitney con respecto a la percepción general de calidad entre los grupos (variable denominada en el experimento: "surveyscore").
- En análisis de las características de la calidad del modelo, realizado al grupo de 42 individuos, se obtuvieron resultados superiores en los cuatro aspectos. Sin embargo, los resultados del test de Mann-Whitney derivaron que los aspectos de calidad semántica (PSQ) y facilidad de uso percibida (PEOU) poseen una diferencia significativa (con valores de significancia bilateral de 0.003 y 0.043 respectivamente).

De los resultados previamente expuestos, se concluye que el grupo que utilizó la propuesta MoProPEI presentó un nivel de satisfacción significativo con respecto al grupo referido como IKDDM.

Interpretamos que los resultados obtenidos se derivan del mayor nivel de detalle e integración de las tareas a realizar, la definición de las técnicas y las dependencias entre las actividades del proyecto, así como la incorporación del proceso orientado a la gestión del proyecto, limitaciones que las propuestas predecesoras (entre ellas IKDDM) poseen.

Desde este contexto, es que entendemos que la relevancia de las características de calidad semántica y facilidad de uso percibida, surgen a partir de una carencia actual de las propuestas vigentes con respecto a brindar un proceso guiado que facilite el desarrollo del proyecto a los usuarios, en lugar de presentar una serie de recomendaciones de posibles acciones. En adición, la introducción de un proceso integrado que contemple los procesos orientados al producto y orientado a la gestión del

proyecto, reduce la cantidad de conceptos externos requeridos (asociado con la facilidad de uso percibido).

Finalmente, el experimento presenta evidencia inicial sobre la mejoría del modelo de proceso propuesto (MoProPEI) con respecto a IKDDM, la relevancia de la propuesta y su contribución a la disciplina. Sin embargo, estos resultados establecen un punto de inicio de una solución satisfactoria, cuya validación requiere ser iterativamente expandida convirtiéndose en un artefacto más fidedigno y relevante [Simon, 1996; Hevner et al., 2004].

6. CONCLUSIONES

En este capítulo se presentan las aportaciones de esta tesis doctoral (sección 6.1) y se destacan las futuras líneas de investigación que se consideran de interés en base al problema abierto que se presenta en este trabajo de tesis (sección 6.2).

6.1. APORTACIONES DE LA TESIS

Un modelo de proceso proporciona al equipo de desarrollo una visión estructurada y estandarizada del procedimiento a realizar para alcanzar los objetivos del proyecto, asegurando la calidad de sus resultados. En este sentido, dicho proceso no solo debe abordar los aspectos relacionados con la construcción del producto resultante, sino también cubrir aquellos aspectos asociados al desarrollo del proyecto; evaluando y reduciendo la posibilidad e impacto de los riesgos y desvíos.

Se ha postulado que las propuestas existentes para proyectos de ingeniería de explotación de información presentan aptitudes reducidas, señalándose como principales carencias identificadas en la literatura, las siguientes (detalladas en la sección 3.1, pág. 53): visión fragmentada del proceso, visión incompleta del proceso y las actividades, no se contempla o se realiza de manera parcial a los recursos en el proceso y no permiten adaptar la ejecución del proyecto de acuerdo a las necesidades específicas del mismo.

A continuación, se responden las problemáticas planteadas en la sección 3.3 (pág. 64) correspondientes al sumario de investigación:

Pregunta 1: ¿Es posible desarrollar un modelo de proceso integral el cual defina las actividades y dependencias para proyectos de explotación de información desde una visión ingenieril que incorpore los conceptos de planificación, administración y control requeridos en todo proceso?

Se ha definido un modelo de proceso el cual comprende los procesos orientados al producto y a la gestión. Dichos procesos proveen una visión integral requerida para el desarrollo de proyectos [Project Management Institute, Inc., 2013a], incorporando al tradicional subproceso de Desarrollo, de un proceso transversal enfocado en la planificación, administración y control del proyecto (subproceso Gestión). La estructura de la propuesta, está conformada por tres niveles jerárquicos de elementos (subprocesos, fases y actividades), cada una de los cuales brinda un mayor grado de especificidad en las tareas a realizar.

Pregunta 2: De ser viable, ¿Es posible definir el modelo de acuerdo a las prácticas y propuestas vigentes?

El modelo de proceso CRISP-DM, definido como el *estándar de facto* [Marbán et al., 2007; Kdnuggets, 2014], define la estructura del proceso orientado al producto que la mayoría de las propuestas modernas se basan (sección 2.2, pág. 12). Dicha estructura está compuesta de seis fases: entendimiento del negocio, entendimiento de los datos, transformación de los datos, modelado, evaluación y despliegue.

En este contexto, su utilización permite hacer uso de las prácticas desarrolladas hasta la fecha, facilitando la comprensión y adaptación de los grupos a la nueva propuesta. Sin embargo, a partir del crecimiento de la disciplina, cuyos proyectos incrementan en complejidad [Mariscal et al., 2010], a causa del incremento de fuentes de información accesibles (en cantidad y tamaño) para un proyecto y la amplia posibilidad de satisfacción de las necesidades que las mismas brindan; y las características intrínsecas de este tipo de proyectos: interdisciplinarios [Fayyad et al., 1996], complejos [Kurgan y Musilek, 2006; Gallardo, 2009] y dinámicos [Brachman & Anand, 1996], se identificaron una serie de carencias (desarrolladas en la sección 3.1, pág. 53) vigentes en las propuestas evaluadas.

Pregunta 3: ¿Existen indicios que deriven en la necesidad de adaptar o redefinir las prácticas existentes?

A partir de lo expuesto en la respuesta a la pregunta 2 (y abordado en detalle en secciones previas), se identifica la necesidad de adaptar la propuesta principalmente utilizada en la disciplina, con el propósito de cubrir las siguientes vacancias señaladas en la literatura:

A) Visión fragmentada del proceso, no contemplándose aquellas actividades asociadas con la gestión del proyecto [Kurgan y Musilek, 2006; Mariscal et al., 2010; do Nascimento y de Oliveira, 2012]. Estando en una situación similar a aquella que originó en la ingeniería del software, la crisis del software, a causa de la carencia de procesos o metodologías formales que den soporte de forma conjunta a los procesos orientados al producto y a la gestión de proyectos [Marbán et al., 2007].

B) Carencia de métodos o técnicas que proporcionen al usuario una guía detallada de las acciones a realizar en cada una de las actividades contempladas en los procesos existentes, dificultando la comprensión de cómo se produce el resultado esperado [Clifton y Thuraisingham, 2001; Charest y Delisle, 2006; Gallardo, 2009; Sharma y Osei-Bryson, 2009; El Sheikh y Alnoukari, 2012; Kdnuggets, 2014] y su implementación [Sharma y Osei-Bryson, 2009].

C) El incorrecto desarrollo del proceso y los desvíos de las metas establecidas en los proyectos, surgen a causa de decisiones deficientes o de la carencia de las mismas, originadas por el desconocimiento de las dependencias entre las numerosas actividades que conforman al proceso; las cuales deben ser claramente identificadas en orden a implementar de forma eficiente las tareas [Sharma, 2008].

D) Limitaciones en la adaptación del proceso a las necesidades específicas del proyecto [Marbán, et al., 2007; Gallardo, 2009; Mariscal et al., 2010] y la falta de consideración de los recursos humanos involucrados en el desarrollo del mismo [El Sheikh y Alnoukari, 2012].

Como resultado del análisis de lo previamente expuesto, se infiere la necesidad de adaptar, ampliar y profundizar los siguientes aspectos abordados en esta tesis:

A partir de la vacancia “A”, se determina la necesidad de incorporar al tradicional proceso orientado al producto (desarrollo), un proceso que incorpore las prácticas requeridas para la gestión de este tipo de proyectos, utilizándose la guía de los fundamentos para la dirección de proyectos (guía PMBOK) [Project Management Institute, Inc., 2013a] como propuesta base. Esta se encuentra conformada por cinco fases: iniciación, planificación, ejecución, control y cierre.

A partir de las vacancias “B” y “C”, se identifica la necesidad de incorporar técnicas que precisen los pasos y resultados requeridos para cada una de las actividades para este tipo de proyectos; y definir las dependencias entre las actividades del proyecto.

A partir de la vacancia “D”, se infiere la necesidad de incorporar herramientas al proceso que permitan comprender la participación de los miembros del equipo de desarrollo durante el transcurso del proyecto; como así también adaptar su estructura de acuerdo a las características del proyecto y de las personas involucradas.

Como resultado del análisis de la integración de los procesos, técnicas y dependencias entre las actividades, se determinó la necesidad de realizar las siguientes adaptaciones a la estructura de los procesos propuestos:

Las fases del proceso gestión se mantuvieron de acuerdo a la propuesta base, sin embargo, el nombre de la fase “Ejecución” fue modificado por “Soporte” para evitar ambigüedades con la fase implementación del proceso orientado al producto. En estas fases se incorporan actividades que permiten ajustar la estructura y ejecución de las mismas, controlar el

progreso del proyecto, estimar, planificar y evaluar los recursos requeridos y su disposición durante el proceso.

El proceso desarrollo fue adaptado de acuerdo a las dependencias entre las actividades. La fase "Modelado", fue separada en dos fases: "Modelado" e "Implementación". La primera de ellas, es introducida por la propuesta, incorporando las actividades que permiten sistematizar el proceso de definición de las técnicas a utilizar para dar solución a los objetivos identificados (es decir, mapear las necesidades del negocio con los modelos requeridos y definir la forma mediante la cual se determinará la calidad de los resultados). La segunda fase cubre las actividades de explotación de información y optimización del modelo y evaluación de la calidad y validez de los patrones obtenidos (consideradas en la fase original). La incorporación de la fase "Modelado", surge a partir de las dependencias identificadas entre la conformación del set de datos y la selección de los algoritmos a utilizar. Dicha etapa favorece la definición temprana del enfoque a utilizar, evitando iteraciones y esfuerzos innecesarios.

Adicionalmente, a partir del reordenamiento de actividades correspondientes a la gestión del proyecto y la desestimación de actividades ajenas al alcance del modelo de proceso (revisión del proyecto, plan de monitoreo y mantenimiento), se integran las fases de "Evaluación" y "Despliegue", modificando el segundo término a "Presentación", para enfatizar en la correcta transmisión los patrones y las piezas de conocimiento extraídas, como objetivo del proceso.

Pregunta 4: ¿Es posible articular técnicas o procedimientos de ingeniería de explotación de información desarrollados *ad hoc* en dicho modelo de proceso?

A partir del análisis de la disciplina, se identifican técnicas desarrolladas *ad hoc* para utilizar en procesos de ingeniería de explotación de información. A continuación se listan las propuestas según el área de interés: entendimiento del negocio [Britos et al., 2008; Gallardo, 2009; Sharma, y Osei-Bryson, 2009], entendimiento de los datos [Britos et al., 2008;], modelado [Britos y García-Martínez, 2009; García-Martínez et al., 2013; Martins et al., 2014], viabilidad del proyecto [Pytel et al., 2015], estimación del esfuerzo del proyecto [Marbán et al., 2008; Pytel et al., 2015], selección de herramientas [Collier et al., 1999; Britos et al., 2006], métricas [Basso et al., 2014] y ciclos de vida [Hofmann, 2003; Alnoukari, 2010; Arboleya, 2013]. La tabla 6.1 lista las actividades cubiertas con las técnicas identificadas a partir de la evaluación de las propuestas previamente mencionadas.

ACTIVIDAD	TÉCNICA
Gestión / Iniciación / Exploración Inicial del Proyecto	Caracterización del desarrollo del proyecto perteneciente a la Metodología para la educción de requerimientos para proyectos de explotación de información [Britos et al., 2008]
Gestión / Iniciación / Evaluación de la Situación	Metodología para la selección de Herramientas de Explotación de Información [Britos et al., 2006]
Gestión / Iniciación / Evaluación de la Situación	Modelo de Evaluación de Viabilidad para Proyectos de Explotación de Información [Pytel et al., 2015]
Gestión / Iniciación / Definición del Ciclo de Vida	Ciclos de vida propuestos en [Hofmann, 2003; Alnoukari, 2010; Arboleya, 2013]
Gestión / Planificación / Planificación de la Mediciones	Métricas para Proyectos de Explotación de Información [Basso et al., 2013]
Gestión / Planificación / Planificación de la Mediciones	Modelo de Estimación para Proyectos de Explotación de Información [Pytel et al., 2015]
Desarrollo / Entendimiento del Negocio / Análisis del Negocio	Definición de los objetivos del proyecto que forma parte de la Metodología para la educción de requerimientos para proyectos de explotación de información [Britos et al., 2008]
Desarrollo / Entendimiento del Negocio / Comprensión del Problema de Negocio	Definición de los Problema de Negocio que forma parte de la Metodología para la educción de requerimientos para proyectos de explotación de información [Britos et al., 2008]
Desarrollo / Entendimiento de los datos / Análisis de los Datos	Identificación de atributos relacionados con el Problema de Negocio definida en la Metodología para la educción de requerimientos para proyectos de explotación de información [Britos et al., 2008]
Desarrollo / Modelado / Modelado del Problema	Proceso de Explotación de Información [Britos y García-Martínez, 2009; García-Martínez et al., 2013]
Desarrollo / Modelado / Modelado del Problema	Derivación del Proceso de Explotación de Información [Martins et al., 2014]

Tabla 6.1. Listado de técnicas de explotación de información utilizadas

Pregunta 5: En caso de poder articular las técnicas con el modelo de proceso propuesto: ¿Existen técnicas faltantes para las actividades? De existir, ¿pueden dichas técnicas ser adaptadas de otras disciplinas o desarrolladas ad hoc?

De la respuesta provista a la pregunta anterior, se desprende la necesidad de incorporar técnicas para las actividades faltantes. La tabla 6.2 lista aquellas actividades para las cuales se utilizan técnicas pertenecientes a otras disciplinas, que puedan ser utilizadas o adaptadas para proyectos de explotación de información.

ACTIVIDAD	TÉCNICA
Gestión / Iniciación / Definición de la Comunicación	Definición de la Comunicación [Verzuh, E., 2015]
Gestión / Planificación / Planificación de las Responsabilidades	Designación de Responsabilidades [Project Management Institute, Inc., 2013a]
Gestión / Control / Gestión del Desarrollo	Reporte de Estado [Project Management Institute, Inc., 2013a; Verzuh, E., 2015]
Gestión / Cierre / Formalización Interna del Cierre del Proyecto	Reporte de Cierre [Project Management Institute, Inc., 2013a; Verzuh, E., 2015]

Tabla 6.2. Listado de técnicas adaptables

Las tablas 6.3 y 6.4, resumen el conjunto de técnicas utilizadas para cada una de las actividades que conforman a los subprocesos Gestión y Desarrollo (respectivamente), incorporándose al listado de técnicas seleccionadas en las tablas previas, aquellas desarrolladas *ad hoc* para cubrir las vacancias del modelo de proceso propuesto. Las herramientas y técnicas identificadas en la propuesta pueden ser utilizadas como medio para resolver las dificultades y problemáticas que podrían ocurrir al desconocer o no aplicar tareas esenciales para el desarrollo de una solución viable y útil.

SUBPROCESO GESTIÓN			
FASE	**ACTIVIDAD**	**TÉCNICA**	**DESCRIPCIÓN**
Iniciación	Exploración Inicial del Proyecto	Caracterización del desarrollo del proyecto perteneciente a la Metodología para la educción de requerimientos para proyectos de explotación de información [Britos et al., 2008]	Se utilizan las herramientas propuestas para formalizar los recursos humanos involucrados, los riesgos y contingencias del proyecto.
	Definición de la Comunicación	Definición de la Comunicación [Verzuh, E., 2015]	Se planifica y formalizan las comunicaciones previstas durante el proceso
	Evaluación de la Situación	Metodología para la selección de Herramientas de Explotación de Información [Britos et al., 2006] y Modelo de Evaluación de Viabilidad para Proyectos de Explotación de Información [Pytel et al., 2015]	Se complementan las técnicas de evaluación de las herramientas de explotación de información y viabilidad del proyecto, permitiendo determinar la posible resolución de las necesidades identificadas
	Definición del Ciclo de Vida	Selección del Ciclo de Vida	Se incorpora un procedimiento y se describen las características a considerar para los recursos [Hofmann, 2003; Alnoukari, 2010; Arboleya, 2013]
Planificación	Planificación de la Mediciones	Métricas para Proyectos de Explotación de Información [Basso et al., 2013] y Modelo de Estimación para Proyectos de Explotación de Información [Pytel et al., 2015]	Se adapta la propuesta de estimación, incorporando el esfuerzo requerido para el subproceso de gestión de acuerdo a lo definido en [Mochal, T., 2006] para proyectos en general
	Planificación de las Actividades	Definición del Programa del Proyecto	Su utilizan las técnicas mapa de actividades y diagramas Gantt, y las estimaciones empíricas de carga de trabajo realizadas en [Rodríguez et al., 2010]
	Planificación de los Recursos	Planificación de los Recursos Necesarios	Lista de roles de recursos humanos [El Sheikh y Alnoukari, 2012]
	Planificación de las Responsabilidades	Designación de Responsabilidades [Project Management Institute, Inc., 2013a]	Se registra la matriz de responsabilidades [Project Management Institute, Inc., 2013a] y se propone un reporte para la propuesta del proyecto
Soporte	Mediciones del Proyecto	Cálculo de Métricas	Registro de las métricas definidas según [Basso et al., 2013]
	Gestión de la Configuración	Configuración del versionado	Registro del cambio de versiones de los productos internos del proyecto
Control	Gestión del Desarrollo	Seguimiento de Avance	Basado en las propuestas de reporte de estado [Project Management Institute, Inc., 2013a; Verzuh, E., 2015]
	Control de las Actividades	Evaluación de Riesgos	Registro de riesgos acontecidos.
	Gestión del Cambio	Evaluación del Cambio	Registro formal de las peticiones de cambio y su resolución
Cierre	Formalización Externa del Cierre del Proyecto	Presentación de Conformidad	Documento de aceptación del proyecto
	Formalización Interna del Cierre del Proyecto	Evaluación del Proceso	Basado en las propuestas de reporte de cierre [Project Management Institute, Inc., 2013a; Verzuh, E., 2015]

Tabla 6.3. Listado de técnicas: subproceso Gestión

SUBPROCESO DESARROLLO			
FASE	**ACTIVIDAD**	**TÉCNICA**	**DESCRIPCIÓN**
Entendimiento del Negocio	Análisis del Negocio	Definición de los objetivos del proyecto que forma parte de la Metodología para la educción de requerimientos para proyectos de explotación de información [Britos et al., 2008]	Se utilizan los formalismos propuestos para dejar registro del conocimiento extraído de los interesados con respecto a las fuentes de información disponibles, la terminología, los objetivos, criterios de éxito y las expectativas, suposiciones y restricciones del proyecto.
	Comprensión del Problema de Negocio	Definición de los Problema de Negocio que forma parte de la Metodología para la educción de requerimientos para proyectos de explotación de información [Britos et al., 2008]	Se utilizan los formalismos propuestos a partir de los cuales se deja registro de los problemas de negocio de interés y sus criterios de éxito.
Entendimiento de los Datos	Análisis de los Datos	Identificación de atributos relacionados con el Problema de Negocio definida en la Metodología para la educción de requerimientos para proyectos de explotación de información [Britos et al., 2008]	Se incorpora el diccionario de datos al formalismo propuesto en [Britos et al., 2008] para la identificación de los campos de interés para los problemas de negocio.
	Exploración de los Datos	Exploración de los Datos	Se utiliza la estadística (resumen de cinco datos [Han et al., 2011]) y visualizaciones para la descripción de los datos
	Evaluación de los Datos	Exploración de la Calidad de los Datos	Se propone el uso de procedimientos de detección de anomalías en los datos (como el propuesto en [Kuna, H. 2013]), visualizaciones y la estadística como herramientas para la evaluación de la calidad de los datos
Modelado	Modelado del Problema	Derivación del Proceso de Explotación de Información [Martins, et. Al, 2014]	Se utilizan los procesos de explotación de información [Britos y García-Martínez, 2009; García-Martínez et al., 2013] junto con la técnica propuesta en [Martins, et. Al, 2014] como herramientas para vincular los requisitos del negocio con los modelos a utilizar
	Configuración del Modelo	Determinación de la Configuración del Modelo	Se formalizan los algoritmos, las variables y las técnicas de evaluación de los resultados a utilizar para la configuración del modelo
Preparación de los Datos	Construcción de la Fuente Temporaria de Datos	Generación de la Fuente Temporaria de Datos	Se generan los datos de acuerdo a las necesidades específicas de las herramientas y técnicas a utilizar
	Adecuación de la Fuente Temporaria de Datos	Adecuación de los Datos	Se transforman los datos de acuerdo a las necesidades específicas de las herramientas y técnicas a utilizar
Implementa-ción	Selección del Modelo	Selección de la Estrategia de Hiperparametrización	Se proponen distintas técnicas de optimización de los parámetros de configuración del modelo
	Explotación de Información	Extracción de Conocimiento	Se implementan los modelos de acuerdo a lo definido en [Britos y García-Martínez, 2009] y se registran los resultados obtenidos por los modelos propuestos
Evaluación y Presentación	Evaluación de los Resultados	Validación del Conocimiento	Se validan los resultados con los interesados definiéndose los próximos pasos
	Presentación de los Resultados	Síntesis del Proyecto	Se proponen las áreas de conocimiento generales del proyecto de interés a formalizar

Tabla 6.4. Listado de técnicas: subproceso Desarrollo

Para comprobar **la viabilidad de la propuesta**, la misma fue aplicada a tres proyectos pertenecientes a distintos dominios: educación, salud y análisis web (presentados en el capítulo 4 y las secciones 5.1 y 5.2, respectivamente), verificando la correcta integración de los elementos propuestos. De forma complementaria, mediante la propuesta del marco comparativo de metodologías para proyectos de minería de datos (sección 5.3, pág. 409) [Moine, 2013], se compararon los aportes del modelo de proceso presentado en esta tesis (MoProPEI) con respecto a las propuestas más relevantes: CRISP-DM, Catalyst, KDD, SEMMA, IKDDM, MPIMD, ASD-BI, FMDS y TDSP (descriptas en la sección 2.2, pág. 12). La tabla 6.5 resume los porcentajes de

cubrimiento para cada uno de los aspectos (calculada a partir de los valores obtenidos en la tabla 5.108). La figuras 6.1 y 6.2 facilitan la visualización de los resultados obtenidos.

ASPECTOS	CRISP-DM	Catalyst	KDD	SEMMA	IKDDM	MPIMD	ASD-BI	FMDS	TDSP	MoProPEI
Nivel de detalle en las actividades de cada fase (5)	80%	80%	20%	20%	100%	20%	20%	0%	80%	100%
Escenarios de aplicación (4)	75%	100%	75%	50%	75%	75%	75%	75%	75%	75%
Actividades específicas en cada fase (26)	77%	81%	46%	38%	77%	42%	65%	62%	62%	85%
Actividades para la dirección del proyecto (17)	47%	47%	0%	0%	0%	35%	0%	0%	18%	88%
TOTAL (52)	**67%**	**71%**	**31%**	**25%**	**54%**	**40%**	**40%**	**37%**	**50%**	**87%**

Tabla 6.5. Comparación de metodologías – Porcentuales por aspecto

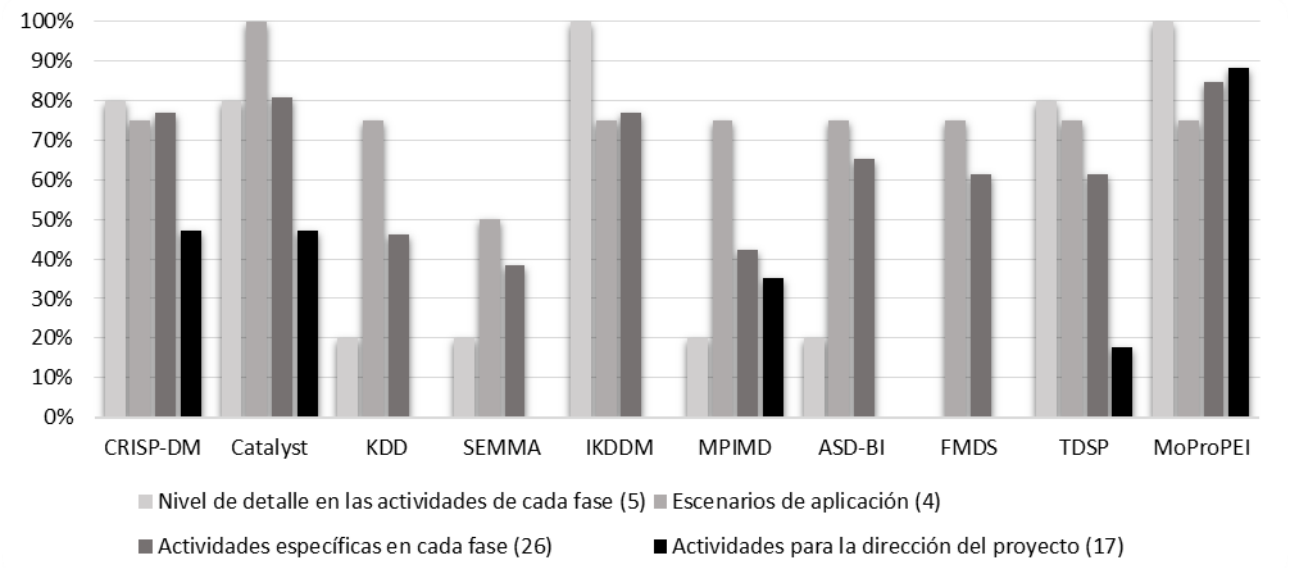


Figura 6.1. Comparación de metodologías– Evaluación general (por aspectos porcentual)

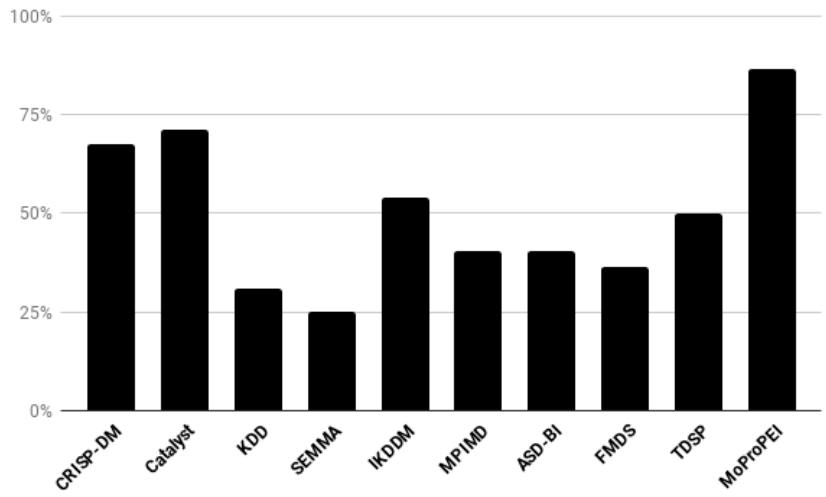


Figura 6.2. Comparación de metodologías– Evaluación general (total porcentual)

A partir de la tabla 6.5, se observa que MoProPEI (con un cubrimiento total del 87%) supera por más de un 15% a las siguientes propuestas con mayor cubrimiento de los aspectos requeridos para una metodología de explotación de información, estas son: Catalyst (71%) y CRISP-DM (67%). Si bien MoProPEI posee el mayor cubrimiento en tres de los cuatro aspectos evaluados (uno compartido con IKDDM), la diferencia más significativa se encuentra en los aspectos orientados en las actividades para la dirección del proyecto, en el cual casi duplica el cubrimiento de las aportaciones de los otros enfoques.

Por último, se realizó un experimento controlado (sección 5.4, pág. 414) para validar las aportaciones del modelo de proceso propuesto con respecto a sus predecesores. Como resultado del análisis sistemático de la literatura (presentado en el Anexo B, pág. 471) se identificó un único diseño experimental replicable [Sharma, 2008], el cual compara al modelo de proceso definido como *estándar de facto* con IKDDM. A partir de lo previamente expuesto, es que se replica el experimento propuesto, comparando el modelo de proceso superador (IKDDM) con el propuesto en este trabajo de investigación.

Para el desarrollo del experimento 42 profesionales fueron encuestados divididos en 2 grupos (asociados a las metodologías en evaluación) de manera aleatoria simple. Los participantes respondieron 16 preguntas (cuya respuesta consistía en una escala de Likert de 7 valores) las cuales brindaban su percepción de la metodología respecto a 4 aspectos: Facilidad de uso percibida, Utilidad Percibida, Calidad Semántica y Satisfacción del usuario. A partir de dichas preguntas, se generó una variable global como indicador de la percepción de calidad de la propuesta. En adición, la experiencia del participante (en la disciplina y en la industria) y el tiempo en responder la encuesta fueron registrados.

Como resultado de aplicar el test no paramétrico de Mann-Whitney para muestras independientes con un nivel de significancia del 5%, se concluyó: que existe una diferencia significativa entre los 2 grupos, es decir, que existe una relación entre la percepción de eficiencia de los usuarios y las metodologías utilizadas. El p valor de significancia bilateral obtenido por el test es de 0.021, siendo inferior al valor de significancia establecido (0.05), identificándose una diferencia positiva para MoProPEI. En adición, del análisis realizado por cada aspecto de percepción, se obtuvo una diferencia significativa para los componentes PSQ (calidad semántica) y PEOU (facilidad de uso percibida) con valores de significancia bilateral de 0.003 y 0.043 respectivamente.

De los resultados derivados del experimento, se concluye que el grupo que utilizó la propuesta MoProPEI presentó un nivel de satisfacción significativo con respecto al grupo referido como IKDDM.

Del análisis de los resultados del experimento, se derivaron las siguientes interpretaciones (detalladas en la sección 5.4.1.6, pág. 426):

- La relevancia de las características de calidad semántica y facilidad de uso percibida, surgen a partir de una carencia actual de las propuestas vigentes con respecto a brindar un proceso guiado que facilite el desarrollo del proyecto a los usuarios, en lugar de presentar una serie de recomendaciones de posibles acciones, dificultando su implementación, pudiendo esto explicar por qué las tareas indicadas no siempre se implementan formalmente [Sharma y Osei-Bryson, 2009].
- La introducción de un proceso integrado que contemple los procesos orientados al producto y orientado a la gestión del proyecto, reduce la cantidad de conceptos externos requeridos (factor asociado con la facilidad de uso percibida).
- Los resultados obtenidos en el experimento presentan evidencia inicial sobre la mejoría del modelo de proceso propuesto (MoProPEI) con respecto a IKDDM, su relevancia y su contribución a la disciplina. Sin embargo, estos resultados establecen un punto de inicio de una solución satisfactoria, cuya validación requiere ser iterativamente expandida convirtiéndose en un artefacto más fidedigno y relevante [Simon, 1996; Hevner et al., 2004].

Se considera de interés señalar que el modelo de proceso propuesto ha sido aplicado en una variedad mayor de proyectos: realizados en la industria sobre una variedad de dominios (entidades financieras, consultoras de RRHH, comercios, entre otros.) y problemáticas (Churn Analysis, predicción de ventas, devoluciones y stock, análisis web, marketing personalizado, entre otros.), desarrollados en la asignatura Tecnologías de Explotación de Información perteneciente al quinto año de la Licenciatura en Sistema de la Universidad Nacional de Lanús (entre los años 2016 y 2019), en dos trabajos finales de carrera y dos tesis de maestría (en progreso).

Por último, se especifican las aportaciones realizadas en esta tesis:

- Se introduce un modelo de proceso el cual presenta una visión integrada de los procesos necesarios para el desarrollo de un proyecto de explotación de información (orientado al producto y gestión); los cuales se conciben mediante una visión ingenieril que define los procesos, fases, actividades, técnicas y las dependencias entre cada uno de los elementos que lo conforman.
- Se incorpora la fase Modelado al proceso orientado al producto con el objetivo de vincular los conocimientos y necesidades del cliente con los modelos (o procesos de explotación de información) a utilizar para dar respuesta a las problemáticas identificadas. Esta fase permite

la identificación temprana del enfoque a utilizar, evitando iteraciones y esfuerzos innecesarios entre las etapas de preparación de los datos y de implementación del modelo.

- El subproceso Gestión presenta un mejor ordenamiento de las actividades existentes según sus objetivos y alcances, dado que permite eliminar las actividades de gestión presentes en el proceso orientado al producto. Además, se incorporan un conjunto de tareas requeridas para el desarrollo de proyectos que no eran contempladas. Las tareas introducidas son:
  - *Definición de la comunicación*: en la cual se realiza un esfuerzo por mejorar las comunicaciones entre los participantes del proyecto (interesados y el equipo de trabajo), siendo la falta o ineficiencia en las comunicaciones una de las principales causas de demoras y fracasos en proyectos [Project Management Institute, 2013b].
  - *Estudio de viabilidad*: donde se determina de manera temprana la posibilidad de éxito del proyecto acorde a las características del mismo y de las herramientas disponibles.
  - *Selección del ciclo de vida*: se determina la forma mediante la cual el proceso será desarrollado, permitiendo adaptar el mismo a las necesidades del proyecto.
  - *Control de indicadores*: Seguimiento de los aspectos de interés para el desarrollo del proyecto (actividades y recursos), permitiendo identificar de manera temprana desvíos en el proyecto.
  - *Evaluación del cambio*: registro y valoración de las peticiones de cambio y su impacto en el proyecto.
  - *Trazabilidad del proyecto*: registro y comunicación de los cambios del proyecto.
  - *Contratación*: definición del acuerdo y registro de finalización de las responsabilidades asumidas.
- La incorporación de las técnicas y formalismos para la ejecución de las actividades precisan los alcances y resultados esperados de cada uno de los elementos de la propuesta; permitiendo así al equipo de desarrollo unificar la visión del proceso y facilitar la comprensión del mismo. En adición, la introducción de formalismos que sistematicen los elementos de salidas definidos incrementa las posibilidades de finalizar de manera exitosa el desarrollo de un proyecto [Microsoft, 2016].
- Se define al modelo de proceso propuesto de forma independiente al flujo de ejecución de las actividades. En tal sentido, se presentan tres modelos de ciclo de vida que permiten al equipo de trabajo determinar, a partir de las características del proyecto, el mejor flujo de trabajo.

6.2. FUTURAS LÍNEAS DE INVESTIGACIÓN

Una aportación secundaria de la tesis, está asociada con la visión integral del proceso para el desarrollo de un proyecto de explotación de información y la valoración del cubrimiento actual de dicho proceso. En este contexto, se evidencia cómo la disciplina se encuentra principalmente enfocada en las tareas asociadas con la explotación de información, mientras que otras tareas de igual importancia (con respecto a la validez y calidad de los resultados) no poseen la misma madurez. Como resultado de esta investigación, se identifican las siguientes futuras líneas de investigación:

- Si bien el modelo de proceso propuesto en esta tesis aporta sistematicidad al desarrollo de proyectos de explotación de información y el mismo ha sido validado en dominios representativos, quedan como líneas abiertas de trabajo las siguientes:
  - Ampliar los casos muestrales utilizados en el experimento realizado para validar el modelo de proceso propuesto, así como las propuestas a comparar.
  - Ampliar la validación empírica del modelo de proceso y las técnicas en un conjunto vasto y representativo, considerando las características de distintos tipos de dominios y datos.
- Ampliar el conjunto de técnicas consideradas en cada actividad, incrementando el cubrimiento de dominios y casos.
- Incorporar aquellas actividades no consideradas en el alcance de la propuesta actual: definición y seguimiento de la tercerización del proyecto (o parte), formación y mejora de recursos, implantación y seguimiento del modelo en producción.

A continuación se listan aquellas técnicas que requieren de un refinamiento con el objetivo de ampliar los alcances o cubrir de manera completa las necesidades de la actividad asociada:

- Refinar el modelo de estimación de esfuerzos e incorporar el esfuerzo requerido para las tareas de gestión.
- Incorporar un procedimiento que permita estimar el grado y la probabilidad de ocurrencia de los riesgos del proyecto.
- Los costos del proyecto se derivan a partir de la necesidad de los recursos en la línea de tiempo; no obstante, se considera necesario incorporar un modelo de estimación de costos que permita predecir dicha información a partir de las características del proyecto.
- Refinar la técnica de evaluación de herramientas acorde a las características actuales de la disciplina (por ejemplo: considerar la posibilidad de distribuir el procesamiento).
- Refinar los procedimientos definidos *ad hoc* para el proceso.

7. REFERENCIAS

Abraham, A. (2003). Business Intelligence from Web Usage Mining. Journal of Information & Knowledge Management, 24: 375-390.

Abran, A., Moore, J. W., Bourque, P., Dupuis, R., Tripp, L. (2004). Guide to the Software Engineering Body of Knowledge (2004 version). IEEE Computer Society Press. ISBN 0-7695-2330-7.

Adriaans, P., & Zantinge, D. (1996). Data Mining. Harlow, UK: Addison-Wesley.

Alnoukari, M. (2010). ASD-BI: A Business Intelligence Modeling and Integration Framework based on Agile Methodologies (Doctoral dissertation, Arab Academy for Banking and Financial Sciences).

Alnoukari, M. (2012). ASD-BI: A Knowledge Discovery Process Modeling Based on Adaptive Software Development Agile Methodology. In Business Intelligence and Agile Methodologies for Knowledge-Based Organizations: Cross-Disciplinary Applications (pp. 183-207). IGI Global.

Alnoukari, M., & El Sheikh, A. (2012). Knowledge Discovery Process Models: From Traditional to Agile Modeling. Business Intelligence and Agile Methodologies for Knowledge-Based Organizations: Cross-Disciplinary Applications, 72-100.

Alnoukari, M., Alzoabi, Z., & Hanna, S. (2008). Applying adaptive software development (ASD) agile modeling on predictive data mining applications: ASD-DM Methodology. In Information Technology, 2008, August. ITSim 2008. International Symposium on (Vol. 2, pp. 1-6). IEEE.

Anand, S. & Buchner, A. (1998). Decision Support Using Data Mining. Financial Times Management, 184.

Arboleya, H. (2013). Propuesta de Ciclo de Vida y Mapa de Actividades para Proyectos de Explotación de Información. Revista Latinoamericana de Ingeniería de Software, 1(3): 107-124, ISSN 2314-2642

Basso, D. (2013). Propuesta de métricas para proyectos de explotación de información. Revista Latinoamericana de Ingeniería de Software, 2(4), 157-218.

Berry, M. J., & Gordon, L. (1997). Data mining techniques: For marketing, sales, and customer support. New York, NY: Wiley.

Berry, M., & Linoff, G. (2004). G Data Mining Techniques: for marketing, sales, and customer relationship management USA.

Brachman, R. J., & Anand, T. (1996). The process of knowledge discovery in databases. In Advances in knowledge discovery and data mining (pp. 37-57). American Association for Artificial Intelligence.

Britos, P., Dieste, O., & García-Martínez, R. (2008). Requirements elicitation in data mining for business intelligence projects. In Advances in Information Systems Research, Education and Practice (pp. 139-150). Springer US.

Britos, P., García-Martínez, R. (2009). Propuesta de Procesos de Explotación de Información. Proceedings XV Congreso Argentino de Ciencias de la Computación. Workshop de Base de Datos y Minería de Datos. Págs. 1041-1050. ISBN 978-897-24068-4-1.

Britos, P., Merlino, H., Fernández, E., Ochoa, M., Diez, E. y García Martínez, R. (2006). Tool Selection Methodology in Data Mining. Proceedings V Ibero-American Symposium on Software Engineering. Pág. 85-90.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1998). Discovering data mining: From concept to implementation. Upper Saddle River, NJ: Prentice Hall.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide.

Charest, M., & Delisle, S. (2006). Ontology-guided intelligent data mining assistance: Combining declarative and procedural knowledge. In Artificial Intelligence and Soft Computing (pp. 9-14).

Cios, K. J., Pedrycz, W., Swiniarski, R. W., & Kurgan, L. A. (2007). Data mining: A knowledge discovery process. Berlin, Heidelberg: Springer.

Cios, K. J., Teresinska, A., Konieczna, S., Potocka, J., & Sharma, S. (2000). Diagnosing myocardial perfusion from PECT bull's-eye maps—A knowledge discovery approach. IEEE Engineering in Medicine and Biology Magazine, Special issue on Medical Data Mining and Knowledge Discovery, 19(4), 17-25.

Clark, W., Polakov, W. N., & Trabold, F. W. (1922). The Gantt chart: A working tool of management. Ronald Press Company.

Clifton, C., & Thuraisingham, B. (2001). Emerging standards for data mining. Computer Standards & Interfaces, 23(3), 187-193.

Collier, K., Carey, B., Sautter, D., & Marjaniemi, C. (1999). A methodology for evaluating and selecting data mining software. In Systems Sciences, 1999. HICSS-32. Proceedings of the 32nd Annual Hawaii International Conference on (pp. 11-pp). IEEE. ISO 690.

Cooley, R. (2003). The Use of Web Structure and Content to Identify Subjectively Interesting Web Usage Patterns. ACM Transactions on Internet Technology, 32: 93-116.

Creswell, J. (2002). ―Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research". Prentice Hall. ISBN 10: 01-3613-550-1.

Curtis, B., Kellner, M., Over, J. (1992). Process Modelling. Communications of the ACM, 359: 75-90.

Debuse, J. C. W., de la Iglesia, B., Howard, C. & Rayward-Smith, V. (2001). Building the KDD Roadmap: A Methodology for Knowledge Discovery. Industrial Knowledge Management. Springer-Verlag, 179–196.

Díaz, L., Ramón Garcia-Martinez, R. (2016). Hacia una Praxis Transformadora de la Comprensión del Estudiante de Educación Superior en Contextos de Masividad. En ―Estado, Política Pública y Acción Colectiva: Praxis Emergentes y Debates Necesarios en la Construcción de la Democracia" (Editor La Serna, C.). Pág. 256-267. Instituto de Investigación y Formación en Administración Pública. Universidad Nacional de Córdoba. ISBN 978-950-33-1255-1.

do Nascimento, G. S., & de Oliveira, A. A. (2012). An agile knowledge discovery in databases software process. In Data and Knowledge Engineering (pp. 56-64). Springer Berlin Heidelberg.

do Nascimento, G. S., & de Oliveira, A. A. (2013). Agilekdd: an agile process model to knowledge discovery in databases and business intelligence systematization. In 10 contecsi.

Edelstein, H. (1998). Data mining: Let's get practical. DB2 Magazine 3(2) 38-40.

El Sheikh, A. A. R., & Alnoukari, M. (2012). Business Intelligence and Agile Methodologies for Knowledge-Based Organizations: Cross-Disciplinary Applications. Business Science Reference.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37.

Feldens, M. A., Moraes, R. L., Pavan, A., & Castilho, J. M. (1998). Towards a methodology for the discovery of useful knowledge combining data mining, Data warehousing and visualisation. Universidade Federal do Rio Grande do Sul.

Ferreira, J., Takai, O., Pu, C. (2005). Integration of Business Processes with Autonomous Information Systems: A Case Study in Government Services. Proceedings Seventh IEEE International Conference on E-Commerce Technology. Pág. 471-474.

Gaber, M., Zaslavsky, A. Krishnaswamy, S. (2010). Data stream mining. En Maimon, O. and Rokach, L., eds. Data mining and knowledge discovery handbook. Springer, Pág. 759-787. ISBN 978-0-387-09823-4.

Gallardo, J. (2009). Metodología para la Definición de Requisitos en Proyectos de Data Mining (ER-DM). Facultad de Informática (UPM), Madrid, España, Tesis Doctoral.

García Martínez, R. (1997). Sistemas Autónomos. Aprendizaje Automático. Editorial Nueva Librería. ISBN 950-9088-84-6.

García Martínez, R., Servente, M. y Pasquini, D. (2003). Sistemas Inteligentes. Editorial Nueva Librería. Buenos Aires. ISBN 987-1104-05-7.

García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P., Vanrell. J. (2011). Towards an Information Mining Engineering. En Software Engineering, Methods, Modeling and Teaching. Sello Editorial Universidad de Medellín. ISBN 978-958-8692-32-6. Páginas 83-99.

García-Martínez, R., Britos, P., Rodríguez, D. (2013). Information Mining Processes Based on Intelligent Systems. Lecture Notes on Artificial Intelligence, 7906: 402-410. ISBN 978-3-642-38576-6.

Gartner (2000). Free methodology and process model for data mining released. https://www.gartner.com/doc/314396/free-methodology-process-model-data (Último acceso 12/03/2017).

Gondar, J. E. (2005). Metodología del Data Mining. Data Mining Institute, SL.

Gopal, R., Marsden, J. R., & Vanthienen, J. (2011). Information mining—Reflections on recent advancements and the road ahead in data, text, and media mining.

Haglin, D., Roiger, R., Hakkila, J., & Giblin, T. (2005). A tool for public analysis of scientific data. Data Science Journal, 4(30), 39–53. doi:10.2481/dsj.4.39

Han, J., & Cercone, N. (2000). RuleViz: A model for visualizing knowledge discovery process. In Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 244-253).

Han, J., & Kamber, M. (2001). Data mining: Concepts and techniques. Morgan Kaufmann.

Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.

Harry, M. & Schroeder, R. (1999). Six Sigma, the Breakthrough Management Strategy Revolutionizing the World's Top Corporations. Currency.

Hevner, A. R., March S. T., Park j., Ram S. (2004). Design Science In Information Systems Research. MIS Quarterly, 28(1), 75-105.

Highsmith, J. (2001). Manifiesto por el Desarrollo Ágil de Software. http://agilemanifesto.org. Página vigente al 20/02/2017.

Hofmann, M. (2003). The development of a generic data mining life cycle (DMLC) (dissertation, School of Computing, Dublin Institute of Technology).

Hossian, A. (2012). Modelo de Proceso de Conceptualización de Requisitos. Tesis Doctoral en Ciencias informáticas. Facultad de Informática. Universidad Nacional de La Plata.

Hsu, W., Lee, M. , Zhang, J. (2002). Image mining: Trends and developments. Journal of Intelligent Information Systems, 19(1): 7-23.

Instituto Nacional de Estadística y Censos. (2011). ENPreCoSP: Encuesta Nacional sobre Prevalencia de Consumo de Sustancias Psicoactivas 2011. Ministerio de Salud Argentina. http://www.indec.gov.ar/nivel4_default.asp?id_tema_1=4&id_tema_2=32&id_tema_3=67. Vigente al 12/12/2016.

Instituto Nacional de Estadística y Censos. (2016). Página web http://www.indec.gov.ar/bases-de-datos.asp?solapa=2. Vigente al 20/12/2016.

ISO, I. (2012). 21500: 2012: Guidance on Project Management. International Organization for Standardization.

Jurney, R. (2017). Agile Data Science 2.0: Building Full-stack Data Analytics Applications with Spark. " O'Reilly Media, Inc.".

Kanungo, S. (2005). Using Process Theory to Analyze Direct and Indirect Value-Drivers of Information Systems. Proceedings of the 38th Annual Hawaii International Conference on System Sciences. Pág. 231-240.

Kdnuggets. (2002). What main methodology are you using for data mining? Encuesta (Junio 2002). http://www.kdnuggets.com/polls/2002/methodology.htm (Último acceso 04/03/2017).

Kdnuggets. (2004). Data Mining Methodology. Encuesta (Abril 2004). http://www.kdnuggets.com/polls/2004/data_mining_methodology.htm (Último acceso 04/03/2017).

Kdnuggets. (2007). Data Mining Methodology. Encuesta (Agosto 2007). http://www.kdnuggets.com/polls/2007/data_mining_methodology.htm (Último acceso 04/03/2017).

Kdnuggets. (2014). What main methodology are you using for your analytics, data mining, or data science projects? Poll (Oct 2014). https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html (Último acceso 04/03/2017).

Kerzner, H. (2013). Project management: a systems approach to planning, scheduling, and controlling. Eleventh Edition. John Wiley & Sons.

Klosgen, W., & Zytkow, J. M. (2002). The knowledge discovery process. In Klosgen, W., & Zytkow, J. M. (Eds.), Handbook of data mining and knowledge discovery (pp. 10–21). New York, NY: Oxford University Press.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Ijcai (Vol. 14, No. 2, pp. 1137-1145).

Kopanakis, I., & Theodoulidis, B. (1999). Visual data mining & modeling techniques. Centre of Research in Information Management (CRIM). UK: Department of Computation. University of Manchester Institute of Science and Technology.

Kosala, R., Blockeel, H. (2000). Web mining research: A survey. ACM SIGKDD Explorations Newsletter, 2(1): 1-15.

Kruse, R., & Borgelt, C. (2003). Information mining. International Journal of Approximate Reasoning, 32(2), 63-66.

Kuna, H. (2013). Procedimientos de Explotación de Información para la Identificación de Datos Faltantes con Ruido e Inconsistentes. Tesis Doctoral en Ingeniería de Sistemas y Computación. Departamento de Lenguajes y Ciencias de la Computación. Escuela Técnica Superior de Ingeniería Informática. Universidad de Málaga.

Kurgan, L. A., & Musilek, P. (2006). A survey of Knowledge Discovery and Data Mining process models. The Knowledge Engineering Review, 21(01), 1-24.

Langseth, J., Vivatrat, N. (2003). Why Proactive Business Intelligence is a Hallmark of the Real-Time Enterprise: Outward Bound. Intelligent Enterprise 518: 34-41.

Li, T., & Ruan, D. (2007). An extended process model of knowledge discovery in database. Journal of Enterprise Information Management, 20(2), 169–177. doi:10.1108/17410390710725751

Lyman, P., Varain, H. (2003). How Much Information?. School of Information Management & Systems. University of California Berkeley. http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/. Ultimo acceso 30/03/2017.

Maes, A., & Poels, G. (2006). Evaluating quality of conceptual models based on user perceptions. In International Conference on Conceptual Modeling (pp. 54-67). Springer, Berlin, Heidelberg.

Maimon, O. y Rokach, L. (Eds.). (2005). Data mining and knowledge discovery handbook. Springer.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity.

Marbán, O., Mariscal, G., Menasalvas, E., Segovia, J. (2007). An Engineering Approach to Data Mining Projects. Lecture Notes in Computer Science, 4881: 578-588. Springer.´

Marbán, O., Menasalvas, E., & Fernández-Baizán, C. (2008). A cost model to estimate the effort of data mining projects (DMCoMo). Information Systems, 33(1), 133-150.

Marbán, O., Segovia, J., Menasalvas, E., & Fernández-Baizán, C. (2009). Toward data mining engineering: A software engineering approach. Information systems, 34(1), 87-107.

Mariscal, G., Marbán, Ó., & Fernández, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. The Knowledge Engineering Review, 25(02), 137-166.

Martins, S., Rodríguez, D., García-Martínez, R. (2014). Deriving Processes of Information Mining Based on Semantic Nets and Frames. Lecture Notes on Artificial Intelligence, 8482: 150-159. ISSN 0302-9743.

McConnell, S. (1997). Desarrollo y gestión de proyectos informáticos. McGraw-Hill Interamericana de España. ISBN: 978-8448112295.

Microsoft. (2016). Team Data Science Process. https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/ (último acceso: 21/05/2018)

Mochal, T. (2006). Use This Process to Estimate Effort Hours. TechRepublic. December 11, 2006. http://www.techrepublic.com/article/use-this-process-to-estimate-effort-hours/ Página Vigente al 25/01/2017.

Moine, J. (2013). Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo [Tesis de Maestría]. Argentina: Universidad Nacional de la Plata.

Moss, L. (2003). Nontechnical Infrastructure of BI Applications. DM Review 131: 42-45.

Moyle, S. & Jorge, A. (2001). Ramsys—a methodology for supporting rapid remote collaborative data mining projects, ECML/PKDD 2001 Workshop on Integrating Aspects of Data Mining, Decision Support and Meta-Learning: Internal SolEuNet Session, 20–31.

Munir, H., Wnuk, K., & Runeson, P. (2016). Open innovation in software engineering: a systematic mapping study. Empirical Software Engineering, 21(2), 684-723.

Negash, S., Gray, P. (2008). Business Intelligence. En Handbook on Decision Support Systems 2, ed. F. Burstein y C. Holsapple Heidelberg, Springer, Pág. 175-193.

Petersen K, Feldt R, Mujtaba S, Mattsson M (2008) Systematic mapping studies in software engineering. In: 12th International Conference on Evaluation and Assessment in Software Engineering, vol 17,p1.

Poels, G., Maes, A., Gailly, F., & Paemeleire, R. (2005). Measuring the perceived semantic quality of information models. In International Conference on Conceptual Modeling (pp. 376-385). Springer, Berlin, Heidelberg.

Pressman, R. S. (2005). Software engineering: a practitioner's approach. Palgrave Macmillan.

Project Management Institute, Inc. (2013a). A guide to the project management body of knowledge (PMBOK® guide). Fifth edition. ISBN: 978-1-935589-67-9.

Project Management Institute, Inc. (2013b). The high cost of low performance: the essential role of communications. Mayo 2013. http://www.pmi.org/-/media/pmi/documents/public/pdf/learning/thought-leadership/pulse/the-essential-role-of-communications.pdf. Página vigente al 12/01/2017.

Pyle, D. (2003). Business modeling and data mining. Morgan Kaufmann.

Pytel, P., Hossian, A., Britos, P., García-Martínez, R. (2015). Feasibility and Effort Estimation Models for Medium and Small Size Information Mining Projects. Information Systems Journal, 47: 01-14. Elsevier. ISSN 0306-4379.

Reinartz, T. (1999). Focusing solutions for data mining. Lecture notes in artificial intelligence, 1623.

Rennolls, K., & AL-Shawabkeh, A. (2008). Formal structures for data mining, knowledge discovery and communication in a knowledge management environment. Intelligent Data Analysis, 12, 147–163.

Ridge, E. (2014). Guerrilla Analytics: A Practical Approach to Working with Data. Morgan Kaufmann.

Riveros, H. y Rosas, L. (1985). ―EMétodo Científico Aplicado a las Ciencias Experimentales". México: Editorial Trillas. ISBN 96-8243-893-4.

Rodríguez, D. (2015). Conceptualización de Espacios Virtuales de Trabajo. Tesis Doctoral en Ciencias informáticas. Facultad de Informática. Universidad Nacional de La Plata.

Rodríguez, D., Pollo Cattaneo, M. F., Britos, P. V., García Martínez, R. (2010). Estimación Empírica de Carga de Trabajo en Proyectos de Explotación de Información. In XVI Congreso Argentino de Ciencias de la Computación.

Rollins, J. B. (2015). Foundational methodology for data science. White Paper. IBM Analytics.

Rubin, K. S. (2012). Essential Scrum: A practical guide to the most popular agile process. Addison-Wesley.

Rudin, K., Cressy, D. (2003). Will the Real Analytic Application Please Stand Up? DM Review 133: 30-34.

Runeson, P., Host, M., Rainer, A., & Regnell, B. (2012). Case study research in software engineering: Guidelines and examples. John Wiley & Sons.

Sabato J, Mackenzie M. (1982). ―LaProducción de Tecnología: Autónoma o Transnacional". Instituto Latinoamericano de Estudios Transnacionales - Technology & Engineering. ISBN 9789684293489.

Saltz, J., & Crowston, K. (2017). Comparing data science project management methodologies via a controlled experiment. In Proceedings of the 50th Hawaii International Conference on System Sciences.

SAS Institute Inc. (1997). Data mining and the case for sampling (white paper).

Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge University Press.

Sharma, S. (2008). An integrated knowledge discovery and data mining process model (Doctoral dissertation).

Sharma, S., & Osei-Bryson, K. M. (2009). Framework for formal implementation of the business understanding phase of data mining projects. Expert Systems with Applications, 36(2), 4114-4124.

Sharma, S., Osei-Bryson, K. M., & Kasper, G. M. (2012). Evaluation of an integrated Knowledge Discovery and Data Mining process model. Expert Systems with Applications, 39(13), 11335-11348.

Siddiqi, N. (2012). Credit risk scorecards: developing and implementing intelligent credit scoring (Vol. 3). John Wiley & Sons.

Simon, H. A. (1996). The sciences of the artificial. MIT press.

Solarte, J. (2002). A Proposed Data Mining Methodology and Its Application to Industrial Engineering, Master's thesis, University of Tennessee, Knoxville.

Sommerville, I. (2011). Software engineering ninth edition. Pearson Education, Inc., publishing as Addison-Wesley. ISBN: 978-0-13-703515-1.

Srivastava, J., Cooley, R., Deshpande, M., Tan, P. (2000). Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. SIGKDD Explorations, 12: 12-23.

Tan, A. (1999). Text mining: The state of the art and the challenges. In Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases. pp. 65-70.

Thomsen, E. (2003). BI's Promised Land. Intelligent Enterprise, 64: 21-25.

Two Crows Corporation. (1998). Introduction to Data Mining and Knowledge Discovery, 2nd edition. Two Crows Corporation. ISBN 892095-00-0.

Vanrell, J. Á., Bertone, R. A., & García Martínez, R. (2010). Un modelo de procesos de explotación de la información. In XII Workshop de Investigadores en Ciencias de la Computación.

Vanrell, J., Bertone, R., & García-Martínez, R. (2012). Un Modelo de Procesos para Proyectos de Explotación de Información. In Proceedings Latin American Congress on Requirements Engineering and Software Testing. Pág (pp. 46-52).

Verzuh, E. (2015). The fast forward MBA in project management. John Wiley & Sons.

Vuori, V. (2006). The Employees as a Source of External Business Information. Proceedings European Productivity Conference EPC´06. Pág. 29-36.

Wieringa, R., Maiden, N., Mead, N., & Rolland, C. (2006). Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. Requirements Engineering, 11(1), 102-107.

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In Proceedings of the 18th international conference on evaluation and assessment in software engineering (p. 38). ACM.

Ye, N. (2003). The handbook of data mining (Vol. 24). Mahwah, NJ/London: Lawrence Erlbaum Associates, Publishers.

ANEXO A: Documentación Casos de Validación

En este capítulo se presentan los versionados iniciales e intermedios de los proyectos desarrollados, los cuales por claridad y simplicidad no fueron introducidos en la sección correspondiente a cada caso. En la sección A.1, se presentan los formalismos de la prueba de concepto "ENPreCoSP-2011", en la sección A.2 aquellos asociados al primer caso de validación "WEB LOG" y en la sección A.3 las versiones del segundo caso de validación "EDUCACIÓN SUPERIOR".

A.1. VERSIONADO PRUEBA DE CONCEPTO: ENPreCoSP-2011

En esta sección se presentan los versionados de los formalismos del proyecto ENPreCoSP-2011. Estos corresponden al subproceso Gestión y se listan a continuación: Plan de acción versión 1.1 y 1.2 (tablas A.1 (a y b) y A.2 (a y b), respectivamente), Diagrama de Gantt actualizado al cierre del proyecto (figura A.1), Plan de Necesidad de Recursos versión 1.0 (tabla A.3) y los Registros de Mediciones versión 1.0 y 1.1 (tablas A.4 y A.5).

Plan de Acción								
Responsable:		Rodriguez H.			**Fecha:**		06/05/16	
ID#:		G.PI.PIA.PIAc			**Versión:**		1.1	
ID Actividad	**Actividad**	**Inicio Estimado**	**Inicio Real**	**Fin Estimado**	**Fin Real**	**Esfuerzo Estimado (Hs)**	**Esfuerzo Real (Hs)**	**Comentarios**
G.In	**Iniciación**	**04/04/16**	**04/04/16**	**22/04/16**	**22/04/16**	**12**	**11**	
G.In.EIP	Exploración Inicial del Proyecto	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DeC	Definición de la Comunicación	04/04/16	04/04/16	18/04/16	18/04/16	2	2	
G.In.EvS	Evaluación de la Situación	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DCV	Definición del Ciclo de Vida	19/04/16	19/04/16	22/04/16	22/04/16	2	1	
G.PI	**Planificación**	**04/04/16**	**04/04/16**	**29/04/16**	**29/04/16**	**10**	**9**	
G.PI.PIM	Planificación de la Mediciones	04/04/16	04/04/16	29/04/16	29/04/16	2	1	
G.PI.PIA	Planificación de las Actividades	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.PI.PIR	Planificación de los Recursos	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.PI.PRe	Planificación de las Responsabilidades	04/04/16	04/04/16	29/04/16	29/04/16	4	4	

Tabla A.1.a. Prueba de Concepto - Plan de Acción versión 1.1

G.So	**Soporte**	**20/04/16**	**20/04/16**	**08/06/16**		**8**	**3**	
G.So.MeP	Mediciones del Proyecto	20/04/16	20/04/16	08/06/16		4	2	Se prevé el registro de las métricas de manera mensual
G.So.GeC	Gestión de la Configuración	20/04/16	20/04/16	08/06/16		4	1	
G.Co	**Control**	**20/04/16**	**20/04/16**	**06/06/16**		**12**	**4**	
G.Co.GeD	Gestión del Desarrollo	20/04/16	20/04/16	06/06/16		4	2	Se prevé la aplicación del reporte de estado de manera mensual
G.Co.CoA	Control de las Actividades	20/04/16	20/04/16	06/06/16		6	2	
G.Co.Gca	Gestión del Cambio	20/04/16	20/04/16	06/06/16		2	0	
G.Ci	**Cierre**	**06/06/16**		**08/06/16**		**4**		
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	06/06/16		06/06/16		2		
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	07/06/16		08/06/16		2		
D.EN	**Entendimiento del Negocio**	**04/04/16**	**04/04/16**	**22/04/16**	**22/04/16**	**44**	**32**	
D.EN.AnN	Análisis del Negocio	04/04/16	04/04/16	20/04/16	20/04/16	28	22	
D.EN.CPN	Comprensión del Problema de Negocio	18/04/16	18/04/16	22/04/16	22/04/16	16	10	
D.ED	**Entendimiento de los Datos**	**25/04/16**	**25/04/16**	**06/05/16**	**06/05/16**	**56**	**52**	
D.ED.AnD	Análisis de los Datos	25/04/16	25/04/16	29/04/16	29/04/16	22	20	
D.ED.ExD	Exploración de los Datos	28/04/16	28/04/16	06/05/16	06/05/16	22	22	
D.ED.EvD	Evaluación de los Datos	02/05/16	02/05/16	06/05/16	06/05/16	12	10	
D.Mo	**Modelado**	**09/05/16**		**13/05/16**		**26**		
D.Mo.MoP	Modelado del problema	09/05/16		10/05/16		14		
D.Mo.CoM	Configuración del Modelo	11/05/16		13/05/16		12		
D.PD	**Preparación de los Datos**	**16/05/16**		**20/05/16**		**36**		
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	16/05/16		18/05/16		20		
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos	17/05/16		20/05/16		16		
D.Im	**Implementación**	**23/05/16**		**27/05/16**		**30**		
D.Im.SeM	Selección del Modelo	23/05/16		24/05/16		8		
D.Im.ExI	Explotación de Información	24/05/16		27/05/16		22		
D.EP	**Evaluación y Presentación**	**30/05/16**		**06/06/16**		**26**		
D.EP.EvR	Evaluación de los Resultados	30/05/16		02/06/16		10		
D.EP.PrR	Presentación de los Resultados	01/06/16		06/06/16		16		

Tabla A.1.b. Prueba de Concepto - Plan de Acción versión 1.1 (continuación)

Plan de Acción								
Responsable:		Rodriguez H.			**Fecha:**		03/06/16	
ID#:		G.PI.PIA.PIAc			**Versión:**		1.2	
ID Actividad	**Actividad**	**Inicio Estimado**	**Inicio Real**	**Fin Estimado**	**Fin Real**	**Esfuerzo Estimado (Hs)**	**Esfuerzo Real (Hs)**	**Comentarios**
G.In	**Iniciación**	**04/04/16**	**04/04/16**	**22/04/16**	**22/04/16**	**12**	**11**	
G.In.EIP	Exploración Inicial del Proyecto	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DeC	Definición de la Comunicación	04/04/16	04/04/16	18/04/16	18/04/16	2	2	
G.In.EvS	Evaluación de la Situación	04/04/16	04/04/16	18/04/16	18/04/16	4	4	
G.In.DCV	Definición del Ciclo de Vida	19/04/16	19/04/16	22/04/16	22/04/16	2	1	
G.Pl	**Planificación**	**04/04/16**	**04/04/16**	**29/04/16**	**29/04/16**	**10**	**9**	
G.Pl.PlM	Planificación de la Mediciones	04/04/16	04/04/16	29/04/16	29/04/16	2	1	
G.Pl.PlA	Planificación de las Actividades	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.Pl.PlR	Planificación de los Recursos	04/04/16	04/04/16	29/04/16	29/04/16	2	2	
G.Pl.PRe	Planificación de las Responsabilidades	04/04/16	04/04/16	29/04/16	29/04/16	4	4	
G.So	**Soporte**	**20/04/16**	**20/04/16**	**15/06/16**		**8**	**9**	
G.So.MeP	Mediciones del Proyecto	20/04/16	20/04/16	15/06/16		4	4	Se prevé el registro de las métricas de manera mensual
G.So.GeC	Gestión de la Configuración	20/04/16	20/04/16	15/06/16		4	5	
G.Co	**Control**	**20/04/16**	**20/04/16**	**13/06/16**		**12**	**15**	
G.Co.GeD	Gestión del Desarrollo	20/04/16	20/04/16	13/06/16		4	6	Se prevé la aplicación del reporte de estado de manera mensual
G.Co.CoA	Control de las Actividades	20/04/16	20/04/16	13/06/16		6	8	
G.Co.Gca	Gestión del Cambio	20/04/16	20/04/16	13/06/16		2	1	
G.Ci	**Cierre**	**06/06/16**	**06/06/16**	**15/06/16**		**4**		
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	06/06/16	06/06/16	13/06/16		2		
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	14/06/16		15/06/16		2		
D.EN	**Entendimiento del Negocio**	**04/04/16**	**04/04/16**	**22/04/16**	**20/04/16**	**44**	**32**	
D.EN.AnN	Análisis del Negocio	04/04/16	04/04/16	20/04/16	20/04/16	28	22	
D.EN.CPN	Comprensión del Problema de Negocio	18/04/16	18/04/16	22/04/16	20/04/16	16	10	
D.ED	**Entendimiento de los Datos**	**25/04/16**	**25/04/16**	**06/05/16**	**06/05/16**	**56**	**52**	
D.ED.AnD	Análisis de los Datos	25/04/16	25/04/16	29/04/16	29/04/16	22	20	
D.ED.ExD	Exploración de los Datos	28/04/16	28/04/16	06/05/16	06/05/16	22	22	
D.ED.EvD	Evaluación de los Datos	02/05/16	02/05/16	06/05/16	06/05/16	12	10	

Tabla A.2.a. Prueba de Concepto - Plan de Acción versión 1.2

D.Mo	**Modelado**	**09/05/16**	**09/05/16**	**13/05/16**	**13/05/16**	**26**	**16**	
D.Mo.MoP	Modelado del problema	09/05/16	09/05/16	10/05/16	10/05/16	14	8	
D.Mo.CoM	Configuración del Modelo	11/05/16	11/05/16	13/05/16	13/05/16	12	8	
D.PD	**Preparación de los Datos**	**16/05/16**	**16/05/16**	**20/05/16**	**20/05/16**	**36**	**28**	
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	16/05/16	16/05/16	18/05/16	18/05/16	20	16	
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos	17/05/16	17/05/16	20/05/16	20/05/16	16	12	
D.Im	**Implementación**	**23/05/16**	**23/05/16**	**27/05/16**	**27/05/16**	**30**	**26**	
D.Im.SeM	Selección del Modelo	23/05/16	23/05/16	24/05/16	24/05/16	8	6	
D.Im.ExI	Explotación de Información	24/05/16	24/05/16	27/05/16	27/05/16	22	20	
D.EP	**Evaluación y Presentación**	**30/05/16**	**30/05/16**	**13/06/16**		**26**	**18**	
D.EP.EvR	Evaluación de los Resultados	30/05/16	30/05/16	02/06/16	02/06/16	10	6	
D.EP.PrR	Presentación de los Resultados	01/06/16	01/06/16	13/06/16		16	12	

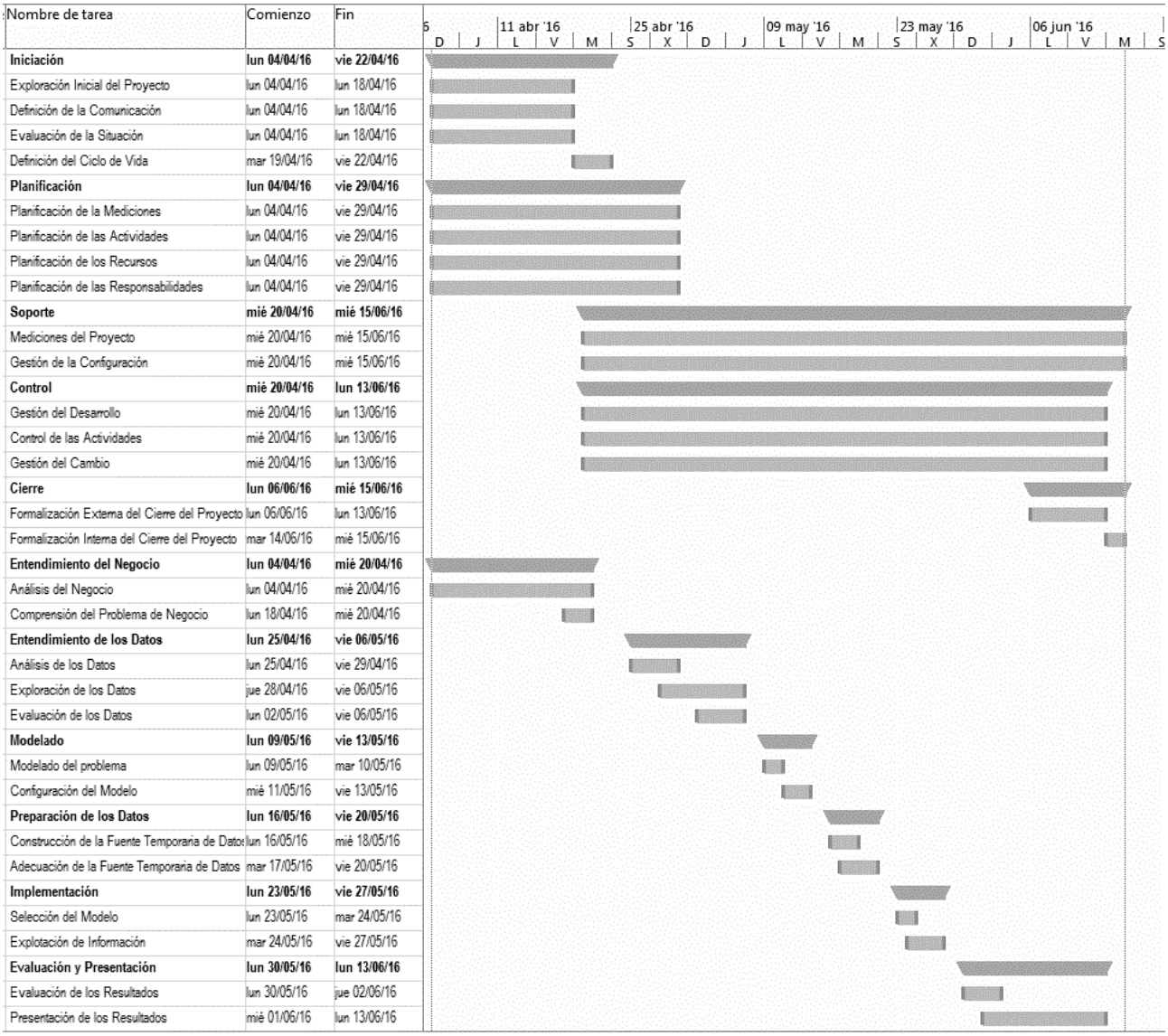
Tabla A.2.b. Prueba de Concepto - Plan de Acción versión 1.2 (continuación)

Plan de Necesidad de Recursos					
Responsable:	Rodriguez H.			**Fecha:**	29/04/16
ID#:	G.Pl.PIR.PINR			**Versión:**	1.0
Recursos Humanos					
ID	**Recurso**	**Cantidad**	**F. Inicio**	**F. Fin**	**Descripción**
rhr.1	Líder de Proyecto	1	04/04/16	08/06/16	
rhr.2	Ingeniero de Explotación de Información Junior	1	04/04/16	08/06/16	
Recursos Materiales					
ID	**Recurso**	**Cantidad**	**F. Inicio**	**F. Fin**	**Descripción**
rmr.1	Computadora Personal	2	04/04/16	08/06/16	SO windows (7 en adelante) RAM 4 gb o más 10GB o más espacio en disco

Tabla A.3. Prueba de Concepto - Plan de Necesidad de Recursos (versión 1.0)

Registro de Mediciones			
Responsable:	Esposito E.	**Fecha:**	06/05/16
ID#:	G.So.MeP.ReMe	**Versión:**	1.0
Indicador	**Medición**	**Descripción**	
Tiempo total requerido para el desarrollo del proyecto	**DRPY** = 111	Tdesarrollo = 84 Tgestion = 27	
Grado de Utilidad de Atributos	**GUA** = 6,71	NA = 392 NASE = 15 NAUD= 2 NO_UTILES = 275	

Tabla A.4. Prueba de Concepto – Registro de Mediciones (versión 1.0)



Nombre de tarea	Comienzo	Fin
Iniciación	**lun 04/04/16**	**vie 22/04/16**
Exploración Inicial del Proyecto	lun 04/04/16	lun 18/04/16
Definición de la Comunicación	lun 04/04/16	lun 18/04/16
Evaluación de la Situación	lun 04/04/16	lun 18/04/16
Definición del Ciclo de Vida	mar 19/04/16	vie 22/04/16
Planificación	**lun 04/04/16**	**vie 29/04/16**
Planificación de la Mediciones	lun 04/04/16	vie 29/04/16
Planificación de las Actividades	lun 04/04/16	vie 29/04/16
Planificación de los Recursos	lun 04/04/16	vie 29/04/16
Planificación de las Responsabilidades	lun 04/04/16	vie 29/04/16
Soporte	**mié 20/04/16**	**mié 15/06/16**
Mediciones del Proyecto	mié 20/04/16	mié 15/06/16
Gestión de la Configuración	mié 20/04/16	mié 15/06/16
Control	**mié 20/04/16**	**lun 13/06/16**
Gestión del Desarrollo	mié 20/04/16	lun 13/06/16
Control de las Actividades	mié 20/04/16	lun 13/06/16
Gestión del Cambio	mié 20/04/16	lun 13/06/16
Cierre	**lun 06/06/16**	**mié 15/06/16**
Formalización Externa del Cierre del Proyecto	lun 06/06/16	lun 13/06/16
Formalización Interna del Cierre del Proyecto	mar 14/06/16	mié 15/06/16
Entendimiento del Negocio	**lun 04/04/16**	**mié 20/04/16**
Análisis del Negocio	lun 04/04/16	mié 20/04/16
Comprensión del Problema de Negocio	lun 18/04/16	mié 20/04/16
Entendimiento de los Datos	**lun 25/04/16**	**vie 06/05/16**
Análisis de los Datos	lun 25/04/16	vie 29/04/16
Exploración de los Datos	jue 28/04/16	vie 06/05/16
Evaluación de los Datos	lun 02/05/16	vie 06/05/16
Modelado	**lun 09/05/16**	**vie 13/05/16**
Modelado del problema	lun 09/05/16	mar 10/05/16
Configuración del Modelo	mié 11/05/16	vie 13/05/16
Preparación de los Datos	**lun 16/05/16**	**vie 20/05/16**
Construcción de la Fuente Temporaria de Datos	lun 16/05/16	mié 18/05/16
Adecuación de la Fuente Temporaria de Datos	mar 17/05/16	vie 20/05/16
Implementación	**lun 23/05/16**	**vie 27/05/16**
Selección del Modelo	lun 23/05/16	mar 24/05/16
Explotación de Información	mar 24/05/16	vie 27/05/16
Evaluación y Presentación	**lun 30/05/16**	**lun 13/06/16**
Evaluación de los Resultados	lun 30/05/16	jue 02/06/16
Presentación de los Resultados	mié 01/06/16	lun 13/06/16

Figura A.1. Prueba de Concepto – Diagrama Gantt (versión final)

Registro de Mediciones			
Responsable:	Esposito E.	**Fecha:**	03/06/16
ID#:	G.So.MeP.ReMe	**Versión:**	1.1

Indicador	**Medición**	**Descripción**
Tiempo total requerido para el desarrollo del proyecto	**DRPY** = 216	Tdesarrollo = 172 Tgestion = 44
Grado de Utilidad de Atributos	**GUA** = 6,71	NA = 392 NASE = 15 NAUD= 2 NO_UTILES = 275

Tabla A.5. Prueba de Concepto – Registro de Mediciones (versión 1.1)

A.2. VERSIONADO CASO DE VALIDACIÓN: WEB LOG

En esta sección se presentan los versionados de los formalismos del primer caso de validación, separados según el subproceso al que pertenecen: Gestión (sección A.2.1) y Desarrollo (sección A.2.2).

A.2.1. Subproceso: Gestión

Durante el desarrollo de las actividades pertenecientes al subproceso gestión, se registraron las siguientes actualizaciones en las versiones de los productos intermedios: Plan de Acción versión 1.0 y 1.1 (tablas A.6 (a y b) y A.7 (a y b), respectivamente), Diagrama de Gantt actualizado al cierre del proyecto (figura A.2) y Registro de Mediciones versión 1.0 (tabla A.8).

Plan de Acción								
Responsable:	Sebastian M.			**Fecha:**		24/02/2017		
ID#:	G.PI.PIA.PIAc			**Versión:**		1.0		
ID Actividad	**Actividad**	**Inicio Estimado**	**Inicio Real**	**Fin Estimado**	**Fin Real**	**Esfuerzo Estimado (Hs)**	**Esfuerzo Real (Hs)**	**Comentarios**
G.In	**Iniciación**	**06/02/17**		**07/04/17**		**16**		
G.In.EIP	Exploración Inicial del Proyecto	06/02/17		17/02/17		4		
G.In.DeC	Definición de la Comunicación	06/02/17		07/04/17		4		
G.In.EvS	Evaluación de la Situación	06/02/17		07/04/17		6		
G.In.DCV	Definición del Ciclo de Vida	20/02/17		22/02/17		2		
G.PI	**Planificación**	**06/02/17**		**14/04/17**		**16**		
G.PI.PIM	Planificación de la Mediciones	06/02/17		24/02/17		2		
G.PI.PIA	Planificación de las Actividades	06/02/17		14/04/17		4		
G.PI.PIR	Planificación de los Recursos	06/02/17		14/04/17		4		
G.PI.PRe	Planificación de las Responsabilidades	06/02/17		14/04/17		6		
G.So	**Soporte**	**06/02/17**		**11/05/17**		**18**		
G.So.MeP	Mediciones del Proyecto	06/02/17		11/05/17		8		Se prevé el registro de mediciones al fin de cada iteración
G.So.GeC	Gestión de la Configuración	06/02/17		11/05/17		10		
G.Co	**Control**	**06/02/17**		**11/05/17**		**22**		
G.Co.GeD	Gestión del Desarrollo	06/02/17		11/05/17		8		Se prevé la aplicación del reporte de estado al fin de cada iteración
G.Co.CoA	Control de las Actividades	06/02/17		11/05/17		10		
G.Co.Gca	Gestión del Cambio	13/02/17		11/05/17		4		
G.Ci	**Cierre**	**09/05/17**		**11/05/17**		**4**		
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	09/05/17		10/05/17		2		
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	09/05/17		11/05/17		2		

Tabla A.6.a. Caso de Validación: Web Log - Plan de Acción versión 1.0

Iteración 1		**06/02/17**		**27/03/17**		**270**		
D.EN	**Entendimiento del Negocio**	**06/02/17**		**24/02/17**		**50**		
D.EN.AnN	Análisis del Negocio	06/02/17		20/02/17		30		
D.EN.CPN	Comprensión del Problema de Negocio	14/02/17		24/02/17		20		
D.ED	**Entendimiento de los Datos**	**22/02/17**		**06/03/17**		**68**		
D.ED.AnD	Análisis de los Datos	22/02/17		01/03/17		30		
D.ED.ExD	Exploración de los Datos	27/02/17		03/03/17		30		
D.ED.EvD	Evaluación de los Datos	01/03/17		06/03/17		8		
D.Mo	**Modelado**	**06/03/17**		**15/03/17**		**42**		
D.Mo.MoP	Modelado del problema	06/03/17		13/03/17		24		
D.Mo.CoM	Configuración del Modelo	10/03/17		15/03/17		18		
D.PD	**Preparación de los Datos**	**13/03/17**		**20/03/17**		**40**		
D.PD.AFI	Construcción de la Fuente Temporaria de Datos	13/03/17		17/03/17		30		
D.PD.CFT	Adecuación de la Fuente Temporaria de Datos	15/03/17		20/03/17		10		
D.Im	**Implementación**	**17/03/17**		**27/03/17**		**56**		
D.Im.SeM	Selección del Modelo	17/03/17		21/03/17		16		
D.Im.ExI	Explotación de Información	21/03/17		27/03/17		40		
D.EP	**Evaluación y Presentación**	**24/03/17**		**27/03/17**		**14**		
D.EP.EvR	Evaluación de los Resultados	24/03/17		27/03/17		14		
D.EP.PrR	Presentación de los Resultados	-		-		0		
Iteración 2		**28/03/17**		**09/05/17**		**184**		
D.EN	**Entendimiento del Negocio**	**28/03/17**		**07/04/17**		**26**		
D.EN.AnN	Análisis del Negocio	28/03/17		03/04/17		10		
D.EN.CPN	Comprensión del Problema de Negocio	03/04/17		07/04/17		16		
D.ED	**Entendimiento de los Datos**	**03/04/17**		**14/04/17**		**40**		
D.ED.AnD	Análisis de los Datos	03/04/17		10/04/17		16		
D.ED.ExD	Exploración de los Datos	03/04/17		12/04/17		20		
D.ED.EvD	Evaluación de los Datos	12/04/17		14/04/17		4		
D.Mo	**Modelado**	**14/04/17**		**20/04/17**		**30**		
D.Mo.MoP	Modelado del problema	14/04/17		18/04/17		18		
D.Mo.CoM	Configuración del Modelo	17/04/17		20/04/17		12		
D.PD	**Preparación de los Datos**	**18/04/17**		**24/04/17**		**26**		
D.PD.AFI	Construcción de la Fuente Temporaria de Datos	18/04/17		21/04/17		18		
D.PD.CFT	Adecuación de la Fuente Temporaria de Datos	21/04/17		24/04/17		8		
D.Im	**Implementación**	**24/04/17**		**02/05/17**		**28**		
D.Im.SeM	Selección del Modelo	24/04/17		26/04/17		8		
D.Im.ExI	Explotación de Información	26/04/17		02/05/17		20		
D.EP	**Evaluación y Presentación**	**02/05/17**		**09/05/17**		**34**		
D.EP.EvR	Evaluación de los Resultados	02/05/17		05/05/17		14		
D.EP.PrR	Presentación de los Resultados	05/05/17		09/05/17		20		

Tabla A.6.b. Caso de Validación: Web Log - Plan de Acción versión 1.0 (continuación)

Plan de Acción								
Responsable:	Sebastian M.			**Fecha:**		28/03/2017		
ID#:	G.PI.PIA.PIAc			**Versión:**		1.1		
ID Actividad	**Actividad**	**Inicio Estimado**	**Inicio Real**	**Fin Estimado**	**Fin Real**	**Esfuerzo Estimado (Hs)**	**Esfuerzo Real (Hs)**	**Comentarios**
G.In	**Iniciación**	**06/02/17**	**06/02/17**	**07/04/17**		**16**	**16**	
G.In.EIP	Exploración Inicial del Proyecto	06/02/17	06/02/17	17/02/17	17/02/17	4	4	
G.In.DeC	Definición de la Comunicación	06/02/17	06/02/17	07/04/17		4	3	
G.In.EvS	Evaluación de la Situación	06/02/17	06/02/17	07/04/17		6	5	
G.In.DCV	Definición del Ciclo de Vida	20/02/17	20/02/17	22/02/17	22/02/17	2	4	
G.PI	**Planificación**	**06/02/17**	**06/02/17**	**14/04/17**		**16**	**19**	
G.PI.PIM	Planificación de la Mediciones	06/02/17	06/02/17	24/02/17	24/02/17	2	4	
G.PI.PIA	Planificación de las Actividades	06/02/17	06/02/17	14/04/17		4	5	
G.PI.PIR	Planificación de los Recursos	06/02/17	06/02/17	14/04/17		4	2	
G.PI.PRe	Planificación de las Responsabilidades	06/02/17	06/02/17	14/04/17		6	8	
G.So	**Soporte**	**06/02/17**	**06/02/17**	**11/05/17**		**18**	**10**	
G.So.MeP	Mediciones del Proyecto	06/02/17	06/02/17	11/05/17		8	5	Se prevé el registro de mediciones al fin de cada iteración
G.So.GeC	Gestión de la Configuración	06/02/17	06/02/17	11/05/17		10	5	
G.Co	**Control**	**06/02/17**	**06/02/17**	**11/05/17**		**22**	**7**	
G.Co.GeD	Gestión del Desarrollo	06/02/17	06/02/17	11/05/17		8	4	Se prevé la aplicación del reporte de estado al fin de cada iteración
G.Co.CoA	Control de las Actividades	06/02/17	06/02/17	11/05/17		10	3	
G.Co.Gca	Gestión del Cambio	13/02/17	13/02/17	11/05/17		4	0	
G.Ci	**Cierre**	**09/05/17**		**11/05/17**		**4**		
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	09/05/17		10/05/17		2		
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	09/05/17		11/05/17		2		
	Iteración 1	**06/02/17**	**06/02/17**	**06/02/17**	**27/03/17**	**270**	**251**	
D.EN	**Entendimiento del Negocio**	**06/02/17**	**06/02/17**	**24/02/17**	**20/02/17**	**50**	**50**	
D.EN.AnN	Análisis del Negocio	06/02/17	06/02/17	20/02/17	20/02/17	30	30	
D.EN.CPN	Comprensión del Problema de Negocio	14/02/17	14/02/17	24/02/17	20/02/17	20	20	
D.ED	**Entendimiento de los Datos**	**22/02/17**	**22/02/17**	**06/03/17**	**06/03/17**	**68**	**70**	
D.ED.AnD	Análisis de los Datos	22/02/17	22/02/17	01/03/17	28/02/17	30	30	
D.ED.ExD	Exploración de los Datos	27/02/17	27/02/17	03/03/17	03/03/17	30	32	
D.ED.EvD	Evaluación de los Datos	01/03/17	01/03/17	06/03/17	06/03/17	8	8	
D.Mo	**Modelado**	**06/03/17**	**06/03/17**	**15/03/17**	**15/03/17**	**42**	**38**	
D.Mo.MoP	Modelado del problema	06/03/17	06/03/17	13/03/17	13/03/17	24	20	
D.Mo.CoM	Configuración del Modelo	10/03/17	10/03/17	15/03/17	15/03/17	18	18	
D.PD	**Preparación de los Datos**	**13/03/17**	**13/03/17**	**20/03/17**	**17/03/17**	**40**	**28**	
D.PD.AFI	Construcción de la Fuente Temporaria de Datos	13/03/17	13/03/17	17/03/17	17/03/17	30	20	
D.PD.CFT	Adecuación de la Fuente Temporaria de Datos	15/03/17	15/03/17	20/03/17	20/03/17	10	8	
D.Im	**Implementación**	**17/03/17**	**17/03/17**	**27/03/17**	**27/03/17**	**56**	**53**	
D.Im.SeM	Selección del Modelo	17/03/17	17/03/17	21/03/17	21/03/17	16	8	
D.Im.ExI	Explotación de Información	21/03/17	21/03/17	27/03/17	27/03/17	40	45	
D.EP	**Evaluación y Presentación**	**24/03/17**	**24/03/17**	**27/03/17**	**27/03/17**	**14**	**12**	
D.EP.EvR	Evaluación de los Resultados	24/03/17	24/03/17	27/03/17	27/03/17	14	12	
D.EP.PrR	Presentación de los Resultados	-	-	-	-	0		

Tabla A.7.a. Caso de Validación: Web Log - Plan de Acción versión 1.1

Iteración 2		**28/03/17**	**28/03/17**	**09/05/17**		**184**		
D.EN	**Entendimiento del Negocio**	**28/03/17**		**07/04/17**		**26**		
D.EN.AnN	Análisis del Negocio	28/03/17		03/04/17		10		
D.EN.CPN	Comprensión del Problema de Negocio	03/04/17		07/04/17		16		
D.ED	**Entendimiento de los Datos**	**03/04/17**		**14/04/17**		**40**		
D.ED.AnD	Análisis de los Datos	03/04/17		10/04/17		16		
D.ED.ExD	Exploración de los Datos	03/04/17		12/04/17		20		
D.ED.EvD	Evaluación de los Datos	12/04/17		14/04/17		4		
D.Mo	**Modelado**	**14/04/17**		**20/04/17**		**30**		
D.Mo.MoP	Modelado del problema	14/04/17		18/04/17		18		
D.Mo.CoM	Configuración del Modelo	17/04/17		20/04/17		12		
D.PD	**Preparación de los Datos**	**18/04/17**		**24/04/17**		**26**		
D.PD.AFI	Construcción de la Fuente Temporaria de Datos	18/04/17		21/04/17		18		
D.PD.CFT	Adecuación de la Fuente Temporaria de Datos	21/04/17		24/04/17		8		
D.Im	**Implementación**	**24/04/17**		**02/05/17**		**28**		
D.Im.SeM	Selección del Modelo	24/04/17		26/04/17		8		
D.Im.ExI	Explotación de Información	26/04/17		02/05/17		20		
D.EP	**Evaluación y Presentación**	**02/05/17**		**09/05/17**		**34**		
D.EP.EvR	Evaluación de los Resultados	02/05/17		05/05/17		14		
D.EP.PrR	Presentación de los Resultados	05/05/17		09/05/17		20		

Tabla A.7.b. Caso de Validación: Web Log - Plan de Acción versión 1.1 (continuación)

Registro de Mediciones			
Responsable:	Sebastian M.	**Fecha:**	28/03/2017
ID#:	G.So.MeP.ReMe	**Versión:**	1.0
Indicador	**Medición**	**Descripción**	
Tiempo total requerido para el desarrollo del proyecto	**DRPY** = 303hs	Tdesarrollo = 251 Tgestion = 52	
Tiempo medio requerido para el desarrollo de un problema de explotación de información	**DRPEI** = 125.5hs	Tdesarrollo = 251 NPEI = 2	
Número medio de atributos significativos por modelo	4	AtS.M(prne.1) = 4 AtS.M(prne.2) = 4 NMOD = 2	

Tabla A.8. Caso de Validación: Web Log - Registro de Mediciones (versión 1.0)

Nombre de tarea	Comienzo	Fin
Iniciación	**lun 06/02/17**	**vie 07/04/17**
Exploración Inicial del Proyecto	lun 06/02/17	vie 17/02/17
Definición de la Comunicación	lun 06/02/17	vie 07/04/17
Evaluación de la Situación	lun 06/02/17	vie 07/04/17
Definición del Ciclo de Vida	lun 20/02/17	mié 22/02/17
Planificación	**lun 06/02/17**	**vie 14/04/17**
Planificación de la Mediciones	lun 06/02/17	vie 24/02/17
Planificación de las Actividades	lun 06/02/17	vie 14/04/17
Planificación de los Recursos	lun 06/02/17	vie 14/04/17
Planificación de las Responsabilidades	lun 06/02/17	vie 14/04/17
Soporte	**lun 06/02/17**	**jue 11/05/17**
Mediciones del Proyecto	lun 06/02/17	jue 11/05/17
Gestión de la Configuración	lun 06/02/17	jue 11/05/17
Control	**lun 06/02/17**	**jue 11/05/17**
Gestión del Desarrollo	lun 06/02/17	jue 11/05/17
Control de las Actividades	lun 06/02/17	jue 11/05/17
Gestión del Cambio	lun 13/02/17	jue 11/05/17
Cierre	**mar 09/05/17**	**jue 11/05/17**
Formalización Externa del Cierre del Proyecto	mar 09/05/17	mié 10/05/17
Formalización Interna del Cierre del Proyecto	mar 09/05/17	jue 11/05/17
Iteración 1	**lun 06/02/17**	**lun 27/03/17**
Entendimiento del Negocio	**lun 06/02/17**	**lun 20/02/17**
Análisis del Negocio	lun 06/02/17	lun 20/02/17
Comprensión del Problema de Negocio	mar 14/02/17	lun 20/02/17
Entendimiento de los Datos	**mié 22/02/17**	**lun 06/03/17**
Análisis de los Datos	mié 22/02/17	mar 28/02/17
Exploración de los Datos	lun 27/02/17	vie 03/03/17
Evaluación de los Datos	mié 01/03/17	lun 06/03/17
Modelado	**lun 06/03/17**	**mié 15/03/17**
Modelado del problema	lun 06/03/17	lun 13/03/17
Configuración del Modelo	vie 10/03/17	mié 15/03/17
Preparación de los Datos	**lun 13/03/17**	**vie 17/03/17**
Construcción de la Fuente Temporaria de Datos	lun 13/03/17	vie 17/03/17
Adecuación de la Fuente Temporaria de Datos	mié 15/03/17	lun 20/03/17
Implementación	**vie 17/03/17**	**lun 27/03/17**
Selección del Modelo	vie 17/03/17	mar 21/03/17
Explotación de Información	mar 21/03/17	lun 27/03/17
Evaluación y Presentación	**vie 24/03/17**	**lun 27/03/17**
Evaluación de los Resultados	vie 24/03/17	lun 27/03/17
Presentación de los Resultados	-	-
Iteración 2	**mar 28/03/17**	**mar 09/05/17**
Entendimiento del Negocio	**mar 28/03/17**	**lun 03/04/17**
Análisis del Negocio	mar 28/03/17	jue 30/03/17
Comprensión del Problema de Negocio	vie 31/03/17	lun 03/04/17
Entendimiento de los Datos	**lun 03/04/17**	**jue 13/04/17**
Análisis de los Datos	lun 03/04/17	mié 05/04/17
Exploración de los Datos	lun 03/04/17	jue 13/04/17
Evaluación de los Datos	mié 12/04/17	jue 13/04/17
Modelado	**vie 14/04/17**	**mié 19/04/17**
Modelado del problema	vie 14/04/17	dom 16/04/17
Configuración del Modelo	lun 17/04/17	mié 19/04/17
Preparación de los Datos	**mar 18/04/17**	**lun 24/04/17**
Construcción de la Fuente Temporaria de Datos	mar 18/04/17	jue 20/04/17
Adecuación de la Fuente Temporaria de Datos	vie 21/04/17	lun 24/04/17
Implementación	**lun 24/04/17**	**vie 28/04/17**
Selección del Modelo	lun 24/04/17	mar 25/04/17
Explotación de Información	mar 25/04/17	vie 28/04/17
Evaluación y Presentación	**mar 02/05/17**	**mar 09/05/17**
Evaluación de los Resultados	mar 02/05/17	vie 05/05/17
Presentación de los Resultados	vie 05/05/17	mar 09/05/17

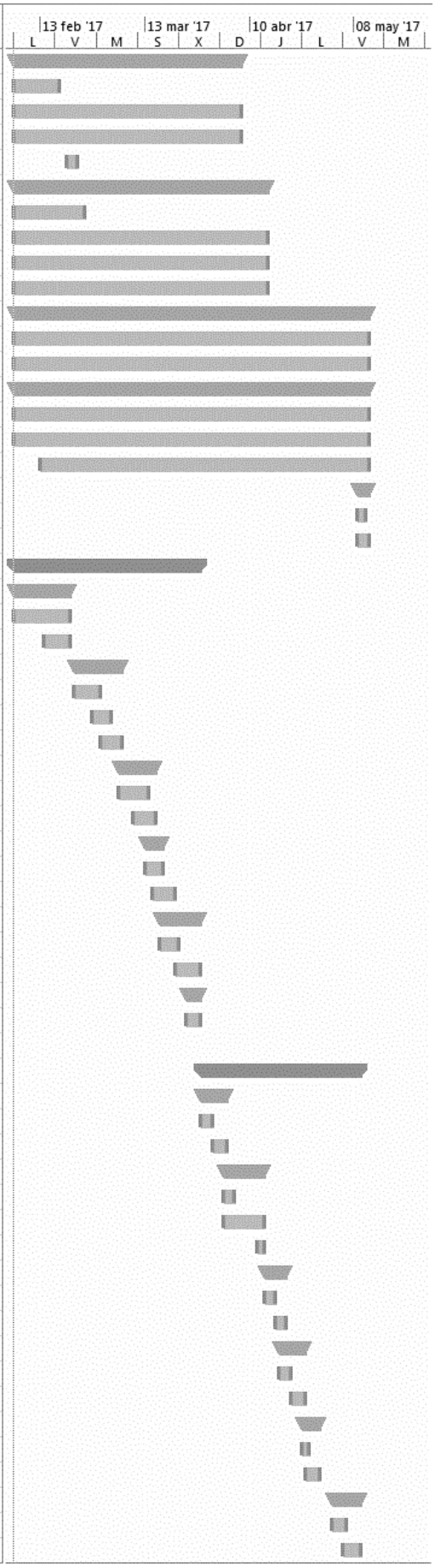


Figura A.2. Prueba de Concepto – Diagrama Gantt (versión final)

A.2.2. Subproceso: Desarrollo

Durante la implementación de las actividades pertenecientes al subproceso Desarrollo, se registraron las siguientes actualizaciones en las versiones de los productos intermedios: Restricciones del Proyecto versión 1.0 (tabla A.9), Problema del Negocio versión 1.0 (tabla A.10), Criterios de Éxito del Problema de Negocio versión 1.0 (tabla A.11) y Reporte de Evaluación de los Resultados versión 1.0 (tabla A.12).

Restricciones del Proyecto				
Responsable:	Sebastian M.	**Fecha:**	14/02/2017	
ID#:	D.EN.ANN.REPR	**Versión:**	1.0	
Restricción	**Tipo**	**Descripción**	**Objetivo asociado**	**Referencia**
repr.1	datos	La versión actual fue puesta en producción los primeros días de febrero del año 2016	(obpr.1) Optimizar la experiencia del usuario a partir del análisis de sus acciones en el sitio web, facilitando su uso y mejorando la disposición de los contenidos de interés	Entrevista 2

Tabla A.9. Caso de Validación: Web Log - Restricciones del Proyecto (versión 1.0)

Problema del Negocio			
Responsable:	Sebastian M.	**Fecha:**	20/02/2017
ID#:	D.EN.CPN.PRNE	**Versión:**	1.0
Objetivo del Proyecto	(obpr.1) Optimizar la experiencia del usuario a partir del análisis de sus acciones en el sitio web, facilitando su uso y mejorando la disposición de los contenidos de interés		
Problema	**Descripción**	**Experto**	**Referencia**
prne.1	Detallar las posibles rutas más frecuentes de navegación que realizan los usuarios	(rehi.3) Dario R.	Entrevista 3
prne.2	Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación	(rehi.3) Dario R.	Entrevista 3

Tabla A.10. Caso de Validación: Web Log - Problema del Negocio (versión 1.0)

Criterios de Éxito del Problema de Negocio			
Responsable:	Sebastian M.	**Fecha:**	20/02/2017
ID#:	D.EN.CPN.CEPN	**Versión:**	1.0
Criterio	**Descripción**	**Problema asociado**	**Referencia**
cepn.1	Las rutas de navegación frecuentes sean representativas de al menos un 15% del total de usuarios	(prne.1) Detallar las posibles rutas más frecuentes de navegación que realizan los usuarios	Entrevista 3
cepn.2	La caracterización de los perfiles tenga una tasa de error inferior al 20%.	(prne.2) Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación	Entrevista 3

Tabla A.11. Caso de Validación: Web Log - Criterios de Éxito del Problema de Negocio (versión 1.0)

Reporte de Evaluación de los Resultados			
Responsable:	Sebastian M.	**Fecha:**	27/03/2017
ID#:	D.EP.EvR.ReER	**Versión:**	1.0
Problema de Negocio	**Criterio de Éxito**	**Resultado**	**Descripción**
(prne.1) Detallar las posibles rutas más frecuentes de navegación que realizan los usuarios.	(cepn.1) Las rutas de navegación frecuentes sean representativas de al menos un 15% del total de usuarios	Ampliatorio	Se confirman los resultados, identificándose la necesidad de estudiar la variación de la navegación de los usuarios según el medio que utilicen.
(prne.2) Identificar y caracterizar perfiles de usuarios de acuerdo a su navegación.	(cepn.2) La caracterización de los perfiles tenga una tasa de error inferior al 20%.	Ampliatorio	El primer grupo está conformado por visitantes ocasionales en busca de algún tipo de recurso que les permita solucionar una problemática específica con respecto al uso del aula virtual. Usualmente, son usuarios sin mucho conocimiento sobre herramientas informáticas. El segundo y tercer grupo, están asociados a usuarios con mayor experticia en el uso del aula virtual, en busca de herramientas o cursos para perfeccionar algún conocimiento específico (en el segundo caso) o con el objetivo de mantenerse informado en los cambios y eventos que realiza el área (para el primer caso). Se acuerda profundizar con la comprensión de las rutas frecuentes de navegación específicas de cada perfil de usuarios.

Tabla A.12. Caso de Validación: Web Log - Reporte de Evaluación de los Resultados (versión 1.0)

A.3. VERSIONADO CASO DE VALIDACIÓN: EDUCACIÓN SUPERIOR

En esta sección se presentan los versionados de los formalismos del segundo caso de validación, separados según el subproceso al que pertenecen: Gestión (sección A.3.1) y Desarrollo (sección A.3.2).

A.3.1. Subproceso: Gestión

Durante el desarrollo de las actividades pertenecientes al subproceso gestión, se registraron las siguientes actualizaciones en las versiones de los productos intermedios: Plan de Acción versión 1.0 y 1.1 (tablas A.13 (a y b) y A.14 (a y b), respectivamente), Diagrama de Gantt actualizado al cierre del proyecto (figura A.3) y Registro de Mediciones versión 1.0 (tabla A.15).

Plan de Acción								
Responsable:		Ramón G.			**Fecha:**		27/07/15	
ID#:		G.PI.PIA.PIAc			**Versión:**		1.0	
ID Actividad	**Actividad**	**Inicio Estimado**	**Inicio Real**	**Fin Estimado**	**Fin Real**	**Esfuerzo Estimado (Hs)**	**Esfuerzo Real (Hs)**	**Comentarios**
G.In	**Iniciación**	**27/07/15**		**28/08/15**		**16**		
G.In.EIP	Exploración Inicial del Proyecto	27/07/15		28/08/15		6		
G.In.DeC	Definición de la Comunicación	27/07/15		28/08/15		3		
G.In.EvS	Evaluación de la Situación	27/07/15		28/08/15		4		
G.In.DCV	Definición del Ciclo de Vida	28/08/15		28/08/15		3		
G.PI	**Planificación**	**31/08/15**		**08/09/15**		**18**		
G.PI.PIM	Planificación de la Mediciones	31/08/15		03/09/15		4		
G.PI.PIA	Planificación de las Actividades	31/08/15		03/09/15		5		
G.PI.PIR	Planificación de los Recursos	03/09/15		04/09/15		4		
G.PI.PRe	Planificación de las Responsabilidades	04/09/15		08/09/15		5		
G.So	**Soporte**	**27/07/15**		**17/11/15**		**8**		
G.So.MeP	Mediciones del Proyecto	27/07/15		17/11/15		4		Se prevé el registro de las métricas a mitad (09/09/15) y final del tiempo estimado para el proyecto
G.So.GeC	Gestión de la Configuración	27/07/15		17/11/15		4		
G.Co	**Control**	**27/07/15**		**20/11/15**		**14**		
G.Co.GeD	Gestión del Desarrollo	27/07/15		20/11/15		6		Se prevé la aplicación del reporte de estado a mitad (09/09/15) y final del tiempo estimado para el proyecto
G.Co.CoA	Control de las Actividades	27/07/15		20/11/15		6		
G.Co.Gca	Gestión del Cambio	27/07/15		20/11/15		2		
G.Ci	**Cierre**	**18/11/15**		**20/11/15**		**12**		
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	18/11/15		19/11/15		4		
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	18/11/15		20/11/15		8		
D.EN	**Entendimiento del Negocio**	**27/07/15**		**31/08/15**		**52**		
D.EN.AnN	Análisis del Negocio	27/07/15		28/08/15		36		
D.EN.CPN	Comprensión del Problema de Negocio	18/08/15		31/08/15		16		
D.ED	**Entendimiento de los Datos**	**28/08/15**		**21/09/15**		**66**		
D.ED.AnD	Análisis de los Datos	28/08/15		11/09/15		30		
D.ED.ExD	Exploración de los Datos	07/09/15		14/09/15		20		
D.ED.EvD	Evaluación de los Datos	14/09/15		21/09/15		16		

Tabla A.13.a. Caso de Validación: Educación Superior - Plan de Acción versión 1.0

D.Mo	**Modelado**	**22/09/15**		**29/09/15**		**24**		
D.Mo.MoP	Modelado del problema	22/09/15		25/09/15		16		
D.Mo.CoM	Configuración del Modelo	28/09/15		29/09/15		8		
D.PD	**Preparación de los Datos**	**01/10/15**		**19/10/15**		**34**		
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	01/10/15		12/10/15		20		
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos	12/10/15		19/10/15		14		
D.Im	**Implementación**	**19/10/15**		**30/10/15**		**40**		
D.Im.SeM	Selección del Modelo	19/10/15		23/10/15		10		
D.Im.ExI	Explotación de Información	26/10/15		30/10/15		30		
D.EP	**Evaluación y Presentación**	**02/11/15**		**16/11/15**		**36**		
D.EP.EvR	Evaluación de los Resultados	02/11/15		06/11/15		14		
D.EP.PrR	Presentación de los Resultados	09/11/15		16/11/15		22		

Tabla A.13.b. Caso de Validación: Educación Superior - Plan de Acción versión 1.0 (continuación)

Plan de Acción								
Responsable:		Ramón G.			**Fecha:**		09/09/15	
ID#:		G.PI.PIA.PIAc			**Versión:**		1.1	
ID Actividad	**Actividad**	**Inicio Estimado**	**Inicio Real**	**Fin Estimado**	**Fin Real**	**Esfuerzo Estimado (Hs)**	**Esfuerzo Real (Hs)**	**Comentarios**
G.In	**Iniciación**	**27/07/15**	**27/07/15**	**28/08/15**	**28/08/15**	**16**	**16**	
G.In.EIP	Exploración Inicial del Proyecto	27/07/15	27/07/15	28/08/15	28/08/15	6	5	
G.In.DeC	Definición de la Comunicación	27/07/15	27/07/15	28/08/15	28/08/15	3	3	
G.In.EvS	Evaluación de la Situación	27/07/15	27/07/15	28/08/15	28/08/15	4	5	
G.In.DCV	Definición del Ciclo de Vida	28/08/15	28/08/15	28/08/15	28/08/15	3	3	
G.PI	**Planificación**	**31/08/15**	**31/08/15**	**08/09/15**	**08/09/15**	**18**	**20**	
G.PI.PIM	Planificación de la Mediciones	31/08/15	31/08/15	03/09/15	03/09/15	4	4	
G.PI.PIA	Planificación de las Actividades	31/08/15	31/08/15	03/09/15	03/09/15	5	6	
G.PI.PIR	Planificación de los Recursos	03/09/15	03/09/15	04/09/15	04/09/15	4	4	
G.PI.PRe	Planificación de las Responsabilidades	04/09/15	04/09/15	08/09/15	08/09/15	5	6	
G.So	**Soporte**	**27/07/15**	**27/07/15**	**17/11/15**		**8**	**4**	
G.So.MeP	Mediciones del Proyecto	27/07/15	27/07/15	17/11/15		4	2	Se prevé el registro de las métricas a mitad (09/09/15) y final del tiempo estimado para el proyecto
G.So.GeC	Gestión de la Configuración	27/07/15	27/07/15	17/11/15		4	2	

Tabla A.14.a. Caso de Validación: Educación Superior - Plan de Acción versión 1.1

G.Co	**Control**	**27/07/15**	**27/07/15**	**20/11/15**		**14**	**3**	
G.Co.GeD	Gestión del Desarrollo	27/07/15	27/07/15	20/11/15		6	2	Se prevé la aplicación del reporte de estado a mitad (09/09/15) y final del tiempo estimado para el proyecto
G.Co.CoA	Control de las Actividades	27/07/15	27/07/15	20/11/15		6	1	
G.Co.Gca	Gestión del Cambio	27/07/15	27/07/15	20/11/15		2	0	
G.Ci	**Cierre**	**18/11/15**	**18/11/15**	**20/11/15**		**12**		
G.Ci.FEC	Formalización Externa del Cierre del Proyecto	18/11/15	18/11/15	19/11/15		4		
G.Ci.FIC	Formalización Interna del Cierre del Proyecto	18/11/15	18/11/15	20/11/15		8		
D.EN	**Entendimiento del Negocio**	**27/07/15**	**27/07/15**	**31/08/15**	**31/08/15**	**52**	**40**	
D.EN.AnN	Análisis del Negocio	27/07/15	27/07/15	28/08/15	28/08/15	36	34	
D.EN.CPN	Comprensión del Problema de Negocio	18/08/15	18/08/15	31/08/15	31/08/15	16	6	
D.ED	**Entendimiento de los Datos**	**28/08/15**	**28/08/15**	**21/09/15**		**66**	**28**	
D.ED.AnD	Análisis de los Datos	28/08/15	28/08/15	11/09/15		30	24	
D.ED.ExD	Exploración de los Datos	07/09/15	07/09/15	14/09/15		20	4	
D.ED.EvD	Evaluación de los Datos	14/09/15	14/09/15	21/09/15		16		
D.Mo	**Modelado**	**22/09/15**	**22/09/15**	**29/09/15**		**24**		
D.Mo.MoP	Modelado del problema	22/09/15	22/09/15	25/09/15		16		
D.Mo.CoM	Configuración del Modelo	28/09/15	28/09/15	29/09/15		8		
D.PD	**Preparación de los Datos**	**01/10/15**	**01/10/15**	**19/10/15**		**34**		
D.PD.CFT	Construcción de la Fuente Temporaria de Datos	01/10/15	01/10/15	12/10/15		20		
D.PD.AFT	Adecuación de la Fuente Temporaria de Datos	12/10/15	12/10/15	19/10/15		14		
D.Im	**Implementación**	**19/10/15**	**19/10/15**	**30/10/15**		**40**		
D.Im.SeM	Selección del Modelo	19/10/15	19/10/15	23/10/15		10		
D.Im.ExI	Explotación de Información	26/10/15	26/10/15	30/10/15		30		
D.EP	**Evaluación y Presentación**	**02/11/15**	**02/11/15**	**16/11/15**		**36**		
D.EP.EvR	Evaluación de los Resultados	02/11/15	02/11/15	06/11/15		14		
D.EP.PrR	Presentación de los Resultados	09/11/15	09/11/15	16/11/15		22		

Tabla A.14.b. Caso de Validación: Educación Superior - Plan de Acción versión 1.1 (continuación)

Nombre de tarea	Comienzo	Fin
Iniciación	**lun 27/07/15**	**vie 28/08/15**
Exploración Inicial del Proyecto	lun 27/07/15	vie 28/08/15
Definición de la Comunicación	lun 27/07/15	vie 28/08/15
Evaluación de la Situación	lun 27/07/15	vie 28/08/15
Definición del Ciclo de Vida	vie 28/08/15	vie 28/08/15
Planificación	**lun 31/08/15**	**mar 08/09/15**
Planificación de la Mediciones	lun 31/08/15	jue 03/09/15
Planificación de las Actividades	lun 31/08/15	jue 03/09/15
Planificación de los Recursos	jue 03/09/15	vie 04/09/15
Planificación de las Responsabilidades	vie 04/09/15	mar 08/09/15
Soporte	**lun 27/07/15**	**mar 17/11/15**
Mediciones del Proyecto	lun 27/07/15	mar 17/11/15
Gestión de la Configuración	lun 27/07/15	mar 17/11/15
Control	**lun 27/07/15**	**vie 20/11/15**
Gestión del Desarrollo	lun 27/07/15	vie 20/11/15
Control de las Actividades	lun 27/07/15	vie 20/11/15
Gestión del Cambio	lun 27/07/15	vie 20/11/15
Cierre	**mié 18/11/15**	**vie 20/11/15**
Formalización Externa del Cierre del Proyecto	mié 18/11/15	jue 19/11/15
Formalización Interna del Cierre del Proyecto	mié 18/11/15	vie 20/11/15
Entendimiento del Negocio	**lun 27/07/15**	**lun 31/08/15**
Análisis del Negocio	lun 27/07/15	vie 28/08/15
Comprensión del Problema de Negocio	mar 18/08/15	lun 31/08/15
Entendimiento de los Datos	**vie 28/08/15**	**lun 21/09/15**
Análisis de los Datos	vie 28/08/15	vie 11/09/15
Exploración de los Datos	lun 07/09/15	lun 14/09/15
Evaluación de los Datos	lun 14/09/15	lun 21/09/15
Modelado	**mar 22/09/15**	**mar 29/09/15**
Modelado del problema	mar 22/09/15	vie 25/09/15
Configuración del Modelo	lun 28/09/15	mar 29/09/15
Preparación de los Datos	**jue 01/10/15**	**lun 19/10/15**
Construcción de la Fuente Temporaria de Datos	jue 01/10/15	lun 12/10/15
Adecuación de la Fuente Temporaria de Datos	lun 12/10/15	lun 19/10/15
Implementación	**lun 19/10/15**	**vie 30/10/15**
Selección del Modelo	lun 19/10/15	vie 23/10/15
Explotación de Información	lun 26/10/15	vie 30/10/15
Evaluación y Presentación	**lun 02/11/15**	**lun 16/11/15**
Evaluación de los Resultados	lun 02/11/15	vie 06/11/15
Presentación de los Resultados	lun 09/11/15	lun 16/11/15

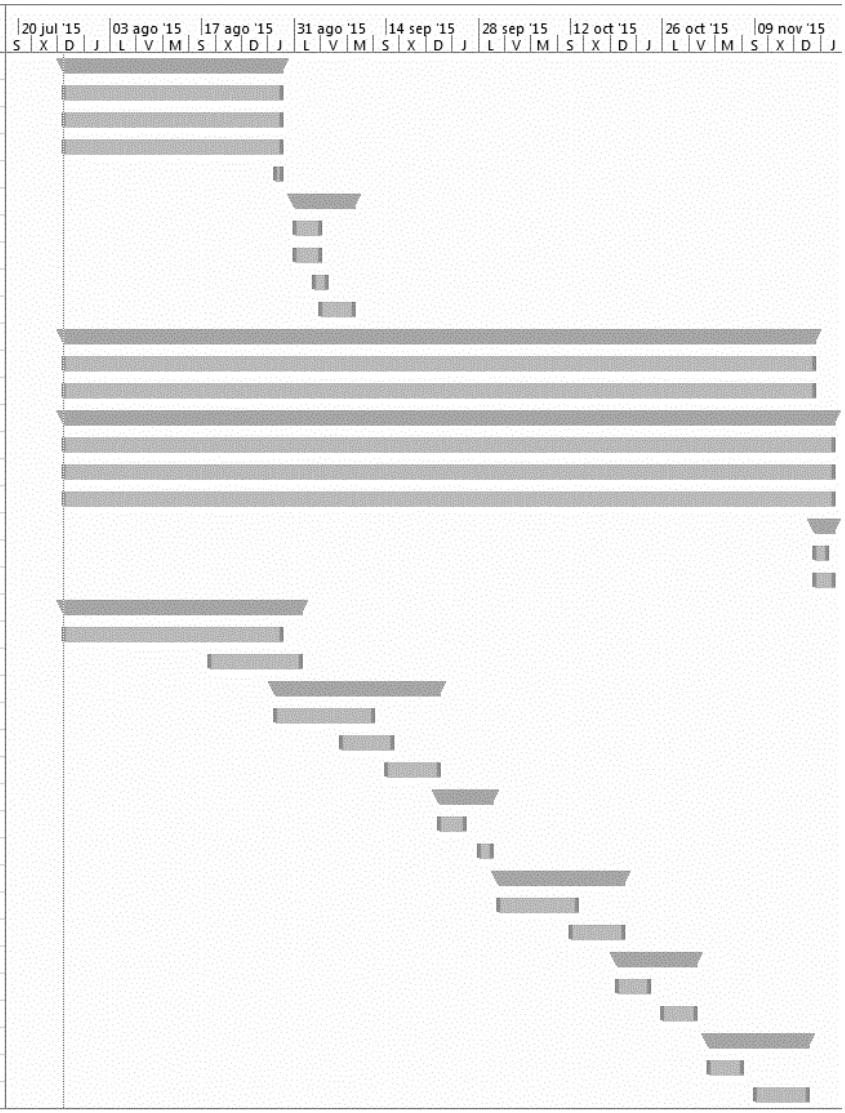


Figura A.3. Prueba de Concepto – Diagrama Gantt (versión final)

Registro de Mediciones			
Responsable:	Ezequiel B.	**Fecha:**	09/09/15
ID#:	G.So.MeP.ReMe	**Versión:**	1.0

Indicador	**Medición**	**Descripción**
Tiempo total requerido para el desarrollo del proyecto	**DRPY** = 114hs	Tdesarrollo = 68 Tgestion = 46
Tiempo medio requerido para el desarrollo de un problema de explotación de información	**-**	
Número total de atributos que no son de utilidad en las tablas	-	
Número medio de atributos significativos por modelo	-	

Tabla A.15. Caso de Validación: Educación Superior - Registro de Mediciones (versión)

A.3.2. Subproceso: Desarrollo

Durante la implementación de las actividades pertenecientes al subproceso Desarrollo, se registró el cambio de versión del formalismo: Fuentes de Información del Cliente versión 1.0 (tabla A.16).

Fuentes de Información del Cliente					
Responsable:		Ezequiel B.		**Fecha:**	28/08/2015
ID#:		D.EN.ANN.FUIC		**Versión:**	1.0
ID	**Nombre**	**Categoría**	**Responsable**	**Descripción**	
fuic.1	SIU_GUARANI	Almacén de datos	(rehi.5) Jorge P.	Sistema de gestión académica que posee información de los estudiantes del tipo académico, como socio-económico y geográficas, relevadas hasta julio de 2014. Se dispone de más de 1500 registros.	
fuic.2	Moodle	Almacén de datos	(rehi.5) Jorge P.	Se registra la información del alumno respecto de la cursada de una materia específica	

Tabla A.16. Caso de Validación: Educación Superior - Fuentes de Información del Cliente

ANEXO B: Mapeo Sistemático de la Literatura

En esta sección presentamos el proceso de Mapeo Sistemático de la Literatura, con el objetivo de identificar y clasificar los elementos existentes en la disciplina, así como identificar las vacancias existentes. Dicho proceso se realizó siguiendo los lineamientos propuestos por Petersen et al., [2008]. Incorporándose la utilización de métricas y criterios de categorización de documentos aplicados en [Munir et al., 2016] con el objetivo de fortalecer los resultados obtenidos.

El resultado derivado del proceso, se utilizó para comprender y describir el estado actual de la disciplina de interés para este trabajo de investigación (capítulo 2 y 3).

El proceso está estructurado en 5 etapas: definición del alcance de la investigación (sección B.1), identificación de artículos de control (sección B.2), definición de la estrategia de búsqueda (sección B.3), definición de los criterios de inclusión y exclusión (sección B.4) y extracción y síntesis (sección B.5).

B.1. ALCANCE DE LA INVESTIGACIÓN

En esta sección se definen las preguntas de investigación que guiarán el proceso de mapeo sistemático de la literatura. El objetivo general del trabajo de investigación consiste en evaluar las propuestas de modelo de proceso para proyectos de ingeniería de explotación de información existentes con el objetivo de definir una propuesta que supla las deficiencias y vacancias identificadas. En este contexto, se definen las siguientes preguntas:

- **RQ1:** ¿Que modelos de procesos/metodologías de Ingeniería de Explotación de Información (IEI) existen?
- **RQ2:** ¿Cómo han cambiado las propuestas en el tiempo?
- **RQ3:** ¿Cuáles son las deficiencias señaladas en los modelos de procesos/ metodologías existentes?

Es relevante señalar, que es interés de la investigación recopilar propuestas generales para proyectos de explotación de información.

B.2. ARTÍCULOS DE CONTROL

Se realiza una búsqueda inicial de la literatura, siguiendo los lineamientos presentados en [Wohlin, 2014] y resumidos a continuación:

- Obtener artículos representativos de distintas comunidades (evitando tendencias a una único lugar de publicación), autores y años de publicaciones.
- El número de artículos inicial no debe ser muy pequeño.
- Si se encuentran demasiados artículos, elegir un subconjunto de acuerdo a su relevancia y alto número de citas.
- Incluir sinónimos de las palabras claves para evitar capturar artículos con una terminología específica.

En esta etapa se identificaron 12 artículos de control, los cuales se listan a continuación:

1. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37.
2. SAS Institute Inc. (1997). Data mining and the case for sampling (white paper).
3. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide.
4. Pyle, D. (2003). Business modeling and data mining. Elsevier.
5. Kurgan, L. A., & Musilek, P. (2006). A survey of Knowledge Discovery and Data Mining process models. The Knowledge Engineering Review, 21(1), 1-24.
6. Mariscal, G., Marban, O., & Fernandez, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. The Knowledge Engineering Review, 25(2), 137-166.
7. Marbán, Ó., Mariscal, G., & Segovia, J. (2009). A data mining & knowledge discovery process model. In Data Mining and Knowledge Discovery in Real Life Applications. InTech.
8. Vanrell, J. Á., & Bertone, R. A. (2010). Modelo de Proceso de Operación para Proyectos de Explotación de Información. In XVI Congreso Argentino de Ciencias de la Computación.
9. Sharma, S. (2008). An integrated knowledge discovery and data mining process model (Doctoral dissertation).
10. Alnoukari, M., & El Sheikh, A. (2012). Knowledge Discovery Process Models: From Traditional to Agile Modeling. In Business Intelligence and Agile Methodologies for Knowledge-Based Organizations: Cross-Disciplinary Applications (pp. 72-100). IGI Global.
11. do Nascimento, G. S., & de Oliveira, A. A. (2012). An agile knowledge discovery in databases software process. In Data and Knowledge Engineering (pp. 56-64). Springer, Berlin, Heidelberg.
12. Gopal, R., Marsden, J. R., & Vanthienen, J. (2011). Information mining—Reflections on recent advancements and the road ahead in data, text, and media mining.

B.3. ESTRATEGIA DE BÚSQUEDA

La estrategia de búsqueda se define en base a los lineamientos presentados en la sección anterior. Se utilizó como base de datos para buscar los artículos Google Scholar, con el objetivo de extraer información de múltiples repositorios representativos de distintas comunidades.

A partir de las preguntas de investigación y los artículos de control, se definieron las palabras de interés (tabla B.1), incluyendo sus sinónimos para evitar capturar artículos con una terminología específica. Las palabras fueron categorizadas en 3 términos de acuerdo a su significado:

- T1: Posibles nombres de la disciplina.
- T2: Posibles nombres del objeto de estudio.
- T3: Posibles tipos de artículos.

Categoría	Término	Sinónimo 1	Sinónimo 2	Sinónimo 3	Sinónimo 4
T1	Data Mining	Knowledge Discovery	Information Mining	Big Data	Data Science
T2	Process Model	Methodology	agile		
T3	survey	review			

Tabla B.1. Listado de términos de interés - Mapeo Sistemático de la Literatura

Las palabras de la categoría T3, son opcionales. Es decir, se considera como tercera opción la ausencia de dicho término, para capturar artículos que no pertenezcan a ninguno de los tipos indicados.

Se realizaron 9 consultas las cuales surgen de la combinación de elementos entre los términos de las categorías T2 y T3. Los términos de la categoría T1 fueron incluidos en todas las consultas, concatenados con ―OR" para obtener todas las posibles terminologías de la disciplina.

1. Process Model AND ("Data Mining" OR "Knowledge Discovery" OR "Information Mining" OR "Big Data" OR "Data Science")
2. Methodology AND ("Data Mining" OR "Knowledge Discovery" OR "Information Mining" OR "Big Data" OR "Data Science")
3. agile AND ("Data Mining" OR "Knowledge Discovery" OR "Information Mining" OR "Big Data" OR "Data Science")
4. survey AND Process Model AND ("Data Mining" OR "Knowledge Discovery" OR "Information Mining" OR "Big Data" OR "Data Science")
5. survey AND Methodology AND ("Data Mining" OR "Knowledge Discovery" OR "Information Mining" OR "Big Data" OR "Data Science")

6. survey agile ("Data Mining" OR "Kn
7. review AND Process Model AND ("Data Mining" OR "Knowledge Discovery" OR "Information Mining" OR "Big Data" OR "Data Science")
8. review AND Methodology AND ("Data Mining" OR "Knowledge Discovery" OR "Information Mining" OR "Big Data" OR "Data Science")
9. review AND agile AND ("Data Mining" OR "Knowledge Discovery" OR "Information Mining" OR "Big Data" OR "Data Science")

Dado que el objetivo de la búsqueda es obtener propuestas generales, se incorporaron a la búsquedas las siguientes palabras claves, para eliminar resultados que poseen dichos términos: clustering, unsupervised, fuzzy, genetic, supervised, web, decision tree, forecasting, health, image, marketing e internet.

Las búsquedas se realizaron haciendo uso de las palabras claves que el motor de búsqueda brinda, filtrando los resultados a partir de los títulos de los recursos. Como resultado se generaron las siguientes consultas:

- allintitle: Process Model "Data Mining" OR "Knowledge Discovery" OR "Information Mining" OR "Big Data" OR "Data Science" -clustering -unsupervised -fuzzy -genetic -supervised -web -"decision tree" -forecasting -health -image -marketing -internet
- allintitle: agile "Data Mining" OR "Knowledge Discovery" OR "Information Mining" OR "Big Data" OR "Data Science" -clustering -unsupervised -fuzzy -genetic -supervised -web -"decision tree" -forecasting -health -image -marketing -internet
- allintitle: Methodology "Data Mining" OR "Knowledge Discovery" OR "Information Mining" OR "Big Data" OR "Data Science" -clustering -unsupervised -fuzzy -genetic -supervised -web -"decision tree" -forecasting -health -image -marketing -internet
- allintitle: survey (process+models) "Data Mining" OR "Knowledge Discovery" OR "Information Mining" OR "Big Data" OR "Data Science" -clustering -unsupervised -fuzzy -genetic -supervised -web -"decision tree" -forecasting -health -image -marketing -internet
- allintitle: survey agile "Data Mining" OR "Knowledge Discovery" OR "Information Mining" OR "Big Data" OR "Data Science" -clustering -unsupervised -fuzzy -genetic -supervised -web -"decision tree" -forecasting -health -image -marketing -internet
- allintitle: survey Methodology "Data Mining" OR "Knowledge Discovery" OR "Information Mining" OR "Big Data" OR "Data Science" -clustering -unsupervised -fuzzy -genetic -supervised -web -"decision tree" -forecasting -health -image -marketing -internet

- allintitle: review (process+models) "Data Mining" OR "Knowledge Discovery" OR "Information Mining" OR "Big Data" OR "Data Science" -clustering -unsupervised -fuzzy -genetic -supervised -web -"decision tree" -forecasting -health -image -marketing -internet
- allintitle: review agile "Data Mining" OR "Knowledge Discovery" OR "Information Mining" OR "Big Data" OR "Data Science" -clustering -unsupervised -fuzzy -genetic -supervised -web -"decision tree" -forecasting -health -image -marketing -internet
- allintitle: review Methodology "Data Mining" OR "Knowledge Discovery" OR "Information Mining" OR "Big Data" OR "Data Science" -clustering -unsupervised -fuzzy -genetic -supervised -web -"decision tree" -forecasting -health -image -marketing -internet

Se siguieron los criterios de optimización de la búsqueda utilizados en [Munir, et al., 2016]. Se utilizaron los 12 artículos de control para medir y refinar las búsquedas. Las métricas utilizadas son precisión y recall. Como resultado de aplicar las consultas se obtuvieron 772 artículos. Los valores obtenidos son:

- Precisión: 66.67% (8/12)
- Recall: 1.04% (8/772).

A partir de la ausencia de 4 papers de control y de acuerdo a la confirmado en [Munir, et al., 2016]: ―**e**hecho que dos de los papers de control no fueron capturados por la cadena final de búsqueda, confirma las observaciones realizadas por Wohlin et al. [2014]: utilizar una única estrategia de búsqueda conlleva a estudios faltantes. Por lo tanto, combinamos las búsquedas en base de datos con snowball sampling.", se utilizó la técnica snowballing [Wohlin, 2014] para ampliar los resultados obtenidos.

B.4. CRITERIOS DE INCLUSIÓN Y EXCLUSIÓN

En esta etapa se introducen los criterios a partir de los cuales se seleccionarán los artículos previamente identificados. La tabla B.2 describe los criterios de inclusión y exclusión utilizados para la selección de los recursos de interés.

Como resultado del proceso (Búsqueda de los términos, snowballing y aplicar los criterios de inclusión y exclusión) se identificaron 68 recursos. A continuación se listan sus referencias, señalando en negrita los 11 artículos que no se tuvieron acceso al documento completo al momento de realizar la evaluación.

1. Alnoukari, M. (2010). ASD-BI: A Business Intelligence Modeling and Integration Framework based on Agile Methodologies(Doctoral dissertation, Arab Academy for Banking and Financial Sciences).

Inclusión:	Libros, artículos, reportes técnicos, literatura gris y white papers (siendo de interés identificar propuestas de formas de trabajos utilizadas por las empresas que no hayan sido publicadas en los medios convencionales).
Exclusión:	Cualquier documento que no indique la identidad de quien lo haya escrito (organización o persona). Aquellas propuestas que sean aplicables a un dominio de negocio específico (es de interés estudiar propuestas generales). Cualquier publicación que no proponga, analiza o evalúa una propuesta de modelo de proceso (o sus términos similares). Cualquier documento en cuyo título o abstract no se indique como tema a tratar la propuesta o el análisis de al menos un modelo de proceso (o sus términos similares). No se posea acceso al material completo

Tabla B.2. Criterios de Inclusión y Exclusión - Mapeo Sistemático de la Literatura

2. Alnoukari, M. (2012). ASD-BI: A Knowledge Discovery Process Modeling Based on Adaptive Software Development Agile Methodology. In Business Intelligence and Agile Methodologies for Knowledge-Based Organizations: Cross-Disciplinary Applications (pp. 183-207). IGI Global.
3. **Alnoukari, M. (2015). ASD-BI: An Agile Methodology for Effective Integration of Data Mining in Business Intelligence Systems. In Integration of Data Mining in Business Intelligence Systems (pp. 61-82). IGI Global.**
4. Alnoukari, M., & El Sheikh, A. (2012). Knowledge Discovery Process Models: From Traditional to Agile Modeling. In Business Intelligence and Agile Methodologies for Knowledge-Based Organizations: Cross-Disciplinary Applications (pp. 72-100). IGI Global.
5. Alnoukari, M., Alzoabi, Z., & Hanna, S. (2008). Applying adaptive software development (ASD) agile modeling on predictive data mining applications: ASD-DM Methodology. In Information Technology, 2008, August. ITSim 2008. International Symposium on (Vol. 2, pp. 1-6). IEEE.
6. Aquino, A. A., Molero-Castillo, G., & Rojano, R. (2018). Hacia un nuevo proceso de minería de datos centrado en el usuario. Pistas Educativas, 36(114).
7. Azevedo, A. I. R. L., & Santos, M. F. (2008). KDD, SEMMA and CRISP-DM: a parallel overview. IADS-DM.

8. Brachman, R., and Anand, T. 1996. The Process of Knowledge Discovery in Databases: A Human-Centered Approach. In Advances in Knowledge Discovery and Data Mining, 37–58, eds. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. Menlo Park, Calif.: AAAI Press
9. Brisson, L., & Collard, M. (2008). An ontology driven data mining process. In International Conference on Enterprise Information Systems (pp. 54-61).
10. Brodley, C. E., & Smyth, P. (1995). The process of applying machine learning algorithms. In proceedings of the ICML-95 workshop on Applying Machine Learning in Practice.
11. Camargo, H., & Silva, M. (2010). Dos caminos en la búsqueda de patrones por medio de Minería de Datos: SEMMA y CRISP. Rev. Tecnol, 9(1).
12. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide.
13. Cios, K. J., Swiniarski, R. W., Pedrycz, W., & Kurgan, L. A. (2007). The knowledge discovery process. In Data Mining (pp. 9-24). Springer, Boston, MA.
14. **Debuse, J. C. W., de la Iglesia, B., Howard, C. M., & Rayward-Smith, V. J. (2001). Building the KDD roadmap. In Industrial Knowledge Management (pp. 179-196). Springer, London.**
15. do Nascimento, G. S., & de Oliveira, A. A. (2012). An agile knowledge discovery in databases software process. In Data and Knowledge Engineering (pp. 56-64). Springer, Berlin, Heidelberg.
16. do Nascimento, G. S., & de Oliveira, A. A. (2013). Agilekdd: An Agile Process Model To Knowledge Discovery In Databases And Business Intelligence Systematization.
17. Džeroski, S. (2006). Towards a general framework for data mining. In International Workshop on Knowledge Discovery in Inductive Databases (pp. 259-300). Springer, Berlin, Heidelberg.
18. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37.
19. Franková, P., Drahošová, M., & Balco, P. (2016). Agile project management approach and its use in big data management. Procedia Computer Science, 83, 576-583.
20. G. Piatetsky, "CRISP-DM still the top methodology for analytics data mining or data science projects", KDD News, 2014.
21. **Gartner. (2000). Free Methodology and Process Model for Data Mining Released.**
22. **Gondar, J. E. (2005). Metodología Del Data Mining. Data Mining Institute, SL.**
23. **González-Aranda, P., Menasalvas, E., Millán, S., Ruiz, C., & Segovia, J. (2008). Towards a methodology for data mining project development: The importance of**

abstraction. In Data Mining: Foundations and Practice (pp. 165-178). Springer, Berlin, Heidelberg.

24. Gopal, R., Marsden, J. R., & Vanthienen, J. (2011). Information mining—Reflections on recent advancements and the road ahead in data, text, and media mining.
25. Gottgtroy, P. (2007). Ontology Driven Knowledge Discovery Process: a proposal to integrate Ontology Engineering and KDD. 11th Pacific-Asia Conference on Information Systems, (pp. 1-7). Auckland, New Zealand.
26. Grady, N. W. (2016). KDD meets big data. In Big Data (Big Data), 2016 IEEE International Conference on (pp. 1603-1608). IEEE.
27. Hofmann, M. (2003). The development of a generic data mining life cycle (DMLC) (dissertation, School of Computing, Dublin Institute of Technology).
28. Hofmann, M., & Tierney, B. (2009). An enhanced data mining life cycle. In Computational Intelligence and Data Mining, 2009. CIDM'09. IEEE Symposium on (pp. 109-117). IEEE.
29. J. S. Saltz, "The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness," in Big Data (Big Data), 2015 IEEE International Conference on, 2015, pp. 2066-2071: IEEE.
30. Klosgen, W., & Zytkow, J. M. (2002). The knowledge discovery process. In Klosgen, W., & Zytkow, J. M. (Eds.), Handbook of data mining and knowledge discovery (pp. 10–21). New York, NY: Oxford University Press.
31. Kurgan, L. A., & Musilek, P. (2006). A survey of Knowledge Discovery and Data Mining process models. The Knowledge Engineering Review, 21(1), 1-24.
32. Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. International Journal of Information Management, 36(5), 700-710.
33. Li, T., & Ruan, D. (2007). An extended process model of knowledge discovery in database. Journal of Enterprise Information Management, 20(2), 169-177.
34. **Ling, C. X. (1999). Perspective of Knowledge Discovery in Database Process Model [J]. Computer Science, 2, 010.**
35. Marbán, Ó., Mariscal, G., & Segovia, J. (2009). A data mining & knowledge discovery process model. In Data Mining and Knowledge Discovery in Real Life Applications. InTech.
36. Marbán, Ó., Mariscal, G., Menasalvas, E., & Segovia, J. (2007). An engineering approach to data mining projects. In International Conference on Intelligent Data Engineering and Automated Learning (pp. 578-588). Springer, Berlin, Heidelberg.
37. Marbán, O., Segovia, J., Menasalvas, E., & Fernández-Baizán, C. (2009). Toward data mining engineering: A software engineering approach. Information systems, 34(1), 87-107.

38. Mariscal, G., Marbán, Ì., & Segovia, F. J. (2013). Un Enfoque Êgil para el Desarrollo de Proyectos de Data Mining.
39. Mariscal, G., Marban, O., & Fernandez, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. The Knowledge Engineering Review, 25(2), 137-166.
40. **Mendes, A. B., Cavique, L., & Santos, J. M. (2013). Data mining process models: a roadmap for knowledge discovery. In Quantitative Modelling in Marketing and Management (pp. 405-433).**
41. Moine, J. M. (2013). Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo (Doctoral dissertation, Facultad de Informática).
42. Moine, J. M., & Haedo, A. S. (2015). Una herramienta para la evaluación y comparación de metodologías de minería de datos. In XXI Congreso Argentino de Ciencias de la Computación (Junín, 2015).
43. Moine, J. M., Gordillo, S. E., & Haedo, A. S. (2011). Análisis comparativo de metodologías para la gestión de proyectos de Minería de Datos. In Congreso Argentino de Ciencias de la Computación (Vol. 17).
44. Moine, J. M., Haedo, A. S., & Gordillo, S. E. (2011). Estudio comparativo de metodologías para minería de datos. In XIII Workshop de Investigadores en Ciencias de la Computación.
45. Moyle, S., & Jorge, A. (2001). RAMSYS-A methodology for supporting rapid remote collaborative data mining projects. In ECML/PKDD01 Workshop: Integrating Aspects of Data Mining, Decision Support and Meta-learning (IDDM-2001).
46. Nogueira, D. R. P. (2014). Agile Data Mining: uma metodologia ágil para o desenvolvimento de projetos de data mining.
47. Osei-Bryson, K. M. (2012). A context-aware data mining process model based framework for supporting evaluation of data mining results. Expert Systems with Applications, 39(1), 1156-1164.
48. Pan, D. (2009). A formal framework for Data Mining process model. In Computational Intelligence and Industrial Applications, 2009. PACIIA 2009. Asia-Pacific Conference on(Vol. 2, pp. 126-129). IEEE.
49. Panov, P., Džeroski, S., & Soldatova, L. (2008). OntoDM: An ontology of data mining. In Data Mining Workshops, 2008. ICDMW'08. IEEE International Conference on (pp. 752-760). IEEE.
50. Panov, P., Soldatova, L., & Džeroski, S. (2013). OntoDM-KDD: ontology for representing the knowledge discovery process. In International Conference on Discovery Science (pp. 126-140). Springer, Berlin, Heidelberg.

51. Pyle, D. (2003). Business modeling and data mining. Elsevier.
52. Rennolls, K., & Al-Shawabkeh, A. (2008). Formal structures for data mining, knowledge discovery and communication in a knowledge management environment. Intelligent Data Analysis, 12(2), 147-163.
53. Saltz, J. (2017). Acceptance Factors for Using a Big Data Capability and Maturity Model.
54. Saltz, J. S., & Shamshurin, I. (2016). Big data team process methodologies: A literature review and the identification of key factors for a project's success. In Big Data (Big Data), 2016 IEEE International Conference on (pp. 2872-2879). IEEE.
55. Saltz, J., & Crowston, K. (2017). Comparing data science project management methodologies via a controlled experiment. In Proceedings of the 50th Hawaii International Conference on System Sciences.
56. **Saltz, J., Hotz, N., Wild, D., & Stirling, K. (2018). Exploring Project Management Methodologies Used Within Data Science Teams.**
57. **Saltz, J., Shamshurin, I., & Connors, C. (2016). A Framework for Describing Big Data Projects. In International Conference on Business Information Systems (pp. 183-195). Springer, Cham.**
58. SAS Institute Inc. (1997). Data mining and the case for sampling (white paper).
59. Shafique, U., & Qaiser, H. (2014). A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). International Journal of Innovation and Scientific Research, 12(1), 217-222.
60. Sharma, N., & Saxena, A. (2018) A Survey Of Data Mining And Knowledge Discovery Process Model And Its Applications In Database.
61. Sharma, S. (2008). An integrated knowledge discovery and data mining process model (Doctoral dissertation).
62. **Sharma, S. (2015). Overview of Knowledge Discovery and Data Mining Process Models. In Knowledge Discovery Process and Methods to Enhance Organizational Performance (pp. 28-41). Auerbach Publications.**
63. **Sharma, S., & Osei-Bryson, K. M. (2010). Toward an integrated knowledge discovery and data mining process model. The Knowledge Engineering Review, 25(1), 49-67.**
64. Sharma, S., Osei-Bryson, K. M., & Kasper, G. M. (2012). Evaluation of an integrated Knowledge Discovery and Data Mining process model. Expert Systems with Applications, 39(13), 11335-11348.
65. Solarte, J. (2002). A proposed data mining methodology and its application to industrial engineering.

66. van Eck, M. L., Lu, X., Leemans, S. J., & van der Aalst, W. M. (2015). PM^2: A Process Mining Project Methodology. In International Conference on Advanced Information Systems Engineering (pp. 297-313). Springer, Cham.
67. Vanrell, J. Á., & Bertone, R. A. (2010). Modelo de Proceso de Operación para Proyectos de Explotación de Información. In XVI Congreso Argentino de Ciencias de la Computación.
68. Vanrell, J. A., Bertone, R., & García-Martínez, R. (2012). A Process Model for Data Mining Projects Un Modelo de Procesos para Proyectos de Explotación de Información. In Proceedings Latin American Congress on Requirements Engineering & Software Testing LACREST (p. 53).

B.5. EXTRACCIÓN Y SÍNTESIS

En esta etapa se realizó una síntesis de la información más relevante (autores, fecha de publicación, abstract, cantidad de referencias) de cada uno de los 56 recursos y se categorizaron de acuerdo a los criterios utilizados en [Munir et al., 2016]. Se definieron dos categorías: tipo de estudio (evaluación, propuesta, validación, opinión, etc.) y tipo de metodología de investigación (caso de estudio, experimentos, survey, etc.). La figura B.1 muestra la distribución de artículos por año de publicación. La figura B.2 (basada en [Wieringa et al., 2006; Runeson et al., 2012]) ilustra la distribución de los artículos por cada intersección de las categorías.

Figura B.1. Distribución de recursos por año de publicación

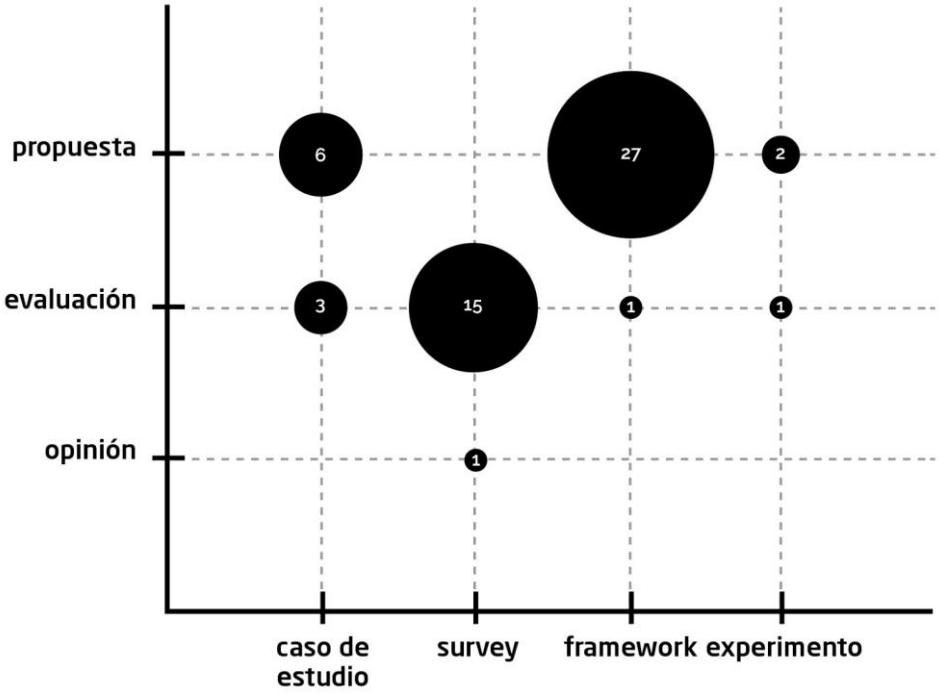


Figura B.2. Distribución de artículos por categorías.

A partir de los recursos obtenidos y la síntesis realizada se observa que el 48% de los artículos identificados son propuestas de modelo de proceso (o sus sinónimos) que definen la estructura de la propuesta desde una perspectiva teórica a partir de deficiencias identificadas en la disciplina, sin mostrar la viabilidad de la propuesta (aplicada a casos de estudio) y/o su eficiencia con respecto a propuestas existentes. De los artículos restantes más de la mitad corresponden a evaluaciones de las propuestas existentes, de las cuales el 10% aplican un proceso sistemático de evaluación.

ANEXO C: Material Complementario Experimentación

Para favorecer la reproducibilidad del experimento, se presenta en esta sección la encuesta realizada a los participantes (sección C.1) y los datos del proceso (sección C.2).

C.1. ENCUESTA

A continuación se presenta la plantilla utilizada para encuestar a cada uno de los participantes del experimento. En la parte superior izquierda se registra la hora de inicio y entrega, mientras que en la parte superior derecha se presenta el patrón utilizado para identificar al individuo (NN) y la propuesta asignada (X).

__ __ . __ __ A.X.1.NN

Gracias por su tiempo para responder esta encuesta. Por favor, indique la respuesta a partir de su experiencia utilizando el modelo de proceso asignado (mencionado como KDDM de manera genérica en adelante).

A) Indique con una cruz "X" la opción seleccionada.

1) Me resultó simple entender lo que el modelo KDDM intentaba modelar.

☐	Totalmente en desacuerdo
☐	En desacuerdo
☐	Levemente en desacuerdo
☐	Indefinido
☐	Levemente de acuerdo
☐	De acuerdo
☐	Totalmente de acuerdo

2) El uso del modelo KDDM fue en ocasiones frustrante.

☐	Totalmente en desacuerdo
☐	En desacuerdo
☐	Levemente en desacuerdo
☐	Indefinido
☐	Levemente de acuerdo

- ☐ De acuerdo
- ☐ Totalmente de acuerdo

3) En general, el modelo KDDM fue fácil de utilizar.

- ☐ Totalmente en desacuerdo
- ☐ En desacuerdo
- ☐ Levemente en desacuerdo
- ☐ Indefinido
- ☐ Levemente de acuerdo
- ☐ De acuerdo
- ☐ Totalmente de acuerdo

4) Entender cómo leer el modelo KDDM me resultó fácil.

- ☐ Totalmente en desacuerdo
- ☐ En desacuerdo
- ☐ Levemente en desacuerdo
- ☐ Indefinido
- ☐ Levemente de acuerdo
- ☐ De acuerdo
- ☐ Totalmente de acuerdo

5) En general, creo que el modelo KDDM es una mejora a la descripción textual de las tareas involucradas en este tipo de procesos.

- ☐ Totalmente en desacuerdo
- ☐ En desacuerdo
- ☐ Levemente en desacuerdo
- ☐ Indefinido
- ☐ Levemente de acuerdo
- ☐ De acuerdo
- ☐ Totalmente de acuerdo

6) En general, el modelo KDDM fue útil para entender el proceso modelado.

- ☐ Totalmente en desacuerdo
- ☐ En desacuerdo
- ☐ Levemente en desacuerdo
- ☐ Indefinido
- ☐ Levemente de acuerdo

- ☐ De acuerdo
- ☐ Totalmente de acuerdo

7) En general, creo que el modelo KDDM mejoró mi rendimiento.

- ☐ Totalmente en desacuerdo
- ☐ En desacuerdo
- ☐ Levemente en desacuerdo
- ☐ Indefinido
- ☐ Levemente de acuerdo
- ☐ De acuerdo
- ☐ Totalmente de acuerdo

8) El modelo KDDM representa correctamente las tareas a realizar en este tipo de proyectos.

- ☐ Totalmente en desacuerdo
- ☐ En desacuerdo
- ☐ Levemente en desacuerdo
- ☐ Indefinido
- ☐ Levemente de acuerdo
- ☐ De acuerdo
- ☐ Totalmente de acuerdo

9) El modelo KDDM es una representación realista de este tipo de procesos.

- ☐ Totalmente en desacuerdo
- ☐ En desacuerdo
- ☐ Levemente en desacuerdo
- ☐ Indefinido
- ☐ Levemente de acuerdo
- ☐ De acuerdo
- ☐ Totalmente de acuerdo

10) El modelo KDDM contiene elementos contradictorios.

- ☐ Totalmente en desacuerdo
- ☐ En desacuerdo
- ☐ Levemente en desacuerdo
- ☐ Indefinido
- ☐ Levemente de acuerdo
- ☐ De acuerdo

- [ ] Totalmente de acuerdo

11) Todos los elementos en el modelo KDDM son relevantes para la representación de este tipo de procesos.

- [ ] Totalmente en desacuerdo
- [ ] En desacuerdo
- [ ] Levemente en desacuerdo
- [ ] Indefinido
- [ ] Levemente de acuerdo
- [ ] De acuerdo
- [ ] Totalmente de acuerdo

12) El modelo KDDM provee una completa representación de este tipo de procesos.

- [ ] Totalmente en desacuerdo
- [ ] En desacuerdo
- [ ] Levemente en desacuerdo
- [ ] Indefinido
- [ ] Levemente de acuerdo
- [ ] De acuerdo
- [ ] Totalmente de acuerdo

13) El modelo KDDM satisfizo adecuadamente las necesidades de información que se me pidió que diera soporte.

- [ ] Totalmente en desacuerdo
- [ ] En desacuerdo
- [ ] Levemente en desacuerdo
- [ ] Indefinido
- [ ] Levemente de acuerdo
- [ ] De acuerdo
- [ ] Totalmente de acuerdo

14) El modelo KDDM no me proveyó de manera eficiente la información que necesitaba.

- [ ] Totalmente en desacuerdo
- [ ] En desacuerdo
- [ ] Levemente en desacuerdo
- [ ] Indefinido

- ☐ Levemente de acuerdo
- ☐ De acuerdo
- ☐ Totalmente de acuerdo

15) El modelo KDDM me proveyó de manera eficaz la información que necesitaba.

- ☐ Totalmente en desacuerdo
- ☐ En desacuerdo
- ☐ Levemente en desacuerdo
- ☐ Indefinido
- ☐ Levemente de acuerdo
- ☐ De acuerdo
- ☐ Totalmente de acuerdo

16) En general, estoy satisfecho con la información que me proveyó el modelo KDDM.

- ☐ Totalmente en desacuerdo
- ☐ En desacuerdo
- ☐ Levemente en desacuerdo
- ☐ Indefinido
- ☐ Levemente de acuerdo
- ☐ De acuerdo
- ☐ Totalmente de acuerdo

B) Completar los siguientes enunciados en cantidad de años

17) Experiencia en la disciplina:

18) Experiencia trabajando en la industria de explotación de información:

C.2. DATOS

En esta sección se presentan los datos crudos obtenidos de las respuestas de los 42 participantes del experimento (tabla C.1) y el set de datos transformado para la implementación del experimento (tabla C.2).

ID	Grupo	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14	p15	p16	Exp.	Ind. Exp.	Time taken
1	ikddm	4	5	5	3	3	5	4	5	4	5	3	2	5	5	5	4	1	0	13
2	ikddm	3	5	7	2	3	4	4	2	3	2	5	3	3	5	3	3	4	2	18
3	ikddm	3	6	5	2	4	4	5	4	5	4	3	3	5	6	4	5	4	0	22
4	ikddm	3	4	5	4	5	5	4	5	5	4	5	5	6	5	5	5	2	2	23
5	ikddm	5	6	4	4	4	5	6	5	5	5	2	3	5	4	5	5	0	0	22
6	ikddm	5	6	3	5	5	5	5	6	5	4	3	3	5	4	5	5	0	0	26
7	ikddm	3	4	5	4	4	4	4	4	3	3	4	3	4	4	4	4	3	1	23
8	ikddm	3	4	5	4	4	4	4	4	4	4	4	3	4	4	4	4	4	0	14
9	ikddm	2	5	7	2	2	3	3	1	3	2	6	4	3	4	3	3	2	0	30
10	ikddm	4	5	5	3	4	4	5	5	5	4	3	3	4	5	4	5	3	0	13
11	ikddm	4	6	5	4	4	4	5	5	5	3	3	2	5	5	5	5	3	2	28
12	ikddm	5	6	4	5	5	4	5	6	5	4	3	3	5	4	5	5	3	4	25
13	ikddm	5	6	2	6	4	4	6	6	5	5	2	2	4	5	5	6	3	0	33
14	ikddm	3	5	5	3	2	4	4	2	3	3	5	3	3	4	3	3	0	0	15
15	ikddm	2	4	5	3	3	3	4	2	3	2	6	2	3	4	3	4	2	0	29
16	ikddm	5	4	4	4	4	5	4	5	4	4	5	5	3	5	4	5	2	0	19
17	ikddm	5	5	4	5	6	3	4	5	4	4	3	4	3	4	5	5	0	0	13
18	ikddm	3	4	3	5	5	4	4	3	4	4	5	5	6	4	4	4	1	0	24
19	ikddm	2	4	6	4	2	4	4	3	2	4	6	4	2	4	4	2	4	1	17
20	ikddm	5	3	3	4	5	5	4	4	3	4	5	5	5	4	4	6	1	0	32
21	ikddm	4	5	5	3	4	4	4	2	3	3	4	3	3	4	4	4	4	2	22
22	mopropei	4	4	5	4	3	3	4	4	5	4	4	3	4	4	4	5	2	0	25
23	mopropei	5	5	3	5	4	5	6	7	5	5	3	3	5	6	5	6	4	2	13
24	mopropei	5	6	2	6	4	6	6	7	6	6	3	2	4	6	6	6	0	0	16
25	mopropei	3	5	4	5	5	4	3	4	4	5	4	3	5	5	4	6	1	0	19
26	mopropei	4	5	3	4	5	5	4	5	5	6	4	2	4	5	4	5	0	0	22

Tabla C.1.a Datos crudos experimento

ID	Grupo	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14	p15	p16	Exp.	Ind. Exp.	Time taken
27	mopropei	4	6	4	4	4	5	4	5	4	6	3	2	5	7	4	6	1	0	27
28	mopropei	6	7	2	7	6	6	5	7	6	6	2	2	6	7	6	7	4	2	25
29	mopropei	3	4	4	4	4	4	3	3	5	4	4	4	4	5	4	5	2	0	21
30	mopropei	4	5	4	4	5	4	5	4	3	5	4	4	5	5	4	5	2	2	23
31	mopropei	5	6	4	5	4	5	4	5	5	6	2	3	5	6	5	6	2	0	20
32	mopropei	5	5	3	5	4	5	5	5	5	6	3	3	5	5	5	6	3	0	14
33	mopropei	3	5	4	3	4	3	5	3	4	3	4	4	4	4	4	4	4	0	12
34	mopropei	6	6	2	5	5	6	6	7	6	5	3	2	5	6	5	6	4	2	28
35	mopropei	4	5	3	4	3	4	5	4	3	5	4	2	4	5	4	6	4	4	17
36	mopropei	4	5	4	4	4	4	5	5	4	5	4	2	4	4	4	5	0	0	21
37	mopropei	4	5	3	4	5	5	4	3	5	4	3	3	4	5	4	5	3	1	19
38	mopropei	3	5	4	3	4	3	5	5	6	3	3	5	5	4	6	4	3	0	32
39	mopropei	4	5	4	4	4	4	5	4	4	4	5	4	4	5	4	5	2	0	26
40	mopropei	4	5	4	4	5	4	5	4	4	5	3	2	5	6	5	4	3	2	32
41	mopropei	5	4	4	3	4	4	4	4	5	3	3	4	4	5	4	5	1	0	23
42	mopropei	5	4	6	4	5	4	4	5	4	4	4	5	4	4	4	4	0	0	21

Tabla C.1.b Datos crudos experimento (continuación)

ID	Grupo	PEOU	US	PU	PSQ	Exp.	Ind. Exp.	surveyscore	Time taken
1	ikddm	14	18	19	20	1	0	71	13
2	ikddm	11	13	12	15	4	2	51	18
3	ikddm	15	16	19	20	4	0	70	22
4	ikddm	15	18	19	17	2	2	69	23
5	ikddm	19	20	21	19	0	0	79	22
6	ikddm	20	22	20	18	0	0	80	26
7	ikddm	14	16	15	16	3	1	61	23
8	ikddm	14	16	16	17	4	0	63	14
9	ikddm	8	11	11	13	2	0	43	30
10	ikddm	16	17	18	19	3	0	70	13
11	ikddm	16	19	20	19	3	2	74	28
12	ikddm	19	21	20	18	3	4	78	25
13	ikddm	21	22	20	22	3	0	85	33
14	ikddm	12	14	12	15	0	0	53	15
15	ikddm	12	12	11	16	2	0	51	29

Tabla C.2.a Datos transformados experimento

ID	Grupo	PEOU	US	PU	PSQ	Exp.	Ind. Exp.	surveyscore	Time taken
16	ikddm	17	18	14	17	2	0	66	19
17	ikddm	19	18	17	17	0	0	71	13
18	ikddm	17	16	17	15	1	0	65	24
19	ikddm	10	15	10	14	4	1	49	17
20	ikddm	19	16	15	17	1	0	67	32
21	ikddm	15	14	14	16	4	2	59	22
22	mopropei	14	15	17	18	2	0	64	25
23	mopropei	20	22	20	22	4	2	84	13
24	mopropei	21	25	21	24	0	0	91	16
25	mopropei	15	18	17	21	1	0	71	19
26	mopropei	18	19	17	22	0	0	76	22
27	mopropei	16	20	18	25	1	0	79	27
28	mopropei	23	27	24	26	4	2	100	25
29	mopropei	14	15	17	18	2	0	64	21
30	mopropei	18	17	16	19	2	2	70	23
31	mopropei	17	21	21	23	2	0	82	20
32	mopropei	19	20	20	22	3	0	81	14
33	mopropei	16	14	16	15	4	0	61	12
34	mopropei	23	24	21	23	4	2	91	28
35	mopropei	17	17	15	22	4	4	71	17
36	mopropei	17	18	16	20	0	0	71	21
37	mopropei	18	17	18	19	3	1	72	19
38	mopropei	16	16	22	14	3	0	68	32
39	mopropei	17	17	15	18	2	0	67	26
40	mopropei	18	17	19	21	3	2	75	32
41	mopropei	17	15	18	17	1	0	67	23
42	mopropei	16	17	16	15	0	0	64	21

Tabla C.2.b Datos transformados experimento (continuación)