

# Propuesta de un Modelo para Evaluar la Viabilidad de Proyectos de Explotación de Información

Pytel, P.<sup>1,2,3</sup>, Britos, P.<sup>4</sup>, García-Martínez, R.<sup>2</sup>

<sup>1</sup> Programa de Doctorado en Ciencias Informáticas. Facultad de Informática. Universidad Nacional de La Plata. Argentina.

<sup>2</sup> Grupo Investigación en Sistemas de Información. Departamento Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús. Argentina.

<sup>3</sup> Grupo de Estudio en Metodologías de Ingeniería de Software. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional. Argentina.

<sup>4</sup> Grupo de Investigación en Explotación de Información del Laboratorio de Informática Aplicada. Sede Andina (El Bolsón). Universidad Nacional de Río Negro. Argentina.

*ppytel@gmail.com, paobritos@gmail.com, rgarcia@unla.edu.ar*

**Resumen.** La Explotación de Información define procesos y las metodologías con el objetivo de extraer conocimiento no-trivial e implícito de los datos disponibles en una organización. Se ha observado la falta de procesos y herramientas que permitan soportar la identificación de los problemas del proyecto al inicio del mismo con sus riesgos asociados. De la gran cantidad de proyectos desarrollados, no todos finalizan con éxito, terminando la mayoría en fracasos. Por lo tanto, este trabajo tiene como objetivo proponer un Modelo que permita la Evaluación de la Viabilidad para Proyectos de Explotación de Información usando la información disponible al comienzo del mismo. Se define entonces un procedimiento de cinco pasos que permite, a partir de la valoración de las características del proyecto, determinar tanto si el proyecto es viable o no como identificar sus puntos débiles.

**Palabras Claves:** Estudio de Viabilidad. Explotación de Información. Ingeniería en Software. PyMEs.

## 1. Introducción

La Explotación de Información consiste en la extracción de conocimiento no-trivial que reside de manera implícita en los datos disponibles en distintas fuentes de información [1]. Dicho conocimiento es previamente desconocido y puede resultar útil para algún proceso [2]. Una vez identificado el problema de Inteligencia de Negocio es posible definir el Proceso de Explotación de Información formado por varias técnicas de Minería de Datos que se ejecuta para lograr, a partir de un conjunto de información con un grado de valor para la organización, en otro conjunto de información con un grado de valor mayor que el inicial [3]. Así la Minería de Datos está relacio-

nada a la tecnología (algoritmos) y la Explotación de Información está relacionada con los procesos y las metodologías. Si bien existen metodologías que acompañan el desarrollo de proyectos de explotación de información que se consideran probadas y tienen un buen nivel de madurez en cuanto al desarrollo del proyecto entre las cuales se destacan CRISP-DM [4], P3TQ [5] y SEMMA [6], estas metodologías dejan de lado aspectos a nivel gestión de los proyectos y de empresa [7]. En estas metodologías se observa la falta de procesos y herramientas que permitan soportar las actividades de gestión al inicio del mismo como es la evaluación de viabilidad. A partir de esta evaluación es posible predecir los principales riesgos del proyecto.

En este contexto, este trabajo tiene como objetivo desarrollar un Modelo que permita la Evaluación de la Viabilidad para Proyectos de Explotación de Información a partir de la información disponible al comienzo del mismo. Para ello primero se realiza una reseña documental sobre modelos de evaluación de viabilidad relacionados a software tradicional, sistemas inteligentes y la explotación de información (sección 2). Luego se define el modelo propuesto (sección 3), realizando dos pruebas de concepto para su validación (sección 4). Finalmente se indican las conclusiones obtenidas y futuras líneas de trabajo (sección 5).

## **2. Modelos para Evaluación de Viabilidad**

Antes de emprender todo proyecto software, la organización debe decidir si es conveniente realizarlo o no. Para poder tomar esa decisión, la cual es compleja y depende de una multitud de factores, se necesita conocer el impacto que ese software va a causar en la organización y los riesgos que la organización posee por su construcción [8]. Para ello es necesario estudiar las características del proyecto a través de una evaluación de la viabilidad técnica y económica del proyecto (también conocida como estudio de factibilidad). Como resultado de esta evaluación se puede determinar si se cumplen las condiciones para garantizar la finalización satisfactoriamente del desarrollo del sistema software. Para realizar dicha evaluación de la viabilidad técnica [9] primero se debe recolectar información tanto sobre la organización (la actitud de los directivos y de los usuarios, la experiencia y los conocimientos previos de los desarrolladores y los procesos organizativos afectados por el software entre otros) como del sistema a desarrollar (el alcance y las características del problema a resolver, la disponibilidad de información necesaria, la integración con otros sistemas software entre otros). Esta información es luego procesada mediante un modelo que genera un informe el cual indica el grado de viabilidad técnica del proyecto.

En los proyectos de construcción de sistemas basados en conocimiento sucede algo similar con la diferencia que la Ingeniería del Conocimiento (INCO) debe evaluar la viabilidad del proyecto considerando varias dimensiones [10]. Dado que las especificaciones iniciales de estos sistemas suelen ser inciertas, incompletas y contradictorias, es necesario desarrollar distintos prototipos para definir coherentemente las funcionalidades, el rendimiento y las interfaces del sistema. Esto produce que los proyectos de la INCO sean más largos y costosos que los de software tradicional [11].

Los Proyectos de Explotación de Información también tienen la necesidad de evaluar la viabilidad del proyecto antes de comenzar el mismo. De la gran cantidad de proyec-

tos desarrollados, no todos finalizan con éxito [12, 13] terminando la mayoría en fracasos (según estudios realizados entre un 60% [14] y 85% [15] han fracasado en la consecución de sus objetivos). Esto se debe, entre otras razones, a problemas no detectados al comienzo del mismo. Mediante una detección temprana de los problemas se podrían reducir los riesgos durante el desarrollo del proyecto. Sin embargo, dada que las características de los proyectos de explotación de información son diferentes a los proyectos de software tradicional y a los proyectos de la INCO, no es posible reutilizar los modelos propuestos para estos tipos de proyectos por lo que sería necesario contar con modelos específicos.

En este dominio, se han realizado diversos trabajos para la identificación de los criterios de éxitos en estos proyectos [16-19], pero no existe un modelo completo que permita analizar al comienzo si el proyecto es viable o no. Si bien en [20] se ha propuesto un modelo que utiliza un sistema experto difuso para medir el nivel de éxito de proyectos a partir de la calidad empleada en cada una de sus fases de CRISP-DM, este estudio puede ser realizado una vez que el proyecto ya está finalizado por necesitar conocer el nivel de calidad empleado en cada fase. Por otro lado, en [21] se aplica un análisis Bayesiano para determinar si la empresa se encuentra calificada para comenzar un proyecto de explotación de información (es decir, se valoran las características de la empresa y los datos para decidir si se puede aplicar minería de datos o no). Sin embargo, deja de lado para evaluar cuestiones importantes para desarrollar el proyecto. Además este análisis deja de lado la clasificación de la viabilidad en diferentes dimensiones por lo que considera la viabilidad como un todo.

En este trabajo, y en forma similar a lo indicado en [10, 23] para proyectos de INCO, se cree útil clasificar las características de proyectos de explotación de información en las siguientes dimensiones:

*Plausibilidad del proyecto:* Esta dimensión incluye todas las características que hacen que sea posible realizar el proyecto de explotación de información.

*Adecuación del proyecto:* Incluye todas las características que determinan que la explotación de información es apropiada para el problema de negocio detectado (es decir es la mejor solución para el problema).

*Éxito del proyecto:* Incluye todas aquellas características que aseguran el éxito del proyecto de explotación de información.

Un proyecto será viable si el proyecto es *apropiado* para ser realizado con explotación de información, y es *posible* desarrollarlo con *éxito*. Nótese que no se ha incluido la evaluación de la *justificación* del proyecto por considerarse incluida en las otras tres dimensiones.

### **3. Modelo Propuesto para Evaluación de Viabilidad**

Para proponer un Modelo de Viabilidad, primero se deben identificar las condiciones que el proyecto debe cumplir para ser considerado viable. Esto se realiza mediante una investigación documental detallada con [15-22]. Las condiciones identificadas luego han sido clasificadas usando las tres dimensiones descriptas en la sección 2.

Sin embargo, debe tenerse en cuenta que estas condiciones normalmente no pueden ser contestadas fácilmente con respuestas del tipo 'sí' / 'no' o dando una valoración

numérica. Para el ingeniero será más sencillo aplicar un rango amplio de valores lingüísticos. Entonces el modelo propuesto debe poder manejar dichos valores para determinar la viabilidad proyecto y luego contar con un proceso que permita convertir los valores de su forma lingüística a una numérica. Una vez que los valores numéricos son obtenidos, estos son ponderados y aplicados con la media armónica para calcular el valor correspondiente para cada dimensión (plausibilidad, adecuación y criterio de éxito). Finalmente, los valores obtenidos para cada dimensión son combinados en una fórmula para obtener la valoración global de la viabilidad del proyecto.

Para formalizar lo explicado anteriormente, a continuación se describen cinco pasos que se deben realizar para evaluar la viabilidad de un proyecto:

**Paso 1: Determinar el valor correspondiente para cada una de las características solicitadas por el modelo.**

Para caracterizar un proyecto de explotación de información y evaluar luego su viabilidad se utilizan las características definidas en la tabla 1. A partir del resultado de las entrevistas realizadas en la organización, el ingeniero debe responder las preguntas asociadas a cada característica. Los valores lingüísticos permitidos para las respuestas son 'nada', 'poco', 'regular', 'mucho' y 'todo'. Donde cuanto más verdadera parezca una característica, mayor valor se le debe asignar (es decir, 'mucho' o 'todo') y cuanto más falsa parezca, menor valor ('nada' o 'poco'). Nótese que cuando el valor asignado a una característica no supera su umbral (indicado en la tabla 1), el proyecto se considera como no viable y no es necesario continuar con los siguientes pasos.

**Paso 2: Convertir los valores en intervalos difusos.**

Una vez que los valores lingüísticos han sido definidos para cada característica de la tabla 1, se deben traducir utilizando el proceso de transformación indicado en el test de viabilidad para proyectos de INCO descrito en [10, 23] que se encuentra basado en sistemas expertos difusos [24]. Para ello, cada valor lingüístico se traduce en un intervalo difuso expresado por cuatro valores numéricos (entre cero y diez) que representan los puntos de ruptura (o puntos angulares) de su función de pertenencia correspondiente. Estos intervalos, junto con la representación gráfica de la función de pertenencia, se indican en la figura 1.

**Paso 3: Calcular la valoración de cada dimensión.**

Para calcular la valoración de cada dimensión del proyecto, los intervalos difusos (convertidos en el paso 2) son ponderados considerando su peso correspondiente (definido en la tabla 1). Para calcular el intervalo que representa la valoración de cada dimensión ( $I_d$ ) se utiliza la fórmula #1 de la tabla 2. Esta fórmula está formada por la combinación de la media armónica y la media aritmética del conjunto de intervalos. De esta forma se busca reducir la influencia de valores bajos en el cálculo de la dimensión. Ya que el resultado de la fórmula anterior es otro intervalo difuso (formado por cuatro valores numéricos), para convertir dicho intervalo en un único valor numérico ( $V_d$ ) se utiliza la media aritmética de los valores del intervalo que se encuentra definido en la fórmula #2 de la tabla 2.

**Paso 4: Calcular la valoración global de la viabilidad del proyecto.**

Finalmente, los valores numéricos calculados en el paso anterior para cada dimensión son combinados a través de una media aritmética ponderada (fórmula #3 de la tabla 2) y así se consigue el valor de la viabilidad global del proyecto ( $EV$ ).

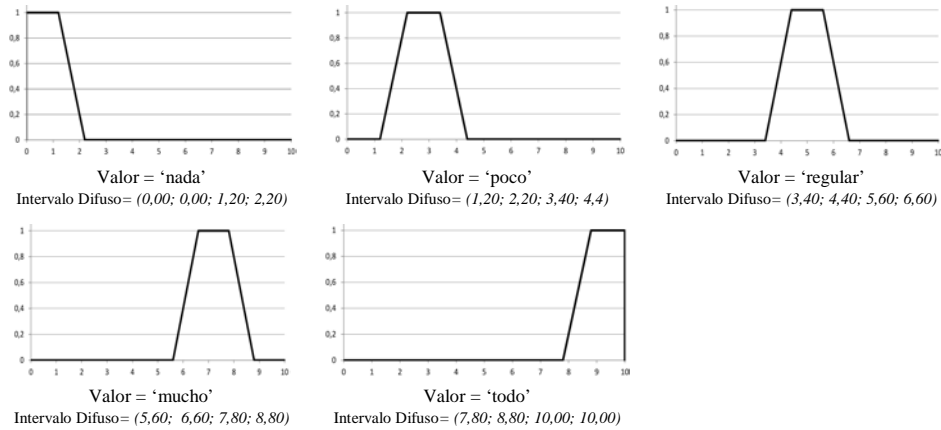
### **Paso 5: Interpretar los resultados obtenidos.**

Una vez que los valores correspondientes a cada dimensión y al proyecto global son calculados (pasos 3 y 4 respectivamente), se deben analizar. Para poder interpretar los resultados de la viabilidad de cada dimensión, se recomienda graficar la función de pertenencia del intervalo difuso ( $I_d$ ) para determinar el valor lingüístico le corresponde. Se considera que la viabilidad de la dimensión está aceptada si supera al intervalo del valor 'regular' (esto es análogo a considerar que el valor de la dimensión  $V_d$  es mayor a 5). Por otro lado, para la viabilidad del proyecto se utiliza el siguiente criterio: si las tres dimensiones son aceptadas (es decir el valor de cada dimensión es mayor a 5) y la valoración global de la viabilidad proyecto,  $EV$ , es mayor a 5 entonces el proyecto es viable. En caso contrario, el proyecto no es viable. En ambos casos, el ingeniero también podrá observar los puntos débiles del proyecto que debe reforzar (en caso de proyecto no viable) y monitorear para el desarrollo del proyecto.

**Tabla 1.** Características a ser evaluadas por el método propuesto

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

En el caso de la característica A4, debido a que se encuentra negada, se recomienda intentar responder la pregunta contraria (es decir, “¿En qué medida el problema de negocio puede ser resuelto aplicando técnicas estadísticas tradicionales?”), y luego asignar el valor complemento para la evaluación de la viabilidad (es decir 'poco' en lugar de 'mucho', 'nada' en lugar de 'todo' y viceversa).



**Fig. 1.** Representación de la Función de Pertinencia y asignación de Intervalo Difuso para los Valores Lingüísticos

**Tabla 2.** Fórmulas utilizadas por el modelo propuesto.

| # | Fórmula  | Donde:  |
|---|--|---|
| 1 | $I_d = \left( \frac{1}{2} \cdot \frac{\sum_{i=1}^{n_d} P_{d_i}}{\sum_{i=1}^{n_d} \left( \frac{P_{d_i}}{C_{d_i}} \right)} \right) + \left( \frac{1}{2} \cdot \frac{\sum_{i=1}^{n_d} (P_{d_i} \cdot C_{d_i})}{\sum_{i=1}^{n_d} P_{d_i}} \right)$ | <p><math>I_d</math>: representa el intervalo difuso calculado para la dimensión <math>d</math> (usando como nomenclatura '<math>P</math>' para plausibilidad, '<math>A</math>' para adecuación y '<math>E</math>' para criterio de éxito).</p> <p><math>P_{d_i}</math>: representa el peso de la característica <math>i</math> perteneciente a la dimensión <math>d</math>.</p> <p><math>C_{d_i}</math>: representa el intervalo difuso asignado a la característica <math>i</math> perteneciente a la dimensión <math>d</math>.</p> <p><math>n_d</math>: representa la cantidad de características asociada a la dimensión <math>d</math>.</p> |
| 2 | $V_d = \frac{\sum_{i=1}^4 I_{d_i}}{4}$   | <p>Donde:</p> <p><math>V_d</math>: representa el valor numérico calculado para la dimensión <math>d</math>.</p> <p><math>I_{d_i}</math>: representa el valor de la posición <math>i</math> del intervalo difuso calculado para la dimensión <math>d</math>.</p>   |
| 3 | $EV = \frac{8 \cdot V_P + 8 \cdot V_A + 6 \cdot V_E}{22}$  | <p>Donde:</p> <p><math>EV</math>: representa el valor global de la viabilidad del proyecto.</p> <p><math>V_P</math>: representa el valor para la dimensión plausibilidad.</p> <p><math>V_A</math>: representa el valor para la dimensión adecuación.</p> <p><math>V_E</math>: representa el valor para la dimensión criterio de éxito.</p>  |

#### 4. Prueba de Concepto

Para validar el modelo para evaluación de la viabilidad propuesto, se utiliza primero una prueba de concepto positiva con un proyecto de explotación de información real finalizada con éxito (y por lo tanto viable), y luego otra negativa modificación algunos datos de dicho proyecto para que no sea viable. El objetivo del proyecto era la detección de evidencias de causalidad entre la satisfacción general de los clientes de una organización proveedora de Internet mediante la detección de evidencias de causalidad entre satisfacción general, servicio contratado y evidencias de baja de clientes. Para ello se utiliza la información de una encuesta realizada por la organización a sus

clientes. Para realizar la prueba de concepto positiva, se aplicaron los pasos propuestos en la sección 3. A partir de diversas sesiones de educación en la organización se ha definido el problema de negocio, las principales características de la organización y de los datos disponibles. Con esta información se respondieron las condiciones requeridas con el valor lingüístico correspondiente como se indica en la tabla 3 (paso 1). Estos valores luego son convertidos en intervalos difusos (paso 2) para calcular el intervalo de cada dimensión (paso 3) con su representación gráfica como se muestra en la tabla 4. Finalmente, se calcula el valor numérico de cada dimensión y la valoración global de la viabilidad del proyecto (paso 4) que son interpretados (paso 5) como se indica en la tabla 5. Para realizar todos estos cálculos se ha utilizado la planilla disponible en [25].

**Tabla 3.** Asignación de las características del proyecto utilizado como prueba positiva.

| Categoría           | ID | Pregunta asociada a la Característica   | Respuesta  | Valor Asig.    |
|---------------------|----|---|--|----------------|
| Datos               | P1 | ¿En qué medida los repositorios disponibles poseen datos actuales?  | Los repositorios poseen datos contestados por clientes actuales de la organización.  | <i>todo</i>    |
|                     | P2 | ¿Qué tan representativos son los datos de los repositorios disponibles para resolver el problema de negocio?                            | No se posee información de todos los clientes que se han dado de baja en los últimos 6 meses.  | <i>regular</i> |
|                     | A1 | ¿En qué medida los repositorios se encuentran disponibles en formato digital?   | Las respuestas de la encuesta se encuentran totalmente digitalizadas.  | <i>todo</i>    |
|                     | A2 | ¿Qué cantidad de atributos y registros tienen los datos disponibles?  | Se cuenta con apropiadamente 5000 registros y 20 atributos para ser utilizados.  | <i>mucho</i>   |
|                     | A3 | ¿Cuánta confianza se posee en la credibilidad de los datos disponibles?   | La encuesta ha sido respondida por los clientes sin supervisión a través de una aplicación web.  | <i>regular</i> |
|                     | E1 | ¿Cuánto facilita la tecnología de los repositorios disponibles las tareas de manipulación de los datos?                                 | La encuesta digitalizada ha sido suministrada en un archivo de texto para ser procesada.   | <i>poco</i>    |
| Problema de Negocio | P3 | ¿Cuánto se entiende del problema de negocio?  | Varias sesiones de educación se han realizado para determinar el objetivo del proyecto para solucionar el problema de negocio detectado.   | <i>todo</i>    |
|                     | A4 | ¿En qué medida el problema de negocio no puede ser resuelto aplicando técnicas estadísticas tradicionales?                              | La organización no cuenta con ningún experto disponible en los datos a ser utilizados que pueda definir alguna hipótesis para ser probadas por técnicas estadísticas. Se considera que la "mejor solución" es la aplicación de técnicas de minería de datos. | <i>mucho</i>   |
|                     | A5 | ¿Qué tan estable es el problema de negocio durante el desarrollo del proyecto?  | No se posee mucha seguridad que el problema de negocio detectado se mantenga durante todo el proyecto ya que depende del comportamiento de los clientes.   | <i>regular</i> |
| Proyecto            | E2 | ¿Cuánto apoyan los interesados (stakeholders) al proyecto?  | El gerente de sistemas y marketing tienen grandes intereses en la finalización exitosa de este proyecto.   | <i>mucho</i>   |
|                     | E3 | ¿En qué medida la planificación del proyecto permite considerar la realización de buenas prácticas ingenieriles con el tiempo adecuado? | La organización desea recibir los resultados en el menor tiempo posible.   | <i>regular</i> |
| Equipo de Trabajo   | P4 | ¿Qué nivel de conocimientos posee el equipo de trabajo sobre explotación de información?  | El equipo de trabajo posee gran conocimientos sobre explotación de información en general y las técnicas de minería de datos en particular.  | <i>todo</i>    |
|                     | E4 | ¿Qué nivel de experiencia posee el equipo de trabajo en proyectos similares?  | El equipo de trabajo posee experiencia en la aplicación de explotación de información en proyectos similares.  | <i>mucho</i>   |

Por otro lado, para realizar la prueba de concepto negativa, se modificaron los valores de las seis características asociadas a los Datos y al Problema de Negocio de la tabla 4 con los valores 'regular' (para las características P1, P3 y A4) y 'poco' (para A1, A2 y A5). Nótese que los nuevos valores no son inferiores a los umbrales para no generar un ejemplo negativo trivial. Luego, se calculan los intervalos y valores correspondientes utilizando la planilla disponible en [26]. Estos resultados se muestran en las tablas

6 y 7 (en la primer tabla, la dimensión Éxito no se muestra porque sus resultados no se modifican con respecto a los de la tabla 4).

**Tabla 4.** Traducción y cálculo de intervalos por dimensión para prueba de concepto positiva.

| Dimensión     | ID | Intervalo Difuso del Valor Asig. | Intervalo Valor Dimensión ( $I_d$ )  | Representación de la Función de Pertenencia de $I_d$ |
|---------------|----|----------------------------------|--|--|
| Plausibilidad | P1 | (7,8; 8,8; 10; 10)               | $(6,05; 7,12; 8,39; 8,82)$<br>Intervalo sobrepasa por pequeña diferencia al valor 'mucho'.   |  |
|               | P2 | (3,4; 4,4; 5,6; 6,6)             |  |  |
|               | P3 | (7,8; 8,8; 10; 10)               |  |  |
|               | P4 | (7,8; 8,8; 10; 10)               |  |  |
| Adecuación    | A1 | (7,8; 8,8; 10; 10)               | $(4,65; 5,68; 6,91; 7,84)$<br>Intervalo entre valores 'regular' y 'mucho'.                   |  |
|               | A2 | (5,6; 6,6; 7,8; 8,8)             |  |  |
|               | A3 | (3,4; 4,4; 5,6; 6,6)             |  |  |
|               | A4 | (5,6; 6,6; 7,8; 8,8)             |  |  |
|               | A5 | (3,4; 4,4; 5,6; 6,6)             |  |  |
| Éxito         | E1 | (1,2; 2,2; 3,4; 4,4)             | $(3,44; 4,62; 5,93; 6,99)$<br>Intervalo sobrepasa por pequeña diferencia al valor 'regular'. |  |
|               | E2 | (5,6; 6,6; 7,8; 8,8)             |  |  |
|               | E3 | (3,4; 4,4; 5,6; 6,6)             |  |  |
|               | E4 | (5,6; 6,6; 7,8; 8,8)             |  |  |

**Tabla 5.** Cálculo por dimensión y viabilidad global para la prueba de concepto positiva.

| Dimensión   | Valor de la Dimensión ( $V_d$ ) | Interpretación   |
|---|---------------------------------|--|
| Plausibilidad                                       | 7,60                            | Dado que los tres valores son superiores al valor mínimo requerido de 5, la viabilidad de todas las dimensiones está aceptada. Sin embargo, debe notarse que a pesar de que la valoración de Plausibilidad y Adecuación es holgada, para el Éxito del proyecto es muy cercana al valor mínimo requerido. Esto significa que durante el proyecto habrá que monitorear con mayor atención a las características evaluadas para el éxito. |
| Adecuación  | 6,27                            |  |
| Éxito   | 5,25                            |  |
| Valor global de la viabilidad del proyecto ( $EV$ ) | 6,47                            | <i>Se considera que el proyecto es viable para ser realizado.</i>  |

**Tabla 6.** Traducción y cálculo de intervalos por dimensión para prueba de concepto negativa.

| Dimensión     | ID | Intervalo Difuso del Valor Asig. | Intervalo Valor Dimensión ( $I_d$ )  | Representación de la Función de Pertenencia de $I_d$ |
|---------------|----|----------------------------------|--|--|
| Plausibilidad | P1 | (3,4;4,4;5,6;6,6)                | $(4,06; 5,08; 6,31; 7,18)$<br>Intervalo sobrepasa por pequeña diferencia al valor 'regular'. |  |
|               | P2 | (3,4;4,4;5,6;6,6)                |  |  |
|               | P3 | (3,4;4,4;5,6;6,6)                |  |  |
|               | P4 | (7,8;8,8;10;10)                  |  |  |
| Adecuación    | A1 | (1,2;2,2;3,4;4,4)                | $(1,99; 3,06; 4,31; 5,34)$<br>Intervalo sobrepasa por pequeña diferencia al valor 'poco'.    |  |
|               | A2 | (1,2;2,2;3,4;4,4)                |  |  |
|               | A3 | (3,4;4,4;5,6;6,6)                |  |  |
|               | A4 | (3,4;4,4;5,6;6,6)                |  |  |
|               | A5 | (1,2;2,2;3,4;4,4)                |  |  |



**Tabla 7.** Cálculo por dimensión y viabilidad global para la prueba de concepto negativa.

| Dimensión   | Valor de la Dimensión ( $V_d$ ) | Interpretación  |
|---|---------------------------------|---|
| Plausibilidad                                       | 5,66                            | Dado que el valor para la Adecuación no supera el valor mínimo de 5, por lo que no se considera que la explotación de información sea la solución adecuada para este proyecto. Esto significa también que el proyecto no es viable para ser realizado. Por otro lado, los valores para Plausibilidad y el Éxito, a pesar de superar el mínimo, no lo hacen por mucho por lo que los riesgos del proyecto son altos. |
| Adecuación  | 3,67                            |   |
| Éxito   | 5,25                            |   |
| Valor global de la viabilidad del proyecto ( $EV$ ) | 4,82                            | <i>Se considera que el proyecto no es viable para ser realizado.</i>  |

## 5. Conclusiones

La Explotación de Información define procesos y las metodologías con el objetivo de extraer conocimiento no-trivial e implícito de los datos disponibles en una organización. Se ha observado la falta de procesos y herramientas que permitan soportar la identificación de los problemas del proyecto al inicio del mismo con sus riesgos asociados. De la gran cantidad de proyectos desarrollados, no todos finalizan con éxito, terminando la mayoría en fracasos. Por lo tanto, este trabajo propone un Modelo que permita la Evaluación de la Viabilidad para Proyectos de Explotación de Información usando la información disponible al comienzo del mismo.

Para ello se define un procedimiento que consta de cinco pasos para registrar y procesar las características del proyecto y de esta forma evaluar su viabilidad considerando tres dimensiones: la plausibilidad que indica si el proyecto es posible ser realizable, si explotación de información es la solución más apropiada y si el proyecto tendrá éxito. A partir de dicha evaluación luego se calcula la viabilidad global del proyecto identificando sus puntos débiles. Si el proyecto es considerado como no viable, el ingeniero deberá realizar alguna acción para tratar de subir su valoración (así puede intentar obtener más datos representativos y/o actuales, o conseguir mayor apoyo de los directivos, entre otras acciones) y de esta forma hacer que el proyecto alcance la valoración de viable. Aun cuando el proyecto sea viable, estos puntos débiles deben ser monitoreados durante el desarrollo del proyecto como riesgos. Es responsabilidad del ingeniero mantener o “subir” su valor para evitar así el fracaso del proyecto.

Queda como futura línea de investigación, presentar un análisis detallado de la variación de las características consideradas por el modelo para los proyectos y su efecto en la evaluación de la viabilidad.

## 6. Referencias

1. Schiefer, J., Jeng, J., Kapoor, S. & Chowdhary, P. (2004). *Process Information Factory: A Data Management Approach for Enhancing Business Process Intelligence*. Proceedings 2004 IEEE International Conference on E-Commerce Technology. Pp. 162-169.
2. Stefanovic, N., Majstorovic, V., Stefanovic, D. (2006). *Supply Chain Business Intelligence Model*. Proceedings 13th International Conference on Life Cycle Engineering. P. 613-618.
3. 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. Pp. 83-99. ISBN 978-958-8692-32-6.

4. Chapman P., Clinton J., Keber R., et al. (2000). *CRISP-DM 1.0 Step by step BI guide Edited by SPSS*. <http://tinyurl.com/crispdm> [Abril 2012].
5. Pyle, D. (2003) *Business Modeling and Business intelligence*. Morgan Kaufmann
6. SAS (2008) *SAS Enterprise Miner: SEMMA*. <http://tinyurl.com/semmaSAS> [Abril 2012].
7. Vanrell, J., Bertone, R., & García-Martínez, R. (2010). *Modelo de Proceso de Operación para Proyectos de Explotación de Información*. Anales del XVI Congreso Argentino de Ciencias de la Computación, Pp. 674-682. ISBN 978-950-9474-49-9.
8. Pressman, R. (2004). *Software Engineering: A Practitioner's Approach*. Editorial Mc Graw Hill.
9. Sapag, N. (2000) *Preparación y evaluación de proyectos*. 4ª Edición. McGraw Hill.
10. García Martínez, R. y Britos, P. (2004). *Ingeniería de Sistemas Expertos*. 649 páginas. Editorial Nueva Librería. ISBN 987-1104-15-4.
11. Gómez, A., Juristo, N., Montes, C. & Pazos, J. (1997) *Ingeniería del Conocimiento*. Centro de Estudios Ramón Areces. S.A., Madrid.
12. Edelstein, H.A. & Edelstein, H.C.. (1997) *Building, Using, and Managing the Data Warehouse*. Data Warehousing Institute. Prentice-Hall PTR, EnglewoodCliffs, NJ.
13. Strand, M. (2000) *The Business Value of Data Warehouses—Opportunities, Pitfalls and Future Directions*. Ph.D. Thesis, Department of Computer Science, University of Skovde.
14. Gondar, J. E. (2005) *Metodología Del Data Mining*. Number 84-96272-21-4. Data Mining Institute, S.L.
15. Fayyad, U.M (2000) *Tutorial report*. Summer school of DM. Monash University, Australia.
16. Sim, J. (2003). *Critical success factors in data mining projects*. Ph.D. Thesis, University of North Texas.
17. Nemati, H. R., & Barko, C. D. (2003). *Key factors for achieving organizational data-mining success*. *Industrial Management & Data Systems*, 103(4), 282-292. doi:10.1108/02635570310470692.
18. Davenport, T. H. (2009). *Make Better Decisions*. *Harvard Business Review*, (November), 117-123.
19. Bolea, U., Jakličb, J., Papac, G., & Žabkard, J. (2011). *Critical Success Factors of Data Mining in Organizations*. Ljubljana.
20. Nadali, A., Kakhky, E. N., & Nosratabadi, H. E. (2011). *Evaluating the success level of data mining projects based on CRISP-DM methodology by a Fuzzy expert system*. *Electronics Computer Technology (ICECT) 2011 3rd International Conference on Kanyakumari* (Vol. 6, pp. 161–165). IEEE. doi:10.1109/ICECTECH.2011.5942073
21. Nie, G., Zhang, L., Liu, Y., Zheng, X., & Shi, Y. (2009). *Decision analysis of data mining project based on Bayesian risk*. *Expert Systems with Applications*, 36(3), 4589–4594.
22. Pipino, L. L., Lee, Y. W. & Wang, R. Y. (2002). *Data quality assessment*. *Communications of the ACM*, 45(4), 211–218.
23. López, D., García-Martínez, R., Marsiglio, A. (1991) *Factibilidad de Construcción de Sistemas Basados en el Conocimiento*. Anales del II Simposio de Inteligencia Artificial y Robótica. Páginas 63-73. Universidad Nacional de Luján, Buenos Aires.
24. Jang, J. S. R. (1997). *Fuzzy inference systems*. Upper Saddle River, NJ: Prentice-Hall.
25. Pytel, P. (2012) *Implementación del Modelo de Viabilidad Propuesto con datos de Prueba de Concepto Positiva*. <http://tinyurl.com/ViabPruConcepto>.
26. Pytel, P. (2012) *Implementación del Modelo de Viabilidad Propuesto con datos de Prueba de Concepto Negativa*. <http://tinyurl.com/ViabPruConceptoNeg>.