



**MINERÍA DE DATOS APLICADA A ESTRATEGIAS
PARA MINIMIZAR EL REZAGO ACADÉMICO Y LA
DESERCIÓN UNIVERSITARIA EN CARRERAS DE
INFORMÁTICA DE LA UNNOBA.**

*Tesis presentada para obtener el grado de Doctor en
Ciencias Informáticas*

Tesista: Mg. Claudia Russo

Director: Ing. Armando De Giusti

Co- Director: Dr. Olivas Varela

FACULTAD DE INFORMÁTICA – UNIVERSIDAD NACIONAL DE LA PLATA

MAYO DE 2019

ÍNDICE

RESUMEN.....

INTRODUCCIÓN

- I. Origen y justificación de la investigación.....
- II. Marco teórico y estado del arte.....
- III. Planteo del problema.....
- IV. Objetivos y aportes.....
- V. Metodología utilizada.....

CAPÍTULO 1: DESERCIÓN ESTUDIANTIL UNIVERSITARIA.

- 1.1 Conceptualización.....
- 1.2 Deserción estudiantil universitaria y estrategias de abordaje en determinados países iberoamericanos.....
- 1.3 La problemática de la deserción estudiantil universitaria en Argentina.....
- 1.4 Deserción estudiantil universitaria y estrategias de abordaje en las carreras del campo de ciencia aplicada en Argentina.....

CAPÍTULO 2: DATA MINING Y EDUCATIONAL DATA MINING

- 2.1 Inteligencia de Negocio o Business Intelligence (BI). Su aporte en torno a Minería de Datos y Pentaho.
- 2.2 Descubrimiento del Conocimiento (KKD), Minería de Datos (DM) y Learning Analytics (LA). Conceptualización y aplicación.....
- 2.3 Minería de datos aplicada al campo educativo. Conceptualización y estado del arte.....
- 2.4 Monitoreo y predicción del rendimiento académico a través de la minería de datos. Estudio de casos.....

CAPÍTULO 3: REZAGO ACADÉMICO Y DESERCIÓN ESTUDIANTIL UNIVERSITARIA EN LAS CARRERAS DEL ÁREA DE INFORMÁTICA DE LA UNNOBA.

- 3.1 Origen y composición de la Escuela de Tecnología de la UNNOBA.....
- 3.2 Oferta académica y matriculación en el área de informática durante el período 2005-2015.....
- 3.3 La problemática del rezago académico y la deserción en las carreras de informática de la UNNOBA. Antecedentes y estrategias llevadas a cabo.....
- 3.4 Análisis de factores causales del rezago y deserción en las carreras de Informática de la UNNOBA.....

CAPÍTULO 4: DEFINICIÓN, DESARROLLO Y APLICACIÓN DE UN MODELO DE INTERVENCIÓN VIRTUAL/ TABLERO DE CONTROL PARA PREDECIR Y MINIMIZAR LA DESERCIÓN EN LAS CARRERAS DE INFORMÁTICA DE LA UNNOBA.

- 4.1 Análisis Predictivo.....
- 4.2 Modelización y aplicación del tablero de control.....

CONCLUSIONES Y APORTES.....

REFERENCIAS BIBLIOGRÁFICAS.....

RESUMEN

La presente investigación se desarrolla en el campo de la Ciencia Informática aplicada en Educación, articulando el área Informática con el subárea de Tecnología informática aplicada al campo educativo.

El objetivo general que orienta esta tesis radica en definir y desarrollar, a partir de la utilización de la minería de datos, un modelo de intervención áulica virtual que permita diseñar, e implementar, estrategias tecnológicas tendientes a realizar un seguimiento de las trayectorias educativas de aquellos estudiantes con rezago académico o riesgos de deserción; se espera que ello contribuya a minimizar la deserción en las carreras de Informática de la Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA).

Los índices de deserción y desgranamiento en estas carreras son muy elevados, sobre todo en los dos primeros años. Ante esta certeza, se espera como aporte de esta tesis poder detectar, utilizando distintos procesos de minería de datos, falencias académicas que permitan predecir el rendimiento de los estudiantes y actuar en los potenciales casos de deserción, contribuyendo en última instancia a minimizar este fenómeno.

Asimismo, los resultados de esta investigación y las herramientas de minería de datos aplicadas en su desarrollo, pueden hacerse extensibles a otros campos del sistema educativo en general o carreras de la UNNOBA en particular, incidiendo positivamente sobre la trayectoria académica de los estudiantes en riesgo de abandono.

INTRODUCCIÓN

I. Origen y justificación de la investigación.

En Argentina, la deserción estudiantil universitaria se presenta como una de las problemáticas más acuciantes de la educación superior.

En la actualidad el sector Software y Servicios Informáticos (SSI) es uno de los principales demandantes de empleo calificado. Sin embargo, la matrícula de las carreras universitarias no acompaña esta demanda y resultan insuficientes los profesionales informáticos (Dapozo *et.al*, 2014); estadísticamente se ha demostrado que se recibe un alumno de cada cuatro que ingresan (Drovandi, 2010).

En relación con ello, en lo que respecta a las carreras de grado relacionadas con el campo de la Ciencia Informática, las estadísticas de los últimos años han estimado que alrededor de un 80% de los estudiantes desertan durante el transcurso de los primeros dos años de las carreras. De hecho se ha demostrado que prevalece una deserción temprana en el primer año de las carreras, cercana al 50 % (Madoz y Gorga, 2006).

II. Marco teórico y estado del arte.

Los estudios pioneros sobre la deserción estudiantil universitaria datan de los años 1970 y 1980 (Spady, 1971; Tinto, 1987), la preocupación por el estudio de esta temática cobró fuerza a partir de la década de 1990 cuando, al compás del crecimiento de la matrícula universitaria, se fue haciendo notorio el incremento de los índices de deserción. En la actualidad, en el mundo occidental, en general, e iberoamericano, en particular, es vasta la literatura que existe sobre este fenómeno.

Existen numerosos estudios y varias definiciones del concepto deserción. Por lo tanto, la conceptualización de este fenómeno es sumamente heterogénea y depende del objeto de estudio y de las características y finalidad de la investigación a la que se aplique. En

relación con ello, en un trabajo empírico, cada institución educativa establece un espacio temporal a partir del cual se califique a un estudiante como desertor.

De acuerdo con esta literatura, la deserción universitaria obedece a múltiples causas, difíciles de jerarquizar en forma generalizada, vinculadas con variables tanto exógenas como endógenas al sistema académico y que no son excluyentes entre sí (Vásquez, 2003). Siguiendo a Páramo y Correa (1999) las variables que contribuyen a explicar la deserción pueden ser diferenciadas en causas 'intrasujeto' y causas 'extrasujeto,' según se trate de actitudes propias de los estudiantes o de factores socio-económicos que determinan o inciden en la decisión de desertar. Así, podemos encontrar estudios que, a partir de análisis empíricos, analizan la deserción universitaria poniendo el foco en una u otra clasificación y otorgan diversos grados de incidencia a los factores asociados a ella.

En relación con las 'causas intrasujeto' (Páramo y Correa, 1999), algunos estudios han postulado la capacidad intelectual del estudiante como la variable predictora más importante en el fenómeno desertivo (Lyengar y Lepper, 2000); otros han puesto el acento en la autorresponsabilidad del estudiante respecto a sus propias metas académicas y al compromiso e integración a la universidad donde ingresó (Tinto, 1975), o en relación con sus hábitos de estudio, señalando que quienes presentan una pobre predisposición para "sentarse a estudiar" tienden a desertar más que los que exhiben mejores hábitos (Johnston, 2002). Otros autores, desde un enfoque psicológico, han hecho hincapié en las aspiraciones y valores portados por los propios estudiantes como condicionantes a la hora de decidir o no el abandono de la carrera (Fishbein y Ajzen, 1975; Attinasi, 1986; Ethington, 1990). También desde una perspectiva individual, la deserción ha sido explicada atendiendo a las metas y propósitos que tienen las personas al momento de incorporarse al sistema de educación superior (Tinto, 1989; Ethington, 1990)

En lo que respecta a las causas 'extrasujeto', y desde un modelo sociológico, algunos trabajos han analizado la influencia de las condiciones socioeconómicas en el desempeño académico. Uno de los referentes de este modelo, y pionero además en los estudios sobre deserción, es Spady (1970). Basándose en la teoría del suicidio de Durkheim, este autor ha sugerido que la deserción es el resultado de la falta de integración social del estudiante en el entorno de la educación superior, determinada por otros condicionantes familiares y del medio en el que se desarrolla. Siguiendo a Spady, si las

influencias familiares y del medio en que vive el estudiante no se producen en una dirección positiva, pueden implicar un bajo rendimiento académico y una deficitaria integración a la institución, lo cual se traduce en mayores probabilidades de deserción. Por el contrario, si el estudiante cuenta con un estímulo familiar y del medio, es muy probable que logre un desarrollo académico y social acorde con sus expectativas personales e institucionales, y permanezca en la carrera y en la institución hasta lograr su titulación.

En relación con la teoría de Spady, Vicent Tinto, uno de los mayores referentes de los estudios sobre deserción, ha señalado a lo largo de sus diversos trabajos (1975; 1989; 2003) que en muchos casos, el fracaso escolar y su correlato, la deserción, dependen del “clima institucional” y de la “integración” del estudiante al mismo. En otros términos, mayoritariamente los casos de deserción se explican por una insuficiente integración personal en los ambientes intelectual y social de la comunidad institucional (Tinto, 1975).

Algunos trabajos recientes, elaborados desde un enfoque organizacional, han analizado la problemática de la deserción atendiendo a los servicios que las instituciones de educación superior ofrecen a los estudiantes, ya sea en cuanto a la calidad docente (Braxton, et.al., 2000) o a la disponibilidad de recursos materiales (bibliografía, comodidades en cuanto a infraestructura, cantidad de profesores por alumnos, etc.) ofrecidos por las instituciones (Tillman, 2002).

Otros trabajos, tributarios también de un enfoque sociológico, han demostrado que es más alta la deserción en el caso de estudiantes con bajos ingresos al momento de iniciar sus estudios (Montoya, 1999) y que la retención de alumnos con padres de menor nivel de educación es muy bajo (Porto, et. al, 2001). Contrariamente a estas posturas, otras investigaciones muestran que el origen social tiene nula o escasa influencia en el éxito en la Universidad (Germani y Sautú, 1965; Latiesa, 1992). Así, señalan que una vez dentro de la enseñanza superior, la motivación y el compromiso con el estudio son las variables que tienen mayor peso al determinar el rendimiento.

En la actualidad, la mayor parte de los estudios sobre deserción presentan una perspectiva más abarcadora que integra variables pertenecientes a diferentes ámbitos, y que son susceptibles de ser modificadas por las instituciones para reducir la deserción (Himmel, 2002). Ejemplo de ello es la obra de Cabrera et al. (2006) en la cual, los autores,

a partir de un recuento de las diversas variables explicativas de la deserción, han elaborado una clasificación de cinco grandes enfoques - psicopedagógico, adaptación, estructural, economicista e interaccionista - en los que se pueden integrar los distintos estudios existentes sobre el tema. Autores inscriptos en estas líneas heterogéneas han analizado la deserción desde la perspectiva de la interacción entre la evaluación previa que hace el estudiante de las opciones de educación superior, la forma de integración social y académica en las instituciones, y la variable económica, por ejemplo el hecho de contar o no con apoyo financiero, con un sistema de becas, etc.

A partir de un relevamiento bibliográfico y empírico sobre el fenómeno de la deserción estudiantil universitaria, en su tesis doctoral, Parrino (2012), sintetiza en cuatro grupos las principales causas y factores que inciden en la deserción:

1. Las *variables individuales* como edad, género, grupo familiar e integración social.
2. Las *variables académicas* como la orientación profesional, el desarrollo intelectual, el rendimiento académico, los métodos de estudio, los procesos de admisión, el grado de satisfacción con la carrera y la carga académica.
3. Las *variables institucionales* como la normativa académica, el financiamiento estudiantil, los recursos universitarios, la calidad del programa y las relaciones con los profesores y pares.
4. Las *variables socioeconómicas* como el estrato socioeconómico de pertenencia, la situación laboral del estudiante, la situación laboral y el nivel educacional de los padres. (Parrino, 2012:66)

Además de la preocupación por dar cuenta de las causas de la deserción, los estudios empíricos sobre el tema se interesan por analizar las etapas o períodos de las carreras en las que se presentan los mayores índices de abandono. En relación con ello, la mayor parte de estos estudios sostiene que los índices más altos de deserción se presentan durante el transcurso del primer año de las carreras (Tinto, 1989; Corominas, 2001; Vivas, 2005) o en la fase postrera del primer año de estudios y antes del comienzo del segundo (Tinto, 1989; Rivera, E. et. al., 2005). Algunos de ellos atribuyen las causas a las dificultades que encuentran los estudiantes para establecer contactos con la comunidad social e intelectual de la institución y lograr amistades en ella. De hecho, algunos de estos estudios han demostrado que las mencionadas dificultades tienden a ser mayores en la

etapa temprana de la carrera que en los últimos años (Tinto, 1989). Según Corominas (2001) “las deserciones se producen mayormente en primero, debido a elecciones inadecuadas de la carrera a elegir, bajo rendimiento académico, no haber aprobado el mínimo de créditos, asignaturas poco motivadoras, y poco esfuerzo y compromiso con el estudio” (p.130)

Data Mining

La Minería de Datos o *Data Mining*, definida como “*a computer-based information system (CBIS)*” (Vlahos, Ferratt, & Knoepfle, 2004) es considerada como una etapa de un proceso más amplio conocido como *Knowledge Discovery in Databases* (Frawley, *et.al.*, 1992). Consiste en “un conjunto de técnicas y herramientas aplicadas al proceso de extraer y presentar conocimiento implícito, previamente desconocido, potencialmente útil y humanamente comprensible, a partir de grandes conjuntos de datos, con objeto de predecir, de forma automatizada, tendencias o comportamientos y descubrir modelos previamente desconocidos” (Frawley, *et.al.*, 1992:58).

A diferencia de otras técnicas, la minería de datos permite construir modelos de manera automática, por lo cual aquellas aplicaciones que no son, en algún grado, auto guiadas están realizando análisis de datos y no minería de datos (Pautsch, *et.al.* , 2011). En los últimos años, la aplicación de la minería de datos en el campo educativo (EDM) ha tenido un fuerte desarrollo, impulsando los sistemas de enseñanza a distancia basados en web o sistemas de e-learning. La EDM trabaja con el objetivo de encontrar patrones y realizar predicciones para caracterizar el comportamiento de los estudiantes y sus logros o sus conocimientos sobre los contenidos del dominio, para detectar estudiantes en riesgo de abandono, o para predecir su rendimiento, entre muchos otros (García Saíz, 2016).

Las plataformas e-learning y el modo de aprendizaje virtual que ello conlleva, están siendo especialmente utilizados en el nivel superior de educación, tanto para complementar con nuevos contenidos y actividades las clases presenciales, conocida como educación semi-presencial o blended learning, como impartir cursos integramente virtuales, lo que comúnmente se denomina e-learning.

Plataformas de aprendizaje virtual como Moodle almacenan toda la información relativa a la actividad que los estudiantes desarrollan en cursos virtuales en ficheros o bases de datos que, procesados correctamente, pueden ofrecer información relevante para el docente. El docente puede conocer el comportamiento que tienen los alumnos en la plataforma y descubrir el proceso de aprendizaje que llevan a cabo. Con esto, el docente podrá adaptar el curso al modo en que trabajan sus alumnos y tomar medidas ante los problemas que se puedan detectar. Es decir, esta información que recopilan los sistemas información educativos puede utilizarse para tomar decisiones y responder a preguntas, buscando la mejora de la calidad del sistema educativo.

III. Planteo del problema.

Al igual que en otras universidades del país y la región, la problemática del desgranamiento y la deserción está también presente en las Carreras de Informática de la Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA).

En esta universidad de reciente creación, si bien el número de estudiantes de las carreras de informática se ha quintuplicado en el término de ocho años, la tasa de egreso ha sido sumamente baja. Esto da cuenta del desgranamiento y deserción estudiantil presente en dichas carreras.

La presente investigación, al desarrollarse en el campo de la Ciencia Informática aplicada en Educación, articula el área Informática con el subárea de Tecnología informática aplicada en Educación

IV. Objetivos y aportes.

El objetivo general que orienta esta investigación radica en diseñar y desarrollar, utilizando herramientas de minería de datos, un modelo de intervención áulica virtual que permita detectar patrones de rendimiento académico y a partir de allí poder predecir, de

forma temprana, potenciales casos de deserción en las Carreras del área de Informática de la UNNOBA.

Como objetivos específicos esta investigación se propone:

- Generar relevamientos que permitan distinguir los casos de rezago académico y deserción interna, de aquéllos casos de deserción absoluta del sistema educativo.
- Utilizar la minería de datos para determinar y analizar, en términos estadísticos los indicadores y causales de desgranamiento en las carreras de informática
- Diseñar y aplicar la minería de datos en el monitoreo de las trayectorias académicas de los alumnos, estableciendo patrones que permitan dar cuenta de las variaciones en su rendimiento a través de los distintos años de su carrera.
- Determinar, a partir de la minería de datos, potenciales casos de fracaso académico con el fin de tomar medidas tendientes a minimizar el problema.
- Diseñar una herramienta software que permitan a los docentes, sin conocimientos sobre minería de datos obtener, interpretar y tomar decisiones para mejorar el proceso de enseñanza-aprendizaje.
- Analizar experimentalmente datos de las carreras de Informática de la UNNOBA a fin de perfeccionar la herramienta y ajustar los algoritmos de Minería de Datos.

En lo que respecta al aporte, se espera que los resultados de esta investigación contribuyan a minimizar los índices de deserción estudiantil universitaria en el área de informática de la UNNOBA.

V. Metodología utilizada.

La presente investigación se llevó a cabo a partir de un estudio retrospectivo – descriptivo y correlacional, caracterizado por un enfoque integrado multimodal, basado en la combinación y triangulación de técnicas cuanti y cualitativas. En lo que concierne al campo de la informática, se enmarca en lo que se conoce como proceso de Extracción de Conocimiento o KDD (Knowledge Discovery in Databases) el cual permite la detección automática de patrones existentes en la información disponible sin requerir una hipótesis especificada a priori. La aplicación de estas herramientas permitió la elaboración de un

análisis predictivo, sustentado y fundamentado en los resultados arrojados por estudios longitudinales del período 2005-2015.

La aplicación de técnicas cuantitativas se implementó a partir de la construcción de base de datos generadoras de fuentes y registros capaces de ser seriables y se hicieron encuestas masivas a estudiantes regulares y desertores de las diversas carreras del área de informática. Más allá de las encuestas, la información disponible provino, mayoritariamente, del sistema SIU-Guaraní de la UNNOBA. Este sistema permite el relevar, para cada alumno, un número importante de características relacionadas con su situación personal, académica y laboral. Otro insumo importante fueron los informes de ingreso a cada una de las carreras del área de Informática, del período 2011-2016. Mediante la aplicación de una de las tareas descriptivas que provee la minería de datos, el agrupamiento (clustering), se procuró obtener grupos o conjuntos de datos, de manera que los elementos asignados al mismo grupo fueran similares. Así se construyó un perfil cognitivo del estudiante basado en los datos filtrados por técnicas de Minería de Datos Educativa. Una vez construida la base de datos, teniendo en cuenta estudios precedentes, se elaboró una serie de variables sobre las causas de deserción.

En lo concerniente a lo cualitativo, identificados los grupos de estudiantes que abandonaron la Universidad, se construyó y aplicó masivamente – vía e-mail o telefónica – una serie de cuestionarios semi-estructurados y se realizaron entrevistas a profundidad a alumnos de los distintos años de las carreras con el fin de efectuar una exploración y descripción densa de las principales problemáticas asociadas a la deserción, desde la perspectiva de los propios actores estudiados.

Luego de comparar los niveles de deserción de cada año de las carreras del área de informática de la UNNOBA, con el fin de determinar cuál presentaba los índices más altos de abandono estudiantil, se determinó que los niveles de deserción más críticos se presentan en el primer año de las carreras analizadas.

Para la construcción de la variable dependiente regular/desertor, se consideró regular a todo aquel alumno que al siguiente cuatrimestre continúa cursando sus estudios universitarios normalmente, y se consideró desertor al alumno que al siguiente cuatrimestre presentó su renuncia voluntaria a la carrera, se retiró de ella sin previo aviso, se cambió a

otra carrera o fue eliminado por la universidad debido al no cumplimiento de exigencias académicas mínimas.

La unidad de tiempo escogida es el cuatrimestre, debido a que corresponde al periodo lectivo mínimo existente para medir la permanencia de un estudiante al interior de la universidad.

CAPÍTULO 1: DESERCIÓN ESTUDIANTIL UNIVERSITARIA.

1.1 CONCEPTUALIZACIÓN

El vocablo **deserción** proviene del latín “desertio”, que significa abandono-abandonado. Aplicado al ámbito educativo, se entiende por deserción al abandono de la formación académica en el sistema de educación formal, independientemente de las condiciones y modalidades de presencialidad, debido a causas endógenas y exógenas al sistema mismo (Páramo y Correa, 1999). Algunos autores distinguen entre una deserción voluntaria y una involuntaria. “La deserción voluntaria puede adoptar la forma de renuncia a la carrera por parte del estudiante o del abandono no informado a la institución de educación superior. La deserción involuntaria, en cambio, se produce como consecuencia de una decisión institucional, fundada en sus reglamentos vigentes, que obliga al alumno a retirarse de los estudios” (Himmel, 2002:95).

Para otros autores, en cambio, la deserción implica siempre una decisión personal del sujeto y no obedece a un retiro académico forzoso (por no acreditar debido a un bajo rendimiento académico) o a un retiro por asuntos disciplinarios. Esto lleva a diferenciar la deserción de otros fenómenos, también inherentes al sistema educativo, como la mortalidad estudiantil, el rezago, el ausentismo y el retiro forzoso (Tinto, 1989; Páramo y Correa, 1999).

Si bien la deserción estudiantil es un fenómeno presente en todos los niveles del sistema educativo, en este trabajo nos concentraremos específicamente en la deserción en el nivel superior universitario.

Los sistemas educativos de América Latina y el Caribe han tomado y utilizan la definición de deserción elaborada por el Instituto Colombiano para el Fomento de la Educación Superior (ICFES):

La deserción está dada por la cantidad de estudiantes que abandona el sistema de educación superior entre uno y otro período académico (semestre o año). La deserción se calcula como el balance entre la matrícula total del primer período, menos los egresados

del mismo período y más los alumnos reintegrados en el período siguiente. (IESALC, 2006:158)

Esta definición da cuenta de una perspectiva de análisis operativa que permite arribar a una cuantificación, pero no posibilita profundizar en las conductas y en las posibles causas del fenómeno. Para ello se requiere de la consideración de otras variables de corte cualitativo.

En este sentido, actualmente existe un consenso académico en definir la deserción estudiantil universitaria como un abandono voluntario que puede ser explicado a partir de diferentes categorías de variables: socioeconómicas, individuales, institucionales y académicas (Páramo y Correa, 1999; Díaz, 2008). El peso que se le adjudica a cada una de estas variables como determinantes de la deserción ha variado según se aborde el estudio de esta problemática desde un paradigma funcionalista o dialéctico. Hasta los años 1970 primó una perspectiva funcionalista que explica la deserción poniendo el foco en los estudiantes y no en las instituciones; de esta manera, se hace hincapié en las variables individuales, en las habilidades, talentos y actitudes de los estudiantes como factores determinantes de la permanencia o abandono de los estudios. Posteriormente comenzaron a cobrar cada vez mayor relevancia los estudios posicionados en un paradigma dialéctico que explican causalmente la deserción estudiantil como el resultado de las contradicciones de los diferentes subsistemas (político, económico y social) que inciden y determinan la decisión de desertar del individuo.

Algunos autores (Diconca, Dos Santos y Egaña, 2011) conciben a la deserción como un “proceso” de abandono. El hecho de poner énfasis en la idea de “proceso” permite ponderar una serie de variables o situaciones que, prolongadas en el tiempo, derivan en el abandono de los estudios. Al mismo tiempo, permite poner en relación la deserción con el fenómeno del rezago académico, entendiendo a este último como antesala del abandono.

Se ha elaborado también una clasificación de la deserción de acuerdo a otras dos variables: la del espacio y la del tiempo en que se deserta (Castaño, Gallón, Gómez y Vásquez, 2004). Con respecto al tiempo, se distinguen los siguientes tipos de deserción:

- a) **Deserción precoz:** cuando se produce la inscripción o admisión de un estudiante pero éste no se matricula.
- b) **Deserción temprana:** cuando se produce el abandono en los primeros semestres de la carrera.
- c) **Deserción tardía:** en los casos en los que se produce el abandono en los últimos semestres de la carrera.

Aquí los estudios sobre el tema presentan discrepancias en torno al número de periodos que se deben tomar como parámetro para considerar que un estudiante universitario se encuentra en estado de deserción académica. Así, mientras algunos consideran que se deben tener en cuenta como mínimo dos semestres consecutivos (Spady, 2006), otros autores ponderan un piso de tres semestres consecutivos (Tinto, 1982; Díaz, 2008). Mayoritariamente, las estadísticas universitarias que miden y analizan la deserción, suelen tomar el piso de dos semestres consecutivos (Viale Tudela, 2014).

En lo que refiere al espacio, se puede distinguir entre:

- a) **Deserción interna o del programa académico:** se refiere al estudiante que decide cambiar su programa académico por otro que ofrece la misma institución universitaria.
- b) **Deserción institucional:** es el caso en el cual el estudiante abandona la universidad para matricularse en otra.
- c) **Total deserción del sistema educativo:** en los casos de abandono definitivo de la formación académica.

Los dos primeros casos referirían a un tipo de **deserción relativa**, mientras que el último a un tipo de **deserción absoluta**, aludiendo al abandono definitivo del sistema universitario.

En cuanto a indicadores y metodología para calcular la deserción, Osorio y Jaramillo (1999) han tenido en cuenta los siguientes:

- **Índices de deserción por semestre:** relación entre el número total de alumnos desertores de un programa en un período y el número total de estudiantes matriculados en dicho programa para el mismo período.

- **Índices de deserción por cohorte:** diferencia entre el número de estudiantes que ingresan a la cohorte en un período y la cantidad de ellos que se matriculan en el mismo período.
- **Índices de deserción promedio por nivel:** promedio simple de los índices de deserción por semestre calendario, calculados como el número total de desertores de cada nivel sobre el total de matriculados en dicho nivel del programa.
- **Tasa ponderada de deserción por nivel:** muestra la expectativa de deserción para el programa. Se calcula ponderando la tasa de deserción con el promedio de la distribución de la población matriculada, en los semestres de duración de la carrera.

Gran parte de los estudios empíricos sobre el tema concuerdan en que los mayores índices de deserción se presentan durante el transcurso del primer año de las carreras (Tinto, 1989 y 2004; Corominas, 2001; Himmel, 2002; Vivas, 2005) o en la fase postrera del primer año de estudios y antes del comienzo del segundo (Tinto, 1989; Rivera, E. et. al., 2005). Algunos de ellos atribuyen las causas a las dificultades que encuentran los estudiantes para establecer contactos con la comunidad social e intelectual de la institución y lograr amistades en ella. De hecho, algunos de estos estudios han demostrado que las mencionadas dificultades tienden a ser mayores en la etapa temprana de la carrera que en los últimos años (Tinto, 1989). Según Corominas (2001)

las deserciones se producen mayormente en primero, debido a elecciones inadecuadas de la carrera a elegir, bajo rendimiento académico, no haber aprobado el mínimo de créditos, asignaturas poco motivadoras, y poco esfuerzo y compromiso con el estudio (p.130)

Estos planteos coinciden con los resultados arrojados por una encuesta realizada por InterUniversidades.com, la red social para estudiantes y universidades de todo el mundo. Fueron encuestados más de 5000 jóvenes que en algún momento pasaron por las aulas de la universidad; de ellos el 58 % abandonó la carrera durante el primer año de estudios.

1.2 DESERCIÓN ESTUDIANTIL UNIVERSITARIA Y ESTRATEGIAS DE ABORDAJE EN DETERMINADOS PAÍSES IBEROAMERICANOS.

Si bien la deserción es un fenómeno inherente a la vida estudiantil, en muchos países del mundo occidental la tasa de deserción se ha empezado a considerar como un indicador de la calidad de la gestión universitaria, más aún cuando agencias nacionales e internacionales están presentes en el proceso de certificación de la calidad de las instituciones y planes de estudios universitarios. Esto convierte a la deserción en un problema a resolver, o por lo menos morigerar, de manera apremiante.

En la región de América Latina y el Caribe el rezago escolar, el fracaso y la deserción son problemas de elevada envergadura. A pesar del incremento cuantitativo de matriculación que se ha dado en los últimos veinte años, menos del 20% de los adultos mayores de 25 años ha accedido a la universidad y menos de un 10% ha completado sus estudios universitarios (Informe sobre la Educación Superior en América Latina y el Caribe 2000-2005). Según este informe, “el incremento previsible de la población estudiantil universitaria ocurrirá en los quintiles de menores ingresos los cuales, debido a su menor capital cultural, en el contexto de la cultura dominante, debiera tender a que la repitencia y la deserción universitaria aumenten en los próximos años si no se establecen políticas y estrategias adecuadas” (Informe sobre la Educación Superior en América Latina y el Caribe 2000-2005: 159).

En lo que refiere al abandono de estudios en la región, los índices se sitúan en valores cercanos al 50 % en Argentina y México, mientras que en Chile y Perú oscilan alrededor de un 30 % (García de Fanelli y Jacinto, 2010). En el caso español las estadísticas universitarias reflejan un porcentaje de abandono cercano al 50% en algunas titulaciones (Cabrera, L, Bethencourt, J, *et.al*, 2006)

Latiesa (1998) al estudiar la deserción en diversas carreras del sistema universitario español concluye que el factor más común, es el bajo rendimiento académico por parte de los estudiantes, que no están lo suficientemente preparados para enfrentarse al mundo universitario, por deficiencias de conocimientos en ciencias básicas.

Estrategias de abordaje

Si bien los estudios comparativos sobre el fenómeno de la deserción, tanto en lo conceptual como en la investigación empírica, son aún incipiente en la región, en general, las universidades están implementando diversas estrategias a fin de favorecer la retención.

El Instituto Internacional de la UNESCO para la Educación Superior en América Latina y el Caribe (IESALC) plantea la necesidad de realizar estudios en toda la región, tendientes a analizar los factores causales de la deserción y el rezago estudiantil y a plantear estrategias para la búsqueda de solución al problema.

A nivel regional, en uno de los Informes elaborados por este organismo, se plantea una propuesta que contempla tres niveles yuxtapuestos de acción:

- A nivel de sistema de educación superior se propuso, entre los más relevante, profundizar en el diagnóstico, realizando estudios nacionales y generando esquemas básicos de medición.
- A nivel institucional y académico: mejorar los mecanismos de detección temprana, identificar grupos de riesgo, otorgar apoyo tutorial integral al estudiante, mejorar la orientación vocacional, realizar seguimiento estudiantil.
- A nivel pedagógico: crear redes de apoyo; considerar los estilos de aprendizaje, la atención, incorporar cursos de nivelación para los estudiantes que no aprueban los exámenes de ingreso y cubrir el desfase del bachillerato en materias fundamentales cuyos contenidos son deficientes; establecer ciclos generales de conocimientos básicos, establecer innovaciones curriculares (perfiles y enseñanza por competencias y fortalecer la metodología de resolución de problemas); establecer currículos más flexibles. (Informe sobre la Educación Superior en América Latina y el Caribe 2000-2005: 164).

García Fanelli (2005) en un estudio sobre acceso, deserción y graduación en la educación superior argentina, plantea que las estrategias para disminuir la deserción

escolar deben ser proyectadas e implementadas “en dos niveles diferentes pero complementarios de análisis e intervención: el macro nivel del sistema educativo en su totalidad y el micro nivel de las organizaciones de enseñanza superior” (2005:12). En lo que respecta a un nivel macro, se debe visualizar que el proceso educativo es acumulativo y los problemas de aprendizaje y de socialización se arrastran de un nivel a otro. Esto implica programar políticas de articulación con el nivel secundario que favorezcan la trayectoria de los egresados y su inserción en el ámbito universitario. En algunos países, tales como Estados Unidos y Canadá, esta tarea está en manos de las instituciones de nivel superior conocidas como “community colleges”, las cuales ofrecen a los estudiantes cursos y talleres especialmente diseñados para brindar capacitación para leer, escribir y operar con números de manera que el egresado de nivel medio pueda continuar sus estudios universitarios con menor dificultad.

Algunos países han incorporado lo que se ha denominado “observatorio de graduados”, que permiten conocer las condiciones de inserción ocupacional, nivel promedio de remuneraciones y trayectoria laboral en las principales profesiones. Ello se puede complementar con el seguimiento de los graduados en su pasaje al mundo laboral por parte de las instituciones universitarias. Esto último es además un muy buen insumo para la actualización y mejora permanente de la calidad de los planes de estudio (Fanelli, 2005).

En lo concerniente al micro nivel de las instituciones universitarias, se propone implementar políticas de orientación vocacional a los estudiantes ingresantes; planes de estudio flexibles que brinden una formación general en los primeros años y que faciliten el pasaje entre carreras de disciplinas afines; cursos, talleres y tutorías orientados a estudiantes con dificultades de aprendizaje. Cabrera et.al. (2006) señalan que debido al retraso y abandono de los estudios, se implementaron los cursos cero (conocidos también como cursos de nivelación) por algunas universidades.

En Colombia, el Ministerio de Educación Nacional ha puesto en funcionamiento la plataforma SPADIES (Sistema de Prevención y Análisis a la Deserción en las Instituciones de Educación Superior) que recoge información socioeconómica y académica de los estudiantes de diferentes IES y permite establecer relaciones con datos sobre deserción. A partir de allí se sugiere implementar estrategias no sólo de tipo financiero en las

universidades, sino también de tipo psicológico, institucional y académico. Algunos investigadores han planteado “continuar con las tutorías relacionadas con las materias en las que se ha visto más afectado el estudiante; durante los primeros semestres, propiciar oportunidades para la orientación frente a las políticas, beneficios y servicios de la universidad y en cuanto a su vocación profesional” (Ariza y Marín, 2009).

También en Colombia, el Grupo de Investigación en Estadística Multivariada - GIEM- con sede en la Universidad Tecnológica de Pereira ha desarrollado un Sistema de Alertas Tempranas (SAT) que permite identificar y pronosticar, a partir de la combinación de variables utilizadas en diversos modelos explicativos referidos al fenómeno de la deserción educativa, los casos de aquellos estudiantes en riesgos de deserción (Carvajal Olaya et.al,)

1.3 LA PROBLEMÁTICA DE LA DESERCIÓN ESTUDIANTIL UNIVERSITARIA EN ARGENTINA.

En Argentina, al igual que en el resto del mundo, durante la segunda mitad del siglo XX se produjo una amplia expansión de la educación superior, tanto universitaria como no universitaria. La educación superior universitaria ha presentado un crecimiento sostenido a lo largo de este tiempo, llegando a contar para el año 2013 con una matrícula de 1.830.743 alumnos (Anuario Estadísticas Universitarias, 2013). No obstante ello, estas altas tasas de cobertura del sistema universitario no se reflejan en las tasas de graduación, de hecho solo 3 de cada 10 estudiantes logra concluir sus estudios, presentándose los índices más elevados de deserción durante el primer año de las carreras. Alrededor de un 40 % de los estudiantes abandonan su carrera en primer año, un porcentaje menor pero todavía importante, lo hacen en el segundo año. Algunos de esos estudiantes cambian de carrera y la mayoría abandona sus estudios. Sumado a ello, se ha demostrado que la deserción impacta mayoritariamente en el 40% de jóvenes con menor ingreso per cápita familiar (García de Fanelli, 2005; Donoso y Schiefelbein, 2007).

Existe en la actualidad una amplia literatura proveniente, en su mayoría, del campo de las ciencias sociales y de la educación tendiente a analizar el fenómeno del desgranamiento y deserción universitaria de grado en el país (García de Fanelli, 2005; Parrino, 2005). No obstante ello, no se ha realizado hasta el presente un análisis integrador

que permita sacar conclusiones generales sobre el estado del conocimiento en el campo (García Fanelli, 2014). En las universidades argentinas, las estadísticas muestran que en términos generales, de cada 100 alumnos que ingresan desertan poco más de ochenta – el valor medio de deserción en el sistema universitario se sitúa en el 82.2 %. Un informe de la SPU del año 2013 estipulaba que de cada 100 inscritos en 20 carreras seleccionadas en las universidades nacionales, sólo se graduaron aproximadamente 19 dentro del plazo de duración normal (Informe SPU 2013).

Contando con el apoyo de la Secretaría de Políticas Universitarias, el Consorcio SIU¹ desde hace más de una década, viene dotando “a cada una de las Instituciones Universitarias que componen el sistema universitario nacional argentino de aplicaciones informáticas que colaboren a mejorar la gestión y la calidad de los datos que día a día se producen en su ámbito” (SIU, 2017). Al mismo tiempo, en relación con la problemática de la deserción, viene realizando una serie de estudios empíricos que, sin partir de hipótesis predeterminadas, intentan encontrar patrones de comportamiento en forma automática en bases de datos de los sistemas de gestión académica de las Universidades.

En lo que respecta a nuestro país, algunas universidades se han valido del uso de algoritmos TDIDT, a partir de la información disponible en el sistema SIU-Guaraní, con el fin de identificar las variables que más inciden en la deserción (Kuna, García Martínez y Villatoro, 2009). Otras, como es el caso de la UNMDP y la UTN, en el ámbito de la Facultad de Ciencias Económicas y Sociales, desarrollaron un “Programa de Retención de Matrícula” basado en la elaboración de Biografías Educativas, que pretende generar información sustantiva para la toma de decisiones por parte de los actores de la Universidad, e implementando sistemas tutoriales que han logrado evitar el desgranamiento de la matrícula (Arana y Bianculli, 2006).

Otro caso interesante es el que presenta la Universidad de Tres de Febrero (UNTREF). En ella, desde el Departamento de Ciencias Sociales se desarrolló un Proyecto de Investigación destinado a estudiar las causas de la deserción en las diversas carreras dictadas por la universidad, y diseñar estrategias de intervención que contribuyan

¹ El SIU es un consorcio conformado en la actualidad por 46 Instituciones de Educación Superior Universidades Nacionales de la Argentina, que tiene como objetivo desarrollar soluciones informáticas y brindar servicios para el sistema Universitario Nacional.

a prevenir o minimizar esta problemática. Los resultados de la investigación demostraron que en el caso de la UNTREF, una universidad de relativamente reciente creación -ya que las primeras cohortes datan de 1999 - los factores extrasistémicos, es decir, externos al sistema académico - origen sociocultural del estudiante, situación laboral, preferencias personales y hábitos culturales, distribución de la población estudiantil por sexo y edad - son determinantes en la deserción. Frente a estos resultados, desde la universidad se reforzaron, en algunos casos, y pusieron en práctica, en otros, una serie de estrategias destinadas a minimizar la deserción. Entre esas estrategias, cabe mencionar las siguientes: implementación de cursos de ingreso y de un programa de tutorías, disposición de grupos pequeños por aula, desarrollo de un programa de mejoramiento administrativo y de gestión (Lorenzano, 2004).

1.4 DESERCIÓN ESTUDIANTIL UNIVERSITARIA Y ESTRATEGIAS DE ABORDAJE EN LAS CARRERAS DEL CAMPO DE CIENCIA APLICADA EN ARGENTINA

En Argentina el bajo rendimiento académico en las universidades se ve mayormente en las carreras técnicas, tal es el caso de las carreras de ingeniería o del área de ciencia aplicada (Drovandi, 2010).

En lo que respecta a las carreras de grado relacionadas con el campo de la Ciencia Informática, las estadísticas de los últimos años han estimado que alrededor de un 80% de los estudiantes desertan durante el transcurso de los primeros dos años de las carreras. De hecho se ha demostrado que prevalece una deserción temprana en el primer año de las carreras, cercana al 50 % (Madoz y Gorga, 2006).

Algunos estudios han demostrado la existencia de un denominador común en los alumnos que abandonan al segundo año: el no haber regularizado materias en el primer año de cursada. Como contraparte, la mayoría de alumnos que regulariza más de una materia continúa sus estudios (Kuna, García Martínez y Villatoro, 2010)

En la Argentina actual, el sector Software y Servicios Informáticos (SSI) es uno de los principales demandantes de empleo calificado. Sin embargo, la matrícula de las carreras universitarias no acompaña esta demanda y resultan insuficientes los profesionales informáticos (Dapozo *et.al*, 2014); estadísticamente se ha demostrado que se recibe un

alumno de cada cuatro que ingresan (Drovandi, 2010). Ya en 2013 Santiago Ceria, Director Ejecutivo de la Fundación Sadosky del Ministerio de Ciencia, Tecnología e Innovación Productiva, planteaba que "Según la Cámara de Empresas de Software y Servicios Informáticos se necesitan más de 7000 profesionales al año del área de informática, pero se están graduando 3500". Incluso, según los datos de la misma CESSI, para 2010, un 31% de la mano de obra especializada en informática abandonó la carrera universitaria.

Sumado a ello, la problemática se agrava y se halla atravesada por una cuestión de género: se observa en todas las carreras relacionadas con la informática una prevalencia masculina, con una tendencia decreciente del porcentaje de mujeres en estas carreras en particular, y en el sector de Software y Servicios Informáticos (SSI), en general. La baja matriculación en este tipo de carreras no se condice con la oferta de las mismas ya que se contabilizan en todo el país un total de cuarenta universidades públicas y 15 privadas que dictan diversas carreras afines a la ciencia informática.

Para afrontar la problemática del desgranamiento y deserción en las carreras de Informática, la mayoría de las universidades han implementado diversas estrategias tanto de monitoreo y seguimiento del rendimiento académico de los alumnos de cursos avanzados, como acciones preventivas, atendiendo a los exámenes de ingreso, para poder detectar potenciales grupos de riesgo y trabajar con ellos. En este sentido, dos antecedentes relevantes lo constituyen la Universidad Nacional de La Plata y la Universidad Tecnológica Nacional. Desde hace más de una década, la Facultad de Informática de la UNLP viene trabajando en el desarrollo y promoción del empleo de herramientas tecnológicas de Educación a Distancia en la articulación Escuela Media-Universidad. Ello ha favorecer la retención de los alumnos en las etapas tempranas de las carreras universitarias (Madoz, Gorga y De Giusti, 2005; González, Madoz y Gorga ;Gorga y Madoz, 2011).

En el caso de la Universidad Tecnológica Nacional, la Facultad Regional General Pacheco, ante la problemática de encontrarse con una tasa de deserción del 40% en el primer año y un promedio de duración de la carrera de más de 10 años, atribuidas sobre todo a variables socioeconómicas y laborales del alumnado, implementó como alternativa el aprendizaje a distancia bajo la modalidad blended-learning a través del desarrollo de un aula de educación virtual (Casanovas, Benegas, Fernández; 2009).

La problemática del desgranamiento y la deserción está también presente en las Carreras de Informática de la Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA).

CAPÍTULO 2: DATA MINING Y EDUCATIONAL DATA MINING.

2.1 Inteligencia de Negocios o Business Intelligence (BI). Su aporte en torno a Minería de Datos y Pentaho.

El elevado nivel de deserción estudiantil, es uno de los problemas principales que enfrentan las instituciones de educación superior de América Latina y el Caribe. Se estima que en la Argentina alrededor del 80% de los alumnos que ingresan a la universidad desertan (Valía, Rostagno, Moine, Bigatti, Riva, y Amar, 2017).

Cada día, las instituciones de educación superior, generan una gran cantidad de datos personales, académicos y socioeconómicos de los estudiantes (Amaya, Barrientos y Heredia, 2015). El contexto de la sociedad de la información ha propiciado la necesidad de tener mejores, más rápidos y más eficientes métodos para extraer y transformar los datos de una organización en información y distribuirla a lo largo de la cadena de valor (Porter, 2008).

La Inteligencia de Negocio o Business Intelligence, en adelante BI, juega un papel fundamental para la solución del problema descrito anteriormente debido a que permite orientar los procesos, tecnologías y herramientas para convertir los datos en información y éstos en conocimiento con el objeto de mejorar la toma de decisiones en pro de aumentar la eficiencia y calidad de los procesos de negocio.

El concepto BI tan de moda en la actualidad y crítico para la mayoría de las organizaciones no es nuevo. Ya en 1958, Hans Peter Luhn, investigador de IBM, acuñó el término en el artículo “A Business Intelligence System”, donde lo define como *“la habilidad de aprehender las relaciones de hechos presentados de forma que guíen las acciones hacia una meta deseada”*.

Sin embargo, no es hasta el año 1989 en que Howard Dresner, analista de Gartner Group, populariza y propone una definición formal para este concepto: *“Conceptos y métodos para mejorar las decisiones de negocio mediante el uso de sistemas de soporte basados en hechos”*.

En la actualidad el concepto de BI incluye una amplia categoría de metodologías, aplicaciones y tecnologías que permiten reunir, acceder, transformar y analizar los datos,

transacciones e información no estructurada (externa e interna), con el propósito de ayudar a los usuarios de una organización a tomar mejores decisiones de negocio. Esto se puede lograr, bien sea mediante la explotación directa (consulta, reportes, etc.) o haciendo uso del análisis y conversión de conocimiento.

El BI puede intervenir en todos los procesos de una organización, actuando en las tareas y actividades de los empleados, mejorando la comunicación entre departamentos e incrementando la capacidad de reacción de la misma. Cabe aclarar que para desarrollar una adecuada y efectiva BI, no es suficiente la utilización de metodologías y la implantación de tecnología de punta. Sí es importante poseer un conocimiento profundo de todos los factores que afectan a la organización, tanto externos como internos. En la *Figura 1* puede observarse el ámbito integral del entorno BI.

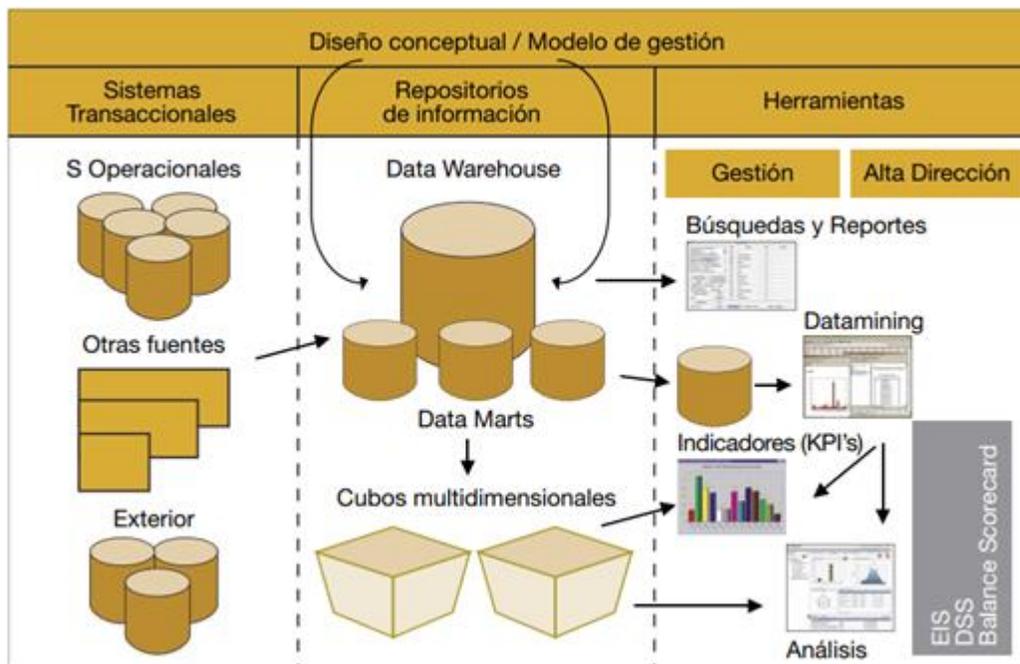


Figura 1. Modelo integral de una solución BI, (PWC, 2008).

Algunas de las tecnologías que forman parte del BI son:

- Data Warehouse.
- Reporting.
- Análisis OLAP (On Line Analytical Processing).
- Análisis visual.

- Análisis predictivo.
- Dashboards.
- Cuadro de mando integral.
- Minería de datos.
- Gestión del rendimiento.
- Previsiones.
- Reglas de negocio.
- Integración de datos (que incluye ETL, Extract, Transform and Load).

La *Figura 2* muestra las distintas tecnologías y aspectos que forman parte de la inteligencia de negocio.

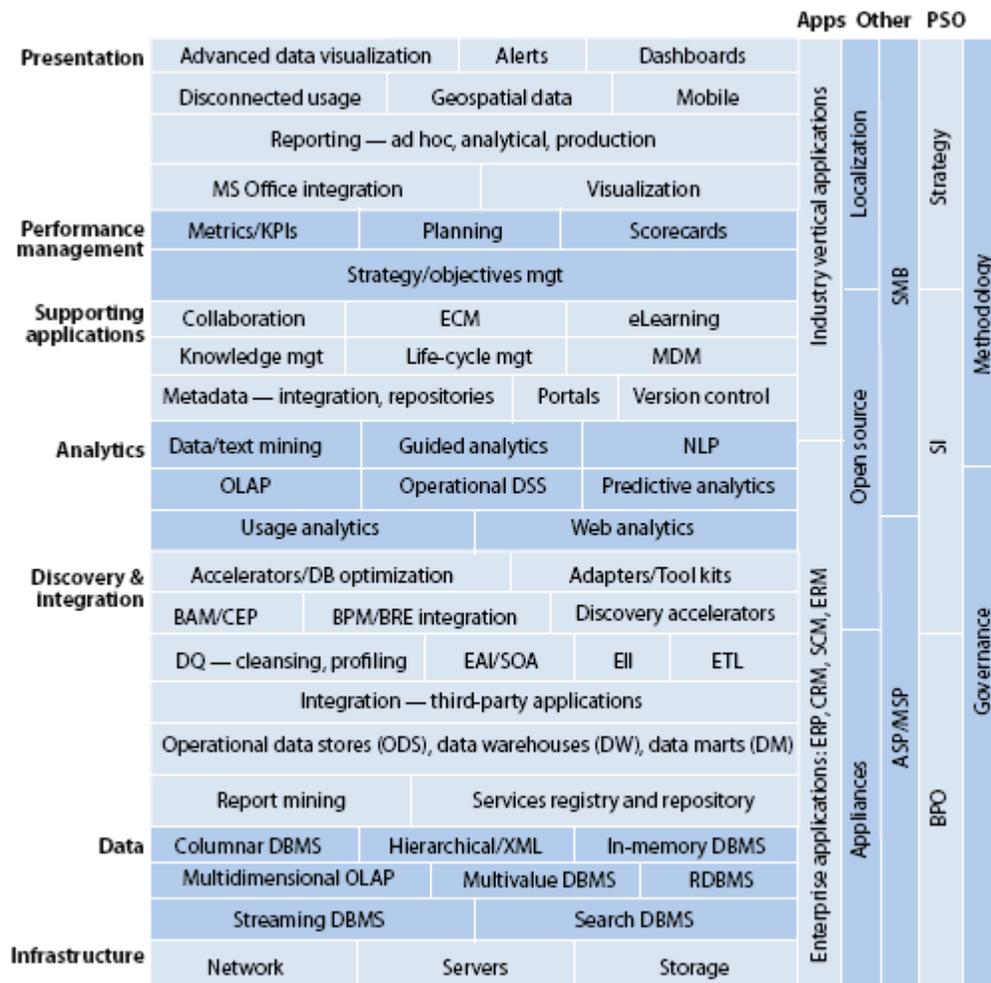


Figura 2. Business Intelligence Stack, (Forrester Research, 2007).

El BI puede llegar a convertirse en una ventaja competitiva fuerte, ya que proporciona conclusiones robustas para la toma de decisiones. Contar con un esquema de este tipo aporta ventajas tales como disponer de la información correcta en el momento adecuado para la toma de decisiones, proveer la capacidad de evaluar distintos escenarios al mismo tiempo, generar capacidad de reacción ante situaciones imprevistas con un nivel de riesgo menor, agrupar la información de distintas áreas en un único repositorio.

El proceso de BI garantiza un ciclo de cuatro fases, que permite generar información analítica constante, ver *Figura 3*. Este proceso dinámico e iterativo empieza con preguntas, y las respuestas dan origen a más preguntas o subsecuentes interacciones del proceso.

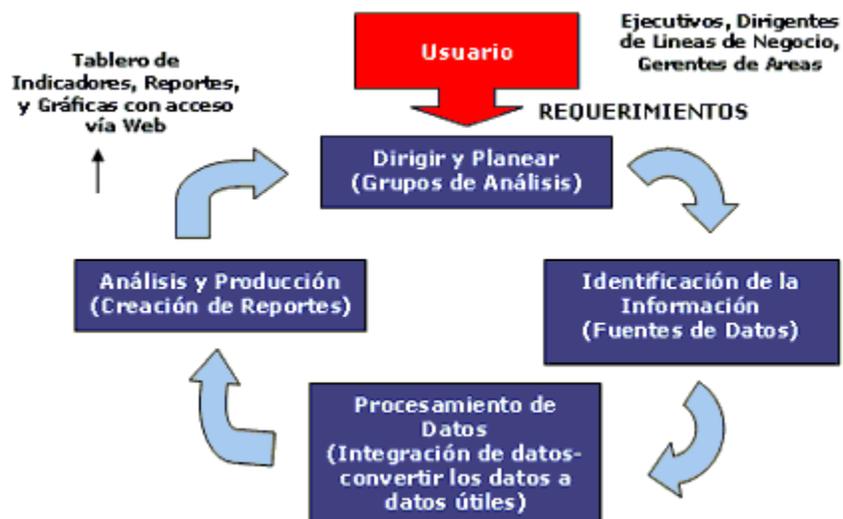


Figura 3. Business Intelligence Process, (LeeWittschen, 2004).

La **fase dirigir y planear**, es el principio y fin del proceso. Es el principio porque involucra redactar los requerimientos específicos y es el final porque contesta preguntas que guían a otros nuevos. El proceso de BI comienza con los usuarios y es aquí donde se generan las preguntas que les ayudará a ellos a alcanzar sus objetivos. El análisis de estas preguntas produce la redacción de necesidades que se plasmarán en una solución BI. Así mismo al finalizar un ciclo del proceso de BI, se obtiene la respuesta a preguntas que estarán sujetas al análisis por parte de los usuarios y que servirán de apoyo para la toma de decisiones.

Una vez formulados los requerimientos de los usuarios y establecido un plan para la recolección de la información que permita dar respuesta a las preguntas planteadas, se comienza con la **fase de identificación de la información**. Dentro de una organización existen diversas fuentes o sistemas de información los cuales crean, procesan y almacenan

diferentes tipos de información. Este es un proceso continuo y es importante entender que los datos de esas fuentes son simplemente información y no inteligencia. Los datos en crudo frecuentemente son incompletos y confusos. La información se convierte en inteligente a través del procesamiento y análisis de la misma. El proceso de recolección de información pone el énfasis en la recolección e identificación de los datos, es decir, en saber dónde encontrar la información que permita dar respuesta a las preguntas planteadas en la etapa anterior.

La **fase de procesamiento de datos** tiene por objetivo la integración de los datos a un formato reutilizable para el análisis. En esta etapa se analiza, diseña y determina en profundidad los datos relevantes, las transformaciones a realizar, y las estructuras de destino como es un Data Warehouse. Esta fase, conocida generalmente como ETL (Extracción, Transformación y Carga), es considerada la base a partir de la cual se implementará el BI.

Una vez finalizada la etapa de procesamiento de datos tiene lugar la **fase de análisis y producción**. El grupo de análisis de la organización utiliza herramientas y técnicas sobre los datos para crear inteligencia. El resultado final es la producción de respuestas “inteligentes”, en un contexto propio. En algunos casos es un proceso simple como la creación de un reporte, en otros, es la creación de indicadores. Tal vez en esta fase, sean generados requerimientos adicionales si los analistas encuentran nuevas preguntas que necesitan ser contestadas.

El proceso de BI finaliza con la **etapa de difusión**, que consiste en entregar productos inteligentes a los distintos usuarios que lo requieren. Esta tarea implica el uso de herramientas BI para la publicación de “tableros de indicadores”, reportes o la posibilidad de tener herramientas de fácil uso para que los mismos usuarios tengan la capacidad de revisar los datos de manera rápida y sencilla (GestioPolis, 2004).

Para que la inteligencia de negocios se lleve a cabo de una forma adecuada se necesita una serie de aplicaciones específicas las cuales se conocen como herramientas BI. Principalmente, se trata de un software donde se apila un gran número de aplicaciones que permiten analizar los datos recogidos y presentarlos de diferentes formas. Con estas herramientas los múltiples usuarios podrán acceder a dicha información sobre el negocio y elaborar informes, analizarlos, crear alertas, entre otras funciones.

Las principales empresas del sector BI son:

- Microsoft (Microsoft, 2017).
- Oracle (Oracle, 2017).
- MicroStrategy (MicroStrategy, 2017).
- IBM Cognos (IBM, 2017).
- Information Builders (Information Builders, 2017).
- SAS (SAS, 2017).
- QlikTech (QlikTech, 2017).
- SAP Business Objects (SAP Business Objects, 2017).
- Teradata (Teradata, 2017).
- Tableau (Tableau, 2017).
- Netezza (Netezza, 2017).
- Pentaho (Pentaho, 2017).

Según Sallam, Richardson, Hagerty y Hostmann (2011), las suite de BI de los proveedores líderes son complejas, caras, y presentan importantes deficiencias en cuanto a rendimiento y usabilidad. Reconocen que los “líderes” no salen muy bien parados en ninguna de las encuestas. Ni SAP, ni IBM, ni Oracle, ni casi nadie pueden estar conformes con estos resultados.

En los últimos años, el mercado del BI se ha enriquecido con soluciones open source que abarcan todo el espectro de necesidades de una organización para la explotación de la información. Algunas de estas herramientas tienen ya varios años en el mercado y se encuentran respaldadas por importantes organizaciones. Es posible encontrar herramientas solventes y maduras desde el nivel de bases de datos hasta el de minería de datos.

La adaptabilidad a todos los entornos, las actualizaciones y mejoras frecuentes, la tendencia a la calidad y utilización de estándares, la independencia del proveedor, sumado a la seguridad y al hecho de no tener que pagar una licencia de software (Smail, 2010), son algunos de los factores por los cuales conviene optar por herramientas BI open source. Existe una gran cantidad de herramientas Open Source Business Intelligence (OSBI) presentes en el mercado actual por lo que se hace necesario diferenciar cada uno de sus componentes y características con el objeto de elegir aquella que mejor se adapte a los

requerimientos del presente proyecto.

En el estudio realizado por Gerolami, Revello y Venzal (2011), se evaluaron herramientas de soporte ETL, operaciones y componentes para análisis OLAP, herramientas de reporting, cuadros de mando (también conocidos como dashboard o tableros de control), seguridad, usabilidad y amigabilidad con el usuario, documentación y aspectos generales. En una primera aproximación al resultado final, las medidas de evaluación definidas y utilizadas por los autores fueron Nivel A, B, C, D y E donde el Nivel A representa el valor máximo esperado y el Nivel E el valor mínimo. El resultado de este estudio se refleja en la *Figura 4*.

Plataforma	Evaluación
Pentaho	A
JasperSoft	B
SpagoBI	A
OpenI	C
Palo	B

Figura 4. Comparación plataformas OSBI, (Gerolami, Revello y Venzal, 2011).

Un estudio más exhaustivo fue realizado por los autores sobre las plataformas Pentaho y SpagoBI. Se tuvieron en cuenta los niveles de facilidad en la instalación de la plataforma y proceso de configuración, el nivel de calidad y completitud de la documentación, nivel de accesibilidad y facilidad de uso de la herramienta. Tales características fueron evaluadas haciendo uso de una nueva métrica que utiliza los valores enteros de 1 a 5 donde a mayor valor mejor es la característica analizada en la plataforma referida y a menor valor peor es dicha característica, ver *Figura 5*.

Plataforma	1	2	3	4	5	Promedio
Pentaho	4	4	4	5	5	4,4
SpagoBI	5	2	3	2	3	3

Figura 5. Tabla comparativa final, (Gerolami, Revello y Venzal, 2011).

Como puede observarse, a través de los valores reflejados en las *Figuras 4 y 5*, Pentaho es la plataforma OSBI más adecuada para llevar adelante la implantación de una solución

BI. La solución OSBI Pentaho pretende ser una alternativa a las soluciones propietarias tradicionales más completas como son Business Objects, Cognos, Microstrategy, Microsoft, etc. por lo que incluye todos aquellos componentes que pueden encontrarse en las soluciones BI propietarias más avanzadas (Chaudhuri, Dayal Narasayya, 2011).

Suite Pentaho

Pentaho es una herramienta de BI desarrollada bajo la filosofía del software libre para la gestión y toma de decisiones empresariales. Basada en Java, se trata de una plataforma con diseño modular compuesta de diferentes programas que satisfacen los requisitos del BI. Ofrece soluciones para la gestión y análisis de la información, incluyendo el análisis multidimensional OLAP, presentación de informes, minería de datos y creación de cuadros de mando para el usuario (Stratebi, 2017; Díaz, 2012; Tuncer, van den Berg, 2010).

Existen en el mercado dos versiones de Pentaho, Pentaho Enterprise Edition (EE) y Pentaho Community (CE). La edición empresarial, que se obtiene a través de una suscripción anual, contiene funciones adicionales que no se encuentran en la edición comunitaria. La versión community, orientada al sector académico, ofrece los productos de Pentaho por separado, por lo que cada uno se debe descargar, instalar y configurar, tarea que no resulta tan sencilla. De igual forma la funcionalidad de estas no son muy limitadas, por lo que gracias a esta versión se tienen poderosas herramientas, para la distintas fases de la inteligencia de negocios, de forma gratuita y abierta (Integrando Datos, 2013). La oferta principal de Pentaho se mejora mediante productos complementarios, generalmente en forma de complementos, que ofrece la propia empresa y la amplia comunidad de usuarios (Openred, 2011).

Plataforma Pentaho

La arquitectura software de esta Suite resulta muy flexible en términos de implantación, ya que permite usar los componentes que el usuario precise y personalizar las soluciones BI dentro de una plataforma completa de desarrollo de BI.

Estos componentes llevan años consolidándose en el mundo del software libre y su uso se está extendiendo rápidamente, además la plataforma cuenta con un importante número

de asistentes y plantillas que facilitan la elaboración de informes por defecto.

Pentaho se sienta sobre las bases de un servidor de aplicaciones J2EE JBoss y JBoss Portal facilitando el acceso a la información a través de Internet o de la Intranet de la empresa usando un navegador Web.

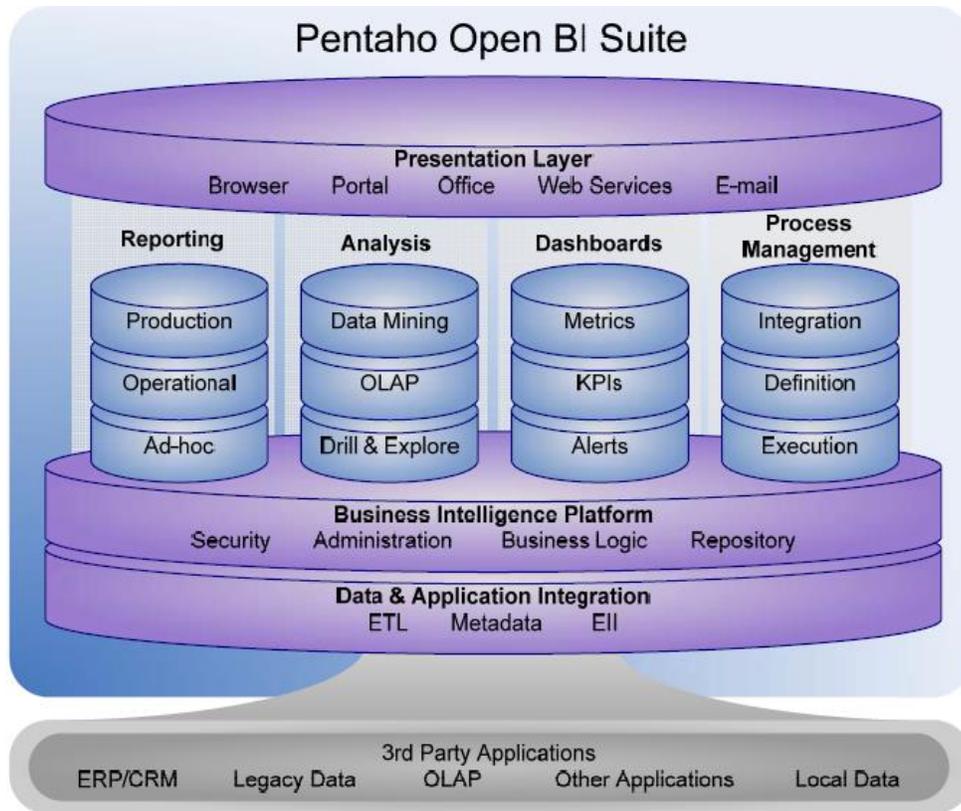


Figura 6. Business Intelligence Process, (LeeWittschen, 2004).

La suite de Pentaho, ver *Figura 6*, además de cumplir con las funcionalidades de una plataforma de BI, también incorpora una capa de integración de datos, una capa de aplicaciones y una capa de presentación de datos que se caracteriza por su usabilidad y posibilidades de acceso a los mismos. Teniendo en cuenta las tendencias del mercado actual resulta de interés destacar la presencia de la capa de integración.

La plataforma de Pentaho provee de servicios críticos incluyendo programación, seguridad, integración, automatización y flujo de trabajo (workflow). Proporcionando

habilidades a los usuarios finales de Pentaho y proveyendo un lugar central para administrar y mantener el despliegue de la organización BI. Entre los principales componentes se destacan:

- **Plataforma:** Pentaho BI es una solución realizada en Java de código abierto flexible y muy potente que cubre prácticamente todas las necesidades de una empresa. Creada con el 100% J2EE, asegura de esta forma la escalabilidad, integración y portabilidad.
- **Servidor:** Pentaho puede correr en servidores compatibles con J2EE como JBOSS AS, IBM WebSphere, Tomcat, WebLogic y Oracle AS.
- **Base de datos:** Vía JDBC, IBM DB2, Microsoft SQL Server, MySQL, Oracle, PostgreSQL, NCR Teradata, Firebird.
- **Sistema operativo:** No existe dependencia; lenguaje interpretado.
- **Lenguaje de programación:** Java, Javascript, JSP, XSL (XSLT / XPath / XSL-FO).
- **Interfaz de desarrollo:** Java SWT, Eclipse, Web-based.
- Repositorio de datos basado en XML
- Todos los componentes están expuestos vía Web Services para facilitar la integración con Arquitecturas Orientadas a Servicios (SOA).

A continuación, la *Figura 7* muestra la interacción entre los distintos componentes de Pentaho.

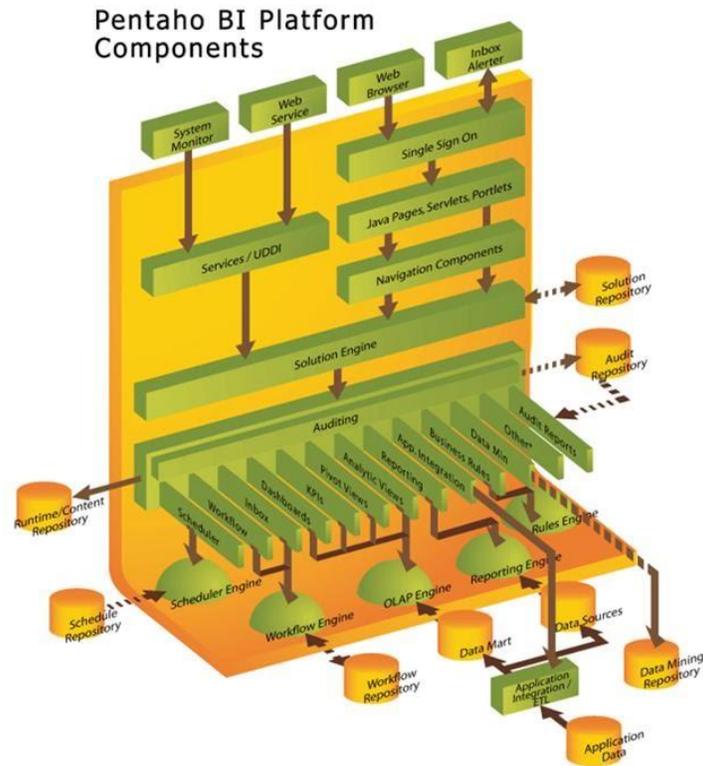


Figura 7. Pentaho BI Platform Components, (Gravitar, 2014).

Workflow de procesos de negocios

El workflow es un conjunto de actividades que abarca la ejecución coordinada de múltiples tareas desarrolladas por diferentes entidades procesadoras para llegar a un objetivo común (Caro, 2005).

Dos son los fundamentos del workflow de procesos de negocio en Pentaho: el motor de workflow Enhydra Shark y el estándar WPD, auspiciado por la WorkFlow Management Coalition (WFMC), organismo que declara tener más de 300 empresas asociadas, incluyendo a las conocidas IBM, Oracle, BEA, Adobe, SAP, TIBCO o SUN. Dentro del proyecto Enhydra es posible encontrar también Enhydra JaWE, un editor de workflow XPDL, según las especificaciones de WfMC.

Enhydra Shark es un framework de motor de flujo de trabajo extensible que incluye una implementación estándar completamente basada en especificaciones WfMC usando XPDL como formato de definición de proceso de flujo de trabajo nativo y la API "ToolAgents" de

WfMC para ejecución de actividades del sistema en el servidor (Enhydra Shark, 2011)

XPDL (XML Process Definition Language) es un lenguaje para la definición de un flujo de trabajo. Basado en XML puede ser usado para intercambiar modelos de procesos de negocio entre distintas herramientas. Se trata de un formato de archivo que representa el “dibujo” de la definición del proceso siendo el objetivo almacenar e intercambiar diagramas de procesos. Permite que una herramienta de diseño de procesos escriba un diagrama y otra lo lea, y la imagen resultante sea lo más parecida posible (Workflow Management Coalition, 2009).

Herramientas de la Suite Pentaho

La Suite Pentaho dispone de herramientas que permiten y facilitan el desarrollo de nuevas aplicaciones de BI. Estas herramientas son de código abierto y pueden ser descargadas de la página oficial de Pentaho al igual que la totalidad de la documentación correspondiente a cada una de ellas.

Pentaho Data Integration

Muchas organizaciones tienen información disponible en aplicaciones y base de datos separados. Pentaho Data Integration (PDI) o Kettle como era conocido anteriormente este proyecto belga antes de ser integrado en la Suite Pentaho, es una poderosa e importante herramienta de ETL (Extracción, Transformación, Transporte y Carga de datos) para un entorno de data warehousing. Kettle es un software libre bajo licencia LGPL (GNU Lesser General Public License) (Pentaho, 2007). El uso de Kettle permite evitar grandes cargas de trabajo manual frecuentemente difícil de mantener y de desplegar.

Conocido por su facilidad de uso y curva de aprendizaje rápido, PDI implementa un enfoque basado en metadatos que significa que el desarrollo se basa en especificar qué hacer y no cómo hacerlo.

Los administradores y desarrolladores de ETL crean sus propios trabajos de manipulación de datos a través de una interfaz gráfica amigable. Usuarios avanzados pueden usar secuencias de comandos avanzadas y crear componentes personalizados.

PDI utiliza un repositorio compartido común que permite la ejecución remota de ETL, facilita el trabajo en equipo y simplifica el proceso de desarrollo.

Kettle permite importar datos desde varios orígenes (Excel, CSV, XML, BBDD, etc.), transformar los datos a través de los pasos y generar una salida a la fuente de datos destino que se necesite. Estas tareas son típicas en procesos de migración, integración de terceros, explotación de Big Data, limpieza de datos, análisis y perfilado de datos (data profiling), etc.

Aunque las herramientas ETL se utilizan con mayor frecuencia en entornos de almacenes de datos, PDI también se puede utilizar para otros fines: migración de datos entre aplicaciones o bases de datos, exportar datos de bases de datos a archivos planos, carga masiva de datos en bases de datos, limpieza de datos, aplicaciones integradas.

PDI se compone de las siguientes herramientas:

- **SPOON**. Es una herramienta de desarrollo y modelado para desarrolladores de ETL. Permite la creación de transformaciones y trabajos.
- **PAN**, ejecuta las transformaciones diseñadas con Spoon.
- **KITCHEN**, permite ejecutar los trabajos diseñados con Spoon.
- **CARTE** es un simple servidor web que permite ejecutar transformaciones y trabajos en forma remota.

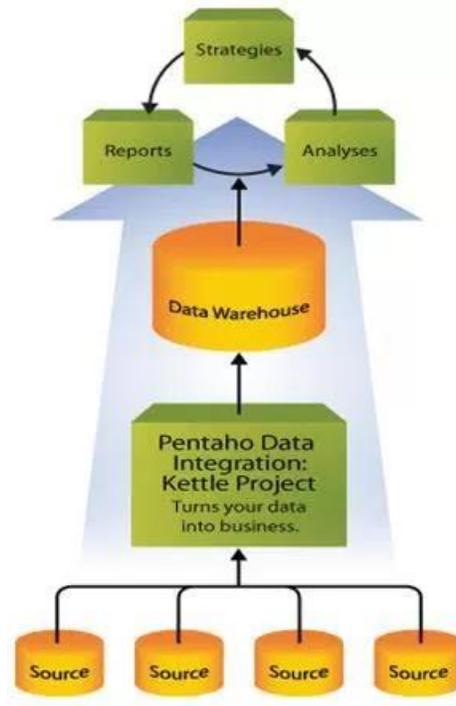


Figura 8. Proceso de ETLT en Kettle, (Gravitar, 2014).

A modo de resumen puede decirse que PDI facilita la construcción, actualización y mantenimiento de Data Warehouses.

SPOON

Diseñado para ayudar en los procesos de ETLT, que incluyen la Extracción, Transformación, Transporte y Carga de datos, Spoon es una Interfaz Gráfica de Usuario (GUI) que permite diseñar transformaciones (transformations) y trabajos (jobs) ETLT para posteriormente ser ejecutados con PAN y KITCHEN respectivamente. Ver Figuras 9 y 10.

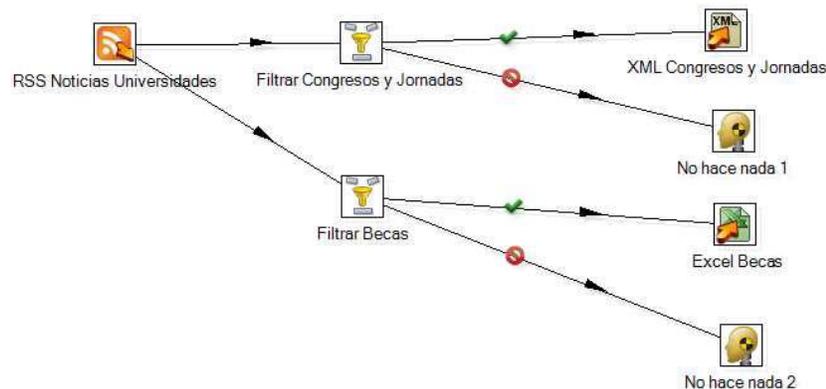


Figura 9. Ejemplo de Transformación en Spoon, (Smail, 2012).

Cuando se trabaja con Spoon, existen dos formas de guardar los elementos que se van diseñando:

- *Repositorio*: se dispone de una base de datos, con una estructura especial, donde son almacenadas las transformaciones y trabajos construidos. Puede ser útil para el trabajo en equipo y para disponer de un lugar centralizado donde se va registrando todo lo realizado.
- *Ficheros*: las transformaciones y trabajos son guardados a nivel del sistema de ficheros, en archivos xml (con extensión .ktr para las transformaciones y .kjb para los jobs).

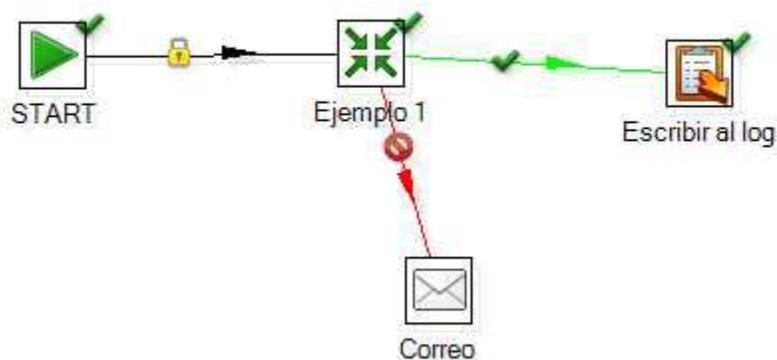


Figura 10. Ejemplo de Trabajo en Spoon, (Smail, 2012).

Spoon soporta distintos tipos de bases de datos (Oracle, MySQL, PostgreSQL, IBM DB2, etc.) y entre sus características más sobresalientes se encuentran:

- Soporta modo de ejecución, previsualización con resultados parciales y repetición.
- Exploración del repositorio de datos (tablas, vistas, sinónimos, etc.) y metadatos.
- Asistente para la creación de conexiones a bases de datos.
- Posibilidad de verificación y análisis del impacto de las transformaciones y trabajos.
- La salida de una transformación o trabajo puede ser una tabla de la base de datos la cual ha sido definida en la conexión.
- Soporta varios tipos de transformaciones básicas las cuales pueden ser llamadas por otras transformaciones y/o trabajos. Los trabajos pueden ser llamados por otros trabajos. Existen mecanismos para pasar información entre transformaciones/trabajos y definir variables.
- Importación y exportación de los metadatos de la transformación o del trabajo a fichero XML.
- Soporta Computer Clustering (Procesamiento distribuido de datos).

PAN

Se trata de un programa que puede ejecutar transformaciones diseñadas con Spoon y almacenadas como XML o en el repositorio de bases de datos. Generalmente las transformaciones se programan en modo por lotes (batch) para ser ejecutadas en intervalos de tiempo regulares. (Pentaho, 2009)

PAN es capaz de realizar multitud de funciones tales como lectura, manipulación y escritura de datos desde y hacia varios tipos de fuentes de datos.

KITCHEN

Kitchen es un programa que puede ejecutar trabajos diseñados con Spoon los cuales, pueden estar almacenados como XML ó en el repositorio de base de datos. Generalmente los trabajos se programan en modo por lotes para ser ejecutadas en intervalos de tiempo regulares. (Pentaho, 2009)

CARTE

Se trata de un simple servidor web que permite la ejecución de transformaciones y trabajos de manera remota. Esto es posible aceptando un fichero XML que contiene la transformación a ejecutar y la configuración para su ejecución. Permite también monitorear, iniciar y parar la transformación remotamente. (Pentaho, 2016)

Pentaho Analysis

Pentaho Analysis provee poder analítico en manos de los usuarios ayudándoles a obtener los conocimientos y la comprensión necesaria para tomar decisiones óptimas de negocio. Esta área de la Suite, ayuda a operar con máxima efectividad para ganar perspicacia y entender todo lo necesario. Basado en el popular motor OLAP Mondrian, proporciona las funcionalidades OLAP para el análisis de la información permitiendo, de una forma interactiva, analizar los datos del Data Warehouse a través de una interfaz de tabla cruzada donde se puede navegar por las diferentes dimensiones definidas en el modelo multidimensional.

Entre las características más relevantes de Pentaho Analysis se pueden enumerar las siguientes (Gravitar, 2014):

- Vista dimensional de datos.
- Navegar y explorar.
- Análisis Ad Hoc.
- Drill-down (profundizar o ir al detalle en una estructura multidimensional).
- Seleccionar un miembro específico para el análisis.

- Interactuar con alto rendimiento.
- Tecnología optimizada para rápida respuesta interactiva.

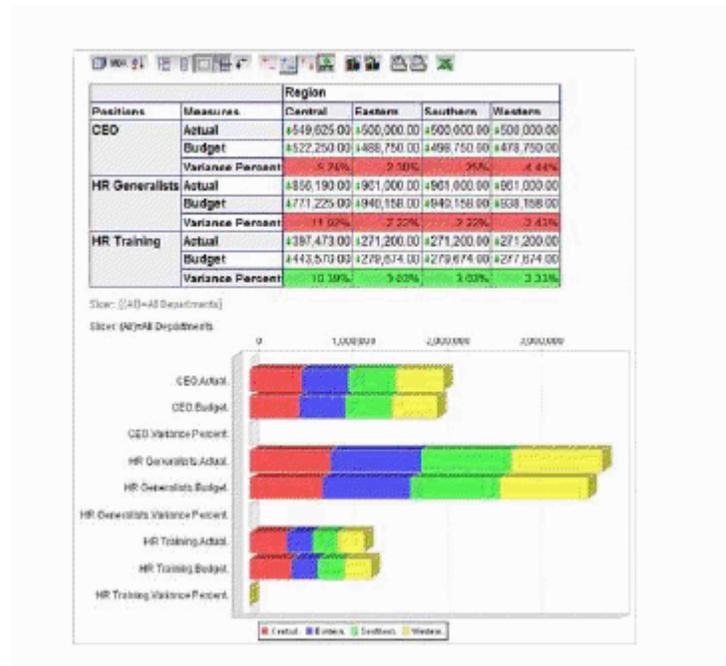


Figura 11. Visualización de los diferentes paneles de análisis con el Pentaho Analysis, (Cutro, 2008).

Cuando un usuario realiza un análisis a través de Pentaho Analysis, se lleva a cabo la siguiente secuencia de acciones entre los diferentes componentes que forman la plataforma:

- Al momento de lanzar la ejecución de un nuevo análisis, o interactuando con un análisis ya visible (haciendo un drill-down por los datos), se produce una request HTTP, que provoca una secuencia de acciones sobre Jpivot.
- El servlet Jpivot recibe la solicitud y la traduce/transforma a una query en lenguaje multidimensional MDX, que es enviada a Mondrian.
- Mondrian interpreta la query MDX y la traduce en una o más sentencias SQL, que son enviadas a la base de datos relacional.
- La base de datos relacional ejecuta las sentencias SQL enviadas por Mondrian y devuelve los resultados en forma de tabla.
- Mondrian procesa los resultados y los traduce a un set de resultados multidimensional. Este set de datos sería la respuesta a la query MDX del punto 2.
- Jpivot utiliza los resultados multidimensionales para construir una página HTML que

permita visualizar la información.

Servidor OLAP

Para obtener la funcionalidad de procesamiento analítico en línea (OLAP) se utilizan dos aplicaciones: el servidor OLAP Mondrian y la librería JPivot. La comunicación entre ambas permite realizar queries (consultas) a Datamarts, que los resultados sean presentados mediante un navegador y que el usuario pueda realizar drill-down y el resto de las navegaciones típicas (Pentaho, 2017).

Mondrian es un motor OLAP escrito en Java que ejecuta consultas escritas en lenguaje MDX, leyendo datos de una base de datos relacional (RDBMS), y presenta los resultados en un formato multidimensional a través de una API de Java (Hyde, 2006).

Las funciones más destacadas de Mondrian son:

- Análisis interactivo de altas prestaciones sobre enormes o pequeños volúmenes de información.
- Exploración dimensional de los datos, por ejemplo para analizar las ventas por línea de producto, región y periodo de tiempo.
- Análisis gramatical para MDX (Multi-Dimensional eXpression) en SQL (Structured Query Language) para recuperar la información en consultas dimensionales.
- Consultas de alta velocidad a través del uso de tablas de agregados en RDBMS.
- Cálculos avanzados usando expresiones de cálculo en lenguaje MDX.

JPivot es un proyecto considerado hermano de Mondrian por lo que al incluirse este último en Pentaho se lo incorporó también a él. Se trata de una biblioteca de etiquetas personalizadas de JSP que representa una tabla OLAP y permite a los usuarios realizar navegaciones OLAP típicas, como cortar y cortar, detallar y enrollar. Utiliza Mondrian como su servidor OLAP (JPivot, 2008).

Pentaho Dashboards

Pentaho Dashboards o CDE, por sus siglas en inglés Community Dashboard Editor, es la herramienta desarrollada por ésta Suite, más concretamente por la empresa portuguesa Webdetails (2016) liderada por Pedro Alves. Permite la creación de dashboards en la

interfaz final de la herramienta web, los cuales permitirán realizar funciones de consulta y análisis de los datos.

Un dashboard se utiliza para presentar el contenido de BI de alto nivel a los usuarios finales. Los indicadores gráficos de alto nivel que aparecen en el dashboard proporcionan un cierto grado de interactividad que permite al usuario navegar por contenidos más detallados (Smail, 2012). Provee inmediata perspicacia en un rendimiento individual, departamental o empresarial. Pentaho Dashboards facilita a los distintos actores de la organización información crítica que necesitan para entender y mejorar el rendimiento organizacional y así facilitar la toma de decisiones.

CDE ofrece un conjunto de herramientas, CTools, para diseñar diversos dashboards interactivos, las cuales están integradas en el business intelligence server. Este conjunto de herramientas incluye entre otras Community Dashboard Framework (CDF), Community Chart Components (CCC) y Community Data Access (CDA).

Entre las características más importantes de Pentaho Dashboards se destacan:

- Amplias mediciones que permiten la definición y seguimiento de medidas críticas (KPIs, Key Performance Indicators) nivel individual, departamental o empresarial.
- Pantallas ricas e interactivas donde observar inmediatamente cuáles mediciones requeridas van bien y cuáles requieren especial atención.
- Integración con Pentaho Reporting y Pentaho Analysis lo que permite profundizar en los reportes y análisis para entender cuáles factores están contribuyendo a un buen o mal desempeño.
- Integración a un portal para facilitar la entrega de mediciones relevantes a un número grande de usuarios.
- Alertas integradas para casos excepcionales y notificación al usuario para que tome una acción.

Para el diseño de un dashboard CDE ofrece 3 perspectivas: Perspectiva de diseño, Perspectiva de componentes y Perspectiva de la fuente de datos (Pentaho, 2017).

La *perspectiva de diseño* permite diseñar el tablero desde cero o utilizando una plantilla CDE. Varias de las propiedades simples proporcionadas por CDE se pueden usar para personalizar cada componente específico. Algunas propiedades adicionales se pueden configurar, como el color de fondo o el estilo para las esquinas de las filas. Esas

propiedades tendrán prioridad sobre cualquier regla de CSS existente que haya incluido en el tablero. En resumen, se puede ajustar la apariencia de del tablero de las siguientes maneras: HTML para crear elementos de tablero, CSS para controlar el estilo y el diseño, JavaScript para agregar interactividad, jQuery para simplificar todas esas tareas.

La *perspectiva de componentes* permite agregar y configurar los diferentes componentes que componen el tablero. Estos componentes son los elementos centrales de un tablero de instrumentos, dado que estos vinculan los elementos de diseño con las fuentes de datos. Existen tres tipos de componentes: componentes visuales y de datos (cuadros de texto, tablas, gráficos, selectores, vistas OLAP e informes), parámetros que representan los valores que comparten los componentes, guiones que son pedazos de código JavaScript que permiten personalizar la apariencia y el comportamiento de otros componentes.

La *perspectiva de fuente* de datos permite encontrar los distintos tipos de orígenes de datos que se pueden emplear en un tablero.

Community Dashboard Framework

Community Dashboard Framework, CDF por sus siglas en inglés, es un framework open source que permite la creación de tableros altamente personalizables sobre el servidor Pentaho Business Intelligence. Basado en estándares de desarrollo web como CSS, HTML5 y JavaScript, ofrece a los desarrolladores de BI una forma fácil y rápida de crear cuadros de mandos dinámicos, a través de los cuales los usuarios podrán explorar y comprender grandes cantidades de datos, haciendo uso de una amplia variedad de gráficos y tablas. Estas características convierten a CDF en una solución efectiva que permite combinar datos con una atractiva capa de visualización.

Las características principales de este framework se presentan a continuación (Goncalves, 2017):

- Basado en tecnologías open source;
- Utiliza AJAX;
- Presenta un ciclo de vida con componentes que interactúan entre sí;
- Separa la lógica (JavaScript) de la presentación (HTML, CSS);
- Brinda a los usuarios un alto nivel de personalización;

- Biblioteca de componentes;
- Posibilidad para los usuarios avanzados de insertar sus propios fragmentos de código JavaScript y jQuery.

Algunas de las ventajas técnicas que presenta CDF son la posibilidad de elegir entre múltiples componentes, tableros de fácil diseño, la seguridad es aplicada a través del modelo de seguridad de Pentaho, etc.

Community Chart Components

CCC es la sigla correspondiente a Community Chart Components, la biblioteca de gráficos de CTools basada en Protovis. Se trata de un kit de herramientas de visualización open source muy potente cuyo objetivo principal es proporcionar a los desarrolladores BI la mecánica para incluir en los cuadros de mando todos tipo tipo de gráficos interactivos y personalizables, sin perder de vista el principio fundamental: la extensibilidad.

Protovis es una librería de visualización gráfica desarrollada por Michael Bostock para visualizaciones complejas y gráficos interactivos. Básicamente, permite manipular documentos basados en datos usando estándares abiertos de la web. Así, los navegadores pueden crear visualizaciones complejas sin depender de un software propietario. Protovis es gratuito y de código abierto, provisto bajo la Licencia BSD. Utiliza JavaScript y SVG para visualizaciones nativas de la web. Cabe aclarar que actualmente no hay desarrollo activo para Protovis ya que su equipo se encuentra trabajando en D3.js. (Protovis, 2010)

Community Data Access

Community Data Access o CDA es un plugin Pentaho diseñado para acceder a datos con gran flexibilidad. Nacido para superar algunos inconvenientes de la implementación más antigua, CDA le permite acceder y entregar datos en diferentes formatos de las muchas fuentes de datos de Pentaho. Desarrollada como una herramienta de abstracción entre las conexiones de bases de datos y CDF, permite recuperar datos de múltiples fuentes de datos y combinarlos en una única salida que se puede pasar fácilmente a los componentes del tablero. Sirve tres propósitos principales (Goncalves, 2017):

- Permite unir datos de diferentes fuentes simplemente editando un archivo XML;
- Las consultas de caché proporcionan un gran impulso en el rendimiento;
- Evita los problemas de inyección de SQL dentro de CDF;
- Permite una forma más fácil de exportar datos de consultas en diferentes formatos (csv, xls, etc.) a través de la Consola de usuario de Pentaho;
- Ordenar y paginar datos del lado del servidor.

Pentaho Data Mining

Pentaho Data Mining o WEKA es una herramienta de aprendizaje automático y data mining, que fue desarrollada en la Universidad de Waikato (WEKA = Waikato Environment for Knowledge Analysis).

Esta aplicación Java contiene las herramientas necesarias para realizar transformaciones sobre los datos, tareas de clasificación, regresión, clustering, asociación y visualización. Los resultados pueden ser visualizados en modo gráfico ya sea agrupado, segmentado, de árbol de decisión, bosque aleatorio, redes neuronales y componentes de análisis (Salinas, 2008). Diseñado como una herramienta orientada a la extensibilidad, añadir nuevas funcionalidades a la misma es una tarea sencilla. Sin embargo, y pese a todas las cualidades que posee Weka, tiene un gran defecto y éste es la escasa documentación orientada al usuario que tiene junto a una usabilidad bastante pobre, lo que la hace una herramienta difícil de comprender y manejar sin información adicional. (Goncalves, 2017)

Weka posee licencia GPL (GNU Public License), lo que quiere significar que es un programa de libre distribución y difusión. Como Weka está programada en Java, es independiente de la arquitectura, es decir, funciona en cualquier plataforma sobre la que esté instalada una máquina virtual de Java.



Figura 12. Visualización distintas tareas realizadas con Pentaho Data Mining, (Gravitar, 2014).

Pentaho Reporting

La Suite Pentaho ofrece la herramienta gráfica Pentaho Reporting que permite la generación de informes. Todas las organizaciones utilizan la presentación de informes de una u otra forma. Como resultado, la presentación de informes es considerada un núcleo necesario y es con frecuencia la primera aplicación de BI en implementarse. Pentaho Reporting permite a las organizaciones acceder fácilmente, dar formato, y distribuir información entre los distintos actores de la organización, ver Figura 13.

Dentro de esta área de la Suite existen tres productos con diferentes enfoques y dirigidos a diferentes tipos de usuarios (Cutro, 2008): Pentaho Report Designer, Pentaho Report Design Wizard, Web ad-hoc reporting.

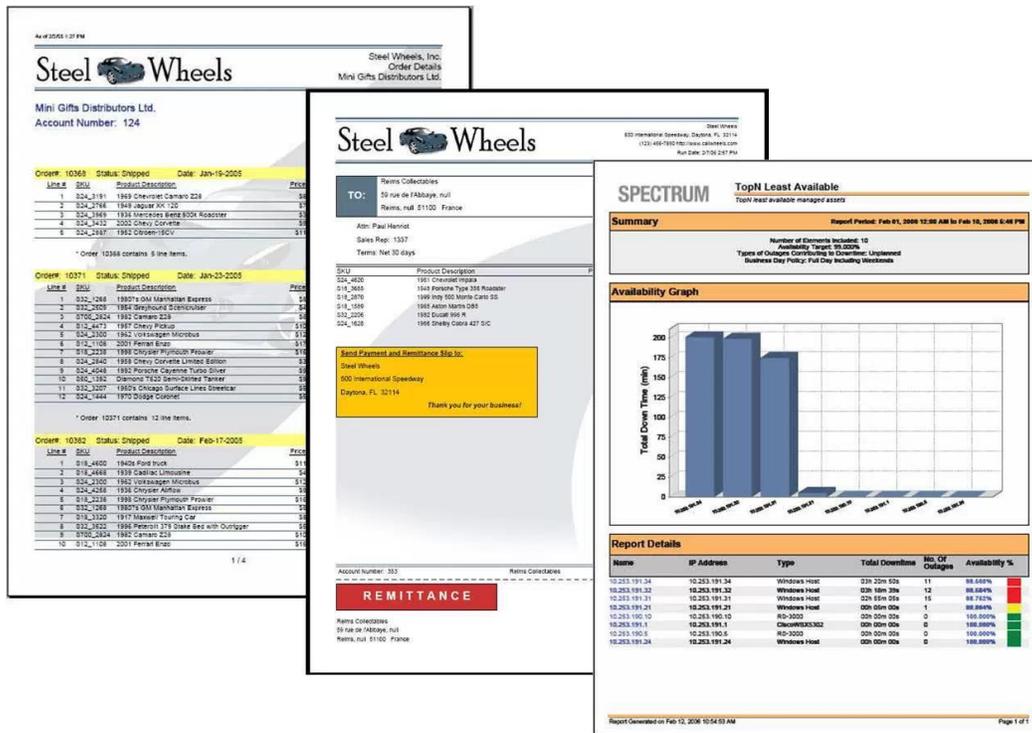


Figura 13. Visualización de los distintos reportes generados por Pentaho Reporting, (Gravitar, 2014).

Pentaho Report Designer.

Destinado a desarrolladores, se trata de un editor basado en Eclipse con prestaciones profesionales y capacidad de personalización de informes de acuerdo a las necesidades de la organización. Estructurada de forma que los desarrolladores puedan acceder a sus prestaciones de forma rápida, incluye un editor de consultas para facilitar la confección de los datos que serán utilizados en un informe.

Pentaho Report Designer permite trabajar con múltiples orígenes de datos como JDBC, Olap4J, Pentaho Analysis, Pentaho Data Integration, XML en un entorno gráfico. Los resultados puede ser presentada en distintos formatos como PDF, HTML, XLS, RTF y CSV.

Pentaho Report Design Wizard.

Pentaho Report Design Wizard es una herramienta de diseño de informes creada para facilitar la creación de informes con JFreeReport. Esta aplicación aunque resulta flexible e intuitiva en la tarea de generación de informes no es capaz de cumplir algunos requisitos específicos, es decir, es una aplicación puente entre la definición pura del informe en un fichero XML y una herramienta de diseño completa.

En el caso de necesitar alguna característica específica no contemplada en este asistente de diseño de informes, el diseñador deberá modificar manualmente el código XML.

Pentaho Reporting Design Wizard es distribuido bajo la licencia MPL (Mozilla Public License) (Pentaho, 2008).

Web ad-hoc reporting.

Web Adhoc Query and Reporting (WAQR), similar a la herramienta Pentaho Report Design Wizard, es un módulo web para la generación de informes personalizados. Diseñada principalmente para facilitar la generación de informes a través de la plataforma web, extiende la capacidad de los usuarios finales permitiendo la creación de informes a partir de plantillas preconfiguradas y siguiendo un asistente de creación.

WAQR divide la creación de informes en cuatro etapas (Stratebi, 2016):

1. Definición de datos contextuales del informe: Descripción, cabecera, pie de página, tamaño y orientación del papel.
2. Selección de la fuente de datos y formato de presentación.
3. Selección de campos y elección de la sección del informe donde ubicarlos.
4. Personalización del informe: Condiciones de filtro, ordenamiento de los datos y formatos de campos.

2.2 DESCUBRIMIENTO DEL CONOCIMIENTO (KDD), MINERÍA DE DATOS (DM) Y LEARNING ANALYTICS (LA). CONCEPTUALIZACIÓN Y APLICACIÓN.

El aprendizaje es producto de la interacción. Estas interacciones abarcan todos los niveles; estudiante-profesor, profesor-estudiante, estudiante-contenido y también entre estudiantes. Independientemente de la metodología utilizada en el diseño instruccional o de aprendizaje surgen de manera constante una serie de preguntas: ¿Qué tan efectivo es el curso? ¿Está satisfaciendo las necesidades de los estudiantes? ¿Cómo pueden apoyar mejor las necesidades de los estudiantes? ¿Qué interacciones son eficaces? ¿Cómo pueden mejorarse aún más? (Elias, 2011) Hoy día es posible responder a estas preguntas gracias a la incorporación de la tecnología a la educación.

En el ámbito educacional existe una amplia variedad de términos para el análisis de datos educativos. Estos términos a su vez han sido definidos por distintos autores desde distintas perspectivas: algunas definiciones, de índole conceptual, indicaban “lo que es” mientras que otras más funcionales indicaban “lo que hace”. Hawkins (2008) y Watson (2011) advierten que la analítica es una práctica dirigida a un objetivo. Esos objetivos, esas necesidades de información dependerán de las instituciones de educación superior. Tal como sostiene Watson (2011) “el análisis significa cosas diferentes para diferentes personas. Hay diferentes tipos de análisis, y las diferencias tienen implicaciones importantes en cuanto a dónde se usan, quiénes las realizan, las habilidades que se requieren y las tecnologías que están involucradas [...] tenga claro qué tipo de análisis está discutiendo”.

Con el objeto de clarificar los términos analíticos usados en la actualidad, se ha elaborado una tabla en la que se plasman las distintas definiciones conocidas para cada uno de estos términos y el nivel en el que se enfocan esos análisis (institución, departamento, alumno).

Analytics	“[The] processes of data assessment and analysis that enable us to measure, improve, and compare the performance of individuals, programs, departments, institutions or enterprises, groups of organizations, and/or entire industries.”	Institución Departamento Profesor Alumno
-----------	--	---

	Data-driven decision making, used to inform decisions at all levels of the enterprise	Institución Departamento Profesor Alumno
Business Analytics	“The whole category is just using data and analysis to understand and manage your business more effectively, as opposed to simply capturing your customer’s address or keeping track of your employees’ vacation balances, that transactional kind of stuff.”	Institución
	“Business Analytics (BA) is the practice of iterative, methodical exploration of an organization’s data with emphasis on statistical analysis. BA is used by companies committed to data-driven decision making. BA is used to gain insights that inform business decisions and can be used to automate and optimize business processes. Data-driven companies treat their data as a corporate asset and leverage it for competitive advantage.”	Institución
Academic Analytics	“Early academic analytics initiatives are seeking to predict which students are in academic difficulty, allowing faculty and advisors to customize learning paths or provide instruction tailored to specific learning needs.”	Alumno
	“[Focused] on academic issues, primarily student access, affordability, and success.”	Alumno
	“[The] imperfect equivalent term for Business Intelligence, which [essentially describes] the use of information technology to support operational and financial decision making.”	Institución
	“[Marrying] data with statistical techniques and predictive modeling to help faculty and advisors determine which students may face academic difficulty, allowing interventions to help them succeed.”	Profesor Alumno
	“Mining data from systems that support teaching and learning to provide customization, tutoring, or intervention within the learning environment.”	Profesor Alumno

	“[A] process for providing [higher education institutions] with the data necessary to respond to the reportage and decision making challenges facing contemporary universities.”	Institución Departamento
	“[It] can refer broadly to data-driven decision making practices for operational purposes at the university or college level, but it can also be applied to student teaching and learning issues.”	Institución Profesor Alumno
	“[Can] identify and even predict students who may be at risk [in a particular course].”	Profesor Alumno
Learning Analytics	“[The] interpretation of a wide range of data produced by and gathered on behalf of students in order to assess academic progress, predict future performance, and spot potential issues.”	Institución Alumno
	“[The] use of predictive modeling and other advanced analytic techniques to help target instructional, curricular, and support resources to support the achievement of specific learning goals” (n.b. Bach presents learning analytics and academic analytics as synonymous concepts).	Departamento Alumno
	“[To] enable teachers and schools to tailor educational opportunities to each student’s level of need and ability.”	Alumno
	“It might be used as well to assess curricula, programs, and institutions.”	Institución Departamento
	“[The] use of data and models to predict student progress and performance, and the ability to act on that information.”	Alumno
	“[The] collection and analysis of usage data associated with student learning; [to] observe and understand learning behaviors in order to enable appropriate intervention.”	Alumno
Predictive Analytics	“[A] set of [business intelligence] technologies that uncovers relationships and patterns within large volumes of data that can be used to predict behavior and events [...] predictive analytics is forward-looking, using past events to anticipate the future.”	Institución Departamento Profesor Alumno

	<p>“Predictive analytics connects data to effective action by drawing reliable conclusions about current conditions and future events.”</p>	<p>Institución Departamento Profesor Alumno</p>
	<p>“Predictive analytics [...] is both a business process and a set of related technologies. Predictive analytics leverages an organization’s business knowledge by applying sophisticated analysis techniques to enterprise data. The resulting insights can lead to actions that demonstrably change how people behave as customers, employees, patients, students, and citizens.”</p>	<p>Institución Departamento Profesor Alumno</p>

El uso de Entornos Virtuales de Aprendizaje (EVA) genera una huella digital que aporta una gran cantidad de información acerca de las actividades de los usuarios (Romero, Ventura y García, 2008). Las plataformas de aprendizaje virtual almacenan toda la información relativa a la actividad que los estudiantes desarrollan en cursos virtuales en bases de datos que, procesados correctamente, pueden ofrecer información relevante para el docente. El docente puede conocer el comportamiento que tienen los alumnos en la plataforma y descubrir el proceso de aprendizaje que llevan a cabo, pudiendo adaptar el curso al modo en que trabajan sus alumnos y tomar medidas ante los problemas que se puedan detectar. Es decir, esta información recopilada por los sistemas de información educativos puede utilizarse para tomar decisiones y responder a preguntas, buscando la mejora de la calidad del sistema educativo.

El gran auge de los Entornos Virtuales de Aprendizaje y el Big Data han permitido la aparición de una serie de paradigmas analíticos cuya aplicación se realiza con vistas a predecir, conocer patrones de comportamiento u optimizar los aprendizajes (individuales o grupales). Esos paradigmas analíticos son: Educational data Mining (EDM), Learning Analytics (LA), Academic Analytics (AA).

- Educational Data Mining. Consiste en aplicar técnicas de minería de datos (clustering, clasificación, reglas de asociación...) en grandes cantidades de datos educativos. Su objetivo es predecir mediante la creación de modelos con el fin de

mejorar los procesos relacionados con los distintos roles educativos. Por ejemplo, puede recomendar a los alumnos qué recurso o tarea consultar a continuación, ayudar a los administradores a mejorar los cursos o detectar y avisar a los profesores de posibles alumnos en riesgo de suspenso. Lectura recomendada: Handbook of Educational Data Mining (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series)

- Learning Analytics. Consiste en un proceso cíclico en cuyo primer paso se definen una serie de métricas. A continuación se coleccionan datos de los alumnos y el contexto a partir de las mismas. Finalmente se analizan y se generan informes decisionales. Su propósito es entender y mejorar los procesos y entornos donde se produce enseñanza-aprendizaje para ayudar a tomar decisiones accionables.
- Academic Analytics. Consiste en aplicar técnicas de Business Intelligence para ayudar al centro a tomar decisiones estratégicas cuyo principal propósito es la mejora del contexto y procesos educativos. El proceso consiste en encontrar insights accionables mediante la definición de métricas, recolección, análisis, informe de datos y/o predicción para actuar en relación a la estrategia de negocio.

Al igual que con la mayoría de las áreas de investigación, la terminología que identifica y diferencia EDM, LA y AA y no es tan homogénea como sería deseable (VanBarneveld; Arnold; Campbell, 2012), sin embargo, las tres áreas tienen en común, además del objetivo de mejorar y comprender el proceso de aprendizaje, la necesidad de una gran cantidad de observaciones de dicho proceso.

Tanto las empresas como las instituciones de educación superior están adoptando prácticas para garantizar el éxito de la organización en todos los niveles al abordar preguntas sobre retención, admisiones, recaudación de fondos y eficiencia operativa. Natsu (2010) informó que los análisis podrían ayudar a los líderes de educación a reducir costos y mejorar la enseñanza y el aprendizaje. Afirmó que el uso de análisis analíticos predictivos podría abarcar desde "mejorar las eficiencias hasta ahorrar dinero para mejorar el rendimiento estudiantil".

El éxito de una organización, así como la ejecución de sus actividades, están centralizados en adoptar decisiones, siendo considerada una de las principales etapas en

el proceso de negocios. De esta forma, la toma de decisiones estratégica en el concepto de Business Intelligence (BI) se convierte en una actividad de extrema importancia, una vez que decisiones equivocadas pueden comprometer el futuro de una organización.

Las necesidades de información de valor, concisa y resumida para la toma de decisiones son día a día más necesarias en todos los contextos de las organizaciones, por ello es que surge la teoría de analítica de la información o de datos. Esta teoría persigue facilitar información objetiva y precisa para soportar la toma de decisiones estratégicas y que permita mantener la competitividad de las organizaciones, mediante la aplicación de modelajes y métodos predictivos que van más allá de simples estadísticas descriptivas.

En el contexto de las universidades no es ajena la urgencia de contar con información precisa y efectiva para dar soporte a la toma de decisiones estratégicas, con miras a mejorar los procesos. Los procesos de gestión académica generalmente toman más del 80% del tiempo productivo de los académicos responsables de unidades o programas de formación universitarios.

La explotación de la información académica es cada vez más necesaria. La analítica académica (J. Campbell y D. Oblinger, 2007) combina los datos institucionales, el análisis estadístico y los modelos predictivos permitiendo la exploración de datos para identificar informaciones nuevas y útiles para atender las expectativas y necesidades estratégicas de las organizaciones de educación superior (A. Rodriguez Almeida y S. da Silva Camargo, 2015).

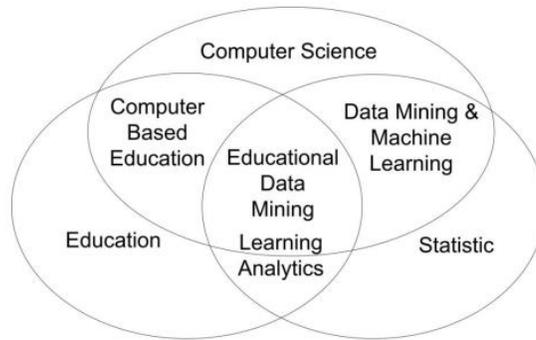
El término Academic Analytics (en castellano, analítica académica, AA) lo acuñaron Goldstein y Katz (2005) definiéndolo como la aplicación de herramientas de inteligencia de negocios para el área de aprendizaje. La principal aplicación de la AA es la de ir más allá de un reporte simple de información y sugerir decisiones.



© Analítica Académica S.A.S

La analítica académica refleja el rol del análisis de datos a nivel institucional, mientras que la analítica del aprendizaje se centra en el proceso de aprendizaje (que incluye el análisis de la relación entre el alumno, el contenido, la institución y el educador).

Por otro lado, si bien la Minería de Datos Educativa (MDE) y Learning Analytics (LA) comparten un mismo objetivo, se solapan y tienen métodos comunes en lo que respecta a la hora de extraer información relevante a través de los datos, existen entre ellos algunas diferencias (Ver Figura 1). Mientras la MDE desarrolla y adapta métodos estadísticos de aprendizaje automático y estudia esos datos educativos generados tanto por estudiantes como por profesores, el Learning Analytics se relaciona con aspectos pedagógicos y de mejora del aprendizaje. Aplica técnicas de las ciencias de la información, la sociología, la psicología, las estadísticas, el aprendizaje automático y la minería de datos (Bienkowski, Feng y Means, 2012). Siemens y Baker (2012) expresan que la *“EDM tiene un considerable mayor enfoque en el descubrimiento automatizado mientras que LAK (Learning Analytics and Knowledge) pone un mayor énfasis en el aprovechamiento del juicio humano”*.



Descubrimiento del Conocimiento y Minería de Datos

Durante los últimos años, la capacidad para generar y recolectar datos ha crecido de manera significativa debido esencialmente al gran poder de procesamiento de las computadoras y al bajo costo de almacenamiento.

Sin embargo, dentro de estas enormes masas de datos existe una gran cantidad de información oculta, a la que no se puede acceder a través de las clásicas técnicas de recuperación de la información. El descubrimiento de esta información es posible gracias a la Minería de Datos o *Data Mining*, campo multidisciplinar que se ha desarrollado en paralelo o como prolongación de otras tecnologías (ver Figura 1). La investigación y los avances en la minería de datos se nutren de los que se producen en estas áreas relacionadas. La minería de datos aplica, entre otras sofisticadas técnicas, la inteligencia artificial para encontrar patrones y relaciones dentro de los datos permitiendo la creación de modelos.

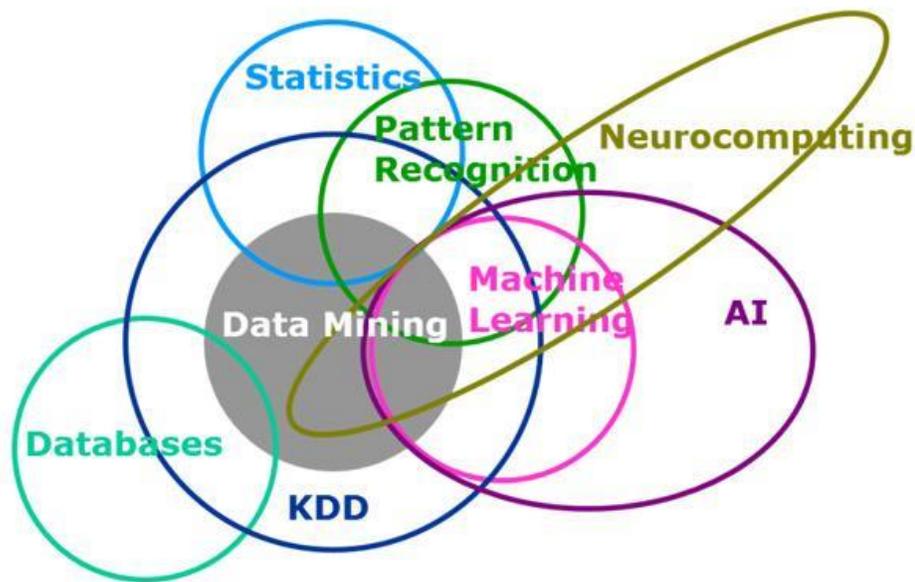


Figura 1. Disciplinas que contribuyen a la minería de datos (SAS, 2014).

La exploración o búsqueda en datos se realizó tradicionalmente a través del análisis estadístico utilizando técnicas estadísticas tales como correlación, regresión, etc. A fines de la década de 1980, el análisis estadístico incorpora la lógica difusa, el razonamiento heurístico y las redes neuronales, utilizadas para lograr mayor eficiencia en la exploración e identificación de datos útiles.

Algunas definiciones que se han dado sobre minería de datos permiten situar a la misma:

“the process of secondary analysis of large databases aimed at finding unsuspected relationships which are of interest or value to the database owners” (Hand, 1998: 112).

“Iterative process of extracting hidden predictive patterns from large databases, using AI technologies as well as statistics techniques” (Mena, 199: 357).

La última definición sin ser la más popular, resalta las raíces de la minería de datos: la Inteligencia Artificial (en particular Machine learning) y la Estadística.

La Minería de Datos es sólo un paso dentro de un proceso más amplio denominado *descubrimiento del conocimiento*, más conocido como KDD por sus siglas en inglés *Knowledge Discovery in Databases* (Fayyad, Piatetsky-Shapiro y Smyth, 1996).

KDD se refiere al proceso general de descubrimiento de información útil en los datos. En

este sentido, se define como:

El descubrimiento de conocimiento en bases de datos es un campo de la inteligencia artificial de rápido crecimiento, que combina técnicas del aprendizaje de máquina, reconocimiento de patrones, estadística, bases de datos, y visualización para automáticamente extraer conocimiento (o información), de un nivel bajo de datos (bases de datos) (Fayyad, 1997: 11).

KDD es un proceso complejo, iterativo e interactivo que incluye no sólo la obtención de los modelos o patrones, sino también la evaluación y posible interpretación de los mismos. Así, los sistemas de descubrimiento de conocimiento permiten la selección, limpieza, transformación y proyección de los datos; posibilita también analizar los datos para extraer patrones y modelos adecuados, evaluar e interpretar los patrones para convertirlos en conocimiento, consolidar el conocimiento resolviendo posibles conflictos con conocimiento previamente extraído, y hacer el conocimiento disponible para su uso (Ver Figura 2). De esta manera se clarifica la relación existente entre el KDD y la Minería de Matos: el KDD es el proceso global de descubrir conocimiento útil desde las bases de datos mientras que la Minería de Datos se refiere a la aplicación de los métodos de aprendizaje y estadísticos para la obtención de patrones y modelos.

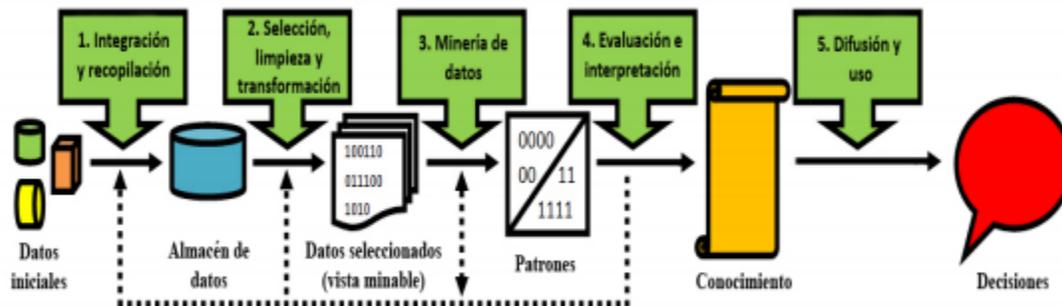


Figura 2. Fases del proceso de descubrimiento de conocimiento en bases de datos, KDD (Cortina, 2015).

Una vez comprendido el dominio de aplicación y el objetivo para el cual se llevará adelante este proceso se comienza con la **fase de integración y recopilación de datos**, fase en la que se hace necesario determinar las fuentes de información que pueden ser

útiles y donde conseguirlas. Para las funciones de análisis, planificación y predicción no siempre son suficientes las bases de datos y las aplicaciones basadas en el procesamiento tradicional de datos en línea (OLTP, del inglés On Line Transaction Processing) dado que estas solo cubren las necesidades diarias de información de una organización. Por este motivo, en algunos casos se requiere obtener datos provenientes de otras áreas de la organización, incluso provenientes de bases de datos públicas (datos censales, datos demográficos, etc.) o privadas (de bancos, compañías de servicios, etc.). Una vez reunida toda la información, los datos deben ser transformados a un formato común, frecuentemente a través de un almacén de datos (Data Warehouse) que consiga unificar de manera operativa toda la información obtenida, detectando y resolviendo previamente las inconsistencias (Orallo, 2010).

Como los datos provienen de distintas fuentes, estos pueden contener valores erróneos o faltantes. Estas situaciones son tratadas en la **fase de selección, limpieza y transformación**, en la que se eliminan y corrigen los datos incorrectos, se decide qué estrategia seguir con los datos incompletos y se proyectan para considerar únicamente aquellas variables que van a ser relevantes, con el objetivo de hacer más fácil la tarea propia de la minería y que los resultados de la misma sean útiles. Una vez procesados los datos, los mismos pueden ser transformados para reducir la dimensión del problema. El número de variables a considerar puede reducirse o se pueden buscar nuevas formas de representar los datos.

En la **fase de minería de datos** se debe decidir qué técnica de minería se ajusta más al objetivo del proceso de KDD bajo estudio. Las distintas técnicas pueden ser *predictivas (aprendizaje supervisado)* o *descriptivas (aprendizaje no supervisado)*, (ver Figura 3).



Figura 3. Técnicas de la minería de datos (Guevara Vega, 2015).

Entre las tareas predictivas se encuentran la clasificación y la regresión. Por su parte, las tareas descriptivas incluyen el agrupamiento (clustering), las reglas de asociación, y las correlaciones.

La *clasificación* es quizá la tarea más utilizada. En ella cada instancia o registro de la base de datos pertenece a una clase, la cual se indica mediante el valor de un atributo denominado clase de la instancia. Este atributo puede tomar diferentes valores discretos, cada uno de los cuales corresponde a una clase. El resto de los atributos de la instancia se utilizan para predecir la clase.

La *regresión* consiste en aprender una función real que asigna a cada instancia un valor real. Esta es la principal diferencia respecto de la clasificación; el valor a predecir es numérico. El agrupamiento, también conocido como *clustering*, es la tarea descriptiva por excelencia. Se construyen grupos de observaciones similares según un criterio prefijado. El proceso de clustering (agrupamiento) consiste en subdividir un conjunto de datos en grupos mutuamente excluyentes de tal manera que cada miembro de un grupo esté lo más cercano posible a otro elemento, y grupos diferentes estén lo más lejos posible entre sí, de modo que la distancia está medida respecto a todas las variables disponibles.

Las *correlaciones* se usan para examinar el grado de similitud de los valores de dos variables numéricas.

Las *reglas de asociación* son usadas para identificar asociaciones no explícitas entre variables. Puede no existir una causa para que los datos estén asociados.

En la **fase de evaluación e interpretación** se evalúan los patrones obtenidos en la etapa anterior y se analizan por expertos. En caso de ser necesario se vuelve a las fases anteriores para una nueva iteración, esto incluye resolver posibles conflictos con el conocimiento que se disponía anteriormente.

Finalmente, en la **fase de difusión** se hace uso del nuevo conocimiento y se hace partícipe de él a todos los usuarios. Este nuevo conocimiento servirá para ejecutar acciones basadas en información.

El proceso de KDD, es un buen mecanismo para hacer lo que hoy damos en llamar **“Data Science”**.

Existen diversas metodologías para concretar proyectos de Minería de Datos. Cada metodología tiene características y etapas o fases que indican cómo lograr el proceso de descubrimiento de conocimiento. Tal como muestra la Figura 4, algunas metodologías son más usadas que otras; una de ellas, CRISP-DM, se ha convertido en un estándar internacional, por sus etapas detalladas y su amplia aplicación en diferentes casos de estudio.

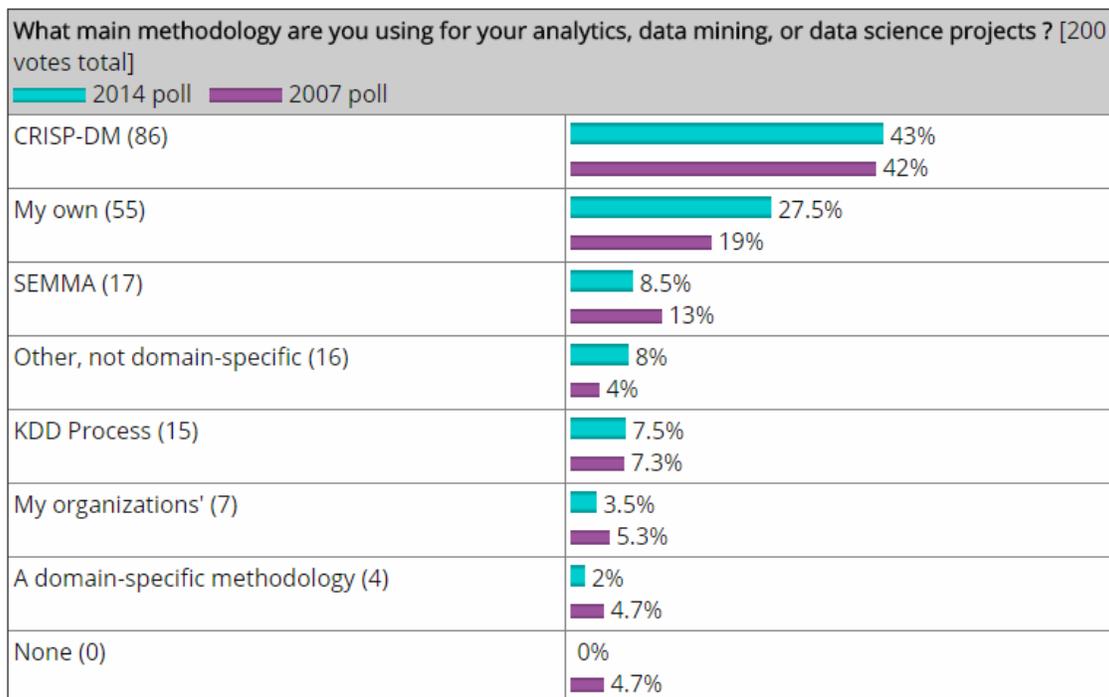


Figura 4. Metodologías más usadas en minería de datos (KDnuggets, 2014).

CRISP-DM (Cross Industry Standard Process for Data Mining) incluye un modelo y una guía, estructurados en seis fases algunas de las cuales son bidireccionales, es decir que de una fase en concreto se puede volver a una fase anterior, por lo que la sucesión de fases no tiene porqué ser ordenada desde la primera hasta la última (Cortina, 2015). (Ver Figura 5)

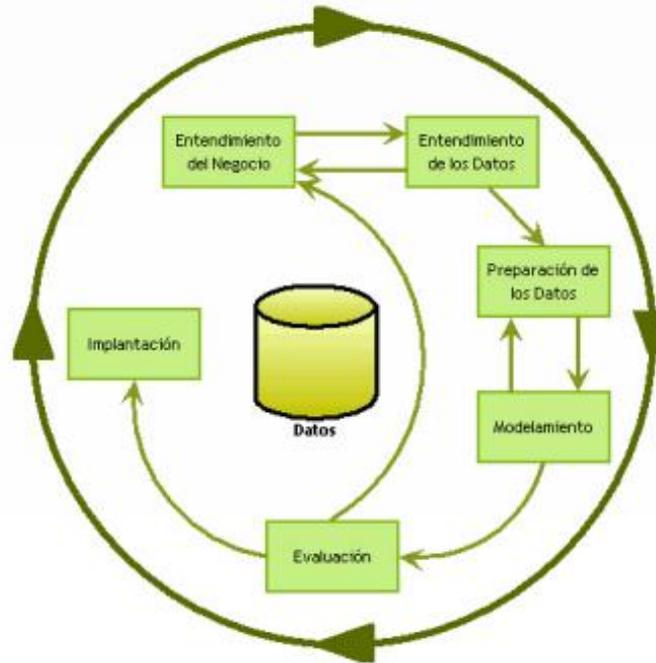


Figura 5. Fases del proceso CRISP-DM (Cortina, 2015).

Áreas de aplicación del DM

La integración de las técnicas de Minería de Datos en las actividades diarias se está convirtiendo en algo habitual.

Según Microsoft (2016), los modelos de minería de datos se pueden aplicar a diferentes escenarios donde los objetivos son variados, como:

- **Pronóstico:** predecir el resultado de un evento o proceso determinado.
- **Riesgo y Probabilidad:** determinación del punto de equilibrio probable para los escenarios de riesgo, y asignación de probabilidades a diagnósticos y otros resultados.
- **Recomendaciones:** sugerir acciones a tomar en base a los resultados del análisis.
- **Búsqueda de secuencias:** descubrir eventos, o sucesos que ocurren juntos o en un cierto orden cronológico, lo que permite extrapolar resultados a momentos no

observados o recomendar acciones a tomar.

- **Agrupación:** agrupar individuos u objetos con características similares, a fin de describir o predecir comportamiento o afinidades.

Existen numerosas y variadas áreas de aplicación con estos objetivos, como:

- *Aplicaciones financieras y banca:* Obtención de patrones de uso fraudulento de tarjetas de crédito. Determinación del gasto en tarjetas de crédito por grupos. Cálculo de correlaciones entre indicadores financieros. Análisis de riesgos en créditos.
- *Análisis de mercado, distribución y comercio:* análisis de los artículos que los clientes han introducido en el carrito de la compra y predicción de posibles eventos. Elección de los mejores clientes para la distribución de correo directo. Evaluación de campañas publicitarias. Análisis de la fidelidad de los clientes. Reducción de fuga. Segmentación de clientes. Estimación de stocks, de costes, de ventas, etc (Berry y Linoff, 2000; Koren, 2010).
- *Seguros y salud privada:* Determinación de los clientes que podrían ser potencialmente caros. Análisis de procedimientos médicos solicitados conjuntamente. Predicción de que clientes contratan nuevas pólizas. Identificación de patrones de comportamiento para clientes con riesgo.
- *Educación:* Selección o captación de estudiantes. Detección de abandonos y de fracasos. Estimación del tiempo de estancia en la institución.
- *Procesos industriales:* Extracción de modelos sobre comportamientos de compuestos. Detección de piezas con trabas, modelos de calidad. Predicción de fallos y accidentes. Estimación de composiciones óptimas en mezclas. Extracción de modelos de coste y producción (Grossman, Kamath, Kegelmeyer, Kumar y Namburu, 2013).
- *Medicina:* Identificación de patologías. Diagnóstico de enfermedades. Detección de pacientes con riesgo de sufrir una patología concreta. Gestión hospitalaria y asistencial. Predicciones temporales de los centros asistenciales para el mejor uso de recursos, consultas, salas y habitaciones. Recomendación priorizada de fármacos para una misma patología (McCormick, Rudin y Madigan, 2012).
- *Biología, bioingeniería y otras ciencias:* Análisis de secuencias de genes y de proteínas. Predecir si un compuesto químico causa cáncer. Clasificación de cuerpos celestes. Predicción de recorrido y distribución de inundaciones. Modelos de calidad

de aguas, indicadores ecológicos (Bergeron, 2002; Friedman, Linial y Nachman, 2000; Burl, Asker, Smyth, Fayyad, Perona, Crumpler y Aubele, 1998).

- *Telecomunicaciones*: Establecimiento de patrones de llamadas. Modelos de carga en redes. Detección de fraude.
- *Otras áreas*: *Correo electrónico y agendas personales*: clasificación y distribución automática de correo, detección de correo spam, gestión de avisos, análisis del empleo del tiempo. *Recursos humanos*: selección de empleados. *Web*: análisis del comportamiento de los usuarios, detección de fraude en el comercio electrónico, análisis de los logs de un servidor web. *Turismo*: determinar las características socioeconómicas de los turistas en un determinado destino o paquete turístico, identificar patrones de reservas, etc. *Tráfico*: modelos de tráfico a partir de fuentes diversas: cámaras y gps. *Hacienda*: detección de evasión fiscal. *Policiales*: identificación de posibles terroristas en un aeropuerto. *Deportes*: estudio de la influencia de jugadores y de cambios. Planificación de eventos. *Política*: diseño de campañas políticas, estudios de tendencias de grupos, etc.

Todos estos ejemplos muestran la gran variedad de aplicaciones en las cuales el uso de la minería de datos puede ayudar a entender mejor el entorno donde se desenvuelve la organización y, en consecuencia, mejorar la toma de decisiones en dicho entorno.

Tendencias en el uso de técnicas y algoritmos en DM

Lo que se sabe acerca de los datos es que, seguirán creciendo. Ya no se habla de terabytes, sino de petabytes, zettabytes y medidas superiores (**Big Data**). Según publicaciones recientes, una de las tendencias actuales en el área de negocios, es el Análisis Predictivo en tiempo real, que da respuesta a interrogantes tales como: ¿Qué pasará? ¿Cómo podemos hacer que suceda? (DATAPINE, 2017). Entre los métodos para Análisis Predictivo que más atraen actualmente se encuentran Artificial Neural Networks (ANN) y Autoregressive Integrated Moving Average (ARIMA).

Las **ANN**, procesan los datos de forma similar que las redes neuronales biológicas. Al igual que en el cerebro humano, el poder de las redes neurales radica en su capacidad de conectar conjuntos de neuronas juntos en capas y crear una red multidimensional. La entrada de la segunda capa es la salida de la primera capa, y la situación se repite con

cada capa. Se basan en una idea sencilla: dados unos parámetros hay una forma de combinarlos para predecir un cierto resultado. Se ha adquirido mayor potencia predictiva migrando a redes multicapas. Últimamente las redes convolucionales están siendo utilizadas en el reconocimiento de voz y procesamiento de imágenes satelitales. Estos algoritmos son parte de lo que hoy se conoce como **“deep learning”**.

ARIMA (modelo autoregresivo integrado de promedio móvil), Se trata de un modelo dinámico de series temporales, es decir, las estimaciones futuras vienen explicadas por los datos del pasado y no por variables independientes. El análisis incluye la inspección de las autocorrelaciones (comparar cómo los valores de los datos actuales dependen de los valores del pasado) en especial la elección de cómo se deben tomar en consideración muchos pasos en el pasado a la hora de hacer predicciones. Cada parte de ARIMA se encarga de diferentes lados de la creación del modelo: “Autoregressive Part” (AR), que trata de estimar el valor actual, considerando la anterior. Cualquier diferencia entre los datos predichos y valor real son utilizadas por la parte de “Moving Average” (MA). Es posible comprobar si estos valores son normales, al azar y estacionarios con una variación constante. Cualquier desviación de estos puntos puede aportar información sobre el comportamiento de las series de datos, la predicción de nuevas anomalías o ayudar a descubrir patrones subyacentes que no son visibles.

Los algoritmos de “Clasificación”, tanto para describir, como para predecir, siguen siendo los más populares, tanto Ad-Hoc como Post-Hoc. La tendencia es usar “múltiples clasificadores” (Boosting, Bagging o variaciones de los mismos), para mejorar el poder predictivo y usar técnicas de remuestreo (undersampling, oversampling o algoritmos híbridos), para afrontar el desbalance de datos situación que se presenta cuando una o más clases mayoritarias se encuentran más representadas por una o más clases minoritarias (Calviño, 2017).

A pesar de la madurez adquirida por las técnicas, algoritmos utilizados y de su constante evolución, una característica en común es su carácter hipotético, es decir, lo aprendido puede, en cualquier momento, ser refutado por evidencia futura. En casi la totalidad de los problemas en que se aplican los modelos obtenidos no aspiran a ser modelos perfectos, sino modelos aproximados. (Hasperué, 2014).

Tendencias en el uso de herramientas para DM

A fines del año 2007 KDnuggets, página dedicada al Data Mining y al Análisis de Datos, creada y desarrollada por **Gregory Piatetsky-Shapiro**, realizó una encuesta en la que consultó a sus seguidores qué software para minería de datos utilizó durante los últimos 12 meses.

De la encuesta participaron 2.895 votantes, quienes eligieron 102 herramientas diferentes. Como puede observarse en la Figura 6, R es la herramienta más utilizada, con un 49% de electores (frente al 46,9% en 2015). Sin embargo, el uso de Python creció más rápido y casi alcanzó a R con un 45,8% de electores (frente al 30,3%). RapidMiner se mantiene como la plataforma general más popular para la ciencia de minería de datos, con alrededor del 33% de participación.

La participación de la región fue: Estados Unidos / Canadá (40%), Europa (39%), Asia (9,4%), América Latina (5,8%), África / Medio Oriente (2,9%), Australia / Nueva Zelanda (2,2%).

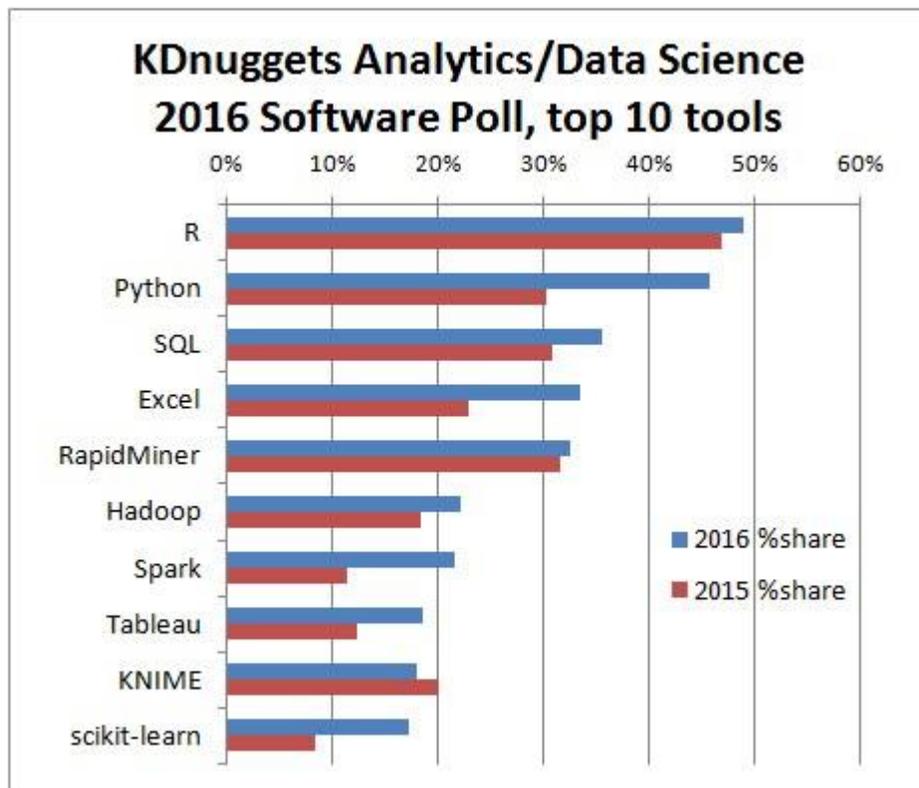


Figura 6. Software más usado en Minería de datos (KDnuggets, 2016).

Learning Analytics

Dietz-Uhler y Hurn (2013) definen “Learning Analytics” como una disciplina emergente relacionada con el desarrollo de métodos para explorar series de datos procedentes de ecosistemas educativos. Y con el uso posterior de los resultados del análisis para entender mejor al alumnado, sus comportamientos y así mejorar el diseño de los entornos en los que aprenden.”

Se están desarrollando servicios que recopilan datos de las herramientas educativas para entender y analizar el proceso de aprendizaje de los estudiantes (LA) o del propio proceso educativo (Academic Analytics, AA) (Long y Siemens, 2011), mostrando los resultados a través de interfaces visuales interactivas de manera que la visualización de los datos y la interacción con estas visualizaciones se convierte en una parte de suma importancia en el proceso de análisis y de la toma posterior de decisiones (Visual Analytics, VA) (Gómez, Therón, y García-Peñalvo, 2009). (Ver Figura 2)

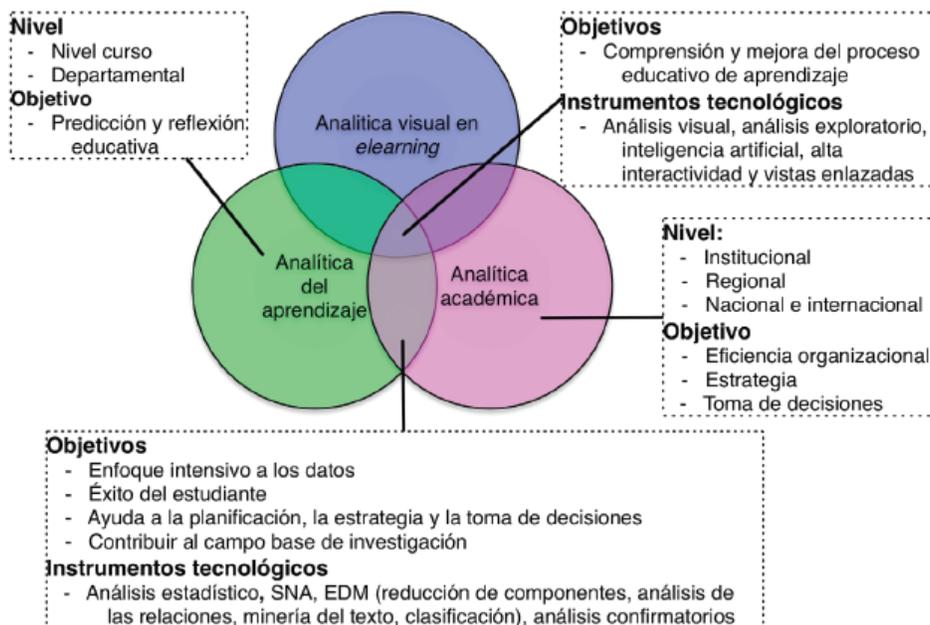


Figura 7. Analítica en educación (Gómez, Therón, y García-Peñalvo, 2009).

El objetivo del LA es el uso de los datos generados por los EVA y de cualquier otra observación adicional de la que se pueda obtener información, para tener un impacto directo sobre los estudiantes, los profesores y el proceso de aprendizaje (Long y Siemens,

2011). Se centra en el desarrollo de métodos para explorar los datos procedentes del contexto educativo y formativo, en busca de la mejora del proceso de aprendizaje. La Figura 3 muestra los pasos básicos para probar una hipótesis relacionada con el proceso de aprendizaje/enseñanza, pasos que se pueden encontrar también en procesos de MDE (Romero, Ventura y García, 2008). Estos pasos, que se describen a continuación, son la recopilación de datos y pre-procesado, análisis y acción y, por último, el post-procesamiento de los datos.

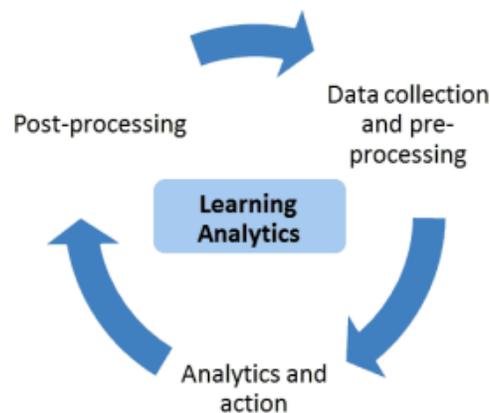


Figura 8. Proceso del Learning Analytics (Chatti, Dyckhoff, Schroeder y Thüs, 2012).

- **Recopilación de datos y pre-procesamiento:** Los datos educativos son la base del proceso de LA. Recopilar datos de diversos entornos y sistemas educativos para el descubrimiento exitoso de patrones útiles de los datos es un paso crítico. Los datos recopilados pueden ser demasiado grandes y / o implicar muchos atributos irrelevantes, que requieren el procesamiento previo de datos (Liu, 2006). El pre-procesamiento de datos también permite transformar los datos en un formato adecuado que se puede utilizar como entrada para un método LA particular. En este paso se pueden utilizar varias tareas de pre-procesamiento de datos, tomadas del campo de minería de datos. Estos incluyen la limpieza de datos, la integración de datos, la transformación de datos, la reducción de datos, el modelado de datos, la identificación de usuarios y sesiones y la finalización de la trayectoria (Han y Kamber, 2006; Liu, 2007; Romero y Ventura, 2007; Chatti, Dyckhoff, Schroeder y Thüs, 2012).

- **Análisis y acción:** una vez pre-procesados los datos diferentes técnicas de LA pueden ser aplicadas para explorarlos con el fin de descubrir patrones ocultos que pueden ayudar a proporcionar una experiencia de aprendizaje más efectiva. El paso de análisis no sólo incluye el análisis y la visualización de la información, sino también las acciones sobre esa información. Tomar acciones es el objetivo principal de todo el proceso analítico. Estas acciones incluyen monitoreo, análisis, predicción, intervención, evaluación, adaptación, personalización, recomendación y reflexión. (Bienkowski, Feng y Means 2012).
- **Post-procesamiento:** Para la mejora continua del ejercicio de análisis el post-procesamiento es una pieza fundamental. Puede implicar la compilación de nuevos datos a partir de fuentes de datos adicionales, el perfeccionamiento del conjunto de datos, la determinación de nuevos atributos necesarios para la nueva iteración, la identificación de nuevos indicadores / métricas, la modificación de las variables de análisis o la elección de un nuevo método analítico (Chatti, Dyckhoff, Schroeder y Thüs, 2012).

Como puede observarse, la dimensión pedagógica se encuentra ausente en el proceso de LA. Greller y Drachsler (2012) en su trabajo describen la relación existente entre estos conceptos. El LA puede dar apoyo a una multitud de estrategias pedagógicas y actividades de aprendizaje gracias al descubrimiento, representación y manifestación del conocimiento oculto tras los datos disponibles. Es decir, solo se puede ver la parte pedagógica a través de los datos. Es por esto que no se incluye como parte del proceso de análisis, pero implícitamente ya está contenida en los conjuntos de datos de entrada que encapsulan el comportamiento pedagógico de los usuarios. Como es sabido, este comportamiento depende en gran medida de la plataforma y la visión pedagógica de los desarrolladores (Anderson y Dron, 2011).

Van Harmelen y Workman (2012) mencionan que tal como se aplica la analítica en el ámbito educativo está proporcionando avances mensurables para el aprendizaje y la enseñanza, y ofrece la esperanza de tomar las decisiones más convenientes basadas en la evidencia, la acción y la personalización en diversas áreas de la educación. En este contexto los beneficiarios son los estudiantes, profesores, departamentos, instituciones educativas y los interesados a niveles regionales y nacionales.

2.3 MINERÍA DE DATOS APLICADA AL CAMPO EDUCATIVO. CONCEPTUALIZACIÓN Y ESTADO DEL ARTE.

Las instituciones de educación superior se han interesado durante mucho tiempo en predecir las trayectorias de los estudiantes y egresados (Luan, 2004). Durante las últimas décadas, las innovaciones más importantes en los sistemas educativos están relacionadas con la introducción de nuevas tecnologías como la educación basada en entornos de aprendizaje virtual (Ha, Bae y Park, 2000).

Las plataformas e-learning y el modo de aprendizaje virtual que ello conlleva, están siendo utilizadas en las instituciones de ámbito educativo, especialmente en el nivel de educación superior, tanto para complementar con nuevos contenidos y actividades las clases presenciales, conocida como educación semipresencial o blended learning, como impartir cursos íntegramente virtuales, lo que comúnmente se denomina e-learning (García Saiz, 2016).

El uso de una plataforma virtual por parte de estudiantes y docentes ha generado una gran cantidad de información (Huapaya, Lizarralde, Arona y Massa, 2012) acerca del acceso y uso de la plataforma, así como la interacción generada a partir del desarrollo de las actividades programadas en el curso. La información generada, procesada con las herramientas adecuadas, permite predecir, prevenir y/o actuar para mejorar el rendimiento académico de los estudiantes. Los entornos de aprendizaje virtual permiten modelar el comportamiento de estos estudiantes mediante aplicación de técnicas de minería de datos.

Bienkowski, Feng, y Means (2012) definen “Educational Data Mining” *como la teoría que desarrolla métodos, aplica técnicas estadísticas y de aprendizaje automático para analizar los datos recogidos durante el proceso de enseñanza y aprendizaje.*

Sin embargo, la MDE puede analizar los datos generados por cualquier sistema de aprendizaje y enfocarse en diversos aspectos tanto individuales como grupales y tener en cuenta datos subyacentes; datos administrativos, datos demográficos, datos motivacionales que a su vez contienen múltiples niveles de jerarquía, contextos, niveles de granularidad y datos históricos.

Como consecuencia del creciente interés en el estudio de la MDE (Peña Ayala, 2014; Romero y Ventura, 2010; Romero y Ventura, 2007), ha sido creada una asociación

internacional, la International Educational Data Mining Society (IEDM) (International educational data mining society, 2017), dedicada a organizar eventos y gestionar medios de publicación enfocados exclusivamente a la MDE.

La EDM involucra técnicas necesarias para la construcción de modelos a partir de la información disponible. Este proceso puede utilizar dos enfoques bien definidos: descriptivo o predictivo. En el primer enfoque, se intenta caracterizar las situaciones de manera tal de poder entender cuáles son las variables principales que describen el comportamiento de una situación particular. El segundo enfoque implica la utilización del modelo para establecer de antemano una situación problemática. Trabajos en el área, dejan ver cómo la aplicación de estas técnicas puede ser útil para mejorar el proceso de enseñanza-aprendizaje en los cursos virtuales, siendo la predicción del rendimiento de los alumnos uno de los problemas más ampliamente demandados y estudiados. Este nuevo conocimiento es útil tanto para los docentes como para los alumnos. A los alumnos se les puede recomendar actividades y recursos que favorezcan su aprendizaje, y los docentes pueden obtener una retroalimentación objetiva para su enseñanza. Los profesores pueden evaluar la estructura del curso y su eficacia en el proceso de aprendizaje, y también, clasificar a los alumnos en grupos en función de sus necesidades de orientación y seguimiento (Vega, Romero y Menéndez, 2016).

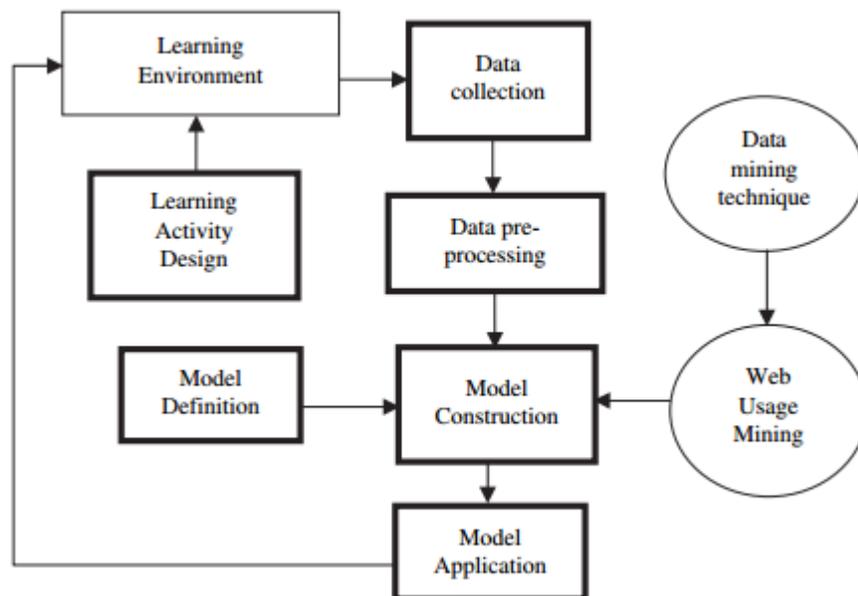


Figura 1. Modelo para el descubrimiento de conocimiento en un entorno e-learning (Wang, 2006).

En la Figura 1, se muestra el proceso de descubrimiento de conocimiento en un entorno

de aprendizaje virtual, en el caso planteado por Wang desde la actividad histórica de los usuarios. Consta de seis tareas, cada una de las cuales deben ser tratadas adecuadamente. La primera es el *diseño de las actividades de aprendizaje*, diseño que depende de los propósitos específicos de las aplicaciones e-learning. Antes de pasar al siguiente paso, se necesita decidir qué actividades de aprendizaje se proporcionan a los estudiantes.

A continuación es necesario realizar un seguimiento de los datos con el fin de facilitar la recopilación de los mismos de las tareas de aprendizaje antes seleccionadas (*Data Tracking*). La persona encargada de esta etapa necesita seleccionar los atributos de seguimiento adecuados como el ID de usuario, la fecha y el período de tiempo de la tarea de aprendizaje que realiza un usuario, etc.

El pre-procesamiento de los datos (*Data Pre-processing*) es una de las tareas más importantes. En esta etapa los datos deben ser limpiados y transformados adecuadamente antes de ser analizados.

El siguiente paso es la *definición y construcción del modelo*. Los investigadores tienen que definir un modelo de comportamiento que refleje los temas que les interesa. Después de la definición del modelo, se derivaran los métodos eficientes y efectivos para construirlo.

Finalmente viene la fase de *aplicación* en la que se definen métricas de evaluación con el objeto de evaluar la aptitud y efectividad del modelo desarrollado.

Algunos de los algoritmos de Data mining más utilizados en esta área de investigación son:

- *Association Mining*: la asociación, más comúnmente conocida como reglas de asociación, es una técnica de minería útil para descubrir las correlaciones entre los elementos de una gran base de datos. Esta técnica explora la probabilidad de que cuando ciertos elementos estén presentes, otros también lo estén, es decir, encontrar hechos que ocurren en juntos dentro de un conjunto de datos.
- *Clustering*: o agrupamiento, es una técnica que permite agrupar los objetos físicos o abstractos en clases de objetos similares. El clustering se basa en la metodología de "divide y conquista" que descompone un sistema de gran escala en componentes más pequeños con el objeto de simplificar el diseño y la implementación (Chen, Han y Yu, 1996). El objetivo del clustering es maximizar la similitud dentro de un grupo

de objetos y minimizar la similitud entre los grupos de objetos.

- *Web usage mining*: el término minería de uso web fue introducido por Cooley, Mobasher y Srivastava (1997) quienes la definen como el "descubrimiento automático de patrones de acceso de usuarios desde servidores web". A través de ella es posible descubrir el comportamiento de navegación de los usuarios y de esta manera personalizar un sitio web según el comportamiento y perfil de cada usuario.

Los entornos de aprendizaje virtual, también requieren de un adecuado **Sistema de Recomendación**, que provea las sugerencias apropiadas al estudiante, basadas en sus requerimientos, preferencias y resultados esperados. En general existen dos tipos de sistemas de recomendación:

- *Content-based filtering systems*: sistemas de recomendación basados en contenido, emplean técnicas de recuperación de información. Por ejemplo, un documento de texto es recomendado basado en una comparación entre su contenido y el del perfil del usuario construido analizando el contenido de elementos que el usuario ha valorado en el pasado y/o la información personal y preferencias del usuario.
- *Collaborative filtering systems*: sistemas de filtrado colaborativo, en estos los artículos se recomiendan a un usuario particular cuando otros usuarios similares también los prefieren. El sistema de filtrado colaborativo recopila toda la información sobre las actividades de los usuarios en el sitio web y calcula la similitud entre los usuarios. Si algunos usuarios tienen un comportamiento similar, se clasificarán en el mismo grupo de usuarios.

Dada la madurez de las técnicas de Data Mining, los investigadores han explotado *Sistemas de Recomendación basados en Reglas de Asociación* (Wang, 2006).

2.4 MONITOREO Y PREDICCIÓN DEL RENDIMIENTO ACADÉMICO A TRAVÉS DE LA MINERÍA DE DATOS. ESTUDIO DE CASOS.

La MDE ha dado lugar a un crecimiento en el número de estudios publicados buscando dar respuesta a situaciones particulares. A continuación se expondrán algunos de estos, cada uno de los cuales ha utilizado diferentes técnicas de minería de datos con

el objetivo de responder a preguntas específicas de educación.

Técnica	Redes Neuronales
Año	1993
Citas	21
Autor/es	T. D. Gedeon, S. Turner
Título	Explaining student grades predicted by a neural network ²
Resumen	Se entrenó una red neuronal para predecir el desempeño del estudiante en una materia de ciencias de la computación de la universidad de New South Wales. El propósito es permitir a los estudiantes predecir la calificación final que es probable lograr con base en el desempeño actual.

Técnica	Redes Neuronales
Año	1994
Citas	30
Autor/es	L.V. Fausett, W. Elwasif
Título	Predicting performance from test scores using backpropagation and counterpropagation ³

² Gedeon, T. D., & Turner, S. (1993, October). Explaining student grades predicted by a neural network. In *Neural Networks, 1993. IJCNN'93-Nagoya. Proceedings of 1993 International Joint Conference on* (Vol. 1, pp. 609-612). IEEE.

³Fausett, L. V., & Elwasif, W. (1994). Predicting performance from test scores using backpropagation and counterpropagation. In *Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on* (Vol. 5, pp. 3398-3402). IEEE.

Resumen	Dos redes neuronales, backpropagation y contrapropagación, son entrenadas para predecir las calificaciones de los estudiantes en la asignatura Cálculo I. Se investiga el efecto del número de unidades ocultas.
----------------	--

Técnica	Reglas de asociación
Año	1995
Citas	23
Autor/es	Arun P. Sanjeev, Jan M. Żytkow
Título	Discovering enrollment knowledge in university databases ⁴
Resumen	El objetivo de este trabajo ha sido descubrir el conocimiento útil en la comprensión de la inscripción universitaria para encontrar distintas maneras de aumentar la matrícula.

Técnica	Reglas de Asociación
Año	2000
Citas	96
Autor/es	Sung Ho Ha, Sung Min Bae, Sang Chan Park
Título	Web mining for distance education ⁵

⁴ Sanjeev, A. P., & Żytkow, J. M. (1995, August). Discovering Enrollment Knowledge in University Databases. In KDD (pp. 246-251).

⁵ Ha, S. H., Bae, S. M., & Park, S. C. (2000). Web mining for distance education. In Management of Innovation and Technology, 2000. ICMIT 2000. Proceedings of the 2000 IEEE International Conference on (Vol. 2, pp. 715-719). IEEE.

Resumen	Aplicación de técnicas de minería de datos con el objetivo de establecer patrones en los gustos académicos de los estudiantes y la eficacia en la estructura de los cursos. En este trabajo se presenta una visión general de la educación a distancia, se describen las posibilidades de aplicación de la minería de datos y, en consecuencia, se demuestra que el uso de ésta para fines educativos es de gran interés.
----------------	---

Técnica	Reglas de Asociación
Año	2001
Citas	59
Autor/es	Chang-Jiun Tsai, Shian-Shyong Tseng, Chih-Yang Lin
Título	A Two-Phase Fuzzy Mining and Learning Algorithm for Adaptive Learning Environment ⁶
Resumen	En este trabajo se propone la integración de un algoritmo de minería de datos y la teoría de conjuntos difusos con el objetivo de descubrir conocimiento antes oculto para que los docentes de los cursos puedan refinar o reorganizar los materiales de aprendizaje y pruebas en entornos de aprendizaje virtual.

Técnica	Reglas de Asociación
Año	2002
Citas	74
Autor/es	K. Yamanishi, Hang Li

⁶ Tsai, C. J., Tseng, S., & Lin, C. Y. (2001). A two-phase fuzzy mining and learning algorithm for adaptive learning environment. Computational Science-ICCS 2001, 429-438.

Título	Mining open answers in questionnaire data ⁷
Resumen	El propósito de este proyecto ha sido el desarrollo de un sistema de análisis de encuestas que permita la extracción de información útil para la toma de decisiones. El sistema se basa en la utilización de dos técnicas de aprendizaje estadístico: las reglas de asociación y el análisis de correspondencias.

Técnica	Reglas de Asociación
Año	2002
Citas	25
Autor/es	Feng-Hsu Wang
Título	On Using Data-Mining Technology for Browsing Log File Analysis in Asynchronous Learning Environment ⁸
Resumen	Desarrollo de una herramienta que permita el análisis de los archivos de log con el objetivo de que el conocimiento extraído permita a los docentes organizar la estructura de enseñanza con mayor eficiencia y además, identificar algunos patrones de aprendizaje interesantes o inesperados. La aplicación de la herramienta se lleva a cabo sobre una base de datos tomada de un curso virtual de la Universidad de Ming Chuan, Taiwán.

Técnica	Reglas de Asociación
Año	2002
Citas	360

⁷ Yamanishi, K., & Li, H. (2002). Mining open answers in questionnaire data. IEEE Intelligent Systems, 17(5), 58-63.

⁸ Wang, F. (2002). On using data-mining technology for browsing log file analysis in asynchronous learning environment. In Proceedings of the Conference on educational multimedia, hypermedia and telecommunication (pp. 2005-2006).

Autor/es	O.R. Zaiane
Título	Building a recommender agent for e-learning systems ⁹
Resumen	Este documento sugiere el uso de técnicas de minería de datos para construir un sistema que podría recomendar, en un curso virtual basado en el historial de acceso de los estudiantes, actividades de aprendizaje o accesos directos con el objeto de mejorar la navegación de materiales del curso, así como también ayudar al proceso de aprendizaje en línea.

Técnica	Redes Neuronales
Año	2002
Citas	29
Autor/es	Timothy Wang, Antonija Mitrovic
Título	Using Neural Networks to predict Student's Performance. ¹⁰
Resumen	Este documento propone para la selección de un problema el uso de un sistema basado en minería de datos, de manera que el mismo identifique el problema apropiado para un determinado estudiante. Con el fin de desarrollar un sistema de este tipo, que permita predecir el número de errores que un estudiante va a cometer, se hizo uso de una red neuronal (con avance y propagación hacia atrás).

Técnica	Reglas de Asociación
Año	2003

⁹ Zaiane, O. R. (2002, December). Building a recommender agent for e-learning systems. In Computers in Education, 2002. Proceedings. International Conference on (pp. 55-59). IEEE.

¹⁰ Wang, T., & Mitrovic, A. (2002, December). Using neural networks to predict student's performance. In Computers in Education, 2002. Proceedings. International Conference on (pp. 969-973). IEEE.

Citas	46
Autor/es	L. dos Santos Machado, Karin Becker
Título	Distance education: a Web usage mining case study for the evaluation of learning sites ¹¹
Resumen	Este trabajo describe un caso de estudio destinado a investigar el potencial de la minería de datos como marco para apoyar la validación de diseños de sitios de aprendizaje.

Técnica	Reglas de Asociación
Año	2004
Citas	77
Autor/es	Behrouz Minaei-Bidgoli, Pang-Ning Tan, and William F. Punch
Título	Mining Interesting Contrast Rules for a Web-based Educational System ¹²
Resumen	Este artículo se centra en el descubrimiento de reglas de asociación, que describen características interesantes de diferentes segmentos de una población. En el contexto de los sistemas educativos virtuales. Las reglas de asociación ayudan a identificar los atributos que caracterizan los patrones de desempeño dispar entre varios grupos de estudiantes. Esta técnica fue aplicada en entorno educativo virtual desarrollado en la Universidad Estatal de Michigan llamado LON-CAPA.

¹¹ Dos Santos Machado, L., & Becker, K. (2003, July). Distance education: A web usage mining case study for the evaluation of learning sites. In *Advanced Learning Technologies, 2003. Proceedings. The 3rd IEEE International Conference on* (pp. 360-361). IEEE.

¹² Minaei-Bidgoli, B., Tan, P. N., & Punch, W. F. (2004, December). Mining interesting contrast rules for a web-based educational system. In *Machine Learning and Applications, 2004. Proceedings. 2004 International Conference on* (pp. 320-327). IEEE.

Técnica	Reglas de Asociación
Año	2004
Citas	158
Autor/es	Cristóbal Romero, Sebastián Ventura, Paul De Bra
Título	Knowledge Discovery with Genetic Programming for Providing Feedback to Courseware Authors ¹³
Resumen	Este artículo muestra cómo descubrir, haciendo uso de las reglas de asociación, relaciones importantes entre los datos tomados durante las sesiones de los distintos usuarios en un entorno de aprendizaje virtual. Esta información puede ser de gran utilidad para el creador del curso, que puede decidir qué modificaciones son las más adecuadas para mejorar el rendimiento de los alumnos.

Técnica	Patrones secuenciales
Año	2004
Citas	82
Autor/es	Enric Mor, Julià Minguillón
Título	E-learning Personalization based on Itineraries and Long-term Navigational Behavior ¹⁴

¹³ Romero, C., Ventura, S., & Bra, P. D. (2004). Knowledge discovery with genetic programming for providing feedback to courseware authors. *User Modeling and User-Adapted Interaction*, 14(5), 425-464.

¹⁴ Mor, E., & Minguillón, J. (2004, May). E-learning personalization based on itineraries and long-term navigational behavior. In *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters* (pp. 264-265). ACM.

Resumen	En este trabajo se describe un marco práctico para el estudio del comportamiento de navegación de los usuarios de un entorno de e-learning integrado en un campus virtual. El objetivo principal es diseñar una herramienta de minería de datos para analizar el comportamiento de navegación de los usuarios y para extraer información relevante que se puede utilizar para validar varios aspectos relacionados con el diseño y usabilidad del campus virtual, Se busca, además, determinar la programación óptima para cada curso dependiendo del perfil de usuario.
----------------	--

Técnica	Patrones secuenciales
Año	2004
Citas	68
Autor/es	Li-ping Shen, Rui-min Shen
Título	Learning content recommendation service based-on simple sequencing specification ¹⁵
Resumen	En este trabajo se propone una forma de organizar los contenidos de aprendizaje en pequeñas unidades denominadas Objetos de Aprendizaje para que puedan ser utilizadas y reutilizadas de manera efectiva. Estas unidades son sistematizadas en una base de conocimiento. Un mecanismo de recomendación inteligente basado en reglas de secuenciación se introduce con detalle, donde las reglas se forman a partir de la base de conocimientos y análisis de la brecha de competencias.

Técnica	Reglas de asociación
Año	2004
Citas	139

¹⁵ Shen, L. P., & Shen, R. M. (2004, August). Learning content recommendation service based-on simple sequencing specification. In International Conference on Web-Based Learning (pp. 363-370). Springer Berlin Heidelberg.

Autor/es	Jie Lu
Título	A Personalized e-Learning Material Recommender System ¹⁶
Resumen	Esta investigación propone un sistema de recomendación de aprendizaje personalizado, que tiene como objetivo ayudar a los estudiantes a encontrar materiales de aprendizaje que tendrían que leer.

Técnica	Clustering
Año	2004
Citas	159
Autor/es	FH Wang, HM Shao
Título	Effective personalized recommendation based on time-framed navigation clustering and association mining ¹⁷
Resumen	Esta investigación propone un nuevo método de recomendación basado en las técnicas de agrupación. Las sesiones de navegación históricas para cada usuario se dividen en marcos de sesiones basadas en un intervalo de tiempo específico. Esta investigación propone un nuevo método de agrupación, denominado HBM (Hierarchical Bisecting Medoids Algorithm) para agrupar usuarios en base a las sesiones de navegación enmarcadas en el tiempo.
Técnica	Reglas de Asociación
Año	2005
Citas	30

¹⁶ Lu, J. (2004, January). Personalized e-learning material recommender system. In International conference on information technology for application (pp. 374-379).

¹⁷ Wang, F. H., & Shao, H. M. (2004). Effective personalized recommendation based on time framed navigation clustering and association mining. *Expert Systems with Applications*, 27(3), 365-377.

Autor/es	Azizul Azhar bin Ramli
Título	Web usage mining using apriori algorithm: UUM learning careportal case ¹⁸
Resumen	Con el fin de producir los patrones de uso de portales y comportamientos de usuario de la universidad E Learning (UUM Educare), este trabajo implementa el proceso de Minería de Datos usando el Algoritmo de A Priori basado en Reglas de Asociación

Técnica	Patrones secuenciales
Año	2005
Citas	23
Autor/es	Andrej Krištofič
Título	Recommender System for Adaptive Hypermedia Applications ¹⁹
Resumen	La mayoría de los sistemas hipermedia adaptables basados en la web adaptan la presentación del contenido y/o la navegación utilizando un conjunto predefinido de reglas. Teniendo en cuenta diferentes comportamientos y preferencias de cada usuario, puede ser difícil generalizar y construir todas las reglas apropiadas de antemano. Este trabajo explora este problema en el contexto de los sistemas educativos de hipermedia adaptativa. Se presenta la arquitectura de un sistema de recomendación que utiliza los resultados del descubrimiento de conocimiento para mejorar la adaptación.

Técnica	Patrones secuenciales
----------------	-----------------------

¹⁸ Bin Ramli, A. A. (2005). Web usage mining using apriori algorithm: UUM learning care portal case. In Proc. of the Int. Conf. on Knowledge Management (pp. 1-19).

¹⁹ Krištofič, A. (2005). Recommender system for adaptive hypermedia applications. In IIT. SRC 2005: Student Research Conference (p. 229).

Año	2005
Citas	222
Autor/es	Pythagoras Karampiperis, Demetrios Sampson
Título	Adaptive learning resources sequencing in educational hypermedia systems ²⁰
Resumen	Con el fin de seleccionar de forma adaptable los recursos de aprendizaje en sistemas de hipermedia educativos adaptativos (AEHS), se requiere la definición de las reglas de adaptación contenidas en el modelo. En este artículo se aborda el problema proponiendo un método de secuenciación alternativo que estima la idoneidad de los recursos de aprendizaje para un alumno en particular.

Técnica	Reglas de asociación
Año	2005
Citas	73
Autor/es	Penelope Markellou, Ioanna Mousourouli, Spiros Sirmakessis, Athanasios Tsakalidis
Título	Using semantic web mining technologies for personalized e-learning experiences ²¹
Resumen	En este artículo se propone un marco para el e-learning personalizado basado en perfiles de uso con el objeto de generar recomendaciones de materiales de aprendizaje en sistemas e-learning.

²⁰ Karampiperis, P., & Sampson, D. (2005). Adaptive learning resources sequencing in educational hypermedia systems. *Educational Technology & Society*, 8(4), 128-147.

²¹ Markellou, P., Mousourouli, I., Spiros, S., & Tsakalidis, A. (2005). Using semantic web mining technologies for personalized e-learning experiences. *Proceedings of the web-based education*, 461-826.

Técnica	Clustering
Año	2006
Citas	53
Autor/es	S. Retalis, A. Papasalouros, Y. Psaromiligkos, S. Siscos, T. Kargidis
Título	Towards Networked Learning Analytics – A concept and a tool ²²
Resumen	El objetivo de este trabajo es proponer un nuevo conjunto de herramientas para que el docente pueda evaluar fácilmente el progreso de sus alumnos y producir informes basados en el comportamiento de los estos dentro de un entorno de aprendizaje virtual. La innovación del conjunto de herramientas propuestas se apoya en el uso de las técnicas de minería de datos.

Técnica	Clustering
Año	2006
Citas	14
Autor/es	Jaime Spacco, Titus Winters, Tom Payne
Título	Inferring Use Cases from Unit Testing ²³
Resumen	El objetivo de este trabajo es analizar, a través de la técnica de minería de datos conocida como clustering, las relaciones existentes en una matriz de evaluación con el propósito de facilitar la toma de decisiones a los docentes.

²² Retalis, S., Papasalouros, A., Psaromiligkos, Y., Siscos, S., & Kargidis, T. (2006). Towards networked learning analytics—A concept and a tool. In Proceedings of the fifth international conference on networked learning (pp. 1-8).

²³ Spacco, J., Winters, T., & Payne, T. (2006, July). Inferring use cases from unit testing. In AAAI workshop on educational data mining (pp. 1-7)

Técnica	Patrones secuenciales
Año	2006
Citas	75
Autor/es	Judy Kay, Nicolas Maisonneuve, Kalina Yacef, Osmar Zaiane
Título	Mining patterns of events in students' teamwork data ²⁴
Resumen	El artículo describe la extracción y análisis de los datos de las interacciones de los estudiantes en un entorno de aprendizaje virtual para identificar secuencias significativas respecto a problemas o logros con el fin de apoyar a los alumnos a solucionar dificultades.

Técnica	Redes Bayesianas
Año	2006
Citas	24
Autor/es	Elizabeth Ayers, Brian Junker
Título	Do skills combine additively to predict task difficulty in eighth-grade mathematics? ²⁵
Resumen	Usado para predecir el resultado del examen de fin de año a través de la actividad del estudiante con tutores en línea.

²⁴ Kay, J., Maisonneuve, N., Yacef, K., & Zaiane, O. (2006, June). Mining patterns of events in students' teamwork data. In Proceedings of the Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems (ITS 2006) (pp. 45-52).

²⁵ Ayers, E., & Junker, B. W. (2006, July). Do skills combine additively to predict task difficulty in eighth grade mathematics. Educational Data Mining: Papers from the AAAI Workshop. Menlo Park, CA: AAAI Press.

Técnica	Algoritmos de clasificación
Año	2007
Citas	16
Autor/es	Hurley T., Weibelzahl S.
Título	Using MotSaRT to support online teachers in student motivation ²⁶
Resumen	Usado para predecir en qué casos el profesor debe recomendar ciertas estrategias de motivación a estudiantes con determinados perfiles establecidos.

Técnica	Clustering / Reglas de asociación
Año	2007
Citas	25
Autor/es	M. Vranic, D. Pintar, Z. Skocir
Título	The use of data mining in education environment ²⁷
Resumen	Usados para mejorar algunos aspectos cualitativos del proceso de enseñanza.

Técnica	Reglas de asociación
Año	2007

²⁶ Hurley, T., & Weibelzahl, S. (2007, September). Using MotSaRT to support on-line teachers in student motivation. In European Conference on Technology Enhanced Learning (pp. 101-111). Springer Berlin Heidelberg.

²⁷ Vranic, M., Pintar, D., & Skocir, Z. (2007, June). The use of data mining in education environment. In Telecommunications, 2007. ConTel 2007. 9th International Conference on (pp. 243-250). IEEE.

Citas	17
Autor/es	Cássia Blondet Baruque, Marília A. Amaral, Alexandre Barcellos, João Carlos da Silva Freitas, Carlos Juliano Longo
Título	Analysing users' access logs in Moodle to improve e learning ²⁸
Resumen	Usado para analizar datos de acceso a Moodle para mejorar el aprendizaje virtual.

Técnica	Patrones secuenciales
Año	2007
Citas	35
Autor/es	Hafidh Ba-Omar, Ilias Petrounias, Fahad Anwar
Título	A Framework for Using Web Usage Mining to Personalise E-learning ²⁹
Resumen	Usado para desarrollar escenarios de aprendizaje personalizados que los estudiantes pueden usar mediante sistemas asistidos alimentados por patrones basados en estilos de aprendizaje.

Técnica	Patrones secuenciales
Año	2007
Citas	30

²⁸ Baruque, C. B., Amaral, M. A., Barcellos, A., da Silva Freitas, J. C., & Longo, C. J. (2007, May). Analysing users' access logs in Moodle to improve e learning. In Proceedings of the 2007 Euro American conference on Telematics and information systems (p. 72). ACM.

²⁹ Ba-Omar, H., Petrounias, I., & Anwar, F. (2007, July). A framework for using web usage mining to personalise e-learning. In Advanced Learning Technologies, 2007. ICALT 2007. Seventh IEEE International Conference on (pp. 937-938). IEEE.

Autor/es	Yang OUYANG, Miaoliang Zhu
Título	eLORM: Learning Object Relationship Mining based Repository ³⁰
Resumen	Descubrimiento de patrones que permitan recomendar objetos de aprendizaje a los estudiantes.

Técnica	Patrones secuenciales
Año	2007
Citas	35
Autor/es	Feng-jung Liu, Bai-jiun Shih
Título	Learning Activity-Based E-Learning Material Recommendation System ³¹
Resumen	Trabajo realizado para elaborar el diseño de un sistema de recomendación de materiales basado en las acciones de aprendizaje de estudiantes anteriores de los cursos.

Técnica	Clustering
Año	2007
Citas	22
Autor/es	F Lu, X Li, Q Liu, Z Yang, G Tan, T He

³⁰ Ouyang, Y., & Zhu, M. (2008). eLORM: learning object relationship mining-based repository. Online Information Review, 32(2), 254-265.

³¹ Liu, F. J., & Shih, B. J. (2007, December). Learning activity-based e-learning material recommendation system. In Multimedia Workshops, 2007. ISMW'07. Ninth IEEE International Symposium on (pp. 343-348). IEEE.

Título	Research on Personalized E-Learning System Using Fuzzy Set Based Clustering Algorithm ³²
Resumen	Provee recomendaciones sobre materiales de cursos basado en los hábitos de aprendizaje.

Técnica	Redes bayesianas
Año	2007
Citas	34
Autor/es	Nguyen Thi Ngoc Hien, Peter Haddawy
Título	A decision support system for evaluating international student applications ³³
Resumen	Usado para predecir el rendimiento de los estudiantes.
Técnica	Redes bayesianas
Año	2007
Citas	36
Autor/es	Zachary A., Pardos Neil, T. Heffernan, Brigham Anderson, Cristina L. Heffernan
Título	The Effect of Model Granularity on Student Performance Prediction Using Bayesian Networks ³⁴

³² Lu, F., Li, X., Liu, Q., Yang, Z., Tan, G., & He, T. (2007). Research on personalized e-learning system using fuzzy set based clustering algorithm. Computational Science–ICCS 2007, 587-590.

³³ Hien, N. T. N., & Haddawy, P. (2007, October). A decision support system for evaluating international student applications. In Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07. 37th Annual (pp. F2A-1). IEEE.

³⁴ Pardos, Z. A., Heffernan, N. T., Anderson, B., & Heffernan, C. L. (2007, July). The effect of model granularity on student performance prediction using Bayesian networks. In International Conference on User Modeling (pp. 435-439). Springer Berlin Heidelberg.

Resumen	Usado para modelar el conocimiento del usuario y predecir el desempeño del estudiante en un sistema de tutoría.
----------------	---

Técnica	Árboles de decisión / Redes bayesianas
Año	2008
Citas	18
Autor/es	J. Ranjan, S. Khalil
Título	Conceptual framework of data mining process in management education in India: An institutional perspective ³⁵
Resumen	Uso de técnicas para apoyar el proceso de admisión, la calidad del proceso de educación y el rendimiento de los estudiantes.

Técnica	Reglas de asociación
Año	2008
Citas	103
Autor/es	A. Merceron, K. Yacef
Título	Interestingness measures for association rules in educational data ³⁶
Resumen	Usado para analizar los datos de aprendizaje y determinar si los estudiantes utilizan los recursos y cuáles de ellos pueden tener mayor impacto.

³⁵ Ranjan, J., & Khalil, S. (2008). Conceptual framework of data mining process in management education in India: An institutional perspective. *Information Technology Journal*, 7(1), 16-23.

³⁶ Merceron, A., & Yacef, K. (2008, June). Interestingness measures for association rules in educational data. In *Educational Data Mining 2008*.

Técnica	Reglas de asociación
Año	2008
Citas	19
Autor/es	Sebastian Ventura, Cristóbal Romero, César Hervás
Título	Analyzing Rule Evaluation Measures with Educational Datasets: A Framework to Help the Teacher ³⁷
Resumen	Usado para analizar las medidas de reglas de evaluación de conjuntos de datos educativos con el fin de identificar patrones interesantes.

Técnica	Reglas de asociación
Año	2008
Citas	28
Autor/es	Farah Habib Chanchary, Indrani Haque, Md. Saifuddin Khalid
Título	Web Usage Mining to Evaluate the Transfer of Learning in a Web-Based Learning Environment ³⁸
Resumen	El objetivo de este artículo es encontrar relaciones en el acceso al LMS y el comportamiento de los estudiantes con el fin de identificar patrones de uso de Internet por parte de los estudiante

³⁷ Ventura, S., Romero, C., & Hervás, C. (2008, June). Analyzing rule evaluation measures with educational datasets: A framework to help the teacher. In Educational Data Mining 2008

³⁸ Chanchary, F. H., Haque, I., & Khalid, M. S. (2008, January). Web usage mining to evaluate the transfer of learning in a web-based learning environment. In Knowledge Discovery and Data Mining, 2008. WKDD 2008. First International Workshop on (pp. 249-253). IEEE.

Técnica	Reglas de asociación
Año	2008
Citas	22
Autor/es	César Vialardi, Javier Bravo, Alvaro Ortigosa
Título	Improving AEH Courses through Log Analysis ³⁹
Resumen	El artículo se propone mejorar el diseño del curso a partir de recomendaciones generadas mediante el análisis de logs de los cursos.

Técnica	Reglas de asociación
Año	2008
Citas	9
Autor/es	Shijue Zheng, Shaojun Xiong, Yin Huang, Shixiao Wu
Título	Using Methods of Association Rules Mining Optimization in Web-Based Mobile-Learning System ⁴⁰
Resumen	Los autores se proponen encontrar relaciones entre atributos, estrategias de solución adoptadas por los alumnos, en un sistema de aprendizaje móvil basado en la web.

Técnica	Clustering / Clasificación / Reglas de asociación
----------------	---

³⁹ Vialardi, C., Bravo, J., & Ortigosa, A. (2008). Improving AEH courses through log analysis. *Journal of Universal Computer Science*, 14(17), 2777-2798.

⁴⁰ Zheng, S., Xiong, S., Huang, Y., & Wu, S. (2008, August). Using methods of association rules mining optimization in web-based mobile-learning system. In *Electronic Commerce and Security, 2008 International Symposium on* (pp. 967-970). IEEE.

Año	2008
Citas	30
Autor/es	Mykola Pechenizkiy, Toon Calders, Ekaterina Vasilyeva, Paul De sujetador
Título	Mining the student assessment data: Lessons drawn from a small scale case study
Resumen	En este trabajo se describe un caso de estudio de MDE basado en los datos recogidos durante la evaluación en línea de los estudiantes. El estudio del caso está dirigido a demostrar que incluso con un conjunto de datos de tamaño modesto y problemas bien definidos es bastante difícil obtener resultados significativos y verdaderamente interesantes. Las técnicas de minería de datos utilizadas fueron clustering, clasificación y análisis de asociación.

Técnica	Patrones secuenciales
Año	2008
Citas	15
Autor/es	Liang Zhang, Xiumin Liu, Xiujuan Liu
Título	Personalized Instructing Recommendation System Based on Web Mining ⁴¹
Resumen	Este artículo analiza los intereses de los usuarios de la web a través de la aplicación de técnicas de minería de datos sobre archivos log y haciendo uso del algoritmo conocido como A priori.

Técnica	Patrones secuenciales
----------------	-----------------------

⁴¹ Zhang, L., Liu, X., & Liu, X. (2008, November). Personalized instructing recommendation system based on web mining. In Young Computer Scientists, 2008. ICYCS 2008. The 9th International Conference for (pp. 2517-2521). IEEE.

Año	2008
Citas	11
Autor/es	John C. Nesbit, Y. Xu, Philip H Winne, Mingming Zhou
Título	Sequential pattern analysis software for educational event data ⁴²
Resumen	Este artículo explica cómo los archivos log se pueden utilizar para recopilar datos detallados y potencialmente críticos sobre lo que hacen los estudiantes al momento de interactuar con un entorno virtual de aprendizaje. También demuestra cómo los algoritmos secuenciales de minería de datos permiten a los investigadores educativos obtener información precisa y significativa acerca de cómo aprenden las personas.

Técnica	Patrones secuenciales
Año	2008
Citas	19
Autor/es	Sebastian Ventura, Cristóbal Romero, César Hervás
Título	Analyzing Rule Evaluation Measures with Educational Datasets: A Framework to Help the Teacher ⁴³
Resumen	En este trabajo se compara el rendimiento y la utilidad de diferentes técnicas de minería de datos que permiten clasificar a los estudiantes que hacen uso del entorno virtual Moodle. Se demuestra que algunos algoritmos mejoran su desempeño cuando se aplican tareas tales como preprocesamiento, discretización y reequilibrio de datos. Se concluye en que un buen modelo clasificador tiene que ser tanto preciso como comprensible para los docentes.

⁴² Nesbit, J. C., Xu, Y., Winne, P. H., & Zhou, M. (2008). Sequential pattern analysis software for educational event data. *Measuring Behavior* 2008, 160.

⁴³ Ventura, S., Romero, C., & Hervás, C. (2008, June). Analyzing rule evaluation measures with educational datasets: A framework to help the teacher. In *Educational Data Mining 2008*.

Técnica	Patrones secuenciales
Año	2008
Citas	35
Autor/es	Feng-Hsu Wang
Título	Content Recommendation Based on Education-Contextualized Browsing Events for Web-based Personalized Learning ⁴⁴
Resumen	Los archivos log muestran cómo los estudiantes utilizan los entornos de aprendizaje virtual. En este trabajo se presenta un nuevo método para la identificación de eventos de navegación construido a través de técnicas estadísticas y minería de datos. El modelo considera varios factores contextuales como los educativos (relevancia y pertinencia grupo de rendimiento). Sobre la base de este modelo, se construye un sistema de recomendación de contenido personalizado.

Técnica	Clasificación / Árboles de decisión
Año	2010
Citas	8
Autor/es	Osvaldo M. Sposito, Martín E. Etcheverry, Hugo L. Ryckeboer, Julio Bossero
Título	Aplicación de técnicas de minería de datos para la evaluación del

⁴⁴ Wang, F. H. (2008). Content recommendation based on education-contextualized browsing events for Web-based personalized learning. *Educational Technology & Society*, 11(4), 94-112.

	rendimiento académico y la deserción estudiantil ⁴⁵
Resumen	Este artículo presenta los resultados de la evaluación del rendimiento académico y de la deserción estudiantil de los estudiantes del Departamento de Ingeniería e Investigaciones Tecnológicas (DIIT) de la Universidad Nacional de La Matanza (UNLaM). La investigación se realizó aplicando el proceso de descubrimiento de conocimiento sobre los datos de alumnos del período 2003-2008.

Técnica	Análisis de correlación
Año	2012
Citas	4
Autor/es	Chalie Patarapichayatham, Akihito Kamata, Sirichai Kanjanawasee
Título	Evaluation of Model Selection Strategies for Cross-Level Two-Way Differential Item Functioning Analysis ⁴⁶
Resumen	Los problemas de especificación del modelo en el funcionamiento de los ítems diferenciales bidireccionales de nivel cruzado, odel, fueron previamente investigados por Patarapichayatham. Su estudio aclaró que una especificación de modelo incorrecta puede conducir fácilmente a estimaciones sesgadas de los parámetros clave. El objetivo de este artículo es proporcionar más información sobre el tema centrándose específicamente en el impacto de las estrategias de selección de modelos. En este estudio se compararon seis estrategias de selección de modelos. A través de análisis de datos simulados repetidamente, se evaluaron las frecuencias de cada modelo que se seleccionó como el mejor modelo y las estimaciones de parámetros.

⁴⁵ Sposito, O., Etcheverry, M., Ryckeboer, H., & Bossero, J. (2010). Aplicación de técnicas de minería de datos para la evaluación del rendimiento académico y la deserción estudiantil. In Novena Conferencia Iberoamericana en Sistemas, Cibernética e Informática, CISCI (Vol. 29, pp. 06-2).

⁴⁶ Patarapichayatham, C., Kamata, A., & Kanjanawasee, S. (2012). Evaluation of model selection strategies for cross-level two-way differential item functioning analysis. *Educational and Psychological Measurement*, 72(1), 44-51.

Técnica	Clasificación / Árboles de decisión
Año	2013
Citas	0
Autor/es	Ricardo Timarán Pereira, Andrés Calderón Romero, Javier Jiménez Toledo
Título	Aplicación de la minería de datos de deserción estudiantil ⁴⁷
Resumen	Este proyecto de investigación tiene por objetivo detectar patrones de deserción estudiantil a partir de los datos socioeconómicos, académicos, disciplinares e institucionales de los estudiantes de los programas de pregrado de la Universidad de Nariño e Institución Universitaria IUCESMAG, ambas instituciones colombianas. Se descubrieron perfiles socioeconómicos y académicos de los estudiantes que desertan utilizando la técnica de clasificación basada en árboles de decisión.

Técnica	Árboles de decisión
Año	2014
Citas	1
Autor/es	Yegny Karina Amaya Torrado, Edwin Barrientos Avendaño, Diana Judith Heredia Vizcaíno
Título	Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos ⁴⁸

⁴⁷ Pereira, R. T., Romero, A. C., & Toledo, J. J. (2013). Aplicación de la minería de datos en la extracción de perfiles de deserción estudiantil [Application of data mining in extracting student dropout profiles]. Ventana Informática, (28).

⁴⁸ Torrado, A., Karina, Y., Barrientos Avendaño, E., & Heredia Vizcaíno, D. J. (2014). Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos.

Resumen	Este artículo muestra la construcción de un modelo predictivo de deserción estudiantil, caracterizando a los estudiantes de la Universidad Simón Bolívar con el objetivo de poder predecir la probabilidad de deserción de los estudiantes; dicho modelo demostró el desempeño de los algoritmos presentados para clasificar datos bajo contextos variables y la precisión de uno con respecto al otro.
----------------	---

Técnica	Clasificación
Año	2015
Citas	1
Autor/es	Angélica Jaramillo, Henry Paz
Título	Aplicación de Técnicas de Minería de Datos para Determinar las Interacciones de los Estudiantes en un Entorno Virtual de Aprendizaje ⁴⁹
Resumen	Este artículo está enfocado en determinar las interacciones de los estudiantes del curso virtual de inglés en la Modalidad de Estudios a Distancia (MED) de la Universidad Nacional de Loja. Se realizó un análisis de la base de datos donde se encontraban los datos de los estudiantes correspondiente al periodo académico 2013-2014, para seleccionar los atributos necesarios que permitieron generar un modelo de minería de datos. Se realizó además un estudio de las diferentes técnicas de minería de datos de las cuales se seleccionó la que más se adapta al proyecto. La técnica de clasificación fue la elegida para generar los modelos a través de los algoritmos pertenecientes a la misma.

Técnica	Árboles de decisión
----------------	---------------------

⁴⁹ Jaramillo, A., & Arias, H. P. P. (2015). Aplicación de Técnicas de Minería de Datos para Determinar las Interacciones de los Estudiantes en un Entorno Virtual de Aprendizaje. Revista Tecnológica-ESPOL, 28.

Año	2015
Citas	5
Autor/es	Diana Heredia, Yegny Amaya, Edwin Barrientos
Título	Student Dropout Predictive Model Using Data Mining Techniques ⁵⁰
Resumen	En este trabajo se aplican diferentes técnicas de Data Mining sobre datos provenientes de un sistema de información académico universitario, para obtener un modelo predictivo que permita conocer de antemano qué estudiantes están en riesgo de abandonar sus estudios. Con la creación del modelo predictivo de deserción estudiantil se busca determinar la probabilidad de que un estudiante abandone la universidad.

Técnica	Árboles de decisión / Clustering
Año	2015
Citas	4
Autor/es	Ritika Saxena
Título	Educational Data Mining: Performance Evaluation of Decision Tree and Clustering Techniques using WEKA Platform ⁵¹

⁵⁰ Heredia, D., Amaya, Y., & Barrientos, E. (2015). Student dropout predictive model using data mining techniques. IEEE Latin America Transactions, 13(9), 3127-3134.

⁵¹ Saxena, R. (2015). Educational data Mining: Performance Evaluation of Decision Tree and Clustering Techniques using WEKA Platform. International Journal of Computer Science and Business Informatics.

Resumen	Las instituciones de educación superior necesitan saber qué estudiante se inscribirá en qué curso, qué estudiante necesita más ayuda. En la minería de datos los usuarios se enfrentan al problema cuando la base de datos consta de un gran número de características e instancias. Este tipo de problema (s) no podría ser manejado usando sólo árboles de decisión solo o la técnica de agrupamiento, debido a que los árboles de decisión dependen del conjunto de datos utilizado y de la configuración de los árboles, del mismo modo, que el agrupamiento por sí solo no funciona para todo tipo de patrones. Por lo tanto, para encontrar la técnica más adecuada, en este trabajo se han evaluado el rendimiento de ambos algoritmos. Los datos educativos se extraen y los algoritmos se aplican a él para predecir los resultados.
----------------	---

Técnica	Árboles de decisión
Año	2015
Citas	37
Autor/es	Sonali Agarwal, G. N. Pandey, and M. D. Tiwari
Título	Data Mining in Education: Data Classification and Decision Tree Approach ⁵²
Resumen	Este trabajo toma los datos de los estudiantes de bases de datos de colegios comunitarios con el objeto de realizar varios enfoques de clasificación y un análisis comparativo. Se propone utilizar la técnica de árbol de decisión dado que éste que puede ser tomado como una base importante para la selección de estudiantes durante cualquier programa del curso. El documento pretende demostrar la importancia de las técnicas de minería de datos para que la educación actual pueda adoptarlas como una herramienta de gestión estratégica.

⁵² Agarwal, S., Pandey, G. N., & Tiwari, M. D. (2015). Data mining in education: data classification and decision tree approach. *International Journal of e-Education, e-Business, e-Management and e-Learning*, 2(2), 140.

Técnica	Árboles de decisión
Año	2016
Citas	0
Autor/es	Jonny Esteban Sotomonte Castro, Cristian Camilo Rodríguez Rodríguez, Carlos Enrique Montenegro Marín, Paulo Alonso Gaona García, John Gabriel Castellanos
Título	Hacia la construcción de un modelo predictivo de deserción académica basado en técnicas de minería de datos ⁵³
Resumen	En este artículo se hace uso de la Minería de Datos con el objeto de generar un modelo de Árbol de Decisión implementando el algoritmo J48 mediante el uso de la herramienta WEKA. El objetivo esta investigación es identificar las causas del elevado índice de deserción.

Técnica	Clasificación / Árboles de decisión
Año	2016
Citas	0
Autor/es	Jhon Jaime Méndez Alandete
Título	Proyección de estudiantes en riesgo de desertar mediante técnicas de minería de datos ⁵⁴
Resumen	El objetivo de esta investigación fue la predicción de estudiantes desertores aplicando técnicas de minería de datos, utilizando algoritmos de clasificación y árboles de decisión sobre diferentes conjuntos de datos. Se obtuvo, como resultado, un conjunto de reglas para predecir el estudiante desertor de la Corporación Universitaria del Caribe – CECAR.

⁵³ Castro, J. E. S., Rodríguez, C. C. R., Marín, C. E. M., García, P. A. G., & Castellanos, J. G. (2016). Hacia la construcción de un modelo predictivo de deserción académica basado en técnicas de minería de datos. *Revista Científica*, 3(26), 35-48.

⁵⁴ Alandete, J. J. M. (2016). Proyección de estudiantes en riesgo de desertar mediante técnicas de minería de datos. *Ingeniería, Innovación y Desarrollo Sostenible*, 1(1), 23-35.

A partir del relevamiento y análisis de los casos expuestos se puede concluir que entre las técnicas de minería de datos utilizadas para abordar las distintas problemáticas planteadas en los diferentes escenarios educativos, las reglas de asociación fueron la técnica de minería de datos más usada, seguida por patrones secuenciales y árboles de decisión, como puede observarse en la Figura 8.

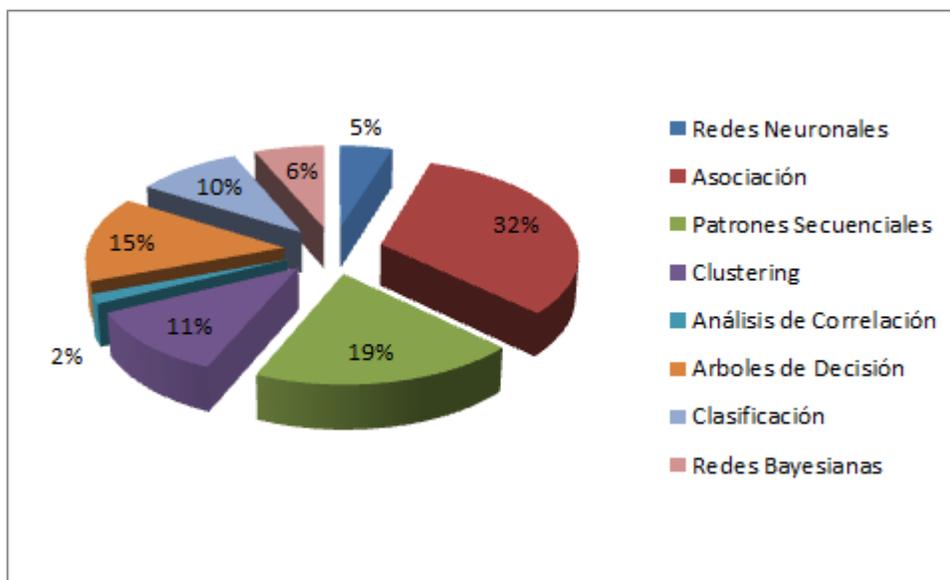


Figura 8. Técnicas de minería de datos más usadas (Elaboración propia, 2017).

CAPÍTULO 3: REZAGO ACADÉMICO Y DESERCIÓN ESTUDIANTIL UNIVERSITARIA EN LAS CARRERAS DEL ÁREA DE INFORMÁTICA DE LA UNNOBA.

3.1 ORIGEN Y COMPOSICIÓN DE LA ESCUELA DE TECNOLOGÍA DE LA UNNOBA.

La Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA) fue creada, por Decreto N°2.617 del Poder Ejecutivo Nacional, el 16 de diciembre de 2002.

Mediante la Ley N°25.824, del 19 de noviembre de 2003, el Congreso Nacional ratificó su creación. Tiene su sede en la ciudad de Junín, donde se encuentra el asiento de sus autoridades centrales, y una sede en la ciudad de Pergamino.

La concreción del proyecto institucional y académico de la UNNOBA tiene como antecedente la organización de dos Centros Universitarios Regionales, el Centro Universitario Regional Junín (CURJ), y el Centro Regional Universitario Pergamino (CRUP), que desarrollaron actividades de educación superior universitaria desde 1990 y 1991 respectivamente, y que en la actualidad constituyen las dos Sedes de la Universidad.

Académicamente está organizada con una estructura convergente de Escuelas y Departamentos, definidos a partir de cuatro áreas estratégicas de intervención que, se corresponden con la organización de las Escuelas: “Ciencias Agrarias, Naturales y Ambientales”; “Ciencias Económicas y Jurídicas”, “Tecnología” y “Ciencias Sociales y Humanas”. Las **Escuelas** son las unidades académicas donde se dictan las diferentes carreras de la universidad. Les corresponde la organización, correcto desarrollo y permanente actualización de las carreras que se dictan en cada una de ellas, siendo responsables de la estructura curricular y de los planes de estudio. El gobierno de cada Escuela es ejercido por un Consejo Directivo y un Director.

Los **Departamentos** se conforman en torno a las disciplinas o áreas del conocimiento específicas. Éstos proveen a las carreras de docentes requeridos en su disciplina o área. Los departamentos cuya temática sea exclusiva de una escuela dependerán de la misma. El Departamento de Ciencias Económicas, Jurídicas y Sociales depende de la Escuela de Ciencias Económicas y Jurídicas. El Departamento de Asignaturas Afines y Complementarias y el Departamento de Tecnología dependen de la Escuela de Tecnología.

Los Departamentos de Ciencias Básicas y Experimentales y de Humanidades cuya temática no corresponde específicamente a una Escuela, dependen de la Secretaría Académica de la Universidad, y proveen los docentes a las Escuelas cuyos planes de estudios así lo requieran. Este esquema de organización en Departamentos permite la movilidad horizontal del cuerpo docente, así como la flexibilidad en la estructura de funcionamiento académico de la Universidad.

En el año 2003 comenzó la etapa de organización y en el 2005 el dictado de su propia oferta académica. A fines del año 2008 se inició la organización de las Escuelas. La normalización de las Escuelas se concretó hacia mediados de 2011.

En lo que respecta a la Escuela de Tecnología, ésta tiene como fin la formación de profesionales con conciencia de compromiso social, en el área de las ingenierías, las informáticas y los diseños. Cuenta con un Director, dos Departamentos, una Secretaría Académica -de la cual dependen tres coordinaciones de carrera- y cinco Comisiones Asesoras. Esta Escuela incluye las carreras que se encuentran dentro de las áreas de Diseño, Informática e Ingeniería.

En relación con los objetivos y la labor que se lleva a cabo desde la Escuela de Tecnología, la UNNOBA ha impulsado y llevado a cabo la creación del Polo Tecnológico del Noroeste de la Provincia de Buenos Aires, una red de empresas e instituciones que reúnen investigaciones e industrias específicas con base tecnológica que tiene como fin promover el desarrollo de software y la inserción de Empresas Informáticas en la región.

A través de la Escuela de Tecnología la UNNOBA integra, a partir del año 2006, la Red de Universidades con Carreras en Informática, la Red de Universidades de Educación a Distancia y la Red provincial de grupos de Investigación y desarrollo en áreas de Ciencia de la Computación e Informática. También, desde el año 2009, forma parte del Consejo Federal de Decanos de Facultades de Ingeniería.

3.2 OFERTA ACADÉMICA Y MATRICULACIÓN EN EL ÁREA DE INFORMÁTICA DURANTE EL PERÍODO 2005-2015

La Universidad del Noroeste de la Provincia de Buenos Aires (UNNOBA) considera a las carreras de Informática como prioritarias en materia de asignación presupuestaria, y esto se evidencia en el crecimiento sostenido que se observa en el presupuesto de la misma desde el año 2008 a la fecha, arrojando un 72% de aumento para el año 2009, y un 150% para el año 2010 respecto del año anterior.

El 23% de los estudiantes de la Universidad cursan sus estudios en la Escuela de

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
ESTUDIANTES	189	214	442	610	717	582	629	652	621	501	528
NUEVOS INSCRIPTOS	---	139	226	326	282	335	165	149	112	102	110
REINSCRIPTOS	-----	50	216	284	435	247	464	503	509	399	418
EGRESADOS	-----	-----	-----	1	1	1	9	14	8	26	22

Tecnología y aproximadamente la mitad de ellos estudia carreras vinculadas a las ciencias tecnológicas, la ingeniería, la informática, la agroindustria, la metalmecánica y la producción de alimentos.

El área de Informática, perteneciente a la Escuela de Tecnología, se encuentra integrada por las siguientes carreras: **Analista de Sistemas, Licenciatura en Sistemas e Ingeniería en Informática**. Todas estas carreras comenzaron a dictarse a partir del año 2005. En el año 2010, en consonancia con la Resolución Ministerial N°786/09, se comenzó a implementar el nuevo Plan de Estudios, vigente en la actualidad, por el cual el título Analista de Sistemas pasó a ser Título intermedio de la Licenciatura en Sistemas. Ante esta modificación en los planes de estudio, se implementó un Plan de Transición (Resolución CS 310/10) estableciendo un régimen de equivalencias entre planes y el reconocimiento como materias optativas de materias que no fueron incorporadas al nuevo plan.

Si bien la carrera Licenciatura en Sistemas se dicta únicamente en la Sede de Junín, las asignaturas que son comunes al Título Intermedio (Analista de Sistemas) pueden ser cursadas por los alumnos en la Sede de Pergamino.

Tabla 1. Cantidad de estudiantes, nuevos inscriptos, reinscriptos y egresados de las carreras de Informática, correspondiente a la rama de estudio Ciencias Aplicadas de la Universidad Nacional del Noroeste de la Provincia de Buenos Aires.⁵⁵

La **Licenciatura en Sistemas** es la carrera, dentro de las que integran el área de Informática, que presenta el mayor número de alumnos durante el período 2005-2015. El siguiente cuadro muestra la relación entre la cantidad de ingresantes por año y la cantidad total de alumnos de la carrera durante dicho período. A partir de 2010 la cantidad de ingresantes se incrementó significativamente ya que se comenzaron a contabilizar los alumnos provenientes de la carrera Analista de Sistemas, que pasó a ser un título intermedio de esta licenciatura.

Año	Ingresantes	Total de alumnos
2005	24	24
2006	32	43
2007	33	55
2008	29	65
2009	25	65
2010	181*	233
2011	80	284
2012	68	294
2013	67	298
2014	66	292
2015	70	313

(*) De los cuales 101 son inscriptos 2010 y 80 han realizado el pase desde la carrera Analista de Sistemas

Tabla 2. Ingresantes y total de alumnos de la Licenciatura en Sistemas entre 2005 y 2015.

⁵⁵ Tabla elaborada en base a la información proporcionada por los Anuarios de Estadísticas Universitarias Argentinas (SPU, Ministerio de Educación, Ciencia y Tecnología de la Nación), con excepción de los años 2014 y 2015 para los cuales se acudió a la base de datos proporcionada por la UNNOBA.

El Plan de Estudio de la carrera contempla una duración de cursada de 5 años con un total de 46 materias. Fue diseñado para que los estudiantes adquieran gradualmente los contenidos a partir de un sistema de correlatividades.

El desarrollo de la teoría y la práctica es integral. Estos dos aspectos permiten la articulación e integración vertical. La integración entre la teoría y la práctica se logra fundamentalmente con la implementación de la modalidad de clases teórico-prácticas. En el desarrollo del teórico-práctico se alternan metodologías de exposición y desarrollos de casos así como también técnicas de trabajo y discusión grupal, tendientes a fomentar el pensamiento crítico, la capacidad de exposición oral y la participación activa. La articulación horizontal se garantiza de dos maneras; por un lado, por la estructura departamental y por el otro por la conformación de áreas y sub-áreas disciplinares (Res CS 322/10) (Ciencias Básicas, Teoría de la computación, Algoritmos y lenguajes, Arquitectura, Sistemas operativos y redes, Aspectos profesionales y sociales e Ingeniería de Software, Base de Datos y Sistemas de Información).

Con respecto a la inclusión en el Plan de Estudio de contenidos no exigidos y tomando como referencia para su determinación los contenidos curriculares básicos para la carrera de Licenciatura en Sistemas Resolución Ministerial 786/09, desde la Escuela de Tecnología se han definido dos trayectos de materias optativas en años superiores. Uno de estos trayectos incluye asignaturas relacionadas con Minería de Datos e Ingeniería en Requisitos y el otro con Programación Web y Comercio Electrónico. Estos trayectos permitirán al alumno completar su formación con un perfil determinado.

Además de los mencionados trayectos se incorporaron dos asignaturas –Antropología de la Comunicación I y Antropología de la Comunicación II- que completan la formación humanística de los futuros egresados.

La carrera **Ingeniería en Informática** comenzó a ser dictada también en el año 2005. Si bien se dicta únicamente en la Sede de Junín, las asignaturas que son comunes al Título Intermedio de la Carrera Licenciatura en Sistemas (Analista de Sistemas) pueden ser cursadas por los alumnos en la Sede de Pergamino.

Para la obtención del título de Ingeniero/a en Informática se requiere la aprobación de las 48 materias, distribuidas en los cinco años de duración del Plan de estudios vigente. Todos los estudiantes se encuentran en el Plan de estudio aprobado por Resolución CS. N° 252/2009 y 324/2010

La estructura del plan de estudios y su régimen de correlatividades contribuye a la articulación vertical. La articulación horizontal se establece por la estructura departamental de la Universidad y la conformación por áreas. A su vez, si bien existen contenidos que son abordados por más de una asignatura, los temas tratados en las asignaturas no se superponen.

La formación de carácter generalista se encuentra reflejada a través de un ciclo complementario con materias de formación general y asignaturas optativas/electivas, que ofrecen una formación con la posibilidad de adecuación por parte de los graduados a las necesidades del mercado laboral. También fue incluido en la currícula un ciclo complementario que tiene por objetivo establecer la cobertura de aspectos formativos relacionados con las ciencias sociales y humanidades que se considera indispensable para la formación integral del futuro egresado

El cuadro que se presenta a continuación muestra la relación entre la cantidad de ingresantes por año y la cantidad total de alumnos de la carrera durante el período 2005-2015.

Año	Ingresantes	Total de alumnos
2005	61	61
2006	39	82
2007	38	105
2008	38	116
2009	43	141
2010	105 (*)	193
2011	69	201
2012	53	199
2013	57	211
2014	57	209
2015	52	215

(*) De los cuales 76 son inscriptos 2010 y 29 han realizado cambio de Plan.

Tabla 3. Ingresantes y total de alumnos de Ingeniería en Informática entre 2005 y 2015.

Un análisis de la matriculación y el ingreso a las carreras del área de informática desde la perspectiva de género permite corroborar la tendencia que se presenta desde hace años y que denota la baja presencia femenina en este campo de estudios (Fundación Sadosky, 2013).

Como se puede observar en el siguiente cuadro, la distancia entre la cantidad de ingresantes varones y mujeres es sumamente notable y se acrecienta vertiginosamente conforme la complejidad de las mismas. Así, vemos que mientras la carrera Analista de Sistemas registra en la inscripción alrededor de un 30% más de varones sobre mujeres, la cifra de Licenciatura en Sistemas ronda el 60% a favor de éstos y la de Ingeniería en Informática supera el 90% de inscripción masculina.

COHORTE	CARRERA	INGRESANTES	
		VARONES	MUJERES
2005	ANALISTA DE SISTEMAS	28	9
	LICENCIATURA EN SISTEMAS	17	3
	INGENIERÍA EN INFORMÁTICA	22	1
2006	ANALISTA DE SISTEMAS	35	9
	LICENCIATURA EN SISTEMAS	21	6
	INGENIERÍA EN INFORMÁTICA	20	4
2007	ANALISTA DE SISTEMAS	33	11
	LICENCIATURA EN SISTEMAS	26	3
	INGENIERÍA EN INFORMÁTICA	29	1
2008	ANALISTA DE SISTEMAS	28	19
	LICENCIATURA EN SISTEMAS	18	4
	INGENIERÍA EN INFORMÁTICA	23	7
2009	ANALISTA DE SISTEMAS	31	22
	LICENCIATURA EN SISTEMAS	14	10
	INGENIERÍA EN INFORMÁTICA	26	3
2010	ANALISTA DE SISTEMAS	5	1
	LICENCIATURA EN SISTEMAS	118	31
	INGENIERÍA EN INFORMÁTICA	66	10
2011	ANALISTA DE SISTEMAS	-----	-----
	LICENCIATURA EN SISTEMAS	53	11
	INGENIERÍA EN INFORMÁTICA	32	3
2012	ANALISTA DE SISTEMAS	-----	-----
	LICENCIATURA EN SISTEMAS	43	6
	INGENIERÍA EN INFORMÁTICA	27	2
2013	ANALISTA DE SISTEMAS	-----	-----
	LICENCIATURA EN SISTEMAS	44	10
	INGENIERÍA EN INFORMÁTICA	32	3
2014	ANALISTA DE SISTEMAS	-----	-----
	LICENCIATURA EN SISTEMAS	55	8
	INGENIERÍA EN INFORMÁTICA	33	2
2015	ANALISTA DE SISTEMAS	-----	-----
	LICENCIATURA EN SISTEMAS	61	5
	INGENIERÍA EN INFORMÁTICA	33	3
TOTALES		973	207

Tabla 4. Matrícula dividida por género de la Licenciatura en Sistemas y de Ingeniería en Informática, entre 2005 y 2015.

3.3 LA PROBLEMÁTICA DEL REZAGO ACADÉMICO Y LA DESERCIÓN EN LAS CARRERAS DE INFORMÁTICA DE LA UNNOBA. ANTECEDENTES Y ESTRATEGIAS LLEVADAS A CABO.

La deserción, el rezago académico, la baja cantidad de materias rendidas por año y los bajos índices de egreso son problemáticas presentes en las diversas carreras que integran el área de Informática de la Escuela de Tecnología.

Si bien durante el período 2005-2015 el número de estudiantes inscriptos en estas carreras se ha incrementado de manera constante, la tasa de egreso ha sido sumamente baja a lo largo de todo este tiempo lo cual refleja los altos índices de desgranamiento y deserción.

La siguiente tabla permite inferir que sobre un total de 1335 ingresantes, sumando las tres carreras durante el período 2005-2012, el número total de egresados de esas cohortes a 2015 ha sido de 100 alumnos, con lo cual a 2015 el índice de egreso fue de un 9.083%.

AÑO ACADÉMICO	LICENCIATURA EN SISTEMAS				ANALISTA DE SISTEMAS				INGENIERÍA EN INFORMÁTICA			
	POSTULANTES	INGRESANTES	ALUMNOS	EGRESADOS	POSTULANTES	INGRESANTES	ALUMNOS	EGRESADOS	POSTULANTES	INGRESANTES	ALUMNOS	EGRESADOS
2005	43	24	24	0	115	56	56	0	51	33	33	0
2006	46	32	43	0	98	65	95	0	35	29	59	0
2007	40	33	55	0	91	61	119	0	47	34	85	0
2008	32	29	65	0	109	83	156	0	48	33	94	0
2009	27	25	66	0	126	90	178	0	47	35	115	0
2010	192	172	224	3	10	10	57*	0	105	95	184	0
2011	105	80	257	4	0	0	24*	0	59	49	200	6
2012	114	68	282	13	0	0	18*	2	63	37	199	5
2013	101	67	290	17	0	0	9*	0	57	44	211	4
2014	102	66	292	22	0	0	13*	0	57	36	209	4
2015	109	70	309	18	1	1	9*	0	52	40	215	2

Fuente: Elaboración propia de acuerdo a la base de datos de la UNNOBA

*A partir del cambio del Plan de Estudios, desde 2010 Analista de Sistema pasó a ser un título intermedio de la Licenciatura en Sistemas por lo cual sus egresados son contabilizados dentro de la carrera Licenciatura en Sistemas.

El rezago académico y la deserción en términos estadístico.

A continuación se muestra el número de egresados de cada una de las tres carreras según cohorte de origen.

ANALISTA DE SISTEMAS		
COHORTE	INGRESANTES	EGRESADOS
2005	56	3
2006	65	2
2007	61	3
2008	83	5
2009	90	5
TOTAL	355	18

LICENCIATURA EN SISTEMAS.		
COHORTE	INGRESANTES	EGRESADOS
2005	24	3
2006	32	10
2007	33	6
2008	29	6
2009	25	2
2010	172	38
2011	80	10
TOTAL	395	75

INGENIERÍA EN INFORMÁTICA		
COHORTE	INGRESANTES	EGRESADOS
2005	33	8
2006	29	4
2007	34	1
2008	33	1
2009	35	2
2010	95	4
2011	49	2
TOTAL	308	22

Desde 2005 hasta 2009 Analista de Sistemas fue una carrera independiente, a partir de 2010 debido a la implementación del nuevo Plan de Estudios comenzó a ser un título intermedio de la Licenciatura en Sistemas, de 3 años de duración. Por este motivo se toma al año 2009 como el último año puro de la carrera. Durante este período, la tasa de egreso ha sido del 5.07%.

Con respecto a las restantes carreras - Licenciatura en Sistemas e Ingeniería en Informática – se ha recortado el período 2005-2011 teniendo en cuenta que, de acuerdo a la duración teórica de las mismas – cinco años – el año 2011 conforma la última cohorte de posibles egresados a 2015. El análisis de las siguientes tablas permite ponderar las diferencias en el índice de ingreso y egreso de cada una de estas carreras.

En relación al porcentaje de ingreso total de las carreras que conforman el área de Informática, un 37.33% corresponde a la Licenciatura en Sistemas y un 29.11% a Ingeniería en Informática. En lo referente a los índices de egreso el porcentaje de egresados de la Licenciatura en Sistemas duplica al de Ingeniería, siendo de un 18.99% y de un 7.14% respectivamente. No obstante esta diferencia, hay que tener en cuenta que prácticamente el total (97.92%) de quienes se contabilizan como egresados de la Licenciatura corresponden al título intermedio de la misma: Analista de Sistemas. Sólo una minoría (2.8%) ha obtenido el título de Licenciado/a en Sistemas.

Para estudiar y afrontar las problemáticas de la deserción y el rezago estudiantil esta tesis propone la creación y puesta en práctica de un tablero de control que, basado en la utilización de herramientas de minería de datos, permita a nivel de gestión institucional realizar un seguimiento de la trayectoria académica de los estudiantes de las carreras de informática y así identificar predictivamente, a través de un sistema de alertas tempranas, factores de riesgo que puedan dificultar la permanencia del estudiante en la carrera o determinar la deserción.

La UNNOBA cuenta con el sistema de gestión SIU- Guaraní de la Secretaría de Políticas Universitarias (SPU) del Ministerio de Educación; utiliza también el programa Pentaho como herramienta para generar informes, minería de datos, etc. Este sistema registra y administra todas las actividades académicas de la universidad desde el ingreso hasta la titulación de cada estudiante, permitiendo identificar el historial académico de cada uno de ellos. Se han utilizado estas bases de datos con el objetivo de determinar y analizar, a la luz de la literatura existente sobre el tema, las problemáticas de rezago académico y deserción en las carreras que integran el área de Informática. Esto ha permitido crear una serie de indicadores educativos.

Se puede definir un **indicador educativo** como “una proposición que identifica un rasgo o característica empíricamente observable, que permite la medida estadística de un concepto o de una dimensión de éste, basado en un análisis teórico previo, e integrado en un sistema coherente de proposiciones vinculadas, cuyo análisis pueda orientarse a describir, comparar, explicar o prever hechos” (Estévez, 2007: 46).

Cada indicador educativo es producto de la información procesada. En la presente investigación fueron desarrollados tomando como insumo las bases de datos y repositorios

institucionales de la UNNOBA así como las tabulaciones de encuestas y análisis de entrevistas semi-estructuradas a estudiantes que han desertado de las carreras mencionadas anteriormente.

La ponderación de los resultados de cada indicador y, en un segundo momento, el entrecruzamiento de los mismos, permitirá identificar factores de riesgo que puedan dificultar la permanencia en dichas carreras.

Los indicadores desarrollados surgieron de un análisis retrospectivo de la deserción en las carreras de informática, específicamente de la Licenciatura en Sistemas y de Ingeniería en Informática, durante el período 2010-2015. Si bien estas carreras comenzaron a implementarse en el año 2005, cuando surgió la UNNOBA, se ha tomado el intervalo 2005-2009 sólo como antecedente dado que las carreras se hallaban en una etapa inicial de normalización académica; otro motivo radica en que el Plan de Estudios vigente data recién de 2010.

En primer lugar se llevó a cabo un relevamiento de información de estudiantes inscriptos, matriculados y graduados cuatrimestre por cuatrimestre, de acuerdo a cada cohorte de las tres carreras que conforman el área de Informática. Esto ha permitido realizar un análisis longitudinal y transversal de la totalidad de los años de cursada pudiendo establecer así el índice de deserción por cada carrera, determinando el momento más crítico de las mismas. Para efectuar dicho análisis se han tenido en cuenta los siguientes indicadores:

- **Índices de deserción por cohorte:** es el resultado de las diferencias, para cada cohorte, entre el número total de estudiantes que ingresan a la carrera y el número de ellos que culmina la cursada del plan de estudio. En muchos casos esta diferencia puede dar cuenta también de una situación de **rezago académico** en tanto un número determinado de alumnos de una cohorte puede no haber egresado en el tiempo teórico correspondiente y haberse reinscripto en años posteriores.
- **Índices de deserción por año académico de cada carrera:** es el número total de desertores de cada año/nivel de una carrera, sobre el total de matriculados en el mismo año/nivel.

Además de los mencionados indicadores de deserción, dada la relevancia de la problemática del desgranamiento y rezago estudiantil, se tendrán en cuenta los siguientes indicadores de duración y retención académica.

- **Índice de duración: duración media real / duración teórica:** indica la relación entre la cantidad de años que les llevó a los egresados concluir la carrera respecto a la cantidad de años establecidos por el plan de estudios. Se considera N (en años) al tiempo mínimo establecido para poder egresar. A partir de ahí se van sumando los años que pasan desde ese momento. Se considera que el alumno se graduó en el tiempo N hasta la 2º convocatoria del año siguiente, y así sucesivamente (N+2), (N+3).
- **Índice de retención de 1º y 2º año:** indica el porcentaje de estudiantes que estando matriculados en 1º y 2º año en cada una de las carreras analizadas, en un determinado año del período 2005-2015, continuaron matriculados en la misma carrera y en la misma cohorte de origen al año siguiente.

Índice de deserción y rezago por cohorte y por año académico en las carreras de Informática de la UNNOBA

La siguiente tabla permite analizar los índices de deserción y egreso por cohorte de cada una de las carreras que conforman el área de Informática de la UNNOBA durante el período 2005-2011. Toma como fecha última ese año en tanto muestra la proporción de alumnos activos que se encontraban en condiciones de egresar (según la duración teórica de la carrera) para fines del año 2015 pero para esa fecha no lo han hecho.

ÍNDICE DE DESERCIÓN Y REZAGO POR COHORTE									
COHORTE	ANALISTA DE SISTEMAS			LICENCIATURA EN SISTEMAS			INGENIERÍA EN INFORMÁTICA		
	INGRESANTES	EGRESADOS	ÍNDICE	INGRESANTES	EGRESADOS	ÍNDICE	INGRESANTES	EGRESADOS	ÍNDICE
2005	56	3	94,64%	24	3	87,50%	33	8	75,76%
2006	65	2	96,90%	32	10	68,75%	29	4	86,20%

2007	61	3	95,08%	33	6	81,82%	34	1	97,05%
2008	83	5	93,96%	29	6	79,31%	33	1	96,97%
2009	90	5	94,45%	25	2	92%	35	2	94,28%
2010				172	38	81,39%	95	4	95,79%
2011				80	10	87,50%	49	2	95,92%
TOTAL	71	18	95,00%	395	75	82,61%	308	22	84,35%

La tabla muestra en términos globales, para todo el período y las tres carreras, un porcentaje total de abandono y rezago de 87.32%, en conjunto.

Ahora bien, como se especificó anteriormente, se deben distinguir los casos de deserción estudiantil de aquellos que configuran rezago académico, es decir, una prolongación o discontinuidad en la trayectoria académica que se observa cuando la misma se extiende más allá de lo previsto en el plan de estudios correspondiente.

Como lo demuestran las siguientes tablas, del total de ingresantes durante el período 2005-2011 un 46.32% abandonó la carrera, en el caso de la Licenciatura en Sistemas, y un 58.39% en el caso de Ingeniería en Informática; mientras que un 34.70% y un 34.23% respectivamente siguió manteniendo actividad académica, con reiteradas reinscripciones más allá de 2015, aunque de forma prolongada y discontinua.

LICENCIATURA EN SISTEMAS				
COHORTE	INGRESANTES	REZAGO	ABANDONO	EGRESO
2005	24	5	16	3
2006	32	6	16	10
2007	33	9	18	6
2008	29	7	16	6
2009	25	12	11	2
2010	172	72	62	38
2011	80	26	44	10
TOTALES	395	137	183	75
	100%	34,70%	46,32%	18,98%

INGENIERÍA EN INFORMÁTICA				
COHORTE	INGRESANTES	REZAGO	ABANDONO	EGRESO
2005	23	8	7	8
2006	29	6	19	4
2007	34	12	21	1
2008	33	15	17	1
2009	35	9	24	2
2010	95	33	58	4
2011	49	19	28	2
TOTALES	298	102	174	22
	100%	34,23%	58,39%	7,38%

Índice de duración de las carreras de Informática de la UNNOBA

La siguiente tabla permite determinar el **índice de duración de las carreras**, es decir, la relación entre la duración media real y la duración teórica de las carreras establecidas por el plan de estudios.

Carrera	Título	Cohorte	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015		
041 - LICENCIATURA EN SISTEMAS	Analista en Sistemas (título intermedio desde 2010)	2005						1	1		1				
		2006							1				1		
		2007									2			1	
		2008						1		2	1	1			
		2009							1	1			2	1	
		2010									4	9	13	11	
		2011										5	4	1	
		2012											1	2	2
		2013													
	Licenciado en Sistemas	2005													
		2006								3	3	1		1	
		2007							1			1		1	
		2008									1				
		2009													
		2010												1	
		2011													
		2012													
2013															
045 - INGENIERÍA EN INFORMÁTICA	Ingeniero en Informática	2005							2	3	1	1	1		
		2006							2				1	1	
		2007											1		
		2008										1			
		2009											1	1	
		2010										2	1	1	
		2011													
		2012													
		2013													

Aquí se evidencia que en su gran mayoría, quienes han egresado lo han hecho luego de varios años de actividad académica, excediendo los tiempos planteados por los planes de estudio. En lo que respecta a la Licenciatura en Sistemas, se percibe que el mayor número de egresados (84.33%) corresponde al título intermedio de Analista de Sistemas; sólo una minoría (15.66%) se ha titulado como Licenciado en Sistemas. En ambos casos se registra una actividad académica de entre 7 y 8 años lo cual refleja el promedio de duración de estas carreras. Algo parecido sucede con Ingeniería en Informática. En esta carrera el número de egresados es menor pero la duración real de la carrera oscila entre los 6 y 7 años.

Índices de retención y de desgranamiento de 1° y 2° año de las carreras de Informática de la UNNOBA

Las tablas que se presentan a continuación permiten analizar el **índice de retención de 1° y 2° año** de cada una de las tres carreras. En interrelación con la retención, en oposición a ella, se presenta el porcentaje de **desgranamiento** tanto de 1° a 2° año como de 2° a 3° año de cada carrera.

LICENCIATURA EN SISTEMAS							
COHORTE	CANTIDAD DE ESTUDINTES			RETENCIÓN		DESGRANAMIENTO	
	1° AÑO	2° AÑO	3° AÑO	1° A 2° AÑO	2° A 3° AÑO	1° A 2° AÑO	2° A 3° AÑO
2005	24	11	8	45,83%	72,73%	54,17%	27,27%
2006	32	14	10	43,75%	71,43%	56,25%	28,57%
2007	33	21	16	63,64%	76,19%	36,36%	23,81%
2008	29	12	11	41,38%	91,66%	58,62%	8,33%
2009	25	18	17	72%	94,44%	28%	5,55%
2010	172	127	114	73,84%	89,76%	26,16%	10,24%
2011	80	54	40	67,50%	74,07%	32,50%	25,93%
2012	68	43	36	63,24%	83,72%	36,76%	16,28%
2013	67	39	35	58,21%	89,74%	41,79%	10,26%
2014	66	47	37	71,21%	78,72%	28,79%	21,28%
2015	70	48	32	68,57%	66,67%	31,43%	33,33%
TOTALES	666	434	356	60,83%	80,83%	39,17%	19,17%

ANALISTA DE SISTEMAS							
COHORTE	CANTIDAD DE ESTUDINTES			RETENCIÓN		DESGRANAMIENTO	
	1° AÑO	2° AÑO	3° AÑO	1° A 2° AÑO	2° A 3° AÑO	1° A 2° AÑO	2° A 3° AÑO
2005	56	30	23	53,57%	76,67%	46,43%	23,33%
2006	65	35	23	53,85%	65,71%	46,15%	34,29%
2007	61	28	20	45,90%	71,43%	54,10%	28,57%
2008	83	32	7	38,55%	21,87%	61,45%	78,13%
2009	90	21	10	23,33%	47,62%	76,67%	52,38%
TOTALES	335	146	83	43,04%	56,66%	56,96%	43,34%

INGENIERÍA EN INFORMÁTICA							
COHORTE	CANTIDAD DE ESTUDINTES			RETENCIÓN		DESGRANAMIENTO	
	1° AÑO	2° AÑO	3° AÑO	1° A 2° AÑO	2° A 3° AÑO	1° A 2° AÑO	2° A 3° AÑO
2005	33	30	29	90,91%	96,67%	9,09%	3,33%
2006	29	22	13	75,86%	59,09%	24,14%	40,91%
2007	34	28	22	82,35%	78,57%	17,65%	21,43%
2008	33	27	20	81,82%	74,07%	18,18%	25,93%
2009	35	20	17	57,14%	85%	42,86%	15%
2010	95	65	52	68,42%	80%	31,58%	20%
2011	49	32	29	65,31%	90,63%	34,69%	9,37%

2012	37	27	25	72,97%	92,59%	27,03%	7,41%
2013	44	38	26	86,36%	68,42%	13,64%	31,58%
2014	36	28	23	77,78%	82,14%	22,22%	17,86%
2015	40	27	20	67,50%	74,07%	32,50%	25,93%
TOTALES	465	344	276	75,13%	80,11%	24,87%	19,89%

Si se comparan de manera global los índices de retención y desgranamiento de las tres carreras se evidencia que en todas ellas los mayores índices de retención se presentan entre el 2° y 3° año y, por el contrario, los mayores porcentajes de desgranamiento entre el 1° y 2° año. De las tres carreras es Ingeniería en Informática la que presenta el mayor porcentaje de retención con respecto al año anterior y, por lo tanto, menor índice de desgranamiento; esto se evidencia sobre todo durante las primeras cuatro cohortes en las cuales llega hasta duplicar los índices de retención de las otras carreras para el mismo período y cohorte de origen. Al mismo tiempo, es la carrera que muestra mayores regularidades en todos los índices. Ahora bien, el análisis comparativo de las diferencias entre los índices de cada carrera muestra que en el caso de la Licenciatura en Sistemas el porcentaje de retención de 2° a 3° año excede en 20 puntos al porcentaje de retención 1° a 2° año; en el caso de Analista de Sistemas y de Ingeniería en Informática la diferencia es de 13.62 puntos y 4.98 puntos respectivamente.

Los siguientes gráficos reafirman lo evidenciado en las tablas anteriores: muestran que en las carreras que componen el área de Informática los mayores índices de desgranamiento y deserción se han presentado y se presentan durante los dos primeros años de estudio.

CARRERA	COHORTE	AÑO ACADÉMICO	ALUMNOS	PORCENTAJE	% DE DESGRANAMIENTO	PROMEDIO DE DESGRANAMIENTO
ANALISTA DE SISTEMAS	2005	2005	56	100%	46.43%	62.47%
		2006	30	53.57%		
	2006	2006	65	100%	46.15%	
		2007	35	53.85%		
	2007	2007	61	100%	54.10%	
		2008	28	45.90%		
	2008	2008	83	100%	61.45%	
		2009	32	38.55%		
	2009	2009	90	100%	76.67%	
		2010	21	23.33%		
	2010	2010	10	100%	90%	
		2011	1	10%		

CARRERA	COHORTE	AÑO ACADÉMICO	ALUMNOS	PORCENTAJE	% DE DESGRANAMIENTO	PROMEDIO DE DESGRANAMIENTO
LICENCIATURA EN SISTEMAS	2005	2005	24	100%	54.17%	38.38%
		2006	11	45.83%		
	2006	2006	77	100%	56.25%	
		2007	52	43.75%		
	2007	2007	38	100%	31.58%	
		2008	26	68.42%		
	2008	2008	31	100%	54.84%	
		2009	14	45.16%		
	2009	2009	25	100%	28%	
		2010	18	72%		
	2010	2010	172	100%	26.16%	
		2011	127	73.84%		
	2011	2011	80	100%	32.50%	
		2012	54	67.50%		
	2012	2012	68	100%	36.76%	
		2013	43	63.24%		
	2013	2013	67	100%	41.79%	
		2014	39	58.21%		
	2014	2014	66	100%	28.79%	
		2015	47	71.21%		
	2015	2015	70	100%	31.43%	
		2016	48	68.57%		

CARRERA	COHORTE	AÑO ACADÉMICO	ALUMNOS	PORCENTAJE	% DE DESGRANAMIENTO	PROMEDIO DE DESGRANAMIENTO
INGENIERÍA EN INFORMÁTICA	2005	2005	54	100%	20.38%	27.20%
		2006	43	79.62%		
	2006	2006	37	100%	21.63%	
		2007	29	78.37%		
	2007	2007	38	100%	21.06%	
		2008	30	78.94%		
	2008	2008	38	100%	21.06%	
		2009	30	78.94%		
	2009	2009	43	100%	53.49%	
		2010	20	46.51%		
	2010	2010	95	100%	31.58%	
		2011	65	68.42%		
	2011	2011	49	100%	34.69%	
		2012	32	65.31%		
	2012	2012	37	100%	27.03%	
		2013	27	72.97%		
	2013	2013	44	100%	13.64%	
		2014	38	86.36%		
	2014	2014	36	100%	22.22%	
		2015	28	77.78%		
2015	2015	40	100%	32.50%		
	2016	27	67.50%			

Estrategias implementadas para promover la permanencia y minimizar la deserción

Para enfrentar y contribuir a minimizar la deserción, a lo largo de los últimos años, se han propuesto una serie de programas tendientes a detectar problemáticas que influyan directa o indirectamente en el desempeño del estudiante, y que contribuyan a mejorar las condiciones de enseñanza y aprendizaje a lo largo de su trayectoria académica.

En articulación con el nivel secundario se han implementado acciones en forma conjunta que permiten optimizar el ingreso de los estudiantes a la universidad y atender las necesidades de formación que requieren para transitar con éxito el primer año de los estudios superiores. Dentro de las acciones realizadas caben mencionar: Voluntariados Universitarios, Capacitación y actualización de docentes secundarios y universitarios, Programa de Apoyo al último año del Nivel Secundario para la Articulación con el Nivel Superior, Convenios con Escuelas Técnicas y Agrarias.

Con el objetivo de disminuir la deserción inicial la UNNOBA dicta un Curso de Ingreso obligatorio, irrestricto y presencial o semi-presencial, donde los alumnos cursan

de lunes a viernes, durante el mes de febrero, dos asignaturas que dependen de la carrera elegida.

También ha desarrollado un Programa de Tutorías (Ordenanza CS 15/2009) que tiene como objetivo principal implementar un sistema integral de asistencia a los estudiantes del primer año de las distintas carreras que permita mejorar los procesos de enseñanza y aprendizaje.

La baja cantidad de materias que los estudiantes rinden por año es otra de las situaciones que se presentan en todas las carreras. Frente a ello, desde la Escuela de Tecnología, se ha implementado un programa de Apoyo a finales, con el objetivo de dar soporte a los alumnos en este tema. La primera etapa del proyecto abarca asignaturas de los dos primeros años de las materias de las Carreras de la Escuela.

La Escuela de Tecnología con Resolución 10/09 crea el Programa de Seguimiento de Alumnos el cual se complementa con el que se viene implementado desde la Secretaría Académica de la Universidad. Tiene como objetivos: mejorar el acceso y la permanencia de los estudiantes garantizando la calidad de la enseñanza, diagnosticar y analizar las características de los estudiantes de la Escuela de Tecnología de la UNNOBA, facilitar el tránsito de los estudiantes del nivel medio al nivel universitario, realizar el seguimiento de los Estudiantes de la Escuela de Tecnología para favorecer la permanencia. Es así como en el marco de este programa se detectan los alumnos que realizado el curso de ingreso no registran actividad académica en el primer cuatrimestre y se deriva esta lista a los tutores para realizar un análisis de los motivos y verificar si los mismos se pueden incorporar nuevamente a la Universidad.

3.4 ANÁLISIS DE FACTORES CAUSALES DE LA DESERCIÓN EN LAS CARRERAS DE INFORMÁTICA DE LA UNNOBA.

Como se planteó en el apartado anterior, el desgranamiento y la deserción son dos problemáticas presentes en las Carreras de Informática de la Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA).

Con el objetivo de poder identificar las causas o variables que inciden y determinan mayoritariamente la deserción en las carreras de Informática se llevó a cabo un relevamiento y, luego, entrecruzamiento de diversas fuentes, algunas disponibles en el sistema SIU-Guaraní de la propia Universidad y otras creadas a partir de la instrumentación de encuestas y entrevistas a alumnos que han desertado de cualquiera de las tres carreras. Cabe aclarar que han sido considerados desertores aquellos alumnos que no cumplieron con los requisitos de regularidad y no están en condiciones reglamentarias de recuperarlas, habiendo abandonado la carrera o cambiado a otra carrera por fuera de las que integran el área de Informática de la UNNOBA.

A partir de estos registros de información, se diseñó un estudio retrospectivo y longitudinal delimitando un arco temporal de diez años 2005-2015, tomando como punto de inicio la creación de la UNNOBA. Dadas las modificaciones en los planes de estudio puestas en práctica a partir de 2010, se ha establecido ese año como bisagra entre dos subperíodos: el primero, entre los años 2005 y 2009, definido y caracterizado por el inicio y normalización de las carreras; el segundo, entre 2010 y 2015, determinado por el comienzo de los planes vigentes en la actualidad. Debido a ello, este estudio se concentró mayoritariamente en el análisis de esta etapa, sobre todo en lo que concierne a la variable de rendimiento académico.

En base a los datos del sistema SIU-Guaraní, se elaboró un registro con un total de 217 alumnos, provenientes de diversas localidades de la región, considerados desertores durante el período en estudio. Se intentó contactarlos ya sea por vía telefónica, mail o redes sociales solicitándoles completaran una encuesta y luego una entrevista semi-estructurada en línea, diseñada con la herramienta “Google Formularios” de Google Drive. Las respuestas obtenidas conformaron una muestra poblacional de un total de 63 alumnos de

los cuales 51 respondieron sólo la encuesta y 12 solamente la entrevista. Solamente 28 de ellos realizaron tanto la encuesta como la entrevista.

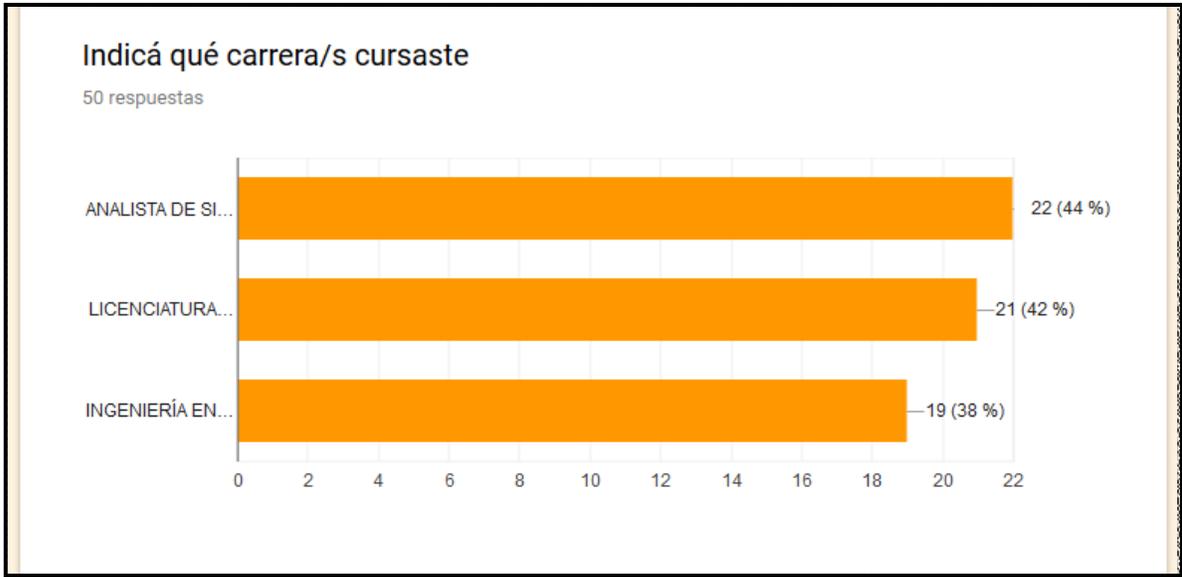
En tanto no todos los estudiantes registrados como desertores fueron localizados o respondieron a la encuesta y entrevista, el diseño de muestra utilizado fue no probabilístico ya que no permitió determinar el error de muestreo y su confiabilidad.

La encuesta permitió, a partir de un cuestionario semi-dirigido y autoadministrado, recoger información estandarizada de un número determinado de individuos para entender a la población o universo al que representan (McMurtry, 2005). Tuvo un enfoque y una finalidad descriptiva ya que se la consideró uno de los medios necesarios para poder construir un perfil de los alumnos que han abandonado las carreras de informática de la UNNOBA.

Por su parte, la entrevista fue semi-estructurada, de carácter centrado o enfocado. En términos metodológicos, este tipo de entrevista combina parte de las dimensiones asociadas con la profundidad y la libertad, que caracterizan a las entrevistas no estructuradas, con el carácter más rígido y preestablecido de las preguntas de las entrevistas estructuradas (Vela Peón, 2001)

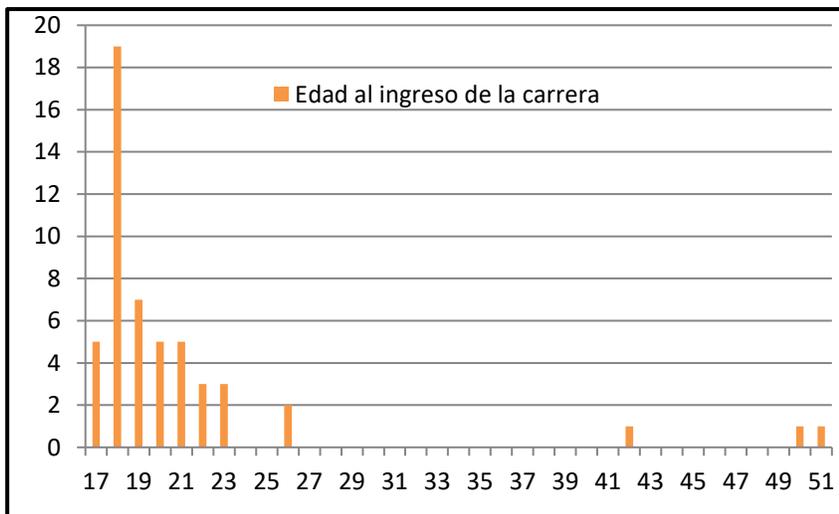
Se acudió a este tipo de entrevista con el objetivo de que sus resultados constituyeran un complemento a la información extraída de las encuestas, proveyendo información cualitativa que permita analizar, con mayor densidad, las causas que actuaron como determinantes de la deserción.

El siguiente gráfico muestra el porcentaje de encuestados que corresponde a cada carrera.



Entre los encuestados más del 80% corresponde al género masculino y un porcentaje menor al 20% al femenino, lo cual condice con el carácter marcadamente masculino de las inscripciones a las carreras del área de informática.

Como lo demuestra el siguiente gráfico, la mayoría de los estudiantes encuestados y entrevistados (48%) ingresó a la carrera tras concluir los estudios secundarios, entre los 18 y los 19 años de edad; sólo un 15% sobrepasaba los 40 años. El restante 37% está conformado por aquellos que comenzaron la carrera entre los 20 y los 39 años.



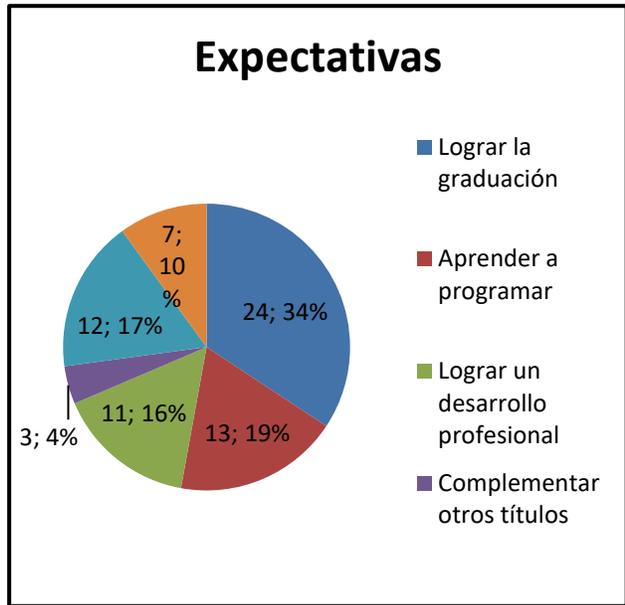
En lo concerniente a la trayectoria académica, el 77.2% cursó sus estudios secundarios en escuela pública mientras que el 22.8% restante lo hizo en establecimientos privados.

Un alto porcentaje (54.67%) provenía de localidades aledañas a la ciudad de Junín; de ellos un 30% viajaba diariamente para cursar, el resto se trasladó a vivir a esta ciudad.

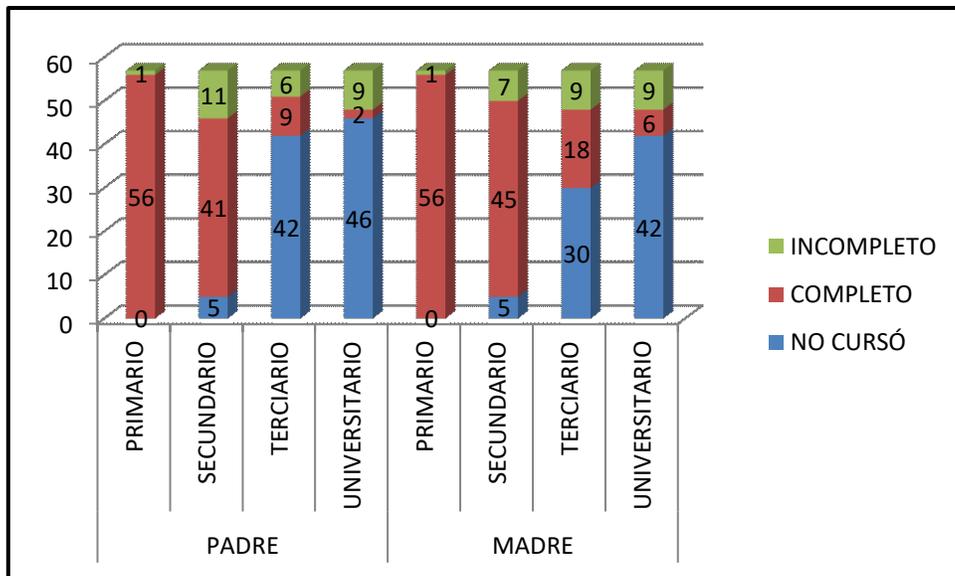
El 91% de quienes fueron encuestados y entrevistados dijeron ser solteros y sin familiares a cargo durante el período en el que cursaron la carrera. Sólo el 9% estaba casado o en pareja, con hijos a cargo.

En lo que respecta a la situación laboral, el 44% no trabajaba durante la cursada de la carrera mientras que el 56% restante sí lo hacía, el 31.5% media jornada y el 24.5% jornada completa.

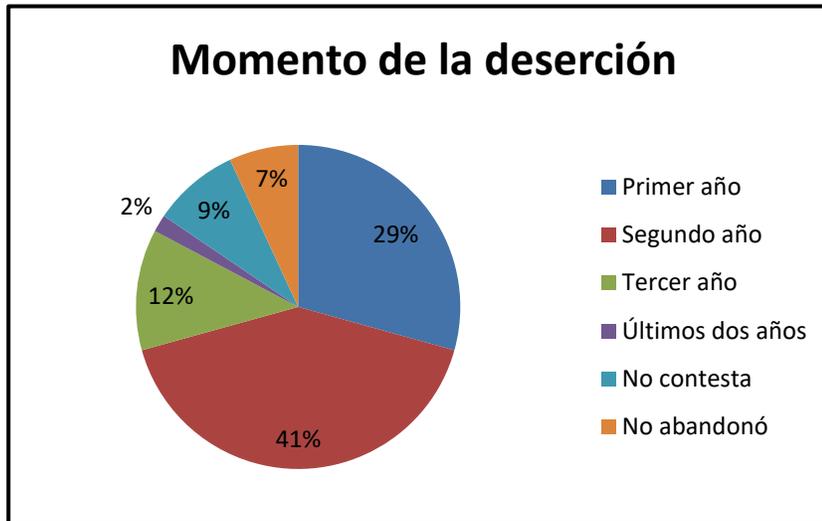
En cuanto a los motivos y expectativas que determinaron la elección de la carrera, las respuestas fueron múltiples y diversas. Una categorización de las mismas permitió llegar a las siguientes conclusiones: en primer lugar (43%) se evocó como motivación, a modo general, el interés por la informática sin hacer alusión específicamente a la programación o análisis de sistemas. El interés por aprender a programar aparece, aunque en segundo lugar y con bajo porcentaje (13.19%), como una de las expectativas de la carrera. Un segundo motivo estuvo determinado por la cercanía de la Universidad al lugar de residencia (16%) lo cual se condice con el alto porcentaje de alumnos (54.67%) provenientes de localidades cercanas al Partido de Junín.



En lo que refiere a las expectativas hacia la carrera elegida, en primer lugar - y superando en más de un 10% a la siguiente - aparece el objetivo de “lograr la graduación” (24,34%). Esta cuestión que, podría parecer llamativa o banal, cobra importancia y complejidad si se tiene en cuenta el nivel máximo de educación obtenido por los padres. Como se puede apreciar en el siguiente gráfico, la mayoría de ellos no obtuvo un título superior, ya sea universitario o terciario, por lo cual sus hijos serían la primera generación, luego de ellos, en obtener una titulación universitaria.



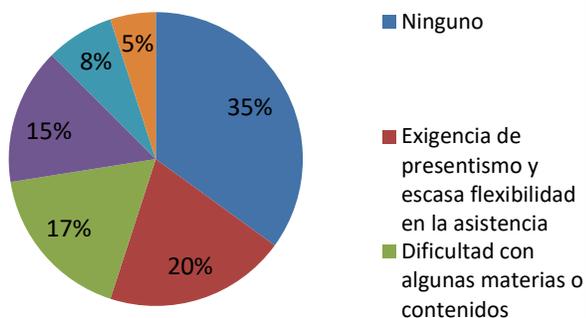
En lo concerniente a la problemática de la deserción, los resultados de la muestra poblacional se condicen y reflejan lo ya expuesto en el apartado anterior, mostrando que los mayores índices de abandono de la carrera se presentan durante los dos primeros años de estudio.



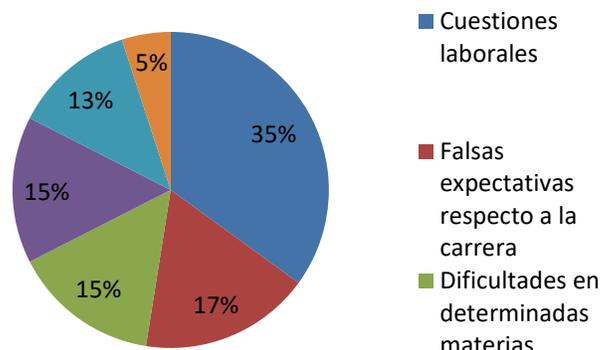
Al

analizar los motivos que, según los entrevistados, llevaron al abandono de la carrera la cuestión laboral aparece como un factor determinante. Esto cobra importancia si se tiene en cuenta que el 56.14% de los estudiantes entrevistados trabajaba mientras cursaba la carrera y que además, más del 50% de quienes trabajaban lo hacían en jornada completa. Este factor de deserción se puede entrecruzar con la exigencia de presentismo y la escasa flexibilidad del régimen de asistencia, considerado por los alumnos como un obstáculo en la carrera. En conjunción, ambas variables explican la imposibilidad de mantener presentismo y con ello la regularidad en muchas materias.

Obstáculos en la carrera



Motivo de la deserción



Otro factor importante que se presenta como determinante de la deserción está dado por las dificultades en determinadas materias (15%). Esto se relaciona, como veremos más adelante, con las problemáticas que surgen a lo largo del primer año en cuanto al régimen de correlatividad de las materias de primer y segundo cuatrimestre y al bajo porcentaje de aprobación de las mismas.

Como se sostuvo al inicio de este capítulo, con el objetivo de poder identificar las causas que determinan mayoritariamente la deserción en las carreras de Informática, además de los resultados recientemente expuestos de las encuestas y entrevistas, se han tenido en cuenta las bases de datos proporcionadas por el Pentaho. A partir de allí, y considerando los principales modelos explicativos sobre el fenómeno de la deserción estudiantil, se han entrecruzado y analizado estas fuentes y se ha creado un sistema de

indicadores. La ponderación de los resultados de cada indicador y, en un segundo momento, el entrecruzamiento de los mismos, permitirá crear el perfil de alumno de las carreras de informática e identificar factores de riesgo que puedan dificultar la permanencia en dichas carreras.

INDICADORES	
PERSONAL	<ul style="list-style-type: none"> - Sexo - Edad al ingreso de la carrera. - Lugar de procedencia. - Lugar de residencia durante el período lectivo (En caso de no residir en sede universitaria, demanda horaria de viaje)
FAMILIAR	<ul style="list-style-type: none"> - Situación familiar al momento del ingreso: <ul style="list-style-type: none"> a) Estado civil. b) Cantidad de familiares a cargo. c) Cantidad de convivientes. - Nivel máximo de educación obtenido por los padres.
ECONÓMICO-SOCIAL	<ul style="list-style-type: none"> - Situación laboral: <ul style="list-style-type: none"> a) Trabajaba al momento de ingresar a la carrera (Horas semanales trabajadas) b) Comenzó a trabajar durante la cursada de la carrera No trabajó durante la cursada. c) No trabajó durante la cursada de la carrera - Fuente de financiamiento de los estudios: <ul style="list-style-type: none"> a) Aportes familiares b) Trabajo propio. c) Becas
ACADÉMICO	<ul style="list-style-type: none"> - Trayectoria académica previa: <ul style="list-style-type: none"> a) Tipo de secundario (Escuela Pública o privada). b) Título secundario obtenido. c) Orientación vocacional d) Otros títulos superiores previos. - Motivo por el cual eligió esta carrera universitaria. - Expectativas hacia la carrera. - Rendimiento académico en la carrera universitaria: <ul style="list-style-type: none"> a) Cantidad de materias con cursadas aprobadas. b) Cantidad de finales rendidos. c) Cantidad de finales aprobados d) Abandono de cursada por materia.

A continuación se analizarán uno de los indicadores claves para poder dar cuenta de la deserción: **el rendimiento académico en la carrera universitaria**.

Uno de los aspectos más importantes del análisis del rendimiento académico está dado por la evaluación y el seguimiento del desempeño del alumnado en las asignaturas, tanto en las cursadas como en la acreditación final de las mismas, sobre todo en aquellas que presentan mayores índices de desaprobación (Porto y Gresia, 2004).

En este sentido, los indicadores más reconocidos y que han sido aplicados en este trabajo son los siguientes:

- Ratio presentados/total: relación porcentual entre el número total de alumnos presentados y el total de matriculados en una asignatura.
- Ratio aprobados/total: relación porcentual entre el número total de alumnos que han superado una asignatura y el total de matriculados en la misma, por año académico.
- Ratio aprobados/presentados: relación porcentual entre el número total de alumnos que han aprobado la asignatura y el número total de alumnos presentados, en las instancias de examen final.
- Ratio desaprobados/ presentados: relación porcentual entre el número total de alumnos que han desaprobado la asignatura y el número total de alumnos presentados, en las instancias de examen final.
- Ratio inscriptos/ausentes: relación porcentual entre el número total de alumnos que, habiéndose inscripto, no se presentaron a rendir en la instancia de examen final.

Para ello, dado que los principales índices de abandono se presentan en los primeros dos años de la carrera, se analiza pormenorizadamente el rendimiento académico en ese trayecto de la carrera.

Los siguientes cuadros muestran el resultado de cursada de tres materias - ARQUITECTURA I (357); INTRODUCCIÓN AL ÁLGEBRA (666) y ANÁLISIS MATEMÁTICO BÁSICO (667) - correspondientes al primer cuatrimestre del primer año de

las carreras que, por el régimen de correlatividades, requieren ser aprobadas para cursar cuatro de las cinco materias del segundo cuatrimestre.

Del análisis de estos cuadros se desprende que sólo un 39.75% de los alumnos de la Licenciatura en Sistemas y un 41.19% de los alumnos de Ingeniería en Informática

RESULTADO DE CURSADAS						
CARRERA	COHORTE	TOTAL ALUMNOS	N° ALUMNOS INSCRIPTOS	N° ALUMNOS APROBADOS	N° ALUMNOS DESAPROBADOS	N° ALUMNOS AUSENTES
LICENCIATURA EN SISTEMAS	2010	224	192	85	52	55
	2011	213	169	86	39	44
	2012	209	135	51	43	41
	2013	219	136	50	50	36
	2014	214	109	39	40	30
	2015	215	133	44	46	53

han aprobado, en el período 2010-2015, las tres materias del primer cuatrimestre

RESULTADOS DE CURSADAS						
CARRERA	COHORTE	TOTAL ALUMNOS	N° ALUMNOS INSCRIPTOS	N° ALUMNOS APROBADOS	N° ALUMNOS DESAPROBADOS	N° ALUMNOS AUSENTES
INGENIERÍA EN INFORMÁTICA	2010	184	161	73	40	48
	2011	201	125	64	37	24
	2012	199	142	63	30	49
	2013	211	148	55	59	34
	2014	209	106	34	38	34
	2015	215	127	47	39	41

necesarias para poder cursar cuatro de las cinco materias que conforman el segundo cuatrimestre. En cuanto a los índices de desaprobación y de ausentismo, como se puede apreciar, la diferencia entre ambos no es sustancial. No obstante ello, en el caso de la Licenciatura en Sistemas se observa una relación de paridad mientras que en Ingeniería en Informática el número de desaprobados es un 30% mayor al número de alumnos ausentes.

Otro dato importante es el hecho de que de la matrícula del total de alumnos que conformaba el primer año de cada una estas cohortes, en promedio, sólo el 67% se inscribieron en las tres materias mencionadas.

A este análisis se le puede sumar que sólo el 49.32% de los alumnos que se inscribieron en las instancias de exámenes finales de esas materias lograron la aprobación. Como muestra la siguiente tabla, si estos datos son desagregados por carrera, a lo largo del período 2005-2015, se puede observar que en ambas los porcentajes más bajos de aprobación corresponden a las materias afines al análisis matemático.

MATERIAS	% APROBACIÓN	
	LICENCIATURA EN SISTEMAS	INGENIERÍA EN INFORMÁTICA
0356- MATEMÁTICA DISCRETA	48%	52%
0357 – ARQUITECTURA I	71%	60%
009 – ÁLGEBRA Y GEOMETRÍA		
00528 – ÁLGEBRA	58%	54%
00666 – INTRODUCCIÓN AL ÁLGEBRA		
00667 – ANÁLISIS MATEMÁTICO BÁSICO	37%	40.5%

CAPÍTULO 4: DEFINICIÓN, DESARROLLO Y APLICACIÓN DE UN MODELO DE INTERVENCIÓN VIRTUAL/ TABLERO DE CONTROL PARA PREDECIR Y MINIMIZAR EL REZAGO ACADÉMICO Y LA DESERCIÓN EN LAS CARRERAS DE INFORMÁTICA DE LA UNNOBA.

La determinación y el análisis de los principales factores causales de rezago académico y abandono de las carreras de informática de la UNNOBA durante el período 2005-2015 – analizados en el capítulo anterior – ha sido el insumo a partir del cual se ha diseñado y construido, aplicando herramientas de *data mining*, un **tablero de control** que actúe como sistema de “alertas tempranas” – SAT – (Carvajal Olaya et. al, 2013), pronosticando aquellos casos de estudiantes con un alto grado de rezago y riesgo de deserción.

Este tablero, por lo tanto, desarrolla y visualiza el entrecruzamiento entre los indicadores y será utilizado a nivel de gestión institucional con el fin de monitorear la trayectoria académica de los estudiantes de las carreras de informática para poder identificar predictivamente, a través de las alertas tempranas, factores de riesgo que puedan dificultar la permanencia del estudiante en la carrera o determinar la deserción. Por ello, una funcionalidad a incorporar es la de la predicción del rendimiento académico de los estudiantes fundamentada en un seguimiento de sus trayectorias académicas.

En última instancia, el tablero permite identificar, a través de estas alertas críticas, a los estudiantes que necesiten atención prioritaria para proceder a su pronta atención.

A continuación se desarrolla un primer apartado abocado a teorizar el análisis predictivo, luego, en un segundo apartado se muestra el desarrollo del tablero de control propiamente dicho.

4.1 ANÁLISIS PREDICTIVO.

El análisis predictivo consiste en la tecnología que aprende de la experiencia para predecir el futuro comportamiento de individuos para tomar mejores decisiones. (Eric Siegel, 2013)

El análisis predictivo es la rama de la minería de datos que trata con la predicción de las probabilidades y las tendencias futuras. El elemento central del análisis predictivo es el predictor, una variable que puede ser medida por un individuo u otra entidad para predecir el comportamiento futuro (Mishra et al., 2010).

El análisis predictivo consiste en la extracción de información existente en los datos y en su utilización para predecir tendencias futuras y patrones de comportamiento pudiendo aplicarse sobre cualquier evento desconocido ya sea en el pasado, presente o futuro. El análisis predictivo se fundamenta en la identificación de relaciones entre variables en eventos pasados, para luego explotar dichas relaciones y predecir posibles resultados en futuras situaciones. (Espino, 2017)

Realizar una predicción no es lo mismo que realizar un pronóstico. Mientras que con un pronóstico se podría predecir la cantidad de helados que se podrían vender el próximo mes, con el análisis predictivo se podría predecir cuales individuos es más probable que se coman un helado.

Machine Learning y Minería de datos.

El Machine Learning o Aprendizaje Automático es una rama de la Inteligencia Artificial que se encarga del diseño y desarrollo de algoritmos que permiten a una computadora mejorar un comportamiento automáticamente a través de la experiencia.

Un algoritmo de aprendizaje automático es capaz de extraer características y patrones comunes de un conjunto de datos (ejemplos o datos de entrenamiento), y aplicarlas a nuevos conjuntos de datos (datos de pruebas).

Los algoritmos de aprendizaje se clasifican en base a la salida de los mismos. Una forma de clasificarlos es agrupándolos en algoritmos de aprendizaje supervisado o predictivos y algoritmos de aprendizaje no supervisado o de descubrimiento del conocimiento (Weiss y Indurkha, 1998). Ver Figura 1.

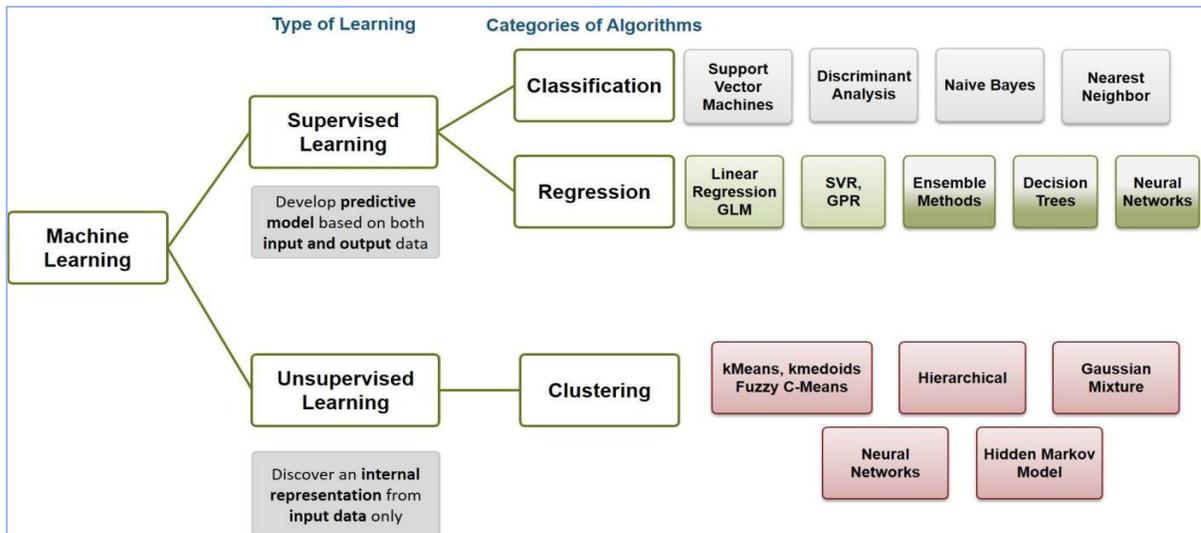


Figura 1. Machine Learning. Extraída de <http://embedded-computing.com/articles/analytics-driven-embedded-systems-part-2-developing-analytics-and-prescriptive-controls/#>

Los algoritmos supervisados predicen el valor de un atributo (etiqueta) de un conjunto de datos, conocidos otros atributos (atributos descriptivos). A partir de datos cuya etiqueta se conoce se induce una relación entre dicha etiqueta y otra serie de atributos. Esas relaciones sirven para realizar la predicción en datos cuya etiqueta es desconocida. Esta forma de trabajar se conoce como aprendizaje supervisado y se desarrolla en dos fases: Entrenamiento (construcción de un modelo usando un subconjunto de datos con etiqueta conocida) y Prueba (prueba del modelo sobre el resto de los datos).

Cuando una aplicación no es lo suficientemente madura no tiene el potencial necesario para una solución predictiva, por lo que en ese caso hay que recurrir a los métodos no supervisados que descubren patrones y tendencias en los datos actuales (no utilizan datos históricos). El descubrimiento de esa información sirve para llevar a cabo acciones y obtener un beneficio (científico o de negocio) de ellas.

De los tipos de tareas de la minería de datos, las predictivas son de las más populares y de uso más extendido en la minería de datos educativa (Romero y Ventura, 2010, 2013; Peña, 2014) debido a que permiten detectar problemas académicos con anticipación y aplicar las medidas necesarias. En la Figura 1 se muestran algunas de las técnicas de minería de ambas categorías.

La aplicación de los algoritmos de minería de datos requiere la realización de una serie de actividades previas encaminadas a preparar los datos de entrada debido a que, en muchas ocasiones dichos datos proceden de fuentes heterogéneas, no tienen el formato adecuado o contienen ruido. Por otra parte, es necesario interpretar y evaluar los resultados obtenidos. El proceso completo conocido como Knowledge Discovery in Databases (KDD) consta de las siguientes etapas (Cabena et al., 1997):

1. Determinación de objetivos.
2. Preparación de datos. Incluye dos subetapas:
 - a. Selección. Identificación de las fuentes de información externas e internas y selección del subconjunto de datos necesario.
 - b. Preprocesamiento. Estudio de la calidad de los datos y determinación de las operaciones de minería que se pueden realizar.
3. Transformación de datos. Conversión de datos en un modelo analítico.
4. Minería de datos. Tratamiento automatizado de los datos seleccionados con una combinación apropiada de algoritmos.
5. Análisis de resultados. Interpretación de los resultados obtenidos en la etapa anterior, generalmente con la ayuda de alguna técnica de visualización.
6. Asimilación del conocimiento. aplicación del conocimiento descubierto.

Si bien los pasos anteriores se llevan a cabo en el orden en que aparecen, se trata de un proceso altamente iterativo cuyos pasos se retroalimentan los unos a los otros. Por otro lado, es importante destacar que no todas las etapas requieren el mismo esfuerzo: la etapa de preprocesamiento es la más costosa ya que representa aproximadamente el 60% del esfuerzo total, mientras que la etapa de minería sólo representa el 10%.

A continuación se presentarán algunos de los algoritmos predictivos más utilizados en la etapa de minería.

Árboles de decisión

Los árboles de decisión se pueden aplicar prácticamente a todo. Los sistemas de aprendizaje basados en árboles de decisión son posiblemente el método más fácil de utilizar y de entender. Un árbol de decisión es un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que la decisión final a tomar se puede determinar siguiendo las condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas. Los árboles de decisión se utilizan desde hace siglos, y son especialmente apropiados para expresar procedimientos médicos, legales, comerciales, estratégicos, matemáticos, lógicos, etc.

Una de las ventajas más importantes de los árboles de decisión es que, en su forma más general, las opciones posibles a partir de una determinada condición son excluyentes. Esto permite analizar una situación y, siguiendo el árbol de decisión apropiadamente, llegar a una sola acción a decisión a tomar. Estos algoritmos se llaman algoritmos de partición o algoritmos de “divide y vencerás”.

Otra característica importante de los primeros algoritmos de aprendizaje de árboles de decisión es que una vez que elijamos la partición, dicha partición no se podrá cambiar, aunque más tarde se pueda llegar a pensar que ha sido una mala elección. Por lo tanto, uno de los aspectos más importantes en los sistemas de aprendizaje de árboles de decisión

es el llamado criterio de partición, ya que una mala elección de la partición (sobre todo, en las partes superiores del árbol) generará un peor árbol.

Lo que hace el algoritmo es simplemente, ir construyendo el árbol (desde el árbol que sólo contiene la raíz) añadiendo particiones y los hijos resultantes de cada partición. Lógicamente, en cada partición, los ejemplos se van dividiendo entre los hijos. Finalmente, se llega a la situación en la que todos los ejemplos que caen en los nodos inferiores son de la misma clase y esa rama ya no sigue creciendo. La única condición que hay que exigir es que las particiones al menos separen ejemplos en distintos hijos.

Como acabamos de explicar, los dos puntos más importantes para que el algoritmo funcione bien son los siguientes:

- *Particiones a considerar*: un conjunto de condiciones exhaustivas y excluyentes. Cuantas más particiones permitamos más expresivos podrán ser los árboles de decisión generados y, probablemente, más precisos. No obstante, cuantas más particiones elijamos, la complejidad del algoritmo será mayor. Por tanto, debemos encontrar un buen compromiso entre expresividad y eficiencia. Debido a esto, la mayoría de algoritmos de aprendizaje de árboles de decisión sólo permiten un juego muy limitado de particiones. Por ejemplo, el C4.5 contiene un solo tipo de partición para los atributos nominales y un solo tipo de partición para los atributos numéricos.
- *Criterio de selección de particiones*: incluso con sólo los dos tipos de particiones sencillas vistas, el número de particiones posibles en cada caso puede dispararse (si existen n atributos y m valores posibles para cada atributo, el número de particiones posibles es de $n*m$). Los algoritmos clásicos de aprendizaje de decisión son voraces, en el sentido de que una vez elegida la partición se continúa hacia abajo la construcción del árbol y no vuelven a plantearse las particiones ya construidas. Por tanto, se debe buscar un criterio que permita realizar una buena elección de la partición que parece más prometedora y que esto se haga sin demasiado esfuerzo computacional.

Las ventajas de un árbol de decisión son (Pérez, 2011):

- Facilita la interpretación de la decisión adoptada.
- Facilita la comprensión del conocimiento utilizado en la toma de decisiones.

- Explica el comportamiento respecto a una determinada decisión.
- Reduce el número de variables independientes.

Esto es lo que diferencia fundamentalmente los distintos algoritmos de árboles de decisión hasta la actualidad, como ID3, C4.5, etc.

Algoritmo de Regresión Logística

El algoritmo de Regresión Logística es un tipo de análisis estadístico orientado a la predicción de una variable categórica en función de otras variables consideradas como parámetros predictores (Fernández, 2011).

Construcción del modelo de regresión

Para construir un modelo de regresión, hay que centrarse en el tipo de variables que se desea introducir (categóricas o continuas) para posteriormente, definir cuál de los métodos que ofrecen los paquetes estadísticos se utilizará para obtener el modelo de regresión más fiable.

- Selección de las variables del modelo:

Existen varios métodos para construir un modelo de regresión, es decir, para seleccionar de entre todas las variables introducidas en el modelo, cuáles son las necesarias para explicarlo.

Para la construcción del modelo de regresión se pueden utilizar las siguientes técnicas:

- Técnica de pasos hacia adelante (Forward)): consiste en ir introduciendo las variables en el modelo únicamente si cumplen una serie de condiciones hasta que no se pueda introducir ninguna más, hasta que ninguna cumpla la condición impuesta;

- Técnica de pasos hacia atrás (Backward): se introducen en el modelo todas las variables y se van suprimiendo si cumplen una serie de condiciones definidas a priori hasta que no se pueden eliminar más, es decir ninguna variable cumpla la condición impuesta;
- Técnica por pasos (Stepwise): combina los dos métodos anteriores, adelante y atrás introduciendo o eliminando variables del modelo si cumplen una serie de condiciones definidas a priori hasta que ninguna variable satisfaga ninguna de las condiciones expuestas de entrada o salida del modelo
- Técnica de introducir todas las variables obligatoriamente (Enter): Esta última técnica de selección de variables para construir el modelo de regresión, produce que el proceso de selección de las variables sea manual, partiendo de un modelo inicial, en el que se obliga a que entren todas las variables seleccionadas, se va evaluando qué variable es la que menos participa en él y se elimina, volviendo a construir un nuevo modelo de regresión aplicando la misma técnica, pero excluyendo la variable seleccionada y aplicando el mismo proceso de selección. Este proceso se repite reiteradamente hasta que se considere que el modelo obtenido es el que mejor se ajusta a las condiciones impuestas y que no se puede eliminar ninguna variable más de las que los componen.

- Métodos de construcción del modelo de regresión

Para evaluar la adecuación de los modelos construidos, es conveniente comenzar a evaluar el modelo saturado, es decir el modelo que contiene todas las variables de interés que queramos evaluar y todas las interacciones posibles. Progresivamente se van eliminando del modelo aquellos términos no significativos, respetando el modelo jerárquico y comenzando por los términos de interacción superiores. Como hemos dicho anteriormente, si un término de interacción es significativo, no podrán eliminarse del modelo los términos de interacción de grado inferior, ni los términos independientes de las variables que participan en la interacción. Las variables introducidas en el modelo se van eliminando progresivamente a cada nuevo modelo que se construye en base a los resultados obtenidos en el modelo anterior, y se van evaluando los nuevos modelos, de la manera que se explicará más adelante. Es importante observar que los coeficientes de las variables que permanezcan en el modelo no varían de forma exagerada tras la eliminación de alguno de los términos del modelo, dado que si así sucediera, podría tratarse de un factor de confusión

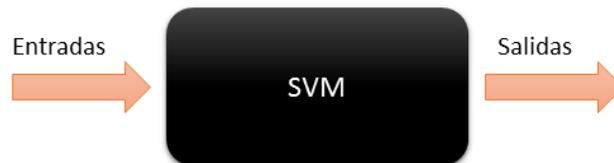
y por tanto debería mantenerse la variable en cuestión en el modelo, para permitir el ajuste del resto de variables y no obtener resultados artificiales.

- Obtención y validación del modelo más adecuado

Los modelos de regresión pueden ser validados en otro conjunto de datos de similares características, extraídos de la misma población, por ejemplo, con el fin de evaluar su fiabilidad. Otra posibilidad, cuando se trabaja con muestras grandes, es dividir aleatoriamente la muestra en dos grupos y utilizarlos para obtener dos modelos con el fin compararlos para comprobar si se obtienen resultados similares. Por otro lado, identificar el modelo más adecuado consistirá en evaluar diferentes parámetros de los modelos de regresión. El modelo de regresión lineal se estima mediante una técnica denominada método de los mínimos cuadrados, mientras que en la regresión logística se utiliza el método de máxima verosimilitud.

Support Vector Machine

Las Máquinas de Vectores Soporte, creadas por Vladimir Vapnik, constituyen un método basado en aprendizaje para la resolución de problemas de clasificación y regresión. En ambos casos, esta resolución se basa en una primera fase de entrenamiento (donde se les informa con múltiples ejemplos ya resueltos, en forma de pares {problema, solución}) y una segunda fase de uso para la resolución de problemas. En ella, las SVM se convierten en una “caja negra” que proporciona una respuesta (salida) a un problema dado (entrada).



La regresión se basa en buscar la curva que modele la tendencia de los datos y, según ella, predecir cualquier otro dato en el futuro.

Redes Neuronales

Una red neuronal puede ser descrita como un modelo de regresión no lineal cuya estructura se inspira en el funcionamiento del sistema nervioso. En términos generales, una red consiste en un gran número de unidades simples de proceso, denominadas neuronas, que actúan en paralelo y están conectadas mediante vínculos ponderados (Hilera y Martínez, 2000; Reed y MarksII, 1999).

Se trata de un modelo matemático inspirado en el comportamiento biológico de las neuronas y en cómo se organizan formando la estructura del cerebro. Las redes neuronales intentan aprender mediante ensayos repetidos como organizarse mejor a sí mismas para conseguir maximizar la predicción.

El primer modelo matemático de una neurona artificial, creado con el fin de llevar a cabo tareas simples, fue presentado en el año 1943 en un trabajo conjunto entre el psiquiatra y neuroanatomista Warren McCulloch y el matemático Walter Pitts. Un modelo de red neuronal se compone de nodos, que actúan como input, output o procesadores intermedios. Cada nodo se conecta con el siguiente conjunto de nodos mediante una serie de trayectorias ponderadas. Basado en un paradigma de aprendizaje, el modelo toma el primer caso, y toma inicial basada en las ponderaciones. Se evalúa el error de predicción y modifica las ponderaciones para mejorar la predicción, a continuación se evalúa el siguiente caso con las nuevas ponderaciones y se modifican para mejorar la predicción de los casos ya evaluados, el ciclo se repite para cada caso en lo que se denomina la fase de preparación o evaluación. Cuando se ha calibrado el modelo, con la muestra test se evalúan los resultados globales.

Las redes neuronales se hicieron operativas por primera vez a finales de los 50. Rosenblat F. (1958) creó el perceptrón, un algoritmo de reconocimiento de patrones basado en una red de aprendizaje de computadora de dos capas usando una simple suma y la resta. A finales de los 60 el proceso de investigación sufrió un estancamiento, y ha sido a partir de los años 80 del siglo pasado, cuando se produjo el mayor desarrollo teórico. Son

muchos los modelos de redes neuronales, el más utilizado es el algoritmo backpropagation fue creado por Werbos P. (1990).

Un ejemplo de modelo neuronal con n entradas, consta de:

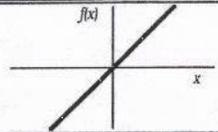
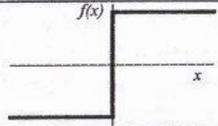
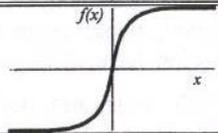
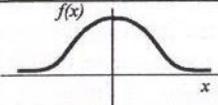
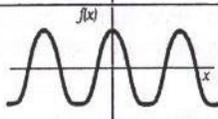
- Un conjunto de entradas x_1, \dots, x_n .
- Los pesos sinápticos w_1, \dots, w_n , correspondientes a cada entrada.
- Una función de agregación, \sum .
- Una función de activación, f .
- Una salida, Y .

Las entradas son el estímulo que la neurona artificial recibe del entorno que la rodea, y la salida es la respuesta a tal estímulo. La neurona puede adaptarse al medio circundante y aprender de él modificando el valor de sus pesos sinápticos, y por ello son conocidos como los parámetros libres del modelo, ya que pueden ser modificados y adaptados para realizar una tarea determinada.

En este modelo, la salida neuronal Y está dada por:

$$Y = f\left(\sum_{i=1}^n w_i x_i\right)$$

La función de activación se elige de acuerdo a la tarea realizada por la neurona. Entre las más comunes dentro del campo de las RNAs podemos destacar:

	Función	Rango	Gráfica
Identidad	$y = x$	$[-\infty, +\infty]$	
Escalón	$y = \text{sign}(x)$ $y = H(x)$	$\{-1, +1\}$ $\{0, +1\}$	
Sigmoidea	$y = \frac{1}{1 + e^{-x}}$ $y = \text{tgh}(x)$	$[0, +1]$ $[-1, +1]$	
Gaussiana	$y = Ae^{-Bx^2}$	$[0, +1]$	
Sinusoidal	$y = A \text{sen}(\omega x + \varphi)$	$[-1, +1]$	

Redes Bayesianas (Sucar, 2006)

Las Redes Bayesianas son parte de las técnicas modernas de la Inteligencia Artificial (AI). Aunque un teorema fundamental de la probabilidad condicional fue presentado por Thomas Bayes en el siglo XVIII, el término "Red Bayesiana" y sus conceptos fueron introducidos en la década de 1980 en el trabajo de Judea Pearl. Investigadores han presentado declaraciones referidas a las Redes Bayesianas, por ejemplo Darwiche en 2009: "Las Redes Bayesianas son tan importantes para la Inteligencia Artificial (AI) y el Aprendizaje Automático como los circuitos booleanos son a la informática", " las Redes Bayesianas han revolucionado la AI".

Las Redes Bayesianas (RB) modelan un fenómeno mediante un conjunto de variables y las relaciones de dependencia entre ellas. Es a través de este modelo que se puede estimar en base a las variables conocidas, la probabilidad de las variables no conocidas. Este enfoque representa una buena estrategia para hacer frente a problemas

relacionados con la incertidumbre, donde las conclusiones no pueden ser construidas sólo a partir de un conocimiento previo sobre el problema.

En un principio, estos modelos eran construidos "a mano" basados en un conocimiento experto, pero en los últimos años se han desarrollado diversas técnicas para aprender a partir de datos tanto la estructura como los parámetros asociados al modelo.

Las RBs se encuentran entre los modelos gráficos más populares. La principal diferencia, con respecto a otros modelos, está en que sus arcos son dirigidos y representan dependencia condicional entre las variables. El nombre de este modelo proviene del hecho que gran parte de la teoría relevante con este tipo de redes se basa en la estadística bayesiana.

Conceptos básicos.

Una RB es un grafo acíclico dirigido en el que cada nodo representa una variable aleatoria y los arcos las relaciones de dependencia directa entre las variables. La estructura de la RB provee información sobre las relaciones de dependencia e independencia condicional existentes entre las variables. Estas relaciones simplifican la representación de la función de probabilidad conjunta como el producto de las funciones de probabilidad condicional de cada variable.

Sea $U = \{X_1, X_2, \dots, X_n\}$ un conjunto de variables aleatorias. Formalmente, una red bayesiana para U es un par $B = \langle G, T \rangle$ en el que:

- G es un grafo acíclico dirigido en el que cada nodo representa una de las variables X_1, X_2, \dots, X_n , y cada arco representa relaciones de dependencia directas entre las variables. La dirección de los arcos indica que la variable "apuntada" por el arco depende de la variable situada en su origen.

- T es un conjunto de parámetros que cuantifica la red. Contiene las probabilidades $PB(x_i | p_{xi})$ para cada posible valor x_i de cada variable X_i y cada posible valor p_{xi} de PX_i , donde éste último denota al conjunto de padres de X_i en G.

Cabe destacar que la topología o estructura de la red no sólo proporciona información sobre las dependencias probabilísticas entre las variables, sino también sobre las independencias condicionales de una variable o conjunto de ellas, dada otra u otras variables. Cada variable es independiente de las variables que no son descendientes suyas en el grafo, dado el estado de sus variables padre.

La inclusión de las relaciones de independencia en la propia estructura del grafo hace de las redes bayesianas una buena herramienta para representar conocimiento de forma compacta (se reduce el número de parámetros necesarios). Además, proporcionan métodos flexibles de razonamiento basados en la propagación de las probabilidades a lo largo de la red de acuerdo con las leyes de la teoría de la probabilidad.

La Figura 2 presenta un ejemplo concreto de una red bayesiana que representa un cierto conocimiento sobre medicina. En este caso los nodos representan enfermedades, síntomas y factores que causan algunas enfermedades. Tal como se mencionó anteriormente la variable a la que apunta un arco es dependiente de la que está en el origen de éste. Por ejemplo, fiebre depende de tifoidea y gripe.

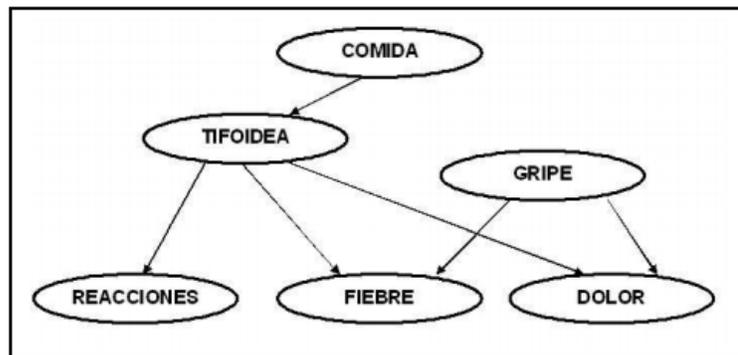


Figura 2. Ejemplo de redes bayesianas.

Pueden observarse las suposiciones de independencia acertadas por la red, por ejemplo, reacciones es condicionalmente independiente de comida, gripe, fiebre y dolor (nodos no descendientes de reacciones) dado tifoidea (su único nodo padre).

Es decir, se observa en la red que: $P(R | C, T, G, F, D) = P(R | T)$, donde R es reacciones, C es comida, T es tifoidea, G es gripe, F es fiebre, y D es dolor.

Inferencia, predicción.

A partir de una red ya construida, y dados los valores concretos de algunas variables de una instancia, podrían tratarse de estimarse los valores de otras variables de la misma instancia aplicando razonamiento probabilístico.

El razonamiento probabilístico sobre las RBs consiste en propagar los efectos de las evidencias (variables conocidas) a través de la red para conocer las probabilidades a posteriori de las variables desconocidas. De esta forma se puede determinar un valor estimado para dichas variables en función de los valores de probabilidad obtenidos.

Cuando se conocen los valores observados para todas las variables de la red excepto para una, obtener una estimación para ésta es inmediato a partir de la fórmula de la distribución de probabilidad conjunta de la red.

En un caso más general interesaría obtener una estimación del valor de alguna variable dados valores observados para un subconjunto de todas las restantes. En general, una red puede usarse para calcular la distribución de probabilidad para cualquier subconjunto de variables dados los valores de cualquier subconjunto de las restantes.

El mecanismo de inferencia sobre RBs permite utilizarlas para construir clasificadores. Para ello, se ha de construir una RB en la que clase y atributos sean las variables interrelacionadas en el grafo. La clase corresponderá a la variable desconocida, objetivo de la inferencia. Dada una instancia cualquiera para la que se conozcan todos sus atributos,

la clasificación se efectuará infiriendo sobre el grafo la probabilidad posterior de cada uno de los valores de la clase, y seleccionando aquel valor que maximice dicha probabilidad.

Clasificadores bayesianos

Un clasificador es una función que asigna un valor de un atributo, llamado clase, a instancias o ejemplos descritos mediante un conjunto de atributos. Cuando una RB se construye con la finalidad de predecir el valor de una variable clase dada cualquier configuración en el resto de las variables, se tiene un "clasificador bayesiano".

Los clasificadores bayesianos son ampliamente utilizados debido a que presentan ciertas ventajas:

- Generalmente, son fáciles de construir y de entender.
- Las inducciones de estos clasificadores son extremadamente rápidas, requiriendo sólo un paso para hacerlo.
- Es muy robusto considerando atributos irrelevantes.
- Toma evidencia de muchos atributos para realizar la predicción final.

Un clasificador bayesiano que suele ser bastante preciso a pesar de su simplicidad es el "clasificador bayesiano simple" (del inglés Naive Bayes classifier).

Aprendizaje de las RBs

El aprendizaje, en general, de redes bayesianas consiste en inducir un modelo, estructura y parámetros asociados, a partir de datos. Este puede dividirse naturalmente en dos partes:

- Aprendizaje estructural. Obtener la estructura o topología de la red.
- Aprendizaje paramétrico. Dada la estructura, obtener las probabilidades asociadas.

Las siguientes son algunas de las ventajas que presentan las RBs:

- Permiten representar al unísono la dimensión cualitativa y cuantitativa de un problema en un entorno gráfico inteligible.
- Pueden trabajar con datos perdidos de una manera eficiente, algo que en la práctica es deseable.
- Permiten reducir el sobre ajuste de los datos.
- Posibilitan el descubrimiento de la estructura causal subyacente en un conjunto de datos.
- Representan toda la información en un único formato (probabilístico y gráfico) lo que hace sencillas las interpretaciones, permite retractarse de conclusiones obtenidas con anterioridad ya que no son razonables a la luz de nuevas evidencias, proporcionan una visión general del problema, generan un conjunto de alternativas ordenadas y facilita la explicación de las conclusiones.
- Permiten realizar inferencias bidireccionales; esto es, desde los efectos a las causas y desde las causas a los efectos, etc.

Series Temporales

Una serie temporal se define como una secuencia de N observaciones ordenadas y equidistantes cronológicamente sobre una característica (serie univariante o escalar) o sobre varias características (serie multivariante o vectorial) de una unidad observable en distintos momentos.

Dependiendo de cómo sean las observaciones una serie temporal puede ser de dos tipos:

- Continuas: cuando los valores se ofrecen de forma permanente, de manera tal que cada uno de ellos representa el estado de la variable en un instante, el cual puede ser tan pequeño como teóricamente se quiera suponer.
- Discretas: cuando los valores se ofrecen para intervalos de tiempo, generalmente homogéneos y donde representan la magnitud acumulada del estado de la variable durante ese intervalo.

Los modelos de series de tiempo tienen un enfoque netamente predictivo y en ellos los

pronósticos se elaborarán solo con base al comportamiento pasado de la variable de interés. Podemos distinguir dos tipos de modelos de series de tiempo (Pindick et al., 2001):

- Modelos deterministas: se trata de métodos de extrapolación sencillos en los que no se hace referencia a las fuentes o naturaleza de la aleatoriedad subyacente en la serie. Su simplicidad relativa generalmente va acompañada de menor precisión. Ejemplo de modelos deterministas son los modelos de promedio móvil en los que se calcula el pronóstico de la variable a partir de un promedio de los n valores inmediatamente anteriores.
- Modelos estocásticos: se basan en la descripción simplificada del proceso aleatorio subyacente en la serie. En términos sencillos, se asume que la serie observada Y_1, Y_2, \dots, Y_T se extrae de un grupo de variables aleatorias con una cierta distribución conjunta difícil de determinar, por lo que se construyen modelos aproximados que sean útiles para la generación de pronósticos.

La serie y_T podrá ser estacionaria o no estacionaria:

- Serie no estacionaria: es aquella cuyas características de media, varianza y covarianza cambian a través del tiempo lo que dificulta su modelamiento. Sin embargo, en muchas ocasiones, si dicha serie es diferenciada una o más veces la serie resultante será estacionaria (procesos no estacionarios homogéneos).
- Serie estacionaria: es aquella cuya media y varianza no cambian a través del tiempo y cuya covarianza sólo es función del rezago. Gracias a estas características podremos modelar el proceso subyacente a través de una ecuación con coeficientes fijos estimados a partir de los datos pasados.

Actualmente diversas organizaciones requieren conocer el comportamiento futuro de ciertos fenómenos con el fin de predecir qué ocurrirá con una variable en el futuro a partir del comportamiento de esa variable en el pasado. Es de mucha utilidad en las organizaciones realizar predicciones de corto y mediano plazo, por ejemplo para predecir qué ocurrirá con la demanda de un determinado producto (Gujarati et al., 2003).

De acuerdo con Nerlove, Grether y Carvalho (1979) las series temporales están conformadas por cuatro componentes, los cuales son:

- Tendencia (T) o movimiento a largo plazo. Representa el comportamiento predominante de la serie.
- Ciclo (C) o fluctuaciones irregulares. caracterizado por oscilaciones alrededor de la tendencia con una larga duración, y sus factores no son claros. Por ejemplo, fenómenos climáticos, que tienen ciclos que duran varios años.
- Estacionalidad (E) o fluctuaciones de mayor o menor regularidad.

Componente Aleatorio(A) o residual. Se trata de movimientos erráticos que no siguen un patrón específico y que obedecen a causas diversas. Este componente es prácticamente impredecible. Este comportamiento representan todos los tipos de movimientos de una serie de tiempo que no son tendencia, variaciones estacionales ni fluctuaciones cíclicas.

4.2 MODELIZACIÓN Y APLICACIÓN DEL TABLERO DE CONTROL

Los datos se extraen desde el sistema SIU – Guaraní de la Universidad y son enviados a al software Pentaho instalado también de la Universidad. Desde allí se realizan consultas que generan archivos de texto que luego son importados en el Tablero de Control.

Uno de los propósitos del tablero es dar respuesta al objetivo, mencionado al inicio de esta tesis, de: Diseñar una herramienta software que permita a los docentes, sin conocimientos sobre minería de datos obtener, interpretar y tomar decisiones para mejorar el proceso de enseñanza-aprendizaje.

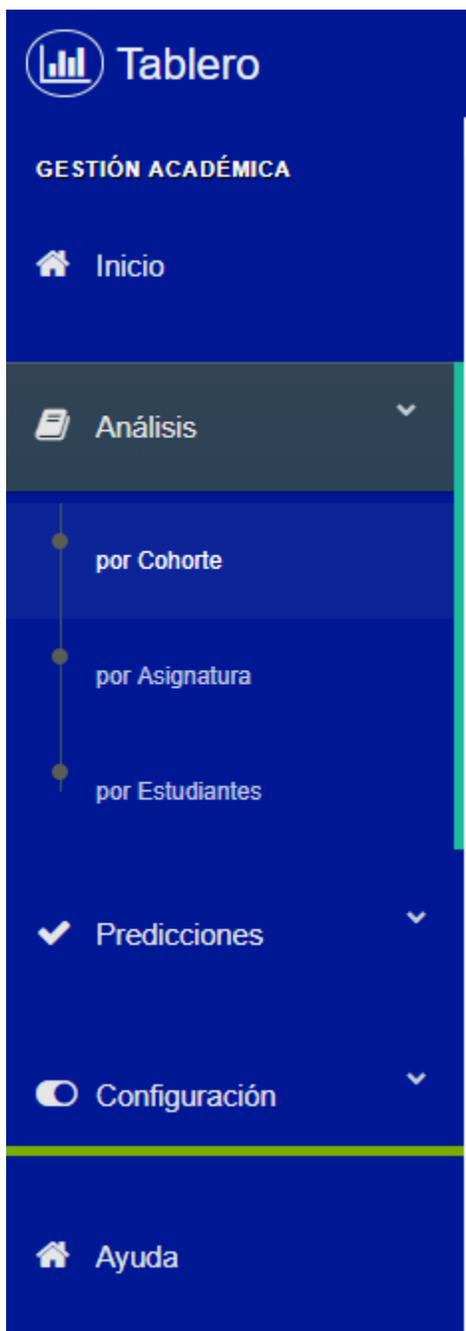
Es por ello que, en relación a la fuente de datos, desde el tablero se brinda la posibilidad de incorporar datos de los estudiantes que no se encuentran en la base de datos de la fuente utilizada para los análisis y predicciones (la que se utiliza es la que se extrae del SIU – Guaraní). El motivo por el cual algunos de estos datos no se obtienen desde el SIU - Guaraní es porque no están cargados dentro del mismo. Dado que la carga de datos de Guaraní responde a criterios fijados por la secretaría Académica de la Universidad, se pretende desde este trabajo observar qué datos pueden ser útiles para realizar modelos de predicciones. Todos aquellos datos que no pueden ser obtenidos desde el Guaraní (ya sea porque no se dispongan o porque no son parte del modelo de datos de dicho sistema -), se proponen para ser agregado directamente al tablero mediante la ficha de Tutoría que tiene cada estudiante.

Niveles de información:

La información se carga y analiza por cada una de las carreras que conforman la disciplina Informática de la UNNOBA, permitiendo discriminar los niveles de análisis e información sobre:

- Cohortes
- Asignaturas
- Estudiantes

La unidad central de información es el estudiante. Los análisis están centrados en el análisis de los indicadores previamente definidos que permiten realizar el control y seguimiento del desempeño del estudiante, pudiendo, mediante cruces de información prever acciones y realizar predicciones.

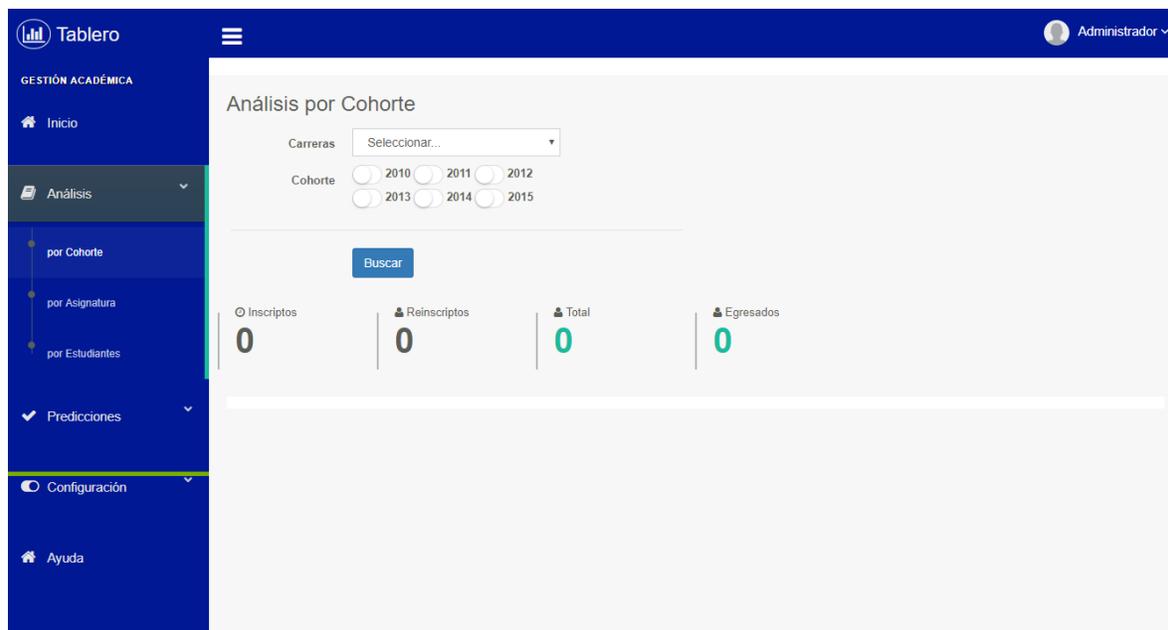


En la sección de *Análisis* aparecen los niveles de información a analizar: por Cohortes, Asignaturas y Estudiantes.

Sección de análisis de información por Cohorte

El filtro de búsqueda permite elegir una carrera y seleccionar, al menos, un año de cohorte.

Se permitirá analizar y comparar hasta 6 años de cohortes determinados por un periodo consecutivo de una determinada carrera. Por ejemplo, se podrá analizar las cohortes 2010 a 2015.



En relación a los estudiantes asociados a una determinada carrera por cohorte o año académico se podrá obtener información sobre:

- **Inscriptos:** ingresantes a la carrera que cumplieron con todos los requisitos en el año académico o están condicionales.
- **Reinscriptos:** estudiantes con actividad en el año académico que no son ingresantes. Aquí se discriminará la cantidad de reinscriptos según cohorte de origen para poder analizar el rezago académico.
- **Total de estudiantes:** se totaliza en base a la suma de la cantidad de inscriptos y reinscriptos
- **Egresados:** se define como egresado al estudiante que se le ha otorgado un título o ha completado las materias del Plan de Estudio.

En el tablero se seleccionará previamente la carrera analizar y luego la/s cohorte/s. Si se selecciona una cohorte se mostrará en números la información y luego un listado de estudiantes.

Análisis por Cohorte

Carreras: Licenciatura en Sistemas

Cohorte: 2010 2011 2012 2013 2014 2015

Buscar

Inscriptos: **85** |
 Reinscriptos: **0** |
 Total: **85** |
 Egresados: **20**

La lista de ESTUDIANTES

Mostrar 10 registros Copy CSV Print Buscar

Apellido	Nombre	DNI	# Cursadas aprobadas	# Finales aprobados	% de avance	Múltiple carrera	Asignaturas pendientes de cursar	Asignaturas con finales pendientes	Año de última cursada aprobada	Año de último final aprobado	Datos de tutoría
Acuña	Lella Mercedes	33547291	0	0	0	NO	44	0	0	0	Ver
Airaldi	Santiago	33096899	29	29	63	NO	15	0	2014	2015	Ver
Amet	Lucrecia	32772633	17	17	36	NO	27	0	2012	2015	Ver
Angelucci	Lucas Agustin	35399910	22	22	47	NO	22	0	2010	2013	Ver
Arzuaga	Nelson Miguel	36050646	0	0	0	NO	44	0	0	0	Ver
Aubalat	Ornela Daiana	33880765	29	26	56	NO	15	3	2015	2016	Ver
Aureli	Gonzalo Alberto	32345003	5	5	10	NO	39	0	2010	2011	Ver
Auza	Martín	34125679	1	1	2	NO	43	0	2011	2011	Ver
Bañfi	Gisela Soledad	33569567	28	24	52	NO	16	4	2015	2016	Ver
Basile	Matias	34240879	43	43	93	NO	1	0	2011	2014	Ver

Mostrando registros del 1 al 10 de un total de 142 registros

Anterior 1 2 3 4 5 ... 15 Siguiente

Seleccionando más de una cohorte, estos valores se totalizarán contemplando de no duplicar información y se mostrarán gráficas comparativas por cohortes (en gráfico de barras) de determinados indicadores asociados a los estudiantes:

- **Comparativa por resultados de cursada de las asignaturas:** Se mostrará el total de inscriptos, reinscriptos, aprobados, desaprobados y ausentes, discriminado por cohorte y totalizado para todas las asignaturas.
- **Comparativa por exámenes finales de las asignaturas:** Se mostrará el total de exámenes rendidos, aprobados, desaprobados y ausentes, discriminado por cohorte y totalizado para todas las asignaturas.



Respecto al listado de estudiantes, el mismo se muestra en forma de planilla, adicionando información de indicadores previamente establecidos en relación a la cohorte que se está analizando. La información que se muestra del estudiante para una determinada cohorte de la carrera seleccionada será:

- **Ficha general del estudiante:**
 - Apellido.
 - Nombre.
 - DNI.

- **Trayectoria académica en la universidad:**
 - Cantidad de cursadas aprobadas.
 - Cantidad de finales aprobados.
 - Porcentaje de avance en la carrera (Diferenciando en base a la cantidad de cursadas aprobadas y finales aprobados)
 - Múltiple inscripción. Permite conocer si está inscripto en más de una carrera.
 - Cantidad de asignaturas pendientes para cursar.
 - Cantidad de asignaturas pendientes para rendir final.
 - Año de última cursada aprobada.
 - Año del último final rendido (independientemente del resultado obtenido).

Los indicadores establecidos trabajan con un nivel de alerta en tres estados: rojo, amarillo, verde. Para cada indicador se definen valores que conforman el segmento de los niveles, indicando el color rojo cuando está por debajo del intervalo mínimo establecida, amarillo para cuando está dentro del intervalo y verde para cuando está por encima del intervalo máximo establecido.

Estos indicadores se pueden configurar desde el propio tablero según la necesidad de alerta que se necesite visualizar, por ejemplo, determinar las cotas mínimas y máximas de la información relacionada con: cantidad de materias pendientes por cursar, cantidad de finales pendientes, año de la última cursada y año del último final aprobado.

A su vez, las columnas permiten ordenarse de forma ascendente o descendente, permitiendo una mejor visualización de la información a analizar. Se podría, por ejemplo, ordenar el listado según el año del último final rendido:

Tablero Administrador

GESTIÓN ACADÉMICA

- Inicio
- Análisis
 - por Cohorte
 - por Asignatura
 - por Estudiantes
- Predicciones
- Configuración
- Ayuda

Análisis por Cohorte

Carreras: Licenciatura en Sistemas

Cohorte: 2010 2011 2012 2013 2014 2015

Buscar

Inscriptos: **85** |
 Reinscriptos: **0** |
 Total: **85** |
 Egresados: **20**

La lista de ESTUDIANTES

Mostrar 10 registros Copy CSV Print Buscar

Apellido	Nombre	DNI	# Cursadas aprobadas	# Finales aprobados	% de avance	Múltiple carrera	Asignaturas pendientes de cursar	Asignaturas con finales pendientes	Año de última cursada aprobada	Año de último final aprobado	Datos de tutoría
Monaco	Hugo Alberto	36575084	45	35	76	NO	0	10	2015	2016	Ver
Quaglia	Esteban Oscar	36098296	44	33	71	NO	0	11	2015	2016	Ver
Benitez	Danilo Ruben	33521458	43	43	93	NO	1	0	2014	2016	Ver
Granero	Alejo	34984794	43	40	84	NO	1	3	2015	2016	Ver
Gomez Luengo	Claudio David Adolfo	35042966	41	40	84	NO	3	1	2015	2016	Ver
Terrarosa	Juan Francisco	36624653	41	38	82	NO	3	3	2015	2016	Ver
De Bello	Maximiliano	36421792	39	36	78	NO	4	3	2015	2016	Ver
Molina	Jorge Daniel	24524094	39	39	82	NO	5	0	2014	2016	Ver
Wehren	Martin Ezequiel	32791334	39	39	84	NO	5	0	2015	2016	Ver
Figueroa	Nahuel Emiliano	34758938	33	30	65	NO	11	3	2015	2016	Ver

Mostrando registros del 1 al 10 de un total de 142 registros

Anterior 1 2 3 4 5 ... 15 Siguiente

O bien según el porcentaje de avance en la carrera:

Tablero Administrador

GESTIÓN ACADÉMICA

Inicio

Análisis

- por Cohorte
- por Asignatura
- por Estudiantes
- Predicciones
- Configuración
- Ayuda

Análisis por Cohorte

Carreras: Licenciatura en Sistemas

Cohorte: 2010 2011 2012 2013 2014 2015

[Buscar](#)

Inscritos: **85**
Reinscritos: **0**
Total: **85**
Egresados: **20**

La lista de ESTUDIANTES

Mostrar registros Copy CSV Print Buscar

Apellido	Nombre	DNI	# Cursadas aprobadas	# Finales aprobados	% de avance	Múltiple carrera	Asignaturas pendientes de cursar	Asignaturas con finales pendientes	Año de última cursada aprobada	Año de último final aprobado	Datos de tutoría
Mendiola	Julia	36574217	46	46	100	NO	0	0	2014	2015	Ver
Gomez	Federico Martin	32363716	44	44	95	NO	0	0	2014	2015	Ver
Benitez	Danilo Ruben	33521458	43	43	93	NO	1	0	2014	2016	Ver
Basile	Matias	34240879	43	43	93	NO	1	0	2011	2014	Ver
Perez	Diego Gonzalo	34332923	45	42	91	NO	0	3	2015	2015	Ver
Blondini	Victor Hernan	30344290	44	43	91	NO	0	1	2015	2015	Ver
Canone	Martin Gabriel	34216754	44	43	91	NO	0	1	2015	2015	Ver
Rossi	Esteban Nicolas	34240811	44	43	91	NO	0	1	2015	2015	Ver
Carea	Mariano Daniel	29416234	43	43	91	NO	1	0	2012	2013	Ver
Negrini	Ariel Mario	32570607	42	42	91	NO	2	0	2012	2013	Ver

Mostrando registros del 1 al 10 de un total de 142 registros

Anterior 1 2 3 4 5 ... 15 Siguiente

Junto con la **ficha general del estudiante** y la **trayectoria académica en la universidad**, que se muestra para una determinada cohorte de la carrera seleccionada, está la opción de acceder a la ficha con los datos de la tutoría, en la cual puede observarse información adicional al desarrollo académico, la cual puede obtenerse de la información que los estudiantes completan en su perfil de usuario en el sistema de SIU-Guaraní, o bien es provista por algún rol de tutoría que obtuvo la información bajo el consentimiento del propio estudiantes. La información que aquí se incorpora es:

- **Personales:**
 - Sexo
 - Edad al ingreso de la carrera.
 - Lugar de procedencia.
 - Lugar de residencia durante el período lectivo (En caso de no residir en sede universitaria, demanda horaria de viaje)
 - Situación familiar al momento del ingreso:
 - a) Estado civil.
 - b) Cantidad de familiares a cargo.
 - c) Cantidad de convivientes.

- **Socioeconómicas:**
 - Situación laboral:
 - a) Trabaja (media jornada o jornada completa)
 - b) Horas de trabajo semanal y año de ingreso laboral
 - c) No trabaja.
 - Fuente de financiamiento de los estudios:
 - a) Aportes familiares.
 - b) Trabajo propio.
 - c) Becas.
 - d) Otros.

- **Académicas:**
 - Trayectoria académica previa
 - Tipo de secundario: Escuela Pública o privada.
 - Título secundario obtenido.
 - Orientación vocacional.
 - Otros títulos superiores previos.
 - Rendimiento académico en la carrera
 - Año de ingreso.
 - Año de primer cursada aprobada.
 - Año del primer final rendido.
 - Año de la última cursada aprobada.
 - Año del último final rendido.

- Cantidad de cursadas aprobadas.
- Cantidad de finales aprobados.
- Porcentaje de avance en la carrera (- diferenciando en base a la cantidad de cursadas aprobadas y finales aprobados -) en relación al total de las materias según el plan de estudios.
- Promedio general de la carrera.
- Cantidad de asignaturas pendientes para cursar.
- Cantidad de asignaturas pendientes para rendir final.

Administrador

Inicio

Análisis

Predicciones

Configuración

Ayuda

Julia Mendiola

Información			
DNI	36574217	Año de ingreso	2010
Año de primer cursada	2010	Año de primer final rendido	2010
Año ultima cursada aprobada	2014	Año de ultimo final rendido	2015
Cursadas aprobadas	46	Finales aprobados	46
Porcentaje de avance	100 %		
Cursadas pendientes de cursar	0	Asignaturas pendientes de final	0
Año de ultima cursada aprobada	2014		
Sexo	FEMENINO	Edad de ingreso	17
Procedencia		Residencia	
Financiamiento			
Aporte familiar		Trabajo propio	
Beca		Nombre	
Situacion laboral			
Hs trabajo semanal	0	Comienzo a trabajar en cursada	
Trabajo en la cursada			
Situacion Familiar			
Estado civil		Familiares a cargo	0
Cantidad convivientes	0		
Educacion de los padres			
Múltiple Carrera			
Nivel maximo			
Detalle			

Volver

Modificar

La información correspondiente a la ficha de tutoría puede ser editada desde el mismo tablero de control, permitiendo así una mayor flexibilidad sobre los datos que se analizan de cada estudiante.

Sección de información por Asignaturas

El filtro de búsqueda permite elegir una carrera y seleccionar, al menos, un año de cohorte.

Al igual que en la Sección de información de Cohorte, se permitirá seleccionar hasta 6 años de cohortes determinados por un periodo de años consecutivo.

Para seleccionar la asignatura se podrán aplicar, previamente, dos filtros alternativos:

1. Un filtro de área de evaluación de CONEAU (determinadas por la Resolución 786/09): *Ciencias Básicas, Teoría de la Computación, Algoritmos y Lenguajes, Arquitectura, Sistemas Operativos y Redes, Ingeniería de Software, Bases de Datos y Sistemas de Información y Aspectos Profesionales y Sociales;*
2. Un filtro por año de ubicación de la asignatura dentro de la carrera previamente seleccionada.

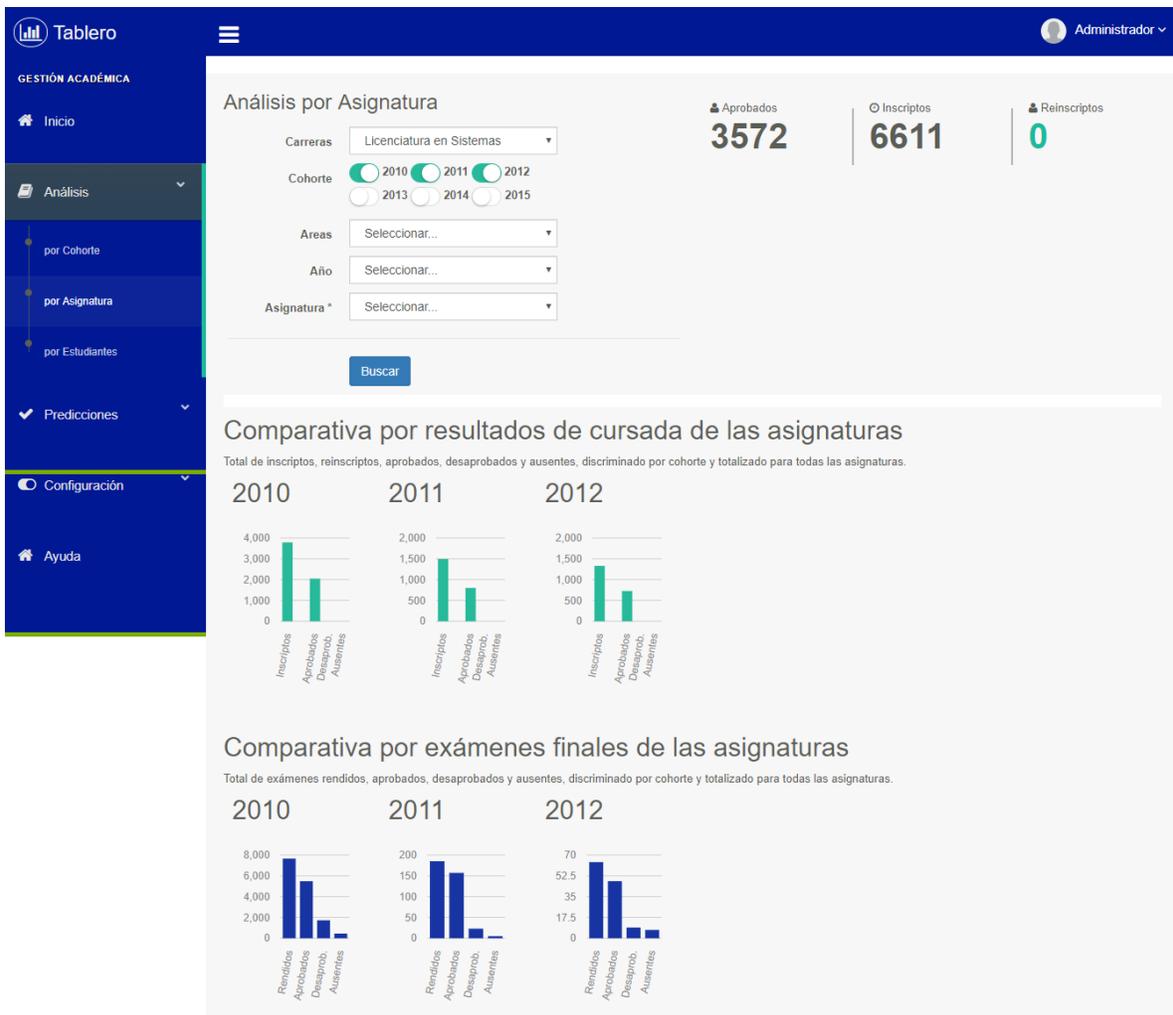
En caso de no aplicarse dichos filtros, se puede seleccionar directamente la asignatura.

The screenshot shows a web application interface for 'Gestión Académica'. The top navigation bar includes the 'Tablero' logo and the user role 'Administrador'. The left sidebar contains a menu with options: 'Inicio', 'Análisis' (expanded), 'por Cohorte', 'por Asignatura', 'por Estudiantes', 'Predicciones', 'Configuración', and 'Ayuda'. The main content area is titled 'Análisis por Asignatura'. It features several filters: 'Carreras' (dropdown), 'Cohorte' (radio buttons for years 2010, 2011, 2012, 2013, 2014, 2015), 'Áreas' (dropdown), 'Año' (dropdown set to 'Segundo Año'), and 'Asignatura *' (dropdown). To the right of these filters are three summary statistics: 'Aprobados' (0), 'Inscriptos' (0), and 'Reinscriptos' (0). A blue 'Buscar' button is positioned below the filters. The interface is clean and modern, with a dark blue header and sidebar.

Si se selecciona más de una cohorte se mostrará en números y gráficos la información comparativa. Estos valores se totalizarán contemplando de no duplicar información y se

mostrarán gráficas comparativas por cohortes (en gráfico de barras) de determinados indicadores asociados a las asignaturas:

- **Comparativa por resultados de cursada de las asignaturas:** Se mostrará el total de inscriptos, reinscriptos, aprobados, desaprobados y ausentes, discriminado por cohorte y totalizado para todas las asignaturas.
- **Comparativa por exámenes finales de las asignaturas:** Se mostrará el total de exámenes rendidos, aprobados, desaprobados y ausentes, discriminado por cohorte y totalizado para todas las asignaturas.



Si se selecciona una carrera, un año de cohorte y una asignatura, se accede a un listado de estudiantes en donde se muestra información relacionada a la asignatura en cuestión junto información academia inherente a cada estudiante.

La lista de ESTUDIANTES para la asignatura

Mostrar 10 registros

Copy CSV Print

Buscar

Apellido	Nombre	DNI	Estado cursada	Estado del final	Múltiple carrera	Año primer inscr.	# Reinsc.	# Inscr. al final	# Ausen. al final	# Aplazos en final	Estado de inscrip año actual	# Ausen. reit.	# Ausen. a Asig.	# Ausen. múltiples	# Final Aprob	Datos de tutoría
Arzuaga	Nelson Miguel	36050646	No	No	NO	0	0	0	0	0		0	0	0	0	Ver
Aureli	Gonzalo Alberto	32345003	Si	No	NO	2010	0	0	0	0		0	0	0	5	Ver
Auza	Martin	34125679	No	No	NO	2011	0	0	0	0		0	0	0	1	Ver
Burga	Enzo Emanuel	35031512	Si	No	NO	2010	0	0	0	0		0	0	0	30	Ver
Campos	Leonel Angel	35099001	No	No	NO	2010	0	0	0	0		0	0	0	1	Ver
Canone	Martin Gabriel	34216754	No	No	NO	2010	0	0	0	0		0	0	0	43	Ver
Cardenas	Enrique Jesus	35085420	No	No	NO	0	0	0	0	0		0	0	0	0	Ver
Cascardo	Román Nahuel	34991246	No	No	NO	2010	0	0	0	0		0	0	0	0	Ver
Castañares	Lucas Alberto	33963398	No	Si	NO	2010	0	0	0	0		0	0	0	12	Ver
Cenacchi	Juan Manuel Cenacchi	34614915	No	No	NO	2010	0	0	0	0		0	0	0	1	Ver

La información que se mostrará del estudiante para una determinada asignatura de una determinada cohorte de una determinada carrera será:

- **Ficha general del estudiante:**
 - Apellido
 - Nombre.
 - DNI.

- **Traectoria académica en relación con las asignaturas:**
 - Estado actual de cursada (- se indica si tiene, o no, la cursada aprobada -).
 - Estado del final (- si indica si tiene, o no, el final aprobado -).
 - Múltiple inscripción. Permite conocer si está inscripto en más de una carrera.
 - Año de la primera inscripción a la asignatura.
 - Cantidad de reinscripciones a la asignatura.
 - Cantidad de inscripciones a final.
 - Cantidad de ausencias a finales
 - Cantidad de aplazos en final.

- Estado de inscripción en el corriente año (- se indica si está, o no, inscripto en la asignatura en el corriente año -).
- Cantidad de ausencias reiteradas por enfermedad/presentación de certificados.
- Cantidad de ausencias reiteradas en una asignatura (determinada por el número de inasistencias que permite el régimen académico).
- Ausencias múltiples en varias asignaturas que actualmente se encuentre inscripto.
- Cantidad de finales aprobados.

Junto con **la ficha general del estudiante** y **la trayectoria académica en relación con las asignaturas**, que se muestra del estudiante para una determinada cohorte de la carrera seleccionada, está la opción de acceder a la ficha con los datos de la tutoría, en la cual puede observarse información adicional al desarrollo académico, la cual puede obtenerse de la información que los estudiantes completan en su perfil de usuario en el sistema de SIU-Guaraní, o bien es provista por algún rol de tutoría que obtuvo la información bajo el consentimiento del propio estudiantes. La información que aquí se incorpora es:

- **Personales:**

- Sexo
- Edad al ingreso de la carrera.
- Lugar de procedencia.
- Lugar de residencia durante el período lectivo (En caso de no residir en sede universitaria, demanda horaria de viaje)
- Situación familiar al momento del ingreso:
 - a) Estado civil.
 - b) Cantidad de familiares a cargo.
 - c) Cantidad de convivientes.

- **Socioeconómicas:**

- Situación laboral:
 - a) Trabaja (media jornada o jornada completa)
 - b) Horas de trabajo semanal y año de ingreso laboral
 - c) No trabaja.

- Fuente de financiamiento de los estudios:
 - a) Aportes familiares.
 - b) Trabajo propio.
 - c) Becas.
 - d) Otros.

- **Académicas:**

- Trayectoria académica previa
- Tipo de secundario: Escuela Pública o privada.
- Título secundario obtenido.
- Orientación vocacional.
- Otros títulos superiores previos.
- Rendimiento académico en la carrera
- Año de ingreso.
- Año de primer cursada aprobada.
- Año del primer final rendido.
- Año de la última cursada aprobada.
- Año del último final rendido.
- Cantidad de cursadas aprobadas.
- Cantidad de finales aprobados.
- Porcentaje de avance en la carrera (- diferenciando en base a la cantidad de cursadas aprobadas y finales aprobados -) en relación al total de las materias según el plan de estudios.
- Promedio general de la carrera.
- Cantidad de asignaturas pendientes para cursar.
- Cantidad de asignaturas pendientes para rendir final.

Administrador

Santiago Airaldi

Información			
DNI	33096899	Año de ingreso	2010
Año de primer cursada	2010	Año de primer final rendido	2010
Año última cursada aprobada	2014	Año de último final rendido	2015
Cursadas aprobadas	29	Finales aprobados	29
Porcentaje de avance	63 %		
Cursadas pendientes de cursar	15	Asignaturas pendientes de final	0
Año de última cursada aprobada	2014		
Sexo	MASCULINO	Edad de ingreso	22
Procedencia		Residencia	
Financiamiento			
Aporte familiar		Trabajo propio	
Beca		Nombre	
Situación laboral			
Hs trabajo semanal	0	Comienzo a trabajar en cursada	
Trabajo en la cursada			
Situación Familiar			
Estado civil		Familiares a cargo	0
Cantidad convivientes	0		
Educación de los padres			
Múltiple Carrera			
Nivel máximo			
Detalle			

Volver

Modificar

La información asociada a la asistencia de los estudiantes no es posible de ser relevada en base a la información provista por el actual sistema del SIU-Guaraní en su versión 3.15.0, por lo que dicha información deberá ser anexada a la información que se cargue en el tablero y la misma será obtenida en base a trabajos de tutorías realizados sobre los estudiantes para un determinado grupo de asignaturas.

El Sistema de asistencia permite discriminar las ausencias y el tipo de elección que hace el estudiante. Distintos y variados trabajos han demostrado la estrecha vinculación entre el grado de asistencia a clase con el fenómeno del abandono, teniendo mayores probabilidades de permanecer aquellos estudiantes que asisten con regularidad (Braxton, Millem y Shaw, 2000; García, Alvarado y Jiménez, 2000; Cabrera, Bethencourt, González y Álvarez, 2006; Iñigo Ochoa de Eribe y Otxoa-Berrio, 2011). Por su parte, Casaravilla et al.

(2012) califican esta variable como predictora del abandono de titulación. En relación con lo anterior Rodríguez y Herrera (2009) y Santos y Vallerado (2013) constatan la relación de esta variable con el rendimiento académico en la universidad.

El seguimiento y visualización de la asistencia en el tablero de control al estandarizar la asistencia para todas las cátedras donde se ha inscripto el estudiante permitirá detectar en forma temprana los potenciales casos de deserción para poder aplicar tempranamente algún procedimiento preventivo.

ASIGNATURA:							
COMISIÓN N°			TEMA				
Día y Horario			MES				
Aula			FECHA				
LEGAJO N°	APELLLIDO Y NOMBRE	Identificador		CLASE 1	CLASE 2	CLASE 3	CLASE 4
		1					
		2					
		3					
		4					
		5					
		6					
		7					
		8					
		9					

En base a la información relevada de las inasistencias se pueden definir indicadores de alerta en base a las siguientes situaciones:

Situación	Tipo de alerta	Procedimiento
Ausencias reiteradas por enfermedad/presentación de certificados	PREVENCIÓN	Realizar seguimiento.

Ausencias por abandono de asignatura: Se determina por el número de inasistencia que permite el régimen académico. Ej: 80% de asistencia.	ALERTA 1	Se establece contacto con el alumno vía <i>online</i> para determinar las posibles causas.
Ausencias múltiples en varias asignaturas.	ALERTA 2	Se establece contacto con el alumno vía <i>online</i> , a través de entrevistas profundas, autoadministradas y semi-estructuradas.

Sección de información por Estudiante

En el tablero de control se detalla y visibiliza por cada estudiante una serie de indicadores relacionados con distintas variables: personales, socioeconómicas y académicas. En el caso de las dos primeras variables, se importarán los datos de las plantillas y formularios dispuestos en el sistema SUI-Guaraní. Dichos datos son completados por los estudiantes al momento de realizar la pre-inscripción a una carrera. Dado que esta información no es de carácter obligatorio, desde el tablero de control se permite realizar una actualización de la misma de forma de mantener dicha información actualizada. Para ello se deberá generar un archivo que contenga dicha información y el tablero posibilitará la carga de la misma mediante las opciones de actualización previstas en la sección de *Configuración*.

El filtro de búsqueda permite elegir una carrera y seleccionar, al menos, un año de cohorte.

Al igual que en la Sección de información de Cohorte, se permitirá seleccionar hasta 6 años de cohortes determinados por un periodo de años consecutivo.

Para seleccionar la asignatura se podrán aplicar, previamente, dos filtros alternativos:

1. Un filtro de área de evaluación de CONEAU (determinadas por la Resolución 786/09): *Ciencias Básicas, Teoría de la Computación, Algoritmos y Lenguajes, Arquitectura, Sistemas Operativos y Redes, Ingeniería de Software, Bases de Datos y Sistemas de Información y Aspectos Profesionales y Sociales*;
2. Un filtro por año de ubicación de la asignatura dentro de la carrera previamente seleccionada.

En caso de no aplicarse dichos filtros, se puede seleccionar directamente la asignatura.

Se adiciona la posibilidad de poder realizar una búsqueda de un determinado estudiante. La misma puede realizarse por: apellido, nombre o DNI.

Mostrar 10 registros

Copy CSV Print

Buscar

Apellido	Nombre	DNI	# Cursadas aprobadas	# Finales aprobados	Múltiple carrera	% de avance	Asignaturas pendientes de cursar	Asignaturas con finales pendientes	Año de última cursada aprobada	Año de último final aprobado	Datos de tutoría	Información de las asignaturas
La tabla no tiene datos cargados												

Mostrando registros del 0 al 0 de un total de 0 registros

Anterior Siguiente

Se puede seleccionar una carrera y luego ir seleccionando los años de las cohortes. Si se selecciona más de un año de cohorte, la información se irá adicionando a la tabla con la información sobre los estudiantes.

A su vez se puede aplicar un único criterio de búsqueda por filtro y se mostrará la información asociada a dicha búsqueda. Como podría ser el caso de realizar una búsqueda de todos los estudiantes inscriptos en una determinada carrera para una determinada cohorte.

Administrador

Analisis por Estudiante

Carreras:

Cohorte: 2010 2011 2012 2013 2014 2015

Areas:

Año:

Asignatura:

Apellido:

Nombre:

DNI:

Mostrar registros

Buscar

Apellido	Nombre	DNI	# Cursadas aprobadas	# Finales aprobados	Múltiple carrera	% de avance	Asignaturas pendientes de cursar	Asignaturas con finales pendientes	Año de última cursada aprobada	Año de último final aprobado	Datos de tutoria	Información de las asignaturas
Addante	Juan Ignacio	34540236	28	26	NO	56	18	2	2015	2015	<input type="button" value="Ver"/>	<input type="button" value="Asignaturas"/>
Agraso	Amilcar David	37379513	0	0	NO	0	44	0	2014	0	<input type="button" value="Ver"/>	<input type="button" value="Asignaturas"/>
Agüero	Albano Alejandro	33096616	41	39	NO	84	3	2	2016	2013	<input type="button" value="Ver"/>	<input type="button" value="Asignaturas"/>
Alvarez	Jonathan	35243804	1	1	NO	2	43	0	2014	2014	<input type="button" value="Ver"/>	<input type="button" value="Asignaturas"/>
Amigó	Matias	31918448	0	0	NO	0	44	0	0	0	<input type="button" value="Ver"/>	<input type="button" value="Asignaturas"/>
Antolin	Nicolas Emmanuel	35342647	7	7	NO	15	37	0	2014	2014	<input type="button" value="Ver"/>	<input type="button" value="Asignaturas"/>
Barra	Gaston Mariano	32150110	33	33	NO	71	11	0	2013	2013	<input type="button" value="Ver"/>	<input type="button" value="Asignaturas"/>
Benzi	Nestor Eduardo	26928610	0	0	NO	0	44	0	2012	0	<input type="button" value="Ver"/>	<input type="button" value="Asignaturas"/>
Bergameschino	Alexis	35892761	0	0	NO	0	44	0	0	0	<input type="button" value="Ver"/>	<input type="button" value="Asignaturas"/>
Bergonzi	Mariana	35505192	0	0	NO	0	44	0	2012	0	<input type="button" value="Ver"/>	<input type="button" value="Asignaturas"/>

Mostrando registros del 1 al 10 de un total de 51 registros

Los indicadores establecidos trabajan con un nivel de alerta en tres estados: rojo, amarillo, verde. Para cada indicador se definen valores que conforman el segmento de los niveles, indicando el color rojo cuando está por debajo del intervalo mínimo establecida, amarillo para cuando está dentro del intervalo y verde para cuando está por encima del intervalo máximo establecido.

Estos valores se pueden configurar desde el propio tablero según la necesidad de alerta que se necesite visualizar. A su vez, las columnas permiten ordenarse de forma ascendente o descendente, permitiendo una mejor visualización de la información a analizar.

La información que se mostrará del estudiante para una determinada asignatura de una determinada cohorte de una determinada carrera será:

- **Ficha general del estudiante:**
 - Apellido
 - Nombre.
 - DNI.

- **Trayectoria académica:**
 - Cantidad de cursadas aprobadas
 - Cantidad de finales aprobados
 - Múltiple inscripción. Permite conocer si está inscripto en más de una carrera.
 - Cantidad de asignaturas pendientes para cursar.
 - Cantidad de asignaturas pendientes para rendir final.
 - Año de última cursada aprobada.
 - Año del último final rendido (independientemente del resultado obtenido).

Junto con la **Ficha general del estudiante** y la **Trayectoria** que se muestra del estudiante para una determinada cohorte de la carrera seleccionada, está la opción de acceder a la ficha con los datos de la tutoría, en la cual puede observarse información adicional al desarrollo académico, la cual puede obtenerse de la información que los estudiantes completan en su perfil de usuario en el sistema de SIU-Guaraní, o bien es provista por algún rol de tutoría que obtuvo la información bajo el consentimiento del propio estudiantes. La información que aquí se incorpora es:

- **Personales:**
 - Sexo
 - Edad al ingreso de la carrera.
 - Lugar de procedencia.

- Lugar de residencia durante el período lectivo (En caso de no residir en sede universitaria, demanda horaria de viaje)
- Situación familiar al momento del ingreso:
 - a) Estado civil.
 - b) Cantidad de familiares a cargo.
 - c) Cantidad de convivientes.

- **Socioeconómicas:**
 - Situación laboral:
 - a) Trabaja (media jornada o jornada completa)
 - b) Horas de trabajo semanal y año de ingreso laboral
 - c) No trabaja.
 - Fuente de financiamiento de los estudios:
 - a) Aportes familiares.
 - b) Trabajo propio.
 - c) Becas.
 - d) Otros.

- **Académicas:**
 - Trayectoria académica previa
 - Tipo de secundario: Escuela Pública o privada.
 - Título secundario obtenido.
 - Orientación vocacional.
 - Otros títulos superiores previos.
 - Rendimiento académico en la carrera
 - Año de ingreso.
 - Año de primer cursada aprobada.
 - Año del primer final rendido.
 - Año de la última cursada aprobada.
 - Año del último final rendido.
 - Cantidad de cursadas aprobadas.
 - Cantidad de finales aprobados.

- Porcentaje de avance en la carrera (- diferenciando en base a la cantidad de cursadas aprobadas y finales aprobados -) en relación al total de las materias según el plan de estudios.
- Promedio general de la carrera.
- Cantidad de asignaturas pendientes para cursar.
- Cantidad de asignaturas pendientes para rendir final.

The screenshot shows a web application interface for a student named Juan Ignacio Addante. The interface includes a navigation sidebar on the left with icons for Inicio, Análisis, Predicciones, Configuración, and Ayuda. The main content area displays the student's name and a table of information.

Información			
DNI	34540236	Año de ingreso	2012
Año de primer cursada	2010	Año de primer final rendido	2010
Año última cursada aprobada	2015	Año de último final rendido	2015
Cursadas aprobadas	28	Finales aprobados	26
Porcentaje de avance	56 %		
Cursadas pendientes de cursar	16	Asignaturas pendientes de final	2
Año de última cursada aprobada	2015		
Sexo	MASCULINO	Edad de ingreso	22
Procedencia		Residencia	
Financiamiento			
Aporte familiar		Trabajo propio	
Beca		Nombre	
Situación laboral			
Hs trabajo semanal	0	Comienzo a trabajar en cursada	
Trabajo en la cursada			
Situación Familiar			
Estado civil		Familiares a cargo	0
Cantidad convivientes	0		
Educación de los padres			
Múltiple Carrera			
Nivel máximo			
Detalle			

At the bottom of the table, there are two buttons: "Volver" and "Modificar".

A su vez se permite visualizar información de las asignaturas asociadas a un estudiante, detallando:

- Lista de Cursadas:

- Carrera
 - Asignatura
 - Fecha de regularidad
 - Resultado
 - Forma de regularidad
-
- Lista de finales
 - Carrera
 - Asignatura
 - Fecha
 - Resultado
 - Nota
 - Forma de aprobación

Y para una determinada carrera, se permite visualizar:

- La cantidad de cursadas aprobadas sobre el total de materias de la carrera
- La cantidad de finales aprobados sobre el total de materias de la carrera
- La cantidad de cursadas que adeuda regularizar en base a la cantidad de materias de la carrera.

Asignaturas de Juan Ignacio Addante

La lista de CURSADAS

Carrera	Asignatura	Fecha de regularidad	Resultado	Forma de regularidad
Licenciatura en Sistemas	Inglés Técnico Nivel I y II	2/2012	Libre	Cursada Equiv.
Licenciatura en Sistemas	Análisis y Diseño de Sistemas I	11/2014	Aprobado	Cursada Equiv.
Licenciatura en Sistemas	Análisis y Diseño de Sistemas I	2/2012	Aprobado	Cursada Equiv.
Licenciatura en Sistemas	Análisis y Diseño de Sistemas II	2/2012	Aprobado	Cursada Equiv.
Licenciatura en Sistemas	Bases de Datos I	2/2012	Aprobado	Cursada
Licenciatura en Sistemas	Bases de Datos II	19/2012	Aprobado	Cursada
Licenciatura en Sistemas	Práctica Profesional y Legislación	2/2012	Aprobado	Cursada Equiv.
Licenciatura en Sistemas	Probabilidades y Estadística	2/2012	Aprobado	Cursada Equiv.
Licenciatura en Sistemas	Programación Orientada a Objetos	2/2012	Reprobado	Cursada Equiv.
Licenciatura en Sistemas	Programación Orientada a Objetos	2/11/2015	Reprobado	Cursada
Licenciatura en Sistemas	Programación Orientada a Objetos	2/2012	Aprobado	Cursada Equiv.
Licenciatura en Sistemas	Sistemas Operativos II	2/2012	Aprobado	Cursada Equiv.
Licenciatura en Sistemas	Sistemas Operativos II	2/2012	Reprobado	Cursada Equiv.
Licenciatura en Sistemas	Antropología de la Comunicación II	2/2012	Libre	Cursada Equiv.
Licenciatura en Sistemas	Antropología de la Comunicación II	19/2012	Aprobado	Cursada
Licenciatura en Sistemas	Gestión de Proyectos	2/2012	Aprobado	Cursada Equiv.
Licenciatura en Sistemas	Lenguajes de Programación	20/12/2012	Aprobado	Cursada Equiv.
Licenciatura en Sistemas	Redes II	2/2012	Aprobado	Cursada Equiv.
Licenciatura en Sistemas	Sistemas de Información	4/12/2012	Aprobado	Cursada Equiv.
Licenciatura en Sistemas	Sistemas de Información	2/2012	Libre	Cursada Equiv.
Licenciatura en Sistemas	Programación para celulares	6/2012	Libre	Cursada
Licenciatura en Sistemas	Bases de Datos I	11/2015	Aprobado	Cursada
Licenciatura en Sistemas	Programación Lógica y Funcional	2/2012	Libre	Cursada Equiv.
Licenciatura en Sistemas	Programación Lógica y Funcional	2/2012	Reprobado	Cursada Equiv.
Licenciatura en Sistemas	Sistemas Administrativos I	2/19/2012	Aprobado	Cursada
Licenciatura en Sistemas	Sistemas Administrativos II	3/12/2012	Aprobado	Cursada
Licenciatura en Sistemas	Redes I	2/2012	Aprobado	Cursada Equiv.

La lista de FINALES

Carrera	Asignatura	Fecha	Resultado	Nota	Forma de aprobación
Licenciatura en Sistemas	Álgebra y Geometría Analítica	2/2012	Aprobado	7	Equivalencia Equiv.
Licenciatura en Sistemas	Análisis Matemático Aplicado	2/2012	Aprobado	6	Equivalencia Equiv.
Licenciatura en Sistemas	Análisis Matemático Básico	2/2012	Aprobado	6	Equivalencia Equiv.
Licenciatura en Sistemas	Arquitectura I	2/2012	Aprobado	9	Examen Equiv.
Licenciatura en Sistemas	Arquitectura II	2/2012	Aprobado	4	Examen Equiv.
Licenciatura en Sistemas	Ciencias de la Computación I	2/2012	Aprobado	8	Examen Equiv.
Licenciatura en Sistemas	Introducción a la Programación Imperativa	2/2012	Aprobado	7	Equivalencia Equiv.
Licenciatura en Sistemas	Introducción al Álgebra	2/2012	Aprobado	7	Equivalencia Equiv.
Licenciatura en Sistemas	Matemática Discreta	2/2012	Aprobado	8	Examen Equiv.
Licenciatura en Sistemas	Programación Imperativa	2/2012	Aprobado	4	Examen Equiv.
Licenciatura en Sistemas	Sistemas y Organizaciones	2/2012	Aprobado	4	Examen Equiv.
Licenciatura en Sistemas	Ciencias de la Computación II	2/2012	Aprobado	9	Examen Equiv.
Licenciatura en Sistemas	Ciencias de la Computación II	2/2012	Reprobado	2	Examen Equiv.
Licenciatura en Sistemas	Comunicación de Datos	2/2012	Aprobado	7	Examen Equiv.
Licenciatura en Sistemas	Estructura de Datos	2/2012	Aprobado	6	Equivalencia Equiv.
Licenciatura en Sistemas	Inglés Técnico Nivel I y II	3/2013	Aprobado	9	Examen
Licenciatura en Sistemas	Introducción a la Programación Orientada a Objetos	2/2012	Aprobado	7	Equivalencia Equiv.
Licenciatura en Sistemas	Introducción a Objetos	2/2012	Aprobado	8	Equivalencia Equiv.
Licenciatura en Sistemas	Sistemas Operativos I	15/2012	Aprobado	8	Examen Equiv.
Licenciatura en Sistemas	Análisis y Diseño de Sistemas I	24/2015	Aprobado	9	Examen
Licenciatura en Sistemas	Análisis y Diseño de Sistemas I	3/11/2014	Ausente	0	Examen
Licenciatura en Sistemas	Análisis y Diseño de Sistemas II	15/2015	Reprobado	2	Examen
Licenciatura en Sistemas	Práctica Profesional y Legislación	2/2012	Aprobado	8	Examen Equiv.
Licenciatura en Sistemas	Probabilidades y Estadística	19/2015	Reprobado	2	Examen
Licenciatura en Sistemas	Sistemas Operativos II	10/2015	Aprobado	8	Examen
Licenciatura en Sistemas	Sistemas Operativos II	25/2014	Ausente	0	Examen
Licenciatura en Sistemas	Antropología de la Comunicación I	2/2012	Aprobado	4	Equivalencia Equiv.
Licenciatura en Sistemas	Redes II	20/2014	Aprobado	8	Examen
Licenciatura en Sistemas	Sistemas de Información	19/2015	Ausente	0	Examen
Licenciatura en Sistemas	Sistemas Administrativos I	2/2012	Aprobado	8	Examen
Licenciatura en Sistemas	Sistemas Administrativos II	2/2015	Aprobado	9	Examen
Licenciatura en Sistemas	Redes I	2/2012	Aprobado	7	Examen Equiv.
Ingeniería en Informática	Ciencias de la Computación II	12/2011	Aprobado	9	Examen
Ingeniería en Informática	Ciencias de la Computación II	19/2011	Reprobado	2	Examen
Ingeniería en Informática	Inglés Técnico Nivel I y II	3/2013	Aprobado	9	Examen Equiv.
Ingeniería en Informática	Introducción a la Programación Orientada a Objetos	7/8/2011	Aprobado	7	Equivalencia
Ingeniería en Informática	Introducción a Objetos	4/8/2011	Aprobado	8	Equivalencia
Ingeniería en Informática	Metodologías de Programación I	17/12/2010	Reprobado	2	Examen
Ingeniería en Informática	Metodologías de Programación I	10/2011	Aprobado	8	Examen
Ingeniería en Informática	Metodologías de Programación II	13/2011	Aprobado	7	Examen
Ingeniería en Informática	Sistemas Operativos I	15/2012	Aprobado	8	Examen
Ingeniería en Informática	Análisis y Diseño de Sistemas I	24/2015	Aprobado	9	Examen Equiv.
Ingeniería en Informática	Análisis y Diseño de Sistemas II	15/2015	Reprobado	2	Examen Equiv.
Ingeniería en Informática	Práctica Profesional y Legislación	15/2011	Aprobado	8	Examen
Ingeniería en Informática	Probabilidades y Estadística	19/2015	Reprobado	2	Examen Equiv.
Ingeniería en Informática	Sistemas Operativos II	10/2015	Aprobado	8	Examen Equiv.
Ingeniería en Informática	Antropología de la Comunicación I	25/2011	Aprobado	4	Equivalencia
Ingeniería en Informática	Redes II	20/2014	Aprobado	8	Examen Equiv.
Ingeniería en Informática	Sistemas de Representación	20/12/2010	Aprobado	9	Examen
Ingeniería en Informática	Redes I	10/12/2010	Aprobado	7	Examen

La cantidad de cursadas aprobadas es 18 sobre un total de 44 materias de la carrera

La cantidad de finales aprobados es 40 sobre un total de 44 materias de la carrera

La cantidad de cursadas que adeuda es 26

Sección de Predicciones

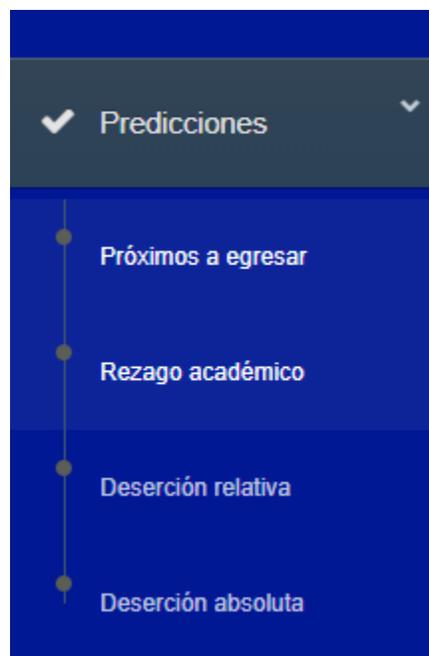
El tablero de control permite desarrollar y visualizar el entrecruzamiento entre los indicadores previamente definidos. La finalidad de las predicciones es que sean utilizadas a nivel de gestión institucional con el fin de monitorear la trayectoria académica de los estudiantes de las carreras de informática para poder identificar predictivamente, a través de las alertas tempranas, factores de riesgo que puedan dificultar la permanencia del estudiante en la carrera o determinar la deserción. Por ello se incorpora la predicción del rendimiento académico de los estudiantes fundamentada en un seguimiento de sus trayectorias académicas.

Es importante aclarar y resaltar que esto constituye una herramienta complementaria basada estrictamente en datos estadísticos, surgidos del análisis retrospectivo y longitudinal del período 2005-2015, y en la experiencia del área académica de la UNNOBA, que no es estrictamente un sistema inteligente de Minería de Datos, sino que mediante el procesamiento de la información (recolección, extracción, almacenamiento, análisis y estadísticas) se genera un “Sistema Informático de Apoyo a Decisiones” para brindar posibilidades a los actores que toman decisiones dentro del área académica.

El tablero permite obtener cuatro tipos de predicciones en base a la información que se le suministra:

- Próximos a egresar: Toma sus bases en campos basados en el año de ingreso, el porcentaje de avance y los registros de la última cursada y final aprobado. Estos datos se cotejan contra el estudio estadístico realizado sobre los datos nominales de los estudiantes de la Universidad en relación a la esperanza del egreso efectivo.
- Rezago académico: Responde a la necesidad de poder visualizar aquellos estudiantes que cuentan con una regularidad activa pero no tienen actividad académica en un tiempo que excede los dos años.
- Deserción relativa: Responde a la necesidad de indicar en base al análisis de datos académicos de cada estudiante, una potencial deserción, como antelación ante el potencial abandono de carrera.

- Deserción absoluta: Responde a la necesidad de relevar lo que puede considerarse como la deserción efectiva del estudiante una vez que se incurre en el abandono de la carrera. Se busca analizar el acontecimiento del abandono para poder fortalecer la predicción de la potencial deserción relativa. Puede utilizarse la funcionalidad de los datos de la tutoría para enriquecer la información de los estudiantes enmarcados en el presente listado a fines de trabajar sobre nuevos datos que pueden tener inferencia en el modelo predictivo y pueden no responder a lo estrictamente académico; como lo pueden ser los contextos socio ambientales de los estudiantes, sus vinculaciones familiares, situación laboral, motivaciones, metas y objetivos del estudiante en relación a la carrera elegida.



La información que se mostrará del estudiante, indistintamente de la predicción seleccionada, será:

- Apellido y Nombre
- DNI
- Año de ingreso
- Edad
- Cantidad de cursadas aprobadas
- Cantidad de finales aprobados
- Múltiple inscripción. Permite conocer si está inscripto en más de una carrera.

- Porcentaje de avance en la carrera (- diferenciando en base a la cantidad de cursadas aprobadas y finales aprobados -) en relación al total de las materias según el plan de estudios.
- Cantidad de asignaturas pendientes para cursar.
- Cantidad de asignaturas pendientes para rendir final.
- Año de última cursada aprobada.
- Año del último final rendido (independientemente del resultado obtenido).

Administrador

Predicción de próximos estudiantes a egresar

La predicción toma sus bases en campos basados en el año de ingreso, el porcentaje de avance y los registros de la última cursada y final aprobado. Estos datos se cotejan contra el estudio estadístico realizado sobre los datos nominales de los estudiantes de la Universidad en relación a la esperanza del egreso efectivo.

Mostrar registros

Buscar

Apellido y Nombre	DNI	Año ingreso	Edad	# Cursadas aprobadas	# Finales aprobados	% de avance	Cursadas pendientes	Finales pendientes	Año de última cursada aprobada	Año de último final aprobado	Datos de tutoría	Información de las asignaturas
Agüero, Albano Alejandro	33096616	2012	31	41	39	84	3	2	2015	2013	Ver	Asignaturas
Basile, Matias	34240879	2010	29	43	43	93	1	0	2011	2014	Ver	Asignaturas
Belles, Ignacio	36278565	2011	26	48	40	83	-3	8	2015	2016	Ver	Asignaturas
Benítez, Danilo Ruben	33521458	2010	30	43	43	93	1	0	2014	2016	Ver	Asignaturas
Biondini, Victor Herman	30344290	2010	36	44	43	91	0	1	2015	2015	Ver	Asignaturas
Blanco, Matias	33096975	2010	30	44	41	89	0	3	2015	2015	Ver	Asignaturas
Calcaterra, Martín	35167850	2011	28	47	45	93	-2	2	2015	2016	Ver	Asignaturas
Canone, Martín Gabriel	34216754	2010	29	44	43	91	0	1	2015	2015	Ver	Asignaturas
Capelli, Matias	35893582	2011	27	45	39	84	-1	6	2015	2016	Ver	Asignaturas
Carea, Mariano Daniel	29416234	2010	36	43	43	91	1	0	2012	2013	Ver	Asignaturas

Mostrando registros del 1 al 10 de un total de 45 registros

Anterior 1 2 3 4 5 Siguiente

Los indicadores establecidos trabajan con un nivel de alerta en tres estados: rojo, amarillo, verde. Para cada indicador se definen valores que conforman el segmento de los niveles, indicando el color rojo cuando está por debajo del intervalo mínimo establecida,

amarillo para cuando está dentro del intervalo y verde para cuando está por encima del intervalo máximo establecido.

Estos valores se pueden configurar desde el propio tablero según la necesidad de alerta que se necesite visualizar. A su vez, las columnas permiten ordenarse de forma ascendente o descendente, permitiendo una mejor visualización de la información a analizar.

Dentro de la visualización de la predicción existe una verificación que puede aplicarse. Para la actual aplicación el proceso de verificación consiste en tomar los datos de las predicciones en base a la información de los años 2010 a 2015 constatada contra la información de los años 2016 y 2017.

Predicción de próximos estudiantes a egresar

La predicción toma sus bases en campos basados en el año de ingreso, el porcentaje de avance y los registros de la última cursada y final aprobado. Estos datos se cotejan contra el estudio estadístico realizado sobre los datos nominales de los estudiantes de la Universidad en relación a la esperanza del egreso efectivo.

Validar Predicción

La lista de ESTUDIANTES incluidos en la predicción:

Mostrar 10 registros Copy CSV Print Buscar

Apellido y Nombre	DNI	Año Ingreso	Edad	# Cursadas aprobadas	# Finales aprobados	% de avance	Cursadas pendientes	Finales pendientes	Año de última cursada aprobada	Año de último final aprobado	Datos de tutoría	Información de las asignaturas
Aguero, Albano Alejandro	33096616	2012	31	41	39	84	3	2	2015	2013	Ver	Asignaturas
Basile, Matias	34240879	2010	29	43	43	93	1	0	2011	2014	Ver	Asignaturas
Belles, Ignacio	36278565	2011	26	48	40	83	3	0	2015	2016	Ver	Asignaturas
Benitez, Danilo Ruben	33521458	2010	30	43	43	93	1	0	2014	2016	Ver	Asignaturas
Biondini, Victor Hernan	30344290	2010	36	44	43	91	0	1	2015	2015	Ver	Asignaturas

El acierto de la verificación en la predicción es indicado con una tilde, caso contrario se muestra una cruz; y a su vez se indica un porcentaje de acierto de verificación respecto a la constatación de la predicción generada.

Blanco, Matias	33096975	2010	30	44	41	89	0	3	2015	2015	Ver
Calcaterra, Martin	35167850	2011	28	47	45	93	-2	2	2015	2016	Ver
Canone, Martin Gabriel	34216754	2010	29	44	43	91	0	1	2015	2015	Ver
Capelli, Matias	35893582	2011	27	45	39	84	-1	6	2015	2016	Ver
Carea, Mariano Daniel	29416234	2010	36	43	43	91	1	0	2012	2013	Ver

Mostrando registros del 1 al 10 de un total de 12 registros

Anterior 1 2

Total: 12 estudiantes

Junto con **la ficha general del estudiante** y la **trayectoria**, que se muestra del estudiante para una determinada cohorte de la carrera seleccionada, está la opción de acceder a la ficha con los datos de la tutoría, en la cual puede observarse información adicional al desarrollo académico, la cual puede obtenerse de la información que los estudiantes completan en su perfil de usuario en el sistema de SIU-Guaraní, o bien es provista por algún rol de tutoría que obtuvo la información bajo el consentimiento del propio estudiantes. La información que aquí se incorpora es:

- **Personales:**

- Sexo
- Edad al ingreso de la carrera.
- Lugar de procedencia.
- Lugar de residencia durante el período lectivo (En caso de no residir en sede universitaria, demanda horaria de viaje)
- Situación familiar al momento del ingreso:
 - a) Estado civil.
 - b) Cantidad de familiares a cargo.
 - c) Cantidad de convivientes.

- **Socioeconómicas:**

- Situación laboral:
 - a) Trabaja (media jornada o jornada completa)
 - b) Horas de trabajo semanal y año de ingreso laboral
 - c) No trabaja.
- Fuente de financiamiento de los estudios:
 - a) Aportes familiares.
 - b) Trabajo propio.
 - c) Becas.
 - d) Otros.

- **Académicas:**
 - Trayectoria académica previa
 - Tipo de secundario: Escuela Pública o privada.
 - Título secundario obtenido.
 - Orientación vocacional.
 - Otros títulos superiores previos.
 - Rendimiento académico en la carrera
 - Año de ingreso.
 - Año de primer cursada aprobada.
 - Año del primer final rendido.
 - Año de la última cursada aprobada.
 - Año del último final rendido.
 - Cantidad de cursadas aprobadas.
 - Cantidad de finales aprobados.
 - Porcentaje de avance en la carrera (- diferenciando en base a la cantidad de cursadas aprobadas y finales aprobados -) en relación al total de las materias según el plan de estudios.
 - Promedio general de la carrera.
 - Cantidad de asignaturas pendientes para cursar.
 - Cantidad de asignaturas pendientes para rendir final.

Administrador

Federico Martin Gomez

Información			
DNI	32363716	Año de ingreso	2010
Año de primer cursada	2010	Año de primer final rendido	2010
Año última cursada aprobada	2014	Año de último final rendido	2015
Cursadas aprobadas	42	Finales aprobados	42
Porcentaje de avance	81 %		
Cursadas pendientes de cursar	3	Asignaturas pendientes de final	0
Año de última cursada aprobada	2014		
Sexo	MASCULINO	Edad de ingreso	23
Procedencia		Residencia	
Financiamiento			
Aporte familiar		Trabajo propio	
Beca		Nombre	
Situación laboral			
Hs trabajo semanal	0	Comienzo a trabajar en cursada	
Trabajo en la cursada			
Situación Familiar			
Estado civil		Familiares a cargo	0
Cantidad convivientes	0		
Educación de los padres			
Múltiple Carrera			
Nivel máximo			
Detalle			

Volver

Modificar

A su vez se permite visualizar información de las asignaturas asociadas a un estudiante, detallando:

- Lista de Cursadas:
 - Carrera
 - Asignatura
 - Fecha de regularidad
 - Resultado
 - Forma de regularidad
- Lista de finales
 - Carrera

- Asignatura
- Fecha
- Resultado
- Nota
- Forma de aprobación

Y para una determinada carrera, se permite visualizar:

- La cantidad de cursadas aprobadas sobre el total de materias de la carrera
- La cantidad de finales aprobados sobre el total de materias de la carrera
- La cantidad de cursadas que adeuda regularizar en base a la cantidad de materias de la carrera.

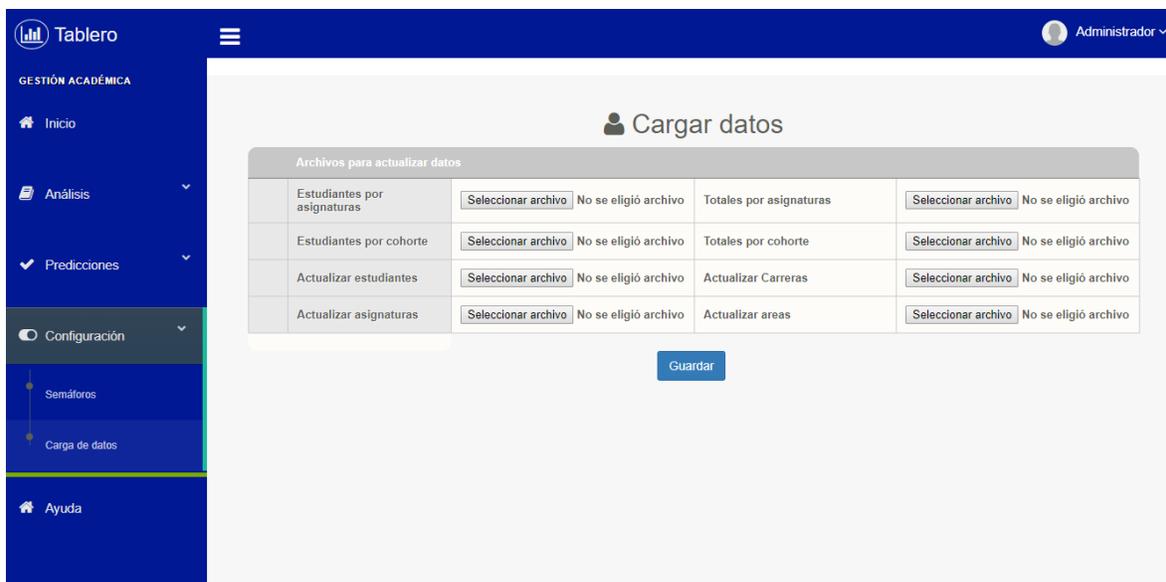
Carga de datos

Desde la configuración del panel se puede realizar el CRUD (*Create, Read, Update and Delete*) de las Carreras, Cohortes, Asignaturas y Estudiantes.

A su vez se permite, mediante un analizador sintáctico (- *parser* -), la carga masiva de datos para las carreras, cohortes, asignaturas y estudiantes. Los datos son suministrados al panel mediante importación de archivos CSV (*comma-separated values*), Estos archivos se obtienen del software PENTAHO, el cual toma los datos de la base de datos del sistema SIU-Guaraní. Este sistema registra y administra todas las actividades académicas de la universidad desde el ingreso hasta la titulación de cada estudiante, permitiendo identificar el historial académico de cada uno de ellos.

En la sección de *Configuración - Carga de datos*, aparecen las diferentes opciones de actualización. Estas son:

- Actualizar estudiantes
- Actualizar asignaturas
- Actualizar Carreras
- Actualizar áreas
- Estudiantes por asignaturas
- Estudiantes por cohorte
- Totales por asignaturas
- Totales por cohorte
- Verificación de Predicciones



La conexión entre el software PENTAHO y el sistema SIU-Guaraní ya se encuentra previamente definida en el ámbito de la Universidad.

Paginación, búsqueda y exportación de resultados

En las diferentes tablas de información que se muestran en el tablero de control se brinda la posibilidad de poder realizar un paginado de la información de forma que la misma responda a un orden de visualización. A su vez se permite realizar una búsqueda general sin criterio definido. Cuando se visualicen las tablas, ya sea porque se realizó un análisis de información o una consulta de predicción, se muestra un cuadro de búsqueda en donde se puede buscar por cualquier texto o año, y se irán automáticamente filtrando los estudiantes que respondan a dicha búsqueda.

El tablero también ofrece la posibilidad de copiar, exportar e imprimir la información que se está visualizando. La exportación se realiza sobre un archivo con formato CSV (*comma-separated values*).

La lista de ESTUDIANTES incluidos en la predicción:

Mostrar registros

Buscar

Apellido y Nombre	DNI	Año ingreso	Edad	# Cursadas aprobadas	# Finales aprobados	% de avance	Cursadas pendientes	Finales pendientes	Año de última cursada aprobada	Año de último final aprobado	Datos de tutoría	Información de las asignaturas
Agüero, Albano Alejandro	33096616	2012	31	41	39	84	3	2	2015	2013	<input type="button" value="Ver"/>	<input type="button" value="Asignaturas"/>
Basile, Matias	34240879	2010	29	43	43	93	1	0	2011	2014	<input type="button" value="Ver"/>	<input type="button" value="Asignaturas"/>

Usuarios de sistema

El tablero permite operar con diferentes niveles de acceso que pueden configurarse según las necesidades que se requieran. Actualmente se cuenta con un nivel de acceso de gestión académica que tiene un total control sobre las secciones de análisis y predicciones y un nivel de acceso tutorial que permite realizar análisis sobre los estudiantes y modificar la ficha del registro de tutoría.

Configuración de indicadores

Los indicadores establecidos trabajan con un nivel de alerta en tres estados: rojo, amarillo, verde. Para cada indicador se definen valores que conforman el segmento de los niveles. El panel de control permite la configuración de alerta de los indicadores previamente definidos. Estos indicadores permiten dos posibles valores correspondientes a los intervalos de mínima y máxima. Todo valor que esté por debajo del mínimo se muestra en rojo, los valores del intervalo en amarillo, y por encima del valor máximo en verde. En cada indicador se especifica la funcionalidad del mismo y a qué sección de información está asociado.



Inicio



Análisis



Predicciones



Configuración



Ayuda

Valores de semáforo

Datos					
Pendientes de cursar	Cota mínima	<input type="text" value="1"/>	Cota Máxima	<input type="text" value="2"/>	
Finales pendientes	Cota mínima	<input type="text" value="3"/>	Cota Máxima	<input type="text" value="4"/>	
Año ult cursada aprobada	Cota mínima	<input type="text" value="3"/>	Cota Máxima	<input type="text" value="5"/>	
Año ult final aprobado	Cota mínima	<input type="text" value="7"/>	Cota Máxima	<input type="text" value="20"/>	

Guardar

CONCLUSIONES Y APORTES

En la actualidad el sector Software y Servicios Informáticos (SSI) es uno de los principales demandantes de empleo calificado. No obstante ello, la matrícula de las carreras universitarias no acompaña esta demanda. En Argentina, al inicio de cada ciclo universitario se anotan unos 20 mil alumnos, pero egresa sólo el 20%. Esta situación pone en evidencia dos problemáticas: el rezago académico y la deserción estudiantil.

Desde el campo de la Ciencia Informática aplicada en Educación, esta tesis ha abordado el estudio de ambas problemáticas en las carreras que conforman el área de Informática de la Escuela de Tecnología (ET) de la UNNOBA. Si bien, en este caso, durante el período 2005-2015 el número de ingresantes y alumnos totales de estas carreras se ha triplicado, la tasa de egreso ha sido sumamente baja lo cual refleja los altos índices de desgranamiento y deserción que, en conjunto, implican un 87%. En correlación con estudios existentes que toman como referencia otras universidades, se ha demostrado que los mayores índices de deserción se presentan durante los dos primeros años de la carrera, especialmente durante el transcurso del segundo año.

Uno de los objetivos específicos de esta investigación radicó en poder distinguir los casos de rezago académico - aquellos estudiantes que cuentan con una regularidad activa pero no tienen actividad académica en un tiempo que excede los dos años - y deserción interna - cambio de carrera - , de aquellos casos en los cuales la deserción implicó la salida de la universidad o del sistema educativo, determinando en términos estadísticos los indicadores y causales de los mismos.

Con el objetivo de afrontar y dar respuestas a estas problemáticas, se ha diseñado un tablero de control que, basado en la utilización de herramientas de minería de datos, permita a nivel de gestión institucional realizar un seguimiento de la trayectoria académica de los estudiantes de las carreras de informática de la UNNOBA y así identificar predictivamente, a través de un sistema de alertas tempranas, factores de riesgo. Para poder desarrollar el tablero de control se requirió previamente que la Universidad contara con el sistema SIU-Guaraní en su versión 3.14.0 y con el software Pentaho.

El software Pentaho toma la base provista por el sistema SIU-Guaraní y genera los cruces de información que son una de la fuente de entrada del tablero de control.

Otra de las fuentes de entrada lo son el registro de asistencias. En base a este punto, una de las problemáticas con las que actualmente se cuenta es que la versión del sistema SIU-Guaraní no cuenta aún con el registro de asistencia. Siendo este uno de los indicadores utilizados por el tablero de control para realizar los análisis y predicciones; por tal se tuvo que estandarizar la asistencia mediante una sistematización de una planilla de registro de asistencia por cátedra, la cual es obtenida en base a trabajos de tutorías realizados sobre los estudiantes para un determinado grupo de cátedras. La información resultante es anexada como como información de entrada del tablero de control. En base estas fuentes de entrada, el tablero la ordena en una sección de Analisis y Predicción con el fin de responder a las necesidades que la gestión académica busca realizar en cuestiones de seguimiento de los estudiantes.

El desarrollo del tablero requirió de la implementación secuenciada y consecutiva de dos tipos de análisis: en primer lugar, un análisis retrospectivo del período 2005-2015 y luego, sustentado y fundamentado en él, un análisis predictivo.

En lo concerniente al análisis retrospectivo, tomando como insumo la información suministrada por el sistema SIU-Guaraní de la UNNOBA se construyó una base de datos sobre la situación académica global de cada una de las cohortes entre los años 2005 y 2015. A modo de complemento, y para obtener información de índole nominal y cualitativa, se llevaron a cabo entrevistas semi –estructuradas tomando como muestra poblacional a un 30% del total de desertores del período estudiado. A partir de la triangulación de los datos obtenidos se elaboraron **indicadores educativos** de acuerdo a la siguiente clasificación: a) personal; b) familiar; c) económico-social; d) académico. La ponderación de los resultados de cada indicador y, en un segundo momento, el entrecruzamiento de los mismos, permitió crear el perfil de alumno de las carreras de informática e identificar los principales factores de riesgo de rezago y deserción académica.

En lo concerniente al primero de los indicadores, de tipo personal, se pudo concluir que del total de ingresantes entre 2005 y 2015, el 82.46% fueron varones y el 17.54% restante mujeres. Esta disparidad se acrecienta en el ingreso a Ingeniería en Informática, en la cual la inscripción masculina supera el 90%. En cuanto a la edad de ingreso a las carreras

informáticas el 48% de los estudiantes lo hace luego de terminar sus estudios secundarios, entre los 18 y los 19 años de edad. Los indicadores de tipo familiar y económico – social dan cuenta de un alto porcentaje de población estudiantil que, si bien no es sostén de familia, se halla inserta en el mercado laboral mientras cursa la carrera. Del 57% de estudiantes que trabajan, más de la mitad lo hace en jornada completa. Esto se vuelve un factor explicativo de los casos de rezago y deserción ya que mayoritariamente implica no poder cumplir con el mínimo de presentimo y regularidad en muchas materias. Vinculado con este indicador, se presentan otros de tipo académico.

Uno de los indicadores claves para poder dar cuenta de la deserción y el rezago ha sido el rendimiento académico en la carrera universitaria. Este indicador se creó a partir de la evaluación y el seguimiento del desempeño del alumnado por cohortes en las distintas materias, tanto en las cursadas como en la acreditación final de las mismas, sobre todo en aquellas que presentan mayores índices de desaprobación. Dado que los principales índices de abandono se presentan en los primeros dos años de la carrera, se analizó pormenorizadamente el rendimiento académico en ese trayecto de las mismas. Si bien, como se sostuvo anteriormente, la investigación cubre el arco temporal 2005 - 2015, se ha establecido el año 2010 como bisagra entre dos subperíodos: el primero, entre los años 2005 y 2009, definido y caracterizado por el inicio y normalización de las carreras; el segundo, entre 2010 y 2015, determinado por el comienzo de los planes vigentes en la actualidad. Debido a ello, este estudio se concentró mayoritariamente en el análisis de esta etapa y en las materias que, por el régimen de correlatividades, deben ser aprobadas durante el primer cuatrimestre del primer año para poder continuar cursando la carrera. La investigación demostró que sólo un 40% de los alumnos aprobaron durante el período 2010-2015 las tres materias del primer cuatrimestre necesarias para poder cursar cuatro de las cinco materias que conforman el segundo cuatrimestre. Por otro lado, alrededor de un 49% de los alumnos que se inscribieron en las instancias de exámenes finales de esas materias lograron la aprobación. Estos resultados abonan al análisis del índice de duración de las carreras, es decir, la relación entre la duración media real y la duración teórica de las mismas. Para cada cohorte se observó que en promedio, los egresados mantuvieron la condición de alumno regular durante ocho años. Esto permite inferir que, si bien el plan de las carreras es de cinco años, son muy pocos los alumnos que logran completarlo en este tiempo. Es notable observar también la condición de cronicidad o rezago académico: los alumnos que abandonan pueden llegar a mantener por once o más años la condición de

estudiante activo, reincorporándose al cursado o recursado de alguna materia, después de dos o tres años de no haber tenido actividad académica.

Fundamentado en los resultados de este estudio longitudinal del período 2005-2015, especialmente en los indicadores de rendimiento académico, se implementó un análisis predictivo que permita pronosticar aquellos casos de estudiantes con un alto grado de rezago y riesgo de deserción. A través de un desarrollo algorítmico, fundamentalmente basado en algoritmos de Regresión Logística, el tablero de control muestra un sistema de alertas tempranas en tres estados: rojo, amarillo, verde vinculadas con los indicadores de rendimiento académico. Por cada indicador se definieron valores que conforman el segmento de los niveles, indicando el color rojo cuando está por debajo del intervalo mínimo establecido, amarillo para cuando está dentro del intervalo y verde para cuando está por encima del intervalo máximo establecido.

A continuación, se plantean los **aportes** que, se espera, surjan como resultado de esta investigación:

- En el futuro mediato, a medida que en el tablero de control se vayan cargando los datos de las nuevas cohortes, se podrá ponderar con criterios estadísticos los factores que incidan en el abandono y rezago académico. Dado el análisis retrospectivo ya elaborado, será factible plantear estudios comparativos entre diversas cohortes, desde la etapa fundacional de la carrera hasta la actualidad. Desde un análisis prospectivo, se podrá alertar sobre potenciales casos de deserción contribuyendo así a mejorar las políticas institucionales para el acompañamiento y retención de los estudiantes.
- Utilizada a nivel de gestión institucional esta herramienta arroja información útil capaz de permitir a las autoridades y docentes tomar decisiones para mejorar el proceso de enseñanza y aprendizaje.
- La metodología de análisis y el tablero de control desarrollado a partir de la utilización de minería de datos, pueden hacerse extensibles al estudio del rezago y

deserción en otras carreras de la UNNOBA, incidiendo positivamente sobre la trayectoria académica de los estudiantes en riesgo de abandono.

REFERENCIAS BIBLIOGRÁFICAS

Sobre rezago académico y deserción estudiantil universitaria

Arana, M., y Bianculli, K. (2006). *Estrategias para la retención de la matrícula universitaria*. La UNMDP y la UTN como análisis de caso.

Ariza, S. y Marín, D. (2009), *Factores intervinientes en la deserción de la facultad de psicología de la Fundación Universitaria Los Libertadores*. Recuperado de: http://www.alfaguia.org/alfaguia/files/1320349840_35.pdf

Braxton, J. M , Johnson, R y Shaw Sullivan, A. (1997). "Appraising Tinto's theory of college students departure" en J.C Smart (Ed.). *Higher Education Handbook of theory and research*. Vol. 12, Agathon Press, New York, Estados Unidos.

Cabrera, L., Bethencourt, J. T., Álvarez Pérez, P., & González Alfonso, M. (2006). El problema del abandono de los estudios universitarios. *Relieve*, 12(2), 171-203.

Castaño, E., Gallón, S., Gómez, K., & Vásquez, J. (2004). Deserción estudiantil universitaria: una aplicación de modelos de duración. *Lecturas de Economía*, 60. pp 39-65.

Corominas, E. (2001). La transición de los estudios universitarios. Abandono o cambio en el primer año de universidad. *Revista de Investigación Educativa*, 19(1), 127-152.

Dapozo, G, Greiner, C, Pedrozo Petazzini, G, Chiapello, J.(2014). "Investigación para fortalecer actividades de promoción y retención de alumnos en carreras de Informática" en XX Congreso Argentino de Ciencias de la Computación, RedUNCI, Buenos Aires.

De Fanelli, A. M. G. (2005). "Acceso, abandono y graduación en la educación superior argentina". en Debate 5. SITEAL.

De Fanelli, A. M. G. (2014). Rendimiento académico y abandono universitario: Modelos, resultados y alcances de la producción académica en la Argentina. *Revista Argentina de Educación Superior*, (8), 9-38.

Diconca, B.; Dos Santos, S.; Egaña, A. (2011). *Desvinculación estudiantil al inicio de una carrera universitaria*. Publicación de la Comisión Sectorial de Enseñanza, Universidad de la República de Uruguay.

Drovandi, María Celina (2010). "Estudio sobre las causas del bajo rendimiento académico en los últimos cursos de carrera del alumnado de Ingeniería Informática de la Universidad de Mendoza, Argentina". Tesis Doctoral, Universidad de Granada, España. Disponible en <http://digibug.ugr.es/bitstream/10481/15707/1/19488920.pdf>

Ethington, C. A. (1990). A psychological model of student persistence. *Research in Higher Education*, 31(3), 266-269.

Germani, G. & R. Sautu (1965) Regularidad y origen social en los estudiantes universitarios, Buenos Aires: Publicaciones del instituto de sociología.

Himmel, E. (2002). Modelos de análisis de la deserción estudiantil en la educación superior. *Revista Calidad en la educación*, 17(2), 91-108.

Lorenzano, César (2004). *La deserción universitaria en la Universidad de Tres de Febrero*. Tesis de grado. Universidad Nacional de Tres de Febrero. Argentina.

Merlino, Ayllón y Escanés (2011) "Variables que influyen en la Deserción de estudiantes universitarios de primer año. Construcción de índices de riesgo de abandono".

Osorio, Ana R. y Catalina Jaramillo, "Deserción Estudiantil en los Programas de] Pregrado 1995-1998," Medellín, Estudio 1999.

Páramo, G., y Correa, C. (1999). Deserción estudiantil universitaria. Conceptualización. Medellín, *Revista Universidad EAFIT*, Abril-Mayo - Junio 1999.

Parrino, M. del. (2005). "Aristas de la problemática de la deserción universitaria" en V Coloquio Internacional sobre Gestión Universitaria en América del Sur. Mar del Plata, 8,9 y 10 de Diciembre de 2005.

Parrino, M. del C. (2015). *¿Evasión o expulsión? Los mecanismos de la deserción universitaria*. Buenos Aires: Biblos.

Rivera, E., Roca, H., Echart, B., Alfaro, E., López, A., Farfán, S., Barrera, B. (2005). Estudio sobre la repitencia y deserción en la educación superior en Bolivia. Secretaría Ejecutiva de Desarrollo Universitario, Universidad Autónoma Juan Misael Saracho, IESALC/UNESCO, Tarija. En www.iesalc.unesco.org.ve

Spady, W. (1971). Dropouts from higher education: toward an empirical model. *Interchange*, 2, 38-62. <http://dx.doi.org/10.1007/BF02282469>

Tinto, V. (1982). Definir la Deserción: Una Cuestión de Perspectiva. *Revista de Educación Superior*, 71, 33-51.

Tinto, V. (1989). Misconceptions Mar Campus Discussions of Student Retention. *The Chronicle of Higher Education*; Sep. 6, 1989; 36, 1; Academic Research Library.

Tinto, V. (1989). Definir la deserción: una cuestión de perspectiva. *Revista de la Educación Superior* (71), 33-51.

Vásquez, C., et al. (2007). La deserción estudiantil en educación superior a distancia: perspectiva teórica y factores de incidencia. [En línea]. *Revista Latinoamericana de Estudios Educativos*, 37.

Vela Peón, F. (2001). "Un acto metodológico básico de la investigación social: la entrevista cualitativa" en Tarrés, M. L. (Coord.) *Observar, escuchar y comprender. Sobre la tradición cualitativa en la investigación social*. FLACSO, México.

Vivas, J. (2005). El abandono de estudiantes universitarios: análisis y reflexiones sobre la experiencia sobre la Universidad Autónoma de Barcelona. Trabajo presentado en el Encuentro Internacional Deserción Estudiantil en Educación Superior. Experiencias Significativas. Bogotá: Colombia.

Sobre Business Intelligence (BI), Knowledge Discovery (KDD) y Minería de Datos (DT)

Aguilar, D. A. G., Therón, R., & García-Peñalvo, F. J. (2009). Semantic spiral timelines used as support for e-learning. *J. ucs*, 15(7), 1526-1545.

Anderson, T., & Dron, J. (2011). Three generations of distance education pedagogy. *The International Review of Research in Open and Distributed Learning*, 12(3), 80-97.

Alexandro Salinas, (2008). *Introducción a Pentaho*. Recuperado de <https://gravitar.biz/bi/introduccion-pentaho-parte-1/>

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief. Recuperado de <http://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf>

Caro, J. (2005). Tecnología Workflow: Estado actual de la Investigación, Departamento de Lenguajes y Ciencias de la Computación–Universidad de Málaga.

Chaudhuri, S., Dayal, U., y Narasayya, V. (2011). *Una visión general de la tecnología de inteligencia de negocios*. Comunicaciones de la ACM, 54 (8), 88-98.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5-6), 318-331.

Cutro, A. (2008). *Minería de Datos Aplicada a la Encuesta Permanente de Hogares*. Trabajo Final de Aplicación de la Licenciatura en Sistemas de Información dirigido por el Prof. David Luis la Red Martínez. Corrientes. Argentina.

Díaz, J. C. (2012). *Introducción al business intelligence*. Editorial UOC.

Dietz-Uhler, B., & Hurn, J. E. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of Interactive Online Learning*, 12(1), 17-26.

Elias, T. (2011). Learning analytics. *Learning*.

Enhydra Shark, (2011). *Documentación*. Recuperado de <http://shark.ow2.org/doc/overview.html>

Forrester Research, (2007). *It's Time to Reinvent Your BI Strategy*. Recuperado de https://www.informationweek.com/software/information-management/its-time-to-reinvent-your-bi-strategy/d/d-id/1061675?page_number=2

Gerolami, N., Revello, E., & Venzal, G. (2011). Implantación de Data Warehouse open free.

GestioPolis (2004). *Procesos de la inteligencia de negocios*. Recuperado de <https://www.gestiopolis.com/procesos-inteligencia-negocios/>

Goncalves, (2017). *Community Chart Components*. Recuperado de <https://community.hds.com/docs/DOC-1009859>

Goncalves, (2017). *Data Mining - Weka*. Recuperado de <https://community.hds.com/docs/DOC-1009857>

Gravitar, (2014). *Pentaho*. Recuperado de <https://gravitar.biz/pentaho/>

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, 15(3), 42.

- Gujarati, D. N., & Porter, D. C. (2003). *Basic Econometrics*. 4th.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques*: Morgan Kaufmann Publishers Inc.
- Heredia, D., Amaya, Y., & Barrientos, E. (2015). Student dropout predictive model using data mining techniques. *IEEE Latin America Transactions*, 13(9), 3127-3134.
- Integrando Datos, (2013). *Pentaho: Introducción*. Recuperado de <http://integrandodatos.blogspot.com.ar/2013/06/pentaho-introduccion.html>
- Hyde, J. (2006). *Pentaho Mondrian Documentation*. Architecture.
- Liu, B. (2007). *Web data mining: exploring hyperlinks, contents, and usage data*. Springer Science & Business Media.
- Lyengar, S. S. y M. R. Lepper. "When choice is demotivating: Can one desire too much of a good thing?", en *Journal of Personality and Social Psychology*, vol. 79, núm. 6, USA, 2000.
- Openred, (2011). *Pentaho BI - Software Libre para la Toma de Decisiones*. Recuperado de <http://www.openred.es/index.php/pentaho>
- Pentaho, (2007). *Data Integration, Spoon Version 2.4.0. User Guide. Manual*. Recuperado de <http://www.pentaho.org/>
- Pentaho, (2009). *Data Integration, Pan Versión 2.4.0. Manual*. Recuperado de <https://wiki.pentaho.com/pages/viewpage.action?pageId=11869063>
- Pentaho, (2009). *Data Integration, Kitchen Versión 2.4.0. Manual*. Recuperado de <https://wiki.pentaho.com/pages/viewpage.action?pageId=11869458>
- Pentaho, (2016). *Data Integration, Carte Versión 2.4.0. Manual*. Recuperado de <https://wiki.pentaho.com/display/EAI/Carte+User+Documentation#CarteUserDocumentation-WhatisCarter>
- Pentaho, (2017). *Multidimensional Data Modeling in Pentaho*. Recuperado de <https://help.pentaho.com/Documentation/8.0/Data/Multidimensional>
- Porter, M. E. (2008). *Competitive advantage: Creating and sustaining superior performance*. Simon and Schuster.
- Protovis,(2010). *A graphical approach to visualization*. Recuperado de <http://mbostock.github.io/protovis/>
- PWC, (2008). *La inteligencia de negocios (Business Intelligence)*. Recuperado de <https://www.pwc.com/ve/es/asesoria-gerencial/boletin/assets/boletin-advisory-edicion-10-2008.pdf>

Reed, R., & MarksII, R. J. (1999). *Neural smithing: supervised learning in feedforward artificial neural networks*. Mit Press.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1), 368-384.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert systems with applications*, 33(1), 135-146.

Sallam, RL, Richardson, J., Hagerty, J., y Hostmann, B. (2011). Cuadrante mágico para plataformas de inteligencia empresarial. Gartner Group, Stamford, CT.

Siemens, G., & d Baker, R. S. (2012, April). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 252-254). ACM.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, 46(5), 30.

Smail, (2010). Pentaho (BI Open Source). Recuperado de <http://slideplayer.es/slide/2542463/>

Smail, (2012). *Taller Pentaho UNNOBA 2012. Pentaho Data Integration*.

Stratebi, (2017). *Análisis de Herramientas BI*. Recuperado de <http://todobi.blogspot.com.ar/2017/04/comparativa-de-herramientas-business.html>

Tuncer, O., & van den Berg, J. (2010). *Implementing BI concepts with Pentaho, an evaluation*. Delft University of Technology. Netherlands.

Van Harmelen, M., & Workman, D. (2012). Analytics for learning and teaching. *CETIS Analytics Series*, 1(3), 1-40.

Wittschen, L. (2004). *Why Business Intelligence?*. January, 2004. [WWW].

Workflow Management Coalition, (2009). *Soporte y recursos de XPD*L. Recuperado de <https://web.archive.org/web/20090512212851/http://wfmc.org:80/xpdl-developers-center.html>

Sobre Minería de Datos aplicado al análisis de la deserción estudiantil universitaria

Gorga, Gladys y Madoz Cristina (2011). "Estrategia metodológica para la contención y retención de alumnos ingresantes. El caso de la carrera Ingeniería en Computación de la

UNLP” en VI Congreso de Tecnología en Educación y Educación en Tecnología, RedUNCI, junio de 2011.

Kuna, H., García Martínez, R., & Villatoro, F. (2009). Identificación de causales de abandono de estudios universitarios. Uso de procesos de explotación de información. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*, 5, 39-44.

Madoz, M.C, Gorga, G y De Giusti, A (2005). Análisis de la utilización de Web Based Learning en un primer curso de Algorítmica. En XI Congreso Argentino de Ciencias de la Computación.

Pautsch, G *et.al* (2011). Resultados preliminares del proceso de minería de datos aplicado al análisis de la deserción en carreras de informática utilizando herramientas open source. *En XVII Congreso Argentino de Ciencias de la Computación, CACIC.*

Valía, L., Rostagno, J., Moine, J. M., Bigatti, C., Riva, F. M., & Amar, E. (2017, August). Minería de datos aplicada a la educación: modelo de deserción universitaria en la Universidad Tecnológica Nacional, Facultad Regional Rosario. In XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, ITBA, Buenos Aires).