

Perfiles de Rendimiento Académico: Un Modelo Basado en Minería de Datos

Academic Performance Profiles: A Model Based on Data Mining

David Luis La Red Martínez¹, Marcelo Karanik¹, Mirtha
Giovannini¹, Noelia Pinto¹

¹ Dpto. de Ingeniería en Sistemas de Información, Facultad Regional Resistencia,
Universidad Tecnológica Nacional, Argentina

laredmartinez@gigared.com

RESUMEN. El rendimiento académico es un factor crítico teniendo en cuenta que, frecuentemente, el bajo rendimiento académico está asociado a una alta tasa de deserción. Esto se ha observado en asignaturas del primer nivel de la carrera de Ingeniería en Sistemas de Información (ISI) de la Universidad Tecnológica Nacional Facultad Regional Resistencia (UTN-FRRRe), situada en la ciudad de Resistencia, provincia del Chaco, Argentina, entre ellas Algoritmos y Estructura de Datos, donde el bajo rendimiento académico se observa en proporciones muy altas (entre el 60% y el 80% aproximadamente en los últimos años). En este trabajo se propone la utilización de técnicas de minería de datos sobre información del desempeño de los alumnos de la asignatura mencionada con el propósito de caracterizar los perfiles de alumnos exitosos (buen rendimiento académico) y de aquellos que no lo son (bajo rendimiento académico). La determinación de estos perfiles permitiría a futuro definir acciones específicas tendientes a revertir el bajo rendimiento académico, una vez detectadas las variables asociadas al mismo. En este artículo se describen los modelos de datos y de minería de datos utilizados y se comentan los principales resultados obtenidos.

ABSTRACT. Academic performance is a critical factor considering that poor academic performance is often associated with an attrition rate. This has been observed in subjects of the first level of Information Systems Engineering (ISI) of the National Technological University, Resistencia Regional Faculty (UTN-FRRRe), situated in Resistencia city, province of Chaco, Argentine. Among them are Algorithms and Data Structures, where the poor academic performance is observed at very high rates (between 60% and about 80% in recent years). In this paper, we propose the use of data mining techniques on performance information for students of the subject mentioned in order to characterize the profiles of successful students (good academic performance) and those that are not (poor performance). In the future, the determination of these profiles would allow us to define specific actions to reverse poor academic performance, once detected the variables associated with it. This article describes the data models and data mining used and the main results are also commented.

PALABRAS CLAVE: Perfiles de rendimiento académico, Almacenes de datos, Minería de datos, Descubrimiento de conocimiento en bases de datos, Algoritmos y estructuras de datos, Patrones de comportamiento.

KEYWORDS: Academic performance profiles, Data warehouses, Data mining, Knowledge discovery in databases, algorithms and data structures, Behavior patterns.

1. Introducción

La velocidad de la evolución tecnológica en la nueva sociedad de la información o cibersociedad implica un gran número de interrogantes de orden técnico, económico, sociológico, cultural y político (Joyanes Aguilar, 1997). Uno de los más importantes es si los sistemas educativos serán capaces de proveer la cantidad y calidad de profesionales para satisfacer las demandas de personal altamente capacitado de esta sociedad de la información y el conocimiento (SIC), especialmente en las áreas relacionadas con las TIC.

Por ello, en el ámbito educativo universitario se plantea el constante desafío de mantener día a día la calidad académica (inclusive mejorarla). Debido a esto, constantemente se revisan contenidos, estrategias y métodos de enseñanza en busca de poder garantizar adecuados estándares de calidad que den como resultado la formación de profesionales altamente calificados útiles para la sociedad.

Claramente, el rendimiento académico es un factor crítico teniendo en cuenta además que, frecuentemente, el bajo rendimiento académico está asociado a una alta tasa de deserción. Esto es precisamente lo que recurrentemente se ha observado en asignaturas del primer nivel de la carrera de Ingeniería en Sistemas de Información (ISI) de la Universidad Tecnológica Nacional Facultad Regional Resistencia (UTN-FRRe), situada en la ciudad de Resistencia, provincia del Chaco, Argentina, entre ellas Algoritmos y Estructura de Datos, donde el bajo rendimiento académico se observa en proporciones muy altas (entre el 60% y el 80% en los últimos años).

Específicamente, se define al rendimiento académico como la productividad del sujeto, matizado por sus actividades, rasgos y la percepción más o menos correcta de los cometidos asignados (Forteza, 1975). El rendimiento académico suele ser objeto de debate, ya que existen muchos factores que pueden afectar el desempeño del alumno y determinar de manera precisa el rendimiento no es una tarea trivial. El rendimiento académico es afectado por una multiplicidad de factores heterogéneos (internos y externos) que condicionan el desempeño del estudiante. Por esto, evaluar el desempeño de todos los alumnos de la misma manera no brinda información que pueda ser utilizada para detectar, y corregir, problemas cognitivos, de aprehensión, de discernimiento, etc. Una alternativa válida, entonces, es intentar establecer si existen particularidades comunes a grupos de alumnos. De esta manera, la determinación de perfiles se convierte en una estrategia de valor significativo a la hora de tomar acciones que permitan mejorar la performance de los alumnos.

Establecer perfiles es una actividad muy difundida en muchas áreas y es análoga al proceso de determinación y clasificación de patrones. En la actualidad existen muchos métodos para determinar y clasificar patrones que se utilizan en el área de la Inteligencia Artificial y del Aprendizaje de máquinas (Mitchell, 1997). Estos algoritmos devuelven información de valor al momento de tomar decisiones. La información se organiza en grandes almacenes de datos (del inglés Data Warehouse o DW) que, luego de una limpieza previa, es analizada por algoritmos que realizan minería de los datos (del inglés Data Mining o DM).

En este contexto, se puede utilizar el proceso de análisis en 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.

En el caso de la determinación de perfiles de rendimiento académico, un modelo descriptivo puede explicar las variables que afectan el desempeño de los alumnos. De esta manera pueden obtenerse varias vistas de un mismo problema y tratar los inconvenientes teniendo una visión global. El modelo predictivo, en cambio, trata de establecer situaciones problemáticas futuras. Es decir, intenta determinar de manera temprana el perfil de rendimiento de un alumno (de acuerdo a los datos procesados) a fin de poder llevar a cabo acciones que ayuden a corregir posibles inconvenientes.

En este trabajo se propone la utilización de técnicas de DM sobre información del desempeño de los alum-

nos de la asignatura Algoritmos y Estructura de Datos antes mencionada con el propósito de caracterizar los perfiles de alumnos exitosos (buen rendimiento académico) y de aquellos que no lo son (bajo rendimiento académico). La determinación de estos perfiles permitiría a futuro definir acciones específicas tendientes a revertir el bajo rendimiento académico, habiéndose detectado las variables asociadas al mismo. Este artículo se estructura de la siguiente manera: en la Sección 2 se detallan algunos conceptos y trabajos relacionados con la medición del desempeño académico. Los conceptos relacionados con DW y DM se presentan en la Sección 3. En la Sección 4 se describe el ámbito de aplicación de la propuesta y el modelo utilizado. En la Sección 5 se muestran los resultados obtenidos. Finalmente, en la Sección 6 se presentan algunas conclusiones en relación al trabajo realizado.

2. Rendimiento Académico

Si bien el rendimiento académico caracteriza individualmente al alumno, es posible obtener a partir de él, ciertas características generales de grupos de estudiantes. Es allí donde radica la importancia de contar con métodos confiables de evaluación del rendimiento. Existen varias maneras de evaluar el rendimiento académico. En términos generales, implica determinar la producción real de un alumno respecto a las actividades formales. No obstante, a la hora de operativizar el rendimiento, se tiende al reduccionismo (González, 1988). Otra manera es utilizar indicadores como las tasas de graduación, diferenciando por tipos de centros y analizando el rendimiento académico a partir de datos individuales (García & San Segundo, 2001) o a través de las calificaciones de entrada a la Universidad, realizando el análisis de los datos mediante la técnica estadística de curva ROC (Receiver Operating Characteristic) (Vivo Molina et al., 2004).

Los aspectos cognitivos fueron la base de las primeras investigaciones sobre el proceso de aprendizaje; luego los investigadores descubrieron la importancia de los componentes afectivos y su influencia decisiva en el aprendizaje; finalmente se conjugaron los aspectos cognitivos y los afectivos, naciendo así el constructo llamado aprendizaje auto-regulado (Herrera Clavero et al., 2004).

El rendimiento académico universitario también ha sido estudiado aplicando el enfoque de función de producción para estimar los determinantes del rendimiento académico (Di Gresia, 2007). Por ejemplo, se han analizado los determinantes del aprendizaje mediante un enfoque de función de producción sugiriendo que los rendimientos escolares dependen de factores genéticos y socioeconómicos, de la calidad del docente, de las condiciones de la escuela y del grupo de alumnos (Delfino, 1989).

Se ha demostrado con varios estudios que el factor más relacionado con la calidad educativa es el propio alumno como co-productor, medido a través del nivel socioeconómico del hogar de donde proviene (Maradona y Calderón, 2004) y se ha evidenciado que la productividad del estudiante es mayor para las mujeres, para los estudiantes de menor edad y para quienes provienen de hogares con padres más educados (Porto & Di Gresia, 2000), teniendo gran importancia la relación entre horas trabajadas y rendimiento académico (Fazio, 2004).

Existe una paradoja respecto a la disponibilidad de tiempo para estudiar. Se ha mostrado el contraste que hay entre las personas que trabajan y estudian y las que solamente estudian, encontrándose que no existen diferencias en el rendimiento académico de los dos conjuntos (Reyes, 2004). Todo esto da como resultado que en general los estudios empíricos confirman la correlación entre mayores niveles de educación y atributos positivos luego de los estudios (McMahon, 2002).

Existen varios estudios que utilizan técnicas matemáticas para la evaluación del desempeño. En este sentido, se ha estudiado la capacidad de la regresión lineal y de la regresión logística en la predicción del rendimiento y del éxito/fracaso académico, partiendo de variables como la asistencia y la participación en clase. Se ha concluido que el rendimiento previo es un buen predictor del rendimiento futuro y que la asistencia y sobre todo la participación son variables con un peso importante en la predicción del rendimiento (García Jiménez et al., 2000).

También se ha demostrado que las variables planificación del estudio, inteligencia, apoyo del profesor, estudio, tiempo, condiciones ambientales de estudio e implicación, formaban parte de la ecuación de predicción de regresión múltiple, explicando un 25,70% de la varianza del rendimiento escolar en cursos de bachillerato (Marcelo García et al., 1987).

Claramente, contar con métodos matemáticos adecuados es una ventaja a la hora de la evaluación del desempeño, la cuestión es usarlos de manera adecuada. El problema de encontrar buenos predictores del rendimiento futuro de manera que se reduzca el fracaso académico en los programas de postgrado ha recibido una especial atención en EE.UU. (Wilson & Hardgrave, 1995), habiéndose encontrado que las técnicas de clasificación como el análisis discriminante o la regresión logística son más adecuadas que la regresión lineal múltiple a la hora de predecir el éxito/fracaso académico.

La diversidad de estudios sobre el rendimiento académico muestra que no existe una manera única para evaluarlo. Por ello, la determinación de grupos o clases de alumnos es un elemento a tener en cuenta para establecer las causas de los problemas relacionados al desempeño de los alumnos. Más aún, los problemas pueden variar dependiendo el contexto regional y la realidad social donde está inserto el alumno. Es decir, no existen herramientas que se puedan aplicar a todos los ámbitos y los resultados tampoco pueden ser extensibles para explicar todas las situaciones posibles. Esto denota claramente la necesidad de determinar perfiles en las instituciones educativas específicas adaptando las herramientas a cada situación particular. En el caso objeto del presente artículo se ha optado por técnicas de minería de datos aplicadas a un almacén de datos cargado con información socio-económica y académica de los alumnos de la asignatura mencionada precedentemente.

3. Manejo de los Datos

La organización correcta de los datos, sumada a un modelo adecuado de manejo de los mismos puede brindar una visión clara de los inconvenientes en el desempeño de los alumnos. En este sentido, existen herramientas del área de la Inteligencia Artificial, específicamente a la Inteligencia de Negocios (del inglés Business Intelligence o BI), tales como los Almacenes de Datos (del inglés Data Warehouse o DW) y la Minería de Datos (del inglés Data Mining), utilizadas para el descubrimiento de conocimiento oculto en grandes volúmenes de datos que pueden ser utilizadas para determinar patrones y perfiles adecuadamente.

En las sub-secciones siguientes se hace una breve reseña de los aspectos más destacados de estas herramientas de almacenamiento y procesamiento de datos.

3.1. Almacén de Datos (Data Warehouse)

Un DW es una colección de datos orientado a temas, integrado, no volátil, de tiempo variante, que se usa para el soporte del proceso de toma de decisiones gerenciales (Inmon, 1992, 1996), (Simon, 1997). Debido a que un DW no se puede adquirir, se tiene que construir siguiendo determinada metodología. La técnica a utilizar en la creación del DW depende de hacia quién se enfoca como punto principal el desarrollo del mismo, puede ser hacia el manejo de datos, de metas o de usuarios (Data-Driven, Goal-Driven y User-Driven) (Gutting, 1994).

Data-Driven: A diferencia de los sistemas clásicos de manejo de requerimientos de usuarios, este enfoque considera que en un DW lo que se manejan son datos, considerando las necesidades de los usuarios en segundo término (Poe, 1996). El modelo de datos consiste de pocas dimensiones y de grupos de hechos. La dimensión representa la estructura básica del diseño. Los hechos son basados en el tiempo y tienen poco nivel de granularidad.

Goal-Driven: Las metas y objetivos establecidos en principio guían el proceso de desarrollo y, a diferencia del modelo Data-Driven, éste contiene más dimensiones y pocos hechos, los cuales son basados en el tiempo y tienen un bajo nivel de granularidad.

User-Driven: Considera que el factor principal a tener en cuenta son las necesidades de los usuarios, pues son quienes utilizarán finalmente el sistema. El modelo consta de pocos hechos, los cuales tienen un nivel moderado de granularidad.

Independientemente de los modelos de desarrollo mencionados, las metodologías a seguir para el desarrollo del DW dependen en gran parte del tamaño del DW a crear y de la prontitud con que se requiera el DW. Dos de las principales metodologías para el desarrollo de un DW son la Rapid Warehousing y la Big Bang.

Rapid Warehousing, también conocida como metodología evolutiva o incremental, considera que la construcción e implantación de un DW es un proceso evolutivo, donde se crea una parte de un DW con la integración de data marts (Widom, 1995). En cambio Big Bang trata de resolver todos los problemas conocidos para crear un DW de gran tamaño, antes de liberarlo para su evaluación y prueba (Harinarayan et al., 1996).

Los bloques funcionales que se corresponden con un sistema de información completo que utiliza un DW se resumen en la Figura 1:

- Nivel operacional: hace referencia a los sistemas operacionales/transaccionales de la organización y a fuentes que forman parte del proceso de Data Warehousing.
- Nivel de acceso a la información: es la capa de interacción del usuario cuya finalidad es la conversión de los datos almacenados en información fácil y transparente para las herramientas de los usuarios finales.
- Nivel de acceso a los datos: comunica el nivel de acceso a la información con el nivel operacional de forma universal.
- Nivel de directorio de datos (metadatos): repositorio de metadatos de los datos almacenados que proporcionan información sobre el origen y sobre la transformación de los mismos en el proceso de Data Warehousing.
- Nivel de gestión de procesos: planificación de las tareas y procesos para la construcción y mantenimiento actualizado del Data Warehouse.
- Nivel de mensaje de la aplicación: determina el transporte de información a lo largo del entorno de computación de la organización a modo de middleware pero más allá de meramente protocolos de red.
- Nivel Data Warehouse (físico): es el repositorio central altamente flexible de información donde residen copias de los datos operacionales y/o externos optimizados para su acceso para la consulta.
- Nivel de organización de datos: incluye todos los procesos necesarios para seleccionar, editar, resumir (normalmente sumarizar), combinar y cargar en el Data Warehouse y en la capa de acceso a la información los datos operacionales y/o externos.

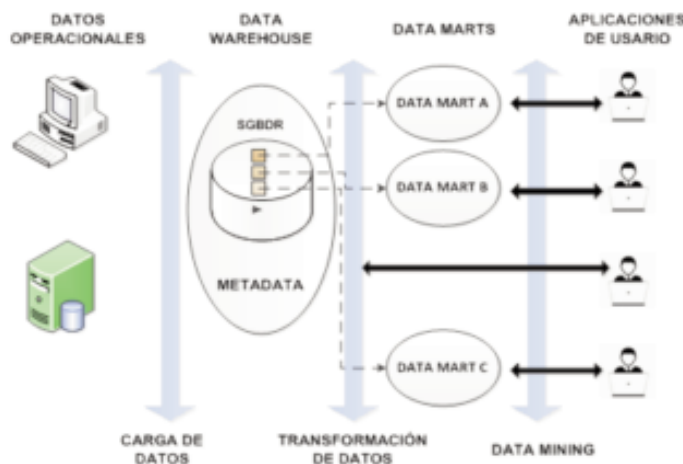


Figura 1. Arquitectura básica de un DW.

Claramente el modelo y la metodología a utilizar dependen del ámbito de aplicación y son decisiones muy importantes a la hora de trabajar con estas tecnologías. Este tipo de almacenes de datos brindan el soporte adecuado para la implementación de técnicas de búsqueda de conocimiento y son adecuados para la determinación de perfiles de alumnos, destacándose con tal fin el uso de técnicas de minería de datos.

3.2. Minería de Datos (Data Mining)

DM es la etapa de descubrimiento de conocimiento en bases de datos (del inglés Knowledge Discovery from Databases o KDD). Es el uso consistente de algoritmos concretos que generan una enumeración de patrones a partir de los datos pre-procesados (Fayyad et al., 2001), (Hand et al., 2000), (Frawley et al., 1992). DM está muy ligada a los DW ya que los mismos proporcionan la información histórica con la cual los algoritmos de minería obtienen la información necesaria para la toma de decisiones (IBM Software Group, 2003). El conjunto de técnicas de análisis de datos que permiten extraer tendencias, patrones y regularidades para describir y comprender mejor los datos forman parte de la DM. Además, permite extraer patrones y tendencias para predecir comportamientos futuros (Simon, 1997), (Berson & Smith, 1997), (White, 2001).

Como se mencionó anteriormente, utilizando DM se pueden generar modelos descriptivos o predictivos (Agrawal & Shafer, 1996). Si bien las técnicas utilizadas son varias, una de las más destacadas es la de clustering (o agrupamiento de datos) (Grabmeier & Rudolph, 1998), (Ballard et al., 2007).

Actualmente existen varias metodologías de DM; las más difundidas son la CRISP-DM y la SEMMA. La metodología CRISP-DM (del inglés Cross-Industry Standard Process for Data Mining: CRISP-DM), se organiza en seis etapas, cada una de ellas a su vez se divide en varias tareas (Chapman, 1999). Las seis etapas son:

- Comprensión del negocio: abarca la comprensión de los objetivos de proyecto y las exigencias desde la perspectiva de negocio.
- Comprensión de los datos: implica la recolección de datos inicial y sigue con las actividades que le permiten descubrir subconjuntos relevantes para formar hipótesis relacionadas a la información oculta.
- Preparación de los datos: cubre todas las actividades necesarias para construir el conjunto de datos final a partir de los datos en bruto iniciales.
- Modelado: implica la selección de varias técnicas de modelado y sus parámetros son calibrados a valores óptimos.
- Evaluación: consiste en evaluar a fondo el modelo y la revisión de los pasos ejecutados para crearlo, para comparar el modelo correctamente obtenido con los objetivos establecidos.
- Despliegue: abarca desde la generación de un informe hasta la realización repetida de un proceso de minería de datos a través de la organización.

Este modelo es una secuencia idealizada de eventos. En la práctica, muchas de las tareas pueden realizarse en orden diferente e inclusive necesitar vuelta atrás hacia alguna de ellas.

La metodología SEMMA se utiliza para descubrir patrones de negocio desconocidos. El nombre refiere a las cinco fases básicas del proceso (del inglés Sample, Explore, Modify, Model, Assess) (Matignon, 2009), (SAS Institute, 2009). Las etapas consisten en lo siguiente:

- Sample: Obtener un subconjunto de los datos suficientemente representativo para poder obtener información útil de ellos y procesarlos fácilmente.
- Explore: Buscar patrones en los datos.
- Modify: Crear y transformar variables o incluso eliminar las que son innecesarias.
- Model: Seleccionar y aplicar el modelo que mejor se ajuste a los datos.
- Assess: Determinar si los resultados son útiles y confiables. Testear los resultados contra datos conocidos.

Ambas metodologías brindan un marco de trabajo adecuado para poder abordar la determinación de perfiles de alumnos. Con estas metodologías se buscó explicar el comportamiento de ciertas variables y poder determinar aspectos relevantes dentro del desempeño académico, para lo cual se desarrolló un modelo que se describe a continuación.

4. Modelo Utilizado

En esta sección se describirá el ámbito de aplicación y brevemente el modelo de datos del proyecto de determinación de perfiles de alumnos según su rendimiento académico; una descripción más detallada puede verse en (La Red Martínez et al., 2014d, 2014e).

4.1. Ámbito de Aplicación

El ámbito de aplicación del proyecto de determinación de perfiles de alumnos fue la Facultad Regional Resistencia de la Universidad Tecnológica Nacional (UTN-FRRe), donde se dicta la carrera de Ingeniería en Sistemas de Información. Esta carrera tiene varias materias en el primer año que son específicas a la profesión del ingeniero en sistemas y sus contenidos no son abordados durante el ciclo secundario. Esto provoca que el alumno se encuentre con una cierta cantidad de temas nuevos y la forma de abordarlos no sea conocida para él. Estas materias por lo general son las que generan el mayor desgranamiento en los primeros años de la carrera ya que su aprobación es obligatoria para cursar otras de años posteriores. Una de estas materias es Algoritmos y Estructura de Datos, cuyo estudio implica un alto contenido de procedimientos lógicos y manejo de estructuras para distintos tipos de variables.

En este sentido se buscó determinar, utilizando técnicas de Data Warehouse y Data Mining, las variables que expliquen el desigual rendimiento académico por parte de los alumnos de Algoritmos y Estructuras de datos de la carrera de Ingeniería en Sistemas de Información de la UTN-FRRe, con el fin de establecer acciones que permitan mejorar dicho rendimiento académico de los alumnos.

En este contexto se consideró rendimiento académico a los resultados logrados en las evaluaciones realizadas durante el cursado de la asignatura en el año 2013 (la carga, depuración y proceso de la información se realizó en el año 2014). Se buscó determinar en qué medida el desigual rendimiento académico de los alumnos de Algoritmos y Estructuras de Datos es influenciado por las siguientes variables: escuela media de procedencia, nivel educativo de los padres, nivel socio-económico, edad, género, actitud general hacia el estudio, existencia del cursillo de ingreso, régimen de cursado (anual - cuatrimestral), uso de herramientas de apoyo (campus virtual).

Se buscaron perfiles de alumnos de rendimiento académico bajo, medio y alto utilizando minería de datos sobre un almacén de datos. En trabajos similares (La Red Martínez et al., 2010, 2012) se propuso un modelo de análisis de datos que integra la información académica y de contexto. En la sección siguiente se describe el modelo de análisis de datos utilizado en este trabajo.

4.2. Modelo de Datos

Como se mencionó en la sección anterior el objetivo general fue determinar, utilizando técnicas de DW y DM, las variables que explican el desigual rendimiento académico por parte de los alumnos de Algoritmos y Estructuras de Datos de la carrera de Ingeniería en Sistemas de Información de la UTN-FRRe. Para lograr esto se realizan las siguientes actividades:

- Relevar información de la situación actual respecto al rendimiento académico de los alumnos.
- Filtrar y depurar la información contenida en las bases de datos actuales.
- Establecer las variables relevantes para describir la situación objeto de estudio.
- Determinar cómo influye cada una de las variables que se fijaron para evaluar la situación del alumno.
- Determinar cómo influye cada una de las variables que se fijaron para evaluar el contexto académi-

co.

- Establecer acciones que tiendan a mejorar los índices de rendimiento académico de los alumnos.

Utilizando la técnica User-Driven se buscó determinar los perfiles de rendimiento (bajo, medio y alto) en base a los resultados obtenidos por los alumnos en las evaluaciones y luego se buscaron relaciones y correlaciones entre las variables mencionadas en la sección anterior. Esto permitió tener una descripción cualitativa de la situación de los alumnos. Es decir, dados los tres tipos de rendimiento, se buscó demostrar la relación del desempeño y las variables analizadas.

En una primera instancia se tomó la información correspondiente a los alumnos de la asignatura Algoritmos y Estructura de Datos de la base de datos del sistema académico, del que se extrajeron los datos específicos de los alumnos y de sus calificaciones, las que se consideraron indicadores de su rendimiento académico. Los datos referidos a la situación socio-económica del alumno y de su familia, como así también los aspectos actitudinales respecto del estudio y de las TIC, fueron recolectados mediante una encuesta realizada utilizando un sistema de formularios en un aplicativo en línea. Esta información fue pre-procesada, haciéndose una limpieza de datos inconsistentes y faltantes. El universo estuvo constituido por los alumnos en condiciones de cursar la asignatura durante el año 2013 (se está trabajando en la carga de información del curado del año 2014 y de años anteriores, aproximadamente 300 alumnos por año) y la unidad de análisis fue cada uno de esos alumnos. Los casos seleccionados fueron los alumnos en condiciones de cursar la asignatura y que habitualmente concurren a clases (no se han considerado los alumnos que se inscriben para cursar pero que luego no lo hacen).

La estructura del DW, como se muestra en la Figura 2, consta de una tabla de hechos y varias tablas de dimensiones. Se pueden distinguir dos tipos de columnas en una tabla de hechos, columnas de hechos y columnas llaves. Las columnas de hechos almacenan las medidas del asunto que se quieren controlar y las columnas llaves, forman parte de la clave de la tabla.

Una tabla de dimensiones o entidad de dimensiones es una tabla o entidad que almacena detalles acerca de hechos. Incluye información descriptiva sobre los valores numéricos de una tabla de hechos. Por ejemplo, las tablas de dimensiones para una aplicación de análisis de mercado pueden incluir el tipo de período de tiempo, región comercial y producto. Asimismo las tablas de dimensiones describen los distintos aspectos de un tema en estudio. Cada tabla de dimensiones contiene una clave simple y un conjunto de atributos que describen la dimensión. Las columnas de una tabla de dimensiones se utilizan para crear informes o para mostrar resultados de consultas. Por ejemplo, las descripciones textuales de un informe se crean desde las etiquetas de las columnas de una tabla de dimensiones.

El modelo que se utilizó en este trabajo, se compone de la tabla de hechos Alumnos y varias tablas de dimensiones asociadas a la misma, que incluyen características que se desean estudiar. En la Figura 2, se representa gráficamente esta estructura.



Figura 2. Modelo del DW utilizado.

La tabla de hechos incluye información específica del alumno y su rendimiento académico, en tanto que las tablas de dimensiones contienen información que hace a la descripción del entorno socio-económico del alumno y familia, sus antecedentes académicos en los estudios secundarios y su actitud hacia el estudio y hacia las TIC.

Para llevar a cabo este trabajo se ha utilizado la suite de herramientas incluidas en el IBM Data Warehouse Edition (DWE) V.9.5, un paquete de productos que combina la potencia de DB2 Universal Database (DB2 UDB) con la robusta infraestructura de inteligencia empresarial de IBM. Se compone de varios productos pero a los fines de este trabajo se han usado los siguientes componentes:

- El Sistema de Gestión de Base de datos DB2 Enterprise Server Edition (DB2 ESE), que permite conexiones multiusuarios y presenta alta escalabilidad.
- Design Studio (DS), una plataforma a modo de herramienta que usan los analistas de negocios y los administradores de almacenes de datos para diseñar las reglas de las cargas de trabajo, los flujos de transformación de datos y los flujos analíticos para la minería de datos y el análisis de textos.
- Intelligent Miner (IM), un conjunto de funcionalidades que permiten realizar análisis de información de acuerdo a conceptos de Business Intelligence (BI).

Se consideró interesante estudiar los datos con técnicas de minería de datos, tanto descriptivas como predictivas, habiéndose empleado la metodología CRISP-DM, mencionada en la sección 3.2; el esquema de trabajo ha sido similar al descrito en (La Red Martínez & Podestá, 2014a, 2014b, 2014c), adaptado a las particularidades propias de la UTN - FRRre. Las acciones desarrolladas se indican en la siguiente sección, donde se comentan los procesos de DM realizados y los principales resultados obtenidos.

5. Detección de Perfiles de Rendimiento Académico

El trabajo realizado fue estructurado en varias etapas tales como la selección, depuración y preparación de datos, minería de datos, descripción de los resultados obtenidos, las que se explican a continuación.

5.1. Etapa de Selección

En esta etapa se seleccionaron las distintas fuentes de información (internas y externas) que sirvieron como base para la etapa de Minería de Datos.

Como fuente de información interna se utilizó la información de la base de datos institucional y de la cátedra, donde se almacenan las calificaciones de los exámenes parciales de los alumnos y su condición al finalizar el cursado (Libre, Regular, Promocionado).

Tal como se planteó en secciones anteriores, para el estudio del desempeño académico de los estudiantes no sólo deben considerarse los resultados de las instancias de evaluaciones previstas por asignatura, sino que también deben analizarse otros factores culturales, sociales y/o económicos que afecten el rendimiento del alumno. Por ello, para la obtención de información externa resultó determinante la participación directa del estudiante, pues era necesario conocer datos sobre aspectos personales que no se podían obtener de otra manera. A tal fin, se utilizó una aplicación web que permitió contar con una Encuesta On-Line compuesta por preguntas relacionadas a situación familiar, historial de estudios secundarios, entre otras cuestiones, incluyendo aspectos actitudinales del alumno respecto del estudio y de las TIC.

Los datos académicos, socio-económicos y actitudinales logrados de la manera antedicha sirvieron para la construcción del DW que luego se utilizó para los procesos de minería de datos.

5.2. Etapa de Depuración y Preparación de los Datos

La calidad de los patrones que se obtienen con la minería de datos es directamente proporcional a la calidad de los datos utilizados (Sposito et al., 2010). En función a esto, el objetivo de esta etapa fue la detección y corrección y/o eliminación de datos anómalos.

Para ello se realizaron tareas exploratorias analizando cada uno de los registros almacenados en la tabla Encuesta (tabla intermedia desde la cual luego de esta etapa se cargaron al DW los datos socio-económicos y actitudinales), logrando la eliminación de inconsistencias a través de correcciones manuales, tales como la existencia de espaciados innecesarios y otros errores tipográficos o incluso la eliminación de registros totalmente en blanco, generados por una inadecuada utilización del sistema on-line por parte de los alumnos.

Una vez depurada la información obtenida se procedió, por cada estudiante, a la carga manual de las calificaciones correspondientes a los 3 exámenes parciales, recuperatorios y la condición final del alumno al concluir el cursado de la asignatura. Resulta necesario aclarar que en caso que el alumno no haya rendido algún examen, el campo correspondiente se dejó en blanco.

Como última actividad en esta etapa, y contando con la información completa, se procedió a la carga del DW mediante flujos de datos desde la tabla Encuesta. Al finalizar el proceso y antes de iniciar con la próxima fase se disponía de 242 registros.

5.3. Etapa de Minería de Datos

En esta etapa se seleccionaron las técnicas de MD a utilizar, creándose los flujos de minería correspondiente, en los cuales se parametrizan los respectivos algoritmos (IBM Academic Initiative, 2009).

En principio se comenzó con la técnica supervisada de Clasificación con Árboles de Decisión, cuyo objetivo es realizar clasificaciones sobre los datos conocidos y así crear modelos que luego puedan ser utilizados para predecir o clasificar valores nuevos o desconocidos (Podestá Gómez, 2013).

5.4. Descripción de los Resultados Obtenidos

El análisis de los resultados se basó en considerar como parámetro de minería la variable relacionada a la situación final del alumno, la cual refleja su estado en la materia al cierre del Ciclo Lectivo.

Se consideró en situación de Libre al alumno que no aprobó ni los exámenes parciales ni los recuperatorios; Regular a quien logró aprobar los 3 exámenes parciales (recuperándolos o no) con nota mayor o igual al 60% pero que no alcanzó al menos el 75% en todos ellos. Finalmente los alumnos en el estado Promocionado, son

aquéllos que aprobaron todos los parciales con nota mayor o igual al 75%.

Teniendo en cuenta lo anterior, se han obtenido los siguientes resultados: 81.42% de alumnos en condición de Libre, 10.62% Regular y sólo 7.96% Promocionado.

Así, y basando siempre el análisis según el parámetro de Situación, se consideraron diferentes criterios de agrupación de los datos para la descripción de clases:

- Dependencia de la Escuela Secundaria.
- Cantidad Horas dedicadas al Estudio.
- Importancia dada al Estudio.
- Estudios alcanzados por la Madre.
- Estudios alcanzados por el Padre.
- Utilización de las TIC.

De esta forma, por ejemplo, en la Figura 3 se muestran los porcentajes considerando la clase Regular según el tipo de Escuela Secundaria desde donde proviene el estudiante.

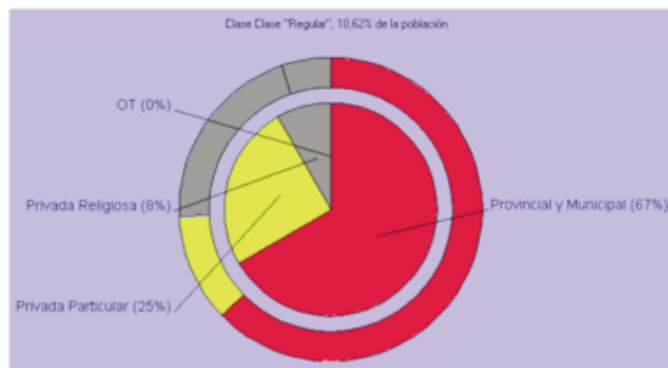


Figura 3. Dependencia de la Escuela Secundaria para alumnos Regulares.

Luego, para una mejor lectura, en la Tabla 1 se describe cómo se caracterizan todas las clases (Regular, Libre y Promocionado) teniendo en cuenta el criterio de Dependencia de la Escuela Secundaria (Privada Religiosa, Provincial y Municipal, Privada Particular, entre otros).

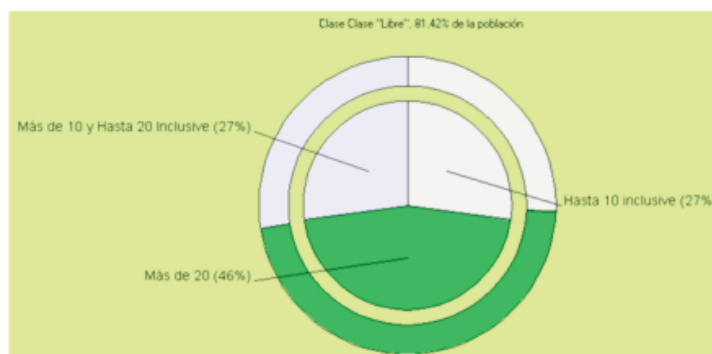


Figura 4. Cantidad de Horas dedicadas al Estudio para alumnos Libres.

Clase	Atributo	Porcentaje
Regular 10.62%	Privada Religiosa	8%
	Provincial y Municipal	67%
	Privada Particular	25%
	Otros	0%
Libre 81.42%	Privada Religiosa	22%
	Provincial y Municipal	61%
	Privada Particular	11%
	Nacional	4%
	Otros	2%
Promocionado 7.96%	Privada Religiosa	22%
	Provincial y Municipal	78%

Tabla 1. Caracterización de Clases teniendo en cuenta la Dependencia de Escuela Secundaria.

En la Figura 4 se presenta un ejemplo gráfico de la distribución de porcentajes obtenidos para la Clase Libre, teniendo en cuenta el criterio de Cantidad de Horas dedicadas al Estudio (Hasta 10, Más de 10 y Hasta 20, Más de 20, etc.).

En la Tabla 2 se incluye la caracterización de todas las clases (Regular, Libre y Promocionado) teniendo en cuenta el criterio de Cantidad de Horas dedicadas al Estudio.

Clase	Atributo	Porcentaje
Regular 10.62%	Hasta 10 inclusive	17%
	Más de 10 y Hasta 20	33%
	Más de 20	50%
Libre 81.42%	Hasta 10 inclusive	27%
	Más de 10 y Hasta 20	27%
	Más de 20	46%
Promocionado 7.96%	Hasta 10 inclusive	22%
	Más de 10 y Hasta 20	22%
	Más de 20	56%

Tabla 2. Caracterización de Clases teniendo en cuenta la Cantidad de Horas dedicadas al Estudio.

Para el caso del criterio Importancia dada al Estudio se muestra en la Figura 5 los resultados obtenidos para la clase "Promocionado" a modo de ejemplo y en la Tabla 3 se incluye la descripción detallada para cada clase teniendo en cuenta este criterio.

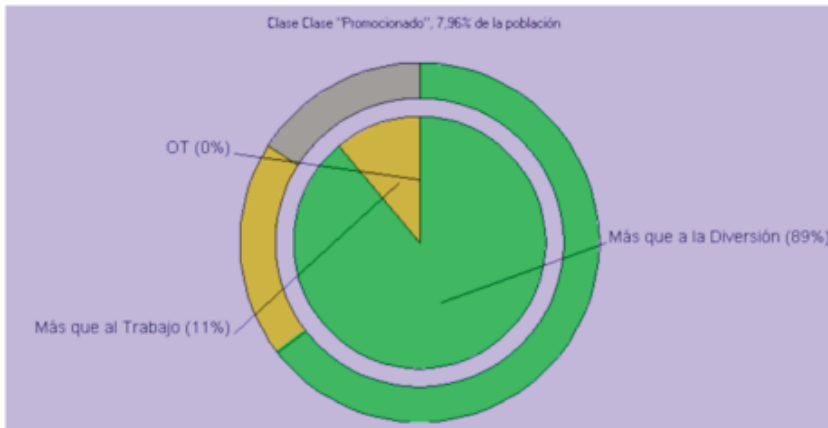


Figura 5. Importancia dada al Estudio para alumnos Promocionados.

Clase	Atributo	Porcentaje
Regular 10.62%	Más que al Trabajo	33%
	Más que a la Diversión	50%
	Más que a la Familia	17%
Libre 81.42%	Más que al Trabajo	18%
	Más que a la Diversión	64%
	Más que a la Familia	17%
Promocionado 7.96%	Más que al Trabajo	11%
	Más que a la Diversión	89%

Tabla 3. Caracterización de Clases teniendo en cuenta la Importancia dada al Estudio.

Ahora, teniendo en cuenta el criterio de Últimos Estudios alcanzados por la Madre, se obtuvo el gráfico que se muestra en la Figura 6 y corresponde al caso de la clase "Regular". Para mayor detalle se incluye una descripción general para todas las clases en la Tabla 4.

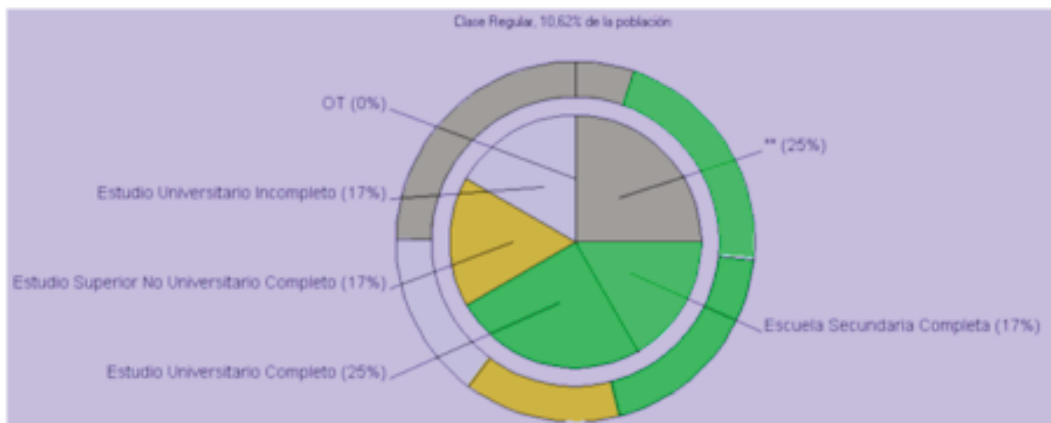


Figura 6. Últimos Estudios alcanzados por la Madre para alumnos Regulares.

Clase	Atributo	Porcentaje
Regular 10.62%	Escuela Secundaria Completa	17%
	Estudio Superior No Universitario Completo	17%
	Estudio Universitario Incompleto	17%
	Estudio Universitario Completo	25%
	No contesta	25%
Libre 81.42%	No hizo Estudios	1%
	Escuela Primaria Incompleta	2%
	Escuela Primaria Completa	5%
	Escuela Secundaria Incompleta	8%
	Escuela Secundaria Completa	23%
	Estudios Superiores No Universitarios Incompletos	3%
	Estudios Superiores No Universitarios Completos	15%
	Estudios Universitarios Incompletos	15%
	Estudios Universitarios Completos	17%
	Estudios de Posgrados	7%
No Contesta	3%	
Promocionado 7.96%	Escuela Secundaria Incompleta	22%
	Escuela Secundaria Completa	11%
	Estudios Universitarios Incompletos	11%
	Estudios Universitarios Completos	33%
	Estudios de Posgrado	22%

Tabla 4. Caracterización de Clases teniendo en cuenta los Últimos Estudios de la Madre.

Teniendo en cuenta el mismo criterio anterior pero analizando la situación del Padre se obtuvieron los resultados detallados en la Tabla 5, y se incluye en la Figura 7, como ejemplo la distribución de porcentajes para la clase “Libre”.

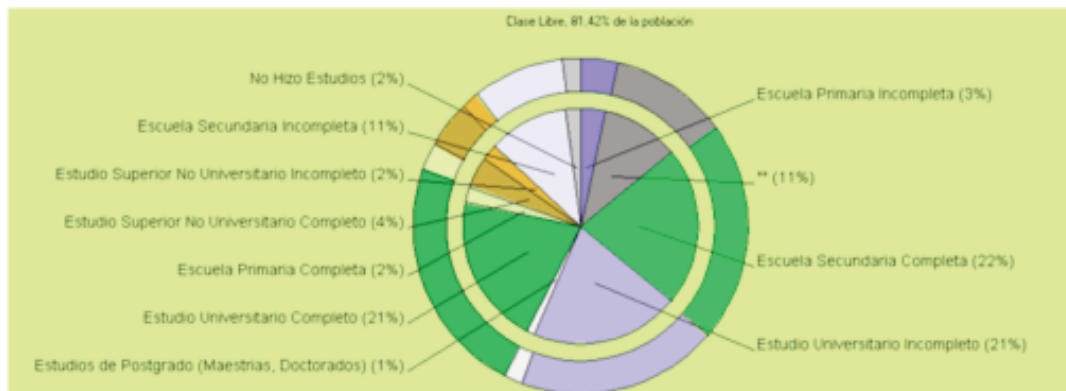


Figura 7. Últimos Estudios alcanzados por el Padre para alumnos Libres.

Clase	Atributo	Porcentaje
Regular 10.62%	Escuela Primaria Incompleta	8%
	Escuela Secundaria Completa	17%
	Estudio Superior No Universitario Completo	8%
	Estudio Universitario Incompleto	25%
	Estudio Universitario Completo	25%
	No Contesta	17%
Libre 81.42%	No hizo Estudios	2%
	Escuela Primaria Incompleta	3%
	Escuela Primaria Completa	2%
	Escuela Secundaria Incompleta	11%
	Escuela Secundaria Completa	22%
	Estudio Superior No Universitario Incompleto	2%
	Estudio Superior No Universitario Completo	4%
	Estudio Universitario Incompleto	21%
	Estudio Universitario Completo	21%
	Estudios de Posgrados	1%
	No Contesta	11%
Promocionado 7.96%	Estudio Primario Completo	11%
	Estudio Secundario Completo	22%
	Estudio Universitario Completo	44%
	Estudio de Posgrado	11%
	No Contesta	11%

En el caso de la Utilización de las TIC los resultados se expresan en la Tabla 6 y como ejemplo se muestra en la Figura 8 la distribución de porcentajes por atributo para la clase "Promocionado".

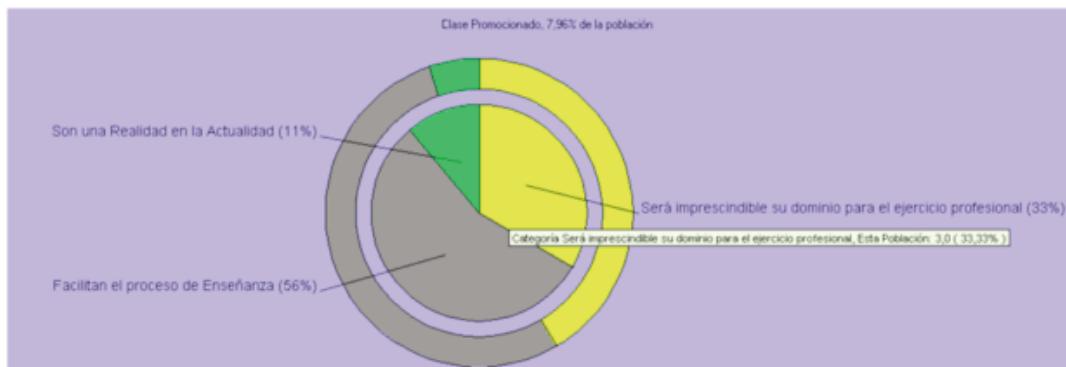


Figura 8. Utilización de las TIC para alumnos Promocionados.

Clase	Atributo	Porcentaje
Regular 10.62%	Son una realidad en la Actualidad	8%
	Facilitan el proceso de Enseñanza	50%
	Será imprescindible su dominio para el ejercicio profesional	42%
Libre 81.42%	Son una realidad en la Actualidad	4%
	Facilitan el proceso de Enseñanza	53%
	Será imprescindible su dominio para el ejercicio profesional	42%
Promocionado 7.96%	Son una realidad en la Actualidad	11%
	Facilitan el proceso de Enseñanza	56%
	Será imprescindible su dominio para el ejercicio profesional	33%

Tabla 6. Caracterización de Clases teniendo en cuenta la Utilización de las TIC.

Por último es importante hacer referencia a la calidad global del modelo utilizado para clasificar la Situación Final del Alumno, la cual resultó ser del 0.944, significando esto que al estimar dicha situación en base a las variables consideradas en el modelo, la estimación resulta correcta en el 94.4% de los casos; un valor de calidad global de un modelo cercano a 0 indica un modelo muy pobre con respecto a la precisión predictiva, fiabilidad y posibilidad de clasificar datos. Un valor de calidad global de un modelo cercano a 1 indica un modelo excelente que clasifica correctamente los datos y los registros y que es fiable al máximo (IBM Knowledge Center, Calidad del Modelo).

6. Conclusiones

Se comenzará con discusiones y comentarios respecto de los principales resultados obtenidos hasta el momento, se indicarán luego las conclusiones propiamente dichas y se finalizará con las principales líneas de trabajo futuras.

Comentarios y Discusiones

Los procesos de minería de datos educacional realizados han producido un considerable volumen de información cuyo detallado estudio insumirá una considerable cantidad de tiempo, no sólo a los integrantes del proyecto de investigación, sino también a otras áreas ya que, tal como se suponía, el rendimiento académico está influenciado por el entorno socio-económico y cultural del alumno y por aspectos actitudinales del mismo respecto del estudio y de la utilización de las TIC.

En los comentarios siguientes se considerará rendimiento académico “alto” al logrado por los alumnos con situación final de “promocionado”, rendimiento “medio” al de los alumnos con situación de “regular” y rendimiento “bajo” al de los alumnos con situación de “libre”; a su vez se considerará “éxito académico” al rendimiento “alto” y “medio” y “fracaso académico” al rendimiento “bajo”.

Seguidamente se comentarán algunos de los aspectos que se considera oportuno destacar.

Teniendo en cuenta el tipo de escuela secundaria de la cual provienen los alumnos (Tabla 1), se observó que para todas las categorías de rendimiento académico la mayoría de los alumnos provienen de escuelas del ámbito Provincial y Municipal, pero con diferencias significativas en los porcentajes según rendimiento académico, alto: 78%, medio: 67%, bajo: 61%. Esto indica que el tipo de escuela secundaria que cursó el alum-

no está relacionado con el rendimiento académico logrado por el mismo, observándose que el porcentaje más alto de participación de escuelas del ámbito Provincial y Municipal (Estatal) corresponde a la categoría de mayor rendimiento académico.

Considerando la cantidad de horas semanales que los alumnos dedicaron al estudio (Tabla 2) se observó que el 56% de quienes han tenido un rendimiento académico alto han dedicado más de 20 hs semanales al estudio, ese porcentaje se reduce al 50% para el rendimiento académico medio y al 46% para el rendimiento académico bajo. Además, el 22% de quienes han tenido un rendimiento académico alto han dedicado hasta 10 hs semanales al estudio, ese porcentaje aumenta al 27% para el rendimiento académico bajo. Esto indica una relación directa entre la dedicación al estudio y el éxito académico.

Considerando la importancia que los alumnos otorgan al estudio (Tabla 3) se observó que el 89% de quienes han tenido un rendimiento académico alto han otorgado más importancia al estudio que a la diversión, ese porcentaje se reduce al 50% para el rendimiento académico medio y al 64% para el rendimiento académico bajo, siendo del 64,6% para el total de la población (Figura 5). Además, el 11% de quienes han tenido un rendimiento académico alto han otorgado más importancia al estudio que al trabajo, ese porcentaje aumenta al 33% para el rendimiento académico medio y al 18% para el rendimiento académico bajo. Esto indica una relación entre el éxito académico y la importancia dada al estudio ante la diversión y el trabajo.

Considerando los últimos estudios de la madre (los de mayor nivel) (Tabla 4) se observó que el 22% de quienes han tenido un rendimiento académico alto tienen madres con estudios de postgrado, ese porcentaje se reduce al 7% para el rendimiento académico bajo, siendo del 7,08% para el total de la población. Además, el 33% de quienes han tenido un rendimiento académico alto son hijos de madres con estudios universitarios completos, ese porcentaje decrece al 25% para el rendimiento académico medio y al 17% para el rendimiento académico bajo. Esto indica una relación entre el éxito académico y el nivel de estudios alcanzado por la madre.

Considerando los últimos estudios del padre (los de mayor nivel) (Tabla 5) se observó que el 11% de quienes han tenido un rendimiento académico alto tienen padres con estudios de postgrado, ese porcentaje se reduce al 1% para el rendimiento académico bajo, siendo del 1,77% para el total de la población. Además, el 44% de quienes han tenido un rendimiento académico alto son hijos de padres con estudios universitarios completos, ese porcentaje decrece al 25% para el rendimiento académico medio y al 21% para el rendimiento académico bajo. Esto indica una relación entre el éxito académico y el nivel de estudios alcanzado por el padre.

Teniendo en cuenta la opinión de los alumnos respecto de la utilización de las TIC (Tabla 6) se observó que el 56% de quienes han tenido un rendimiento académico alto consideraron que las TIC facilitan el proceso de enseñanza, ese porcentaje se reduce al 50% para el rendimiento académico medio, siendo del 53% para el bajo. Además, el 33% de quienes han tenido un rendimiento académico alto consideraron que será imprescindible el dominio de las TIC para el ejercicio profesional, ese porcentaje aumenta al 42% para los rendimientos académicos medio y bajo. Esto estaría indicando que la mayoría de los alumnos con mayor rendimiento académico estarían concentrados más en el proceso de enseñanza – aprendizaje que en el posible ejercicio futuro de la profesión.

Lo indicado precedentemente permite afirmar que el éxito y el fracaso académicos están relacionados con el tipo de escuela secundaria en la que cursó el alumno, la dedicación del alumno medida en horas semanales de estudio, la importancia otorgada al estudio frente a la diversión y al trabajo, el nivel de estudios de los padres y madres y la percepción que de las TIC tienen los alumnos.

Conclusiones

En este artículo se ha presentado un modelo eficiente para la determinación de perfiles de alumnos según rendimiento académico utilizando almacenes de datos y técnicas de minería de datos. Esto permitirá tomar

medidas que tiendan a reducir el fracaso académico, actuando tempranamente con un acompañamiento especial de los alumnos cuyo perfil indique alta probabilidad de fracaso académico.

Claramente, el modelo presentado en este artículo es adecuado para la determinación de perfiles y constituye una herramienta válida de elevada utilidad para la gestión académica. El modelo propuesto puede ser implementado en diversas instituciones.

Líneas de Trabajo Futuras

Actualmente se está trabajando en la evaluación de información previa al cursado del año 2013 y con la correspondiente al cursado del año 2014, para su posterior incorporación al DW y análisis con modelos de minería de datos de clasificación con árboles de decisión y de clusterización; también se comenzará la reunión de la información correspondiente al cursado del año 2015. Asimismo se continuará estudiando la influencia que otras variables tienen sobre el rendimiento académico, tales como la situación laboral de los padres y del propio alumno, la cobertura de salud del alumno, la relación del trabajo con la temática de la carrera, etc.

Agradecimientos

Este trabajo se ha realizado en el marco del proyecto de investigación acreditado "Determinación de perfiles de estudiantes y de rendimiento académico mediante la utilización de minería de datos", código 25/L059 – UTI1719.

Cómo citar este artículo / How to cite this paper

La Red Martínez, D. L., Karanik, M., Giovannini, M., y Pinto, N. (2015). Perfiles de Rendimiento Académico: Un Modelo Basado en Minería de Datos. *Campus Virtuales*, Vol. IV, Num. 1, pp. 12-30. Consultado el [dd/mm/aaaa] en www.revistacampusvirtuales.es

Referencias

- Agrawal, R.; Shafer, J. C. (1996) Parallel Mining of Association Rules. *IEEE Transactions on Knowledge and Data Engineering*, December 1996. USA.
- Ballard, Ch.; Rollins, J.; Ramos, J.; Perkins, A.; Hale, R.; Dorneich, A.; Cas Milner, E. & Chodagam, J. (2007) *Dynamic Warehousing: Data Mining Made Easy*. IBM International Technical Support Organization. IBM Press. USA.
- Berson, A. & Smith, S. J. (1997) *Data Warehouse, Data Mining & OLAP*. Mc Graw Hill. USA.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Renartz, T., Shearer, C., Wirth, R. (1999) *CRISP-DM 1.0. Step-by-step data mining guide*.
- Delfino, J. A. (1989) Los determinantes del aprendizaje. In Petrei, A. H., editor, *Ensayos en economía de la educación*. Educational Evaluation and Policy Analysis.
- Di Gresia, L. (2007) *Rendimiento Académico Universitario*. Tesis Doctoral. Universidad Nacional de La Plata. Argentina.
- Fayyad, U.M.; Grinstein, G. & Wierse, A. (2001) *Information Visualization in Data Mining and Knowledge Discovery*. Morgan Kaufmann. Harcourt Intl.
- Frawley, W., Piatetsky-Shapiro, G. & Matheus, C. (1992) *Knowledge Discovery in Databases: An Overview*. *AI magazine*, 13(3), 57.
- Fazio, M. V. (2004) Incidencia de las horas trabajadas en el rendimiento académico de estudiantes universitarios argentinos. *Documentos de Trabajo UNLP*, 52. Argentina.
- Forteza, J. (1975) Modelo instrumental de las relaciones entre variables motivacionales y rendimiento. *Revista de Psicología General y Aplicada*, 132, 75-91. España.
- García Jiménez, M. V.; Alvarado Izquierdo, J. M.; Jiménez Blanco, A. (2000) La predicción del rendimiento académico: regresión lineal versus regresión logística. *Psicothema* Vol. 12, Supl. n° 2, pp. 248-252. España.
- García, M. M.; San Segundo, M. J. (2001) El Rendimiento Académico en el Primer Curso Universitario. *X Jornadas de la Asociación de Economía de la Educación*. Libro de Actas, págs. 435-445. España.
- González, A. J. (1998) Indicadores del rendimiento escolar: relación entre pruebas objetivas y calificaciones. *Revista de Educación*, 287,

31-54. España.

- Grabmeier, J. & Rudolph, A. (1998) *Techniques of Cluster Algorithms in Data Mining version 2.0*. IBM Deutschland Informationssysteme GmbH. GBIS (Global Business Intelligence Solutions). Germany.
- Gutting, R. (1994) An Introduction to spatial database systems. *VLDB Journal*, 3, 357- 399.
- Hand, D.J.; Mannila, H. & Smyth, P. (2000) *Principles of Data Mining*. The MIT Press. USA.
- Harinarayan V., Rajaraman, A., Ullman, J. (1996) Implementation data cubes efficiently. *ACM SIGMOD Record*, 25 (2), 205 - 216.
- Herrera Clavero, F. et al. (2004) ¿Cómo Interactúan el Autoconcepto y el Rendimiento Académico en un Contexto Educativo Pluricultural?. *Revista Iberoamericana de Educación*. España.
- IBM Software Group. (2003) *Enterprise Data Warehousing whit DB2: The 10 Terabyte TPC-H Benchmark*. IBM Press. USA.
- Inmon, W. H. (1992) *Data Warehouse Performance*. John Wiley & Sons. USA.
- Inmon, W. H. (1996) *Building the Data Warehouse*. John Wiley & Sons. USA.
- Joyanes Aguilar, L. (1997) *Cibersociedad*. Mc Graw Hill. España.
- La Red Martínez, D. L.; Acosta, J. C.; Uribe, V. E.; Rambo, A. R. (2012) Academic Performance: An Approach From Data Mining. *Journal of Systemics, Cybernetics and Informatics*; V. 10 N° 1 2012, págs. 66-72; USA.
- La Red Martínez, D. L.; Acosta, J. C.; Uribe, V. E.; Rambo, A. R.; Cutro, A. L. (2010) *Data Warehouse y Data Mining Aplicados al Estudio del Rendimiento Académico*. CИСCI 2010 (9na. Conferencia Iberoamericana en Sistemas, Cibernética e Informática); Memorias, Volumen I, págs. 289-294; ISBN N° 978-1-934272-94-7; Orlando, Florida. USA.
- La Red Martínez, D. L.; Podestá, C. E. (2014a). *Data Mining to Find Profiles of Students*; Volume 10 – N° 30; *European Scientific Journal (ESJ)*; pp. 23-43; ISSN N° 1857-7881; University of the Azores, Portugal.
- La Red Martínez, D. L.; Podestá, C. E. (2014b). *Metodología de Estudio del Rendimiento Académico Mediante la Minería de Datos*; Volume III – N° 01; *Revista Científica Iberoamericana de Tecnología Educativa - Scientific Journal of Educational Technology*; pp. 56-73; ISSN N° 2255-1514; España.
- La Red Martínez, D. L.; Podestá, C. E. (2014c). *Contributions from Data Mining to Study Academic Performance of Students of a Tertiary Institute*; Volume 02 – N° 9; *American Journal of Educational Research*; pp. 713-726; ISSN N° 2327-6126; U.S.A.
- La Red Martínez, D. L.; Karanik, M.; Giovannini, M. (2014d). *Determinación de Perfiles de Estudiantes y de Rendimiento Académico Mediante la Utilización de Minería de Datos en la UTN – FRRE*. XVI Workshop de Investigadores en Ciencias de la Computación - WICC - 2014; Universidad Nacional de Tierra del Fuego, Antártida e Islas del Atlántico Sur, Ushuaia, Tierra del Fuego, Argentina.
- La Red Martínez, D. L.; Karanik, M.; Giovannini, M.; Pinto, N. (2014e). *Estudio del perfil de rendimiento académico: un abordaje desde Data Warehousing*. 2° Congreso Nacional de Ingeniería Informática / Sistemas de Información - CoNalISI - 2014; ISSN N° 2346-9927; pág. 604-612; Universidad Nacional de San Luis, San Luis, Argentina.
- Maradona, G. & Calderón, M. I. (2004) Una aplicación del enfoque de la función de producción en educación. *Revista de Economía y Estadística*, Universidad Nacional de Córdoba, XLII. Argentina.
- Marcelo García, C.; Villarín Martínez, M.; Bermejo Campos, B. (1987) Contextualización del rendimiento en bachillerato. *Revista de Educación*, 282, 267-283. España.
- Matignon, R. (2009) *Data Mining Using SAS Enterprise Miner*. U.S.A.: Wiley.
- McMahon, W. W. (2002) *Education and Development*. Oxford University Press.
- Mitchell, T. (1997) *Machine Learning*. Mc Graw Hill.
- Poe, V. (1996) *Building a Data Warehouse for Decision Support*. New Jersey: Prentice Hall.
- Podestá Gómez, C. (2013) *El rendimiento académico de los alumnos de la cátedra Sistemas Operativos en la Tecnicatura Superior Analista Programador del Instituto Superior de Curuzú Cuatiá*. Tesis académica de maestría en Ciencias Aplicadas. Universidad Nacional de Pilar, Paraguay.
- Porto, A. & Di Gresia, L. (2000) *Características y rendimiento de estudiantes universitarios. El caso de la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata*. Documentos de Trabajo UNLP, 24.
- Reyes R, S. L. (2004) *El Bajo Rendimiento Académico de los Estudiantes Universitarios. Una Aproximación a sus Causas*. *Revista Theoretikos*. Año VI, N° 18, Enero-Junio, 2004. El Salvador.
- SAS Institute (2009) Disponible en: <http://www.sas.com/technologies/analytics/datamining/miner/semma.html>: Fecha de Consulta: 20/06/2009.
- Simon, A. (1997) *Data Warehouse, Data Mining and OLAP*. John Wiley & Sons. USA.
- 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, CИСCI (Vol. 29, pp. 06-2)*.
- Vivo Molina, J. M.; Franco Nicolás, M.; Sánchez de la Vega, M. del M. (2004) *Estudio del rendimiento académico universitario basado en curvas ROC*. *Revista de Investigación Educativa*. RIE, Vol. 22, N° 2, 2004, págs. 327-340. España.
- White, C. J. (2001) *IBM Enterprise Analytics for the Intelligent e-Business*. IBM Press. USA.
- Widom J. (1995) *Research Problems in data warehousing*. Conf. Information and Knowledge Management, Baltimore. U.S.A.
- Wilson, R. L.; Hardgrave, B. C. (1995) *Predicting graduate student success in an MBA program: Regression versus classification*. *Educational and Psychological Measurement*, 55, 186-195. USA.