

Determinación de perfiles de rendimiento académico utilizando minería de datos en la UTN – FRRe

David L. la Red Martínez, Marcelo Karanik, Mirtha Giovannini, Noelia Pinto

Grupo de Investigación Educativa / Departamento de Ingeniería en Sistemas de Información / Facultad Regional Resistencia / Universidad Tecnológica Nacional
French 414, (3500) Resistencia, Argentina, +54-379-4638194

laredmartinez@gigared.com mkaranik@gmail.com meg_c51@yahoo.com.ar ns.pinto@gmail.com

Resumen

Este proyecto analiza el rendimiento académico de los alumnos de la Cátedra Algoritmos y Estructuras de Datos de la Carrera Ingeniería en Sistemas de Información de la Facultad Regional Resistencia de la Universidad Tecnológica Nacional. La idea principal es la utilización de técnicas de manejo de Almacenes de Datos y Minería de Datos para establecer los aspectos sociales, familiares y de antecedentes académicos que son comunes a los distintos perfiles de rendimiento en dicha Cátedra. Para ello se ha trabajado con los alumnos utilizando encuestas que recopilan la información general necesaria, asociándola con el desempeño en los exámenes parciales. Con este trabajo se pretende obtener modelos no solamente descriptivos y explicativos, sino también modelos predictivos que permitan abordar la problemática de la deserción de los alumnos de manera temprana.

Palabras clave: perfiles; rendimiento académico; almacenes de datos; minería de datos.

Contexto

En búsqueda de obtener respuestas al alto grado de desgranamiento en los primeros años, se propuso el desarrollo de este proyecto de investigación. La idea principal es tratar de responder las inquietudes básicas respecto del bajo desempeño de los alumnos en los primeros años de la carrera de Ingeniería en Sistemas de Información. De esta manera se intenta establecer los perfiles de los alumnos según sus condición social, familiar, ante-

cedentes académicos, relacionándolos con su rendimiento académico.

El presente proyecto inicia su tercer año de ejecución y como se ha mencionado aborda la problemática del rendimiento académico de los alumnos, específicamente, en la cátedra Algoritmos y Estructuras de Datos de la carrera Ingeniería en Sistemas de Información de la Facultad Regional Resistencia (FRRe) de la Universidad Tecnológica Nacional (UTN).

Es un proyecto de investigación aplicada con fuerte componente de análisis de datos y está inserto en el ámbito del Grupo de Investigación Educativa (GIE) de la FRRe. La coordinación está a cargo de docentes investigadores de la FRRe quienes trabajan con los alumnos de primer año de la mencionada carrera de ingeniería.

El proyecto es financiado por la UTN, homologado e inserto en el programa de incentivos del Ministerio de Educación.

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, se busca mediante la revisión frecuente de contenidos, estrategias y métodos de enseñanza, poder garantizar adecuados estándares de calidad que den como resultado la formación de profesionales altamente calificados y útiles para la sociedad.

Claramente, el rendimiento académico es un factor crítico teniendo en cuenta además que, generalmente, 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 de la UTN FRRe, entre ellas Algoritmos y Estructuras 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 que 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 váli-

da, 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 (González, 1988), (Di Gresia, 2007).

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 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 es posible obtener 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.

Líneas de Investigación, Desarrollo e Innovación

Como se mencionó, el ámbito de aplicación del proyecto de determinación de perfiles de alumnos es la UTN FRRe, donde se dicta la carrera de Ingeniería en

Sistemas de Información. Esta carrera tiene varias materias correspondientes al primer nivel, 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 asignaturas 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 resolver una amplia gama de problemas.

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 socioeconómico, 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.

El objetivo general es 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 académico.
- 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 posibilitó 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 cuales 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 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.

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 (La Red Martínez et al., 2014).

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 consideró interesante estudiar los datos con técnicas de minería de datos, tanto descriptivas como predictivas, habiéndose empleado la metodología CRISP-DM (Chapman et al., 1999); 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 - FRRe.

Resultados y Objetivos

Algunas cuestiones observadas en el análisis de los datos se mencionan a continuación:

Teniendo en cuenta el tipo de escuela secundaria de la cual provienen los alumnos, 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. Esto indica que el tipo de escuela secundaria en la que cursó el alumno 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 se observó que quienes han tenido un rendimiento académico alto han dedicado más de 20 hs semanales al estudio. Además, el 22% de quienes han tenido un rendimiento académico alto han dedicado hasta 10 hs semanales al estudio. Esto indica una relación directa entre la dedicación al estudio y el éxito académico.

Considerando la importancia que los alumnos otorgan al estudio se observó que el 89% de quienes han tenido un rendimiento académico alto han otorgado más importancia al estudio que a la diversión.

Considerando los últimos estudios de la madre y del padre se observa influencia positiva en el rendimiento de los alumnos cuando esos estudios son de postgrado.

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 y el nivel de estudios de los padres y madres.

Formación de Recursos Humanos

El equipo de trabajo está integrado por un Director (Doctor, Categoría II P.I., Categoría A UTN), un Co-director (Doctor, Categoría IV P.I., Categoría C UTN), un investigador (Especialista) realizando su tesis de maestría y dos becarios. En

diciembre/2013 fue aprobada en la Universidad Nacional de Pilar, Paraguay, una tesis de maestría relacionada con la temática del proyecto de investigación y dirigida por el Director del mismo, quien además, en la actualidad, dirige una tesis de maestría temáticamente relacionada en la Universidad Nacional del Este, Paraguay y una tesina de grado en la Universidad Nacional del Nordeste, Argentina.

Referencias

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.

Di Gresia, L. (2007) Rendimiento Académico Universitario. Tesis Doctoral. Universidad Nacional de La Plata. Argentina.

Forteza, J. (1975) Modelo instrumental de las relaciones entre variables motivacionales y rendimiento. *Revista de Psicología General y Aplicada*, 132, 75-91. España.

González, A. J. (1998) Indicadores del rendimiento escolar: relación entre pruebas objetivas y calificaciones. *Revista de Educación*, 287, 31-54. España.

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.; Karanik, M.; Giovannini, M.; Pinto, N. (2014). Estudio de Perfil de Rendimiento Académico: Un Abordaje desde Datawarehousing; 2° Congreso Nacional de Ingeniería Informática / Sistemas de Información - CoNaIIISI - 2014; ISSN N° 2346-9927; pág. 604-612; Universidad Nacional de San Luis, San Luis, Argentina.

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.

Mitchell, T. (1997) Machine Learning. Mc Graw Hill.