

Hacia un modelo predictivo de rendimiento académico utilizando minería de datos en la UTN – FRRe

David L. la Red Martínez, Marcelo Karanik, Mirtha Giovannini, Reinaldo Scappini

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 rscappini@gmail.com

Resumen

Durante el cursado de la carrera de Ingeniería en Sistemas de Información en la Facultad Regional Resistencia de la Universidad Tecnológica Nacional (UTN-FRRe), el alumno se enfrenta con la necesidad de cursar y regularizar ciertas materias que le generan restricciones de correlatividad. Ese es el caso de la cátedra de primer año Algoritmos y Estructuras de Datos, cuya regularización es necesaria para cursar varias asignaturas de segundo y tercer año. Tomando como base los resultados del proyecto “Determinación de perfiles de estudiantes y de rendimiento académico mediante la utilización de minería de datos”, 25/L059 - UTI1719, implementado en la mencionada cátedra (2013-2015), se está comenzando un nuevo proyecto que tiene como objetivo tomar como punto de partida el análisis descriptivo (lo que pasó), y utilizar la analítica avanzada, con el objetivo de explicar el por qué, el qué va a pasar y cómo podemos abordarlo. Para el estudio se utilizarán distintas herramientas de Minería de Datos: clustering, redes neuronales, redes bayesianas, árboles de decisión, regresión y series temporales, etc. Estas herramientas permiten obtener resultados desde distintas perspectivas del problema abordado. De esta manera se podrán detectar situaciones problemáticas potenciales al inicio del cursado y tomar las medidas necesarias para solucionarlas.

Palabras clave: rendimiento académico; almacenes de datos; minería de datos; modelos predictivos.

Contexto

En búsqueda de obtener respuestas al alto grado de desgranamiento en los pri-

meros años, se propuso el desarrollo del proyecto “Determinación de perfiles de estudiantes y de rendimiento académico mediante la utilización de minería de datos”, 25/L059 - UTI1719, implementado en la cátedra “Algoritmos y Estructuras de Datos” de la Ingeniería en Sistemas de Información de la UTN – Regional Resistencia (2013-2015). Como consecuencia de ello se establecieron los perfiles de rendimiento académico de los alumnos según su condición social, familiar, antecedentes académicos, etc., relacionándolos con su rendimiento académico en la asignatura mencionada.

Como continuación del proyecto mencionado se ha iniciado el proyecto “Diseño de un modelo predictivo de rendimiento académico mediante la utilización de minería de datos”, UTI3808TC (2016-2018), inserto en el ámbito del Grupo de Investigación Educativa (GIE) de la FRRe. El proyecto es financiado por la UTN, homologado e inserto en el programa de incentivos del Ministerio de Educación.

Introducción

La Universidad enfrenta actualmente el desafío de mejorar su calidad académica enfocándose no sólo en el sistema de enseñanza-aprendizaje, sino contemplando otras variables, como la sistematiza-

ción de procesos de evaluación permanentes (Briand et. al., 1999). Entre estas variables, se destaca el estudio del perfil de rendimiento académico de los estudiantes.

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 (Maletic et. al., 2002).

Generalmente al evaluar el rendimiento académico, se analizan en mayor o menor medida los elementos que influyen en el desempeño como ser, entre otros, factores socioeconómicos, la amplitud de programas de estudio, las metodologías de enseñanza, conocimientos previos del alumno (Marcus, 2003).

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, 2007) 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, 2003).

También 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).

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 mismos. 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.

Surge, entonces, la necesidad de implementar un mecanismo que permita determinar las características propias del estudiante, analizando la existencia de relaciones y patrones de comportamiento estudiantiles que posibilite la definición clara de perfiles de alumnos. Para ello una alternativa es utilizar técnicas de minería de datos para el modelado descriptivo (La Red Martínez et. al., 2014a, 2014b, 2015).

A su vez, el modelado predictivo puede usarse para analizar una base de datos y determinar ciertas características esenciales acerca del conjunto de datos que permitan predecir el comportamiento de alguna variable (Connolly & Begg, 2005).

Líneas de Investigación, Desarrollo e Innovación

En la UTN-FRRe se dicta la carrera de Ingeniería en Sistemas de Información (ISI). Esta carrera tiene varias asignaturas en el primer año que son específicas a la profesión del Ingeniero en Sistemas, las cuales por lo general son las que generan

el mayor desgranamiento. Una de estas asignaturas es Algoritmos y Estructura de Datos.

En el marco del proyecto “Determinación de perfiles de estudiantes y de rendimiento académico mediante la utilización de minería de datos”, se trabajó en la identificación de las variables que explican el desigual rendimiento académico por parte de los estudiantes de la citada asignatura, lográndose modelos descriptivos del rendimiento académico; en el contexto del proyecto “Diseño de un modelo predictivo de rendimiento académico mediante la utilización de minería de datos” iniciado este año se buscará desarrollar modelos predictivos de rendimiento académico.

Teniendo en cuenta los resultados logrados en las evaluaciones realizadas durante el cursado de la asignatura, se buscó determinar en qué medida el desigual rendimiento académico es influenciado por variables socioeconómicas y actitudinales tales como: escuela media de procedencia, nivel educativo de los padres, nivel socioeconómico, edad, género, actitud general hacia el estudio, uso de herramientas de apoyo (campus virtual) (La Red Martínez et. al., 2015).

El universo estuvo constituido por los alumnos en condiciones de cursar la asignatura durante el año 2013, 2014 y 2015 y la unidad de análisis fue cada uno de esos alumnos.

El análisis de los resultados se basó en considerar como parámetro de minería 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 y Promocionado a aquel que aprobó

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

Resultados y Objetivos

Resultados

Se mostrarán a continuación algunos de los resultados del proyecto anterior que constituyen el punto de partida del proyecto que se inicia.

Se han obtenido los siguientes resultados: 81.42% de alumnos en condición de Libre, 10.62% Regular y sólo 7.96% Promocionado.

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, 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 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 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 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. 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) 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) 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 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 y el nivel de estudios de los padres y madres (La Red Martínez et al, 2015b, 2015c).

Objetivo general del nuevo proyecto

Obtener el conocimiento necesario para desarrollar un modelo predictivo para determinar rendimientos académicos potencialmente problemáticos, a partir de la implementación de técnicas de Minería de Datos.

Objetivos específicos del nuevo proyecto

- Utilizar técnicas de Minería de Datos sobre el Almacén de Datos construido con los datos de alumnos relevados durante 2013, 2014 y 2015 y extender el relevamiento de los datos de alumnos durante 2016 y 2017 para obtener conocimiento referido a las variables que predominantemente influyen en el rendimiento académico.

- Analizar y observar el comportamiento del modelo bajo condiciones típicas de rendimiento académico, obteniendo el conocimiento necesario para depurar el modelo predictivo.

- Validar el modelo mediante el seguimiento del rendimiento académico de estudiantes-tipo seleccionados, obteniendo un modelo predictivo confiable.

- Utilizar el modelo predictivo de rendimiento académico desarrollado para predecir la probabilidad de que un estudiante abandone el cursado de una asignatura, dadas sus características socioeconómicas y académicas.

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, un investigador (Magister) y dos becarios. Actualmente se está trabajando en la definición de planes de tesis de maestría con temáticas afines a la del proyecto.

Referencias

Briand, L. C., Daly, J., y Wüst, J. (1999) A unified framework for coupling measurement in object oriented systems, *IEEE Transactions on Software Engineering*, 25, 1.

Connolly, T. y Begg, C. (2005) *Sistemas de bases de datos. Un enfoque práctico para diseño, implementación y gestión*. 4ta Edición. Ed. Pearson Addison Wesley.

La Red Martínez, D.; Karanik, M.; Giovannini, M. Y Pinto, N. (2015) *Academic Performance Profiles: A Descriptive Model Based on Data Mining*; Volume 11 – N° 9; *European Scientific Journal (ESJ)*; pp. 17-38; ISSN N° 1857-7881; University of the Azores, Portugal.

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) *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.

Maletic, J. I., Collard, M. L., y Marcus, A. (2002) *Source Code Files as Structured Documents*, in *Proceedings 10th IEEE International Workshop on Program Comprehension (IWPC'02)*, Paris, France, pp. 289-292.

Maradona, G. y Calderón, M. I. (2007) *Una aplicación del enfoque de la función de producción en educación*. *Revista de Economía y Estadística*, Universidad Nacional de Córdoba, XLII. Argentina.

Marcus, A. (2003) *Semantic Driven Program Analysis*. Kent State University, Kent, OH, USA, Doctoral Thesis.

Porto, A. y Di Gresia, L. (2003) *Características y rendimiento de estudiantes universitarios. El caso de la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata*. Documentos de Trabajo, Universidad Nacional de La Plata.

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, El Salvador.

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.