

Detección de Perfiles de Rendimiento Académico en la Universidad Nacional del Este de Paraguay

David L. la Red Martínez¹, Gabriela Bobadilla²

¹Facultad Regional Resistencia / Universidad Tecnológica Nacional
French 414, (3500) Resistencia, Argentina, +54-379-4638194
laredmartinez@gigared.com

²Facultad Politécnica / Universidad Nacional del Este
Campus Km 8 Acaray, Calle Universidad Nacional del Este y Rca. del Paraguay, Ciudad del Este, Paraguay, 59561-577-261

Resumen

El problema de la deserción temprana en las Universidades se ve agravado cada día más por diferentes causas. La situación académica de pobre rendimiento se ve afectada por múltiples factores que conforman un escenario complejo de análisis. En general no existe un solo aspecto que determine el fracaso de los alumnos, por el contrario, es la interacción de varias circunstancias la que provoca el abandono de las carreras universitarias. Claramente, la situación socio-económica es clave a la hora del análisis y no puede dejarse de lado cuando se intenta estudiar el fenómeno de la deserción. Es importante, por tanto, estudiar y determinar cuáles son las variables que inciden en el rendimiento académico a fin de poder establecer estrategias de acción pedagógicas que permitan mejorar dicho rendimiento. Claramente, la interacción de dichas variables es un problema complejo de abordar con técnicas de análisis tradicionales. Por esto se propone realizar un análisis de los datos utilizando herramientas computacionales y algoritmos adecuados para tal fin.

Este proyecto tiene como propósito intentar establecer los perfiles de rendimiento académico de los alumnos de la Universidad Nacional del Este (UNE) de Ciudad del Este – Paraguay a fin de poder describir la situación de deserción e intentar explicar las causas que motivan el abandono de las carreras que se dictan allí. Para ello, se utilizarán técnicas de minería sobre los datos académicos y socio-económicos de los alumnos, aplicando algoritmos de búsqueda de conocimiento en grandes volúmenes de información. Entre esos algoritmos de Minería de Datos (del inglés Data Mining o DM) se destaca la utilización de redes neuronales, algoritmos genéticos, predicción dinámica, agentes inteligentes, clustering, reglas de

asociación, árboles de decisión, análisis de correlación, análisis semántico, análisis de regresión, entre otros.

Si bien el proyecto se llevará a cabo analizando carreras de la UNE, ha sido desarrollado en el ámbito de la Red de Cooperación Interuniversitaria en TICs del Mercosur (ReCITic) integrada por la Universidad Tecnológica Nacional, Facultad Regional Resistencia (UTN-FRRe, Resistencia, Chaco, Argentina), la Universidad Gastón Darachary (UGD, Posadas, Misiones, Argentina), la Universidade Estadual do Oeste do Paraná (UNIOESTE, Foz do Iguaçu, Brasil) y la Universidad Nacional del Este (UNE, Ciudad del Este, Paraguay). En el proyecto trabajarán docentes investigadores de las mencionadas universidades, lo que lo hace de carácter interdisciplinario e internacional.

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

Contexto

La Universidad Nacional del Este fue creada en el año 1993 por Ley de la Nación Paraguaya N° 250/93. Actualmente cuenta con las Facultades de: Ciencias Agrarias; Filosofía; Politécnica; Ciencias Económicas; Derecho y Ciencias Sociales y la de Ciencias de la Salud. Cuenta con tres Escuelas Superiores y con filiales en varias localidades.

Cuenta con más de 10.000 alumnos distribuidos en las distintas unidades académicas y escuelas. La UNE enfrenta el reto expuesto por (Montero Rojas et al., 2007) de la pro-

blemática de las Instituciones públicas de educación superior, que es el de mejorar su calidad académica con recursos cada vez más escasos, y a la vez, hacer frente a las demandas de los nuevos contextos sociales y económicos de una sociedad globalizada.

En el sondeo exploratorio realizado en la FPUNE (Bobadilla y La Red Martínez, 2015), se muestra la acumulación de alumnos en los primeros semestres de las carreras tecnológicas y el bajo porcentaje de egresados en función al año de cohorte, que llevan a la situación de agotar recursos que debieran estar disponibles para atender la demanda natural generada por los alumnos provenientes de la educación media. A fin de atender el problema de la calidad académica y la escasez de recursos, este proyecto propone identificar las variables que inciden en el rendimiento académico de los alumnos de la UNE. Esto permitirá determinar perfiles de alumnos exitosos (los que promocionan en su cohorte), como así también los perfiles de alumnos que no lo logran (los que quedan desfasados a su cohorte o desertan). Una vez determinados los perfiles de alumnos, se podrán plantear acciones tendientes a evitar potenciales fracasos académicos. Se considera oportuno utilizar las técnicas de Almacenes de Datos (Data Warehouse: DW) y Minería de Datos (Data Mining: DM), para la obtención de los perfiles.

El proyecto “Estudio del rendimiento académico y detección temprana de perfiles de alumnos en la Universidad Nacional del Este de Paraguay, aplicando técnicas de minería de datos” fue presentado en la convocatoria 2015 del CONACYT (Consejo Nacional de Ciencia y Técnica de Paraguay) y fue aprobado por Resolución N° 310 del 22/07/16, comenzando su ejecución en el año 2017.

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 sistematización 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 & 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., 2014, 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).

El presente proyecto intenta establecer las causas del bajo rendimiento académico de los alumnos de la Universidad Nacional del Este del Paraguay (UNE), utilizando técnicas de Minería de Datos. Para ello se propone detectar patrones de deserción estudiantil a partir de los datos socio económicos, académicos, actitudinales e institucionales de los estudiantes universitarios, haciendo uso de técnicas de DM. Inicialmente se utilizarán los datos de los estudiantes que ingresaron en los años anteriores, con el fin de hacerles un seguimiento hasta la actualidad, determinando las características de los casos particulares. Con estos datos se construirá un repositorio de datos que se pre-procesará y se transformará con el fin de obtener un conjunto de datos limpios y listos para aplicarle las técnicas de minería de datos.

Para la etapa de análisis se utilizarán algoritmos de DM a fin de poder construir modelos descriptivos que expliquen las circunstancias que llevan al bajo rendimiento académico que frecuentemente termina en deserción de los alumnos o a un rendimiento académico bueno o muy bueno que generalmente los lleva a concluir sus estudios. Los modelos descriptivos se utilizarán para explicar la

situación de rendimiento académico de los alumnos.

Los resultados serán analizados, evaluados e interpretados para determinar la validez del conocimiento obtenido. De esta manera se requerirá una metodología de trabajo sólida con un alto componente de interdisciplinariedad que brinde resultados consistentes y útiles para la toma de decisiones directivas.

Se estima que este proyecto tendrá un alto impacto en el ámbito académico debido a que se pretende mejorar la situación de bajo rendimiento académico y el desgranamiento en las carreras de la UNE. Además, se considera que, en el ámbito científico-tecnológico, también se obtendrán resultados significativos ya que se pretende generar un modelo sólido de análisis de datos que pueda ser extensible a otros campos de aplicación y contribuir a la formación de recursos humanos en el área de investigación del proyecto.

Líneas de Investigación, Desarrollo e Innovación

En el marco del proyecto “Determinación de perfiles de estudiantes y de rendimiento académico mediante la utilización de minería de datos”, desarrollado en la UTN-FRRe entre los años 2013 y 2015, 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 en el año 2016 en la UTN-FRRe, se buscará desarrollar modelos predictivos de rendimiento académico.

En la tesis de maestría “Estudio del Rendimiento Académico de Estudiantes de Análisis de Sistemas, Ingeniería Eléctrica e Ingeniería de Sistemas, de la Facultad Politécnica de la Universidad Nacional del Este, Paraguay”, realizada por la Ing. Bobadilla con la dirección del Dr. La Red Martínez, presentada y aprobada en el año 2015, se trabajó en la misma línea de investigación desarrollada en los proyectos mencionados precedentemente.

Esta misma línea de investigación se tiene previsto desarrollar y profundizar en el proyecto aprobado por la CONACYT para la UNE.

La minería de datos ha sido utilizada para la búsqueda de perfiles de alumnos con el propósito de identificar potenciales éxitos o fracasos académicos, teniendo en cuenta su rendimiento académico, situación demográfica y socio económico (La Red Martínez et al., 2012), (La Red Martínez et al., 2014).

En una universidad del sudeste de los Estados Unidos, se pudo determinar qué alumnos tienen tendencia a tomar cursos online o no, mediante el uso de DM teniendo en cuenta el perfil de los estudiantes por medio de las inscripciones a cursos, rendimiento escolar y situación demográfica (Chong Ho Yu et al., 2012).

En (Kabakcheva et al., 2011) se explica cómo utilizaron DM para predecir el rendimiento de los alumnos de la Universidad de Bulgaria, basados en datos personales y características pre-universitarias.

En (Brijesh Kumar Baradwaj & Saurabh Pal, 2012) utilizan DM para evaluar el rendimiento de los estudiantes universitarios utilizando árboles de decisión para la clasificación de datos, pudiendo extraer de los mismos el rendimiento del examen final del semestre, lo cual ayudó a reconocer de manera temprana estudiantes que necesitarán apoyo pedagógico.

En (Bobadilla & La Red Martínez, 2015) se utilizó DM para la identificación de variables características de los perfiles de alumnos con distinto rendimiento académico, habiéndose encontrado que el grado educacional de los padres, la actitud general hacia el estudio, el género, la utilización de las Tecnologías de la Información y Comunicación y los promedios generales de sus calificaciones del segundo y primer semestre inciden en la situación final de las calificaciones (predominantemente), en el rendimiento académico de los alumnos de las carreras tecnológicas de la FPUNE.

Resultados y Objetivos

Resultados

El presente proyecto intenta identificar las variables que influyen en el rendimiento académico, de los alumnos de las carreras de la Universidad Nacional del Este del Paraguay (UNE), utilizando técnicas de DM. Para ello se propone detectar patrones de rendimiento estudiantil a partir de los datos socio económicos, académicos, actitudinales e institucionales de los estudiantes universitarios. Para la identificación de los perfiles se utilizarán algoritmos de DM a fin de poder construir modelos descriptivos que expliquen las circunstancias del rendimiento académico que frecuentemente termina en deserción de los alumnos o un rendimiento académico bueno o muy bueno que generalmente los lleva a concluir sus estudios. Los resultados serán analizados, evaluados e interpretados para determinar la validez del conocimiento obtenido. Se estima que este proyecto tendrá un alto impacto en el ámbito de la calidad académica debido a que se pretende mejorar la situación de bajo rendimiento académico y el desgranamiento en las carreras de la UNE. Además, se busca contribuir a la formación de recursos humanos en la línea de investigación en Tecnologías de Datos y Gestión del Conocimiento de la Facultad Politécnica de la Universidad Nacional del Este.

Objetivo general del proyecto

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

Objetivos específicos del proyecto

- Utilizar técnicas de DM 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 Doctor, una Magister y varios maestrandos. Actualmente se está trabajando en la definición de planes de tesis de maestría con temáticas afines a la del proyecto.

Referencias

- Bobadilla, G., La Red Martínez, D. L. (2015) Estudio del rendimiento académico y determinación temprana de perfiles de alumnos en la Facultad Politécnica de la Universidad Nacional del Este de Paraguay; N° 10; FPUNE Scientific; pp. 43-48; ISSN N° 2222-2286; Ciudad del Este, Paraguay.
- Briand, L. C., Daly, J., Wüst, J. (1999) A unified framework for coupling measurement in object oriented systems, IEEE Transactions on Software Engineering, 25, 1.
- Brijesh Kumar Baradwaj, Saurabh Pal (2011) Mining Educational Data to Analyze Student's Performance. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6.
- Chong Ho Yu, Digangi, S., Kay Jannasch-Pennell, A., Kaprolet, C. (2012) Profiling Students Who Take Online Courses Using Data Mining Methods. Online Journal of Distance Learning Administration, Volume XI, Number II, Summer 2008, University of West Georgia, Distance Education Center, U.S.A.
- Connoly, T., Begg, C. (2005) Sistemas de bases de datos. Un enfoque práctico para diseño, implementación y gestión. 4ta Edición. Ed. Pearson Addison Wesley.
- Kabakcheva, D., Stefanova, K., Kisimov, V. (2011) Analyzing University Data for Determining Student Profiles and Predicting Performance. Conference: Proceedings of the 4th International Conference on Educational Data Mining, Eindhoven, The Netherlands, July 6-8.
- La Red Martínez, D. L.; Acosta J. C., Uribe V. E., Rambo A.; (2012) Academic Performance: An Approach From Data Mining; V. 10 N° 1; Journal of Systemics, Cybernetics and Informatics; pp. 66-72; ISSN N° 1690-4524; U.S.A.
- La Red Martínez, D. L.; 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 Ss "Cyril and Methodius" Skopje, Macedonia.
- La Red Martínez, D. L.; Podestá, C. E. (2014) 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 Theorethikos. 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.