

Minería de datos para un Sistema de alerta temprana de deserción en carreras de Ingeniería

Ahumada, Hernán César; Dip, Hugo Rubén; Herrera, Carlos Gabriel; Leguizamón Almendra, Juan Carlos

Departamento de Formación Básica / Facultad de Tecnología y Ciencias Aplicadas / Universidad Nacional de Catamarca
Maximio Victoria 55 – 4700 Catamarca, Teléfonos
{hcahumada, hrdip, cgherrera, daa}@tecno.unca.edu.ar

Resumen

Las carreras de Ingeniería de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca tienen estadísticas de deserción y desgranamiento de magnitud significativa y se producen fundamentalmente en los primeros años de estudios. La mayoría de los análisis descriptivos de este fenómeno revelan demasiado tarde el abandono de estudiantes en diferentes tramos de la carrera.

En este proyecto se busca mediante la aplicación de métodos de minería de datos, caracterizar la trayectoria académica de los estudiantes, con la finalidad de detectar patrones compatibles con situaciones de dificultades en el aprendizaje, que puedan derivar posteriormente en abandono de los estudios universitarios.

La determinación de patrones que describan el rendimiento académico como así también las dificultades prevalentes que se les presentan a los alumnos en la resolución de situaciones prácticas de matemáticas correspondientes al primer año de las carreras de Ingeniería, permitirá adoptar en tiempo y forma,

medidas didácticas e institucionales tendientes a disminuir los índices de deserción y desgranamiento de las carreras de Ingeniería.

Para el desarrollo de los modelos descriptivos y predictivos se aplicarán principalmente técnicas de agrupamiento (clustering), árboles de decisión, reglas de asociación y clasificación, y máquinas de vectores soporte (SVM).

Palabras clave: Minería de datos, Alerta Temprana, Deserción

Contexto

Las líneas de investigación forman parte del proyecto acreditado por SECyT - UNCA. 02/J715 - *Hacia La Búsqueda De Un Sistema De Alerta Temprana De Deserción En Carreras De Ingeniería*. El proyecto se desarrolla en el Departamento de Formación Básica de la Facultad de Tecnología y Ciencias Aplicadas (Universidad Nacional de Catamarca).

Introducción

Las estadísticas de deserción y desgranamiento que se observan en las carreras de Ingeniería de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca, son de magnitud significativa y se producen fundamentalmente en los primeros años de estudios. Si bien la baja tasa de egresados en estas carreras tiene distintas causales (económicas, afectivas, laborales por citar solo algunas), muchos han sido los intentos por tratar de revertir estos indicadores a través de diferentes estrategias, algunas con mayor éxito que otras (tutorías, orientación vocacional, entre otras). La mayoría de estas acciones revelan demasiado tarde el abandono de estudiantes en diferentes tramos de la carrera. En función de esta problemática, se propone, utilizando herramientas informáticas de Minería de Datos Educativos, encontrar patrones entre la gran cantidad de información disponible que permitan establecer medidas didácticas e institucionales a los efectos de poder diseñar un Sistema de Alerta Temprana de Deserción. Se utilizará la información disponible en la historia académica de cada estudiante de las cohortes 2005-2012 como así también la que se logre al evaluar los procedimientos o habilidades generales matemáticas en cada alumno que cursa las asignaturas de primer año de las carreras de Ingeniería (Informática, Electrónica, Agrimensura, Minas) de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca.

La “Minería de Datos” (Witten, & Frank, 2005), es la denominación de una novedosa rama de la ciencia informática que se compone de técnicas avanzadas de procesamiento de datos para extraer conocimiento significativo y hasta entonces desconocido acerca de la parte

de la realidad que una base de datos almacena. Minería de datos involucra técnicas de Aprendizaje Automatizado (Machine Learning) (Alpaydin, 2014), para lograr la construcción de modelos computacionales que permitan realizar descripciones y/o predicciones acerca de los hechos que se almacenan en los diferentes sistemas informáticos. Se encarga además de la preparación de los datos y de la interpretación de los resultados obtenidos.

La Minería de Datos Educativos (Educational Data Mining -EDM), (Romero, Ventura, Pechenizkiy, & Baker, (Eds.) 2010), es una disciplina relacionada con el desarrollo de métodos para extraer información útil a partir de los datos que se generan en los entornos educativos, y utilizarla para mejorar dicho entorno. La información así obtenida se convierte en el insumo indispensable para la toma de decisiones. (Romero, Ventura, Pechenizkiy, & Baker, 2010).

Publicaciones previas han demostrado que la minería de datos se puede utilizar para detectar alumnos en riesgo de deserción. Luan (2002) aplicó exitosamente técnicas de minería de datos para predecir qué grupos de alumnos podrían abandonar los estudios. En un trabajo relacionado, Lin (2012) usó técnicas de minería de datos para optimizar los esfuerzos para retener estudiantes. Para ello se generaron modelos predictivos basados en datos de los estudiantes, que permitieron detectar con gran precisión aquellos alumnos que debían recibir ayuda de los sistemas de retención de alumnos.

Investigadores de la Universidad Estatal de Bowie (Chacon, Spicer, & Valbuena, 2012) desarrollaron un sistema basado en minería de datos que permite a dicha Institución identificar y atender a estudiantes en riesgo de abandono de estudios.

Líneas de Investigación, Desarrollo e Innovación

El proyecto comprende las siguientes líneas de investigación.

1. Minería de datos académicos

En base a registros académicos institucionales (SIU GUARANI) de las cohortes comprendidas entre 2005-2012 y de las cátedras de Álgebra y Geometría Analítica de primer año de las carreras de Ingeniería de la Facultad de Tecnología y Ciencias Aplicadas de la U.N.Ca., se busca mediante la aplicación de métodos de minería de datos, caracterizar la trayectoria académica de los estudiantes, con la finalidad de detectar patrones compatibles con situaciones de dificultades en el aprendizaje, que puedan derivar posteriormente en abandono de los estudios universitarios.

2. Análisis de habilidades matemáticas

En esta línea de investigación se busca analizar y evaluar el desarrollo de habilidades o procedimientos generales matemáticos de alumnos que cursan las asignaturas Álgebra y Geometría Analítica de Primer Año de carreras de ingeniería en la Facultad de Tecnología y Ciencias Aplicadas.

Se desarrollarán las etapas clásicas de la Minería de Datos. Iniciando con la recopilación de datos desde el SIU Guarani (datos de inscripción del alumno, curso de ingreso, actas de regularidades y de exámenes) y de las evaluaciones (parciales y finales) de las cátedras de Álgebra y Geometría Analítica (trabajos prácticos, parciales, etc). Los mismos serán agrupados en un único archivo. Luego se realizará la selección y remoción de registros inconsistentes, y las transformaciones necesarias para el procesamiento posterior. Para el desarrollo de los modelos descriptivos y predictivos se aplicarán principalmente técnicas de agrupamiento (clustering), árboles de decisión, reglas de asociación y clasificación, y máquinas de vectores soporte (SVM). El protocolo de trabajo de modelización incluirá la separación de datos para entrenamiento (75%) y test (25%) siguiendo un esquema de validación cruzada para evitar resultados producto del azar. Se dará luego la etapa de evaluación e interpretación de los patrones obtenidos a fin de caracterizar las trayectorias académicas que pueden derivar en abandonos de estudios. Estos resultados serán dados a conocer a las autoridades de la Facultad para la definición de acciones didácticas e institucionales tendientes a disminuir los índices de deserción y desgranamiento pertinentes. También se publicarán las conclusiones en eventos científicos

Resultados y Objetivos

Aplicando Minería de Datos, se pretende encontrar patrones que puedan describir el rendimiento académico como así también las dificultades más comunes que se les presentan a los alumnos en la resolución de situaciones prácticas de matemáticas correspondientes al primer año de las carreras de Ingeniería.

Formación de Recursos Humanos

El equipo de trabajo actualmente está conformado por 6 docentes, 4 de ellos categorizados: Lic. Hugo R. Dip (III), Dr. Ing. Hernán C. Ahumada (IV), Mgter. Ing. Carlos G.R. Herrera (IV), Lic. María V. Póliche (IV), Prof. María Inés

Cisterna Fernández y Lic. Juan Carlos Leguizamón Almendra. También participan 2 alumnos avanzados de Ingeniería en Informática: Carolina Rodríguez Salemi y Mariano Ubaid.

Hay 2 Tesis de posgrado en curso de las docentes Póliche (Maestría en Informática) y Cisterna Fernández (Especialización En Enseñanza De Ciencias Exactas). Mientras que Rodríguez Salemi desarrolla su Tesina de grado en el contexto del presente proyecto de investigación

Referencias

Alpaydin, E. (2014). *Introduction to machine learning*. MIT press.

Chacon, F., Spicer, D., & Valbuena, A. (2012). Analytics in support of student

retention and success. *Research Bulletin*, 3, 1-9.

Lin, S. H. (2012). Data mining for student retention management. *Journal of Computing Sciences in Colleges*, 27(4), 92-99.

Luan, J. (2002). Data mining and its applications in higher education. *New directions for institutional research*, 2002(113), 17-36.

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. (Eds.). (2010). *Handbook of educational data mining*. CRC Press.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.