

METODOLOGÍA PROPUESTA PARA LA PREDICCIÓN DE DESERCIÓN UNIVERSITARIA MEDIANTE EXPLOTACIÓN DE INFORMACIÓN

Raus, N. A., Vegega, C., Pytel, P., Pollo-Cattaneo, M. F.

Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS)
Universidad Tecnológica Nacional. Facultad Regional Buenos Aires.

Medrano 951 (C1179AAQ) Ciudad Autónoma de Argentina. Buenos Aires Tel +54 11 4867-7511
nicolasraus@gmail.com ,cinthiavg@yahoo.com.ar , ppytel@gmail.com
,fpollo@posgrado.frba.utn.edu.ar

Resumen

Se ha comprobado que la aplicación de procesos de Explotación de Información en el ámbito de la educación tiene resultados sumamente positivos. Por otro lado, en nuestro país la cantidad de graduados es baja, el Sistema Universitario Argentino maximiza la cantidad de estudiantes, pero minimiza la cantidad de graduados; especialmente en las carreras científicas y tecnológicas, que son esenciales para el mundo moderno de la producción. En este contexto, el objetivo principal del presente trabajo es proponer una metodología a través de la cual las Universidades puedan identificar aquellos factores que son determinantes en la deserción universitaria dentro del ámbito de las carreras ingenieriles.

Palabras Clave: *Minería de Datos, Explotación de Información, Deserción Universitaria, Metodología.*

Contexto

Este proyecto articula líneas de trabajo del Grupo de Estudio en Metodologías de

Ingeniería de Software (GEMIS) de la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional (UTN-FRBA).

Introducción

La Inteligencia de Negocio [Morik & Rüping, 2002], se centra en generar conocimiento que contribuya con la generación de planes estratégicos y la toma de decisiones de gestión en las organizaciones, mediante los recursos de información disponibles y el uso de herramientas analíticas y de síntesis; proponiendo un abordaje multidisciplinario, dentro del cual está incluida la Informática [Thomsen, 2003]. La Explotación de Información es la subdisciplina Informática [Langseth & Vivatrat, 2003], que brinda a la Inteligencia de Negocio las herramientas que esta necesita para realizar dicha transformación [Mobasher *et al.*, 1999]. Por lo siguiente, se puede definir a la Explotación de Información como la búsqueda de regularidades importantes y de patrones interesantes en grandes volúmenes de información [Fayad *et al.*,

1996] utilizando, en el marco de un proceso ingenieril, técnicas provenientes de la Inteligencia Artificial [Pollo-Cattaneo *et al.*, 2012]. Incluso, como se menciona en [Britos & García Martínez, 2009], la Minería de Datos se utiliza reiteradas veces como sinónimo de Explotación de información, lo cual implica un error conceptual. La primera brinda las tecnologías y algoritmos que dan soporte a los procesos y metodologías, propias de la Explotación de Información.

Econométrica S.A en su informe [Guadagni, 2011], indica que cada 100 niños que ingresan a primer grado menos de 14 se gradúan en la Universidad. En nuestro país sólo se gradúan 2,4 de cada 1000 habitantes; el Sistema Universitario Argentino maximiza la cantidad de estudiantes, pero minimiza la cantidad de graduados en comparación con sistemas de otros países como por ejemplo Brasil. Esta situación puede observarse principalmente en carreras científicas y tecnológicas, como son las carreras de ingeniería (esenciales en el mundo moderno de la producción). Por otro lado, dicho informe menciona que un estudio realizado por Alberto Porto y publicado por la Universidad Nacional de La Plata (2005) indica que, el principal problema radica en el bajo rendimiento de aquellos estudiantes que no trabajan y dependen de la ayuda familiar. Por el contrario, los estudiantes que dependen de su propio esfuerzo laboral, o de becas, muestran un rendimiento superior en sus estudios.

Actualmente, en las carreras de ingeniería, se gradúan alrededor del 20% de los ingresantes [Departamento de Información Universitaria de la Secretaría de Políticas Universitarias, 2010]. Si bien la cantidad de graduados se encuentra en aumento, lo hace muy lentamente, más aún, en comparación con la demanda, la cual crece de manera exponencial

[Samela, 2011]. Una de las razones fundamentales de esta baja cantidad de egresados, es la deserción de los estudiantes en estas carreras. Por lo tanto, resulta esencial, para intentar solucionar este problema, descubrir las causas de dicha deserción. En este sentido, se han realizado estudios, en el ámbito de la educación, sobre el desempeño académico en determinadas asignaturas o carreras, utilizando procesos de Explotación de Información. Como resultado de dichos estudios se ha comprobado que la eficiencia de la aplicación de estas técnicas resulta sumamente positiva:

-En [Pautsch, 2009] y [Pautsch *et al.*, 2010], se aplicaron técnicas de Minería de Datos para un estudio sobre la deserción estudiantil realizado en la Universidad Nacional de Misiones (Argentina). El objetivo principal de dicho estudio fue maximizar la calidad de clasificación y agrupación, que tienen los modelos de la Facultad de Ciencias Exactas, Químicas y Naturales. Para clasificar y agrupar los estudiantes que desertaron de la Carrera Analista en Sistemas de Computación de dicha facultad (en base a sus características académicas, factores sociales y demográficos). Con las variables que se utilizaron en [Pautsch *et al.*, 2010], se concluyó que si bien se obtenía una muy buena distinción entre los estudiantes activos y pasivos, no se podía afirmar que la información obtenida fuera concluyente para determinar si un alumno puede, o no, desertar de la carrera.

-En [Pautsch *et al.*, 2011], se aplicaron técnicas de Minería de Datos para poder pronosticar alumnos desertores de la Carrera Analista en Sistemas de Computación de la Facultad de Ciencias Exactas, Químicas y Naturales de la Universidad Nacional de Misiones; a partir de patrones académicos, factores

sociales y demográficos, que caracterizaron a los estudiantes. Este trabajo se realizó siguiendo la metodología CRISP-DM [Chapman *et al.*, 2000], y utilizando herramientas open source.

-En [Sposito *et al.*, 2010], con el fin de encontrar un clasificador del rendimiento académico y para detectar los patrones determinantes de la deserción estudiantil, se realizó una investigación aplicando el proceso de descubrimiento de conocimiento sobre los datos de los estudiantes del Departamento de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de La Matanza. No obstante, este trabajo no logró encontrar un clasificador del rendimiento académico y de la deserción estudiantil con un alto grado de precisión y comprensibilidad.

-En [La Red *et al.*, 2010], se aplicaron técnicas de Minería de Datos, basadas en segmentación, a un Data Warehouse con datos socioeconómicos y demográficos de los alumnos de la asignatura Sistemas Operativos, de la Licenciatura en Sistemas de Información de la Facultad de Ciencias Exactas y Naturales y Agrimensura de la Universidad Nacional del Nordeste, con el fin de determinar, a priori, situaciones potenciales de éxito o de fracaso académico. Se obtuvieron diversos modelos referidos a diversos aspectos de los alumnos de la asignatura mencionada, lo que permitió descubrir el perfil de dichos estudiantes. También se pudo comprobar la gran ventaja de la utilización de tecnologías y software de última generación que soporta sistemas multiplataforma.

-En [Raus *et al.*, 2013], se aplicaron procesos de Explotación de Información sobre exámenes finales de la asignatura Sistemas y Organizaciones, del primer año de la carrera Ingeniería en Sistemas de Información de la Facultad Regional

Buenos Aires de la Universidad Tecnológica Nacional. Como resultado se identificaron grupos de estudiantes con características homogéneas, y se centró en aquellos grupos con mayor cantidad de finales desaprobados, o aprobados con baja nota para luego, homogeneizar el dictado de los temas y mejorar el desempeño de los alumnos. En este trabajo se pudo identificar 5 grupos representativos de alumnos, de los cuales 2 grupos se caracterizaban por tener finales desaprobados, o aprobados con baja nota.

-En [Argüello *et al.*, 2013], se aplicó el procedimiento de Explotación de Información CRISP-DM, en la detección de patrones de rendimiento académico de los alumnos de la Universidad Tecnológica Nacional, Facultad Regional Concepción del Uruguay. Este trabajo obtuvo buenos resultados ya que abarcó un gran porcentaje de la totalidad de los casos de alumnos desertores.

A pesar de la existencia de estudios en este dominio, algunos incluso exitosos, actualmente no se sigue una metodología general que sea aplicable a estudios de este tipo, debido a que cada caso se analiza en forma particular y se realiza de manera diferente. Esto conlleva el insumo de gran cantidad de tiempo y trabajo para decidir qué datos recolectar y así también definir los procesos de Explotación de Información que se deban aplicar. En este contexto, es necesario proponer un conjunto de pasos que guíen a los ingenieros en Explotación de Información para realizar dicho tipo de estudio.

Objetivos e Hipótesis de Investigación

Este trabajo se enmarca en una línea de investigación que pretende ser útil a

instituciones universitarias, para identificar las características de los alumnos con altas probabilidades de abandonar sus estudios, o retrasarse considerablemente en ellos, en comparación con el plan ideado por la casas de altos estudios.

Los supuestos (o hipótesis) que guían este trabajo son:

-*Hipótesis I*: Es posible aplicar técnicas de Minería de Datos sobre los datos que posee la Universidad para identificar aquellos factores que son determinantes en la deserción universitaria.

-*Hipótesis II*: Los datos que posee la Universidad serán suficientes para poder aplicar la metodología propuesta.

-*Hipótesis III*: Las características a ser evaluadas para analizar la deserción de alumnos son similares en todas las Universidades, por lo que se podría aplicar el procedimiento propuesto en cualquiera de ellas.

Considerando dichas hipótesis se buscará como *Objetivo General* proponer una metodología, a través de la cual las Universidades puedan identificar aquellos factores que son determinantes en la deserción universitaria para carreras ingenieriles. Con el fin de alcanzar dicho objetivo, se fijarán como *objetivos particulares*:

-Identificar los tipos datos que deben ser recolectados para realizar el estudio.

-Identificar los resultados generalmente esperados en este tipo de estudios.

-Identificar los procesos de Explotación de Información que deban ser aplicados para cada tipo de resultado esperado.

Resultados Esperados

La finalidad de este proyecto es brindar a las Instituciones Universitarias un procedimiento, que permita identificar aquellos factores que son determinantes en la deserción universitaria. Teniendo dichos factores, la Universidad podrá detectar a los alumnos con más probabilidades de abandonar sus estudios, o retrasarse en éstos. Este resultado permitirá enfocarse en dicho grupo de alumnos y, habiendo identificado las razones que podrían llevarlos a desertar, o retrasarse en sus estudios, la institución podrá trabajar sobre esas cuestiones mejorando así la calidad educativa y disminuyendo el porcentaje de deserción universitaria.

Formación de Recursos Humanos

Este proyecto pretende tanto la obtención de nuevos conocimientos, como la mejora en la formación del recurso humano. El grupo de trabajo se encuentra formado por dos investigadores tesistas de doctorado, un investigador tesista de maestría y un investigador en formación.

Por otro lado, la presente propuesta de trabajo ha sido preseleccionada para el programa de intercambio UTN-DAAD (Deutscher Akademischer Austauschdienst o Servicio Alemán de Intercambio Académico en español), en el cual la Universidad Tecnológica Nacional y el Servicio Alemán de Intercambio Académico realizan un aporte significativo al acercamiento científico, tecnológico y cultural entre ambos países [UTN, S/A].

Referencias

- Argüello, M., Domé, T., Gorosito, G. & Ramos, L. (2013) Explotación de Información aplicada en el rendimiento de alumnos de la Universidad Tecnológica Nacional Facultad Regional Concepción del Uruguay. Trabajo realizado en el marco de la asignatura “Tecnologías para la Explotación de Datos” (UTN FRCU) y presentado en *el 1er Congreso Nacional de Ingeniería Informática / Sistemas de Información (CoNaIISI 2013) - Categoría Alumnos*.
- Britos, P. V., & García Martínez, R. (2009). Propuesta de Procesos de Explotación de Información. En *XV Congreso Argentino de Ciencias de la Computación*.
- Chapman P., Clinton J., Keber R., Khabaza T., Reinartz, T., Shearer, C., Wirth, R. “CRISP-DM 1.0 Stepbystep BI guide”. Edited by SPSS. 2000. <http://tinyurl.com/crispdm>
- Departamento de Información Universitaria de la Secretaría de Políticas Universitarias (2010) Anuario de Estadísticas Universitarias Argentina 2010. *Ministerio de Educación de la Nación Argentina. ISSN 1850-7514*. Disponible en: <http://portales.educacion.gov.ar/spu/investigacion-y-estadisticas/anuarios/>
- Fayad, U. M., Piatetsky-Shapiro, G., Smyth, P. & Uhturudsamy, R. (1996). *Advances in Knowledge Discovery and Data Mining*. AAAI Press.
- Guadagni, A. (2011). Deserción, desigualdad y calidad educativa. *Econométrica SA Economic Research and Forecasts, Informe Económico Especial*, (415), 1-20.
- La Red, David, Acosta, Julio; Cutro, Luis; Uribe, Valeria. & Rambo, Alice (2010). Data Warehouse y Data Mining Aplicados al Estudio del Rendimiento Académico. En *Novena Conferencia Iberoamericana en Sistemas, Cibernética e Informática, CISCI 2010, CISCI 2010 (Volumen I, p. 289-294)*.
- Langseth, J. & Vivatrat, N. (2003). Why Proactive Business Intelligence is a Hallmark of the Real Time Enterprise: Outward Bound. *Intelligent Enterprise* 5(18), Pág.34-41.
- Mobasher, B., Cooley, R. & Srivastava, J. (1999). Creating adaptive web sites through usage-based clustering of URLs. *Proceedings Workshop on Knowledge and Data Engineering Exchange*, Pág.19-25.
- Morik, K. & Rüping, S. (2002). A Multistrategy Approach to the Classification of Phases in Business Cycles. *Lecture Notes in Computer Science*, 2430: 307-318.
- Pautsch, J. G. A. (2009). Minería de datos aplicada al análisis de la deserción en la Carrera de Analista en Sistemas de Computación. *Facultad de Ciencias Exactas, Químicas y Naturales*.
- Pautsch, J. G. A., La Red Martínez, D. L., & Cutro, L. A. (2010). Minería de Datos aplicada al análisis de la deserción en la Carrera de Analista en Sistemas de Computación. *Facultad de Ciencias Exactas, Químicas y Naturales*.
- Pautsch, J., Kuna, H., & Godoy, A. E. (2011). Resultados preliminares

- del proceso de minería de datos aplicado al análisis de la deserción en carreras de informática utilizando herramientas open source. En *XVII Congreso Argentino de Ciencias de la Computación*.
- Pollo-Cattaneo, M., García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Rodríguez, D., Merlino, H., Pytel, P., Vanrell, J. 2012. Elementos para una Ingeniería de Explotación de Información. *Proyecciones 10(1): 67-84. ISSN 1667-8400*.
- Raus, N. A., Lujan, F. N., Deroche, A., Vegega, C., Pytel, P., Pollo-Cattaneo, M. F. (2013) Aplicación del Proceso de Ponderación de Reglas de Pertenencia a Grupos en Evaluaciones Finales en Carreras de Grado. *Memorias del 1er Congreso Nacional de Ingeniería Informática / Sistemas de Información (CoNaIISI 2013). ID 22-440-1-DR. ISSN: 2346-9927*.
- Samela, G. (2011) Estudiantes de Ingeniería: alta deserción al final de la carrera. *Diario Clarín.com, sección Economía*. Disponible en: http://www.ieco.clarin.com/economia/Estudiantes-Ingenieria-desercion-final-carrera_0_497950435.html
Disponible online marzo del 2014.
- 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. En *Novena Conferencia Iberoamericana en Sistemas, Cibernética e Informática, CISCI* (Vol. 29, pp. 06-2).
- Thomsen, E. (2003). BI's Promised Land. *Intelligent Enterprise*, 6(4), Pág. 21-25.
- UTN (S/A). Programa de Intercambio UTN-DAAD.
http://www.utn.edu.ar/programasalemania/05_internacionales.utn
Disponible online marzo del 2014.