

Hacia un modelo de deserción universitaria: Análisis exploratorio de variables socioeconómicas y académicas de alumnos de primer y segundo año de la carrera Ingeniería en Sistemas de Información, Universidad Tecnológica Nacional Facultad Regional Rosario.

Luciano Valía, Juan Miguel Moine, José Rostagno,
José Milán, Claudio Arfuso, Cristian Bigatti,
Fabiana María Riva, Eduardo Amar

*Departamento de Ingeniería en Sistemas de Información
Facultad Regional Rosario
Universidad Tecnológica Nacional
E. Zeballos 1342, 2000 Rosario, Argentina*

*lucianovalia@hotmail.com, juanmiguelmoine@gmail.com, joserostagno@hotmail.com,
josmilan@gmail.com, arfuso.claudio@gmail.com, cristianbigatti@gmail.com,
fabianamriva@gmail.com, eduardoamar000@gmail.com*

Abstract

Actualmente la deserción de alumnos es uno de los problemas más importantes que enfrentan las Universidades de nuestro país. La minería de datos es una disciplina que puede colaborar con esta problemática, permitiendo detectar patrones de comportamiento en grandes volúmenes de información. Este trabajo representa la primera etapa de un estudio cuyo objetivo es el descubrimiento de patrones de deserción universitaria en la carrera Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional Facultad Regional Rosario.

Palabras clave: deserción universitaria, minería de datos, estudiantes universitarios, patrones de comportamiento

1. Introducción

Se estima que en la Argentina alrededor del 80 % de los alumnos que ingresan a la universidad desertan, mientras que en países más desarrollados, como en Estados Unidos, esta cifra desciende al 50 %[1]. La carrera de Ingeniería en Sistemas de Información en la Universidad Tecnológica Nacional

Facultad Regional Rosario, al igual que la mayoría de las carreras universitarias de nuestro país, no está exenta a este problema.

Los sistemas de información en la Universidad almacenan día a día una gran cantidad de datos sobre el estado académico del alumno, los cuales pueden ser explorados por disciplinas como la Minería de Datos para obtener nuevo conocimiento. Minería de Datos o Explotación de Información, es el proceso de extraer conocimiento útil, comprensible y novedoso de grandes volúmenes de datos, siendo su principal objetivo encontrar información oculta o implícita que no es posible obtener mediante métodos estadísticos convencionales. El resultado del proceso es un conjunto de patrones (modelos), los cuales pueden ser convertidos en información valiosa para la toma de decisiones [2]. El problema de deserción universitaria ha sido analizado con métodos de minería de datos a partir de registros académicos, tomando variables como el grado de avance en la carrera y el promedio de calificaciones[3, 4] o sólo trabajando exclusivamente con variables socioeconómicas como el ingreso familiar, horas de trabajo y edad[5, 6, 7]. La mayor parte de los estudios que utilizan simultáneamente variables académicas y socioeconómicas fueron realizados con información proveniente del sistema de gestión

académica, señalando sus autores la mala calidad de la misma, especialmente aquella referida a aspectos sociales y económicos[8, 9].

En este trabajo se lleva a cabo un análisis exploratorio de datos que no sólo integra ambos tipos de variables, sino que utiliza otras fuentes de información (como encuestas personales), permitiendo de esta forma la generación de modelos más precisos.

Los resultados de este estudio representan la etapa inicial de un proyecto de investigación cuyo objetivo es la identificación de patrones de deserción en alumnos de los primeros años de la carrera Ingeniería en Sistemas de Información en la Universidad Tecnológica Nacional Facultad Regional Rosario.

En el presente trabajo la meta es encontrar relaciones entre variables socioeconómicas y académicas que permitan describir los diferentes perfiles de los alumnos que cursan sus primeros años en la carrera Ingeniería en Sistemas de Información, en la Facultad Regional Rosario de la Universidad Tecnológica Nacional.

2. Metodología

Para realizar el estudio se trabajó con una muestra de 168 alumnos, los cuales se encontraban cursando primer y segundo año en la carrera Ingeniería en Sistemas de Información, en la Universidad Tecnológica Nacional Facultad Regional Rosario. Los mismos corresponden a las cohortes 2015 y 2014 respectivamente.

Los datos fueron recolectados mediante encuestas presenciales. Las variables que se han evaluado son las que se muestran en el Cuadro 1.

Para el análisis se utilizó el software RapidMiner Studio, empleando técnicas descriptivas estadísticas univariantes y multivariantes.

En una etapa posterior se espera recolectar información acerca de la permanencia del alumno en la carrera, para comenzar la búsqueda de patrones de deserción universitaria.

3. Resultados

El análisis exploratorio realizado sobre las encuestas ha evidenciado relaciones interesantes entre las variables en estudio.

El 80 % de los estudiantes son de género masculino y en su gran mayoría viven en Rosario (84 %). Generalmente viven con sus familias, especialmente aquellos que no residen en la ciudad (Figura 2).

Variable	Tipo	Posibles Valores
Año de cursado	Cualitativa	1° o 2° año
Turno	Cualitativa	Mañana/Tarde/Noche/Más de uno
Sexo	Cualitativa	Masculino/Femenino
Edad	Númerica	Entero positivo
Tiene hijos	Booleana	Si/No
Tiene secundaria técnica	Booleana	Si/No
Debe materias de la secundaria	Booleana	Si/No
Cursó previamente alguna carrera Terciana o universitaria	Booleana	Si/No
Máximo nivel de educación alcanzado por sus padres	Cualitativa	Ninguno/Primaria/Secundario/Terciario/Universitario
Nació en Rosario	Booleana	Si/No
Vive en Rosario	Booleana	Si/No
Con quién vive	Cualitativa	Familia/Solo/Con Amigos/Pareja
Recibe incentivo del estado	Cualitativa	No/Beca/Plan Social
Trabaja	Booleana	Si/No
Cantidad de horas semanales que trabaja	Cualitativa	Menos de 12 hs, entre 12 y 40 hs, más de 40 hs.
Empleo del rubro informático	Booleana	Si/No
Cantidad de horas que estudia	Númerica	
Cantidad de materias aprobadas	Númerica	Entero positivo
Asiste consulta regularmente	Booleana	Si/No
Encuentra interesante la carrera	Booleana	Si/No
Piensa que la carrera tiene salida laboral	Booleana	Si/No

Cuadro 1. Variables identificadas

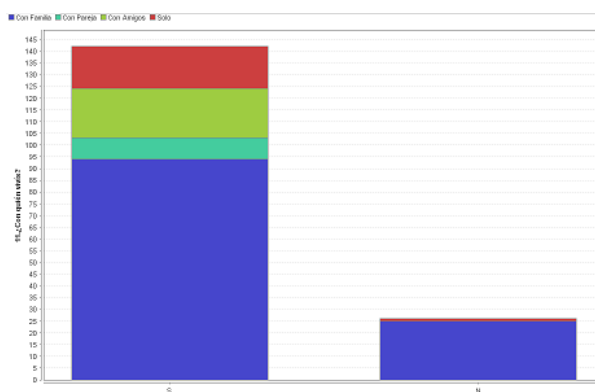


Figura 2. Vive en Rosario vs con quién vive

Sólo el 27 % cursó sus estudios secundarios en una escuela técnica. El desempeño académico del alumno (como cantidad de materias aprobadas) no resulta significativamente influenciado por esta variable.

Respecto a la **situación laboral**, los resultados han indicado que sólo el 20 % de los alumnos trabaja y la mayor proporción lo hace entre 12 y 40 horas semanales. Este grupo de alumnos generalmente asiste al turno noche y la mayoría de sus empleos no están relacionados con el rubro informático. A su vez se detectó que los alumnos que trabajan por lo general nacieron en Rosario (Figura 3), situación que permite suponer que aquellos estudiantes que provienen de otra localidad se dedican exclusivamente a estudiar.

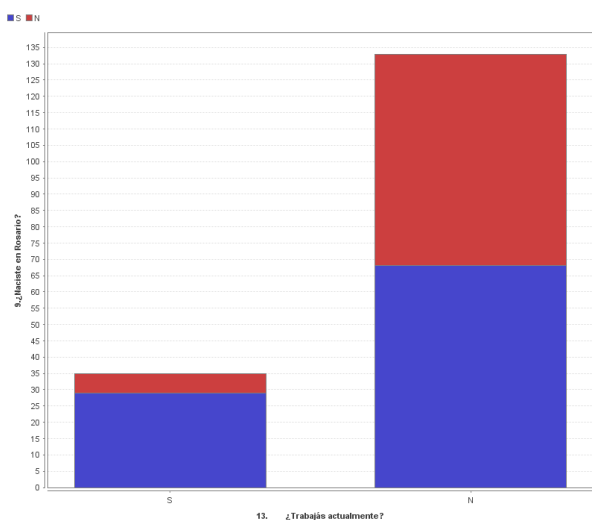


Figura 3. Trabajas actualmente vs Nació en Rosario

En cuanto a los **hábitos de estudio**, los alumnos estudian 2.5 hs en promedio aparte del horario de cursado. Este valor es superior cuando provienen de otra localidad (Figura 4), viven solos o cursan 2º año. El horario de cursado, condición laboral, edad o formación secundaria no influye en esta variable. Asisten a consulta en forma regular sólo el 13 % de los alumnos. Los resultados indican que aquellos que provienen de otra localidad o bien no trabajan, suelen asistir con mayor frecuencia.

En cuanto a la **situación académica**, sólo pudo ser evaluada en 2º año ya que al momento de recolectar los datos los alumnos de primer año no presentaban materias rendidas. Analizando los datos detectamos que aquellos alumnos cuyos padres cuentan con estudios Universitarios/Terciarios presentan una mayor

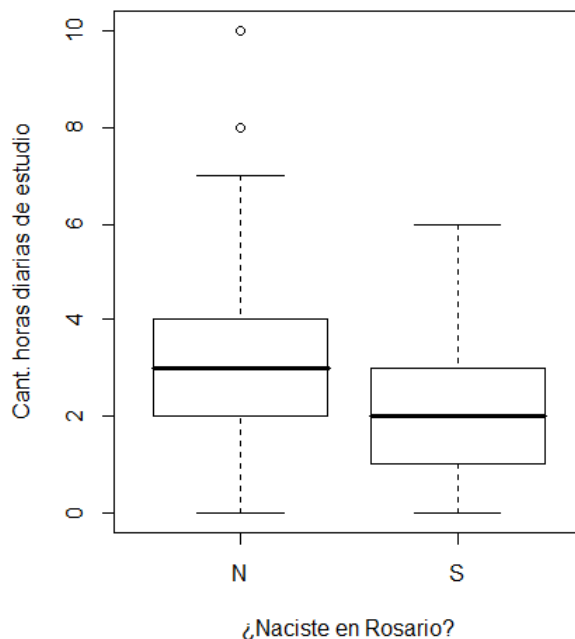


Figura 4. Nació en Rosario vs Cantidad de horas de estudio

cantidad de materias aprobadas (Figura 5). Esta variable también varía significativamente entre los diferentes turnos de cursado.

Por otra parte aquellos que reciben una beca presentan un desempeño inferior al resto de los alumnos (35 % menor) (Figura 6) , mientras que aquellos que cursaron otra carrera anteriormente tienen un 30 % más de materias aprobadas que el resto.

Si analizamos el lugar donde viven y los turnos que cursan, observamos que todos aquellos que no viven en Rosario cursan en un único turno. La situación laboral no influye significativamente en la cantidad de materias aprobadas.

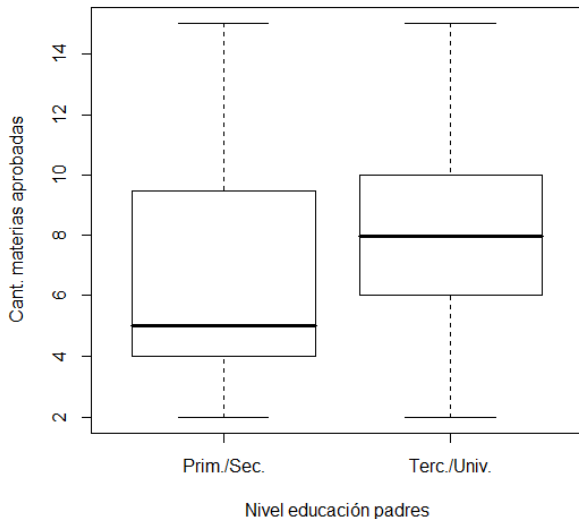


Figura 5. Nivel de estudio padres vs Cantidad materias aprobadas

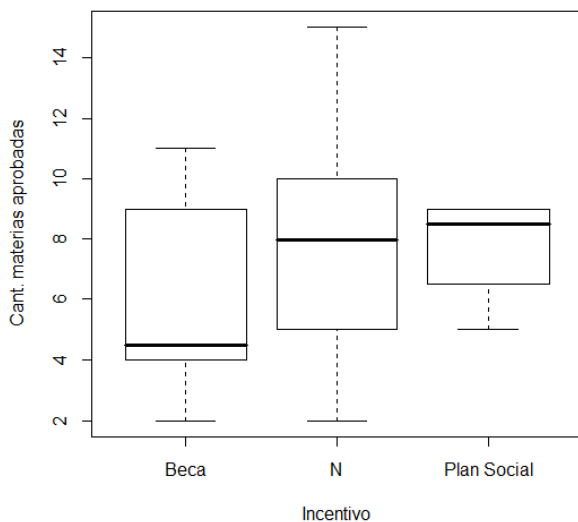


Figura 6. Recibe incentivo vs Cantidad materias aprobadas

4. Conclusiones

Por lo analizado podemos ver que es claro que los hábitos de estudio, son diferentes entre los alumnos de 1º y 2º año. Los primeros estudian menos horas y asisten menos a consulta.

La variable “Nació en Rosario” es muy importante, ya que los alumnos que son de otra localidad suelen tener hábitos diferentes: en general no trabajan, estudian más horas y asisten frecuentemente a consulta.

También se encontró un patrón interesante en los alumnos que tienen becas, donde sus rendimientos son inferiores a quienes no la tienen o reciben subsidio del Estado.

Además, se pudo observar que el desempeño académico (cantidad de materias aprobadas) varía según el nivel de educación de los padres, y no resulta significativamente influenciado por la situación laboral o el tipo de estudio secundario del alumno.

En base a los resultados obtenidos podemos ver la importancia de la inclusión de variables socio-económicas y las relaciones entre las mismas, en virtud de lo que enriquecen el análisis preliminar y el potencial que pueden llegar a tener respecto de las próximas etapas del estudio de búsqueda de patrones en lo que respecta a deserción universitaria.

5. Reconocimientos

Este trabajo fue realizado en el marco del Proyecto de Investigación y Desarrollo: Modelo de Minería de datos para la identificación de perfiles de deserción universitaria en alumnos de ISI (PID-UTN 3802).

Referencias

- [1] C. Lorenzano, “La deserción universitaria en la universidad de tres de febrero,” <http://www.untref.edu.ar/documentos/AutoevaluacionLadesercion.pdf> (disponible al 2/3/2015).
- [2] H. Orallo, J. RAMIREZ, C. R. QUINTANA, M. J. H. Orallo, M. J. R. Quintana, C. F. Ramírez, H. W. Boyd, S. F. W. Stasch, F. Gould, C. GD Schmidt, *et al.*, *Introducción a la Minería de Datos*. Pearson Prentice Hall., 2004.
- [3] C. Balestieri, “Modelo neuronal para la estimación del riesgo de deserción de alumnos de grado,” 2013.
- [4] S. T. Karamouzis y A. Vrettos, “An artificial neural network for predicting student graduation outcomes,” in *Proceedings of the World Congress on Engineering and Computer Science*, pp. 991–994, Citeseer, 2008.
- [5] S. Formia, *Evaluación de técnicas de Extracción de Conocimiento en Bases de Datos y su aplicación a la deserción de alumnos universitarios*. PhD thesis, Facultad de Informática, 2012.

- [6] P. I. Giovagnoli, *Determinantes de la deserción y graduación universitaria*. PhD thesis, Facultad de Ciencias Económicas, 2001.
- [7] S. V. Orea, A. S. Vargas, y M. G. Alonso, “Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos,” *Ene*, vol. 779, no. 73, p. 33, 2005.
- [8] R. T. Pereira, A. C. Romero, y J. J. Toledo, “Descubrimiento de perfiles de deserción estudiantil con técnicas de minería de datos,” *Vínculos*, vol. 10, no. 1, pp. 373–383, 2013.
- [9] F. Angúlo y E. Sergio, “Modelo para la automatización del proceso de determinación de riesgo de deserción en alumnos universitarios,” 2012.