

Minería de Datos y una Aplicación en la Educación Superior

Mag. Graciela Elida Beguerí, Mag. María Alejandra Malberti

Instituto de Informática / Departamento de Informática / Facultad de Ciencias Exactas
Físicas y Naturales / Universidad Nacional de San Juan

Av. Ignacio de la Roza 590 (O), Complejo Universitario "Islas Malvinas", Rivadavia, San Juan, Teléfonos:
4260353, 4260355 Fax 0264-4234980, Sitio Web: <http://www.exactas.unsj.edu.ar>
grabeda ,amalberti @gmail.com

RESUMEN

Se propone abordar la problemática que afecta a la educación en general y en particular a la universitaria, en cuanto a deserción, rezago y abandono. La idea, si bien ambiciosa, es contemplar dos líneas de acción. Por un lado las características de los alumnos, extrayendo conocimiento a partir del análisis de instrumentos propios por medio de algoritmos de aprendizaje de máquina, para contribuir con las acciones de apoyo de los tutores asignados. Por otro lado, el análisis y revisión curricular de las carreras de grado del Departamento de Informática-Facultad de Ciencias Exactas, Físicas y Naturales- Universidad Nacional de San Juan DI- FCFEUN-UNSJ con el fin de vincular las áreas disciplinares e incorporar, integrar y explotar simultáneamente distintas herramientas libres, provenientes de los aportes de las Tecnologías de la Información y la Comunicación –TIC-.

Palabras clave: Desempeño Académico, Aprendizaje Automático, TIC, Software Libre.

CONTEXTO

La línea de investigación se enmarca en el proyecto bianual 2016-2017 “La Ciencia de Datos en grandes colecciones de datos” código E1014 que se desarrolla en el ámbito de la FCFEUN-UNSJ, aprobado por el Consejo de Investigaciones Científicas Técnicas y de Creación Artística (CICITCA), financiado por la propia Universidad y ajustado a evaluación externa.

En el ámbito de la misma se pretende dar continuidad a los proyectos “Minería de Datos en la Determinación de Patrones de Uso y Perfiles de Usuario” y “Búsqueda de Conocimientos en Datos Masivos” llevados adelante a partir de 2012.

Es idea del grupo para la presente propuesta, profundizar el conocimiento y utilización de algoritmos específicos de aprendizaje de máquina que permitan la determinación de perfiles de alumnos universitarios, utilizando técnicas de caracterización multidimensional. Respecto al abordaje de problemas actitudinales se integran a la presente propuesta, participantes del proyecto “El comportamiento académico de los alumnos de primer año de las carreras de informática de la FCFEUN de la UNSJ Estrategias para mejorar su rendimiento” desarrollado en el periodo 2011-2013, Cod. 21/E875 res037/11CS, quienes han realizado experiencias teóricas y de campo en la difícil tarea de la caracterización de comportamientos de alumnos, detectando una compleja confluencia de factores que tienen que ver con la personalidad, aspectos vocacionales y su situación social, económica y familiar.

El estudio realizado por Clara Patricia Buentello y un equipo de colaboradores en 2013, titulado “Deserción Escolar, Factores que Determinan el Abandono de la Carrera Profesional” Estrategias y Condiciones para el Desarrollo del Estudiante, demostró que el principal factor es la distancia que existe entre las expectativas personales y las metas de la institución, así como los planes de estudio, el

tipo de docentes que son considerados como tradicionalistas y rígidos.

1. INTRODUCCIÓN

Algunos modelos psicológicos relativos a la deserción señalan que existen diferencias entre los alumnos que terminan sus estudios universitarios respecto de aquellos que no logran finalizarlos. La diferencia se halla en los rasgos de la personalidad. La teoría de la Acción Razonada, desarrollada por Ajzen y Fishbein (1975), explica que el debilitamiento de las intenciones iniciales de los estudiantes en pos de concluir sus estudios es -en rasgos generales- lo que provoca la deserción. Esta teoría de la Acción Razonada es considerada un predictor (Breva Asensio y Carpi Ballester, 2002), un juicio probabilístico en el que puede estimarse la intención de la conducta directamente mediante una escala de probabilidad (Reyes Rodríguez, 2007).

Gabriela Edith Huesca Ramírez y Belén Castaño Corvo, en “Causas de Deserción de Alumnos de Primeros Semestres de una Universidad Privada” citan la declaración de Vincent Tinto: *“Es importante reconocer que ciertas personas deciden a veces abandonar las instituciones de educación superior, sin alcanzar la graduación, sencillamente porque no tienen la meta o la intención de permanecer en ella hasta lograr esa graduación”*. También señalan que Romo y Fresán (Romo y Fresan, 2001) identifican como elementos explicativos de las dificultades para integrarse al medio académico y social de la institución: la falta de personalidad y madurez intelectual del estudiante, así como la falta de conocimientos y habilidades previas necesarias para realizar estudios superiores.

En la actualidad habilidades de aprendizaje de máquina son fuertemente requeridas en el mercado laboral, a raíz de que las empresas tienden cada vez más a construir sistemas de decisión automatizados. Esto conlleva a promover la incorporación en las currículas de las carreras dependientes del DI, de los saberes requeridos para la formación de

habilidades inherentes a un científico de datos. Datos concernientes a la identificación de las causas potenciales y los principales rasgos de personalidad que puedan interpretarse como riesgos para el buen desempeño académico serán relevados, en particular, entre los alumnos ingresantes y del primer año de las carreras del DI-FCEFNU-UNSJ, Para el tratamiento y análisis de los datos, se prevé utilizar, entre otras herramientas libres, Weka, R y RapidMiner.

- Weka: Waikato Environment Knowledge Analysis es un desarrollo en Java de la Universidad de Waikato, Hamilton, Nueva Zelanda desde 1993. Se descarga de <http://www.cs.waikato.ac.nz/~ml/weka/>
- R: (librería Rattle): desarrollado en la Universidad de Auckland, Nueva Zelanda, en 1993 y la librería Rattle, creada por el Dr. Graham Williams. www.r-project.org
- RapidMiner: (originalmente YALE) versión: 5.3.15, desarrollado, en Java (plataforma Eclipse) en la Universidad de Dortmund, Alemania, en lenguaje Java, desde 2001 <http://sourceforge.net/projects/rapidminer/>

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

En el marco del proyecto se pretende vincular las áreas disciplinares e incorporar, integrar y explotar simultáneamente distintas herramientas libres, provenientes de los aportes de las Tecnologías de la Información y la Comunicación –TIC

También se procura:

- Analizar y describir el conjunto de estudios y prácticas requeridos en Ciencia de Datos.
- Construir y validar instrumentos tendientes a recabar datos inherentes a los estudiantes.
- Estudiar y analizar diferentes conjuntos de datos masivos a procesar. Evaluar herramientas de software libre para arquitecturas secuenciales, paralelas y distribuidas.

3. RESULTADOS OBTENIDOS Y ESPERADOS

Con el propósito de analizar las posibles relaciones existentes entre desempeño académico y tipos de personalidad, se aplica Minería de Datos a datos de 112 alumnos de una asignatura de primer año de las carreras dependientes del Departamento de Informática de la FCEF. Estos datos contienen, entre otros, la situación de cada alumno en la asignatura (Aprobado, Reprobado, Abandono, Promocionó) y los resultados de encuestas que caracterizan su personalidad en 9 tipos (Vargas, 2015)

La deserción y el rezago son problemáticas que preocupan cada vez más a autoridades y a docentes de las instituciones de educación superior. Con el propósito de generar acciones tendientes a prevenir situaciones que alteren el “normal” desempeño académico, se emplean algoritmos de clasificación para descubrir posibles relaciones entre las tendencias actitudinales según el tipo de personalidad y el desempeño académico en las carreras tratadas. Dada una colección de registros, en el que cada registro contiene un conjunto de atributos, siendo uno de ellos la clase, los algoritmos de clasificación hallan un modelo para el atributo de clase como una función de los valores de otros atributos. En este caso el atributo de clase es la situación académica del alumno después de cursar la asignatura, mientras que los restantes atributos tienen que ver con la caracterización del tipo de personalidad (desde 1 hasta 9).

Al momento de la elaboración de este documento, se ha obtenido, usando algoritmo J48 de Weka, lo expuesto en la figura 1. En el modelo se muestra la posible relación existente entre el tipo de personalidad y situación académica, plausible de ser validada análogamente en otros ambientes.

```

=== Run information ===
Scheme:      weka.classifiers.trees.J48 -U -M 2
Relation:    DataTable-weka.filters.unsupervised.attribute.Remove-R1-10,12-21,23-296
Instances:   112
Attributes:  2
             Estado
             Dimension tipo
Test mode:   10-fold cross-validation
=== Classifier model (full training set) ===
J48 unpruned tree
-----
Dimension tipo = 1: REPROBADO (12.0/6.0)
Dimension tipo = 2: REPROBADO (10.0/4.0)
Dimension tipo = 3: REPROBADO (13.0/6.0)
Dimension tipo = 4: APROBO-BOLETA (2.0/1.0)
Dimension tipo = 5: PROMOCIONA (19.0/12.0)
Dimension tipo = 6: REPROBADO (14.0/5.0)
Dimension tipo = 7: REPROBADO (24.0/9.0)
Dimension tipo = 8: REPROBADO (8.0/2.0)
Dimension tipo = 9: REPROBADO (10.0/5.0)

```

Figura 1- Modelo de Clasificación generado por el algoritmo J48.

Algunos de los resultados esperados son:
 Sugerir un conjunto de saberes convenientes para la formación de un científico de datos, para que los mismos sean incorporados, según corresponda, en las currículas de las carreras pertenecientes al DI-FCEF-UNSJ.
 Realizar aportes con los datos tratados en las diferentes aéreas del saber.
 Encontrar perfiles de alumnos universitarios utilizando técnicas de caracterización multidimensional que oriente en la posibilidad de analizar la existencia o no, de relación entre tipos de personalidad, conducta y desempeño académico y desde allí contribuir con acciones de apoyo.
 Obtener una segmentación con el listado de los alumnos, según el riesgo en la regularidad académica.
 Caracterizar a los estudiantes de acuerdo al tipo de personalidad o en términos de posibles dimensiones.
 Contribuir, ante una situación de contingencia, con el trabajo del tutor en el área de tutorías de carreras.

4. FORMACIÓN DE RECURSOS HUMANOS

Los miembros del equipo de investigación, son docentes pertenecientes a las áreas curriculares de Ciencias Básicas, Algoritmos y Lenguajes, Teoría de la Computación e Ingeniería de Software-Base de Datos- Sistemas de Información, de las carreras del

DI, e integrantes de sucesivos proyectos de investigación en el área temática de minería de datos, desarrollados en el ámbito de la FCFN. Se destaca también que para el proyecto en ejecución se cuenta con tres adscriptos: una Lic. en Psicología y dos alumnos avanzados.

5. BIBLIOGRAFÍA

- Buentello Martínez, Clara Patricia "Deserción escolar, factores que determinan el abandono de la carrera profesional" Estrategias y condiciones para el desarrollo del estudiante. XVI Congreso Internacional sobre Innovaciones en Docencia e Investigación en Ciencia.
- Kargupta, H., Han, J., Philip, S. Y., Motwani, R., & Kumar, V. (Eds.). Next generation of data mining. CRC Press. (2008).
- Merlino, A., Ayllón, S., & Escanés, G. (2011). Variables que influyen en la deserción de estudiantes universitarios de primer año. Construcción de índices de riesgo de abandono/Variables that influence first year university students' dropout rates. Construction of dropout risk indexes. *Actualidades Investigativas en Educación*, 11(2).
- Peña, Daniel (2002), *Análisis de datos multivariantes*. Ed. Mc. Graw Hill, España.
- R <http://www.r-project.org/>
- Rapid-I . <http://rapid-i.com/api/rapidminer-5.1/com/rapidminer/tools>. 2011.
- Ramírez, G., & Corvo, M. (2007). Causas de deserción de alumnos de primeros semestres de una universidad privada. *Revista Mexicana de Orientación Educativa*, 5(12), 34-39.
- Vargas, A. (2015). Eneagrama (Nueva edición):¿ Quién soy?. Alamah.