

# Analítica del Aprendizaje y la personalización de la Educación

Javier Díaz<sup>1</sup>, Laura Lanzarini<sup>2</sup>, M. Emilia Charnelli<sup>1</sup>, Guillermo Baldino<sup>3</sup>,  
Alejandra Schiavoni<sup>1</sup>, Paola Amadeo<sup>1</sup>

<sup>1</sup>Laboratorio de Investigación en Nuevas Tecnologías Informáticas.  
Facultad de Informática. Universidad Nacional de La Plata  
{javierd, ales}@info.unlp.edu.ar,  
{pamadeo, mcharnelli}@linti.unlp.edu.ar

<sup>2</sup>Instituto de Investigación en Informática LIDI (III-LIDI).  
Facultad de Informática. Universidad Nacional de La Plata  
laural@lidi.info.unlp.edu.ar

<sup>3</sup>LINSI - Laboratorio de Innovaciones en Sistemas de Información.  
Dpto de Sistemas. UTN Facultad Regional La Plata  
gbaldino@linsi.edu.ar

## Resumen

La línea de investigación actual se centra en la Minería de Datos Educativa donde se ha puesto énfasis en la etapa del preprocesamiento, el análisis y aplicación de diferentes técnicas no supervisadas y de técnicas de visualización de datos masivos con el objetivo de identificar las características más relevantes de los alumnos de la Facultad de Informática de la UNLP y de la UTN-Facultad Regional La Plata. A su vez, se espera poder contribuir en el área educativa a través de la determinación de perfiles de los alumnos de la Facultad de Informática en lo que se refiere a su interacción con recursos educativos de acceso libre. Dichos perfiles podrán ser utilizados para caracterizar su comportamiento actual y asistirlo en forma automática a fin de que pueda alcanzar el comportamiento esperado.

**Palabras clave:** analítica del aprendizaje, minería de datos educativa.

## Contexto

Este trabajo de investigación se encuadra en el Proyecto de Incentivos acreditado “Innovación en TIC para el desarrollo de aplicaciones en educación, inclusión, gobierno y salud” desarrollado en el Laboratorio de Investigación en Nuevas Tecnologías Informáticas, LINTI.

## Introducción

La Analítica del Aprendizaje (Learning Analytics) es un campo de investigación emergente que analiza información referida a personas que están aprendiendo y sus contextos, con el propósito de comprender y optimizar el aprendizaje y los entornos en los que dicho aprendizaje sucede.

En los últimos años, las instituciones educativas se han embarcado en su propia exploración de grandes conjuntos de datos para mejorar los índices de retención y proporcionar una experiencia personalizada y de mayor calidad para los

estudiantes. La aplicación de técnicas de Minería de Datos en el ámbito educativo ha permitido caracterizar a los distintos actores que intervienen en los procesos de enseñanza aprendizaje [1].

En el III-LIDI, se trabaja en este tema desde 2008. Las investigaciones realizadas han permitido evaluar la pertinencia y calidad del material desarrollado para un curso dado [2] [3]. También se estudiaron técnicas aplicables a la modelización del estudiante en lo referido a su proceso de aprendizaje [4] [5].

En el LINTI, desde el año 2011 se aplican herramientas y técnicas de inteligencia de negocio a datos almacenados en los sistemas académicos de gestión universitaria [6]. También se realizaron integraciones con plataformas de enseñanza-aprendizaje con otros sistemas de información [7].

A continuación se describen las tareas realizadas recientemente.

### **Preprocesamiento de datos**

En esta etapa, el énfasis estuvo puesto en el análisis y estudio de distintas técnicas de transformación de atributos. Se trabajó con la información publicada por la Fundación Sadosky referida a una encuesta realizada a 628 alumnos de colegios secundarios de la ciudad de Buenos Aires. Esta encuesta fue realizada con el fin de analizar los motivos por los cuales los jóvenes deciden estudiar o no una carrera relacionada con la Informática. Por tal motivo, dicha encuesta contenía preguntas sobre habilidades tecnológicas y uso de la computadora. A través del análisis realizado, de un total de 200 atributos iniciales, se eliminaron aquellos con exceso de datos faltantes o no generalizables y se crearon otros de

acuerdo a los requerimientos de las técnicas de Minería de Datos a utilizar.

- ***Minería de Texto aplicada a las preguntas abiertas***

Debido a la complejidad de procesar texto no estructurado, las preguntas abiertas contenidas en la encuesta requirieron un tratamiento especial. En este caso, dado que fue preciso operar con información textual se utilizaron técnicas de Minería de Texto a fin de identificar los términos más frecuentemente empleados por los encuestados. Este tipo de técnicas fueron empleadas anteriormente en [8].

### **Análisis de relevancia de atributos**

A partir de la encuesta realizada por la Fundación Sadosky, aplicando métodos de selección de características, se obtuvieron los 12 atributos más relevantes que permitieron caracterizar a los alumnos de colegios secundarios que eligen seguir carreras de informática. Esto significó una reducción del 90% sobre la cantidad de preguntas totales de la encuesta original. Los resultados de esta investigación fueron publicados en [9].

Actualmente, se está trabajando a partir de la información de los alumnos de la Facultad de Informática de la UNLP recolectados a través del sistema SIU-Guaraní y de la UTN Facultad Regional La Plata. En este caso, a diferencia de la encuesta de Sadosky, toda la información se encuentra estructurada. Debido a esto, la etapa de preprocesamiento resultó más sencilla.

- **Técnicas de agrupamiento**

Se estudiaron y aplicaron técnicas de aprendizaje no supervisado para evaluar la performance de los alumnos a lo largo

de los primeros cinco años de la carrera. Se aplicaron técnicas de clustering para obtener los grupos de alumnos con un desempeño similar en la carrera [10].

- Técnicas de visualización

Se analizaron técnicas de visualización para representar la información de los alumnos. Se utilizó una visualización de conjuntos paralelos para identificar tendencias en las respuestas [9]. Mediante un gráfico de dispersión se analizó la manera en que los alumnos completaron la encuesta observando que la mayoría de ellos no respondían a las últimas preguntas.

Se realizó un diagrama de coordenadas paralelas para observar el progreso de la carrera de cada alumno a lo largo de los primeros cinco años. Uno de los resultados obtenidos muestra que de los alumnos egresados, la mayoría de ellos en 5to año sólo tenían aprobado el 50% de la carrera. Así también se realizaron diagramas de desviación sobre los clusters obtenidos. Como resultado se observó que se produce un punto de inflexión en el segundo año. A partir de ese momento una gran cantidad de alumnos detienen su progreso en la carrera.

### **Modelado de Perfiles de Alumno**

Uno de los temas que más preocupa a las distintas unidades académicas es la deserción universitaria. Por tal motivo, se han estudiado distintas técnicas no supervisadas para caracterizar los datos del problema [11]. Uno de los objetivos más importantes es poder modelar los perfiles de los alumnos de las carreras de la Facultad de Informática de la UNLP y la UTN Facultad Regional La Plata.

- Sistema Recomendador

Actualmente, uno de los puntos centrales de esta línea de investigación consiste en modelar perfiles dinámicos a través de información provista por los entornos de enseñanza y aprendizaje que usan las diferentes cátedras de la Facultad de Informática, información de los préstamos de la biblioteca, actividad en las redes sociales, entre otros. Por tal motivo interesan soluciones capaces de operar tanto sobre información estructurada como no estructurada. En este último caso el énfasis está puesto en la extracción de conocimiento a partir de información textual. Como resultado se espera obtener modelos dinámicos capaces de adaptarse fácilmente a los cambios de la información.

Luego se desarrollará un sistema recomendador como una extensión de los entornos de aprendizaje, que utilizando la información de los perfiles obtenidos pueda recomendar materiales educativos [12].

### **Líneas de Investigación, Desarrollo e Innovación**

- Estudio de distintas técnicas de preprocesamiento aplicables a Minería de Textos.
- Estudio, análisis y comparación de diferentes técnicas de visualización.
- Estudio y desarrollo de métodos para la identificación de los atributos más relevantes de un conjunto de datos.
- Estudio de técnicas de agrupamiento aplicables a información numérica y categórica.
- Revisión y análisis de técnicas específicas de Learning Analytics.
- Estudio y desarrollo sobre modelado de perfiles dinámicos de los

alumnos según sus características más relevantes.

- Estudio y desarrollo de un sistema recomendador de materiales educativos utilizando la información de los perfiles obtenidos.

## Resultados y Objetivos

- Descripción de características más relevantes de los estudiantes aportando información útil en relación a los datos personales y académicos.
- Desarrollo de una prueba de concepto que arroja información preliminar relevante respecto a la problemática de la deserción universitaria.
- Como objetivos futuros se pretende modelar perfiles dinámicos de los alumnos y desarrollar un sistema recomendador.

## Formación de Recursos Humanos

Dentro de los temas involucrados en esta línea de investigación, actualmente hay 2 investigadores realizando su doctorado, un investigador realizado una maestría y un alumno ha concluido su tesina de grado.

## Referencias

[1] Romero C., Ventura S. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, Volume 33, Issue 1, July 2007, Pages 135-146, ISSN 0957-4174.

[2] Grossi, M.D., Lanzarini, L. Reglas de Predicción aplicables al Diseño de un Curso de Computación. III Congreso de

Tecnología en Educación y Educación en Tecnología. TE&ET'08. Mayo 2008. Bahía Blanca.

[3] Lanzarini, Denazis, Grossi Estrategias Inteligentes aplicables a un Sistema Educativo. X Workshop de Investigadores en Ciencias de la Computación (WICC 2008), Educación. Mayo de 2008. La Pampa

[4] Arona, Huapaya, Lanzarini, Lizarralde. Lógica Difusa aplicada al Modelo del Estudiante de un Sistema Tutorial Inteligente. IV Congreso de Tecnología en Educación y Educación en Tecnología. TE&ET'09. Julio 2009. La Plata. Bs.As

[5] Lanzarini, Huapaya. Diagnóstico Adaptativo del Estudiante en Sistemas Tutoriales Inteligentes. XI Workshop de Investigadores en Ciencias de la Computación (WICC 2009). Mayo de 2009. San Juan

[6] Diaz, F. Osorio, A. Amadeo, P. Romero, D. Aplicando estrategias y tecnologías de Inteligencia de Negocio en sistemas de gestión académica. WICC, Paraná, 2013.

[7] M. E. Charnelli. Integrando repositorios digitales de recursos educativos abiertos con plataformas virtuales de aprendizaje. 2014.

[8] Villa Monte, A. Estrebou, C., Lanzarini, L. E-mail processing using data mining techniques. Publicado en el Libro *Computer Science & Technology*

Series – XVI Argentine Congress of Computer Science Selected Papers, ISBN 978-950-34-0757-8. EDULP, 2011.

[9] Charnelli, E. Lanzarini, L. Baldino, G. Diaz, F. Determining the profiles of young people from Buenos Aires with a tendency to pursue computer science studies. XX Congreso Argentino de Ciencias de la Computación. La Matanza, 2014.

[10] Asif, R. Merceron, A. Pathan K. Investigating Performances's Progress of Students. CEUR Workshop Proceedings. ISSN 1613-0073, 2014

[11] Formia S. Evaluación de técnicas de Extracción de Conocimiento en Bases de Datos y su aplicación a la deserción de alumnos universitarios. Tesis de Especialista en Tecnología Informática aplicada en Educación. Dic 2012.

[12] J. Díaz, A. Schiavoni, P. Amadeo, M. E. Charnelli, “Búsqueda personalizada de recursos educativos abiertos basada en el perfil del alumno dentro de un entorno educativo”, WICC, Tierra del Fuego, 2014.