

## Extracción de Conocimiento en el Cursado del Ciclo Común de Articulación de Carreras de Ingeniería

Oscar Eduardo Quinteros<sup>1</sup>, Ana Funes<sup>2</sup>, Hernán César Ahumada<sup>1</sup>

<sup>1</sup>Universidad Nacional de Catamarca

Maximio Victoria 55, 4700 Catamarca, Argentina

{oequinteros,hcahumada}@tecno.unca.edu.ar

<sup>2</sup>Universidad Nacional de San Luis

Ejército de los Andes 950, 5700 San Luis, Argentina

afunes@unsl.edu.ar

### Resumen

Sobre datos de regularidades y exámenes rendidos de las cátedras pertenecientes al Ciclo Común de Articulación (CCA) de las carreras de Ingeniería de la Facultad de Tecnología y Ciencias Aplicadas, se puede aplicar métodos de minería de secuencias para obtener el recorrido académico de los alumnos y de esta manera poder descubrir reglas de secuencias (patrones) frecuentes en la historia académica de cada uno de ellos. El análisis de las secuencias obtenidas permitirá caracterizar a los grupos de alumnos con diferentes rendimientos académicos y grado de avance en los estudios. La interpretación de los resultados posibilitará conocer las materias que mayor tiempo insumen hasta que son aprobadas. El conocimiento obtenido brindará la posibilidad de entender mejor un aspecto de la problemática del abandono temprano de los estudios universitarios.

**Palabras clave:** Minería de Datos, Minería de Secuencias, Rendimiento Académico, Deserción.

### Contexto

Esta línea de trabajo se viene llevando a cabo como un trabajo colaborativo entre proyectos de investigación de dos universidades nacionales: el Proyecto de Incentivos código 22/F222 *Aspectos de alta sensibilidad en el ejercicio de la profesión de Ingeniero de Software*, de la Universidad Nacional de San Luis, y el proyecto acreditado por SECyT - UNCA. 02/J715 - *Hacia La Búsqueda De Un Sistema De Alerta Temprana De Deserción En Carreras De Ingeniería*, que se desarrolla en el Departamento de Formación Básica de la Facultad de Tecnología y Ciencias Aplicadas (Universidad Nacional de Catamarca).

En particular, esta propuesta de investigación se desarrolla como una tesis de posgrado de la carrera Maestría en “Ingeniería de Software” de la Universidad Nacional de San Luis.

### Introducción

Las autoridades y equipos docentes de universidades se plantean la mejora continua de la calidad de los procesos y servicios que ofrecen, de esta manera buscan incrementar su contribución al desarrollo social y económico de su entorno. Un estudio realizado por

Ahumada [3], indica que las carreras de Ingeniería de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca (U.N.Ca.) tienen estadísticas de deserción y desgranamiento de magnitud significativa y se producen, fundamentalmente, en los primeros años de estudios, en donde la mayoría de los análisis descriptivos de este fenómeno revelan demasiado tarde el abandono de estudiantes en diferentes tramos de la carrera.

La Facultad de Tecnologías y Ciencias Aplicadas<sup>1</sup>, de la Universidad de Catamarca, junto a otras facultades del país, se ha adherido a un convenio con el fin de dictar un conjunto común de materias, denominado Ciclo Común de Articulación (CCA) [14], en la etapa inicial de las carreras de ingeniería. El estudio del rendimiento académico de los alumnos en el cursado del CCA puede brindar información útil para disminuir la deserción temprana de los alumnos en las carreras.

El sistema de gestión de información académica de las Universidades Nacionales denominado, SIU-GUARANI, lleva un registro actualizado de los exámenes rendidos por los alumnos, junto con otros datos académicos referidos básicamente a inscripciones de materias para cursar y actas de regularidades. En los últimos años, la Facultad de Tecnología y Ciencias Aplicadas, ha implementado sistemas de gestión que integran procesos y áreas, donde estos sistemas producen datos que se almacenan en bases de datos, con dimensiones considerablemente grandes y diversidad de temas. Los trabajos realizados por Ahumada [3] y Pautsch [11] muestran la aplicación del uso de Minería de Datos (MD) en estos

entornos educativos, particularmente enfocados en la identificación de las causales de abandono.

Más específicamente, existe una rama en la MD que se denomina Minería de Datos Educativos (EDM), la cual 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. Romero [15] expresa que la información así obtenida se convierte en el insumo indispensable para la toma de decisiones. La EDM tiene varios objetivos, tales como los descriptos por Jindal [9]. Uno de estos objetivos es el estudio, desde el punto de vista académico, de la performance del alumno. La aplicación de MD para extracción de conocimiento desde los datos que gestionan las Universidades es factible, ya que desde hace varios años, las Universidades Nacionales vienen trabajando en la implementación de sistemas de gestión o transaccionales que integran procesos y áreas. Estos sistemas producen datos que se almacenan en bases de datos, con dimensiones considerablemente grandes y diversidad de temas, como indica Menéndez en [10].

Los métodos de MD principalmente usados en el EDM se dividen en 2 grupos, los orientados a la verificación y los orientados al descubrimiento: dentro de estos últimos tenemos a su vez métodos de Clasificación, Estadísticas, Agrupamiento, Predicción, Redes Neuronales, Minería de Reglas de Asociación y Minería Web [9].

Galindo[6] también describe categorías de métodos de EDM, donde una de ellas es la Minería de Relaciones y, dentro de ésta, los métodos de minería de patrones de secuencias. La Minería de Secuencia [1] consiste en la búsqueda de patrones secuenciales frecuentes en una base de datos de eventos con fecha y hora [13][5].

---

<sup>1</sup><http://www.tecno.unca.edu.ar>

Existen además, diversos métodos para minar secuencias, Ahola [2] explicita varios, tales como SPADE [16], o PrefixSpan [12], entre otros. Como antecedente en la aplicación de minar secuencias sobre datos provenientes de entornos educativos tenemos el trabajo de Guerra [7] en donde usa los métodos SPAM [4] y PexSPAM [8] para descubrir patrones comunes en secuencias de ejercicios parametrizados usados en las herramientas de aprendizaje on-line.

Ante lo expuesto, se considera que es posible obtener los datos relacionados con exámenes rendidos y modelarlos como una serie y luego aplicar métodos para minar las secuencias. El sistema SIU-GUARANI lleva un registro de los exámenes rendidos por los alumnos durante el cursado del CCA, que combinado con otros datos académicos puede conformar un dataset para ser analizado por métodos de MD.

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

La línea de investigación aquí presentada trata del un tema surgido para el desarrollo de una tesis de Maestría en Ingeniería de Software de la Universidad Nacional de San Luis.

El proyecto comprende la investigación de Minería de Datos en base a registros académicos obtenidos del SIU-GUARANI del cursado de las cátedras que se dictan en el CCA 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 y analizar el recorrido académico de los estudiantes, con la finalidad de detectar patrones que describan la demora en la terminación del

cursado, que puedan derivar posteriormente en abandono de los estudios universitarios.

## **Resultados y Objetivos**

### **Objetivo General**

Caracterizar y analizar el recorrido académico de exámenes rendidos por alumnos del Ciclo Común de Articulación (CCA) de las Carreras de la Facultad de Tecnologías y Ciencias Aplicadas de la Universidad Nacional de Catamarca (U.N.Ca.).

### **Objetivos Específicos**

- Descubrir reglas que caractericen la demora en la terminación del cursado del CCA.
- Encontrar características en una secuencia de exámenes rendidos que permitan explicar algunas de las causales que puedan generar que un alumno tenga una demora excesiva.
- Descubrir relaciones entre las materias y los patrones de demoras en la rendición de los exámenes.
- Buscar patrones frecuentes en el conjunto de exámenes rendidos de alumnos con mejor y peor rendimiento.
- Brindar elementos para poder realizar un análisis institucional y que permita tomar decisiones para definir nuevas estrategias dentro de la institución, como por ejemplo re-definir planes de estudio, o realizar cursos de nivelación.

## **Formación de Recursos Humanos**

El presente trabajo es la base para el desarrollo de una tesis de posgrado de la

carrera “Maestría en Ingeniería de Software” de la Universidad Nacional de San Luis. La misma está siendo llevada adelante por el Licenciado en Sistemas de Información Oscar Eduardo Quinteros, y siendo dirigida por la investigadora de la Universidad Nacional de San Luis Magister Ana Funes, y codirigido por el investigador de la Universidad Nacional de Catamarca, Dr. Hernán Cesar Ahumada. El trabajo de tesis se encuentra en la etapa de elaboración del proyecto de tesis.

## Referencias

[1] Rakesh Agrawal and Ramakrishnan Srikant. Mining sequential patterns. In Data Engineering, 1995. Proceedings of the Eleventh International Conference on, pages 3\_14. IEEE, 1995.

[2] Jussi Ahola. Mining sequential patterns. 2001.

[3] Hernán César Ahumada, Hugo Rubén Dip, Carlos Gabriel Herrera, and Juan Carlos Leguizamón Almendra. Minería de datos para un sistema de alerta temprana de deserción en carreras de ingeniería. In Anales XVII Workshop de Investigadores en Ciencias de la Computación WICC 2015), Abril 16-17, 2015, Salta, Argentina. Red UNCI, 2015.

[4] Jay Ayres, Jason Flannick, Johannes Gehrke, and Tomi Yiu. Sequential pattern mining using a bitmap representation. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 429\_435. ACM, 2002.

[5] Dante Conti and Fco Javier Martínez de Pisón Ascacíbar. Reglas de asociación en series temporales: panorama referencial y tendencias.

[6] Álvaro Jiménez Galindo and H García. Minería de datos en la educación. Universidad Carlos III, 2010.

[7] Julio Guerra, Shaghayegh Sahebi, Yu-Ru Lin, and Peter Brusilovsky. The problem solving genome: Analyzing sequential patterns of student work with parameterized exercises. In Educational Data Mining 2014, 2014.

[8] Joshua Ho, Lior Lukov, and Sanjay Chawla. Sequential pattern mining with constraints on large protein databases. In Proceedings of the 12th International Conference on Management of Data (COMAD), pages 89\_100, 2005.

[9] Rajni Jindal and Malaya Dutta Borah. A survey on educational data mining and reserarch trends. International Journal of Database Management Systems, 5(3), 2013.

[10] Mariano Menéndez and María de Lujan Gurmendi. Sistemas para la toma de decisiones en el ámbito universitario. In 6o Simposio Argentino De Informática En El Estado, 2012.

[11] JGA Pautsch. Minería de datos aplicada al análisis de la deserción en la carrera de analista en sistemas de computación. Misiones, Argentina: Universidad Nacional de Misiones, Facultad de Ciencias Exactas, Químicas y Naturales, 2009.

[12] Jian Pei, Jiawei Han, Behzad Mortazavi-Asl, Jianyong Wang, Helen Pinto, Qiming Chen, Umeshwar Dayal, and Mei-Chun Hsu. Mining sequential patterns by pattern-growth: The pre\_xspan approach. Knowledge and Data Engineering, IEEE Transactions on, 16(11):1424\_1440, 2004.

[13] Prof Pinkal Shah and AK Dua. Algorithm for sequence mining using gap constraints. International Journal of Engineering Research and Development, pages 37\_49, 2014.

[14] Facultad de Tecnologías y Ciencias Aplicadas. Historia - Facultad - CCA. [Web:www.tecno.unca.edu.ar/index.php/facu/historia; accedido el 29-02-2016].

[15] Cristobal Romero, Sebastian Ventura, Mykola Pechenizkiy, and Ryan SJD Baker. Handbook of educational data mining. CRC Press, 2010.

[16] Mohammed J. Zaki. Spade: An efficient algorithm for mining frequent sequences. Mach. Learn., 42(1-2):31\_60, January 2001.