

# MINERÍA DE DATOS EDUCACIONAL PARA DETERMINAR PERFILES DE ALUMNOS RECURSANTES EN CARRERAS DE INGENIERÍAS

C. Claudio, E. Jramoy, F. Barrera, F. Cardona, R. Navarro Peláez, S. Racca

Grupo de Investigación Base de Datos

Universidad Tecnológica Nacional - Facultad Regional San Francisco

Av. de la Universidad 501 - San Francisco, Córdoba, Argentina

{cjcarrizo77, emilianolucas, facubarrera123, ferdcardona, ingrnavarropeleaz, raccasofi}@gmail.com

## RESUMEN

Las instituciones universitarias tienen el desafío de asegurar y mantener el nivel de calidad académica con el fin de proveer profesionales altamente capacitados que respondan a las demandas del mercado laboral actual, especialmente en el área de las TICs. Uno de los aspectos en donde las universidades deben poner mayor énfasis es en el rendimiento académico, ya que generalmente un bajo rendimiento académico está asociado con una alta tasa de deserción de alumnos. Para evaluar el rendimiento académico de un alumno es necesario conocer si existen patrones o perfiles comunes a grupos de alumnos, esto es de significativa importancia para definir acciones que permitan mejorar el desempeño de los alumnos.

En este trabajo se propone el uso de minería de datos educacional para la construcción de modelos que permitan identificar perfiles de alumnos que recursan materias básicas en el primer año de las carreras de ingenierías.

Los resultados de este proyecto serán un aporte para el área de gestión académica, ya que podrán contar con un instrumento objetivo que les permitirá definir acciones a futuro en pos de lograr la mejora en el rendimiento académico de los alumnos en materias básicas del primer año.

**Palabras clave:** minería de datos educacional – modelos – patrones – alumnos recursantes - ingeniería

## CONTEXTO

La Unidad Científico Tecnológica donde se enmarca el presente proyecto es el Grupo de Investigación "Base de Datos". Esta Unidad desarrolla sus líneas de investigación en concordancia con las áreas prioritarias correspondientes a la carrera de Ingeniería en Sistemas de Información, que se encuentran formalizadas a través de la Resolución de Consejo Directivo N° 353/2016 de la Facultad Regional San Francisco. El presente proyecto de investigación y desarrollo, de tipo sin incorporación en programa de incentivos, se encuentra homologado y financiado por la Secretaría de Ciencia, Tecnología y Posgrado de la Universidad Tecnológica Nacional bajo el código UTN7854 y según la Disposición SCTyP N° 221/19, el periodo de ejecución es desde Enero de 2020 hasta Diciembre de 2021, el mismo está incluido en el Programa I&D + i de Tecnología de las Organizaciones de la Universidad Tecnológica Nacional.

## 1. INTRODUCCIÓN

Las universidades tienen el constante desafío de asegurar y mantener el nivel de calidad académica en pos de proveer a la sociedad profesionales que estén altamente capacitados para responder a las demandas actuales del mercado laboral, especialmente en el área de las tecnologías de la información y las comunicaciones

(TICs). Para ello, es necesario sistematizar procesos que permitan evaluar en forma permanente aspectos relacionados a la calidad académica de la Universidad (Briand et al., 1999). Uno de los aspectos sobre los cuales se debe realizar un constante seguimiento y monitoreo es en el rendimiento académico, debido a que comúnmente el bajo rendimiento académico está asociado a una alta tasa de deserción.

El rendimiento académico se puede definir como la productividad del sujeto, matizado por sus actividades, rasgos y la percepción más o menos correcta de los cometidos asignados (Maletic et al., 2002). Generalmente el rendimiento académico se evalúa teniendo en cuenta aspectos que influyen en el desempeño, pudiendo mencionar entre ellos: factores socioeconómicos, la amplitud de programas de estudio, las metodologías de enseñanza, conocimientos previos del alumno (Marcus, 2003). Según datos estadísticos brindados por el Área de Gestión Académica de la Universidad Tecnológica Nacional Facultad Regional San Francisco, en las carreras de ingenierías existe históricamente una alta tasa de alumnos con bajo rendimiento académico en instancias de evaluación en materias básicas del primer año (Álgebra y Geometría Analítica, Análisis Matemático I y Física I), esto provoca que los alumnos cursen varias veces este tipo de materias con el fin de alcanzar la regularidad o la aprobación directa. Esto ocasiona no sólo el atraso del alumno en la carrera, sino que además muchos alumnos optan por abandonar la carrera al sentirse frustrados por no poder regularizar o aprobar materias que pertenecen al área básica.

El rendimiento académico suele estar afectado por múltiples factores de tipo socio-educativos, socio-económicos, demográfico, actitudinales, que sin duda condicionan el desempeño del alumno

frente al cursado de las materias. Para evaluar el rendimiento académico de un alumno es necesario conocer si existen patrones o perfiles comunes a grupos de alumnos, de esta manera la determinación de perfiles se convierte en una estrategia de valor significativo a la hora de tomar acciones que permitan mejorar el desempeño de los alumnos (González, 1988; Di Gresia, 2007). El establecimiento de perfiles consiste en el proceso de determinación y clasificación de patrones.

En el área de Inteligencia Artificial y del Aprendizaje de máquinas (Mitchell, 1997), existen muchos métodos que permiten determinar y clasificar patrones, los cuales permiten devolver información de valor para la toma de decisiones, uno de ellos es la minería de datos o DM (Data Mining). DM es el uso consistente de algoritmos concretos que generan una enumeración de patrones a partir de los datos pre-procesados (fayyad et al., 2001; hand et al., 2000; frawley et al., 1992). El conjunto de técnicas de análisis de datos que permiten extraer tendencias, patrones y regularidades para describir y comprender mejor los datos forman parte de la DM. Además, permite extraer patrones y tendencias para predecir comportamientos futuros (Simon, 1997; berson & Smith, 1997; White, 2001).

Para poder extraer conocimiento usando minería de datos, se requiere la aplicación de uno o varios modelos o algoritmos. Estos modelos pueden ser descriptivos o predictivos (L.-D. Chen, T. Sakaguchi, and M. N. Frolick, 2000). Los modelos descriptivos usan modelos estadísticos como, por ejemplo: distribución probabilística, correlación, regresión, análisis de clusters y análisis de discriminación. En los modelos descriptivos se conocen las características principales del conjunto de datos a analizar. Por otra parte, los modelos predictivos se basan mayormente en

técnicas de aprendizaje automático (machine learning) y emplean inteligencia artificial (J. L. Hung and K. Zhang, 2006). Estos modelos se crean para predecir valores desconocidos de variables llamadas variables objetivo.

La variedad de aplicaciones para análisis de información en sectores comerciales, financieros, médicos, educativos o científicos, ha permitido el desarrollo de gran cantidad de métodos y algoritmos para predecir tendencias y descubrir patrones. Entre los más importantes se encuentran: clustering, clasificación, regresión, análisis de desviación y análisis de cesta de mercado (M. J. Jafar, 2010). Cuando se aplica minería de datos en instituciones educativas la disciplina se conoce como minería de datos educacionales (MDE). La MDE es una disciplina en evolución que usa tecnologías informáticas como son almacenes de datos y herramientas de inteligencia de negocios para descubrir tendencias y patrones sobre datos educacionales. El conocimiento que DM genera actúa como soporte para las autoridades de centros de educación superior en la toma de decisiones oportunas y a los profesores para analizar el comportamiento y aprendizaje de sus alumnos (R. Jindal and M. D. Borah, 2013). La disciplina se enfoca en el diseño de modelos para mejorar las experiencias del aprendizaje y la eficiencia organizacional (R. A. Huebner, 2013).

Al ser una disciplina nueva, no existe una definición ampliamente aceptada de MDE. Una definición que se ajusta a los objetivos de nuestra investigación es proporcionada por la International Society of Educational Data Mining que define a MDE como “una disciplina en evolución, que tiene que ver con el desarrollo de métodos para explorar los tipos únicos de datos que provienen de ambientes educativos y por medio de la aplicación y

uso de estos métodos permite una mejor comprensión de los estudiantes y el entorno en el que aprenden”.

La MDE usa métodos, herramientas y algoritmos de DM para investigar datos de interacciones de estudiantes y docentes con el sistema educativo, colaboración entre estudiantes, datos administrativos y datos demográficos. La MDE integra métodos, algoritmos y técnicas con los cuales se puede realizar diferentes experimentos y diseñar modelos para predecir u obtener patrones de datos educacionales.

## **2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO**

Los principales ejes de trabajo de esta línea de investigación son los siguientes:

- Obtener y explorar los datos académicos de alumnos de ingenierías de UTN FR San Francisco.
- Identificar y seleccionar los alumnos que han recurrido materias básicas en el primer año de acuerdo a la muestra definida.
- Encuestar a alumnos recurrentes de materias básicas para obtención de datos socioeconómicos, socioeducativos, demográficos y actitudinales.
- Seleccionar características del alumno recurrente basándose en aspectos personales, académicos, socioeducativos, socioeconómicos, socio-demográficos y actitudinales.

- Caracterizar la disciplina de modelos descriptivos.
- Caracterizar y comparar los diferentes métodos, herramientas y algoritmos de minería de datos educativos.
- Analizar y seleccionar el método y la tipología más acorde para la construcción del modelo descriptivo.
- Construir el modelo descriptivo de acuerdo al método y tipología de MDE seleccionada.
- Determinar el perfil del alumno de ingeniería recursante en materias básicas en el primer año.

### **3. RESULTADOS OBTENIDOS/ESPERADOS**

Dentro de los resultados esperados/obtenidos para esta línea de investigación, se han logrado los siguientes:

#### **Resultados Obtenidos**

- Se han obtenido los datos académicos de alumnos de ingenierías de UTN FR San Francisco y se ha realizado una exploración de los mismos a través de la herramienta RapidMiner.
- Se ha identificado la población de alumnos y se han seleccionado, de acuerdo a una muestra, aquellos que han cursado materias básicas en el primer año.

- Se ha comenzado con el diseño de la encuesta a realizar a alumnos recursantes de materias básicas para obtener datos socioeconómicos, socioeducativos, demográficos y actitudinales.

#### **Resultados Esperados**

- Obtener los resultados de las encuestas y realizar el registro correspondiente.
- Obtener el modelo descriptivo que permita determinar el perfil del alumno de ingeniería recursante en materias básicas en el primer año.
- Difundir y divulgar los resultados obtenidos.

### **4. FORMACIÓN DE RECURSOS HUMANOS**

En esta línea de investigación se trabaja desde el Grupo de Investigación "Base de Datos" perteneciente a la UTN Facultad Regional San Francisco, Córdoba. El equipo de trabajo está constituido por un Director de Proyecto, un Co-Director, tres docentes investigadores, un graduado y dos alumnos de la especialidad Sistemas de Información. A su vez, de los docentes que participan, 2 de ellos están llevando adelante sus tesis de maestrías, a través de la Universidad Nacional de San Luis.

### **5. BIBLIOGRAFÍA**

- Berson, A. & Smith, S. J. (1997). Data Warehouse, data Mining & OLAP. MC GRAW HILL. USA.
- Briand, L., Daly, J. & Wüst, J. (1999). A unified framework for coupling measurement in objectoriented systems. IEEE Transactions on Software Engineering, 25 (1), 91-121.

- Data Mining and OLAP”, Tata McGraw – Hill Berson & Smith, 1997. “Data Warehousing, Edition.
- Di Gresia, L. (2007). Rendimiento académico universitario. Asociación Argentina de Economía Política. Extraído el 5 de marzo de 2012 desde: <http://www.aaep.org.ar/anales/works/works2007/digresia.pdf>.
- Fayyad, U.M.; grinstein, g. & Wierse, A. (2001) Information Visualization in data Mining and Knowledge discovery. Morgan Kaufmann. harcourt Intl.
- Frawley, W., Piatetsky-Shapiro, g. & Matheus, C. (1992) Knowledge discovery in databases: An Overview. AI magazine, 13(3), 57.
- González, A. J. (1998) Indicadores del rendimiento escolar: relación entre pruebas objetivas y calificaciones. Revista de Educación, 287, 31-54. España.
- Hand, d.J.; Mannila, h. & Smyth, P. (2000) Principles of data Mining. The MIT Press. USA.
- International Society Educational Data Mining. “Educational Data Mining”, 2015. <http://www.educationdatamining.org/>.
- J. L. Hung and K. Zhang, “Data Mining Applications to Online Learning,” World Conf. E-Learning Corp. Gov. Heal. High. Educ., pp. 2014–2021, 2006.
- L.-D. Chen, T. Sakaguchi, and M. N. Frolick, “Data Mining Methods, Applications, and Tools,” Inf. Syst. Manag., vol. 17, no. 1, pp. 65–70, Jan. 2000.
- Maletic, J., Collard, M. & Marcus, A (2002). Source Code Files as Structured Documents. Proceedings 10th IEEE International Workshop on Program Comprehension (IWPC’02), 289- 292.
- Marcus, A. (2003). Semantic Driven Program Analysis, Kent, OH, USA, Kent State University Doctoral Thesis.
- Mitchell, T. (1997) Machine Learning. Mc Graw Hill.
- M. J. Jafar, “A Tools-Based Approach to Teaching Data Mining Methods.,” J. Inf. Technol. Educ., vol. 9, pp. 1–29, 2010.
- Oswaldo Moscoso-Zeam, Sergio Luján-Mora - Minería de Datos Educativos: una visión holística Educational Data Mining: an holistic view
- R. A. Huebner, “A Survey of Educational DataMining Research,” Res. High. Educ. J., pp. 1–13, 2013.
- R. Jindal and M. D. Borah, “A Survey on Educational Data Mining and Research Trends,” Int. J. Database Manag. Syst., vol. 5, no. 3, pp. 53–73, 2013.
- S. A. Kumar and Dr.Vijayalakshmi.M.N, “A Novel Approach in Data Mining Techniques for Educational Data.
- S.Anupama,” 3rd Int. Conf. Mach. Learn. Comput. (ICMLC 2011) A, no. Icmlc, pp. 152–154, 2011.
- Siemens, G., Baker, R.S.J.d. (2012) Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. Proceedings of the 2nd International Conference on Learning Analytics and Know.
- Simon, A. (1997) data Warehouse, data Mining and OLAP. John Wiley & Sons. USA.
- White, C. J. (2001) IbM Enterprise Analytics for the Intelligent e-business. IbM Press. USA.