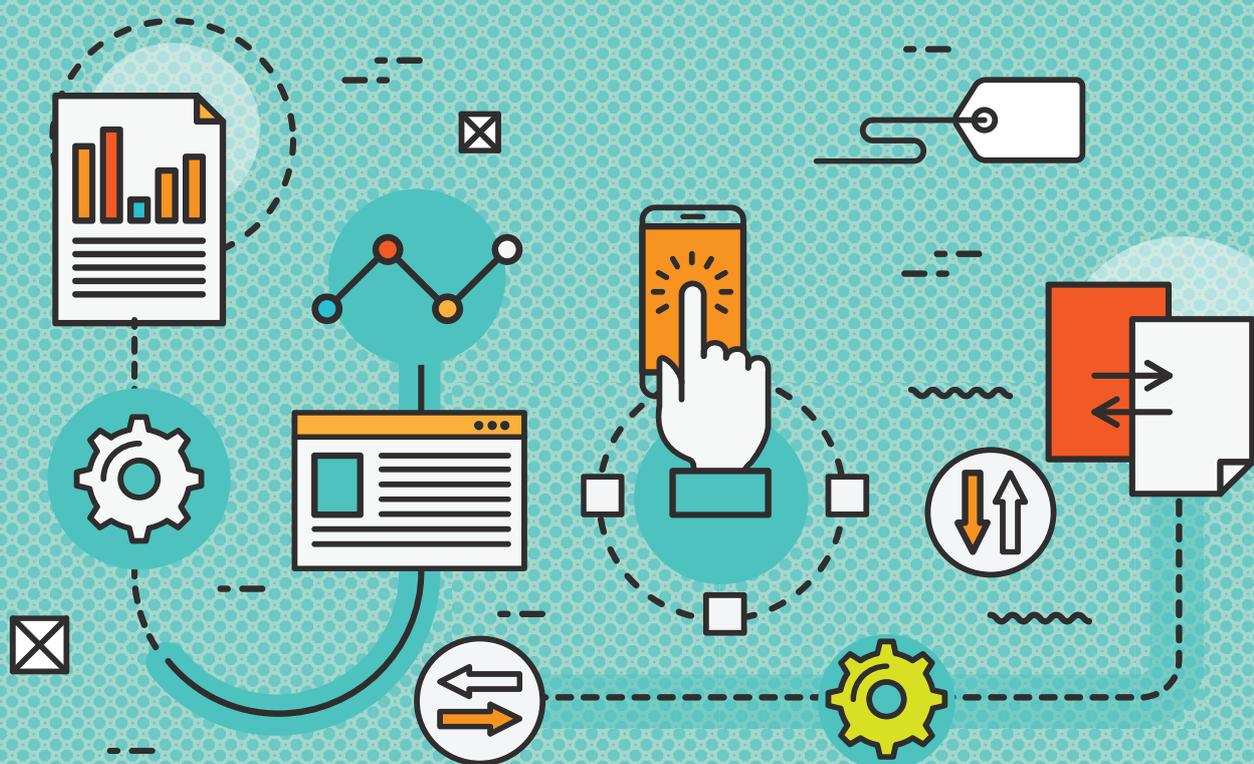


# CONAISI 2017

5to. Congreso Nacional de  
Ingeniería Informática /  
Sistemas de Información

**2 y 3 de Noviembre Santa Fe**



**Actas**

**5to Congreso Nacional de Ingeniería Informática  
/ Sistemas de Información**

**CONAIISI 2017**

**2-3 Noviembre, Santa Fe - Argentina**

**Actas**

**5to Congreso Nacional de Ingeniería Informática  
/ Sistemas de Información**

**CONAIISI 2017**

**2-3 Noviembre, Santa Fe – Argentina**



# **Comité Organizador CONAISI 2017**

## **Comisión Ejecutiva**

### **Coordinador**

Ing. Roberto Muñoz, UTN Regional Córdoba

### **Coordinador Suplente**

Ing. Osvaldo Marcovecchio, Universidad de la Marina Mercante

### **Secretarios Titulares**

Dr. Carlos Neil, Universidad Abierta Interamericana

Lic. Augusto José Nasrallah, UTN FR Tucumán

### **Secretarios Suplentes**

Ing. Nelson Roberto Sotomayor, UFASTA

Ing. Juan Calloni, UTN Regional San Francisco

## **Comité Organizador local**

Dr. Aldo Vecchietti

Dr. Horacio Leone

Esp. Marta Castellaro

Dra. Ma. de los Milagros Gutiérrez

Dra Ana Rosa Tymoschuk

Ing. Javier Gomez

Dr. Pablo Villarreal

Dra. María Laura Caliusco

Dr. Mariano Rubiolo

Ing. Diego Alarcón

# **Educación en Ingeniería**

Dra. Mariel Ale, Ing. Patricia Nazar  
Coordinadoras

**CONAIISI 2017**

# Predicción del Rendimiento Académico con Minería de Datos Buscando Reducir el Bajo Rendimiento Académico en Asignatura de la UTN-FRRe

David L. la Red Martínez, Mirtha E. Giovannini, Reinaldo Scappini

*Grupo de Investigación Educativa / Departamento de Ingeniería en Sistemas de Información /  
Facultad Regional Resistencia / Universidad Tecnológica Nacional  
French 414, (3500) Resistencia, Argentina, +54-9379-4638194  
[laredmartinez@gigared.com](mailto:laredmartinez@gigared.com) [meg\\_c51@yahoo.com.ar](mailto:meg_c51@yahoo.com.ar) [rscappini@gmail.com](mailto:rscappini@gmail.com)*

## Resumen

*A menudo, las universidades no son capaces de lidiar con la variedad de factores que pueden afectar el rendimiento académico de los estudiantes. Este tipo de situación genera la necesidad de herramientas que determinen patrones de desempeño académico, y permitan establecer perfiles como base para detectar posibles casos de bajo rendimiento de los estudiantes, es decir, detectar los alumnos que necesitan apoyo en sus actividades académicas.*

**Palabras clave:** *rendimiento académico, patrones de rendimiento académico, perfiles de rendimiento académico, minería de datos.*

## Abstract

*Often, universities are not able to deal with the variety of factors that may affect the academic performance of students. This kind of situation generates the need of tools that establish academic performance patterns, setting profiles as a basis to detect potential cases of underachieving students who need support in their academic activities.*

**Keywords:** *academic performance, academic performance patterns, academic performance profiles*

## 1. Introducción

La deserción de los estudiantes tiene una alta relación con su rendimiento académico, es decir la productividad del alumno [1]. El rendimiento académico está ligado con la asimilación correcta de los contenidos, las demás actividades que realiza y sus características personales (sociales e individuales), siendo un elemento fundamental de análisis ya que puede reflejar muchas características de las instituciones.

Durante los primeros años universitarios el rendimiento impacta fuertemente en la decisión de continuar o abandonar los estudios y es por esto que las

Universidades deben orientar sus esfuerzos en motivar y retener a aquellos estudiantes que muestran, desde un principio, insuficiencias en su rendimiento académico [2].

El rendimiento académico se puede medir observando las calificaciones obtenidas por los estudiantes cuando sus conocimientos, capacidades y aptitudes son evaluados. Pero esta evaluación difícilmente permite obtener información útil que permita detectar y corregir problemas cognitivos, de aprehensión, etc. Por este motivo deben considerarse otros factores incidentes, directa o indirectamente, como ser factores sociales, económicos y experiencias educativas previas, y así poder establecer perfiles de desempeño estudiantil [3].

Considerando que el desempeño durante el primer año es un muy buen predictor de la trayectoria académica posterior del estudiante, este trabajo examina el rendimiento académico en la asignatura troncal del primer nivel de la carrera de Ingeniería en Sistemas de Información (ISI) de la UTN-FRRe, Algoritmos y Estructura de Datos (AED). Esta materia posee una tasa muy alta de alumnos "libres", es decir, aquellos que deberán volver a cursar la materia por no haber aprobado las instancias de evaluaciones parciales. Por este motivo se plantea la necesidad de analizar la existencia de patrones socioeconómicos y de comportamiento que permitan distinguir diferentes perfiles de rendimiento académico.

Existen varias maneras de evaluar el desempeño académico: utilizando indicadores como las tasas de graduación, diferenciando por tipos de centros y analizando el rendimiento a partir de datos individuales [4]; a través de las calificaciones obtenidas en los seminarios de ingreso a la Universidad, realizando el análisis de los datos mediante la técnica estadística de curva ROC (Receiver Operating Characteristic) [5]; o aplicando el enfoque de función de producción para estimar los determinantes del rendimiento académico [6]. También existen varios estudios que utilizan técnicas matemáticas para la evaluación del desempeño. En este sentido, se ha estudiado la capacidad de la regresión lineal y de la regresión logística en la predicción del

rendimiento y del éxito/fracaso académico, partiendo de variables como la asistencia y la participación en clase [7].

Una alternativa es desarrollar los métodos de evaluación aprovechando las capacidades de las tecnologías de información disponibles. En este sentido las técnicas de Data Warehouse (DW) y Data Mining (DM) son herramientas de suma utilidad para la obtención de conocimiento en grandes volúmenes de datos. Un DW es una colección de datos orientado a temas, integrado, no volátil, variable en el tiempo, que se usa para el soporte del proceso de toma de decisiones gerenciales [8]. El DM es la etapa de descubrimiento de conocimiento en bases de datos. Es el paso que consiste en el uso de algoritmos concretos que generan una enumeración de patrones a partir de los datos pre-procesados [9].

En este trabajo se propone la utilización de técnicas de DW y DM sobre la información socioeconómica y del desempeño de los alumnos de la cátedra Algoritmos y Estructura de Datos de la Carrera de Ingeniería en Sistemas de Información de la Facultad Regional Resistencia de la UTN. A partir de un análisis descriptivo obtenido en la mencionada cátedra durante los ciclos lectivos 2013 a 2015, este proyecto tiene como objetivo la determinación de patrones de rendimiento académico, desarrollando un modelo predictivo que pronostique la probabilidad de que un estudiante abandone el cursado, dadas sus características socioeconómicas y académicas.

El artículo se encuentra estructurado de la siguiente manera: en la Sección 2 se explican los conceptos básicos sobre DM haciendo hincapié en las técnicas utilizadas. En la Sección 3 se exhibe el diseño del modelo propuesto. En la Sección 4 se muestran los resultados obtenidos y finalmente, en la Sección 5, las conclusiones.

## 2. Data Mining

Data Mining es un proceso que tiene como propósito descubrir, extraer y almacenar información relevante de amplias bases de datos, a través de programas de búsqueda e identificación de patrones y relaciones globales, tendencias, desviaciones y otros indicadores. El objetivo fundamental es aprovechar el valor de la información localizada y usar los patrones preestablecidos para tener mejor conocimiento de los datos y optimizar la toma de decisiones [10].

Muchos autores consideran a la Data Mining como un paso esencial en el proceso de descubrimiento de conocimiento en base de datos, el cual se compone de una secuencia iterativa de pasos [11]:

- Limpieza de datos, para eliminar el ruido o datos irrelevantes.

- Integración de datos, donde múltiples fuentes de datos pueden ser combinados.
- Selección de datos, donde los datos relevantes para la tarea de análisis se recuperan de la base de datos.
- Transformación de datos, cuando los datos se transforman o consolidan en formas apropiadas para la minería mediante la realización de operaciones de resumen o de agregación, por ejemplo.
- Data Mining, un proceso esencial en la que se aplican métodos inteligentes con el fin de extraer los patrones de datos.
- Evaluación de patrones, para identificar los patrones de interés que representan el conocimiento sobre la base de algunas medidas interesantes.
- Presentación de conocimiento, donde se utilizan técnicas de visualización y la representación de conocimiento para presentar el conocimiento extraído para el usuario.

La Data Mining comprende los algoritmos básicos que permiten obtener información y conocimientos fundamentales a partir de datos masivos. Es un campo interdisciplinario de conceptos de áreas afines, tales como los sistemas de bases de datos, estadísticas, aprendizaje automático y reconocimiento de patrones. De hecho, la Data Mining es parte de un proceso de descubrir mayor conocimiento. El conocimiento descubierto se puede aplicar para la toma de decisiones, control de procesos, gestión de la información, procesamiento de consultas, entre otros [12].

### 2.1. Algoritmos de Data Mining

Un algoritmo de Data Mining es un conjunto de cálculos y reglas heurísticas que permite crear un modelo de Data Mining a partir de los datos. Para crear un modelo, el algoritmo analiza primero los datos proporcionados, en busca de tipos específicos de patrones o tendencias. El algoritmo usa los resultados de este análisis para definir los parámetros óptimos para la creación del modelo de Data Mining. A continuación, estos parámetros se aplican en todo el conjunto de datos para extraer patrones procesables y estadísticas detalladas [13]. Si bien existen varios algoritmos de análisis de datos que se utilizan en las soluciones de Data Mining, la elección del algoritmo adecuado es todo un desafío ya que se pueden utilizar diferentes alternativas para realizar la misma tarea. Cada uno de ellos puede generar un resultado diferente, y algunos pueden generar más de un tipo de resultado. Por ejemplo, se puede usar el algoritmo

de árboles de decisión no sólo para la predicción, sino también como una forma de reducir el número de columnas de un conjunto de datos, ya que el árbol de decisión puede identificar las columnas que no afectan al modelo de Data Mining final. Por tales motivos para este trabajo debido a que permiten crear grupo de personas con perfiles concretos en función de atributos definidos como datos demográficos y comportamientos se utilizan los siguientes algoritmos de Data Mining [14]:

- *Árboles de Decisión*, este algoritmo constituye uno de los métodos del aprendizaje supervisado más utilizados. Una de sus principales virtudes, es la sencillez de los modelos obtenidos [15]. El algoritmo genera el modelo mediante la creación de una serie de divisiones en el árbol. Estas divisiones se representan como nodos. El algoritmo agrega un nodo al modelo cada vez que una columna de entrada tiene una correlación significativa con la columna de predicción [16]. Un paso crítico en la construcción de un árbol de decisión es definir cómo estructurar estos nodos en el árbol. Una buena guía para esto es la entropía, la cual mide la incertidumbre asociada a un conjunto de datos, y es usada para ayudar a decidir qué atributo debe ser el siguiente en seleccionarse. Básicamente, un atributo es elegido como el siguiente nivel del árbol si puede ayudar a discriminar más objetos, es decir tiende a reducir la entropía [17]. Un árbol de decisión es un conjunto de condiciones o reglas organizadas en una estructura jerárquica, de tal manera que la decisión final se puede determinar siguiendo las condiciones que se cumplen desde la raíz hasta alguna de sus hojas. Un árbol de decisión tiene unas entradas las cuales pueden ser un objeto o una situación descrita por medio de un conjunto de atributos y a partir de esto devuelve una respuesta la cual en definitiva es una decisión que es tomada a partir de las entradas. Los valores que pueden tomar las entradas y las salidas pueden ser valores discretos o continuos. Un árbol de decisión lleva a cabo un test a medida que éste se recorre hacia las hojas para alcanzar así una decisión. El árbol de decisión suele contener nodos internos, nodos de probabilidad, nodos hojas y arcos. Un nodo interno contiene un test sobre algún valor de una de las propiedades. Un nodo de probabilidad

indica que debe ocurrir un evento aleatorio de acuerdo a la naturaleza del problema, este tipo de nodos es redondo, los demás son cuadrados. Un nodo hoja representa el valor que devolverá el árbol de decisión. Las ramas brindan los posibles caminos que se tienen de acuerdo a la decisión tomada [18].

- *Clustering Demográfico*: Proporciona una agrupación rápida y natural de las bases de datos de gran tamaño. El número de agrupaciones que se deben crear se determina automáticamente, y éstas se caracterizan por las distribuciones de valor de sus miembros [19]. El algoritmo de Clustering Demográfico construye conjuntos mediante la comparación de cada registro con todos los grupos creados anteriormente asignando el registro de la agrupación que maximiza una puntuación de similitud [20]. Éste es un proceso iterativo que realiza múltiples pasadas sobre el conjunto de registros antes de converger a un nivel de agrupamiento óptimo. La calidad de una partición se evalúa por una medida global, lo que favorece los grupos con alta similitud. En cada paso del proceso, el algoritmo utiliza este criterio para decidir si asignar un registro en un clúster existente, o si se debe crear uno nuevo. Dicho proceso finaliza cuando los resultados de la iteración no presentan cambios en las agrupaciones [21].

### 3. Modelo Propuesto

El proceso de obtención de los perfiles de rendimiento consta de las siguientes fases: Obtención de datos, Armado del Data Warehouse y Data Mining (Figura 1).

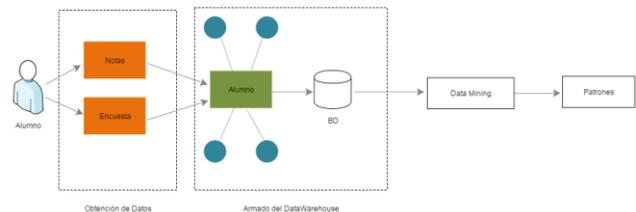


Figura 1. Proceso de obtención de perfiles

#### 3.1. Obtención de Datos

La primera fase consiste en la recopilación de toda la información necesaria para poder alimentar las fases posteriores que permitan determinar los perfiles de

rendimiento académico. Dichos perfiles permiten conectar alumnos con determinadas características socioeconómicas con un determinado rendimiento académico (éxito o fracaso).

Para determinar dichos perfiles primero se requiere de información acerca de los factores académicos previos y socioeconómicos que puedan afectar el rendimiento académico del alumno. Estos aspectos fueron definidos por el grupo de estudio en base a investigaciones realizadas. Algunos de los aspectos abarcados por la misma son: estudios secundarios, residencia actual, tiempo dedicado al estudio, situación laboral de los padres y del alumno, estudios de los padres y consideraciones respecto al uso de las TICs.

Además, se requiere información relativa al rendimiento académico de los alumnos durante el año lectivo, compuesta por las calificaciones que obtuvieron los mismos en las distintas instancias de exámenes parciales y de trabajos prácticos. Aquí sólo se tuvo en cuenta el rendimiento sobre la cursada y no se tomaron en cuenta las calificaciones definitivas de exámenes finales ya que ambas instancias son independientes entre sí y generalmente los alumnos demoran varios años en rendir el examen final luego de haber regularizado la materia.

Por tal motivo en la fase de obtención se precisa la participación tanto de los alumnos como de la cátedra. En primera instancia los estudiantes participaron en forma activa, pues era necesario conocer aspectos personales que no se podían obtener de otra manera que a través de respuestas directas por parte de los alumnos. A tal fin completaron una encuesta online sobre su situación académica y socioeconómica. Los aspectos abarcados por la misma son factores que se ha demostrado a través de los años influyen en el desarrollo académico y que son los anteriormente mencionados.

Al finalizar el cursado la cátedra de Algoritmos y Estructuras de Datos facilita las calificaciones obtenidas por los alumnos en las distintas instancias de exámenes y su condición final (promovido, regular o libre).

### 3.2. Armado del DW

Una vez que se obtuvo la información necesaria, pudo ocurrir que se presentaran incompletitudes, inconsistencias e incoherencias entre los datos. Esto pudo darse porque los valores de los contenidos de muchas características no poseen ciertas restricciones. Por lo tanto, los datos obtenidos en la etapa anterior debieron ser sometidos a un proceso de depuración: eliminación o relleno de campos nulos, corrección de errores tipográficos, e integración con las notas de los exámenes. Este proceso asegura la consistencia y coherencia de los datos que son cargados en el DW.

La estructura del modelo de DW utilizado es muy sencilla, sólo se compone de la tabla de hechos Alumno y varias tablas para las distintas dimensiones asociadas a la misma. La tabla de hechos incluye información específica del alumno y su rendimiento académico, en ella se especifica su situación final, es decir si promovió, regularizó o bien quedó libre en la materia de análisis. Las dimensiones son las características que se desean estudiar; contienen información descriptiva del entorno socio-económico del alumno, obtenida en la encuesta.

### 3.3. Data Mining

Luego de la carga y armado del DW se siguió por la fase de Data Mining. Para ello se seleccionaron las técnicas a utilizar, creándose los flujos de minería correspondiente, en los cuales se parametrizan los respectivos algoritmos.

La selección de los distintos algoritmos utilizados se basó en las ventajas que proveían cada uno de ellos. Por tal motivo se trabajó con Clustering (Cluster Demográfico) el cual permite encontrar caracterizaciones útiles para construir clasificadores, permite también el descubrimiento de grupos y subgrupos que revele la naturaleza de la estructura del problema. El objetivo de esta técnica es obtener grupos o conjuntos entre los elementos, de tal manera que los elementos asignados al mismo sean similares. También se utilizó en mayor medida la técnica de Clasificación con Árboles de Decisión.

Los Árboles de Decisión son fáciles de usar, admiten atributos discretos y continuos, tratan bien los atributos no significativos y los valores faltantes. Su principal ventaja es la facilidad de interpretación, son útiles en problemas de alta dimensionalidad, plantean el problema para que todas las opciones sean analizadas. Su objetivo es realizar clasificaciones sobre los datos conocidos y con estas clasificaciones crear modelos que luego puedan ser utilizados para predecir o clasificar valores nuevos o desconocidos.

Para determinar los perfiles de rendimiento se utilizaron las técnicas descriptas anteriormente, que permitieron realizar un análisis dimensional de los datos considerando como parámetro de minería la variable relacionada con la situación final del alumno, determinada por su estado en la materia al cierre del ciclo lectivo (promovido, regular o libre)

Como resultado se obtuvieron los patrones que determinan el modelo descriptivo de los datos, a partir de los cuales se estimaron los perfiles de rendimiento.

## 4. Resultados

Con el objetivo de determinar patrones de rendimiento académico, se realizaron pruebas sobre los

datos de los alumnos del ciclo lectivo 2013 al 2015, un total de 615 alumnos. Para el análisis se tomó su condición final, al cierre del ciclo lectivo, como parámetro principal, es decir, si el alumno se encontraba regular, libre o promovido. Se consideró en situación de libre a los alumnos que no aprobaron los exámenes parciales o bien abandonaron la materia. Los estudiantes que regularizaron la materia aprobaron los 3 exámenes parciales con una nota igual o superior al 60% pero no alcanzó el 75%. Finalmente, los alumnos que la promovieron son aquellos que en las 3 instancias de exámenes aprobaron con nota superior al 75%.

Según los datos recabados, de los 615 alumnos analizados, el 60% corresponde a libres, 28% a regulares y el 12% a promovidos.

Para definir los perfiles además de la condición final también se tuvieron en cuenta los distintos factores mencionados anteriormente, estudios secundarios, residencia actual, tiempo dedicado al estudio, situación laboral de los padres y del alumno, estudios de los padres y consideración respecto al uso de las TICs.

Analizando cada uno de los factores socioeconómicos y actitudinales y comparándolos con la condición final del alumno, se procedió a aplicar los algoritmos de Clustering Demográfico y Árbol de Decisión y se obtuvieron resultados que se resumieron en la tabla 1, donde pueden observarse los patrones resumidos mayoritariamente observados (observados en la mayoría de los casos de las respectivas categorías de éxito académico -promovidos y regularizados- y de fracaso académico -libres-) según las dimensiones de análisis consideradas.

**Tabla 1. Patrones de características para cada perfil de rendimiento**

Éxito académico	Fracaso académico
Estudian más de 10 horas semanales.	Estudian hasta 10 horas semanales.
Trabajan hasta 35 horas semanales.	No trabajan o Trabajan 36 o más horas semanales.
Sus madres trabajan 36 o más horas semanales.	Sus madres no trabajan.
Sus padres trabajan de 36 o más horas semanales.	Sus padres no trabajan o trabajan hasta 20 horas semanales.
Otorgan más importancia al estudio que a la familia.	Otorgan más importancia al estudio que a la diversión.

Disponen de obra social propia.	No poseen obra social o poseen obra social de los padres.
El mayor porcentaje de éxito académico corresponde al grupo femenino.	Los porcentajes más altos de alumnos libres corresponden al grupo masculino.
Residen en forma independiente.	Residen con familiares o en alguna situación no prevista de residencia.
Sus madres tienen estudios universitarios de postgrado o superiores no universitarios completos.	Sus madres tienen estudios secundarios o universitarios incompletos.
Sus padres tienen estudios universitarios o superior no universitario completo.	Sus padres tienen estudios secundarios o universitarios incompletos o no hicieron estudios.
Consideran que el dominio de las TIC será imprescindible para el ejercicio profesional.	Consideran que las TIC son una realidad en la actualidad y que están de moda.
Su motivación para estudiar es aprender integralmente y aprobar.	Su motivación para estudiar es aprobar la asignatura.

## 5. Conclusiones

El presente trabajo propone un modelo para definir perfiles de rendimiento académico utilizando técnicas de DW y DM basado en la recolección de datos sobre los alumnos durante el cursado de AED en los ciclos lectivos 2013, 2014 y 2015. El trabajo incluye tanto la condición académica final del alumno como la influencia que tiene el entorno socioeconómico, cultural y actitudinal frente a sus estudios a fin de establecer perfiles característicos de éxito o de fracaso académico.

Los perfiles obtenidos permiten determinar, de manera general, que los alumnos con tendencia al éxito son aquellos de sexo femenino, que estudian más de 10 horas a la semana, trabajan hasta 35 horas por semana, cuyos padres y madres trabajan más de 36 horas semanales y poseen los estudios superiores completos, dicen otorgar más importancia al estudio que a la familia, residen de forma independiente, consideran a las TIC como herramientas imprescindibles para el ejercicio profesional, buscan aprender integralmente y aprobar y poseen obra social propia.

En contrapartida, los alumnos con tendencia al fracaso son aquellos de sexo masculino, que estudian hasta 10 horas semanales, no trabajan o trabajan más de 36 horas por semana, cuyos padres no trabajan o trabajan hasta 20 horas semanales y no terminaron sus estudios secundarios y universitarios, dicen otorgar más importancia al estudio que a la diversión, residen con sus familiares, consideran a las TIC sólo como una realidad, buscan sólo aprobar la materia y no poseen obra social.

La determinación de estos perfiles brinda la posibilidad de predecir el futuro rendimiento académico de los alumnos a partir del conocimiento de los factores que los afectan, verificando la correspondencia que tengan cada uno de estos con los perfiles determinados anteriormente.

Estas predicciones son una poderosa herramienta para la cátedra para poder determinar a principio del año lectivo cuáles son aquellos alumnos con alta probabilidad de fracaso académico, lo cual le permite a la misma preparar y definir distintas estrategias pedagógicas para los alumnos identificados, que los ayuden a superar un posible fracaso académico. Como consecuencia, también se pretende contribuir a disminuir el alto nivel de deserción que existe en los primeros años de la carrera.

Se tiene previsto seguir con la incorporación de datos correspondientes a las cohortes 2016 y sucesivas, a los efectos de ir ajustando el modelo predictivo con la incorporación de cada cohorte. También es posible incorporar al DW a alumnos de otras asignaturas y generar para ellos también el modelo predictivo correspondiente y las acciones tendientes a evitar el posible fracaso académico.

## Referencias

- [1] Maletic, J. I. Collard, L. Marcus, A. "Source Code Files as Structured Documents". 10th IEEE International Workshop on Program Comprehension. Paris, France. June 27-29, 2002.
- [2] Oloriz, Mario. Lucchini, María L. Ferrero, Emma. "Relación entre el Rendimiento Académico de Ingresantes en Carreras de Ingeniería y el Abandono de los Estudios Universitarios". Mar del Plata, Argentina. 29 de Noviembre 2007.
- [3] Tinto, V. "Leaving College. Rethinking the Causes and Cures of Student Attrition". Chicago: The University of Chicago Press. 1993.
- [4] Vivo Molina, J. M.; Franco Nicolás, M. y Sánchez de la Vega, M. del M. "Estudio del rendimiento académico universitario basado en curvas ROC". Revista de Investigación Educativa, RIE, vol. 22, n.2. España. 2004.
- [5] García-Jiménez, M. Visitación; Alvarado Izquierdo, Jesús M; Jiménez Blanco, Amelia. "La predicción del rendimiento académico: regresión lineal versus regresión logística". Psicothema, vol. 12, n.2. 2000.
- [6] Di Gresia, L. "Rendimiento Académico Universitario". Tesis Doctoral. Universidad Nacional de La Plata, Argentina. 2007.
- [7] García, M. M.; San Segundo, M. J. "El Rendimiento Académico en el Primer Curso Universitario". X Jornadas de la Asociación de Economía de la Educación. Libro de Actas. España. 2001.
- [8] Inmon, W. H. "Data Warehouse Performance". John Wiley & Sons. USA. 1992.
- [9] Fayyad, U. M.; Grinstein, G.; Wierse, A. "Information Visualization in Data Mining and Knowledge Discovery". Morgan Kaufmann. Harcourt Intl. 2001.
- [10] Larrieta, M. I. Ángeles; Santillán Gómez, Angélica M. "Minería de datos: Concepto, características, estructura y aplicaciones". [En línea, 10/09/16]. Disponible en: <http://www.ejournal.unam.mx/rca/190/RCA19007.pdf>
- [11] Han, Jiawei; Kamber, Micheline; Pei, Jian. "Data Mining: Concepts and Techniques". 3rd edition. Morgan Kaufmann. 2011.
- [12] Zaki, M. J.; Meira, Wagner. "Data mining and analysis: fundamental concepts and algorithms". Rensselaer. Cambridge University Press. 2014.
- [13] Microsoft. TechNet. "Algoritmos de minería de datos (Analysis Services: Minería de datos)". [En línea, 13/04/17]. Disponible en: [https://technet.microsoft.com/es-es/library/ms175595\(v=sql.105\).aspx](https://technet.microsoft.com/es-es/library/ms175595(v=sql.105).aspx).
- [14] Baker, R.; Yaceff, K. "The State of Educational Data Mining in 2009: A Review and Future Visions". Journal of Educational Data Mining. Vol 1, No 1. 2009.
- [15] Roche, Ariel. "Árboles de decisión y Series de tiempo". Tesis de Maestría en Ingeniería Matemática. Facultad de Ingeniería, UDELAR. Montevideo, Uruguay. 21 de diciembre de 2009.
- [16] Microsoft. "Algoritmo de árboles de decisión de Microsoft". [En línea, 13/04/17] Disponible en: [https://msdn.microsoft.com/es-es/library/ms175312\(v=sql.120\).aspx](https://msdn.microsoft.com/es-es/library/ms175312(v=sql.120).aspx).
- [17] Theophano Mitsa. "Temporal Data Mining". 1st Ed. Chapman & Hall/CRC. 2010.
- [18] Vizcaino Garzon, Paula A. "Aplicación de técnicas de inducción de árboles de decisión a problemas de clasificación mediante el uso de WEKA (Waikato Environment for Knowledge Analysis)". Fundación Universitaria Konrad Lorenz. Facultad de Ingeniería de Sistemas. Bogotá, Colombia. 2008.
- [19] IBM Knowledge Center. "Distribution-based Clustering". [En línea, 13/04/17]. Disponible en: [http://www.ibm.com/support/knowledgecenter/SSEPGG\\_9.5.0/com.ibm.im.model.doc/c\\_c\\_distribution\\_based\\_clustering.html](http://www.ibm.com/support/knowledgecenter/SSEPGG_9.5.0/com.ibm.im.model.doc/c_c_distribution_based_clustering.html)
- [20] Brause, Rüdiger; Hanisch, Ernst. "Medical Data Analysis". First International Symposium. ISMDA. Frankfurt, Germany. September 2000.
- [21] Manganaris, S.; Christensen, M.; Zerkle, D.; Hermiz, K. "A Data Mining Analysis of RTID Alarms". IBM. 1999.



UNIVERSIDAD TECNOLÓGICA NACIONAL - FACULTAD REGIONAL SANTA FE  
LAVALSE 610 - S30004EWB SANTAFE - ARGENTINA  
TE +54 (342) 460 1579  
conaiisi2017.frfsf.utm.edu.ar | e-mail: conaiisi@frsf.utm.edu.ar