

## **Sistema Informático para la Detección Temprana de Deserción Estudiantil Universitaria**

Estudio sobre ingresantes de la UTN Regional La Plata

Istvan, Romina <sup>1,3</sup>; Lasagna, Valeria <sup>2</sup>

<sup>1</sup> Grupo de I&D Aplicado a Sistemas Informáticos, GIDAS UTN FRLP

<sup>2</sup> Secretaría de TIC - UTN FRLP / <sup>3</sup> ristvan@firlp.utn.edu.ar

**Resumen** Argentina se caracteriza por ser uno de los países con menor graduación en proporción a la cantidad de estudiantes universitarios (CEA, 2018). Es por ello, que dentro de la Secretaría de Políticas Universitarias (SPU) surge el Plan Estratégico de Formación de Ingenieros (PEFI, 2012) cuyo fin es aumentar los egresados de las distintas carreras de Ingeniería.

Dentro de este contexto, la UTN Facultad Regional La Plata (UTN-FRLP) desde principios del año 2016 se encuentra trabajando en el Proyecto de Investigación y Desarrollo (PID) denominado Estudio Sistemático de Deserción Estudiantil Universitaria (ESDEU) homologado por Rectorado UTN con código TEINNLP0003786. Dicho PID tiene una relevante importancia institucional ya que formaliza el primer estudio sobre deserción en la UTN-FRLP y responde a la necesidad actual de elevar la tasa de graduación efectiva en cada una de las especialidades que brinda la Regional.

El producto final del PID es un software de gestión tutorial basado en un modelo de indicadores de riesgo de deserción estudiantil obtenido mediante Minería de Datos. Dicho sistema identifica a aquellos alumnos próximos a desertar y de esta manera, posibilita la determinación de estrategias de retención ajustadas a las particularidades del estudiante y de la institución.

El Proyecto presenta como objetivo general elevar la tasa de promoción efectiva de Ingenieros, en apoyo a las Políticas Públicas y en concordancia con el Plan PEFI, siendo sus objetivos específicos: (i) Consolidar la primera base de datos unificada para el estudio sistemático de la deserción, la cual recopila información de diversas fuentes: académica, encuestas y asistencias. (ii) Identificar a posibles desertores mediante la visualización de alertas tempranas. (iii) Registrar el estado de avance y seguimiento de los alumnos. (iv) Brindar un modelo de indicadores que represente los principales factores causales de la deserción.

Las actividades llevadas a cabo con el objetivo de realizar el desarrollo del PID profundizan en las características y funcionalidades de herramientas existentes en Latinoamérica, obteniendo puntos de referencia sobre el potencial del nuevo software, enmarcándolo en sistemas de tutorías con impacto en la deserción estudiantil.

En función de lo anterior, se optimizan los objetivos del sistema contextualizando los módulos y fuentes de datos necesarias, identificando los actores con sus roles y perfiles, modelando así la interactividad con el sistema.

Seguidamente se obtiene el Modelo de Indicadores de Deserción, mediante técnicas de Minería de Datos. Y por último se realiza la incorporación de dicho modelo al sistema para realizar la predicción de deserción de cada estudiante.

**Deserción Universitaria; Deserción Estudiantil; Indicadores de Deserción; Ingeniería; Universidad; Retención Estudiantil.**

Recibido: 01/07/2019 Aceptado: 12/11/2019

DOI 10.24215/26838559e006

## **Computer System for the Early Detection of University Students' Abandonment**

### **A Study about Beginner Students at UTN Facultad Regional La Plata**

**Abstract** Argentina is characterised as being one of the countries with the lowest rate of graduate students, considering the number of university students (CEA, 2018). Due to this, the Secretary of University Policies (SPU) develops a Strategic Plan of Development for Engineers (PEFI, 2012) whose aim is to increase the number of graduates of the different engineering courses of studies.

Within this context, from the beginning of 2016, the UTN Faculty Regional La Plata (UTN-FRLP) has been working on a Research and Development Project (PID) called Systematic Study of University Student's Abandonment (ESDEU) homologated by UTN Rectorado under code TEINNLP0003786. This PID has great institutional importance for it formalises the first study on abandonment in the UTN-FRLP as well as it is intended to answer the current need of raising the graduate rate in every single field the regional university offers.

The PID final product is a software of tutorial management based on a model of indicators of a risk of student's abandonment which has been obtained by a Data Miner. This system identifies those students who are about to drop out and it, then, enables strategies of students' permanence adjusted to each student and to the institution.

The main purpose of the project is to increase the effective graduate rate of Engineers, in agreement with the Public Policies and with the PEFI Plan, being the following its specific objectives: (i) Consolidating the first unified data base for the systematic study of dropping out, which gathers information from different sources: academic, questionnaires and attendance. (ii) Identifying future dropouts by considering early alerts. (iii) Registering the students' development and follow-up. (iv) Providing indicator models which represent the main factors causing the dropout.

The activities performed to develop PID deepen the features and functions of the existing tools in Latin America, gaining references about the new potential software, framing it under tutorial systems with impact on students' dropout.

Regarding the above points, the objectives of the system are optimised considering the modules and the necessary data sources, identifying the actors with their roles and profiles, modelling, then, the interactivity with the system.

Then, a Model of Desertion Indicator is obtained from techniques of Date Miner. As a final moment, the model is included to the system to predict a possible dropout.

**College Desertion; Student Desertion; Desertion Indicators; Engineering; University; Higher Education; Student Retention.**

### **Novedad u originalidad local en el conocimiento:**

Dentro de la UTN, el fenómeno de la deserción estudiantil genera un gran número de estudios y programas tendientes a minimizar su impacto, tanto a nivel local como de forma cooperativa con otras instituciones.

El Grupo Interinstitucional de Tutorías de Buenos Aires (GITBA) tiene como objetivo crear un ámbito de intercambio entre las unidades académicas de la región para trabajar en el mejoramiento de los sistemas de tutorías. La Red Iberoamericana de Mentoría (RIME) genera un espacio de comunicación donde participan por Argentina la UNC, la UNSAM y la UTN. También en el Proyecto de Gestión Universitaria Integral del Abandono (GUIA) participan la UTN y la UNC.

A nivel facultad existen diferentes estudios realizados por las regionales en los cuales se analizan las principales causas de deserción: Concepción del Uruguay (Haudemand, Echazarreta, Gianera, Carbone y Osti, 2016); La Plata (Ronconi, Chancel, Del Zotto y Zerbino, 2018; Istvan, Falco y

Antonini, 2017); La Rioja (Alanis y Cova, 2017); Resistencia (La Red Martínez, Giovaninni, Pinto, Frisone, y Báez, 2015); Villa María (Rosso, Oddino, Peralta, Aimar y Vaira, 2014); Córdoba (Oliver, Eimer, Bálsamo y Crivello, 2011); Rosario (Valía et al., 2017) y Buenos Aires (Pollo-Cattaneo et al., 2016) entre otras.

Es así como es factible observar que dentro de la Universidad existe una diversidad de estudios, programas y proyectos sobre esta temática; sin embargo, en este contexto no se encontraron herramientas que haciendo uso de las TIC, integren en un sistema informático de gestión tutorial indicadores tempranos de riesgo de deserción.

### **Grado de relevancia:**

La evolución de la industria, los mercados internacionales aún no explotados y una creciente vinculación entre empresas y universidades, hacen que las Ingenierías sean esenciales para consolidar el desarrollo industrial, económico y científico del país.

Con el fin de aumentar los egresados en estos campos, las carreras de Ingeniería se enmarcan dentro del Plan Estratégico de Formación de Ingenieros (PEFI, 2012) impulsado por la Secretaría de Políticas Universitarias (SPU).

El plan se articula mediante tres ejes principales, el primero de ellos se centra en los “Proyectos de Mejoramiento de Indicadores Académicos” y reúne distintas estrategias con el fin de incrementar la cantidad de graduados en ingeniería en un 50% en 2016 y en un 100% en 2021, con relación al año 2009.

En concordancia con las Políticas Públicas que dieron origen a este plan, la UTN-FRLP trabaja desde el año 2016, en el Proyecto de Investigación y Desarrollo (PID): “Estudio Sistemático de Deserción Estudiantil Universitaria (ESDEU)”. Mediante este PID intenta acentuar su papel en la construcción de la educación universitaria, asumiendo una participación activa en el desarrollo de la industria del país.

### **Grado de pertinencia:**

El Proyecto antes mencionado cuenta con el apoyo institucional UTN-FRLP: Secretaría Académica, Secretaría de Ciencia y Tecnología, Dirección de Curso de Ingreso y el trabajo conjunto de la Secretaría de TIC y el GIDAS (Grupo de Investigación & Desarrollo Aplicado a Sistemas informáticos) donde se desarrolla la línea de investigación.

Particularmente, brinda un aporte sustancial al Plan de Acción de Tutorial que se encuentra en vigencia en la institución, destinado a aquellos alumnos que presentan riesgo de deserción estudiantil. El mismo realiza un seguimiento personalizado e integral de los estudiantes próximos a desertar mediante contactos personales, entrevistas, apoyo pedagógico y acompañamiento académico. Por consiguiente, el sistema propuesto resulta de aplicación inmediata en el ámbito de la institución en el que se desarrolla.

### **Grado de demanda:**

El Equipo Interdisciplinario de la UTN-FRLP, dependiente de la Secretaría Académica, es un espacio de intercambio profesional donde se generan entrecruzamientos de saberes para la toma de decisiones pedagógicas.

Dentro de la gestión tutorial realiza la coordinación, capacitación y monitoreo del Plan de Acción Tutorial para alumnos de primer año, llevando el relevamiento estadístico de ausentismo de

grupos testigos de las distintas especialidades. Asimismo realiza la orientación y acompañamiento de los estudiantes que requieran asistencia personalizada a lo largo de su trayectoria educativa. La nueva herramienta informática brinda un aporte sustancial a las actividades de tutorías vigentes en la institución, donde la Secretaría Académica y el Equipo Interdisciplinario de la UTN-FRLP serán sus máximos beneficiarios.

## Desarrollo del producto

Con el objetivo de realizar el desarrollo del PID se profundiza en las características y funcionalidades de herramientas existentes en Latinoamérica. Colombia y México se encuentran entre los países que mayor desarrollo tecnológico presentan en relación a la temática, destacándose entre ellos: el Sistema de Alerta Temprana - SAT (Carvajal Olaya, Montes García, Trejos Carpintero y Cárdenas, 2016); el Sistema para la Prevención de la Deserción de la Educación Superior - SPADIES (Ministerio de Educación de la República de Colombia - a, 2019); el Sistema de Información para el Monitoreo, Prevención y Análisis de la Deserción Escolar - SIMPADE (Ministerio de Educación de la República de Colombia - b, 2019); el Sistema de Análisis, Seguimiento y Evaluación de la Deserción - SASSED (Reyes Sarmiento, Caballero Domínguez, Rodríguez Pautt, Bolívar Troncoso, 2013) y el Sistema PASPE (Combita, 2014). Del análisis de los sistemas precedentes, se obtienen puntos de referencia sobre el potencial del nuevo software, enmarcándolo en sistemas de tutorías con impacto en la deserción estudiantil. En función de lo anterior, es posible optimizar los *objetivos* del nuevo sistema contextualizando los *módulos* y *fuentes de datos* necesarias. Seguidamente, se identifican los actores con sus roles y perfiles, determinando la interactividad con el sistema. Además se obtiene un *Modelo de Indicadores de Deserción*, mediante técnicas de Minería de Datos y en una etapa posterior se realiza la *Incorporación del Modelo al Sistema* con el fin de predecir el riesgo de deserción de cada estudiante.

### I - Objetivos

Los objetivos generales del proyecto plantean el diseño y desarrollo de un sistema informático para la detección y caracterización temprana de posibles desertores; con el fin de identificar a aquellos alumnos próximos a desertar, y así brindar herramientas adecuadas para la determinación de estrategias de retención ajustadas a las particularidades de la institución.

El sistema presenta como objetivos específicos: (i) Consolidar la primera base de datos unificada para el estudio sistemático de la deserción, la cual recopila información de diversas fuentes: académica, encuestas y asistencias. (ii) Identificar a posibles desertores mediante la visualización de alertas tempranas. (iii) Registrar el estado de avance y seguimiento de los alumnos. (iv) Brindar un modelo de indicadores que represente los principales factores causales de la deserción.

### II - Fuente de Datos

La *Base de Datos de Gestión Académica* perteneciente al sistema SysAcad, provisto por Rectorado UTN, dispone de un número importante de atributos personales y académicos. En ella se registra: materias recursadas, libres, aprobadas, notas de parciales, finales, año de ingreso, asistencias, encuestas y toda otra información derivada de los mismos.

Las *Asistencias* son registradas para todos los cursos con el fin de poder identificar tempranamente las ausencias consecutivas a una materia por parte de los alumnos.

Para poder definir los perfiles de los alumnos desertores y no desertores, la UTN-FRLP utiliza los datos registrados en las *Encuestas* PEFI (Plan Estratégico de Formación de Ingenieros, 2012) llevadas a cabo en el período 2013 / 2017. Las mismas respondían al pedido de la Secretaría de Políticas Universitarias (SPU) y se completaban a través de una plataforma web de forma obligatoria por los mismos estudiantes. Recogían información complementaria a los datos

académicos, registrándose datos Laborales, Familiares y Económicos, como ser: Horas de Trabajo, Relación del Trabajo con la Carrera, Situación Económica de los Padres, Nivel de Instrucción de los Padres, Distancia a la institución y Becas, entre otros.

De forma conjunta, los Datos Académicos, las Asistencias y las Encuestas, conforman una fuente de atributos relevantes cuya disponibilidad es un importante activo en la institución, en la medida en que pueden ser transformados en información de interés para la gestión tutorial.

### III - Módulos del Sistema

El sistema cuenta con los siguientes módulos:

- *Módulo Alertas*: este módulo permite identificar el grado de riesgo de cada estudiante, visualizar las alertas por materia/comisión, generar listas de estudiantes según los riesgos, e identificar los alumnos que necesitan atención prioritaria. El sistema muestra a través de distintos colores alertas visuales con el nivel de prioridad asignada a cada alumno mediante un semáforo de cuatro colores, en relación a la deserción o posible deserción asociada: verde (sin riesgo), amarillo (riesgo leve), naranja (riesgo moderado) y rojo (crítico).

- *Módulo de Alumnos*: permite al Coordinador de Tutores, Profesores, Jefes de Cátedra, Jefes de Departamento y Equipo Interdisciplinario visualizar los datos personales, de contacto y perfil académico del alumno, junto con su categorización respecto de la deserción y el acceso al historial de seguimientos o entrevistas realizadas.

Para el estudiante visualiza alertas relacionadas con su situación académica, como ser vencimientos de cursadas, regularidad y cantidad de horas restantes para quedar libre.

- *Módulo Tutor*: este módulo presenta el listado de estudiantes asignados a un tutor con la posibilidad de filtro, resalta las alertas de estudiantes posibles desertores identificando con distintos colores cuáles de sus tutorados se encuentran en una situación de deserción crítica, permite el acceso al perfil del estudiante: datos básicos de contacto y perfil académico, junto a su seguimiento.

El tutor puede visualizar el indicador y los factores de deserción ya calculados, con la posibilidad de ordenar descendientemente según su indicador de deserción para priorizar la gestión a los registros más críticos. Asimismo facilita la comunicación con el alumno, con docentes y la carga de asistencias/inasistencias, como así también el ajuste manual del semáforo luego de la evaluación personal realizada al estudiante.

El tutor genera cada vez que lo requiera reportes de la situación del alumno y la actividad tutorial llevada a cabo.

- *Seguimiento de Alumnos*: en este módulo el Tutor es el encargado de registrar la información de seguimiento en la cual se planifican las acciones para mejorar el desempeño académico del estudiante.

El tutor durante su gestión puede registrar: fecha y hora de inicio y fin del contacto establecido con el alumno, sus resultados, observaciones y conclusiones.

- *Materia/Comisión/Asistencia*: visualiza el listado de alumnos por materia/comisión, a fin de poder evaluar la cantidad de desertores y posibles desertores para un curso, posibilitando la visualización y carga de asistencias e inasistencias y cambios de comisión.

- *Módulo Comunicación / Mensajería Interna*: permite de manera rápida comunicarse tanto a docentes, alumnos y tutores entre sí.

- *Módulo de Reportes y Estadísticas*: Los reportes son generados en archivos Excel, con filtros avanzados para los usuarios.

- *Módulo Administrador*: gestiona los accesos y permisos por rol a cada módulo del sistema.

### IV - Modelo de Indicadores de Riesgo de Deserción Estudiantil. Perfil de Riesgo

Con el fin de poder definir un modelo de indicadores de riesgo de deserción estudiantil (perfil de riesgo), la UTN-FRLP utiliza los datos almacenados en la base de datos académica SysAcad y los datos recabados por las Encuestas PEFI, aplica la técnica de Explotación de Datos (Minería



Innovación y Desarrollo Tecnológico y Social (2019) 2: 1-15

de Datos o Data Mining) con el fin de detectar patrones y relaciones entre variables de manera automática.

Esta técnica, basada en la extracción de conocimiento, está principalmente relacionada con el proceso de descubrimiento conocido como “Descubrimiento de Conocimiento en Bases de Datos” (Knowledge Discovery in Databases, KDD), el cual se refiere al proceso no-trivial de descubrir conocimiento e información potencialmente útil dentro de los datos contenidos en repositorios de información (Han, Kamber y Pei, 2011).

El proceso de KDD se organiza en base a cinco fases.

*I) Selección de datos.* En esta etapa se determinan las fuentes de datos y el tipo de información a utilizar.

*II) Preprocesamiento de datos*

Esta etapa consiste en la preparación y limpieza de los datos extraídos desde las distintas fuentes de datos. En este paso se utilizan diversas estrategias para manejar datos faltantes o en blanco, datos inconsistentes o que estén fuera de rango, obteniéndose una estructura de datos adecuada para su posterior transformación.

*III) Transformación de datos*

Consiste en el tratamiento preliminar de los datos, transformación y generación de nuevas variables. Aquí se realizan operaciones de agregación o normalización consolidando los datos acorde a la necesidad de la fase siguiente.

*IV) Minería de Datos*

La minería de datos se define como un proceso que reúne un conjunto de técnicas y herramientas de diversas ciencias, especialmente estadística e informática, para extraer conocimiento oculto y patrones no observables en grandes volúmenes de datos (Han et al., 2011).

Esta etapa es la fase de modelamiento en donde diversos métodos son aplicados con el objetivo de extraer patrones previamente desconocidos, válidos, nuevos, potencialmente útiles y comprensibles y que estén contenidos u ocultos en los datos.

*V) Interpretación y Evaluación de Resultado*

En este punto se realiza la interpretación y evaluación de los modelos obtenidos en la etapa anterior a fin de comprobar si cumplen los objetivos planteados en las fases preliminares.

Para los fines de la investigación, se completan las 5 etapas para el grupo de aspirantes e ingresantes a la institución en el período 2013 / 2015, dando como resultado un modelo de indicadores que luego es comparado con el modelo resultante para el período 2016 / 2017. A partir de esto, se obtiene un tercer modelo convergente que se toma como base para la *Estimación del Riesgo de Deserción* en el sistema.

#### *Desarrollo del Modelo*

Luego de la fase de Selección de Datos, en la fase de Preprocesamiento y Transformación de Datos se organiza y representa la vista minable, a la cual se le aplica las técnicas concretas de Minería de Datos. Siguiendo las metodologías de preparación de datos relevadas en la bibliografía y por el conocimiento del dominio, se siguen las siguientes estrategias:

Se eliminan atributos con exceso de datos faltantes y con atributos constantes, como la regional con un valor “FRLP” para todos los registros.

Se eliminan atributos no generalizables como el domicilio del alumno y localidad de nacimiento.

Se eliminan atributos redundantes como el identificador alumno y el legajo.

Se numerizan atributos, como en el caso de: Nivel de Instrucción de la Madre y del Padre (donde 0 representa Nivel Primario Incompleto y 9 Postgrado), Relación del Trabajo con la Carrera (representa la relación del trabajo del alumno, si lo tuviere con la especialidad elegida en su carrera, 0 representa ninguna relación, 1 parcial y 2 total) y Horas de Trabajo Semanales (que

**Innovación y Desarrollo Tecnológico y Social (2019) 2: 1-15**

pasó a tener dos valores: 0-20 horas representa a los alumnos que trabajan medio tiempo y 20-35+ representa a los alumnos que trabajan tiempo completo).

Se utiliza la generalización para transformar atributos de alta cardinalidad como la Edad que pasa a estar definida como rango etario al momento del comienzo de la carrera: 18-20 años / 21-25 años / 26-30 años / 31-40 años / 40+ años. También se generaliza el atributo Estado Civil que pasa a tener solo dos valores posibles: Soltero y Casado/Separado (este último estado incluye a los Unidos de Hecho, Unidos Civilmente, Casados, Separados, Divorciados y Viudos).

Se reemplaza el atributo Localidad de Residencia por el atributo Distancia a la UTN con dos valores posibles hasta 15 km y más de 15 km.

Por otro lado se construyeron nuevos atributos mediante funciones de sumarización, como en el caso de Cantidad de Familiares a cargo que suma tanto a los hijos, como a los familiares a cargo del alumno.

El nuevo atributo Ocupación Familiar representa la combinación de la Ocupación de la Madre y del Padre mediante tres valores posibles: Trabajo Permanente (en el caso de que de que al menos uno de los dos padres tenga un trabajo estable), Trabajo Transitorio (en el caso de que los dos padres tengan trabajo temporario o uno de los dos presente trabajo temporario y el otro desconocido) y Desconoce (si son desconocidos los trabajos de ambos padres).

Finalmente, se establece un atributo de estado Deserción que diferencia a los alumnos que ya han desertado, de los que cursan con regularidad.

La Tabla 1 resume los nombres, descripción y valores de los atributos de la vista minable luego de las etapas de Preprocesamiento y Transformación de Datos.

**Tabla 1. Estructura y los nombres de los atributos de la vista minable**

Atributos	Descripción	Valores
Edad	Rango etario al momento de comenzar la carrera: 18-20 años / 21-25 años / 26-30 años / 31-40 años / 40+ años	18-20 / 21-25 / 26-30 / 31-40 / 40+
Estado Civil	Denota la situación familiar: Soltero / Casado-Separado (incluye los estados: Unido de hecho, Unido civilmente, Casado, Separado, Divorciado, Viudo)	Soltero Casado/Separado
Técnico	Especifica si el título bachiller es Técnico o no.	SI / NO
Horas de Trabajo	Indica las horas de trabajo semanales. Divididas en dos grupos: 0-20 horas representa a los alumnos que trabajan medio tiempo y 20-35+ representa a los alumnos que trabajan tiempo completo.	0-20 hrs. / 20-35+ hrs.
Relación con Carrera	Representa la relación del trabajo del alumno, si lo tuviere, con la especialidad elegida en su carrera.	0: Ninguna / 1: Parcial / 2: Total
Beca	Indica si el alumno posee alguna de las siguientes becas: Contraprestación, Investigación,	SI / NO

	Económica, Universidad, Internacional, Nacional, Provincial, Municipal u Otra Beca.	
Cantidad de Familiares a cargo	Número entero que representa la cantidad de hijos y familiares a cargo total. Incluye: hijos, abuelos, padres, hermanos, ...	Entero (0, ...)
Nivel de Instrucción del Padre / Nivel de Instrucción de la Madre	Especifica el nivel educativo de la madre y del padre.	- (-1) Desconoce - 0 No hizo estudios - 1 Primario incompleto - 2 Primario completo - 3 Secundario incompleto - 4 Secundario completo - 5 Terciario incompleto - 6 Terciario completo - 7 Universitario incompleto - 8 Universitario completo - 9 Postgrado
Ocupación Padre / Ocupación Madre	Tipo de trabajo de la madre y del padre.	0: Trabajo Permanente / 1: Trabajo Transitorio / 2: Desconoce
Ocupación Familiar	Indica Trabajo Permanente en el caso de que de que al menos uno de los dos padres tenga un trabajo estable. Trabajo Transitorio en el caso de que los dos padres tengan trabajo temporario o uno de los dos presente trabajo temporario y el otro desconocido. Y Desconoce si son desconocidos los trabajos de ambos padres.	0: Trabajo Permanente / 1: Trabajo Transitorio / 2: Desconoce
Distancia a UTN	Distancia desde la localidad de residencia a la institución	hasta 15 km / más de 15 km

El análisis predictivo consiste en la extracción de información existente en los datos y su utilización para predecir tendencias y patrones de comportamiento; de esta manera, obtiene un modelo capaz de dar una respuesta ante la presencia de nuevos casos. Si la respuesta que se obtiene es discreta la tarea predictiva se denomina clasificación. En ella, cada registro de la vista minable pertenece a una clase, la cual se indica mediante el valor de un atributo nominal que se denomina etiqueta. Esta característica permite obtener el modelo a través de una estrategia supervisada que, operando sobre el resto de los atributos de cada instancia, busca maximizar la tasa de acierto sobre el conjunto de ejemplos de entrada. Al finalizar el proceso, el clasificador



obtenido es capaz de determinar la clase para cada nuevo ejemplo sin etiquetar (Formia, 2014). Tal como ocurre con el fenómeno de la deserción en el cual un nuevo estudiante puede clasificarse con “perfil de riesgo” o “perfil sin riesgo” de deserción.

Dentro de las tareas de clasificación existen distintas variantes. Para este estudio en particular se utiliza la técnica basada en árboles de decisión; la cual pertenece a la categoría ‘aprendizaje basado en similitudes’ y genera árboles de decisión de manera recursiva, tomando el o los atributos que mejor clasifican a los datos. Un problema que se presenta en la generación de los árboles es que en muchos casos el ajuste es demasiado estricto. La solución a este problema se resuelve con la denominada poda del árbol obtenido, mediante la cual se eliminan los nodos inferiores por ser considerados demasiado específicos.

La herramienta elegida para el análisis de datos es Rapid Miner, RM (RapidMiner Studio 7.1, 2018), la cual implementa más de 500 técnicas de pre-procesamiento de datos, modelación predictiva y descriptiva y visualización de datos.

Se aplica al problema estudiado el algoritmo W-J48 del componente Weka (Weka, 2007). Este algoritmo es una versión del clásico algoritmo de árboles de decisión C4.5 propuesto por Quinlan en 1993.

Para la poda del árbol se tiene en cuenta el Factor de Confianza C (Confidence Level) que incide en la capacidad de predicción y tamaño del árbol resultante. El valor por defecto de este factor es del 25% y conforme este va disminuyendo, se permiten más operaciones de poda y por lo tanto llegar a árboles cada vez más pequeños (García y Álvarez, 2010). Otro factor que se utilizó para variar el tamaño del árbol fue a través del parámetro M que especifica el mínimo número de instancias o registros por nodo del árbol (Witten y Frank, 2000).

A continuación se presentan los resultados obtenidos al ejecutar el proceso con el algoritmo de clasificación en árboles W-J48, implementado en WEKA y ejecutado desde la interface de RM para los dos períodos de estudio. En ambos estudios se utiliza un valor del Factor de Confianza para la poda del 20%.

*a) Modelo de Indicadores de Riesgo de Deserción Estudiantil para el período 2013 / 2015. Perfil de riesgo.*

Los resultados obtenidos mediante el método elegido para el conjunto de estudiantes en el período 2013 / 2015 (Istvan, Chong Arias y Antonini, 2016) demuestran que:

El Nivel de Instrucción de la Madre es por sobre todo uno de los factores más relevante. El algoritmo demuestra que si la madre posee un nivel igual o mayor a Secundario Completo (Nivel 4), el alumno posee mayor probabilidad de no desertar. El Nivel de Instrucción del Padre presenta en este modelo una baja incidencia.

Otro factor clasificador con nivel de incidencia alto para los alumnos son las Horas de Trabajo semanales. Si el alumno no trabaja o trabaja medio tiempo, tiene mayor probabilidad de no desertar por sobre los alumnos que trabajan tiempo completo. Este factor influye de manera significativa en el modelo.

Por otro lado, la Relación del Tipo de Trabajo con la Especialidad de la Carrera es otro atributo a considerar. Cuanto mayor es esta relación, menor es la probabilidad de deserción. Este atributo presenta un nivel de incidencia medio sobre el modelo.

El nivel económico familiar reflejado en la Ocupación de la Madre, del Padre y Familiar influye sobre los grupos desertores. Las familias con trabajo Transitorio o Permanente, tienen una mayor probabilidad de no desertar. La Ocupación del Padre presenta un nivel de incidencia medio sobre el modelo, mientras que la Ocupación de la Madre y la Ocupación Familiar influyen de manera más relevante.

Innovación y Desarrollo Tecnológico y Social (2019) 2: 1-15

La Distancia a la Institución es otro factor con alta incidencia; si la distancia entre la institución y la localidad de residencia es mayor a 15 kilómetros los estudiantes presentan una mayor probabilidad de desertar.

El atributo Técnico que representa el Tipo de Bachillerato posee un nivel de incidencia bajo en el modelo obtenido.

En resumen, el modelo obtenido que presenta las causas de deserción para los estudiantes ingresantes a la UTN-FRLP entre los años 2013 / 2015 se puede resumir de la siguiente manera: Factores de incidencia ALTA: Nivel de Instrucción de la Madre, Horas de Trabajo, Ocupación Familiar y de la Madre y Distancia a la institución.

Factores de incidencia MEDIA: Relación del Trabajo con la Especialidad, Ocupación del Padre. Factores de incidencia BAJA: Instrucción del Padre, Técnico (Tipo de Bachillerato).

*b) Modelo de Indicadores de Riesgo de Deserción Estudiantil para el período 2016 / 2017. Perfil de riesgo.*

Luego de correr el proceso para el conjunto de ingresantes en el período 2016 / 2017 se obtienen los siguientes resultados:

El Nivel de Instrucción de la Madre y del Padre son factores relevantes en el perfil. Si el Nivel de Instrucción de la Madre es mayor a Secundario Incompleto (nivel 3) y a Terciario Completo (nivel 6) y el Nivel de Instrucción del Padre es mayor a Terciario Incompleto (nivel 5), aunque combinados con otros factores, demuestran que los alumnos en estos casos poseen mayor probabilidad de no desertar. Ambos niveles de instrucción presentan un nivel de incidencia alto sobre el modelo obtenido.

Cuanto menor es la Relación del Trabajo con la carrera, mayor es la probabilidad de no desertar. Los alumnos ya insertos en trabajos relacionados, con horarios de tiempo completo, poseen mayor probabilidad de desertar en relación a los alumnos que trabajan en tareas no relacionadas con la especialidad elegida. De esta manera, las Horas de Trabajo y la Relación del mismo con la Carrera se presentan como factores altamente influyentes en la deserción.

La Ocupación Familiar también constituye un factor alto sobre los perfiles obtenidos; cuando la misma es Transitoria la probabilidad de deserción del estudiante es mayor.

La Distancia a la UTN y Cantidad de Familiares a Cargo se presentan como factores de importancia media.

En resumen, es factible observar:

Factores de incidencia ALTA: Instrucción del Padre, Instrucción de la Madre, Ocupación Familiar, Horas de Trabajo y Relación del Trabajo con la carrera.

Factores de incidencia MEDIA: Cantidad de Familiares a Cargo y Distancia a la UTN.

*c) Modelo de Indicadores de Riesgo de Deserción Estudiantil Convergente. Confrontación de resultados. Perfil de Riesgo.*

La Tabla 2 visualiza para los modelos obtenidos la incidencia de cada factor sobre el fenómeno de la deserción y la convergencia resultante.

**Tabla 2. Confrontación de modelos obtenidos. Incidencia de Factores.**

	Ingresantes 2013 / 2015	Ingresantes 2016 / 2017	Convergencia
Instrucción Padre	BAJA	ALTA	MEDIA
Instrucción Madre	ALTA	ALTA	ALTA
Técnico (Tipo de Bachillerato)	BAJA	NULA	BAJA
Horas de Trabajo	ALTA	ALTA	ALTA
Relación del Trabajo con la Carrera	MEDIA	ALTA	ALTA
Distancia a la Institución	ALTA	MEDIO	ALTA
Ocupación Familiar	ALTA	ALTA	ALTA
Ocupación de la Madre	ALTA	NULA	MEDIO
Ocupación del Padre	MEDIA	NULA	BAJA
Cantidad Familiares a Cargo	NULA	MEDIA	BAJA

Al confrontar los modelos, se puede observar que los mismos convergen, dando como resultado los siguientes niveles de incidencia de los indicadores:

Factores de incidencia ALTA: Instrucción de la Madre, Horas de Trabajo, Relación del Trabajo con la Carrera, Distancia a la Institución y Ocupación Familiar.

Factores de incidencia MEDIA: Instrucción del Padre y Ocupación de la Madre.

Factores de incidencia BAJA: Técnico (Tipo de Bachillerato), Ocupación del Padre y Cantidad de Familiares a Cargo.

**Instrucción Familiar:** Nivel de Instrucción de la Madre y del Padre

En el primer modelo es posible discernir que si la madre del alumno posee un nivel igual o mayor a ‘Secundario Completo’, tiene mayor probabilidad de no desertar. Esto coincide básicamente con los resultados obtenidos en el segundo modelo, donde además se incorpora el Nivel de Instrucción del Padre como factor importante en el fenómeno.

**Ocupación Familiar.** Ocupación del Padre y de la Madre

En el modelo convergente se presenta la Ocupación Familiar con nivel de incidencia alto, la Ocupación de la Madre con nivel medio y la Ocupación del Padre con nivel bajo. Más allá del

Innovación y Desarrollo Tecnológico y Social (2019) 2: 1-15

tipo de incidencia, la situación económica reflejada en estos tres atributos incide sobre el modelo obtenido de riesgo de deserción estudiantil.

**Horas de Trabajo y Relación del Trabajo con la Carrera**

En el modelo final se observa que estos factores inciden en la deserción. Estos dos atributos tienen una alta incidencia.

**Distancia a la institución**

La distancia a la institución se presenta con factor de incidencia alto en el primer modelo y medio en el segundo, motivo por el cual se la considera con un nivel alto en el modelo final.

**Cantidad de familiares a cargo**

Este factor no se hace presente en el primer modelo. En contraposición a esto, en el segundo modelo se presenta con un nivel medio. Por tal motivo, se incorpora a los indicadores convergentes con un nivel bajo de incidencia.

**Técnico. Tipo de Bachillerato**

Este atributo no se hace presente en el segundo estudio pero se presenta con incidencia baja en el primer grupo. Debido a esto, se considera con influencia baja en el modelo resultante.

*V - Estimación de Riesgo de Deserción*

Para determinar la *Estimación de Riesgo de Deserción de cada Estudiante* se suman al modelo anteriormente obtenido dos enfoques complementarios: las Inasistencias y el Rendimiento Académico. De esta manera, se constituyen tres ejes de análisis:

(I) Modelo de Indicadores de Riesgo de Deserción Estudiantil Convergente. Perfil de Riesgo.

(II) Inasistencias (Cursadas Libres y Faltas para el ciclo lectivo en curso).

(III) Rendimiento Académico (Promedio, Finales y Parciales).

Cada uno de estos ejes se asocia con una medida de importancia relativa o peso, que representa la incidencia sobre la estimación. A mayor peso, mayor incidencia. Dichas medidas se encuentran en etapa de análisis.

La *Estimación de Riesgo de Deserción* obtenida para cada alumno en particular se visualizará en el sistema de gestión tutorial mediante un semáforo de colores: verde (sin riesgo), amarillo (riesgo leve), naranja (riesgo moderado) y rojo (crítico), junto a su valor numérico representativo.

## **Agradecimientos**

---

El equipo de investigación agradece a la Profesora María Inés Giuliano de la Vega, por su valiosa contribución al realizar las traducciones de este trabajo al idioma inglés.

## **Referencias bibliográficas**

---

Alanis, María Eugenia; Cova, Walter. (2017). Consideraciones preliminares para un análisis de deserción y rezago en la carrera de Ingeniería Electrónica (UTN-FRLR); IX Congreso Iberoamericano de Educación Científica y del I Seminario de Inclusión Educativa y Sociodigital (CIEDUC 2017)

ALFA-GUIA; Proyecto GUIA (Gestión Universitaria Integral del Abandono). Recuperado en mayo de 2019 de <http://clabes-alfaguia.org/clabes-2018/presentacion.html>

La Red Martínez, David Luis; Giovaninni, Mirta Eve; Pinto, Noelia; Frisone, Noelia; Báez, María Eugenia. (2015). Determinación de Perfiles de Rendimiento Académico en la UTN-FRRe. CONAIIISI 2015. UTN Facultad Regional Buenos Aires.

Carvajal Olaya, P., Montes García, H., Trejos Carpintero, Álvaro, & Cárdenas, J. (2016). Sistema de Alertas Tempranas: una herramienta para la identificación de riesgo de deserción estudiantil, seguimiento académico y monitoreo a estrategias. Congresos CLABES. Recuperado a partir de <https://revistas.utp.ac.pa/index.php/clabes/article/view/914>

CEA Centro de Estudios de la Educación Argentina. (2018). Universidad de Belgrano. Recuperado en junio de 2019 de [http://boletin.ub.edu.ar/comunicaciones/flyers/cea\\_junio\\_2018.pdf](http://boletin.ub.edu.ar/comunicaciones/flyers/cea_junio_2018.pdf)

Combata, Harold. (2014), Plataforma Tecnológica para Disminuir la Deserción Estudiantil en la Universidad de la Costa. Conferencia TICAL 2014, Colombia.

Formia, Sonia (2014); La deserción en cursos universitarios. Construcción de modelos sobre datos de la Universidad Nacional de Río Negro usando técnicas de extracción de conocimiento. Tesis de Magister en Tecnología Informática Aplicada en Educación. Institución otorgante: Universidad Nacional de La Plata. <http://sedici.unlp.edu.ar/handle/10915/33855> - Último acceso agosto de 2019.

García, M.; Álvarez, A. (2010). Análisis de Datos en WEKA –Pruebas de Selectividad”. <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/28.pdf>. Accedido en marzo de 2018.

Grupo Interinstitucional de Tutorías de la Provincia de Buenos Aires (GITBA). Recuperado en mayo de 2019 de <https://www.gitba.org>

Han, J. Kamber, M. y Pei, J. (2011). Data mining concepts and techniques [Conceptos y técnicas de minería de datos]. EE.UU.: Morgan Kaufmann.

Haudemand, Norma Yolanda; Echazarreta, Darío Rodolfo; Gianera, Jorge Eduardo; Carbone, Daniel Alberto; Osti, María Javiera. (2016). Deserción y desgranamiento de alumnos en los dos primeros años de las carreras de grado de la FRCU; Jornada Interna de Difusión y de Investigación en Ciencia Y Tecnología de la FRCU; <http://hdl.handle.net/123456789/2142>

Istvan Romina, Chong Arias Carlos, Antonini Sergio (2016); “Sistema de Indicadores de riesgo de deserción para la UTN-FRLP”, , CYTAL 2016 – 7º Jornadas de Ciencia y Tecnología, 12 al 14 de octubre de 2016, UTN FRVM.

Istvan, Romina; Falco, Mariana; Antonini, Sergio. (2017). Análisis y Detección Temprana de Deserción Estudiantil en UTN FRLP; (WICC 2017, ITBA, Buenos Aires); Red de Universidades con Carreras en Informática (RedUNCI); ISBN: 978-987-42-5143-5.

Ministerio de Educación Nacional de la República de Colombia - a. Recuperado en junio de 2019 de <http://www.mineducacion.gov.co/sistemasdeinformacion/1735/w3-article-254648.html>



Innovación y Desarrollo Tecnológico y Social (2019) 2: 1-15

Ministerio de Educación Nacional de la República de Colombia - b. Recuperado en junio de 2019 de <http://www.mineducacion.gov.co/sistemasdeinformacion/1735/w3-propertyvalue-48107.html>

Oliver, María C.; Eimer, Griselda; Bálsamo, Nancy; Crivello, Mónica. (2011). Permanencia y abandono en química general en las carreras de ingeniería de la Universidad Tecnológica Nacional - Facultad Regional Córdoba (UTN-FRC), Argentina; Avances en Ciencias e Ingeniería, ISSN-e 0718-8706, Vol. 2, N°. 2, 2011, pág. 117

PEFI Plan Estratégico de Formación de Ingenieros (2012/2016); Secretaría de Políticas Universitarias, Ministerio de Educación, Presidencia de la Nación Argentina; <http://pefi.siu.edu.ar/>, 2012. Accedido en abril de 2018.

Pollo-Cattaneo, María Florencia; Pytel, Pablo; Vegega, Cinthia; Ramón, Hugo Dionisio; Deroche, Ariel; Straccia, Luciano; Bernal, Luciano; Acosta, Mariana Paola. (2016). Implementación de sistemas inteligentes para la asistencia a alumnos y docentes de la carrera de Ingeniería en Sistemas de Información; Red de Universidades con Carreras en Informática (RedUNCI); XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016, Entre Ríos, Argentina), ISBN: 978-950-698-377-2

RapidMiner Studio 7.1. <https://rapidminer.com/>, Accedido en marzo de 2018.

Red Iberoamericana de Mentoría (RIME). Recuperado en junio de 2019 de <https://mentoriaiberoamerica.org/es/>

Reyes Sarmiento, María; Caballero Dominguez, Carmen; Rodríguez Pautt, Amir; Bolívar Troncoso, Adriana. (2013). Sistema de Análisis y Seguimiento a la Deserción como instrumento integrador de esfuerzos en pro de la permanencia estudiantil en la UNIMAGDALENA. III Conferencia Latinoamericana sobre el Abandono en la Educación Superior. Universidad Nacional Autónoma de México

Ronconi, Jorge; Chancel, Myrian; Del Zotto, Ruben; Zerbino, Lia. (2018). Ingreso y deserción, motivos condicionantes; V Jornadas Nacionales y I Latinoamericanas de Ingreso y Permanencia en carreras Científico-Tecnológicas; <http://hdl.handle.net/123456789/2306>

Rosso, Martha; Oddino, Sonia; Peralta, José; Aimar, Jaquelina; Vaira, Stella. (2014). Características socioeducativas y desgranamiento temprano de estudiantes de ingeniería en la FRVM - UTN; Revista Cuartas Jornadas de Ingreso y Permanencia en Carreras Científico Tecnológicas; Editor Asociación de Profesores de la Facultad de Ciencias Exactas e Ingeniería de la Universidad Nacional de Rosario

Valía, Luciano; Rostagno, José; Moine, Juan Miguel; Bigatti, Cristian; Riva, Fabiana María; Amar, Eduardo. (2017). Minería de datos aplicada a la educación: modelo de deserción universitaria en la Universidad Tecnológica Nacional, Facultad Regional Rosario; XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, ITBA, Buenos Aires); Red de Universidades con Carreras en Informática (RedUNCI); ISBN: 978-987-42-5143-5; p. 301-304

Weka (2007); Data Mining Software in Java; <http://www.cs.waikato.ac.nz/ml/weka/> Acceso marzo de 2018.

**Innovación y Desarrollo Tecnológico y Social (2019) 2: 1-15**

Witten I, Frank E.; (2000); Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. San Francisco: Morgan Kaufmann Publishers; pag. 365