

Modelo para Predecir la Cantidad de Graduados de Ingeniería de UTN Aplicando Técnicas de Minería de Datos

Carrizo Claudio, Saldarini Javier, Ribotta Gabriela, Cardona Fernando, Marotti Juan Ignacio
 Grupo de Investigación Base de Datos - Facultad Regional San Francisco
 Universidad Tecnológica Nacional
 Av. de la Universidad 501 - San Francisco - Córdoba - Tel. 03564-421147
 {cjarrizo77, saldarinijavier, garibotta, ferdcardona, nachomarotti} @gmail.com

RESUMEN

Desde la década del 60, la población mundial ha crecido en forma exponencial, este crecimiento ha generado problemas en sistemas energéticos, sanitarios, telecomunicaciones, infraestructura, etc.

La ingeniería cumple un factor fundamental para el desarrollo económico y el bienestar social de un país. En el ámbito de la República Argentina, será necesario contar, entre otros, con una mayor cantidad de graduados en carreras de ingeniería para incrementar el desarrollo industrial, la innovación productiva y la expansión económica.

En nuestro país, la Universidad Tecnológica Nacional (UTN) aporta el 42,75% de los ingenieros, por lo que esta cifra la posiciona como una de las principales Universidades formadora de ingenieros de Argentina.

El propósito del proyecto consiste en la elaboración de un modelo que permita, por un lado, predecir la cantidad de graduados de carreras de ingeniería en el ámbito de la UTN, y por otro lado, que este mismo modelo permita identificar

cuáles son los patrones que determinarán la graduación de estudiantes de ingeniería en un plazo promedio de 8 años.

Los resultados de este proyecto representarán un aporte para la gestión académica en lo respecta a la planificación, seguimiento y control de las cohortes de los estudiantes de carreras de ingeniería.

Palabras clave: Modelo Predictivo - Minería de Datos - Ingeniería - UTN

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 de la UTN Facultad Regional San Francisco, las cuales son:

- Gestión de procesos de negocios
- Ingeniería de software

- Gestión y tecnologías de las organizaciones
- Calidad de Software
- Seguridad de la información
- Bases de datos

Estas áreas 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 se encuentra homologado y financiado por la Secretaría de Ciencia, Tecnología y Posgrado de la Universidad Tecnológica Nacional bajo el código UTN3936 y según la Disposición SCTyP N° 380/15, el periodo de ejecución es desde Enero de 2016 hasta Diciembre de 2018, el mismo está incluido en el Programa I&D + i de Tecnología de las Organizaciones de la Universidad Tecnológica Nacional.

1. INTRODUCCIÓN

Desde el año 1960, la población mundial ha crecido más del doble [1], esto ocasionó un incremento en la demanda de servicios como agua potable, energía, comunicaciones, infraestructura, etc. Este crecimiento además trajo serios inconvenientes en lo que respecta al cuidado del medio ambiente, desafíos importantes en lo que hace a telecomunicaciones y el colapso de sistemas energéticos y sanitarios.

La ingeniería es un factor clave y prioritario para el desarrollo económico y el bienestar social [2]. Sin dudas, los ingenieros son una parte fundamental a la

solución de estos problemas, a través del desarrollo tecnológico y la innovación. Según se puede observar [3], los países que apuesten a la formación de ingenieros, serán aquellos con mayor índice de industrialización y desarrollo. También hay aportes de datos estadísticos que indican que en China hay 1 ingeniero cada 2000 personas, en Alemania 1 cada 2300, en Brasil 1 cada 6000 y finalmente en Argentina 1 de cada 6600.

En nuestro país, la evolución de la industria, mercados internacionales aun no explotados, y una creciente vinculación entre empresas y universidades, hacen que las ingeniería sean esenciales para consolidar el desarrollo industrial, relacionar conocimiento con innovación productiva, y disminuir los niveles de dependencia tecnológica [4]. En este sentido, es necesario incrementar la cantidad de graduados de ingeniería por cantidad de habitantes, según está establecido en el "Plan Estratégico de Formación de Ingenieros (PEFI)", cuyo objetivo principal es lograr 1 ingeniero cada 4000 habitantes [4].

Según un informe realizado por la Secretaría de Políticas Universitarias (SPU), la Universidad Tecnológica Nacional (UTN) forma alrededor del 42.75% de los ingenieros que se gradúan en nuestro país, esta cifra marca la importancia relativa que tiene la UTN respecto de las demás universidades en carreras de Ingeniería y la posiciona a como la Universidad de Ingeniería más grande del país [5].

En vista de la necesidad de contar con un mayor número de graduados de Ingeniería en nuestro país, el objetivo principal del proyecto consiste en elaborar un modelo que permita predecir

estimativamente la cantidad de graduados de ingeniería que puede proveer la UTN para el año 2021. Además este modelo permitirá identificar cuáles son los patrones que determinarán la graduación de estudiantes de ingeniería en un plazo promedio de 8 años.

Para llevar adelante este proyecto se ha seleccionado la temática "Minería de Datos", específicamente lo que respecta a técnicas de minería de datos de tipo predictivas [6]. La minería de datos se origina a principios de los años 80 y es un campo de las ciencias de la computación referido al proceso del descubrimiento de patrones en grandes volúmenes de conjuntos de datos. Está compuesta por técnicas, las cuales provienen de la rama de la inteligencia artificial y de la estadística, y no son más que algoritmos que se aplican sobre un conjunto de datos para obtener resultados [7]. Estos algoritmos se clasifican en Supervisados o Predictivos (entre los más representativos se encuentran los modelos de regresión, árboles de decisión, redes neuronales) y No Supervisados o del descubrimiento del conocimiento (entre los más representativos se encuentran los la segmentación y el análisis clúster). . Dentro de los algoritmos supervisados o predictivos, según lo expuesto por investigaciones de diferentes autores [8] [9] [10], la técnica de minería de datos que más se adapta para la construcción de un modelo predictivo es la de árboles de decisión con sus algoritmos C4.5 y ID3.

Existen en el mercado herramientas de minería de datos que permiten el análisis de datos y construcción de modelos predictivos, algunas de ellas son de tipo comerciales, como por ejemplo, SPSS Clementine [11], SAS Enterprise Miner [12], mientras que otras son de tipo Open

Source, como ejemplo, Weka [13] y Rapid Miner [14], entre otras.

Si bien existen muchos trabajos de investigación referidos a la predicción del rendimiento y deserción universitaria aplicando técnicas de minería de datos [15] [16], hay muy pocos trabajos enfocados en la obtención de modelos que permitan, por un lado, predecir la cantidad de graduados en un plazo de tiempo determinado para el contexto universitario y más específicamente para carreras de ingeniería, y por otro, que también estos modelos permitan identificar cuáles son los patrones que determinarán que los estudiantes de ingeniería puedan graduarse en un plazo promedio de 8 años.

El modelo resultante se transformará en un instrumento para la gestión, seguimiento, planificación y control en el área académica, posibilitando no sólo la predicción de la cantidad de graduados en carreras de ingeniería, sino que también, constituirá una herramienta de utilidad para ejecutar acciones concretas tomando como referencia los patrones identificados en el modelo.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Una de las áreas prioritarias establecidas por la especialidad Sistemas de Información en la política de desarrollo de Ciencia y Tecnología de la Facultad Regional San Francisco es la de Base de Datos. Dentro de este contexto, el presente proyecto aborda la temática "Minería de Datos".

Los ejes de trabajo para esta línea de investigación se detallan a continuación:

- Caracterización y selección de las técnicas de minería de datos más adecuadas para llevar a cabo la predicción.
- Caracterización y selección de las herramientas que permitan analizar y procesar los datos para llevar a cabo la predicción.
- Construcción del modelo predictivo a través de la herramienta y técnica de minería de datos seleccionada.

3. RESULTADOS OBTENIDOS/ESPERADOS

El resultado esperado es un modelo que permita determinar estimativamente la cantidad de graduados en carreras de Ingeniería que puede proveer la UTN para el año 2021 e identificar los patrones que permiten determinar la graduación de estudiantes de ingeniería en un plazo promedio de 8 años. Para poder lograr este resultado, se han propuesto los siguientes objetivos:

- Construir el perfil del estudiante de ingeniería de UTN.
- Caracterizar y seleccionar las técnicas de minería de datos más adecuadas para llevar a cabo la predicción.
- Caracterizar y seleccionar las herramientas que permitan analizar y procesar los datos para llevar a cabo la predicción.
- Determinar la cantidad de graduados de ingeniería de UTN para el año 2021
- Identificar los patrones que permiten que los estudiantes puedan graduarse en

carreras de ingeniería en un plazo de 8 años.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está constituido por un director de proyecto, un co- director, dos docentes investigadores y un alumno de la especialidad Sistemas de Información que está cursando actualmente el quinto nivel de la carrera Ingeniería en Sistemas de Información y antes de fin de año comenzará con su tesina de grado.

Uno de los docentes investigadores tiene a su cargo la cátedra Gestión de Datos en la especialidad Sistemas de Información, en donde en una de las unidades del contenido curricular se trata la temática Minería de Datos. Los demás docentes que participan del proyecto también están dictando cátedras que tienen relación con la temática Base de Datos. Por su parte, el alumno mencionado anteriormente, viene participando y desarrollando tareas de investigación desde el año 2015. Dos de los docentes están desarrollando actualmente su tesis de maestría en calidad de software.

5. BIBLIOGRAFÍA

- [1] Día Mundial de la Población. Recuperado el 02 de febrero de 2016. https://www.unicef.org/lac/flash/DW/dia_mundial_de_la_poblacion.htm
- [2] La Ingeniería y su papel fundamental en la sociedad. Recuperado el 15 de febrero de año 2016. <http://www.ingenieriyempresa.org/ingenieria-sociedad/>

- [3] ¿Por qué faltan?. Recuperado el 21 de Marzo de 2016. <http://www.revistanueva.com.ar/portal/verNota/248>
- [4] Plan Estratégico de Formación de Ingenieros (PEFI). Recuperado el 4 de abril de 2016. <http://pefi.siu.edu.ar/>
- [5] La UTN forma más del 40% de los ingenieros que se gradúan en el país. Recuperado el 20 de abril de 2016. <https://www.frba.utn.edu.ar/dia-de-la-ingenieria-la-utn-forma-mas-del-40-de-los-ingenieros-que-se-graduan-en-el-pais/>
- [6] María Pérez Márquez. Minería de datos a través de ejemplos. México, 2015. Edit. Alfaomega.
- [7] Weiss y Indurkha. Predictive Data Mining: A Practical Guide, San Francisco, 1998.
- [8] D. Heredia y W. Nieto. Generación de un modelo predictivo para determinar el desempeño académico en la asignatura fundamentos de programación II del programa de Ingeniería de Sistemas. Colombia, 2011
- [9] S. Valero Oreal, A. Salvador Vargas y M. García Alonso. Minería de datos: predicción de la deserción es colar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos, 2010
- [10] E. Rodallegas R., G. Torres y B. Gaona C. Modelo Predictivo para la determinación de causas de reprobación mediante minería de datos. México, 2010.
- [11] Software SPSS Clementine de IBM Recuperado el 10 de marzo de 2017. <http://www-03.ibm.com/software/products/es/spss-modeler>
- [12] Software SAS Enterprise Miner Recuperado el 20 de mayo de 2016. https://www.sas.com/en_us/software/enterprise-miner.html
- [13] Weka 3: Data Mining Software in Java. Recuperado el 24 de mayo de 2016. <http://www.cs.waikato.ac.nz/ml/weka/>
- [14] Data Science Behind Every Decision. Recuperado el 6 de junio de 2016. <https://rapidminer.com/>
- [15] Yegny Karina Amaya Torradoa, Edwin Barrientos Avendañoa, Diana Judith Heredia Vizcaíno. Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos. Colombia.
- [16] Sergio Celis, Luis Moreno, Patricio Poblete, Javier Villanueva, Richard Weber. Un modelo analítico para la predicción del rendimiento académico de estudiantes de ingeniería. Chile, 2015.