

Determinación de Perfiles de Rendimiento Académico en la UTN - FRRe

David L. la Red Martínez, Mirtha E. Giovannini, Noelia Pinto, Martín Frisone, M. Eugenia Báez
Grupo de Investigación Educativa / Departamento de Ingeniería en Sistemas de Información /
Facultad Regional Resistencia / Universidad Tecnológica Nacional
French 414, (3500) Resistencia, Argentina, +54-379-4638194
laredmartinez@gigared.com meg_c51@yahoo.com.ar
ns.pinto@gmail.com martinf032@gmail.com maeugeniabaez@gmail.com

Resumen

El rendimiento académico es un factor crítico en toda institución educativa debido a que se encuentra directamente relacionado con la deserción estudiantil. Es decir, el bajo rendimiento está asociado a una alta tasa de deserción. Por este motivo es altamente deseable poder definir perfiles de estudiantes que puedan ayudar a establecer patrones de rendimiento que conduzcan al éxito o fracaso académico. Con el avance de las tecnologías de la información, han aparecido técnicas y metodologías que permiten el análisis de grandes volúmenes de datos para explicar sus patrones de comportamiento, evolución y singularidades. Estas herramientas utilizan almacenes de datos (Data Warehouses) y minería de datos (Data Mining) para encontrar relaciones no evidentes entre los datos.

En este artículo se describe un modelo basado en Data Warehousing y Data Mining para determinar los perfiles de rendimiento académico en la asignatura Algoritmos y Estructura de Datos de la carrera Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional-Facultad Regional Resistencia (UTN-FRRe), con el objetivo de caracterizar los perfiles de estudiantes con un alto y bajo rendimiento académico. La importancia de la determinación de estos perfiles radica en que permiten definir acciones concretas para revertir el bajo desempeño académico. Finalmente, en este trabajo se muestran los resultados del análisis realizado con alumnos que han cursado la asignatura Algoritmos y Estructuras de Datos durante el ciclo lectivo 2014, comparándolos con los resultados obtenidos durante el ciclo lectivo 2013.

1. Introducción

Mantener la calidad académica en las instituciones educativas es un reto constante. Aún más si se trata del nivel universitario. Por ello, se revisan constantemente

los contenidos, las estrategias y métodos de enseñanza en busca de asegurar estándares adecuados de calidad que den lugar a la formación de profesionales altamente calificados y con un perfil acorde a las necesidades de la sociedad. Claramente el rendimiento académico es un factor crítico asociado a este objetivo. Específicamente, se define al rendimiento académico como la productividad del sujeto, matizado por sus actividades, rasgos y la percepción más o menos correcta de los cometidos asignados [1]. Es decir, el rendimiento académico está ligado a muchos factores que determinan si un alumno adquiere los conocimientos y capacidades de manera adecuada. Obviamente este rendimiento está ligado a las calificaciones que obtiene cuando esos conocimientos y capacidades son evaluados. Desde el punto de vista netamente cuantitativo, la revisión de los resultados obtenidos define el perfil del alumno, pero muchas veces lo que importa realmente son los factores que se hallan detrás de sus calificaciones. Por ello, no es suficiente caracterizar el perfil de rendimiento académico mediante las calificaciones obtenidas en las evaluaciones, sino que hay que indagar en las características sociales, económicas y de formación previa del alumno.

En la asignatura del primer nivel de la carrera de Ingeniería en Sistemas de Información (ISI) de la UTN-FRRe, Algoritmos y Estructura de Datos, se dictan contenidos relacionados con la lógica necesaria para resolver problemas con un enfoque práctico y formal. Esta asignatura sitúa al alumno frente al desafío de representar soluciones a problemas secuenciando los pasos de resolución de manera lógica, cuestión para la cual, por lo general, no está preparado. Esto, sumado a la escasa capacidad de los alumnos para realizar razonamientos críticos, genera un ámbito de suma complejidad para el estudiante, donde el bajo rendimiento se observa en proporciones muy elevadas. Por ejemplo, el porcentaje de alumnos libres (es decir que deben recursar la materia) en el año 2014 ha sido de 74,04%.

Como se ha mencionado anteriormente, existen múltiples factores que pueden afectar el rendimiento académico de los alumnos como ser, entre otros, factores socioeconómicos, la amplitud de programas de estudio, las metodologías de enseñanza, conocimientos previos del alumno [2]. Por este motivo no es adecuado evaluar el desempeño general de los alumnos a través de notas obtenidas, porcentajes de aprobación, etc., ya que este proceso de evaluación no brinda información que permita detectar y corregir problemas cognitivos, de aprehensión, de discernimiento, etc.

Una alternativa es tratar de establecer la existencia de características comunes a ciertos grupos de estudiantes que posibiliten la definición clara de perfiles de alumnos. En ese 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 orientada a un dominio, integrada, no volátil y variante en el tiempo para ayudar a tomar decisiones [3]. Los DW surgieron por dos razones: primero, la necesidad de proporcionar una fuente única de datos limpia y consistente para propósitos de apoyo para la toma de decisiones; segundo, la necesidad de hacerlo sin afectar a los sistemas operacionales [4].

DM [5] [6] [7] es un campo de las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos [8]. Utiliza los métodos de Inteligencia Artificial y del Aprendizaje de Máquinas para extraer patrones y tendencias de forma que posibiliten la descripción y comprensión de los datos y también la predicción de comportamientos futuros.

Si bien existen diversos estudios y publicaciones que abordan la evaluación de rendimiento académico utilizando técnicas de DM [9] [10] [11], este artículo presenta un enfoque que incluye variables socio-económicas de los alumnos y de su entorno.

Específicamente, en este trabajo se propone la utilización de técnicas de DW y DM sobre la información 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. El artículo se ha estructurado de la siguiente manera: en la Sección 2 se describen los conceptos introductorios relacionados con DW y DM. En la Sección 3 se describe el diseño del modelo propuesto de DW, junto al proceso de DM utilizado para la clasificación y determinación de perfiles. En la Sección 4 se muestran algunos resultados obtenidos y finalmente, en la Sección 5, se presentan las conclusiones obtenidas.

2. Data Warehouse y Data Mining

El término Data Warehouse fue proporcionado por primera vez por Bill Inmon, el cual lo define como un repositorio de datos con las siguientes características [12]:

- Orientado a temas: Los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.
- Integrado: La base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.
- Variante en el tiempo: Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones
- No volátil: La información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de *sólo lectura*, y se mantiene para futuras consultas.

Los bloques funcionales que corresponden con un sistema de información que utiliza un DW son [13]:

1. *Nivel Operacional*: hace referencia a los sistemas operacionales/transaccionales de la organización y a fuentes que forman parte del proceso de Data Warehousing.
2. *Nivel de acceso a la información*: es la capa de interacción del usuario cuya finalidad es la conversión de los datos almacenados en información fácil y transparente para las herramientas de los usuarios finales.
3. *Nivel de acceso a los datos*: comunica el nivel de acceso a la información con el nivel operacional de forma universal.
4. *Nivel de directorio de datos (metadatos)*: repositorio de metadatos de los datos almacenados que proporcionan información sobre el origen y sobre la transformación de los mismos en el proceso de Data Warehousing.
5. *Nivel de gestión de procesos*: planificación de las tareas y procesos para la construcción y mantenimiento actualizado del DW.
6. *Nivel de mensaje de la aplicación*: determina el transporte de información a lo largo del entorno de computación de la organización a modo de middleware pero más allá de meramente protocolos de red.
7. *Nivel DW (físico)*: es el repositorio central altamente flexible de información donde residen copias de los datos operacionales y/o externos optimizados para su acceso para la consulta.
8. *Nivel de organización de datos*: incluye todos los procesos necesarios para seleccionar, editar, resumir (normalmente sumarizar), combinar y cargar en el

DW y en la capa de acceso a la información los datos operacionales y/o externos.

Detrás de la arquitectura de componentes del DW, existen un conjunto de procesos asociados, como el proceso ETL (Extracción, Transformación y Carga). Éste hace referencia al proceso de transferir, formatear, limpiar y cargar datos desde aplicaciones de producción (ERP, CRM, BDR, archivos) a los sistemas de DW.

El proceso ETL se lo puede visualizar en la figura 1, el mismo se encarga de [14]:

- Extraer los datos de las fuentes.
- Transformar y limpiar los datos.
 - Resumir los datos.
 - Detectar cambios en las fuentes.
 - Reestructurar las claves.
- Cargar los datos en el almacén.
 - Refreshar el almacén con datos actualizados.
 - Indexar los datos.
- Mantener los metadatos.

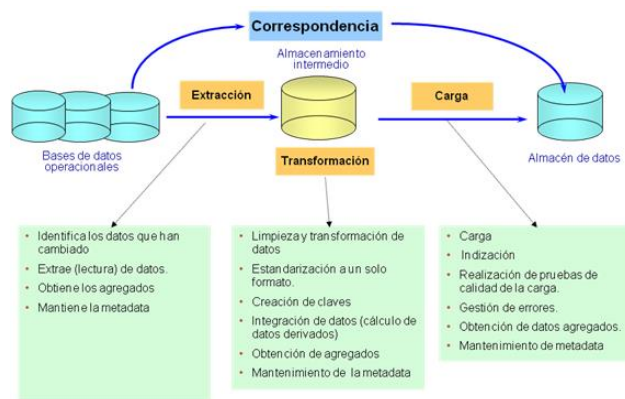


Figura 1: Proceso ETL [14].

Una vez realizado el proceso ETL y cargados los datos al DW se procede a aplicar alguna técnica para realizar el análisis de los datos almacenados. El método más utilizado es el proceso de DM.

DM es la etapa de análisis de Knowledge Discovery in Databases o KDD – descubrimiento del conocimiento. Es un campo de las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de datos [15]. Existen varias alternativas del DM, por ejemplo la Minería de Datos en Educación (Educational DM, EDM).

El objetivo de la EDM es el desarrollo de métodos para la exploración de tipos de datos únicos provenientes de plataformas educativas, y usando estos métodos para entender mejor a los estudiantes en el aprendizaje [16].

DM como tecnología general, puede ser aplicada a cualquier tipo de datos. Sin embargo, los tipos básicos de datos donde puede aplicarse DM son las bases de datos,

los DW y los datos transaccionales. Cuando se hace minería de datos sobre bases relacionales, se buscan patrones y tendencias, esto permite describir situaciones que representan actividades o características básicas de los datos. Inclusive se pueden detectar desviaciones de lo que se consideran situaciones normales. Como se mencionó anteriormente, un DW es un repositorio de información de varias fuentes que se unifica bajo un mismo esquema, de esta manera pueden utilizarse algoritmos de DM para buscar patrones en este tipo de almacenamiento. Por lo general un DW posee una estructura multidimensional de datos, lo que permite tener varias visiones de un mismo problema. Los algoritmos de DM se aplican de manera multidimensional, obteniendo patrones en cada una de esas dimensiones y relacionándolos luego. Finalmente, los algoritmos de DM pueden ser aplicados sobre datos transaccionales buscando características y patrones de comportamiento.

La característica principal es que cada transacción es analizada dentro de un conjunto mayor y busca encontrar las particularidades que son comunes a esas transacciones.

Entre esos algoritmos de DM se destaca la utilización de redes neuronales, algoritmos genéticos, predicción dinámica, agentes inteligentes, clustering, reglas de asociación, árboles de decisión, análisis de correlación, análisis semántico, análisis de regresión, entre otros. Una lista de métodos y técnicas de DM más utilizados pueden encontrarse en [17].

Las técnicas de DM permiten hacer distintos tipos de tareas:

- *Caracterizaciones y distinciones.* Aquí se intenta asociar las entradas de datos con las clases a la que pertenecen discriminando cuáles son las características básicas que las representan.
- *Minería de patrones frecuentes, asociaciones y correlaciones.* La idea aquí es encontrar repeticiones de entradas, tanto en los datos como en las estructuras. Luego asociar esas características a los conjuntos de datos y realizar correlaciones entre los mismos.
- *Clasificaciones y regresiones.* Aquí se intenta encontrar modelos que expliquen los datos. Comúnmente se utilizan métodos de aprendizaje supervisado para llevar a cabo esta tarea.
- *Clustering.* A diferencia de las clasificaciones, el clustering analiza los datos sin tener supervisión, es decir se intentan establecer las características comunes que definen las clases en base a los atributos de los datos.

- *Análisis de valores atípicos.* A diferencia del ruido en los datos, los valores atípicos corresponden a situaciones anómalas pero que son de interés. En algunas aplicaciones contar con algoritmos que detecten y expliquen estas situaciones es altamente deseable.

Finalmente, para poder llevar a cabo todas estas tareas DM se basa en varias tecnologías como ser modelos estadísticos [18], aprendizaje de máquina [19] [5], visualización de datos [20], computación de alta performance y recuperación de información [21], entre otras. Todas ellas son cada vez más utilizadas en el análisis de grandes volúmenes de datos.

3. Modelo Propuesto

Antes de definir el modelo primero se debe comprender los factores claves que intervienen en el proyecto, para luego poder traducirlos en las consideraciones de diseño. Para ello se cuenta con varias herramientas, una de ellas Bus Matrix, la cual relaciona los procesos organizacionales a las entidades u objetos que participan en el proceso [14]. Cada fila es un proceso y cada columna una dimensión. La matriz aquí se encuentra representada por un proceso Alumno y las dimensiones asociadas a él son:

- Situación Laboral del Alumno.
- Situación Laboral del Padre.
- Situación Laboral de la Madre.
- Horas Dedicadas al Estudio.
- Importancia a las TIC.
- Importancia que da al Estudio.
- Residencia Actual.
- Lugar de Procedencia.
- Dependencia de la Escuela Secundaria.

A partir de la matriz descripta se realiza el proceso de diseño, en el cual se identifican las tablas involucradas en el modelo propuesto, las tablas de Hechos y las tablas de Dimensiones.

Las tablas de Hechos contienen medidas que surgen de los procesos de negocios. Una medida es un atributo (campo) de una tabla que se desea analizar, sumando o agrupando sus datos, usando los criterios de corte conocidos como dimensiones. Cada tabla de Hechos tiene como atributos una o más medidas de un proceso organizacional, de acuerdo a los requerimientos [22].

Las tablas de dimensiones surgen naturalmente de la matriz de procesos/dimensiones. Las tablas de dimensiones tienen un conjunto de atributos (generalmente textuales) que brindan una perspectiva o forma de análisis sobre una medida en una tabla de Hechos [22].

El modelo propuesto se muestra en la Figura 2, en el cual se visualiza una tabla de hechos y varias tablas de dimensiones asociadas.



Figura 2: Modelo Propuesto del DW.

Para la implementación del modelo propuesto se utilizó las herramientas incluidas en el IBM Data Warehouse Edition (DWE) V.9.7, un paquete de productos que combina la potencia de DB2 Universal Database (DB2 UDB) con la robusta infraestructura de inteligencia empresarial de IBM [23]. Se compone de varios productos pero a los fines de este trabajo se han usado los siguientes componentes:

- El Sistema de Gestión de Base de datos DB2 Enterprise Server Edition (DB2 ESE), que permite conexiones multiusuarios y presenta una alta escalabilidad.
- Design Studio (DS): Una plataforma a modo de herramienta que usan los analistas de negocios y los administradores de almacenes de datos para diseñar las reglas de las cargas de trabajo, los flujos de transformación de datos y los flujos analíticos para la minería de datos y el análisis de textos.
- Intelligent Miner (IM): Conjunto de funcionalidades que permiten realizar análisis de información de acuerdo a conceptos de Business Intelligence (BI).

4. Resultados

Para lograr el objetivo de descubrir los perfiles de rendimiento académico se requería tanto información interna como externa. La cátedra de Algoritmos y Estructuras de Datos facilitó las calificaciones de los exámenes de los alumnos y su condición al finalizar el cursado (libre, regular, promocionado).

Dicha fuente fue completada con una encuesta, la cual contaba con preguntas relacionadas con aspectos demográficos, situación familiar, historial de estudios

realizados por sus familiares, aspectos actitudinales del alumno respecto al estudio y las TIC (Tecnologías de Información y Conocimiento), entre otras cuestiones.

Esta información se usó como datos de entrada del DW para luego aplicar procesos de DM. En principio se utilizó la técnica de Clasificación con Árboles de Decisión, cuyo objetivo es realizar clasificaciones sobre los datos conocidos y así crear modelos que luego puedan ser utilizados para predecir o clasificar valores nuevos o desconocidos.

El análisis de los resultados se basó en considerar como parámetro de minería la variable relacionada a la situación final del alumno, la cual reflejó su estado en la materia al cierre del Ciclo Lectivo.

Según lo planteado anteriormente se obtuvieron en el ciclo lectivo 2014 los siguientes resultados: 74,04% de alumnos en condición de Libre, 18,27% Regular y 7,69% Promocionados.

En la Figura 3 se describe cómo se caracterizan todas las clases (Regular, Libre y Promocionado) teniendo en cuenta el criterio de Dependencia de la Escuela Secundaria del alumno y comparando los resultados obtenidos en el año 2013 y los obtenidos en el 2014.

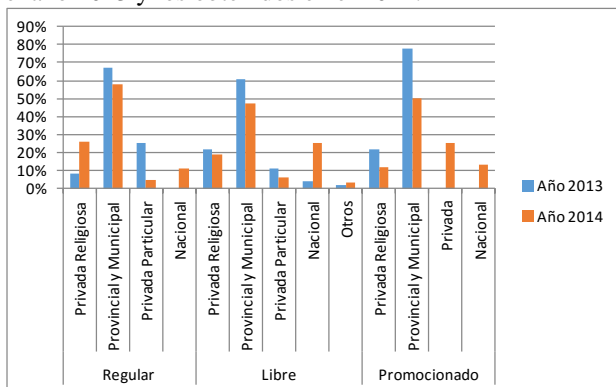


Figura 3: Dependencia de la Escuela Secundaria.

En el tipo de escuela secundaria del alumno se puede observar un marcado descenso en la cantidad de alumnos Regulares y Promocionados que han egresado de un colegio secundario provincial y en contraposición una participación de los colegios secundarios que son administrados por el estado nacional; eso impacta directamente en la comunicación que se tenga con las diferentes entidades.

En la Figura 4 se toma como criterio la Cantidad de Horas Dedicadas al Estudio.

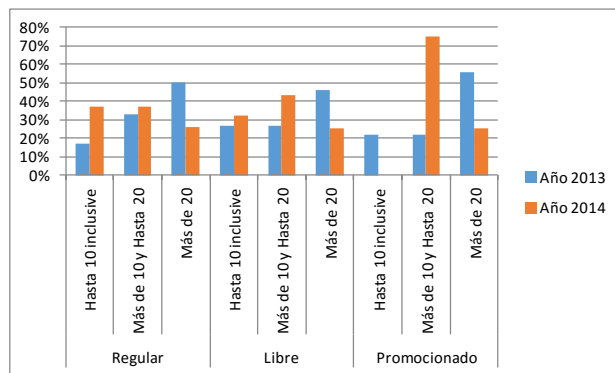


Figura 4: Cantidad de Horas Dedicadas al Estudio.

En el criterio tiempo dedicado al estudio se puede vislumbrar varias aristas en esta variable, por ejemplo la totalidad de alumnos que invierten más de 20 horas al estudio ha descendido, tanto en libres, como regulares y promocionados, tendiendo todos a una dedicación de entre 10 y 20 horas más allá de rendimiento académico. Lo que podría alertar sobre la calidad de estudio que aplican los estudiantes a la materia.

En la Figura 5 se detalla la Importancia Dada al Estudio.

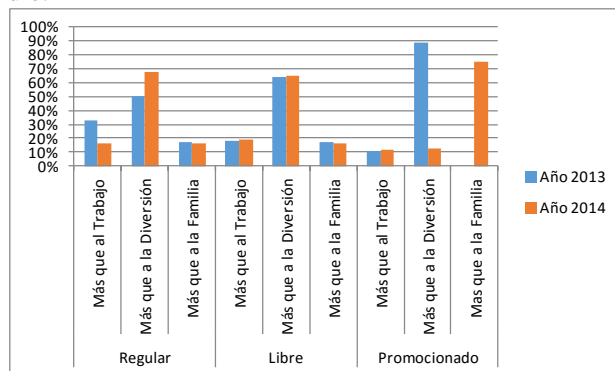


Figura 5: Importancia Dada al Estudio.

En la importancia dedicada al estudio, el escenario muestra una variación significativa en cuanto al estrato de alumnos promocionados, cuya respuesta se incrementa notablemente con respecto al año anterior en dedicarle más importancia al estudio que a la familia. Desde esta perspectiva el alumno promocionado es consciente del tiempo requerido por el estudio y el tiempo dedicado a la familia.

En la Figura 6, 7 y 8 el criterio utilizado es Últimos Estudios Alcanzados por la Madre.

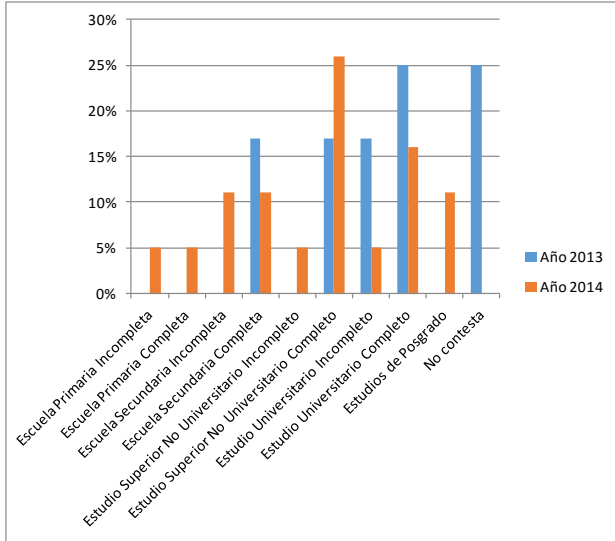


Figura 6: Últimos Estudios Alcanzados por la Madre por los Alumnos Libres.

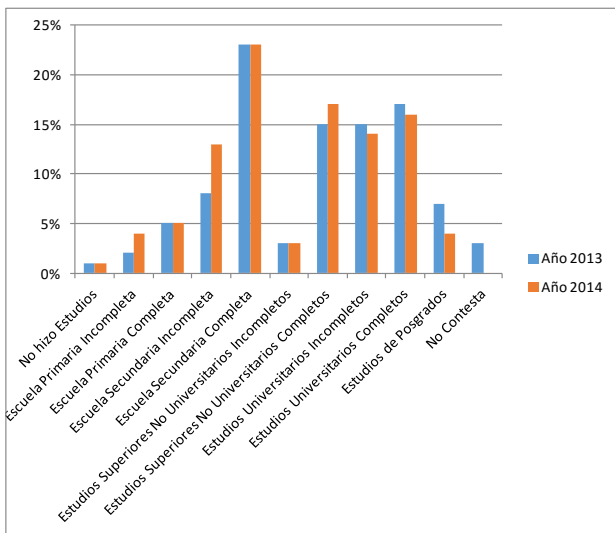


Figura 7: Últimos Estudios Alcanzados por la Madre por los Alumnos Regulares.

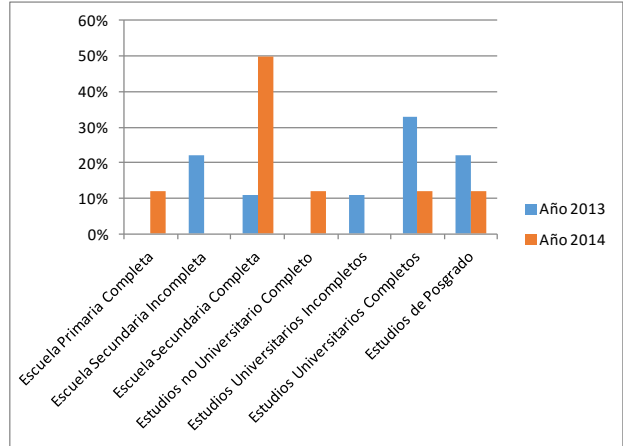


Figura 8: Últimos Estudios Alcanzados por la Madre por los Alumnos Promovidos.

En los últimos estudios alcanzados por la madre la alteración más notable está en los estudiantes promocionados cuyas madres tienen alcanzado hasta un nivel secundario completo; pasando de un 11% a un 50% que puede inferirse como un cambio en la visión de las madres respecto de la importancia de la educación universitaria, aún sin haberla alcanzado.

En la Figura 9, 10 y 11 el criterio utilizado es Últimos Estudios Alcanzados por el Padre.

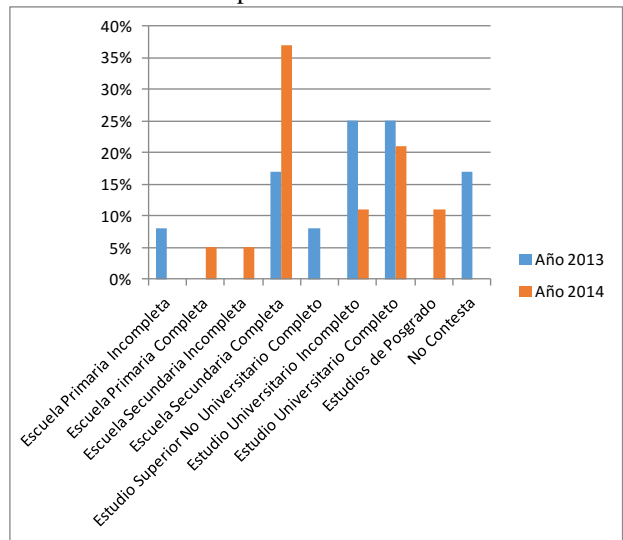


Figura 9: Últimos Estudios Alcanzados por la Padre por los Alumnos Libres.

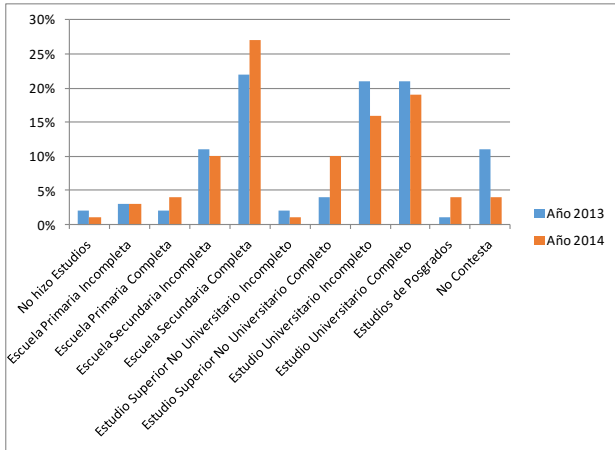


Figura 10: Últimos Estudios Alcanzados por la Padre por los Alumnos Regulares.

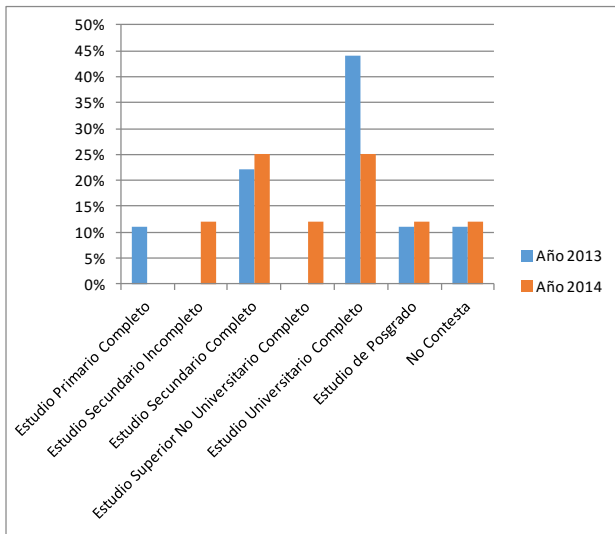


Figura 11: Últimos Estudios Alcanzados por la Padre por los Alumnos Promovidos.

En los últimos estudios alcanzados por el padre nuevamente se puede ver una tendencia repetida con el punto anterior, con el incremento de los padres con estudio secundario como máximo nivel. En las demás variables no se presentan grandes cambios respecto del año anterior.

En la Figura 12 el criterio utilizado es la Utilización de las TIC.

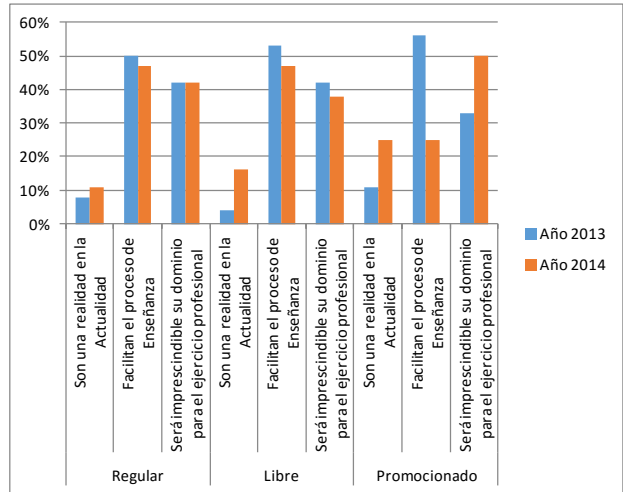


Figura 12: Utilización de las TIC.

Se visualiza un notorio cambio en cuanto a los alumnos promocionados por el incremento de 33% a 50% a la elección de: “será imprescindible su dominio para el ejercicio profesional”, lo que habla de una visión de futuro del alumno respecto de los estudios universitarios y situándose en el ejercicio profesional, algo deseable desde el punto de vista académico ya que pone a la educación como un medio y no como un fin en sí mismo.

5. Conclusiones

Este trabajo describe un modelo eficiente que posibilita la determinación de perfiles según el rendimiento académico de los alumnos utilizando las herramientas de DW y las técnicas de Minería de Datos.

Con los resultados obtenidos en el año 2014 y comparándolos con los obtenidos en el 2013 se puede observar que el rendimiento académico se encuentra influenciado por el entorno socio económico, cultural y actitud del alumno frente al estudio y al uso de las TIC. Por tal motivo se presta una atención especial en las relaciones de estas variables con los diferentes rendimientos de los alumnos.

Los resultados obtenidos confirman que los alumnos finalizan sus estudios secundarios en escuelas Provinciales y Municipales en su gran mayoría. Además, se presenta un alto rendimiento académico en aquellos alumnos que le dedican más de 10 horas de estudio semanales. Otro aspecto destacable es que aquellos alumnos que brindan más importancia al estudio sobre la familia y la diversión pudieron promocionar la materia Algoritmo y Estructura de datos. También, se puede verificar que los alumnos cuya madre ha finalizado los estudios secundarios o universitarios y su padre tiene estudio universitario completo presentan un mejor rendimiento académico. Además, aquellos alumnos que ven a las TIC como im-

prescindible para el ejercicio profesional tienen un alto rendimiento académico.

Claramente se puede apreciar que con la utilización del modelo propuesto se pueden tomar medidas correctivas sobre los casos de los alumnos que presenten perfiles de bajo rendimiento académico, procurando disminuir así el fracaso académico. De esta manera, la herramienta presentada aquí y las acciones subsiguientes concordantes, permitiría disminuir la deserción temprana de los alumnos en la cátedra Algoritmos y Estructura de Datos.

6. Agradecimientos

Este trabajo ha sido desarrollado en el marco del proyecto “Determinación de Perfiles de Estudiantes y de Rendimiento Académico Mediante la Utilización de Minería de Datos”. Universidad Tecnológica Nacional, Facultad Regional Resistencia. Código 25/L059 (programa de Incentivos a Docentes Investigadores del Ministerio de Educación). Código EIINRE571. (01/01/2013 – 31/12/2015). Se destaca el apoyo prestado por el Dr. Marcelo Karanik en la revisión de este trabajo.

Referencias

- [1] J. I. Maletic, M. L. Collard, and A. Marcus, “Source Code Files as Structured Documents”, in *Proceedings 10th IEEE International Workshop on Program Comprehension (IWPC'02)*, Paris, France, June 27-29 2002, pp. 289-292.
- [2] A. Marcus, *Semantic Driven Program Analysis*, Kent State University, Kent, OH, USA, Doctoral Thesis, 2003.
- [3] G. Salton, *Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer*, Addison-Wesley, 1989.
- [4] J. Molina López, J. García Herrero, *Técnicas de Análisis de Datos*. Universidad Carlos III. Madrid, España, 2006.
- [5] Ian H. Witten, and Frank Eibe. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.
- [6] David J. Hand, Mannila Heikki and Smyth Padhraic. *Principles of data mining*. MIT press, 2001.
- [7] Tan Pang-Ning, Michael Steinbach, and Vipin Kumar. “Introduction to data mining”. Library of Congress. 2006.
- [8] Oded Maimon and Lior Rokach. *Data Mining and Knowledge Discovery Handbook*. Springer, New York. 2010. ISBN 978-0-387-09823-4.
- [9] S. Formia, L. Lanzarini, W. Hasperué: *Characterization of University Drop-Out at UNRN Using Data Mining. A Study Case*, CACIC, Mar del Plata, Buenos Aires, Argentina, 2013.
- [10] D. L. La Red Martínez; C. E. Podestá. Data Mining to Find Profiles of Students; Volume 10 – N° 30; *Euro-pean Scientific Journal (ESJ)*; pp. 23-43; ISSN N° 1857-7881; University of the Azores, Portugal; 2014.
- [11] R. Timarán Pereira: “Una lectura sobre deserción universitaria en estudiantes de pregrado desde la perspectiva de la minería de datos”. *Revista Científica Guillermo de Ockham*. Vol. 8, N° 1, pp. 121-130, 2010.
- [12] J. Curto Días. *Introducción al Business Intelligence*. Editorial UOC. Barcelona, España, 2010. ISBN: 978-84-9788-886-8.
- [13] *Arquitectura de un data warehouse* [en línea]. Disponible en: <http://informationmanagement.wordpress.com/2007/06/03/arquitectura-de-un-datawarehouse/>
- [14] Ralph Kimball; Joe Caserta. *The Data Warehouse ETL Toolkit*. Wiley, 2004. ISBN 0-7645-6757-8.
- [15] Oded Maimon and Lior Rokach. *Data Mining and Knowledge Discovery Handbook*. Springer, New York, 2010. ISBN 978-0-387-09823-4.
- [16] R. Baker, K. Yaceff: “The State of Educational Data Mining in 2009: A Review and Future Visions”. *Journal of Educational Data Mining*, Artículo 1, Vol 1, No 1, 2009.
- [17] Shu-Hsien L., Pei-Hui C., Pei-Yuan H. “Data mining techniques and applications – A decade review from 2000 to 2011”. *Expert Systems with Applications* 39. 11303–11311. 2012.
- [18] R. O. Duda, P. E. Hart, D. G. Stork. *Pattern Classification*. 2nd Edition. John Wiley & Sons, 2001.
- [19] T. M. Mitchell. *Machine Learning*. MacGraw Hill. 1997.
- [20] W. Cleveland. *Visualizing Data*. Hobart Press. 1993.
- [21] C. D. Manning, P. Raghavan, H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press. 2008.
- [22] Gustavo R. Rivadera. *La metodología de Kimball para el diseño de almacenes de datos (Data Warehouses)*. [en línea]. Disponible en: <http://www.ucasal.edu.ar/htm/ingenieria/cuadernos/archivos/5-p56-rivadera-formateado.pdf>
- [23] *IBM DB2 Universal Database Data Warehouse Edition* [en línea]. Disponible en: ftp://public.dhe.ibm.com/ps/products/dwe/V8.2.1/es_ES/getstart821.pdf