

Una lectura sobre deserción universitaria en estudiantes de pregrado desde la perspectiva de la minería de datos*

An insight on university drop-out of undergraduate students from the perspective of data mining

Ricardo Timarón Pereira

Resumen

En este artículo se describe el proceso de descubrimiento de conocimiento que se llevó a cabo en la Universidad de Nariño para determinar en la comunidad universitaria perfiles de bajo rendimiento académico y deserción estudiantil, para lo cual se utilizó la base de datos histórica de los estudiantes de pregrado. Este proceso se apoyó con TaryKDD, una herramienta de minería de datos de distribución libre, desarrollada en los laboratorios KDD del grupo de investigación Grias, del Departamento de Sistemas de la Facultad de Ingeniería de la Universidad de Nariño.

Palabras clave: deserción estudiantil, descubrimiento de patrones, minería de datos.

Abstract

This article describes the process that was developed at Universidad de Nariño in order to identify low academic performance and dropout rates in the university, for which the historical database of its undergraduate students was used. This process was supported on TaryKDD, a free distribution data mining tool developed in the KDD laboratories of the research group Grias from the Department of Systems Engineering of the Faculty of Universidad de Nariño.

Keywords: student dropout, pattern discovery, data mining.

Introducción

La deserción estudiantil en los programas de pregrado de la gran mayoría de institu-

• Fecha de recepción del artículo: 25 de septiembre de 2009 • Fecha de aceptación: 17 de diciembre de 2009.

RICARDO TIMARÓN PEREIRA. Doctorado en Ingeniería, Universidad del Valle, Colombia. Master of Science en Ingeniería, Universidad Politécnica de Donetsk. Pregrado en Ingeniería de Sistemas y Computación, Universidad Politécnica de Donetsk. Profesor Asociado del Departamento de Sistemas, Universidad de Nariño-Colombia. Correo e: ritimar@udenar.edu.co

* Este artículo es producto del proyecto de investigación *Detección de patrones de bajo rendimiento y/o deserción de los estudiantes de la Universidad de Nariño con técnicas de minería de datos.*

ciones de educación superior (IES) tanto de Colombia como de Latinoamérica es un problema que tiene un impacto multidimensional en el desarrollo social y económico de un país.

Se entiende por deserción estudiantil al hecho de que un número de estudiantes matriculados no siga la trayectoria normal del programa académico, bien sea por retirarse de ella, por repetir cursos o por retiros temporales (UPN, 2009). La disolución del vínculo que se estipula a través de la matrícula académica, ya sea por parte del estudiante o de la universidad, tiene efectos de tipo financiero, académico y social que implican la pérdida de esfuerzos y recursos en un país como Colombia en donde más de la mitad de los estudiantes que comienzan una carrera universitaria no la concluyen (Rojas *et al.*, 2008).

La Universidad de Nariño es una institución pública de educación superior, su área de influencia es el suroccidente de Colombia y su sede principal se encuentra en la ciudad de Pasto, capital del departamento de Nariño. En ella se encuentra la mayoría de estudiantes universitarios de la región. Los estudiantes de educación secundaria aspiran obtener un cupo en ella, por su calidad educativa y el prestigio de sus egresados. En algunos casos, cuando un estudiante se matricula en un determinado programa, su rendimiento no es el esperado, lo que genera índices de deserción altos y bajo rendimiento académico. Por lo tanto, se abre un interrogante acerca de cuáles son las causas que motivan la deserción y/o el bajo rendimiento y qué perfiles tienen este tipo de estudiantes.

De acuerdo con los datos de la Oficina de Planeación de la Universidad de Nariño, en el año 2006 se encontraban matriculados en los diferentes programas de pregrado 8136 estudiantes distribuidos en 11 facultades (ver Tabla 1), de los cuales el 27.94% han perdido dos veces la misma asignatura, el 3.44% han perdido tres veces y el 0.19%, cuatro veces. Así mismo el 6.71% tienen un promedio menor a 3.0, el 66.85% tienen un promedio entre 3.0 y 4.0, y el 26.22%

Tabla 1
Número de estudiantes por facultad año 2006

| Facultades | No. Estudiantes |
|---------------------------------------|-----------------|
| Artes | 902 |
| Ciencias Agrícolas | 618 |
| Ciencias de la Salud | 243 |
| Ciencias Económicas y Administrativas | 1408 |
| Ciencias Humanas | 1341 |
| Ciencias Naturales y Matemáticas | 687 |
| Ciencias Pecuarias | 523 |
| Derecho | 510 |
| Educación | 435 |
| Ingeniería | 1188 |
| Ingeniería Agroindustrial | 281 |
| Total | 8136 |

tienen un promedio mayor que 4.0 sobre 5.0. Igualmente, en este año el promedio del índice de deserción fue del 30%, teniendo en cuenta el último corte de los programas (ver Tabla 2).

A pesar de esto, son muy pocos los estudios que se han realizado en la Universidad de Nariño con respecto a este problema que permitan aplicar estrategias efectivas que ayuden a minimizar este fenómeno y conlleven el mejoramiento de la calidad educativa en la universidad.

A través de técnicas de minería de datos aplicadas a los datos históricos almacenados en las bases de datos de una IES, es posible predecir las características del estudiante que va a abandonar sus estudios o predecir quienes están propensos a desertar. El determinar perfiles de deserción permite predecir qué estudiantes son los que desertan. Teniendo estos perfiles la IES puede tomar acciones anticipadas que le permitan disminuir el índice de deserción.

La minería de datos es un proceso automático en el que se combinan descubrimiento y análisis. El proceso consiste en extraer patrones en forma de reglas o funciones, a partir de los datos, para que el usuario los analice. Esta tarea implica generalmente preprocesar los datos, hacer minería de datos

Tabla 2
Índice de deserción por programas año 2006

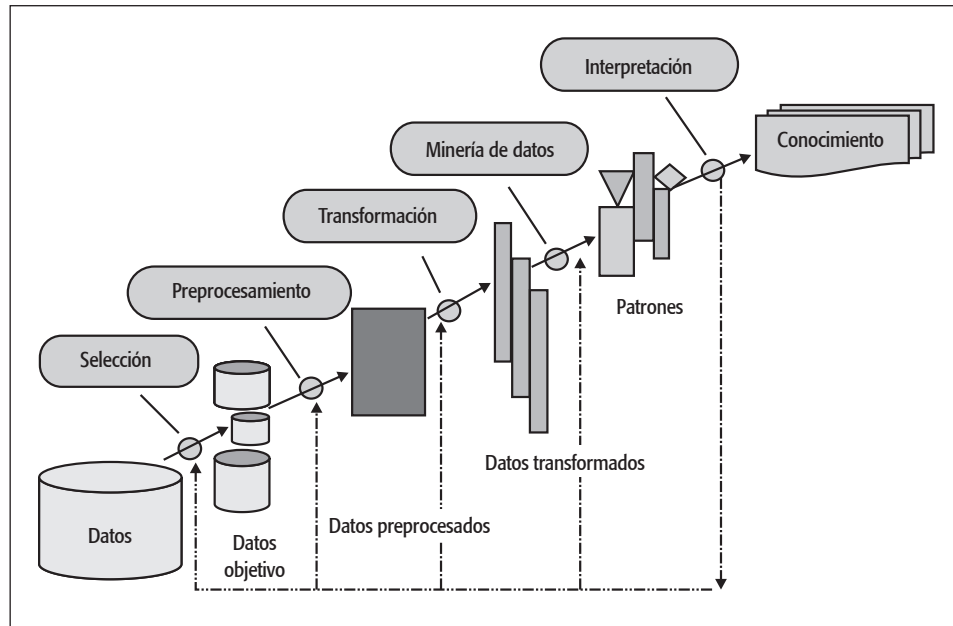
| Facultades | Programas | Índice Deserción |
|----------------------------------|-------------------------------------|------------------|
| Ciencias Agrícolas | Ingeniería Agroforestal | 33,3 |
| | Ingeniería Agronómica | 12,5 |
| Ciencias Pecuarias | Medicina Veterinaria | 20,0 |
| | Zootecnia | 25,5 |
| | Ingeniería en Producción Acuícola | 36,4 |
| Educación | Lic. Educación Básica | 37,7 |
| | Lic. Lengua Castellana y Literatura | 22,0 |
| Ingeniería Agroindustrial | Ingeniería Agroindustrial | 14,3 |
| Derecho | Derecho | 14,7 |
| Artes | Arquitectura | 38,0 |
| | Artes Visuales | 43,5 |
| | Lic. Artes Visuales | 23,1 |
| | Diseño Industrial | 2,6 |
| | Diseño Gráfico | 51,3 |
| | Lic. en Música | 9,4 |
| Ciencias Económicas | Administración de Empresas | 34,4 |
| | Comercio Internacional | 10,0 |
| | Economía | 11,3 |
| Ciencias Humanas | Geografía Aplicada. | 15,9 |
| | Lic. Ciencias Sociales | 47,1 |
| | Lic. Filosofía y Letras | 42,2 |
| | Sociología | 10,4 |
| | Lic. Inglés Francés | 55,6 |
| | Lic. Educación Básica | 45,5 |
| | Psicología | 2,2 |
| Ciencias Naturales y Matemáticas | Biología | 79,1 |
| | Física | 73,9 |
| | Lic. Informática | 38,5 |
| | Lic. Matemáticas | 63,8 |
| | Química | 5,1 |
| Ingeniería | Ingeniería Civil | 21,4 |
| | Ingeniería de Sistemas | 39,8 |
| Ciencias de la Salud | Tecnología en Promociónon Salud | 9,5 |

y evaluar e interpretar los resultados (Chen *et al*, 1997; Imielinski *et al*, 1996; Han *et al*, 2001). Este proceso es iterativo e interactivo. Es iterativo ya que la salida de alguna de las fases puede hacer volver a pasos anteriores y porque a menudo son necesarias varias iteraciones para extraer conocimiento de alta calidad. Es interactivo porque el usuario (experto en el dominio del problema) interviene en la toma de muchas decisiones (Hernández

et al, 2004). En la Figura 1 se muestran las etapas del proceso de minería de datos.

En este artículo se describe el proceso de minería de datos que se llevó a cabo en la Universidad de Nariño para determinar en la comunidad universitaria perfiles de bajo rendimiento académico y deserción estudiantil utilizando la base de datos histórica de los estudiantes de pregrado, compuesta por información personal y académica de 46.173 estudiantes entre activos, egresados

Figura 1
Etapas del proceso de minera de datos



y retirados, acumulada en un periodo de dieciocho aos.

Esta investigacion se apoyo con TariyKDD, una herramienta de minera de datos de distribucion libre, desarrollada en los laboratorios KDD del grupo de investigacion GRIAS del Departamento de Sistemas de la Facultad de Ingeniera de la Universidad de Nario (Timaran *et al.*, 2007). Esta herramienta implementa todas las etapas del proceso de descubrimiento de conocimiento. Su funcionalidad ha sido probada con varios repositorios de datos tanto snteticos como reales cuyos resultados demostraron su fiabilidad (Timaran *et al.*, 2007).

La arquitectura de TariyKDD esta compuesta por cuatro modulos: *el modulo de conexion*, que permite la recuperacion de datos desde archivos planos y bases de datos relacionales; *el modulo de utilidades* con clases y libreras comunes; *el modulo kernel*, donde se encuentran los filtros que permiten realizar los procesos de limpieza y transformacion de datos, los algoritmos de minera de datos para las tareas de asociacion y clasificacion y los programas de visualizacion de datos; y *el modulo de interfaz grafica de usuario*,

que facilita la interaccion del usuario con la herramienta de una manera amigable.

En TariyKDD se encuentran implementados los algoritmos de minera de datos: Apriori (Agrawal *et al.*, 1994), *FPGrowth* (Han *et al.*, 2000) y *EquipAsso* (Timaran *et al.*, 2005a; Timaran *et al.*, 2005b) para la tarea de asociacion y los algoritmos *C4.5* (Quinlan, 1993) y *Mate-tree* (Timaran, 2007) para la tarea de clasificacion.

Metodologa

Con el objeto de extraer conocimiento a partir de los datos almacenados en las bases de datos de la Universidad de Nario y determinar patrones de desercion estudiantil se cumplio con cada una de las etapas del proceso de minera de datos:

Etapas de seleccion

El objetivo de esta etapa es obtener las fuentes de datos internas y externas que sirven de base para el proceso de minera de datos. Como fuente interna se selecciono la base de datos historica de los estudiantes de

la Universidad de Nariño, compuesta por información personal y académica de 46.173 estudiantes entre activos, egresados y retirados, acumulada en un periodo de dieciocho años. Como fuente externa se seleccionó el archivo con la información de los colegios de educación secundaria del país, que se obtuvo con el Ministerio de Educación Nacional de Colombia, con el fin de obtener algunas características de los colegios entre las que se cuentan: si el colegio es público o privado, el tipo de calendario A o B, la jornada de estudio, si el colegio es urbano o rural. Estas características son importantes para incluirlas dentro de los atributos relevantes que permitan determinar el perfil de los estudiantes que abandonan sus estudios o tienen bajo rendimiento académico. Estas fuentes de datos se integraron en la base de datos UDENARDB, construida con el sistema gestor de base de datos PostgreSQL. UDENARDB la componen siete tablas, cuyas descripciones se pueden ver en la Tabla 3.

Etapa de preprocesamiento de datos

El objetivo de esta etapa es obtener datos limpios, i.e. datos sin valores nulos o anómalos que permitan obtener patrones de calidad. Por medio de consultas *ad-hoc* sobre la base de datos UDENARDB se analizó minuciosamente la calidad de los datos contenidos en cada uno de los atributos de las tablas y se seleccionaron los atributos más relevantes para la investigación. De la tabla alumnos se seleccionaron 19 atributos; de la

tabla carreras, 4; de la tabla facultades, 2; de la tabla materias, 3; de la tabla notas, 8; de la tabla liquidación, 12 y de la tabla colegios 3 atributos. Los atributos seleccionados de las diferentes tablas en su gran mayoría no contenían valores nulos ni anómalos (*outliers*), pero en aquellos casos que se presentaban, estos fueron reemplazados utilizando técnicas estadísticas tales como la media y la moda o derivando sus valores a través de otros de acuerdo con el dominio de donde provienen los datos. Por citar algunos casos: se utilizó la moda para reemplazar los valores nulos del atributo estado_civil, teniendo en cuenta la moda de los valores no nulos del estado civil de aquellos estudiantes que tenían la misma edad y género que los analizados; se reemplazó con la media los valores nulos del atributo ingresos_familiares, teniendo en cuenta la media de los valores no nulos de los ingresos familiares de los estudiantes que tenían el mismo valor de matrícula y estrato; se reemplazó los valores nulos del atributo edad_ingreso del estudiante, derivándola de la fecha de ingreso y la fecha de nacimiento. Como resultado de esta etapa quedaron únicamente datos de 20.329 estudiantes para su posterior análisis.

Etapa de transformación de datos

En la etapa de transformación se buscan características útiles para representar los datos según la meta del proceso de minería de datos. Se utilizan métodos de reducción de dimensiones o de transformación para

Tabla 3
Descripción de tablas de la base de datos UDENARDB

| Tablas | No. atributos | Descripción. |
|-------------|---------------|---|
| Alumnos | 69 | Se encuentran todos los datos personales del estudiante. |
| Carreras | 10 | Se encuentra información de todas las carreras existentes en la Universidad de Nariño. |
| Facultades | 4 | Contiene información de las facultades de la Universidad de Nariño. |
| Materias | 4 | Se encuentra toda la información de las materias existentes en el plan académico de cada carrera. |
| Notas | 8 | Contiene información de las notas por materia de cada estudiante. |
| Liquidación | 27 | Se encuentra toda la información financiera del estudiante. |
| Colegios | 7 | Contiene información de los colegios del país. |

disminuir el número efectivo de variables en consideración o para encontrar representaciones invariantes de los datos (Fayyad *et al.*, 1996).

En esta etapa se construyó el conjunto de datos UDENAR.DAT integrando los atributos de las diferentes tablas de la base de datos UDENARDB. Se eliminaron los atributos que eran llaves primarias de las tablas, se construyeron nuevos atributos (ver Tabla 4) y se discretizaron los atributos con valores continuos, es decir, se transformaron los valores numéricos en valores discretos o nominales, con el fin de disminuir el número

de valores distintos de estos atributos. La discretización de estos atributos se realizó teniendo en cuenta el hecho de que las diferencias en ciertas zonas del rango de valores eran más importantes que otras, lo que resultaba en intervalos más significativos que otros dentro del conjunto de valores de cada atributo. Algunos de los atributos discretizados se muestran en las Tablas 5, 6 y 7.

Por conveniencia en el manejo de memoria y rendimiento de los algoritmos de minería de datos, el atributo Clase_al, que determina el tipo de estudiante, se numerizó;

Tabla 4
Descripción de nuevos atributos del conjunto de datos UDENAR.DAT

| Atributo | Descripción |
|---------------|---|
| Ingresos | Establece un valor real actualizado de ingresos familiares del estudiante. Para ello relaciona los campos ingresos_familiares de la tabla alumnos e ingresos de la tabla liquidación. |
| Edad | Determina qué edad tiene actualmente el estudiante; para ello se relacionaron los campos fecha nacimiento de la tabla alumnos y la fecha actual. |
| edad_ing | Establece la edad en la que ingresó el estudiante. Para crearlo se relacionó el campo fecha de ingreso y la fecha de nacimiento del estudiante. |
| val_matrícula | Determina el valor real que paga el estudiante por concepto de matrícula financiera; relaciona los valores de los campos nueva_matrícula y de nuevos servicios de la tabla liquidación. |
| Clase_al | Determina qué estudiantes han reingresado, se han retirado o no cumplen con ninguna de las condiciones anteriores. |
| Claserend | Determina la cantidad de materias perdidas por el estudiante. |
| Clasepromedio | Determina el promedio acumulado del estudiante. |

Tabla 5
Discretización del atributo edad

| Edad | Valor | No. Registros |
|---|-------|---------------|
| Menores e iguales a 18 | A | 827 |
| Mayores de 18 y menores de 22 | B | 3.634 |
| Mayores e iguales de 22 y menores de 26 | C | 4.856 |
| Mayores e iguales de 26 | D | 11.012 |

Tabla 6
Discretización del atributo Fecha de Ingreso

| Fecha Ingreso | Valor | No. Registros |
|---|-------|---------------|
| Antes de 1990 | A | 1.022 |
| Después o igual a 1990 a menores de 1995 | B | 4.852 |
| Después o igual a 1995 a menores de 2000 | C | 5.978 |
| Después o igual al 2000 y menores de 2003 | D | 5.046 |
| Mayores o iguales de 2003 | E | 3.431 |

Tabla 7
Discretización del atributo Clasepromedio

| Clasepromedio | Valor | No.Registros |
|-------------------------------|-------|--------------|
| Menor a 2 | A | 2.391 |
| Mayor o igual a 2 hasta 3 | B | 2.934 |
| Mayor o igual a 3 hasta 3.5 | C | 5.166 |
| Mayor o igual a 3,5 hasta 4,0 | D | 6.850 |
| Mayor o igual a 4.0 hasta 5.0 | E | 2.988 |

i. e. se convirtió de un valor nominal a entero, como se muestra en la Tabla 8.

Por otra parte, el conjunto de datos UDE-NAR.DAT se adecuó al formato ARFF (*Attribute Relation File Format*) utilizado por la herramienta TaryKDD para importar los datos. La estructura del formato ARFF (Witten *et al*, 2000) es la siguiente:

- Cabecera: se define el nombre de la relación y su formato es el siguiente: @relation <nombre-de-la-relación>
- Declaraciones de los atributos. En esta sección se declaran los atributos que compondrán el archivo arff con su tipo. La sintaxis es la siguiente:
@attribute <nombre-del-atributo>
<tipo>
- Sección de datos. Se declaran los datos que componen la relación, separando entre comas los atributos y con salto de líneas las relaciones.

Finalmente, se obtuvo el conjunto de datos UDENAR.ARFF con 26 atributos y 20.329 registros, listo para aplicarle las técnicas de minería de datos utilizando la herramienta TaryKDD que permitan obtener los patrones de bajo rendimiento académico y /o deserción de los estudiantes de la Universidad de Nariño.

Tabla 8
Discretización del atributo Clase_al

| Clase_al | Valor | No. Registros |
|-----------|-------|---------------|
| Normal | A | 18.641 |
| Reingreso | B | 3.634 |
| Retirado | C | 4.856 |

Etapa de minería de datos

El objetivo de esta etapa es la búsqueda y descubrimiento de patrones insospechados y de interés utilizando diferentes técnicas de descubrimiento tales como clasificación, clustering, patrones secuenciales, asociación, entre otras. Para el descubrimiento de patrones de bajo rendimiento académico y deserción estudiantil se utilizaron las tareas de clasificación y asociación. Para generar las reglas de clasificación se utilizó el algoritmo C4.5 y para las reglas de asociación, el algoritmo EquipAsso, disponibles en la herramienta TaryKDD. Los patrones descubiertos se describen en la sección de resultados.

Etapa de Interpretación y evaluación de resultados

En esta etapa se interpretan los patrones descubiertos y posiblemente se retorna a los anteriores pasos o etapas para posteriores iteraciones. Esta etapa puede incluir la visualización de los patrones extraídos, la remoción de los patrones redundantes o irrelevantes y la traducción de los patrones útiles en términos que sean entendibles para el usuario. Los resultados de esta etapa se analizan en la siguiente sección.

Resultados

Para predecir los perfiles de bajo rendimiento académico, el conjunto de datos UDENAR.ARFF se clasificó escogiendo como clase el atributo Clasepromedio cuyas etiquetas se muestran en la Tabla 7. Este atributo indica el rendimiento académico del

estudiante basado en el promedio acumulado de las notas hasta el semestre cursado.

Entre las reglas de clasificación más representativas están:

- Si: el estrato socioeconómico es 2, el ponderado de exámenes de estado ICFES está entre 50 y 70, es del sur del departamento de Nariño, está en primer semestre y pertenece a la Facultad de Artes, entonces su rendimiento es de tipo A (ver tabla 6). El 68% de los estudiantes con estas características se clasifican de esta manera.
- Si: la edad de ingreso es menor o igual a 18 años, proviene de un colegio privado, el calendario del colegio es septiembre a junio, es de género femenino, es del sur del departamento de Nariño, está en primer semestre y pertenece a la facultad de Ciencias Naturales y Matemáticas, entonces su rendimiento es de tipo A. El 70% de los estudiantes con estas características se clasifican de esta manera.
- Si: el estrato socioeconómico es 2, es de género masculino, el tipo de residencia es propia, cursa primer semestre y pertenece a la Facultad de Ciencias Económicas y Administrativas, entonces su rendimiento es de tipo B. Un 67% de los estudiantes con estas características se clasificarán en este grupo.

Para predecir los perfiles de deserción estudiantil se escogió como clase el atributo Clase_al (sus valores se muestran en la Tabla 8). Este atributo indica si el estudiante no se ha retirado, si ha reingresado o si se retiró definitivamente de la Universidad.

Entre las reglas de clasificación más representativas están:

- Más del 50% de los estudiantes retirados que pertenecen a la Facultad de Ingenierías reingresan.
- Más del 50% de los estudiantes retirados que pertenecen a las facultades de Ciencias Naturales y Matemáticas y Ciencias Humanas no reingresan.

Entre las reglas de asociación más representativas que permiten identificar relaciones no explícitas entre los atributos del conjunto de datos UDENAR.ARFF que involucran bajo rendimiento y deserción están:

- El 95% de los estudiantes que tienen promedio bajo están en primer semestre. El 10% de todos los estudiantes son de primer semestre y tienen promedio bajo.
- El 84% de los estudiantes retirados son de estrato socioeconómico 2 y provienen de municipios del sur de Nariño. El 2.5% de todos los estudiantes se han retirado, son de estrato 2 y provienen del sur de Nariño.
- El 89% de los estudiantes retirados son de primer semestre, tienen un ponderado ICFES entre 50 y 70 y proceden del sur de Nariño. El 2.5% de todos los estudiantes se han retirado, son de primer semestre, tienen un ponderado ICFES entre 50 y 70 y provienen del Sur de Nariño.
- El 88% de los estudiantes retirados tienen una edad de ingreso menor que 18 años y provienen del sur de Nariño. El 2.5% de todos los estudiantes se han retirado, tienen una edad de ingreso menor de 18 años y son del Sur de Nariño.
- El 86% de estudiantes retirados terminaron su bachillerato en colegios públicos, son de primer semestre y provienen del sur de Nariño. El 2.5% de todos los estudiantes se han retirado, terminaron su bachillerato en colegios públicos, son de primer semestre y provienen del sur de Nariño.

De acuerdo con los resultados obtenidos, la mayoría de los estudiantes de primer semestre, provenientes de la zona sur del departamento de Nariño, de estratos socioeconómicos bajos y matriculados en algún programa de la Facultad de Ciencias Naturales y Matemáticas o en la Facultad de Ciencias Humanas, presentan un bajo rendimiento académico. Este perfil es similar al perfil de la mayoría de estudiantes que se retiran. Por

otra parte, la mayoría de estudiantes que se retiran de estas dos facultades no reingresan, lo que no sucede en la Facultad de Ingeniería, a la cual casi la mayoría de estudiantes retirados reingresan.

Conclusiones y recomendaciones

Se han presentado los resultados del primer proyecto de investigación realizado en la Universidad de Nariño aplicando técnicas de minería de datos para determinar perfiles de bajo rendimiento y deserción estudiantil en sus programas de pregrado.

En este proyecto las fases de preprocesamiento y transformación de datos fueron las más costosas en tiempo, debido a la mala calidad de los datos de la base de datos de la población estudiantil utilizada en esta investigación. Se encontraron muchos datos nulos o faltantes y otros redundantes. Además, en el cambio de un semestre a otro se adicionan nuevos atributos y otros se abandonan, lo que altera los datos históricos de la base de datos. Esto significó que de 46.173 registros seleccionados, se analizaran solamente 20.329, lo que incidió negativamente en los resultados del estudio.

Es necesario que los patrones de bajo rendimiento y deserción obtenidos se analicen detenidamente por las directivas de la Universidad de Nariño, con el fin de tomar decisiones y proponer estrategias conducentes a prevenir que estudiantes con estos perfiles deserten o caigan en bajo rendimiento. Específicamente, se debe hacer un seguimiento a los estudiantes de primer semestre que provienen de los municipios de la zona sur de Nariño que ingresan a programas de las facultades de Ciencias Naturales y Matemáticas e Ingeniería, para disminuir el alto grado de deserción. Por otra parte, se recomienda la construcción de una bodega de datos y el rediseño del sistema de información de Registro Académico que permita obtener datos de calidad que soporten futuros proyectos encaminados al fortalecimiento de la toma de decisiones con

respecto al problema de deserción estudiantil en la Universidad de Nariño.

Bibliografía

- AGRAWAL, R.; SRIKANT, R. (1994). *Fast algorithms for mining association rules*. In: Proceedings of VLDB conference. Santiago, Chile.
- CHEN, M.; HAN, J.; YU, P. (1996). *Data mining: An overview from database perspective*. IEEE Transactions on Knowledge and Data Engineering.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. (1996). *The KDD process for extracting useful knowledge from volumes of data*. Communications of the ACM, Vol. 39, (No 11).
- HAN, J.; KAMBER, M. (2001). *Data mining concepts and techniques*. San Francisco (CA): Morgan Kaufmann Publishers.
- HAN, J.; PEI, J.; YIN, Y. (2000). *Mining frequent patterns without candidate generation*. In: Proceedings of ACM SIGMOD. Dallas (TX).
- HERNÁNDEZ, O.J.; RAMÍREZ, Q. M.; FERRI, R.C. (2004). *Introducción a la minería de datos*. Madrid (España): Editorial Pearson Prentice Hall.
- IMIELINSKI, T.; MANNILA, H. (1996). *A database perspective on knowledge discovery*. Communications of the ACM, Vol. 39, (No. 11).
- QUINLAN, J.R. (1993). *C4.5: Programs for machine learning*. San Francisco (CA): Morgan Kaufmann Publishers.
- ROJAS, B.M.; GONZÁLEZ, D.C. (2008). *Deserción estudiantil en la Universidad de Ibagué*. Revista Zona Próxima (No 9), pp. 70-83, ISSN 1657-24, Universidad del Norte, Colombia.
- TIMARÁN, R.; CALDERÓN, A.; RAMÍREZ, F.; GUEVARA, F.; ALVARADO, J. (2007). *TariyKDD una herramienta de minería de datos débilmente acoplada con un SGBD*. En: Memorias de VII Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería del

- Conocimiento. Guayaquil, Ecuador: Escuela Superior del Litoral. Guayaquil, Ecuador, pp. 3-11.
- TIMARÁN, R.; MILLÁN, M. (2005a). *EquipAsso: un algoritmo para el descubrimiento de reglas de asociación basado en operadores algebraicos*. En: memorias de 4ª Conferencia Iberoamericana en Sistemas, Cibernética e Informática CISCI 2005. Orlando (Florida), pp. 343-348.
 - TIMARÁN, R.; MILLÁN, M. (2005b). *EquipAsso: an algorithm based on new relational algebraic operators for association rules discovery*. In: Proceedings of the Fourth IASTED International Conference on Computational Intelligence. ACTA Press, Calgary (Canada).
 - TIMARÁN, R. (2007). *Mate-tree: un algoritmo para el descubrimiento de reglas de clasificación basado en operadores algebraicos relacionales*. En: Memorias de 6ª Conferencia Iberoamericana en Sistemas, Cibernética e Informática CISCI 2007. Orlando (Florida), pp 196-201.
 - UPN (2009). *La deserción estudiantil: reto investigativo y estratégico asumido de forma integral por la UPN*. [Consulta: marzo 16 de 2009]. Disponible en: URL:http://www.mineducacion.gov.co/1621/articles-85600_Archivo_pdf3.pdf.
 - WITTEN, I.H, FRANK, E. (2000). *Data mining practical machine learning tools and techniques with Java Implementations*. San Francisco (CA): Morgan Kaufmann Publishers.