MODELO DE DETECCIÓN DE ESTUDIANTES EXCLUIDOS EN CARRERAS DE INGENIERÍA UTILIZANDO MINERÍA DE DATOS

Investigación

Fecha de recepción:
10 de julio de 2013

Aprobación:
2 de agosto de 2013

Alveiro Alonso Rosado Gómez
MsC.
Investigador del Grupo de Investigación en Tecnología y Desarrollo en Ingeniería-GITYD, Facultad de Ingenierías
Universidad Francisco de Paula Santander Ocaña, Colombia
aarosadog@ufpso.edu.co

Resumen

Este artículo describe el proceso de descubrimiento de conocimiento de la Base de Datos del Sistema de Información Académico de la Universidad Francisco de Paula Santander de Ocaña, con la finalidad de detectar que estudiantes pueden ser excluidos de los programas de pregrado de la Facultad de Ingeniería por bajo rendimiento académico. Utilizando el aprendizaje supervisado se pudo determinar los factores que permiten predecir la deserción, como son: a) la edad, b) la fecha de ingreso y c) el semestre que cursa. Igualmente los resultados demuestran que los estudiantes con edades superiores a los 26 años, con permanencia de más de cinco años en la universidad y que además se encuentran cursando los primeros semestres son los más propensos a la exclusión.

Palabras Clave

Aprendizaje supervisado, Minería de Datos, KDD, rendimiento Académico.

Abstract

This paper describes the knowledge discovery process of the Database System Academic Information Francisco de Paula Santander University of Ocaña, in order to detect that students may be excluded from the undergraduate programs of the Faculty of Engineering underachievement. Using supervised learning is able to determine the factors that predict attrition, such as: a) age, b) the date of admission and c) the semester courses. Also the results show that students with above 26 years age, with tenure of more than five years in college and also are studying the first semester are more likely to exclusion.

Keywords

Supervised Learning, Data Mining, KDD, Academic Performance.

Introducción

El bajo rendimiento académico es considerado como un tipo de abandono, que contribuye al aumento de la deserción estudiantil en las instituciones de educación superior (Guzmán Ruiz, y otros, 2009). Al caracterizar los estudiantes que pueden ser excluidos del programa por bajo rendimiento académico de la UFPSO, se espera que sea insumo para la definición ó redefinición de mecanismos que permitan fortalecer los apoyos académico, económicos, y de bienestar que ofrece la Universidad a sus estudiantes.

En Colombia las investigaciones sobre la deserción estudiantil, son hechas por el Ministerio de Educación Nacional (MEN) y por las Instituciones de Educación Superior (IES) Oficiales y No Oficiales. La tasa de deserción en las IES, hace que se disminuya la eficiencia del sistema de educación y aumenten las dificultades para el cumplimiento de la función social de la educación, particularmente en aspectos de equidad social y utilización eficiente de recursos estatales, institucionales y familiares (Ministerio de Educación Nacional, 2008).

La deserción estudiantil en la educación superior en Colombia, ha sido una de las principales preocupaciones del MEN, es por ello que como meta se ha propuesto reducir al 25% la deserción para el año 2019 (Departamento Nacional de Planeación (DNP), 2005). De acuerdo al seguimiento hecho por el MEN a la deserción estudiantil en la educación superior, se obtuvo como resultado que el factor determinante del abandono de estudios en Colombia se sitúa en la dimensión académica asociado al potencial o capital cultural y académico con el cual ingresan los estudiantes a la educación superior seguido de los factores financieros, socioeconómicos, institucionales y los de orientación vocacional y profesional (Guzmán Ruiz, y otros, 2009). Como parte de la solución el MEN implemento el Sistema de Prevención de la Deserción en Educación Superior (SPADIES), el cual es una herramienta informática que permite hacer seguimiento al problema de la deserción mediante la generación de estadísticas que permitan identificar los motivos que llevarían a un estudiante abandonar sus estudios y por otro lado permite evaluar las estrategias aplicadas por las IES para evitar este problema (Ministerio de Educación Nacional, 2008).

Además de los esfuerzos realizados por el MEN, existen estudios realizados por IES de forma independiente uno de ellos es el realizado en la Universidad de Nariño por el investigador Ricardo Timarán, el cual busca determinar en la comunidad universitaria los perfiles de bajo rendimiento académico y deserción estudiantil, para lo cual se utilizó la base de datos histórica de los estudiantes de pregrado, uno de los resultados que se obtuvo con este análisis, fue que el 89% de los estudiantes retirados son de primer semestre, tienen un ponderado Saber 11 entre 50 y 70 y proceden del sur de Nariño (Timarán Pereira, Una Lectura sobre deserción universitaria en estudiantes de pregrado desde la perspectiva de la minería de datos, 2010) (Timarán Pereira, Detección de Patrones de Bajo Rendimiento Académico y Deserción Estudiantil con Técnicas de Minería de Datos, 2009), otro estudio realizado por la Universidad Autónoma de Colombia, relaciono las técnicas de moldeamiento predictivo y moldeamiento descriptivo, análisis de retención y sus algoritmos respectivos, en el cual se definió que no todos los modelos y algoritmos desarrollados dan solución a los problemas planteados o satisfacen sus necesidades con respecto al tratamiento de la deserción (Castillo Santos, 2006).

En el ámbito internacional la Universidad Politécnica de Valencia, realizó un análisis que relaciona el rendimiento académico con las características socioeconómicas y académicas de los alumnos, estos atributos son obtenidos en el momento de su matrícula, y se almacenan en la base de datos de la universidad, la utilización de técnicas de Minería de Datos determino factores que podrían influir en el rendimiento, como las ocupaciones y estudios de los padres, o la edad de ingreso del alumno, aunque estos dependen de la técnica utilizada (Alcover, Benlloch, Blesa, Ferri, & Orallo). En Argentina la Universidad Nacional de Misiones, adelanto un estudio de medición de la deserción encontró que existe una relación entre el bajo rendimiento y los factores socioeconómicos (Andrés, 2009). En otro estudio realizado en México se compararon, el método de clasificación con árboles de decisión con la técnica de agrupamiento a través de la generación de clústeres, en ella se encontró que las causas de la deserción son la edad, el ingreso familiar y el nivel de inglés del estudiante (Valero Orea, Salvador Vargas, & García Alonso, 2010). En general los estudios mostrados están enfocados a la deserción dentro de la modalidad Presencial y el nivel educativo Pregrado.

Nan-hua utilizó técnicas de Minería de Datos, para soportar la toma de decisiones en las acciones que permitan controlar el abandono de los estudiantes de sus programas académicos (Hsu, 2008). Otro estudio investigó el comportamiento de la deserción en la modalidad distancia, el cual es mayor que en presencial, en este estudio se propone una serie de factores que inciden en la permanencia de los estudiantes y a cada uno de estos factores se le asignan costos (Kotsiantis, 2009).

El auge que tiene el uso de la Minería de Datos para medir, clasificar y predecir la deserción estudiantil, ha generado una nueva corriente de investigación, conocida como La Minería de Datos Educativos EDM (Educational Data Mining, por sus siglas en ingles) la cual es una disciplina relacionada con el desarrollo de métodos para extraer información útil a partir de los datos que se generan en los entornos educativos, y utilizar dichos métodos para mejorar los ambientes académicos (International EDM Society). Las investigaciones en EDM, están dirigidas a temas como la mejora del software educativo; gracias a los análisis de los datos, se puede medir el efecto de las intervenciones de los profesores en entornos de e-learning, la caracterización del comportamiento y emociones de los estudiantes, y la integración con las teorías pedagógicas (Baker & Yacef, 2009).

Los aspectos que se evalúan como factores de la deserción se agrupan en aspectos Individuales, Académicos, Institucionales y Socioeconómicos. Dentro de los factores Individuales están; la edad, el género, el estado civil y el entorno familiar. Por el lado de los factores Académicos están; el Rendimiento Académico, el tipo de colegio y resultados de examen de ingreso. Dentro de los aspectos Institucionales se tienen; Orden Publico, Becas, Formas de Financiación y por ultimo esta el aspecto Socioeconómico que agrupa atributos como; Estrato, Situación Laboral y Nivel Educativo de los Padres (Castaño, Gallón, Gómez, & Vásquez, 2004).

Las investigaciones anteriores demuestran que se pueden utilizar técnicas de minería de datos para determinar y prestar apoyo a los estudiantes que pueden ser excluidos de la carrera, es por eso que debido a los porcentajes de deserción de la Universidad Francisco de Paula Santander, se pretende realizar e implementar un modelo que permita predecir el abandono de los estudiantes por bajo rendimiento académico.

El documento se estructura de la siguiente forma: en el ítem uno se presenta el desarrollo experimental o metodología, en el ítem dos los resultados, en el ítem tres se presentan las conclusiones, y finalmente las referencias.

Metodología

Para extraer el conocimiento sobre los datos seleccionados y determinar patrones de deserción estudiantil, se organizaron las acciones según el proceso de extracción de conocimiento KDD (del ingles Knowledge Discovery from Databases), en donde se establece que la Minería de Datos, es una de las fases de este proceso (Hernandez Orallo, Ramirez Quintana, & Ramirez Ferri, 2004).

El proceso KDD está formado por cinco fases, cada una de las cuales necesita de los resultados de la fase anterior para poder cumplir con sus responsabilidades, los pasos que se definen en cada una de las fases son los siguientes:

Fase de integración y recopilación de datos

Esta fase consiste en determinar las fuentes de información que pueden ser útiles. Como fuente de datos se determino la base de datos del Sistema de Información Académico, de la Universidad Francisco de Paula Santander; para no violar las restricciones de seguridad y de confidencialidad de la información, definidas por la universidad, debido a esta problemática se diseño una hoja de cálculo que tuviera los atributos e instancias necesarios para realizar el estudio que se extrajeron de la base de datos, se seleccionaron 1863 estudiantes correspondiente a los programas académicos de Ingeniería Civil, Mecánica y de Sistemas, que hacen parte de la Facultad de Ingenierías de la Institución. La Tabla 1 muestra cuáles son los atributos que fueron solicitados para el proyecto, de conformidad con lo definido por (Castaño, Gallón, Gómez, & Vásquez, 2004).

Atributo	Descripción
CIUDAD	Ciudad de procedencia
ESTU_SEXO	Genero del Estudiante M, para masculino y F, para femenino
ESTU_PROMEDIOGENERAL	Promedio acumulado del estudiante (rango de 0 a 5)
ESTU_PROMEDIOPERIODO	Promedio del último semestre cursado por el estudiante (rango de 0 a 5)
ESTU_SEMESTRE	Indica el último semestre aprobado por el estudiante
ESTU_FECHANACIMIENTO	Indica el día, mes y año de nacimiento del estudiante
ESTU_FECHAINGRESO	Fecha de ingreso del estudiante a la universidad
ESTU_ESTRATO	Estrato económico del estudiante (valores de 1 a 4)
ESTU_CREDITOSAPROBADOS	Cantidad de créditos aprobados por el estudiante en los semestres cursados
ESTU_PROGRAMA	Programa en el cual está inscrito el estudiante dentro de la universidad
ESTU_DESERTOR	Indica si el estudiante esta activo académicamente o por el contrario está excluido por bajo rendimiento académico (SI, para los desertores y NO, para los estudiantes activos)

Tabla 1: Atributos trabajados

Fase de selección, limpieza y transformación

Se procesan los datos, buscando eliminar el ruido, transformación de los datos a un formato común, manejo de datos incompletos, dejando los datos listos para ser minados. El nombre de los atributos se estandarizaron con el fin de hacerlos más cortos, quedando de la siguiente manera; ciudad, sexo, promedio_general, promedio_periodo, semestre, edad, fecha_ingreso, estrato, creditos_aprobados, programa, desertor.

Las tablas 2, 3, 4 y 5 muestran como fueron discretizados los valores de los diferentes atributos usados.

Promedio General y Promedio Periodo	Valor
Entre 0 y 3	BAJO
Entre 3.5 y 4	BUENO
Entre 3 y 3.5	ACEPTABLE
Entre 4 y 5	EXCELENTE
Otro valor fuera de rango o nulo	ERROR

Tabla 2: Discretización del atributo Promedio General y Promedio Periodo

Edad	Valor
Menor de 18 años	MENOR_18
Entre 18 y 22	ENTRE_18_22
Entre 22 y 26	ENTRE_22_26
Mayores de 26	MAYOR_26
Datos nulos o valores erróneos	ERROR

Tabla 3: Discretización del atributo edad

Fecha de Ingreso	Valor
Fecha menor al año 2000	ANTES_2000
Fecha entre 2000 y 2005	ENTRE_2000_2005
Fecha entre 2005 y 2008	ENTRE_2005_2008
Fecha entre 2008 y 2010	ENTRE_2008_2010
Años erróneos o valores nulos	ERROR

Tabla 4: Discretización del atributo Fecha de Ingreso

Créditos Aprobados	Valor
Menores o iguales a 0	NINGUNO
Entre 0 y 40	BAJO
Entre 40 y 80	MEDIO
Entre 80 y 120	ALTO
Mayores de 120	MUY_ALTO
Valores erróneos o nulos	ERROR

Tabla 5: Discretización del atributo Créditos Aprobados

Por otra parte, la Figura 1 muestra que existen valores anómalos con respecto a la edad de los estudiantes en los casos en los cuales hay edades entre 40 y 88 años.

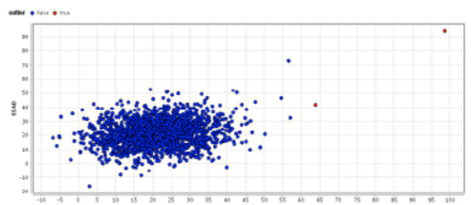


Figura 1: Valores anómalos en el atributo edad

También se encontraron estudiantes con 4 años de edad, en los demás atributos se detecto que existen valores faltantes, los cuales fueron reemplazados por la palabra error con el fin de eliminar sus instancias más fácilmente.

Debido a la cantidad de ciudades que existen, se decidió trabajar solo con las ciudades con mayor número de estudiantes; Ocaña con 679, Aguachica con 105, Río de Oro con 64, Valledupar con 44, Fonseca con 32, Curumaní con 32, Ábrego con 29, Cúcuta con 27 instancias. Y el resto de ciudades se remplazó con el nombre de COLOMBIA. Para el descubrimiento de los candidatos a desertar de la institución se utiliza el algoritmo J48 que permite generar arboles de decisión y el algoritmo de reglas de clasificación llamado JRIP, los cuales están contenidos dentro de la herramienta de Minería de Datos Weka (Timarán Pereira, Una Lectura sobre deserción universitaria en estudiantes de pregrado desde la perspectiva de la minería de datos, 2010).

Fase de evaluación e interpretación

Se evalúan los patrones y son analizados por expertos y, si es necesario, se vuelve a las fases anteriores para una nueva iteración. Los resultados de la aplicación de los algoritmos mencionados se describen en la siguiente sección.

Resultados y análisis

Para poder mirar desde distintos puntos de vista el comportamiento de los datos que se tenían, se dividió el proceso en dos partes en la primera se evalúa la información con todos los programas, atributos y luego por cada programa y solo algunos atributos.

La configuración de Weka para ejecutar ambos algoritmos consistió en definir un 66% del total de los atributos como conjunto de entrada (Percentagesplit) y con el factor de confianza (confidencelevel) para la poda en 25%. El atributo que se selecciono como clase fue desertor, este atributo indica si un estudiante académicamente se encuentra o no activo.

Se evaluó el conjunto de datos agrupando todos los programas académicos, se clasificaron correctamente el 90,6 % de las instancias. Entre las reglas de clasificación más representativas están:

Si el promedio general es aceptable y la ciudad de origen es Aguachica, la edad es mayor de 26 y está en los primeros semestres, entonces el estudiante desertará.

Si el promedio general es aceptable y ciudad de origen es Aguachica, el programa es Ingeniería Mecánica y el promedio del periodo es aceptable, entonces el estudiante es desertor.

Si el promedio general es aceptable, es de los primeros semestres, la ciudad de origen es Abrego, la fecha de ingreso fue entre el 2005 y 2008, entonces el estudiante puede desertar.

Si el promedio general es aceptable, los créditos aprobados son pocos y la fecha de ingreso está entre el 2000 y 2005, entonces el estudiante es desertor.

La figura2, muestra los resultados obtenidos con árboles de decisión; los cuales expresan que los estudiantes con edades mayores de 25 años y menores de 18 son los que más desertan.

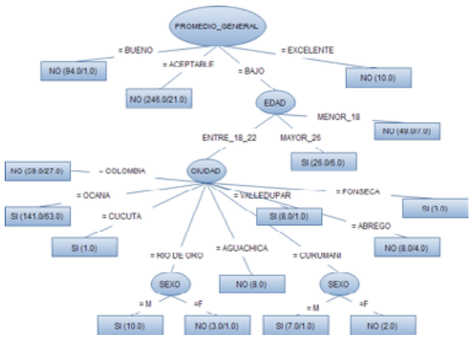


Figura 2. Árbol de Decisión General.

Adicionalmente se hicieron análisis a cada conjunto de datos del mismo programa y como es el comportamiento con diferentes atributos.

Ingeniería Civil

Para el programa ingeniería Civil el proceso arrojo las siguientes clasificaciones y reglas:

En la figura 3, se muestra el árbol de decisión donde el estudiante tiene promedio bajo, tiene una edad que oscila entre 18 y 22 años, es de sexo masculino y proviene de Rio de Oro o de Curumani, con un grado de clasificación correcta de 76%, el estudiante va a ser desertor.

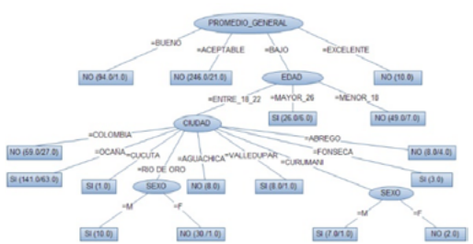


Figura 3. Atributos ciudad, sexo, promedio_general, semestre, edad, estrato, con respecto al atributo desertor

En la figura 4, se muestra como los estudiantes con promedio general bajo cuya edad está entre 18 y 22 años o mayores de 26 son candidatos a desertar, mientras que los menores de 18 no lo harán. Con un grado de clasificación correcta de 78%.

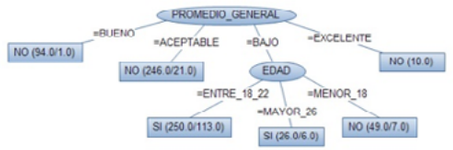


Figura 4. Atributos promedio_general,edad, semestre, con respecto al atributo desertor

Si los créditos aprobados son bajos y la fecha de ingreso esta entre 2005 y 2008, si la fecha de ingreso esta entre 2000 y 2005, entonces son candidatos a desertar. Con un grado de clasificación correcta de 79%.

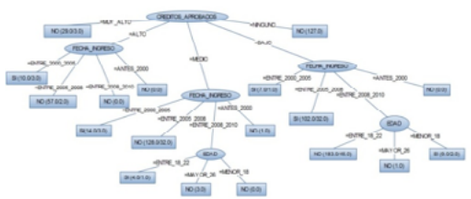


Figura 5. Atributos sexo, fecha_ingreso, edad, creditos_ aprobados, con respecto al atributo desertor

Los estudiantes de promedio bajo cuya fecha de ingreso está comprendida entre el año 2000 al 2008 van ser desertores, mientras que los estudiantes que ingresaron antes del 2000 y entre el 2008 y 2010 no lo serán. Con un grado de clasificación correcta de 81%. Esto se puede observar en la figura 5.

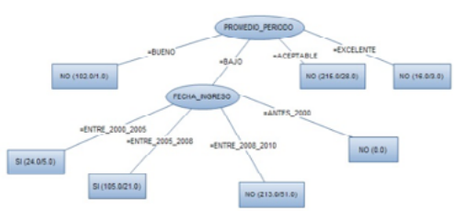


Figura 6. Atributos sexo, fecha_ingreso, edad, promedio_periodo, con respecto al atributo desertor

Ingeniería Mecánica

Para el programa Ingeniería Mecánica el proceso arrojo las siguientes clasificaciones y reglas:

La figura 6, muestra los estudiantes con promedio general bajo, con edades entre los 18, 22 y superiores de 26 años, son candidatos a desertar, al contrario de los estudiantes menores que no desataran, para los otros promedios aceptable, bueno y excelente, no desertaran. Con un grado de clasificación correcta de 77%.

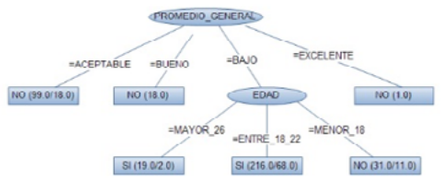


Figura 7. Atributos ciudad, sexo, promedio_general, semestre, edad, estrato, con respecto al atributo desertor

Si promedio general es aceptable o bueno entonces el estudiante no va a desertar. Con un grado de clasificación correcta de 74%. Como se muestra en la figura 7.

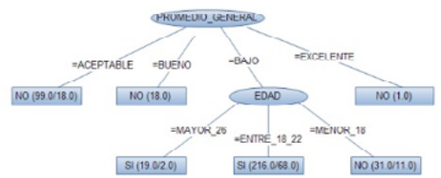


Figura 8. Atributos promedio_general,edad, semestre, con respecto al atributo desertor

Se seleccionan los atributos; sexo, fecha_ingreso, edad, creditos_aprobados, con respecto al atributo desertor.

Como lo muestra la figura 8, para los estudiantes con la mitad de los créditos de su programa aprobados y edad mayor de 26 años, será un desertor, para los que tengan edades entre 18, 22 y mayores de 18 no. Los estudiantes que tengan créditos aprobados bajos se convierten en desertores. Con un grado de clasificación correcta de 76%.



Figura 9. Atributos sexo, fecha_ingreso, edad, creditos_ aprobados, con respecto al atributo desertor

Si fecha de ingreso está entre 2000-2005 y 2008-2010, entonces el estudiante no desertara. Con un grado de clasificación correcta de 75%. Este comportamiento lo ilustra la figura 9.

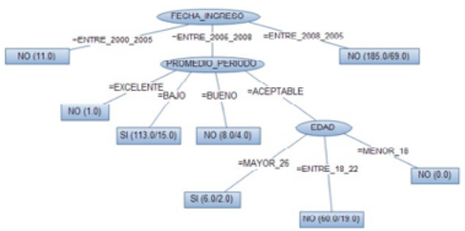


Figura 10. Atributos sexo, fecha_ingreso, edad, promedio_periodo, con respecto al atributo desertor

Ingeniería de Sistemas

Para el programa Ingeniería de Sistemas el proceso arrojo las siguientes clasificaciones y reglas:

Cuando el estudiante tiene un promedio general bajo, y curse los primeros cuatro semestres, tenga una edad entre 18, 26 o sea mayor de 26 años es candidato a desertar, pero al contrario si es menor de 18 años y es menor o igual al estrato dos, el estudiante no desertara del programa. Con un grado de clasificación correcta de 80%.

En la figura 10, se puede apreciar el comportamiento que se produce cuando el promedio general es aceptable y el estudiante proviene de rio de oro entonces puede ser desertor. Con un grado de clasificación correcta de 80%.

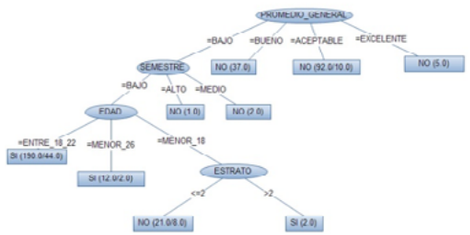


Figura 11. Atributos ciudad, sexo, promedio_general, semestre, edad, estrato, con respecto al atributo desertor

Se seleccionan los atributos; promedio_ general,edad, semestre, con respecto al atributo desertor.

Si el estudiante tiene un promedio general bajo, y está en los primeros semestres del plan de estudios puede desertar. Con un grado de clasificación correcta de 80%



Figura 12. Atributos promedio_general,edad, semestre, con respecto al atributo desertor

Se seleccionan los atributos; sexo, fecha_ingreso, edad, creditos_aprobados, con respecto al atributo desertor.

Si el estudiante tiene aprobado la mitad de los créditos del programa y es de sexo masculino entonces puede ser desertor. Con un grado de clasificación correcta de 77%. Este comportamiento se muestra en la figura 12.



Figura 13. Atributos sexo, fecha_ingreso, edad, creditos_ aprobados, con respecto al atributo desertor

En la imagen 13, se muestra como el estudiante que tenga un promedio en el periodo bajo, ingreso entre el año 2008 y 2010, tiene la edad entre 18 y 22 años y es de sexo masculino, puede desertar. Con un grado de clasificación correcta de 77%.

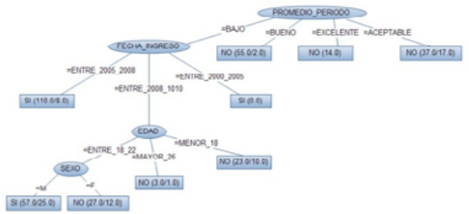


Figura 14. Atributos sexo, fecha_ingreso, edad, promedio_ periodo, con respecto al atributo desertor

De acuerdo con los resultados obtenidos, la mayoría de los desertores tienen un promedio general bajo, una edad que oscila entre los 18 y 22 años, cursa los primeros semestres del programa académico y los créditos aprobados son pocos. De forma contraria los no desertores son estudiantes con suficientes créditos aprobados y mayores de 18 años.

Conclusiones

Se demuestra que los estudiantes con edades superiores a los 26 años, con permanencia de más de cinco en la universidad y que además se encuentran cursando los primeros semestres tienen una alta probabilidad de desertar o ser excluidos, y las variables que más influyen son: la edad, la fecha de ingreso y el semestre que cursa. Se puede concluir que el modelo permite predecir el abandono de los estudiantes por bajo rendimiento académico en la Universidad Francisco de Paula Santander Ocaña.

Referencias

Acosta, M. d. (2009). Los procesos de aprendizaje y su incidencia en la deserción estudiantil en el programa de química farmacéutica de la universidad de cartagena. Psicología desde el Caribe(24), 33.

Alcover, R., Benlloch, J., Blesa, P., Ferri, C., & Orallo, H. (s.f.). Análisis del rendimiento académico en los estudios de informática de la Universidad Politécnica de Valencia aplicando técnicas de minería de datos.

Andrés, J. G. (2009). Minería de Datos aplicada al análisis de la deserción en la Carrera de Analista en Sistemas de Computación. Universidad Nacional de Misiones.

Baker, R., & Yacef, K. (Octubre de 2009). The State of Educational Data Mining in 2009: A Review and Future Visions. JEDM - Journal of Educational Data Mining, págs. 3-17.

Castaño, E., Gallón, S., Gómez, K., & Vásquez, J. (enero - junio de 2004). Deserción estudiantil universitaria: una aplicación de modelos de duración. Lecturas de Economía(60), 39-65.

Castillo Santos, R. (2006). Estrategias y Procesos Relacionados con la Adquisición y Retención de Estudiantes en una Universidad Privada Utilizando Herramientas y Tecnologías de Inteligencia de Negocios. Clepsidra, 77-85.

Departamento Nacional de Planeación (DNP). (2005). Visión Colombia II Centenario: 2019, Propuesta para discusión (Tercera Edicion ed.). Bogota, Colombia: Editorial Planeta Colombiana S. A.

Guzmán Ruiz, C., Durán Muriel, D., Franco Gallego, J., Castaño Vélez, E., Gallón Gómez, S., Gómez Portilla, K., y otros. (2009). Deserción estudiantil en la educación superior colombiana (Primera edición ed.). (V. d. Superior, Ed.) Bogotá D.C, Colombia: ministerio de educación nacional Viceministerio de Educación Superior.

Hernandez Orallo, J., Ramirez Quintana, M. J., & Ramirez Ferri, C. (2004). Introduccion a la Mineria de Datos. Madrid: Pearson Educacion S.A.

Hsu, Y.-c. (2008). Nhu e Thesis. Recuperado el 20 de Agosto de 2011, de http://libserver2.nhu.edu.tw/ETD-db/ETD-search/view_etd?URN=etd-0701109-153053

International EDM Society. (s.f.). International Educational Data Mining Society. Recuperado el 19 de Agosto de 2011, de http://www.educationaldatamining.org/index.html

Kotsiantis, S. (2009). Educational data mining: a case study for predicting dropout-prone students. International Journal of Knowledge Engineering and Soft Data Paradigms, págs. 101-111.

Lamos Díaz, H., & Giraldo Sagra, J. A. (2011). Un modelo conceptual para el análisis del desempeño académico de los estudiantes de cálculo i en la unab. Revista Educación en Ingeniería(12), 11.

Mateus Rodríguez, M. J., Herrera Hernández, C., Perilla Suárez, C., Parra Quecan, G., & Vera Maldonado, A. (2011). Factores presentes en la deserción universitaria en la facultad de psicología de la universidad de san buenaventura, sede bogotá en el periodo comprendido entre 1998-2009. Psychologia, 5(1), 13.

Ministerio de Educación Nacional. (Abril de 2008). Obtenido de SPADIES Sistema de Prevención y Análisis a la Deserción en las Instituciones de Educación Superior: http://www.mineducacion.gov.co/1621/article-156292.html

Ministerio de Educación Nacional. (2008). Deserción Estudiantil en la Educación Superior Colombiana, Elementos para su Diagnostico y Tratamiento. Obtenido de Ministerio de Educación Nacional.

Ministerio de Educación Nacional. (06 de 2012). Obtenido de Deserción estudiantil, Consulta Personalizada, SPADIES: http://spadies.mineducacion.gov.co/spadies/JSON.html

Rodríguez Núñez, L. H., & Londoño Londoño, F. J. (Mayo de 2011). Estudio sobre deserción estudiantil en los programas de Educación de la Católica del Norte Fundación Universitaria. Revista Virtual Universidad Católica del Norte, 33, 28.

Rojas Betancur, M., & González, D. C. (2008). Deserción estudiantil en la Universidad de Ibagué, Colombia: una lectura histórica en perspectiva cuantitativa. Zona Próxima(9), 14.

Salcedo Escarria, A. (Marzo-Abril de 2010). Deserción universitaria en Colombia. Revista Académia y Virtualidad, 3(1), 11.

Timarán Pereira, R. (2009). Detección de Patrones de Bajo Rendimiento Académico y Deserción Estudiantil con Técnicas de Minería de Datos. Octava Conferencia Iberoamericana en Sistemas, Cibernética e Informática: CISCI 2009, (págs. 1-5). Orlando.

Timarán Pereira, R. (2010). Una Lectura sobre deserción universitaria en estudiantes de pregrado desde la perspectiva de la minería de datos. Revista Científica Guillermo de Ockham, 121-130.

Valero Orea, S., Salvador Vargas, A., & García Alonso, M. (2010). Minería de Datos: Predicción de la Deserción Escolar Mediante el Algoritmo de Arboles de Decisión y el Algoritmo de los k Vecinos más Cercanos. Recursos Digitales para la Educación y la Cultura, 33-39.