

CIENCIAS COMPUTACIONALES

MEMORIAS 13° CONGRESO COLOMBIANO DE COMPUTACIÓN

ISBN: 978 958 8862 54 - 2

Juan Carlos Martinez Santos
Jairo Enrique Serrano Castañeda

Sociedad Colombiana de Computación

Universidad Tecnológica de Bolívar
CARTAGENA DE INDIAS

VIGILADA MINEDUCACIÓN

RECTOR
Alberto Roa Várelo

VICERRECTOR ACADÉMICO
Daniel Toro González

VICERRECTORA ADMINISTRATIVA Y FINANCIERA
María del Rosario Gutiérrez de Piñeres

DECANO DE INGENIERÍAS
José Luis Villa

COMPILADORES
Juan Carlos Martínez Santos
Jairo Enrique Serrano Castañeda

ISBN: 978-958-8862-54-2 (digital)

www.utb.edu.co

Campus Tecnológico

Parque Industrial y Tecnológico Carlos Vélez Pombo Km 1 Vía Turbaco
Tel:+57 5 6535200 - 6619240
(+57) 323 566 8733 – 323 566 8732 - 323 566 8730 – 323 566 8729

Campus Casa Lemaitre

Calle del Bouquet Cra.21 #25-92 Barrio Manga
Tel:+57 5 6606041

Prólogo

El Congreso Colombiano de Computación es el encuentro informático anual organizado por la Sociedad Colombiana de Computación (www.sco2.org). Tiene como objetivo propiciar el fortalecimiento de la comunidad colombiana en informática, acercando a investigadores, estudiantes y practicantes, tanto nacionales como internacionales.

La Sociedad Colombiana de Computación (SCO2) lleva organizando este congreso desde el año 2005 (Cali) junto con el CLEI 2005, y se ha mantenido en sucesivas ediciones en Bogotá (2007), Medellín (2008), Bucaramanga (2009), Cartagena (2010), Manizales (2011), Medellín (2012) junto con el CLEI 2012, Armenia (2013), Pereira (2014), Bogotá (2015), Popayán (2016) y Cali (2017).

El 13avo Congreso Colombiano de Computación (13CCC) se realizó en Cartagena entre el 26 y el 28 de Septiembre de 2018 y contó con la participación de investigadores nacionales e internacionales. Este año, el congreso estará organizado por la Sociedad Colombiana de Computación y la Universidad Tecnológica de Bolívar. Este congreso pretende crear un espacio de intercambio de ideas, técnicas, metodologías y herramientas con un enfoque multidisciplinar, fomentando las sinergias entre los investigadores, profesionales y empresas relacionadas a las temáticas de interés del Congreso.

El evento central del congreso está compuesto por la presentación de trabajos de investigación con (i) un aporte significativo al conocimiento o (ii) que presenten experiencias innovadoras en las diferentes áreas de la informática. Entre otros, el congreso incluye conferencias plenarias, foros de discusión, tutoriales y un simposio para estudiantes de maestría y doctorado.

Se cubrieron las siguientes áreas:

- Ingeniería de Software y Arquitecturas De TI
- Interacción Humano Computador
- Seguridad de la Información
- Procesamiento de Imágenes, Visión por Computador y Multimedia
- Sistemas Inteligentes y Robótica
- Informática Educativa
- Gestión de la Información y el Conocimiento
- Sistemas Distribuidos y Arquitecturas de Gran Escala
- Métodos Formales, Lógica Computacional y Teoría de la Computación

Nuestros más sinceros agradecimientos a los miembros del comité técnico, a los autores que enviaron sus artículos y a todos los conferencistas y participantes del 13CCC.

Presidentes de la conferencia

- Juan Carlos Martínez Santos, jcmartinezs@utb.edu.co, Universidad Tecnológica de Bolívar, Colombia
- Jairo Serrano Castañeda, jserrano@utb.edu.co, Universidad Tecnológica de Bolívar, Colombia

Comité del Programa

- Mauricio Alba, malba@autonoma.edu.co, Universidad Autónoma de Manizales, Colombia
- Luis Fernando Castro, lufer@uniquindio.edu.co, Universidad del Quindio, Colombia
- César Collazos, ccollazo@unicauca.edu.co, Universidad del Cauca, Colombia
- Toni Granollers, antoni.granollers@udl.cat, Universidad de Lleida, España
- Leonardo Flórez, florez-l@javeriana.edu.co, Pontificia Universidad Javeriana de Bogotá, Colombia
- María Patricia Trujillo, maria.trujillo@correounivalle.edu.co, Universidad del Valle, Colombia
- Nestor Duque, ndduqueme@unal.edu.co, Universidad Nacional de Colombia, Colombia
- Iván Cabezas, imcabezas@usbcali.edu.co, Universidad de San Buenaventura, Colombia
- Carlos Hernán Gómez, carlosh@ucaldas.edu.co, Universidad de Caldas, Colombia
- Harold Castro, hcastro@uniandes.edu.co, Universidad de los Andes, Colombia

Sociedad Colombiana de Computación (SCO2)

- Enrique González, egonzal@javeriana.edu.co
- María Clara Gómez, mcgomez@udem.edu.co
- Yenny Alexandra Méndez Alegría, yenny.mendez@unad.edu.co
- Iván M. Cabezas T., imcabezas@usbcali.edu.co
- Juan Carlos Martinez, jcmartinezs@utb.edu.co
- Andrés Solano, afsolano@uao.edu.co
- Jorge Iván Ríos Patiño, jirios@utp.edu.co

Miembros del programa

- Gerardo M. Sarria M., Pontificia Universidad Javeriana Cali, gsarria@javerianacali.edu.co
- Néstor Darío Duque Méndez, Universidad Nacional de Colombia, ndduqueme@unal.edu.co
- Silvana Aciar, Instituto de Informática, Universidad Nacional de San Juan, Argentina, silvanav.aciar@gmail.com
- Maria Villegas, Universidad del Quindío, mlvillegas@uniquindio.edu.co

- Hector Florez, Universidad Distrital Francisco Jose de Caldas, haflorezf@udistrital.edu.co
- Fabio Martinez Carrillo, Bioingenium Research Group - National University of Colombia, famarcar@saber.uis.edu.co
- Paula Lago, Uniandes, pa.lago52@uniandes.edu.co
- Francisco Alvarez, Universidad Autnoma de Aguascalientes, fjalvar@correo.uaa.mx
- Mauricio Alba-Castro, Universidad Autonoma de Manizales UAM, malba@autonoma.edu.co
- Sonia Contreras Ortiz, Universidad Tecnológica de Bolívar, scontreras@unitecnologica.edu.co
- Victor M. Gonzalez, Instituto Tecnológico Autónomo de México, vmgonz@acm.org
- Edwin Puertas, Universidad Tecnologica de Bolivar, epuerta@unitecnologica.edu.co
- Cristina Manresa-Yee, University of Balearic Islands, cristina.manresa@uib.es
- Pablo Ruiz, Unicomfacauca, phruiz27@gmail.com
- Fabio González, Universidad Nacional de Colombia, fagonzalezo@unal.edu.co
- Luis Fernando Castro Rojas, Universidad Nacional de Colombia - Universidad del Quindío, luferpanfer@gmail.com
- Harold Castro, Communications and Information Technology Group (COMIT), Department of Systems and Computer Engineering, Universidad de Los Andes, Bogotá, Colombia, hcastro@uniandes.edu.co
- Helga Duarte, Universidad Nacional de Colombia, hduarte@unal.edu.co
- Vanessa Agredo Delgado, Unicauca, vanessa.agredo.delgado@gmail.com
- Jorge Villalobos, University of los Andes, jvillalo@uniandes.edu.co
- Andres Moreno, School of Engineering, University of los Andes, dar-more@uniandes.edu.co
- Philippe Palanque, ICS-IRIT, University Toulouse 3, palanque@irit.fr
- Pablo Torres-Carrion, UTPL, pvtorres@utpl.edu.ec
- Patricia Paderewski, University of Granada, patricia@ugr.es
- Juan Pavón, Universidad Complutense de Madrid, jpavon@fdi.ucm.es
- José Antonio Macías Iglesias, Universidad Autónoma de Madrid, j.macias@uam.es
- Ana Isabel Molina Díaz, University of Castilla-La Mancha, AnaIsabel.Molina@uclm.es
- Enrique González, none, egonzal@javeriana.edu.co
- Marta Rosecler Bez, UFRGS, martabez@gmail.com
- Ricardo Azambuja Silveira, Universidade Federal de Santa Catarina, ricardo.azambuja.silveira@gmail.com
- Ivan Cabezas, Universidad de San Buenaventura, imcabezas@gmail.com
- Alicia Mon, Universidad Nacional de La Matanza, alicialmon@gmail.com
- Carina Gonzalez, Universidad de La Laguna, cjgonza@ull.es
- Andrés Adolfo Navarro Newball, Pontificia Universidad Javeriana, Cali, aannewball@hotmail.com
- Olga Marino, Universidad de los Andes, olmarino@uniandes.edu.co
- Jose Luis Villa, U. Tecnológica de Bolívar, jvilla@unitecnologica.edu.co
- Gabriel Pedraza, Universidad Industrial de Santander, gpedraza@uis.edu.co
- Angela Carrillo-Ramos, Pontificia Universidad Javeriana, angela.carrillo@javeriana.edu.co

- Wilson Javier Sarmiento, Universidad Militar Nueva Granada, wilson.sarmiento@unimilitar.edu.co
- Carlos Mario Zapata Jaramillo, Universidad Nacional de Colombia, cmzapata@unal.edu.co
- Daniela Quiñones, Pontificia Universidad Católica de Valparaíso, Chile, danielacqo@gmail.com
- Yannis Dimitriadis, University of Valladolid, yannis@yllera.tel.uva.es
- Jaime Muñoz-Arteaga, Universidad Autónoma de Aguascalientes, jmauaa@gmail.com
- Víctor Bucheli, Universidad del Valle, victor.bucheli@correounivalle.edu.co
- Jaime Chavarriaga, University of Los Andes, jaime.chavarriaga@gmail.com
- José Antonio Pow-Sang, Pontificia Universidad CatÃ³lica del PerÃº, japowsang@pucp.edu.pe
- Fernando De La Rosa R., Universidad de los Andes, fde@uniandes.edu.co
- Cristian Rusu, Pontificia Universidad Catolica de Valparaiso, cristian.rusu@ucv.cl
- Andrea Rueda, Pontificia Universidad Javeriana, rueda-andrea@javeriana.edu.co
- Leonardo Flórez-Valencia, Pontificia Universidad Javeriana, florez-l@javeriana.edu.co
- Gisela T. de Clunie, Universidad Tecnológica de Panamá, gisela.clunie@utp.ac.pa
- Mario Alberto Moreno Rocha, Universidad Tecnológica de la Mixteca, sirpeto@gmail.com
- Jorge E. Camargo, Universidad Antonio Nariño, jorgecamargo@uan.edu.co
- Norha M. Villegas, Universidad Icesi, Cali, Colombia, nvillega@icesi.edu.co
- Omar S. Gómez, Technical School of Chimborazo, ogomez@espoch.edu.ec
- Claudia Roncancio, Grenoble University, Claudia.Roncancio@imag.fr
- Cesar Collazos, Full Professor, ccollazo@unicauca.edu.co
- Tiago Primo, Federal University of Pelotas, tiagoprimo@gmail.com
- Federico Botella, UMH, federico@umh.es
- Maria Patricia Trujillo, Universidad del Valle, maria.trujillo@correounivalle.edu.co
- Kyungmin Bae, Pohang University of Science and Technology (POSTECH), kmbae@postech.ac.kr
- William Caicedo, Universidad Tecnológica de Bolívar, wcaicedo@unitecnologica.edu.co
- Lyda Peña, Universidad Autonoma de Occidente, lpena@uao.edu.co
- Artur Boronat, University of Leicester, aboronat@mcs.le.ac.uk
- Fáber Danilo Giraldo Velásquez, none, fdgiraldo@uniquindio.edu.co
- Mauricio Ayala-Rincon, Universidade de Brasilia, ayala@unb.br
- Andrés Sicard-Ramírez, EAFIT University, asr@eafit.edu.co
- Mayela Coto, Universidad Nacional, Costa Rica, mayela.coto.chotto@una.cr
- Gustavo Isaza, University of Caldas, gustavo.isaza@ucaldas.edu.co
- Miguel Redondo, University of Castilla-La Mancha, Miguel.Redondo@uclm.es
- Carlos Hernan Gomez, Universidad de Caldas, carlosh@ucaldas.edu.co
- Mauricio Toro-Bermudez, Universidad Eafit, mauriciotorob@gmail.com
- Leandro Krug Wives, Universidade Federal do Rio Grande do Sul (UFRGS), wives@inf.ufrgs.br
- Hugo Jair Escalante, INAOE, hugo.jair@gmail.com
- Xabiel García Pañeda, Universidad de Oviedo, xabiel@uniovi.es
- Alfonso Infante Moro, Universidad de Huelva, alfonso.infante@uhu.es
- Toni Granollers, University of Lleida, antoni.granollers@udl.cat

- Luis Freddy Muñoz Sanabria, fundacion Unversitaria de Popayan, lfreddyms@hotmail.com
- Angela Villareal, Universidad del Cauca, angelavillareal@gmail.com
- Juan Francisco Diaz, Universidad del Valle, juanfco.diaz@correounivalle.edu.co
- Leonardo Arturo Bautista Gomez, Barcelona Supercomputing Center, leobago@gmail.com

Sociedad Colombiana de Computación

Universidad Tecnológica de Bolívar
CARTAGENA DE INDIAS

Tabla de Contenido:

- Reconocimientos II
- Prologo III

1. Artículos VIII
    1. Recuperación de información en el Registro Poblacional de Cáncer. Revisión bibliográfica 1
    2. LogSkills: aplicación personalizada basada en actividades lúdicas de apoyo al desarrollo del pensamiento lógico 16
    3. LogSkills: aplicación personalizada basada en actividades lúdicas de apoyo al desarrollo del pensamiento lógico 25
    4. Horus: aplicación personalizada de apoyo al aprendizaje de fraccionarios en niños de básica primaria 42
    5. SPAL: Sistema Personalizado para el Apoyo de Lectura basado en actividades tipo pasatiempo 57
    6. Divitic: una aplicación personalizada para el apoyo en el aprendizaje de divisiones mediante el uso de actividades lúdicas 73
    7. MULTIC: Sistema personalizado de apoyo al aprendizaje de la multiplicación a través de actividades lúdicas 88
    8. ChildDebugging: Practicas, Estrategias y Recursos para la Depuración de Programas en ChildProgramming 105
    9. Procesos algorítmicos en la formación de Contadores Públicos 117
    10. A Metadata Scheme for Educational Competencies: an Extension of the Standard IEEE RCD 121
    11. Sistema adaptativo de apoyo a la toma de decisiones, orientado al ganadero del ecosistema lechero en el departamento de Boyacá 137
    12. Customizable Software Processes: An Exploratory Case Study in a Small Organization 153
    13. Entendiendo el Modding como un Proceso de Ingeniería 165
    14. Evolución en Líneas de Productos de Software: Una Comparación de Técnicas ... 177

15. Documentación del Rationale Arquitectónico usando anotaciones de código fuente: Un estudio exploratorio. 192

16. Análisis de la satisfacción en pruebas de usabilidad, a partir de la expresión facial 203

17. Development Method of a Technological Proposal for User Interface Generation of Interactive Systems based on Video Content Distribution 219

18. Técnicas y metodologías propuestas para la programación de interfaces gráficas de usuario 233

19. COLIBRÍ: Herramienta para la conversión de textos a audios 242

20. Representation of semantic memory unit in cognitive architecture CARINA 256

21. Tales of Etrya: A Digital Game for Motivating Students to Learn and Rehearse English Vocabulary 264

22. Delitos y conductas ilícitas en las redes sociales más utilizadas en Colombia 281

23. Diseño e Implementación de un Cluster Apache Hadoop para la Gestión de Big Data y la Prestación de Servicios de Data Analytics 295

24. BOINC – MGE (Mobile Grid Extension) 312

25. Arquitectura del cluster GRID-SoC para el procesamiento de datos 320

26. Procesos de descubrimiento de conocimiento aplicado al análisis tecnológico y estilístico en hallazgos de una excavación arqueológica: Un estudio de caso 336

27. Uso de sistemas inteligentes para la predicción del desarrollo de falla renal aguda en pacientes hospitalizados 351

28. Shell para Aplicación de Algoritmos Genéticos Simples SheGA 362

29. A prediction model for recommending tags in Stack Overflow 370

30. Escenario IoT para el soporte de agricultura de precisión 383

31. Evaluation of Student Performance Data through Repertory Grid based Method ... 393

32. Analysis of the Software Implementation Process from the Comparison of ERP Systems Implementation Methodologies 400

33. From Digital Signal Processing to Deep Neural Networks: a journey towards modern Automatic Speech Recognizer systems 415

34. Sistema Inteligente de Bajos Recursos Para el Análisis Biomecánico de Ciclistas de Ruta 430

35. Software Educativo para apoyar los procesos de lectoescritura 434

36. KeyScore: musical theory teaching by using serious games 443

37. Modelos de predicción de dianas de microARN basados en métodos de aprendizaje automático y una aproximación al enfoque de aprendizaje profundo 454

38. Freiya, a Gamified Application for Promoting Peace Culture among Teenagers 466

39. An Artificial Neural Network based-system for detection of Android Malware 470

40. Modelo de predicción de deserción de alumnos 474

41. Asistente de conducción basado en visión artificial 478

42. Soy Ciudadano TIC, prototipo de plataforma web educativa para el desarrollo de competencias TIC en ambientes mixtos de aprendizaje dirigida a la población adulta mayor 482

43. Aplicación Móvil de Colaboración Abierta Distribuida (Crowdsourcing) como Apoyo a la Vigilancia Entomológica de Aedes Aegypti 486

44. Diseño e Implementación de un Sistema de Posicionamiento Automático de una Cámara de Video para Pacientes con Esclerosis Lateral Amiotrófica Usando Técnicas de Visión por Computador 490

45. REQCOLLAB, Una solución colaborativa para el desarrollo de software enfocada en la disciplina de la ingeniería de requisitos 494

46. Solución TIC que contribuye a mejorar la eficiencia del proceso de visita de campo por parte de los técnicos de la CVC 498

Modelo de predicción de deserción de alumnos

E.A. Ruiz Tabarez, CI2S Labs, D. López De Luise, CAETI

Buenos Aires, Argentina

Resumen

El presente trabajo introduce un modelo de perfilado de deserción escolar utilizando los datos biográficos dentro la franja etaria de adolescencia, tomando como caso de referencia un colegio privado ubicado en la zona de Urabá región costera de Antioquia Colombia. El modelo se refiere a estudiantes de educación básica secundaria y a educación media definidos por el MEN (Ministerio de Educación Nacional de Colombia) para obtener el título de bachiller. Se predice las condiciones en los alumnos con mayor probabilidad de abandonar sus estudios. La contribución de este trabajo consiste en la identificación de variables esenciales del modelo y el modo en que deben interpretarse, conforme técnicas de aprendizaje automático. Este trabajo comprende una introducción a la problemática de deserción escolar en Urabá, el modelado del perfil de deserción escolar y descripción del modelo inicial, justo con algunas estadísticas preliminares.

Palabras clave: Deserción escolar, aprendizaje automático, modelo predictivo, árbol de inducción.

I. Introducción

La deserción de los estudiantes del sistema formal evidencia la existencia de situaciones de riesgo que llevan a los alumnos a no continuar su formación académica en un centro educativo. [01] Éste es un fenómeno presente tanto en países industrializados, como en vías de desarrollo. [02] Para el presente trabajo se realiza un modelo donde se utilizan datos de un colegio privado ubicado en la zona de Urabá - Colombia. Según [03], la deserción escolar es un fenómeno cultural y socio-económico. Varios estudios en Colombia han encontrado estadísticas y altos índices de deserción que afectan significativamente los esfuerzos realizados para aumentar la cobertura y la calidad de la educación. De manera similar, cerca de uno de cada dos estudiantes que ingresan al sistema educativo no lo completan con éxito o lo hacen en tiempos más alto de lo esperado que corresponde a otra definición o explicación de la deserción escolar dada como [04]. La deserción es opción del estudiante, influenciado positiva o negativamente por circunstancias internas o externas definidas como variables. [05] El Ministerio de Educación Nacional de Colombia, define la deserción como una situación a la que se enfrenta un estudiante cuando aspira y no logra concluir su proyecto educativo, considerándose como desertor a aquel individuo.

Existen diversos mecanismos para manejar y predecir ésta problemática. Entre ellos la minería de datos permite, entre otras cosas, predecir cualquier fenómeno dentro del ámbito educativo. [05] En el presente estudio se determina un conjunto de variables a fin de construir un modelo predictivo obtenido a través de minería de datos. Una característica importante de este modelo es que está ajustado para los estudiantes del ciclo secundario colombiano de una zona apartada de la capital de Colombia y así poder medir que tan notable es en la deserción escolar la migración interna del país, la educación secundaria de Colombia corresponde a dos partes, el ciclo de básica secundaria y educación media [06] [07] [08] [09] como se muestra en la Tabla 1. Para así poder obtener el título de bachiller en Colombia y evaluar así quienes son los posibles desertores.

Lo que resta de este documento se organiza como sigue: Propuesta (sección II), Base de Datos(sección III), Análisis de datos (sección IV), modelo (sección V), conclusiones y trabajos futuros (sección VI).

II. Propuesta

Tomando como punto de partida el entorno internacional como en América Latina, se han desarrollado varios estudios que han fundamentado la teoría de predecir la deserción escolar a través de la minería de datos, el propósito de desarrollar un del perfil del alumno desertor es poder prevenir la deserción, creando estrategias que puedan ser utilizadas y así lograr el objetivo.

La base de datos considerada consiste en la captura de datos del colegio privado ubicado en la zona de Urabá región costera de Antioquia Colombia los datos como el nombre y dirección del colegio se mantienen en reserva por temas de confidencialidad, durante un período desde 2000 hasta 2005 consecutivamente correspondientes a los grados sexto, séptimo, octavo y noveno de la educación básica secundaria, y los grados décimo y undécimo a la educación media como se describe en la Tabla 2. Se aplican técnicas de derivación de reglas básicas a los datos depurados mínimamente, además de extraer estadísticas básicas.

III. Base de Datos

Los datos utilizados corresponden a los años desde el 2000 iniciando con el grado sexto (6), culminando en el año 2005 correspondiente al grado undécimo (11), adicional se realiza el estudio del año 2006 del grado 11(undécimo), que corresponden a los años de la educación básica y media para así obtener el título de bachiller como está estipulado por el Ministerio de Educación Colombiano (MEN). El colegio en estudio es confidencial, se encuentra ubicado en Urabá región geográfica en el departamento de Antioquia Colombia. Es un colegio privado, categoría asignada por el Ministerio de Educación Colombiano (MEN).

III.A) Selección de datos

Divididos entre básica secundaria y educación media. Donde la básica secundaria corresponde los grados de sexto, séptimo, octavo y noveno de manera consecutiva y la educación media que comprenden los grados décimos y undécimo cada grado de un año de duración para obtener así el título de Bachiller. Los datos seleccionados por año se describen y su distribución de frecuencias en la Tabla 1 y Tabla 2.

Tabla 1: Años utilizados

Básica Secundaria				**Educación Media**		
Sexto	Séptimo	Octavo	Noveno	Décimo	Undécimo	Otro*
2000	2001	2002	2003	2004	2005	2006

**Fuente: Colegio utilizado para el análisis.*

Tabla 2: Datos Utilizados

Año	**Grado**	**Núm. De estudiantes**
2000	Sexto (6)	115
2001	Séptimo (7)	110
2002	Octavo (8)	117
2003	Noveno (9)	119
2004	Décimo (10)	110
2005	Undécimo (11)	95
2006**	Otros (11)	101***

**Fuente: Colegio utilizado para el análisis.*
*** Otro estudiantes que se egresaron en 2006 como décimo año.*
**** En el año 2006 el total de estudiantes matriculados que coinciden en los estudiantes matriculados del 2000 son 4.*

III.B. Limpieza de la base de datos

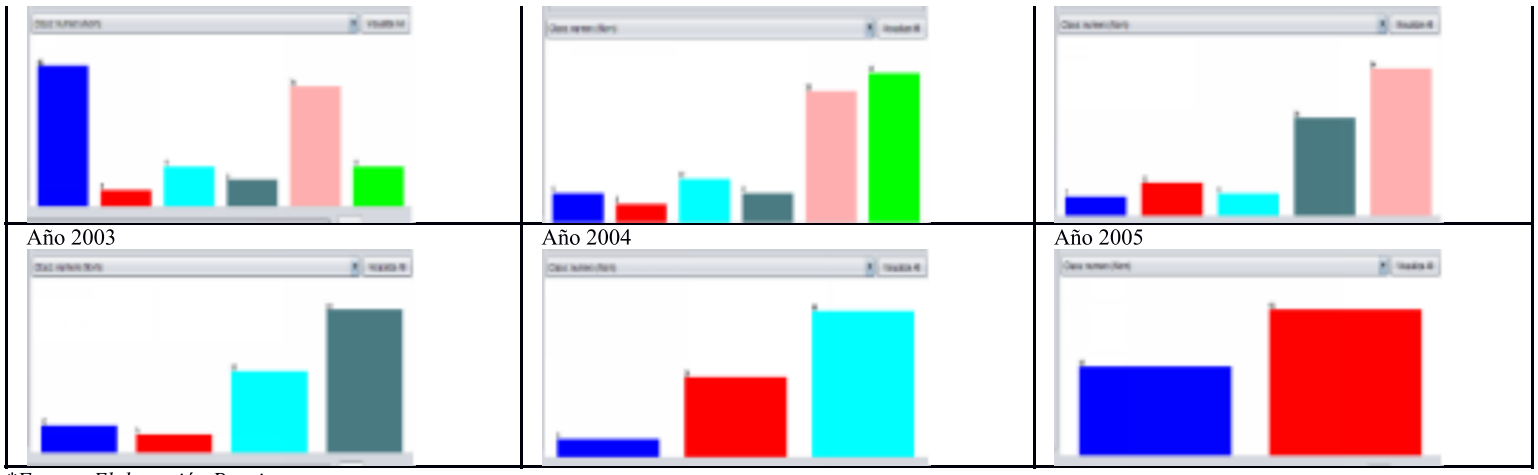
Se limpian los datos no reportados y algunos datos personales se omiten para proteger la identidad tanto de los alumnos y sus familiares responsables. Se agrupan datos y se crea una base de datos "deserción escolar" con un total de 6 tablas distribuidas en anio2000, anio2001, anio2002, anio2003, anio2004, anio2005, anio2006. Los atributos utilizados son: nro(id de identificación), sección(la distribución dentro de los grados que corresponde a: A, B, C), nombre_completo (que se oculta), departamento (Lugar de nacimiento), ciudad (lugar de residencia), fecha_nacimiento, anio (corresponde al año de nacimiento), direccion (domicilio), profesion_padre (profesión del padre), profesion_madre (profesión de la madre), religion (religión que profesa), sexo (F= Femenino, M=Masculino), numero (0,1,2,3,4,5 ó 9 para determinar la continuidad del estudiante). El procesamiento de los datos se realiza por medio del software SPSS y Weka.

IV. Análisis de datos

En la Tabla 3. Se selecciona el atributo número que determina la cantidad de estudiantes que cursan consecutivamente los años correspondientes, y aquellos que van desertando a través de la línea de tiempo.

Tabla 3: Análisis de histograma por año atributo numero.

Análisis Informe año		
Año 2000	Año 2001	Año 2002



**Fuente: Elaboración Propia.*

IV.A. Clasificación

Las estadísticas sobre el modelo muestran que es necesario trabajar con mayor cantidad de casos a fin de clasificar (y predecir) la cantidad de años de cursados previos a la deserción, ésto se verifica usando los algoritmos OneR y J48.

A continuación, se detallan las estadísticas de ambos algoritmos para el año 2000 como año de inicio en ciclo de la secundaría. De manera similar los años 2001, 2002, 2003, 2004, 2005 presentan resultados similares que no se detallan aquí por cuestión de espacio. Por ejemplo, OneR presenta un nivel de confiabilidad más alto al trabajar con cantidades mayores de datos (ver Tabla 4). Las estadísticas OneR y J48 del año 2000 y OneR se repiten para los años restantes.

Como se puede observar, el algoritmo J48 para el año 2000 no alcanza a generar más que un nodo, y la clasificación resulta en un 36% de acierto. Al clasificar con OneR, a pesar de cambiar drásticamente el sesgo, se observa el mismo porcentaje. Similares resultados se obtienen con los años 2001 al 2005 evaluados consecutivamente (ver Tabla 4). Ésto indica que para construir un modelo clasificador requiere más información y lo que sí se obtiene es una descripción de las características típicas de los alumnos en proceso de deserción en base a EM, derivado a reglas. Dicho modelo deberá luego ser contrastado con más casos y eventualmente ajustado, si fuera necesario, con más atributos.

Tabla 4: Estadísticas algoritmos OneR y J48 año 2000 y OneR para para los años correspondientes.

AÑO	**TIPO DE ALGORITMO**	
	OneR	**J48 pruned tree**
2000	Correctly Classified Instances 42 36.5217 % Incorrectly Classified Instances 73 63.4783 %	NODO: 1 (115.0/73.0) Number of Leaves : 1 Size of the tree : 1 Correctly Classified Instances 42 36.5217 % Incorrectly Classified Instances 73 63.4783 %
2001	Correctly Classified Instances 41 37.2727 % Incorrectly Classified Instances 69 62.7273 %	
2002	Correctly Classified Instances 54 46.1538 % Incorrectly Classified Instances 63 53.8462 %	
2003	Correctly Classified Instances 63 52.9412 % Incorrectly Classified Instances 56 47.0588 %	
2004	Correctly Classified Instances 66 60 % Incorrectly Classified Instances 44 40 %	
2005	Correctly Classified Instances 59 62.1053 % Incorrectly Classified Instances 36 37.8947 %	

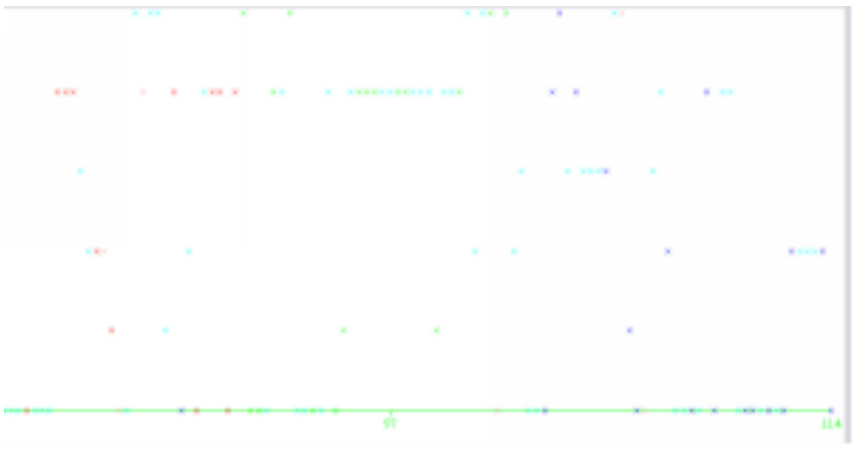
IV.B. Clasificación, Clustering y demostración de resultados

Para el presente análisis se toma como referencia el año 2000. Se observa (Fig. 1a) que Cluster 0, 1 y 4 sólo detentan una instancia en el grupo clasificado como 9 (valor que denota los alumnos que sólo cursan un año). Son relevantes para determinar las características de deserción. Por el contrario, el cluster 2 y 3 son los mayoritarios. El atributo anio_nacimiento para el año 2000 corresponde a 115 estudiantes del grado sexto de la básica secundaria.

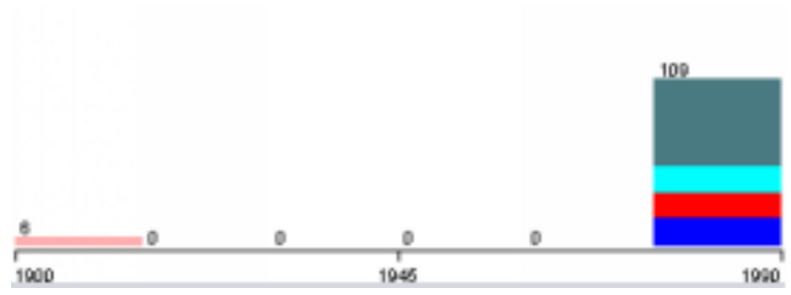
V.Modelo

A continuación se observan las características de los estudiantes con alta deserción para el año 2000 (tabla 5 y figura 2). Resultados similares se observan para los años 2001 a 2005.

Figura 1: Análisis de histograma por año atributo numero.



a) Característica de dispersión año 2000



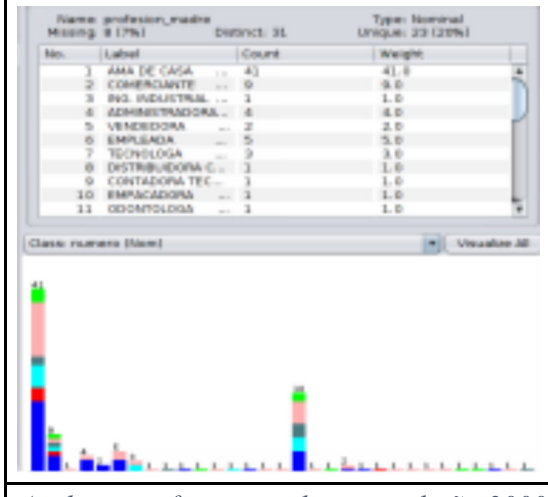
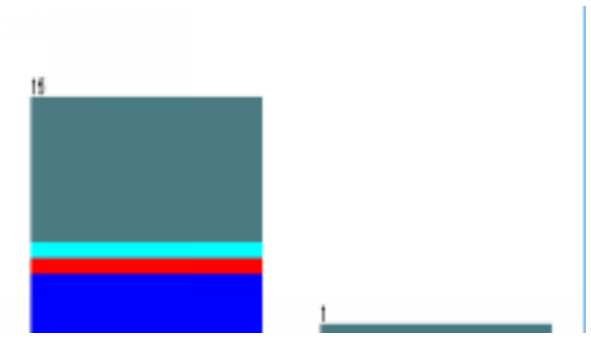
b) anio_nacimiento para el año 2000

Tabla 5: Resultados del modelo predictivo

Atributos	**Descripción**	
anio_nacimiento	El atributo no es esencial para determinar el cluster, lo que indicaría que la edad en sí como atributo en el análisis no influye en el modelo de deserción por lo que se descarta, lo que constituye un hallazgo.	Características de quienes no realizan el cursado completo.
profesion_madre	Mayormente *Ama de casa* tiene una tasa de 35,65%. Este atributo es importante en el modelo ya que determina un porcentaje alto en los alumnos desertores.	
profesion_padre:	Atributo no determinante.	
Sexo	M=*Masculino* (42%) especialmente en cluster 2, y F=*Femenino*(22%) en cluster 3	
Barrio	Típicamente los alumnos de situación socio económica baja ubicados en los barrios *Obrero*, *Policarpa*, *Pueblo nuev*o, y *Chinita* entre otros.	
religion	Hace referencia a la religión que profesa el estudiante, es una variable homogénea.	
Ciudad	Los estudiantes que viven fuera de la ciudad donde está ubicado el colegio. Típicamente las mujeres tienen un porcentaje de deserción del 38.5%	
Profesión_madre	Este atributo es homogéneo tanto madres que realizan una actividad económica o cuentan con una carrera profesional. Como hallazgo se observa que el valor *Ama de*	Características de quienes realizan el cursado completo (atributo numero=5). Corresponden fundamentalmente a los

	casa se encuentra presente en el atributo en un 0,65%.	clasificados como cluster 1.
Profesión_padre	Este atributo es determinante para el alumno que finaliza completamente los años para obtener el título de bachiller. Los valores del atributo son variados (*profesional*, *comerciante*, *empleado*, *oficiales de policía*, entre otros)	
Departamento	El 84,036% de los estudiantes nacieron en el departamento donde se encuentra ubicado el colegio. Si bien no se puede identificar una migración interna, sí es visible que el departamento de origen no corresponde a un atributo determinante para la deserción.	
anio_nacimiento	Atributo no determinante	
Sexo	Es un atributo altamente calificado en el modelo donde se aprecia que la frecuencia mayor se da en el sexo F. (F=*Femenino*, M=*Masculino* valores asignados al atributo).	

Figura 2: Características alumnos con deserción precoz

	
Atributo profesion_madre para el año 2000	*Distribución del atributo sexo para el año 2000*

En función de los resultados expuestos, es evidente que se debe extender el modelo actual. Pero como primer resultado de la presente investigación, se puede identificar que el modelo está compuesto por el conjunto de las reglas descritas en la tabla 6.

Tabla 6: Resultados del modelo predictivo

IF	THEN
profesion_madre='Ama de Casa' AND anio in(1988,1989, 1990) AND sexo=M AND barrio in('Obrero', 'Policarpa', 'Pueblo nuevo', 'Chinita') AND numero<> 5 AND ciudad<>'Apartado' AND departamento= 'Antioquia'	DESERCION-PREMATURA
profesion_madre='Ama de Casa' AND anio in(1988,1989, 1990) AND sexo=M AND numero =9 AND ciudad<>'Apartado' AND departamento= 'Antioquia'	DESERCION-PREMATURA
profesion_madre<>'Ama de Casa' AND profesion_padre in('Profesional' , 'Comerciante','oficiales de Policía') AND anio in(1988,1989, 1990) AND sexo=F AND barrio in('Velez', 'Ortiz', 'villa del rio') AND numero<>5 AND ciudad<>'Apartado' AND departamento='Antioquia'	DESERCION-NO-PREMATURA

En esencia, el modelo consta de las reglas fundamentales para predecir que un alumno potencialmente pueda desertar precozmente.

VI. Conclusiones y Trabajo a Futuro

En este trabajo se han presentado los pasos previos para la generación de una base de datos mínima con información sobre alumnos de educación básica secundaria y media.

Las estadísticas preliminares muestran que la cantidad de registros debe incrementarse sustancialmente a fin de obtener un modelo confiable que permita predecir la deserción. A pesar de ésto, es posible generar un conjunto de reglas que funcionan básicamente para detención grosera de potenciales alumnos con deserción precoz.

Entre los primeros hallazgos se encuentra que la edad del menor no es relevante pero sí el barrio donde vive, profesión de los padres. Así mismo la religión y el departamento de proveniencia tampoco son relevantes.

Queda aún pendiente mejorar las reglas, generar una categorización de los barrios por clase (para no trabajar con bolsas de valores), y ajustar el modelo con más atributos.

Referencias

[01] Moreira-Mora, T. E. (2007). Perfil sociodemográfico y académico de estudiantes en deserción del sistema educativo. *Actualidades en psicología*, *21*(108), 145-165.

[02] Lugo, B. (2013). La deserción estudiantil:¿ realmente es un problema social. *Revista de postgrado FACE-UC*, *7*(12), 289-309.

[03] Pérez-Gama, A., Hoyos, G., Parra-Espitia, L., Ortegón, M., Rozo-Pardo, L. G., & Perez-Gutierrez, B. (2010, September). Education software architecture: facing student desertion in Colombia higher education with an intelligent knowledge based coaching system. In *ANDESCON, 2010 IEEE* (pp. 1-5). IEEE.

[04] Gabriel, P., & Carlos, C. (1999). Deserción Estudiantil Universitaria. *Revista Universidad EAFIT, Medellin*.

[05] Pereira, R. T., Romero, A. C., & Toledo, J. J. (2013). Descubrimiento de perfiles de deserción estudiantil con técnicas de minería de datos. *Revista vínculos*, *10*(1), 373-383.

[06] Antúnez, S., Bolívar, A., Córdoba, F., Del Rey, R., Escaño, J., Rodríguez, M., ... & Viñao, A. (2010). *Procesos y contextos educativos: enseñar en las instituciones de educación secundaria* (Vol. 1). Ministerio de Educación.

[07] Castro, J. E. S., Rodríguez, C. C. R., Marín, C. E. M., García, P. A. G., & Castellanos, J. G. (2016). Hacia la construcción de un modelo predictivo de deserción académica basado en técnicas de minería de datos. *Revista Científica*, *3*(26), 35-48.

[08] Pereira, S. R. T., & Toledo, J. J. (2015). EXTRACCIÓN DE PERFILES DE DESERCIÓN ESTUDIANTIL EN LA INSTITUCIÓN UNIVERSITARIA CESMAG. *InvestigiumIRE*, *6*(1), 30-44.

[09] Pereira, R. T., Romero, A. C., & Toledo, J. J. (2013). Descubrimiento de perfiles de deserción estudiantil con técnicas de minería de datos. *Revista vínculos*, *10*(1), 373-383.

Universidad Tecnológica de Bolívar

CARTAGENA DE INDIAS