



**“DETECCIÓN DE PATRONES DE DESERCIÓN EN LOS PROGRAMAS DE PREGRADO DE LA UNIVERSIDAD MARIANA DE SAN JUAN DE PASTO, APLICANDO EL PROCESO DE DESCUBRIMIENTO DE CONOCIMIENTO SOBRE BASE DE DATOS (KDD) Y SU IMPLEMENTACIÓN EN MODELOS MATEMÁTICOS DE PREDICCIÓN”**

Línea Temática 1: Factores asociados al abandono. Tipos y perfiles de abandono.  
Tipo de comunicación: Derivada de Investigación.

ARGOTE, Ivan  
JIMÉNEZ, Robinson  
GÓMEZ, Jair  
Universidad Mariana - COLOMBIA  
e-mail: iargote@umariana.edu.co

**Resumen.** La investigación titulada: “Detección de Patrones de Deserción en los Programas de Pregrado de la Universidad Mariana de San Juan de Pasto, aplicando el proceso de descubrimiento de conocimiento sobre base de datos (KDD) y su implementación en modelos matemáticos de predicción”. Es un proyecto que tiene como objetivo primordial Aportar al Proceso de Toma de Decisiones para disminuir los niveles Deserción Estudiantil en los programas de pregrado de la Universidad Mariana, mediante el descubrimiento de Conocimiento sobre Bases de Datos (KDD) e implementación en el diseño de unos modelos matemáticos de predicción, en relación con los 4 factores propuestos por el Ministerio de Educación de Colombia concerniente a lo socioeconómico, personal, académico e institucional de cada estudiante, que permitieran visualizar el nivel de deserción de cada educando en los diferentes programas académicos de la Universidad Mariana. Para la construcción de dicho propósito se llevó a cabo 3 objetivos específicos citados como: a). Construcción de un repositorio de datos unificado con la información socioeconómica, académica, personal e institucional de los educandos de los diferentes programas de pregrado de la Universidad Mariana. b). Determinar el perfil del educando desertor mediante una descripción comparativa entre los patrones evidentes y los patrones no evidentes que inciden en la deserción. Y c). Elaboración de unos modelos matemáticos con base en los patrones que presentan una mayor incidencia en la deserción estudiantil en cada uno de los programas de pregrado que se ofrece en la Universidad Mariana. Finalmente, esta investigación se desarrolló bajo un paradigma cuantitativo, haciendo uso de un enfoque empírico analítico y de tipo cuasi experimental, posteriormente se elaboran unas conclusiones y recomendaciones, con el propósito de que sean estos un aporte significativo en la minimización de los índices de deserción estudiantil en la Universidad Mariana y en la educación superior.

**Descriptor o Palabras Clave:** Deserción Estudiantil, Minería de Datos, Factores, Educación Superior, Modelos Matemáticos.

## 1. Introducción

Según el Ministerio de Educación Nacional – MEN, el 52% de los estudiantes colombianos que comienzan una carrera universitaria no la concluyen, así mismo las promociones de estudiantes que terminaron estudios entre 1999 y el 2004, el 48% en promedio no finalizaron sus estudios. Además, de cada dos estudiantes que se matriculan en un programa de pregrado, solo uno culmina su carrera. La preocupación es mayor si se tiene en cuenta que el 39.52% de quienes abandonan sus estudios lo tienen que hacer por razones económicas. [1]

Con base en la problemática descrita anteriormente y los altos índices de deserción estudiantil que se presentan en los diferentes programas de pregrado en la Universidad Mariana de San Juan de Pasto, conllevaron al desarrollo de este proyecto que tiene como objetivo identificar los patrones de deserción estudiantil aplicando el proceso de descubrimiento de Conocimiento sobre Bases de Datos (KDD) y su implementación en el diseño de modelos matemáticos de predicción, los cuales permitieron visualizar e informar el estado actual de los educandos y su relación con la deserción estudiantil en los programas inscritos a la Universidad Mariana.

Como es sabido en Colombia se han realizado diferentes estudios sobre deserción estudiantil, con base en los parámetros grado de deserción y el estado socioeconómico del estudiante, pero sus resultados han sido enfocados al planteamiento de estrategias para el financiamiento de créditos por parte del ICETEX, mas no aplicados a la búsqueda de métodos para el mejoramiento del nivel educativo en todas sus etapas o el de determinar perfiles de estudiantes que estén propensos a caer en bajo rendimiento, lo que ocasiona el abandonar la Universidad, así mismo los estudios observados en el país se realizan con base a encuestas, análisis de datos basado en estadísticas, y porcentajes

calculados con historiales de datos académicos. [1]

Por lo anterior, se consideró que la presente investigación haga una incorporación del proceso KDD, proceso que permite, entre otras cosas, predecir cualquier fenómeno que se manifieste en distintas áreas, para este proyecto se aplicó de acuerdo al ámbito educativo, con el propósito de predecir, con un nivel de confiabilidad la probabilidad de desertar de cualquier estudiante [2]. Para lograr el objetivo deseado se planteó la posibilidad de precisar los referentes teóricos del proceso KDD y deserción estudiantil, con el fin de hacer una correcta implementación del proceso y de esta manera seleccionar los patrones descriptivos de deserción estudiantil de tipo socioeconómico, cultural, académico, disciplinar e institucional a partir de las bases de datos internas, donde se aplicaron técnicas de pre procesamiento y transformación de datos para lograr obtener datos que sean concretos, consistentes y categorizados, que permitieran encontrar información adecuada para la toma de decisiones, por parte de los funcionarios responsables en los diferentes programas de pregrado de la Universidad Mariana con respecto al problema de la deserción estudiantil.

## 2. Descripción del Problema.

Algunas percepciones a la problemática de la deserción estudiantil en los programas de pregrado en la Universidad Mariana, puede sintetizarse en la siguiente descripción de síntomas, causas, diagnóstico y pronóstico.

- **Síntomas**

En Colombia, como en Latinoamérica, la educación superior presenta altas tasas de deserción estudiantil, especialmente en los primeros semestres académicos, a lo cual la los programas de la Universidad Mariana no son ajenos.

Se observa que el índice de deserción es alto en los diferentes programas de pregrado de la Universidad Mariana.

Programa académico	Porcentaje de deserción
Administración de negocios	35,1%
Administración de negocios internacionales	34,3%
Auxiliares de enfermería	0,33%
Comunicación social y periodismo	32,5%
Contaduría pública (Diurno)	23,8%
Contaduría pública (Nocturno)	50,1%
Derecho (Diurno)	14,0%
Derecho (Nocturno)	21,7%
Enfermería	17,5%
Fisioterapia	15,8%
Ingeniería ambiental	30,7%
Ingeniería de procesos	12%
Ingeniería de sistemas	44,3%
Psicología	24,5%
Radiología y radiodiagnóstico	9,78%
Regencia de farmacia	22,2%
Tecnología en automatización electrónica	20,7%
Terapia ocupacional	22,8%
Trabajo social	23,3%
<b>MEDIA UNIVERSIDAD MARIANA</b>	<b>23,5%</b>

Fuente: esta investigación

Los estudiantes no siguen una trayectoria normal del programa académico, por motivos de retiros temporales o absolutos. [3]

#### • Causas

Decisión del estudiante de abandonar su carrera profesional, debido a factores sociales económicos, académicos, culturales, y sociales.

Los estudiantes presentan bajo rendimiento académico, originando que estos repitan y pierdan las asignaturas, lo cual ocasiona retraso en su trayectoria académica.

No es evidente un estudio donde se identifique los motivos reales de porque los estudiantes están abandonando los programas que se ofrecen en la Universidad Mariana.

No es visible en la Universidad Mariana, mecanismos por disminuir estos índices tales como el incremento de los cupos universitarios, sistemas alternativos de financiación, flexibilidad académica, programas de seguimiento y apoyo a posibles desertores, mejoramiento del bienestar universitario, entre otros.

#### • Diagnostico

El diagnóstico que se visualizó en el desarrollo de la investigación fue la escases de sistemas de

información que generen el perfil de un estudiante desertor, con base en los patrones de tipo socioeconómico, académico, culturales, disciplinares e institucionales, para cada uno de los programas de pregrado que ofrece la Universidad Mariana.

#### • Pronostico

Con base en lo anterior, se podía pronosticar que de continuar con el problema de la no identificación de los diferentes patrones de deserción, los niveles de deserción estudiantil no se minimizarían, ocasionando posibles aumentos en estos índices debido a la ausencia de estrategias de retención estudiantil. Es por ello que era de carácter urgente, el desarrollo de un proyecto de investigación que permita identificar patrones de tipo socioeconómico, cultural, académico, disciplinar e institucional, para los diferentes programas de pregrado en la Universidad Mariana utilizando técnicas y algoritmos de minería de datos, y así mantener informada a la comunidad universitaria con el objetivo de generar planes de retención de educandos que beneficien a la Universidad.

### 3. Objetivos

#### ✓ Objetivo general

Aportar al proceso de toma de decisiones para disminuir los niveles deserción estudiantil en los programas de pregrado de la Universidad Mariana mediante el descubrimiento de Conocimiento sobre Bases de Datos (KDD) e implementación en el diseño de modelos matemáticos de predicción.

#### ✓ Objetivo específicos

- Construir un repositorio de datos unificado con la información socioeconómica, académica, personal e institucional de los

educandos de los diferentes programas de pregrado de la Universidad Mariana.

- Determinar el perfil del educando desertor mediante una descripción comparativa entre los patrones evidentes y los patrones no evidentes que inciden en la deserción.
- Elaboración de un modelo matemático con base en los patrones que presentan una mayor incidencia en la deserción estudiantil en cada uno de los programas de pregrado que se ofrece en la Universidad Mariana

#### 4. Marco Metodológico

Este proyecto se desarrolló con base en un paradigma cuantitativo, porque permite la recolección de datos para evaluar la hipótesis, de igual manera se puede utilizar una perspectiva cuantitativa la cual se enfatiza en lo exterior, es decir lo válido o externo, posible de observación [3]. Por lo anterior la actual investigación basa su estudio en el paradigma cuantitativo al estudiar el fenómeno social de la deserción en la educación superior a través de la recolección de información que permitan determinar el nivel de incidencia de cada uno de los patrones de deserción estudiantil de tipo socioeconómico, académico, personal e institucional en los programas académicos de la Universidad Mariana de San Juan de Pasto aplicando el proceso KDD y su posterior análisis inferencial de los resultados.

Además, este proyecto tiene un enfoque empírico – analítico, porque la investigación tuvo como objetivo determinar el perfil de un estudiante desertor en un contexto determinado, el cual tiene existencia propia, independiente de quién lo estudia, el estudio es potenciado por leyes que permiten explicar, predecir y controlar, como también se utiliza la lógica hipotético – deductivo para la demostración del fenómeno social de la deserción estudiantil.

Por otra parte, el tipo de investigación empleado fue cuasi-experimental debido a que el objetivo de la investigación fue seleccionar y analizar muestras con base en características intrínsecas de los sujetos que no son directamente manipulables [4] como son los factores de la deserción estudiantil (sexo, edad, nivel educativo,...). Se habla entonces de manipulación por selección. Además con este tipo de investigación se permite aproximar los resultados de una investigación experimental en situaciones en las que no es posible el control y manipulación absolutos de las variables [5].

Finalmente, la población para la aplicación del proyecto fueron los datos recolectados de los Sistema de información SPADIES y del Sistema de información de la Universidad Mariana que son los entes encargados de llevar los registros de todos los estudiantes de cada uno de los programas académicos para las cohortes del año dos mil cinco (2005) hasta dos mil once (2010) que han sido matriculados en la Universidad Mariana de San Juan de Pasto.

#### 5. Resultados

Como se ha mencionado durante todo el texto, los resultados de la investigación fueron los patrones evidentes y no evidentes que inciden en la deserción estudiantil, y su implementación en el diseño de unos modelos matemáticos de predicción, en los diferentes programas que se ofrecen en la Universidad Mariana.

Es conveniente resaltar que el modelo matemático de predicción permitirá tomar decisiones por parte de los directores y diseñar estrategias o proyectos concernientes a la retención de educandos que minimicen los niveles de deserción en la institución educativa.

Para la presentación de los modelos de predicción para cada factor de tipo socioeconómico, personal, académico e institucional, se hizo descripción de los porcentajes en cada uno de los indicadores que los contiene y su relación con su nivel de incidencia en la deserción del educando.

Es decir, con base en la información, que contiene los porcentajes del estudio de los patrones evidentes y no evidentes, para cada uno de los factores que hicieron parte fundamental de la investigación, se procedió a la creación de las reglas de inferencias, las cuales contienen los modelos matemáticos para cada uno de los factores anteriormente mencionados.

Es importante resaltar que el nivel de incidencia que posee un educando en relación con la deserción estudiantil tiene como fundamento los cuatro factores que cita el Ministerio de Educación, por lo cual se hizo un cálculo matemático para cada uno de ellos y su posterior sumatoria, indicaría el nivel total de deserción que posee un estudiante.

A continuación se hace la presentación de los cuatro modelos que resultaron en la investigación.

Ecuación 1. Modelo Matemático Factor Socioeconómico

$$\frac{E}{\%IE} + \frac{SLE}{\%ISLE} + \frac{SLM}{\%ISLM} + \frac{SLP}{\%ISLP} + \frac{IF}{\%IIF} + \frac{DE}{\%IDE} + \frac{PC}{\%IPC} + \frac{NEM}{\%INEM} + \frac{NEP}{\%INEP} * 100$$

Ecuación 4. Modelo Matemático Factor Académico

$$\frac{RAP}{\%IRAP} + \frac{RAU}{\%IRAU} + \frac{NAP}{\%INAP} + \frac{NAU}{\%INAU} + \frac{TC}{\%ITC} + \frac{PA}{\%IPA} + \frac{F}{\%IF} * 100$$

Ecuación 2. Modelo Matemático Factor Personal

$$\frac{A}{\%IA} + \frac{G}{\%IG} + \frac{EC}{\%IEC} + \frac{NIF}{\%INIF} + \frac{IS}{\%ITS} + \frac{D}{\%ID} * 100$$

Ecuación 3. Modelo Matemático Factor Institucional

$$\frac{NA}{\%INA} + \frac{F}{\%IF} + \frac{RU}{\%IRU} * 100$$

Finalmente, el modelo que permite predecir el perfil de un estudiante desertor en cualquier etapa, en el transcurso de su carrera académica, consiste en la sumatoria de los resultados obtenidos en cada uno de ellos, cabe resaltar que para hacer uso de estos modelos matemáticos es necesario poseer las tablas de normalización que fueron resultado de la investigación para cada uno de los 4 factores que cita el Ministerio de Educación. Por ejemplo para el factor de socioeconómico producto de la investigación resultaron 7 tablas normalizadas, las cuales contienen el porcentaje de incidencia para cada uno de los ítems que componen este factor entre los cuales se encuentran, la estratificación social, la situación laboral del educando, ingresos familiares, dependencia económica, personas a cargo, nivel educativo de la madre y del padre del educando, de igual manera para el factor personal se obtuvieron 6 tablas de normalización entre las cuales se encuentra información acerca de la edad, estado civil, genero, número de integrantes en su núcleo familiar, integración social; así mismo, en el factor institucional existen 3 tablas que contienen información relacionada con la normalidad académica, financiación y recursos académicos, finalmente el factor académico contiene 7 tablas de normalización, las cuales contienen información en relación con el rendimiento académico durante su carrera dividida en dos momentos, número de asignaturas para los primeros y últimos semestres, tipo de colegio, programa académico, facultad académica; todo lo anterior es fruto de la investigación, la cual demuestra que el perfil de un estudiante desertor no se pueden determinar con el mismo nivel de incidencia para cada uno de los 4 factores, esto debido a que cada carrera académica posee una estructura curricular diferente, además de que cada persona adscrita a uno de los programas académicos de la Universidad Mariana posee una

información diferente y única en el momento de su ingreso, como también el transcurso de su carrera universitaria.

A continuación se hace una presentación de algunas de las tablas donde se visualiza el porcentaje de incidencia, en relación con cada ítem presente en los cuatro factores citados por el Ministerio de Educación.

Porcentaje de Incidencia en relación con la Facultad Académica

Facultad	% de Incidencia vs Indicador de Facultad Académica
Facultad de Ciencia Contables	36,1
Facultad de Ciencias de la Salud	17,6
Facultad de Humanidades y Ciencias Sociales	23,2
Facultad de Ingeniería	26,9

Fuente: esta investigación

Porcentaje de Incidencia en relación con el Tipo de Colegio

Programa Académico	% de Incidencia vs Indicador de Tipo de Colegio	
	Privado	Público
Adm. Negocios Internacionales	32,4	19,6
Contaduría Pública Diurno	19,6	24,1
Contaduría Pública Nocturno	11,7	18,7
Enfermería	5,7	16,7
Fisioterapia	29	32,3
Tec. en Radiodiagnóstico y Radioterapia	11,1	55,6
Tec. en Regencia de Farmacia	4,2	26
Terapia Ocupacional	7,3	17,7
Comunicación Social y Periodismo	6,1	18,2
Derecho Diurno	26,8	42,4
Derecho Nocturno	23,3	37,2
Psicología	17,5	18,4
Trabajo Social	12,6	26,4
Ingeniería Ambiental	16,2	21,2
Ingeniería de Procesos	0	33,3
Ingeniería de Sistemas	13,4	18,7
Tec. en Automatización Electrónica	18,8	62,5

Fuente: esta investigación

Porcentaje de Incidencia en relación con la Financiación

Programa Académico	% de Incidencia vs Indicador de Financiación	
	Si	No
Adm. Negocios Internacionales	11,5	88,5
Contaduría Pública Diurno	11,6	88,4
Contaduría Pública Nocturno	14,3	85,7
Enfermería	6,3	93,7
Fisioterapia	32,3	67,7
Tec. en Radiodiagnóstico y Radioterapia	11,1	88,9
Tec. en Regencia de Farmacia	15,6	84,4
Terapia Ocupacional	6,3	93,7
Comunicación Social y Periodismo	12,7	87,3
Derecho Diurno	15,3	84,7
Derecho Nocturno	27,9	72,1
Psicología	10,7	89,3
Trabajo Social	13,8	86,2
Ingeniería Ambiental	13,1	86,9
Ingeniería de Procesos	66,7	33,3
Ingeniería de Sistemas	9,7	90,3
Tec. en Automatización Electrónica	12,5	87,5

Fuente: esta investigación

Porcentaje de Incidencia en relación con el Género

Programa Académico	% de Incidencia vs Indicador de Genero	
	Femenino	Masculino
Adm. Negocios Internacionales	39,9	60,1
Contaduría Pública Diurno	51,3	48,7
Contaduría Pública Nocturno	54,1	45,9
Enfermería	84,5	15,5
Fisioterapia	74,2	25,8
Tec. en Radiodiagnóstico y Radioterapia	77,8	22,2
Tec. en Regencia de Farmacia	66,7	33,3
Terapia Ocupacional	80,2	19,8
Comunicación Social y Periodismo	54,5	45,5
Derecho Diurno	50,8	49,2
Derecho Nocturno	48,8	51,2
Psicología	62,1	37,9
Trabajo Social	90,8	9,2
Ingeniería Ambiental	25,0	75,0
Ingeniería de Procesos	33,3	66,7
Ingeniería de Sistemas	12,7	87,3
Tec. en Automatización Electrónica	0	100,0

Fuente: esta investigación

Con base en estas tablas y la información suministrada por el educando, en cualquier

momento de su estadía en la Universidad Mariana, el personal encargado de la toma de decisiones en relación con la deserción estudiantil, puede monitorear el nivel de deserción del educando, de igual manera determinar el perfil de un estudiante desertor dependiendo de la carrera profesional y de la información histórica del educando que posee la Universidad Mariana, Es decir cada carrera académica posee un perfil diferente en relación con un estudiante que posiblemente tiende a desertar de la Universidad.

En definitiva, el perfil de un estudiante desertor, depende de la información suministrada por el educando u obtenida por el historial del estudiante en su estadía en la Universidad, en relación con cada uno de los ítems presentes en los cuatro factores que cita el Ministerio de Educación, los cuales son el objeto de sustitución en los modelos matemáticos y su producto con cada uno de los valores de incidencia resultado de las tablas de normalización para los factores de tipo Socioeconómico, Académico, Personal e Institucional; por tal razón el perfil de un educando desertor no es el mismo para todos los estudiantes adscritos a la Universidad Mariana.

## 6. Conclusiones y Contribuciones

Las principales conclusiones que se obtuvieron de la investigación fueron:

- Esta investigación se convirtió en un punto de partida en la aplicación sistemas de soporte para la toma de decisiones en relación con el ámbito educativo, basado en la construcción y ejecución de los modelos de minería de datos, para analizar y evaluar los factores que influyen en la deserción universitaria. De igual manera la elaboración de los modelos matemáticos de predicción permitieron a los decanos y directores de programas tomar una decisión adecuada, con base en estudio temprano de la información de los

indicadores que presentaron una mayor incidencia en la deserción estudiantil de sus programas académicos, con el propósito de diseñar estrategias o proyectos concernientes en la retención de los educandos.

- En esta investigación se descubrió que la Universidad Mariana, no hace una recolección de información suficiente referente a la caracterización del estudiante al momento de ingresar a cursar un programa académico, que permitiera establecer modelos de predicción de retención de educandos.

Finalmente, y concluido el trabajo de investigación es necesario establecer algunas recomendaciones para trabajos posteriores en relación con la temática de la deserción estudiantil, entre las cuales se destacan:

- ❖ Recoger un gran conjunto de datos reales incorporando nuevos indicadores, los cuales carecen de calidad en las bases de datos de estudiantes universitarios.
- ❖ Aplicar otros métodos de clasificación para validar el método más adecuado que se adapte a la estructura de los datos suministrados por los sistemas de información que posee la Universidad Mariana y dar una mejor precisión en la clasificación.
- ❖ Profundizar en la categorización y determinación de los indicadores que poseen una mayor incidencia en el problema de la deserción y establecer un procedimiento de captura de dichas indicadores, en el momento que el alumno registra su matrícula en la institución.
- ❖ En cuanto a los modelos de predicción como de los patrones que inciden en la deserción estudiantil en cada uno de los programas académicos adscritos a la Universidad Mariana, es necesario tomar decisiones y proponer estrategias de seguimiento de los educandos que poseen

un alto índice de deserción, para prevenir que se convierta en un estudiante desertor lo cual aumentaría los niveles de deserción estudiantil de la institución.

### Agradecimientos

Por sus conocimientos y valiosos aportes, agradecemos al decano de la Facultad de Ingeniería Dr. Francisco Mafla y al director del programa de Ingeniería de Sistemas Esp. Andrés Muños.

Y finalmente, agradecemos a la Universidad Mariana por propiciar los escenarios para nuestra formación avanzada en la labor que es el eje de nuestra vida: la docencia.

### Referencias

- [1]. M. d. E. Nacional, (2007). Educación Superior en Colombia. Bogotá, Colombia.
- [2]. Valero O. S. (2009). Aplicación de técnicas de minería de datos para predecir la deserción. Matamoros: Universidad Tecnológica de Izúcar de Matamoros.
- [3]. Calvache. L. J. (2006). Paradigma cuantitativo y cualitativo en la investigación socio-educativa: proyección y reflexiones. vol. XIV al XVII.
- [4]. Meltzoff. J. (2000). Crítica a la investigación, Psicología y campos afines, Madrid: Alianza Editorial.
- [5]. Gambara. H. (2002). Métodos de Investigación en Psicología y Educación. Madrid, España.