



**REVISTA**

*educare*

*Órgano Divulgativo de la Subdirección de Investigación y Postgrado  
del Instituto Pedagógico de Barquisimeto "Luis Beltrán Prieto  
Figueroa"*

**BARQUISIMETO – EDO. LARA – VENEZUELA**

**NUEVA ETAPA  
FORMATO ELECTRÓNICO  
DEPOSITO LEGAL: ppi201002LA3674  
ISSN: 2244-7296**

**Volumen 18 N° 2  
Mayo – Agosto 2014**

**MINERÍA DE DATOS APLICADA A LA DESERCIÓN ESTUDIANTIL. CASO:  
LICENCIATURA EN COMPUTACIÓN-UNIVERSIDAD DEL ZULIA-NPF**

***DATA MINING APPLIED TO SCHOOL DROPOUT  
CASE: BACHELOR OF COMPUTING AT UNIVERSIDAD DEL ZULIA-NPF***

**Yelitza J. Marcano\***  
**Rodney H. Rodríguez\***

**\* Universidad del Zulia – Núcleo Punto Fijo (LUZ-NPF)**

**MINERÍA DE DATOS APLICADA A LA DESERCIÓN ESTUDIANTIL.  
 CASO: LICENCIATURA EN COMPUTACIÓN DE LA UNIVERSIDAD DEL  
 ZULIA-NPF**

***DATA MINING APPLIED TO SCHOOL DROPOUT  
 CASE: BACHELOR OF COMPUTING AT UNIVERSIDAD DEL ZULIA-NPF***

<b><u>TRABAJO DE INVESTIGACIÓN</u></b>	<b>Yelitza J. Marcano.* Rodney H. Rodríguez** LUZ - NPF</b>
Recibido:16/11/2013	Aceptado: 20/03/2014
<b>RESUMEN</b>	<b>ABSTRACT</b>
<p>El objetivo de la investigación fue obtener patrones sobre los estudiantes que no han podido concluir sus estudios universitarios, en la Licenciatura en Computación de LUZ-NPF; aplicando para ello la Minería de datos. En tanto, corresponde a una investigación de tipo descriptiva de campo y se desarrolló bajo la metodología computacional Crisp-DM, con apoyo de Weka; los datos provienen de tres poblaciones: estudiantes entre el primer y tercer semestre, profesores entre el primer y quinto semestre del período I-2012, así como los reportes de matrícula 2008-2011 suministrados por Control de Estudios. Se construyó un modelo computacional para la predicción de la deserción estudiantil, empleando: Árboles de decisión C4.5 y la técnica de los k vecinos más cercanos. Entre los resultados se presentan pocos conocimientos previos en el área de lógica y matemática, escasos recursos económicos para proveerse de equipos de computación, falta de concentración en los estudios y pocas horas dedicadas al estudio.</p> <p><b>Descriptores:</b> Minería de Datos, Deserción estudiantil, Crisp-DM</p>	<p>The aim of this study was to obtain patterns on students who could not finish their university studies, in the bachelor of computing at LUZ-NP; by applying data mining. A field-descriptive research was conducted based on Crisp-DM computing methodology with Weka support. Sample data was collected from students between first and third term, teachers from first and fifth term during the period I-2012; as well as the reports from control study's office between 2018 and 2011. A computing model was built to predict student's dropout. C4.5 decision and k neighbors' technique were applied. Results demonstrated not too much previous knowledge in the logic and mathematical areas, reduced economical resources to buy computing equipments, lack of concentration on studies and a few hours dedicated to study.</p> <p><b>Keywords:</b> Data mining, students dropout, Crisp-Dm</p>

\* Ingeniero en Sistemas, Msc. en Gerencia de Empresas de la Univ. Del Zulia. Programa de Ciencia y Tecnología en Computación. Profesora titular a Dedicación Exclusiva de la Universidad del Zulia. Docente Investigadora adscrita al CONDES y PEII categoría B. correo electrónico: [ymarcano@hotmail.com](mailto:ymarcano@hotmail.com)

\*\*Licenciado en computación, egresado de la Universidad del Zulia. Correo electrónico: [rodnehyhr@gmail.com](mailto:rodnehyhr@gmail.com)

## INTRODUCCIÓN

Las instituciones de Educación Universitaria en Venezuela han evidenciado un interés progresivo en dotar gran parte de sus dependencias administrativas y académicas con tecnologías de punta y procesos cada vez más automatizados, a fin de hacerse competitivas y eficientes; para ello hace uso de los beneficios que ofrece la computación, las telecomunicaciones y la informática. En este orden de ideas, Valero (2009), plantea que la Minería de datos es una subdisciplina de las ciencias de la computación que presta apoyo a otras ciencias; su mayor fortaleza radica en la búsqueda de patrones de datos que sean válidos, novedosos, potencialmente útiles y comprensibles.

Es por ello que a través de la Minería de Datos, apoyada en métodos matemáticos de análisis, es posible llevar a cabo el análisis de grandes volúmenes de información almacenada en las bases de datos de las instituciones académicas; y descubrir patrones, asociaciones, cambios o anomalías, sobre datos que tratados de manera aislada, no aportan elementos que permitan tomar decisiones para analizar las razones de deserción universitaria; derivadas por ejemplo del solo análisis de los expedientes académicos de los estudiantes.

Cabe destacar que en el campo de la educación, la Minería de Datos no es un tópico nuevo, su estudio y aplicación ha sido muy relevante en los últimos años. El uso de estas técnicas permite, entre otras cosas, predecir fenómenos dentro del ámbito educativo. En tal sentido, sería valioso poder identificar los perfiles del estudiante durante su permanencia en la universidad y por ende el perfil del estudiante desertor; por tanto el estudio de los factores que afectan a la deserción estudiantil, ha cobrado mayor importancia en los últimos años.

Es por ello que la necesidad de obtener datos para identificar las causas de deserción de los estudiantes en los primeros semestres de su carrera, es indispensable para tomar las acciones pertinentes y poder disminuir este índice, y no menos importante, predecir su deserción en cualquier momento, a efectos de su monitoreo y así poder tomar de acciones correctivas. En este contexto, la Licenciatura en Computación perteneciente al Programa de Ciencia y Tecnología del Núcleo LUZ Punto Fijo, institución científica educativa de

tradición centenaria, paradigma en la formación universitaria a nivel nacional, no escapa de esta realidad. Cabe destacar que según los datos suministrados por Control de Estudios, se aprecia un elevado nivel de deserción en los primeros semestres de la carrera, bien sea por falta de competencias de los estudiantes para asumir los retos del nivel universitario, por no tener una buena orientación vocacional o por la situación socioeconómica imperante que afecta a los jóvenes, coartando su capacidad intelectual y el tiempo disponible para los estudios (Morazan, 2011).

Por lo antes mencionado, esta investigación aplicó la Minería de Datos, a través del uso de técnicas supervisadas y no supervisadas; para obtener datos que permitan determinar y predecir las causas de deserción estudiantil en la Licenciatura en Computación, del Núcleo LUZ-NPF (Universidad del Zulia – Núcleo Punto Fijo, empleando algoritmos mineros, según el diseño de un modelo computacional previo que permitió extraer el conocimiento de una base de datos y así, predecir con cierto grado de certeza y en base a patrones académicos, factores sociales y demográficos, que indicadores pueden aumentar la probabilidad de desertar de la Licenciatura en Computación.

### **ARGUMENTACIÓN TEÓRICA**

La deserción puede entenderse como el abandono de las asignaturas/cursos o la carrera en los que se ha inscrito el estudiante, dejando de asistir a las clases y de cumplir con las obligaciones establecidas previamente, lo cual tiene efectos sobre los índices de la eficiencia terminal de una cohorte. Esta percepción ubica la deserción, como el abandono, entre los problemas más complejos y frecuentes que enfrentan las Instituciones de Educación Superior, así como también el rezago estudiantil y los bajos índices de eficiencia terminal; tanto la deserción como el rezago estudiantil son condiciones que afectan el logro de una alta eficiencia terminal en las instituciones de educación superior, lo que obliga a considerarlos en el marco de la compleja dinámica de la educación superior.

Cabe destacar que la trayectoria académica es un proceso durante el cual cada alumno está sometido a un conjunto de reglas que le permiten avanzar de forma diferenciada, en la medida en la que cumpla o no los requerimientos establecidos por la

institución. En tal sentido, la deserción estudiantil, entendida como una forma de abandono de los estudios superiores, adopta distintos comportamientos en los estudiantes, situación que afecta la continuidad de sus estudios y generalmente estos comportamientos se caracterizan por: Retiro de los alumnos debido a las deficiencias académicas y cambio de carrera o de institución.

Según Anuiés (2007) se ha detectado que la deserción responde a una multiplicidad de factores que afectan a los estudiantes. Entre ellos se encuentran:

- Las condiciones económicas desfavorables de los estudiantes
- El deficiente nivel cultural de la familia al que pertenece
- Las expectativas del estudiante con respecto a la importancia de la educación
- La incompatibilidad del tiempo dedicado al trabajo y a los estudios
- Las características personales del estudiante, por ejemplo, la falta de actitud de logro
- El poco interés por los estudios en general, por la carrera y la institución
- Las características previas del estudiante, como los bajos promedios obtenidos en la educación media que reflejan la insuficiencia de los conocimientos y las habilidades con que egresan los estudiantes, en relación con los requeridos para mantener las exigencias académicas del nivel universitario.
- La deficiente orientación vocacional recibida, antes de ingresar a la educación universitaria, que provoca que los alumnos se inscriban en las carreras profesionales, sin sustentar su decisión en una sólida información sobre la misma.

Para complementar, se puede mencionar a Chalabe, Pérez y Truninger (2001) quienes incluyen los problemas económicos como causales de la deserción en Argentina. Por su parte, Cabral (2005) apunta a factores como: los aspectos socioeconómicos y culturales de los estudiantes y las horas de dedicación al estudio. También figuran los problemas motivacionales, personales psicoafectivos que pueden incidir en el fracaso académico, tales como los sentimientos de frustración, desorientación vocacional, baja autoestima, la posible y precaria adaptación al medio universitario. En Venezuela Castellano (2002), señala además poca identificación vocacional, con respecto a la oferta

específica de las instituciones de educación superior, entre quienes egresan de la educación media, lo que pudiera corresponder a la realidad del caso estudiado.

En correspondencia con el planteamiento de esta investigación, se plantea a la Minería de Datos como aquel proceso mediante el cual se puede obtener un modelo que sirva para la predicción sobre el comportamiento de los datos, específicamente en lo relacionado a la deserción estudiantil. Este modelo se operacionaliza con el empleo o uso de bases de datos, aunado a la aplicación de algún algoritmo que construya el modelo.

Sin duda alguna, actualmente las bases de datos contienen una gran cantidad de datos susceptibles a ser “minables”, datos que muchas veces exceden las capacidades humanas de reducción y análisis, a fin de obtener información útil. Debido a esto, frecuentemente las decisiones importantes se toman en base a la intuición y experiencia en lugar de tomar como referencia la riqueza de estos datos almacenados, provocando que seamos vistos como ricos en datos, pero pobres en información (Kamber, 2006). En aras de abordar esa problemática, se intenta solucionar a través del proceso de KDD (Knowledge Discovery from Databases), para descubrir patrones en una gran cantidad de datos susceptibles a ser minados. Básicamente este proceso se basa en cinco fases (Hernández, Fernández y Ramírez, 2004):

- Integración y recopilación de datos. Integrar múltiples bases de datos en un almacén de datos (data warehouse –colección de datos de las bases de datos transaccionales y otras fuentes diversas). Se determinan las fuentes de información que pueden ser útiles y dónde conseguir las.
- Selección, limpieza y transformación. El objetivo es mejorar la calidad de los datos. Algunos datos son irrelevantes o necesarios para la tarea de minería que se desea realizar. Se eliminan o corrigen los datos incorrectos.
- Minería de Datos. El objetivo es producir conocimiento que pueda utilizar el usuario, realizando un modelo predictivo basado en datos recopilados.
- Evaluación e interpretación. Se evalúan patrones y se analizan para que, de ser necesario, se vuelva a las fases anteriores para una nueva iteración.

- Difusión y uso. Una vez construido y validado el modelo, es usado por los analistas para datos iniciales, almacenamiento de datos, selección de los datos, recomendar acciones y hacer partícipe de él a todos los posibles usuarios.

En síntesis, las técnicas de Minería de Datos, permiten construir modelos predictivos, basados en datos históricos almacenados en distintas fuentes: bases de datos, archivos de texto plano, documentos impresos, reportes, entre otros. Usando todos estos datos, es posible predecir un fenómeno dado, a partir de las herramientas que la minería ofrece, obteniendo conocimiento que ayuda en la toma de decisiones.

Cabe destacar, que la Minería de Datos es un campo en pleno desarrollo en el que se aplican métodos de varias disciplinas como los presentes en sistemas de bases de datos, data warehousing, estadística, el aprendizaje automático, visualización de datos, obtención de información y computación de alta performance. Además, se utilizan métodos de las áreas de redes neuronales, reconocimiento de patrones, análisis espacial de datos, bases de datos de imágenes, procesamiento de señales y programación lógica inductiva; en tal sentido, numerosos especialistas señalan que la Minería de Datos necesita de la integración de enfoques de múltiples disciplinas. Una de sus características principales es que invierte la dinámica del método científico. Es decir, primero se coleccionan los datos y luego se los “escucha” para que de ellos emerjan las hipótesis. Luego se validan esas hipótesis en los datos mismos.

En tal sentido, las técnicas de Minería de Datos (una etapa dentro del proceso completo de KDD) intentan obtener patrones o modelos a partir de los datos recopilados y se clasifican en dos grandes categorías: supervisadas o predictivas y no supervisadas o descriptivas. A continuación se describen de manera general dos de las técnicas empleadas en este estudio:

- Árboles de decisión C4.5: esta técnica está categorizada como aprendizaje basado en similitudes (Kamber, ob. cit), los árboles de decisión son uno de los algoritmos más sencillos y fáciles de implementar y a su vez uno de los más poderosos. Este algoritmo genera un árbol de decisión de forma recursiva al considerar el criterio de

la mayor proporción de ganancia de información (Britos y Hossian, 2005), es decir, elige al atributo que mejor clasifica a los datos.

- Técnica de los k vecinos más cercanos: Conocido como algoritmo de aprendizaje basado en instancias, su funcionamiento es muy simple: se almacenan los ejemplos de entrenamiento de datos históricos y cuando se requiere clasificar a un nuevo objeto, se extraen los objetos más parecidos y se usa su clasificación para clasificar al nuevo objeto. Según Morales (2009) y Kamber (ob. cit), los vecinos más cercanos a una instancia se obtienen, para el caso de los atributos continuos, utilizando la distancia Euclidiana sobre los n posibles atributos y el resultado de la clasificación derivado del uso de este algoritmo, puede ser de tipo discreto o continuo.

### **ASPECTOS METODOLÓGICOS**

El estudio en cuestión según la posición de Sampieri, Fernández y Baptista (2003) es de tipo descriptivo, puesto que busca analizar sistemáticamente características semejantes de los fenómenos estudiados sobre la realidad de las variables en estudio; para este caso específico “Deserción Estudiantil” y “Minería de Datos”, en el ámbito de la Licenciatura en Computación del Núcleo Punto Fijo de LUZ, sin pretender modificar o intervenir en su comportamiento; solo se buscó observarla y describirla. De igual manera, es un estudio de campo debido a que se basa en la recolección de los datos de la realidad sin manipularlos deliberadamente.

La investigación aborda el uso de la minería de datos, como método para evaluar las causas que generan la deserción, tomando como base, fuentes de información, categorizadas en tres grupos: una está representada por los estudiantes de la Licenciatura en Computación que cursaron materias entre el primer y tercer semestre en el período I-2012 (Cuadro 1), como segunda muestra 12 profesores que dictaron clases entre los semestres I y V, en ese mismo período; para efectos de recopilar la información en ambas muestras se diseñó y aplicó dos cuestionarios estructurados, sometidos a una validación de experto y por último se consideraron los reportes de la matrícula estudiantil, clasificada por estatus (desertor, retirado y cambio de ubicación) y períodos (I, II y único), según la base

de datos perteneciente a Control de Estudios de LUZ Núcleo Punto Fijo entre los años 2008 – 2011; estos datos fueron registrados y clasificados en la Tabla 1.

La metodología seleccionada para el desarrollo del modelo de minería fue Crisp-DM, debido a que el proceso de descubrimiento de conocimiento en bases de datos y la metodología Crisp-DM, establecen una serie de fases bien definidas para el diseño y ejecución de un proyecto de Minería (Morales, ob. cit); éstas se desarrollan en la sección de resultados.

Cuadro 1

**Estudiantes inscritos en el I, II, III semestre período I-2012. Programa de Ciencia y Tecnología – LUZ NPF**

I Semestre	II Semestre	III Semestre	Total Población	Muestra
80	20	30	130	44

Tabla 1

**Lista de Estatus por año y por períodos. Control de Estudios. LUZ NPF (2013).**

Años	2008		2009		2010	2011		Total
Período	I	II	I	II	U	I	II	-
Desertor	16	10	26	25	23	38	40	208
Retirado	8	8	16	10	8	7	6	63
Cambio de Ubicación	33	9	25	6	5	2	13	93

**Resultados y Discusión**

Posibles causas que inciden en la deserción estudiantil: Visión del estudiante.

A través del análisis descriptivo e inferencial, en función de la aplicación de un cuestionario aplicado a 44 estudiantes que cursan estudios en la Licenciatura en Computación, entre el primer y tercer semestre (período I-2012), se efectuó el juicio crítico de las respuestas emitidas, tomando como referencia las frecuencias absolutas, para

establecer la relación porcentual de cada ítem, clasificando las preguntas en tres dimensiones, a saber: académica, económica-social y ética.

***Dimensión: Académica***

De acuerdo con los resultados obtenidos, un 61% de la población entrevistada expresó conocer el perfil de ingreso para la Licenciatura en Computación, mientras que 39% de los estudiantes entrevistados afirma no conocerlo. En tal sentido, el hecho de que una parte significativa de las unidades de información conozca el perfil de ingreso para la carrera, permite deducir que el estudiante está preparado y tiene el enfoque suficiente para cursar estudios en el área de computación.

Con respecto a las bases cognitivas, o competencias, un 70% de los estudiantes manifestaron que los conocimientos adquiridos en la educación media no son suficientes para cursar estudios en la Licenciatura en Computación; esto permite inferir que su habilidades o destrezas en matemática y lógica, no cubren los requerimientos de los primeros semestres de la carrera; a diferencia de un 30% quienes manifiestan que si son suficientes. Tomando en cuenta los resultados obtenidos, la muestra encuestada señala en un 75% que no dedican las horas suficientes al estudio aun cuando conocen de los niveles de exigencia de la carrera, mientras que un 25% señala que si dedica el tiempo necesario y suficiente para cumplir a cabalidad con sus compromisos académicos.

Estos resultados evidencian la poca disposición de los estudiantes para dedicar horas al estudio, tanto así que durante la aplicación del cuestionario se evidenciaron dilaciones como: "Un día más no importa, empezaré mañana" y "Sólo revisaré Facebook un cuarto de hora más". Situación que coincide con los resultados de una investigación realizada por Anuies (ob. cit), sobre la deserción estudiantil, donde se obtuvo que en efecto las características previas del estudiante, como los bajos promedios obtenidos en la educación media superior que reflejan la insuficiencia de los conocimientos y las habilidades con que egresan los estudiantes, en relación con los requeridos para mantener las exigencias académicas del nivel superior. De igual manera sucede con el tiempo dedicado a los estudios; el cual se enumera como otra posible causa de deserción.

***Dimensión: Económica-Social.***

Según Fuenmayor, Perozo y Rietveldt (2010) afirman que los aspectos económicos son una posible causa de deserción estudiantil universitaria; situación que coincide con los resultados que arroja la presente investigación, en la cual se obtuvo que un 52% de los estudiantes encuestados certifican que los aspectos económicos si influyen en la deserción estudiantil en la Licenciatura en Computación, mientras que 48% opina lo contrario. Es importante destacar que al momento de iniciar los estudios en la Licenciatura en Computación, no todos los estudiantes cuentan con los recursos tecnológicos necesarios, por lo costosos que estos pueden ser; para ser usados fuera del recinto universitario; entre estos están por ejemplo los equipos de computación.

En cuanto a la procedencia de los estudiantes, un 80% de los entrevistados afirman que residen en el Municipio Carirubana del Estado Falcón; sitio donde se encuentra ubicada la universidad. Mientras que un 20% corresponde a estudiantes foráneos. Es importante destacar que la universidad cuenta con el servicio de rutas estudiantiles urbanas (Municipio Carirubana) e intra-urbanas; beneficiando así a los estudiantes foráneos pertenecientes a los Municipios Falcón y Los Taques.

Entre los indicadores de Deserción estudiantil está el factor socio-económico, dentro de este se construye una nueva categoría: La situación económica – social precaria del estudiante, resultante de la convergencia de las variables el desempleo y la incompatibilidad entre trabajo y estudio. (Sánchez, Navarro y García, 2005). De acuerdo con los resultados obtenidos, un 75% de la población entrevistada manifestó que no tienen inconvenientes para asistir en el horario de clases establecido (diurno), el otro 25 % aprecia lo contrario por lo cual afirman que sería de gran ayuda un horario nocturno, para poder compaginar trabajo y estudio.

En el mismo orden de ideas, un 61% de la muestra entrevistada no posee empleo por lo que no tiene ningún inconveniente respecto al horario de clases, el otro 39% testifica que si posee empleo y además esto le atrae inconvenientes con el horario de clases.

***Dimensión: Ética***

Al consultarse a las unidades de información acerca de su ingreso a la carrera, se evidenció que la mayoría manifestó hacerlo por motivación personal (73%); mientras que el

27% restante optó ingresar a la carrera solo a efecto de obtener un cupo universitario, para luego cambiarse de carrera. Estas aseveraciones permiten inferir que la mayoría de los encuestados están motivados por el estudio de la Licenciatura en Computación y un porcentaje menor optan a futuro por obtener un cambio de carrera.

Las actitudes de estudio que posee un estudiante es un posible indicador de deserción estudiantil (Fuenmayor et. al, ob. cit). En tal sentido, los resultados del presente estudio revelan que un 37% dedica mayor tiempo a Juegos, 24% Redes Sociales, 23% Programación y 16% Investigación, observándose que las opciones relacionadas con la carrera (Investigación y Programación) son las de menor porcentaje y por ende a las que dedica menor tiempo; situación que puede afectar el rendimiento académico y constituirse en una posible causa de deserción.

De igual manera, una mala relación del estudiante con el profesor se puede considerar como un factor de deserción estudiantil (Sánchez et. al, ob. cit), tomando en cuenta los resultados obtenidos por este estudio, un 63% manifiesta que existe una buena relación con los profesores de la carrera, 23% excelente, 14% regular y ninguno de los entrevistados señaló deficiente. Con estos resultados se puede determinar que la relación entre los estudiantes y los profesores es buena, construyendo un trato armónico dentro y fuera de clases, sin ningún tipo de inconvenientes personales con el profesor o viceversa.

Así mismo Sánchez (ob. cit) aportan, que el nivel de relación social que tenga un estudiante con sus compañeros de clases puede afectar o beneficiar el rendimiento académico de un estudiante, depende cual sea el caso, es por esto que una mala relación del estudiante con sus compañeros es considerada como una potencial causa de deserción estudiantil. En base a esto se consultó a los sujetos informantes acerca de cómo es la relación social con sus compañeros de estudio, dando como resultado que un 46% dice que existe una buena relación con compañeros de estudio, 46% excelente, 6% regular y 2% deficiente, con estas aseveraciones se observa que el nivel de integración de un estudiante con sus compañeros de estudios está entre buena y excelente.

Posibles causas que inciden en la deserción estudiantil: Visión del docente.

La deserción universitaria es un fenómeno inherente a la vida estudiantil, en tal sentido al consultar a 12 docentes que dictan clases entre los semestres I – V del primer

período del 2012, para conocer su opinión sobre la deserción estudiantil en la Licenciatura en Computación, se evidenció lo siguiente:

***Dimensión: Académica***

Tomando en cuenta los resultados obtenidos, la muestra entrevistada señala en un 100% que en el primer, segundo y tercer semestre de la carrera es donde se genera mayor deserción estudiantil, demostrando que en estos semestres es donde el estudiante se da cuenta del perfil de la carrera y que éste no se corresponde con sus gustos, intereses actitudes entre otros. De la revisión que se realizó al pensum de estudios, en estos semestres convergen áreas como la de Matemáticas, Programación, Lógica; áreas de conocimiento indispensables, para determinar el comportamiento y desempeño del estudiante durante toda la carrera.

Los profesores entrevistados indicaron que la unidad curricular Algoritmo y Programación I, de la Licenciatura en Computación es la que posee mayor exigencia para el estudiante, es importante señalar que las cuatro materias con mayor ponderación pertenecen a los semestres I, II y III, lo cual afirma que es en los primeros semestres donde existe mayor exigencia para el estudiante y por ende mayor deserción.

Al consultar a las unidades informantes sobre el nivel exigencia, se evidencia que la opinión de los profesores está dividida por igual, 50% de los consultados considera que el nivel de exigencia de las unidades curriculares puede ser causa de deserción estudiantil, el 50% restante opinó lo contrario, pero acotaron que las unidades curriculares del primer y segundo semestre tienen el nivel de exigencia suficiente como para considerarlo factor de deserción estudiantil. Cabe destacar, que los profesores que indicaron que cátedras como Algoritmo y Programación I y II, Estructura de Datos y Cálculo I, son las que demandan mayor nivel de exigencia, puesto que constituyen la base de la lógica formal matemática de la carrera.

En relación al índice de reprobados en las materias ubicadas entre el primer y tercer semestre, el 67% de los profesores entrevistados indicó que no posee un alto índice de estudiantes reprobados, mientras que el 33% restante señala que si; situación que incide en el índice de repitencia en las cátedras antes señaladas. Cabe destacar, que los docentes

manifiestan que se han aplicado estrategias como rotar los profesores, empleo de preparadores y cambios en la dinámica de evaluación, pero no se han conseguido los resultados esperados.

***Dimensión: Social.***

De acuerdo a los resultados obtenidos en este planteamiento, se puede observar que un 84% de total de profesores entrevistados certifican que su relación con los alumnos es excelente coincidiendo con lo planteado por los estudiantes en una pregunta similar. Estos resultados confirman que la relación entre los estudiantes y los profesores de la carrera es Buena por lo que se descarta este ítem como posible causa de deserción estudiantil.

Por último, pero no menos importante, se le pregunta al profesor de la Licenciatura en Computación si aplica técnicas para identificar a los posibles desertores, ya que los profesores son los que orientan a los alumnos de un curso o asignatura y son los primeros en identificar a un posible desertor (González, 2006). Los resultados obtenidos reflejan que un 25% de los profesores no aplica técnicas para identificar a los posibles estudiantes desertores, mientras que solo un 75% dice aplicarlas pero de manera empírica, sin ningún tipo de soporte tecnológico y procedimiento validado, más que todo hacen uso de la conversación de manera informal para indagar sobre sus limitaciones y capacidades; esta es una de las razones por lo que se hace necesario la aplicación de técnicas de Minería de Datos para identificar, cuáles son las causas que afectan a un determinado estudiante para desertar de la Licenciatura en Computación.

Tabla 2

**Unidades Curriculares con mayor exigencia. (Las autoras, 2013).**

Cátedras	Frecuencia	Frecuencia	Total	
	Absoluta	Relativa		
Algoritmo y Programación I	6	50	6	50%
Algoritmo y Programación II	3	25	3	25%
Estructura de Datos	2	17	2	17%
Cálculo I	1	8	1	8%

### **Modelo Minable**

A continuación se muestra de manera detallada las fases de la metodología Crisp-DM, que se aplicaron para diseñar el modelo minable para la Deserción estudiantil en la Licenciatura en Computación.

#### *a) Fase de integración y recopilación.*

Las fuentes de datos con las que se trabajó fueron las siguientes:

- Reportes de la base de datos del estatus por período de la Licenciatura en Computación proporcionado por Control de Estudios de LUZ –NPF; pertenecientes al año 2008 hasta el 2011 con sus respectivos períodos.
- Cuestionario aplicado a los estudiantes de la Licenciatura en Computación que estén cursando materias en el primer, segundo y tercer semestre pertenecientes al primer período del año 2012
- A efectos de contrastar los resultados arrojados por el modelo predictivo se utilizó el cuestionario aplicado a los profesores de la Licenciatura en Computación que dictan clases entre los semestres I–V pertenecientes al primer período del año 2012.

En lo que respecta a la base de datos, Control de Estudios proporcionó 7 reportes de lista de matrícula por estatus, estos reportes pertenecen a los años 2008, 2009, 2010 y 2011 con sus respectivos períodos lectivos, la estructura del reporte contiene la descripción del estatus y el total de alumnos por cada estatus, entre las descripciones del estatus se encuentran: Inscripción suspendida, Inscrito, Fin del RR en el año, Desertor, Retirado, Egresado, Cambio de ubicación y Período.

#### *b) Fase de selección, limpieza y transformación*

Luego de ejecutar la fase de integración y recopilación se procedió a diseñar la estructura de un repositorio de datos, para gestionar los datos derivados de la investigación. Dado que la calidad del modelo predictivo está en función de la calidad de los datos minados (Hernández et. al, 2004), el siguiente paso en el desarrollo del modelo predictivo fue diseñar el repositorio de datos, para poder seleccionar y preparar los datos que se iban a minar, los cuales constituyen lo que se conoce como vista minable.

El manejador de base de datos utilizado para el diseño del repositorio fue MySQL, puesto que por ser de licencia libre no generó costos por el uso del mismo. Se procedió a cargar cada uno de los reportes proporcionados por Control de Estudios, de la lista de matrícula sólo se seleccionaron los estatus Desertor, Cambio de Ubicación y Retirado, para efectos del repositorio de datos los reportes se dividieron en 2 tablas, en la tabla 1: deserción\_ce se encuentran los campos Año, Periodo, Cantidad\_desertor y como clave primaria se estableció el campo Cod\_deserción. En la tabla 2: razón\_deserción, se estructuró con los campos cod\_razon y razón deserción.

Después de analizar los resultados obtenidos de los cuestionarios aplicados, se procedió a discriminar o eliminar los ítems cuyos resultados son irrelevantes o innecesarios para efectos de la fase de Minería de Datos, en el repositorio 4 ítems del cuestionario aplicado a los estudiantes, para obtener así un total de 6 tablas cuyo modelo Entidad Relación se presenta en la Figura 1.

Para el autor Storti (2007) el Modelo de Entidad Relación es un modelo de datos basado en una percepción del mundo real que consiste en un conjunto de objetos básicos llamados entidades y relaciones entre estos objetos, implementándose en forma gráfica a través del Diagrama Entidad Relación. En ese sentido, se muestra las diferentes relaciones que se dan en la base de datos denominada "Listado"; en la misma se aprecia la correspondencia entre las diferentes tablas que la conforman; después de este paso, se obtuvo una segunda vista minable.

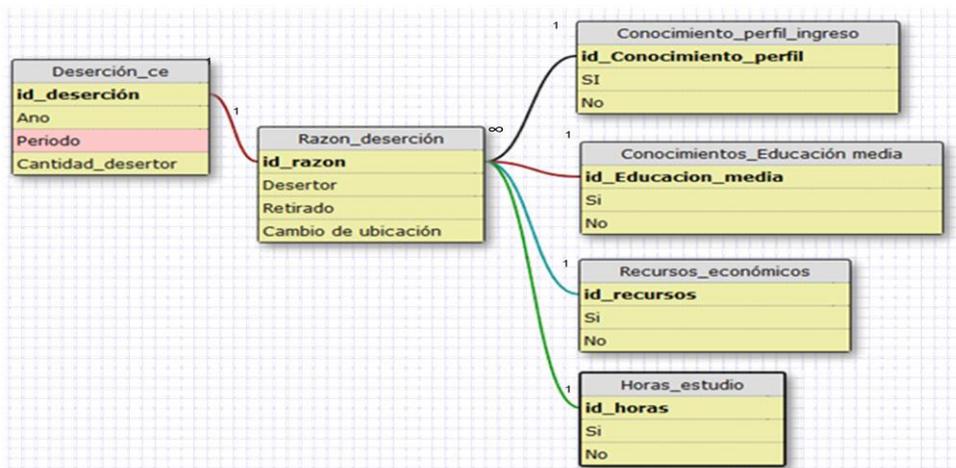


Figura 1. Modelo Entidad Relación del Modelo Propuesto. Las autoras, 2013.

*c) Fase de minería de datos*

Una vez obtenida la vista minable (base de datos cargada), se tuvo que elegir el análisis de datos requerido. Para lograr los objetivos planteados inicialmente, el modelo de minería de datos que se desarrolló es de tipo predictivo. Las técnicas de minería de datos seleccionadas fueron la de clasificación y clustering, utilizando un árbol de decisión mediante el algoritmo C4.5 y el método de aprendizaje basado en vecindad, conocido como técnica de los k vecinos más cercanos.

Además se empleó el software denominado Weka (versión 3.6). Se alimentó la vista minable mediante un archivo. Arff (Attribute-RelationFile Format), el cual fue creado mediante un archivo de texto sin formato. En consecuencia, se empleó para la creación de árboles de decisión que Weka ofrece, el algoritmo J48, el cual es la implementación de Weka para el algoritmo C4.5 mencionado anteriormente (Witten y Frank, 2005). Esto se logró mediante la división de los atributos propuestos en nuestra vista minable, usando la técnica divide y vencerás. Para su construcción, se crearon varios prototipos (árboles) en los que se podía identificar los atributos más representativos para la creación de los nodos.

Igualmente, un segundo prototipo se construyó utilizando la técnica del vecino más próximo (k nearestneighbor). De la misma manera que el modelo anterior, se utilizó la vista minable y se alimentó al minero (software) para generarlo. Para la construcción de este modelo se empleó el método use training set. Con este método se representa un número de vecinos más próximos a una instancia dada. A continuación en la figura 2, se presenta el árbol resultante en función de los datos introducidos al modelo minable procesado por Weka. Luego de cargar el archivo de relaciones en Weka, se procede a seleccionar la tarea de Classify (Clasificación) y a la técnica de árbol de decisión, en Weka esta técnica se representa con el nombre de J48 como se puede apreciar en la figura 2; tal como se evidencia en ella, se evalúan las categorías: Conocimiento del perfil de ingreso a la carrera, sexo, conocimientos de educación media, recursos económicos y actitud para el estudio; conectados por los nodos de decisión: sexo (femenino y masculino) y año de ingreso para ser evaluado en dos rangos ( $\leq 1990$  y  $> 1990$ ). En tal sentido el algoritmo realiza subdivisiones de las instancias válidas, y la instancia que más se repita se le asigna el valor más alto.

Realizada la tarea de clasificación, se procedió a seleccionar el clúster y la técnica de los k vecinos más cercanos; en Weka esta técnica se representa con el nombre de SimpleKMeans; en ella se dividen los atributos en clústeres, con el objetivo de determinar cuáles atributos tienen relaciones iguales (vecindad) y así establecer los atributos que más predominaron. De esta forma, se generó el modelo y se compararon los resultados obtenidos con las dos técnicas.

d) Pruebas y verificación de resultados

En esta fase se generaron los modelos con la ayuda del minero de datos (Weka). Se realizaron un conjunto de pruebas que se verificaron al momento de crear el modelo. Por otro lado, para la construcción del árbol de decisión, de las 44 instancias (registros) que formaba la primera vista minable, Weka tomó 30 instancias (66.6%) para construir el modelo y 14 instancias (33.4) para probarlo, con una precisión del 67.07%.

De igual manera, se probó el segundo modelo con la ejecución del algoritmo de los k vecinos más cercanos, utilizando el método de user training set con 10 evaluaciones, y se pudo notar que al establecer el valor de k en 10, se obtuvo una precisión del 64.07%, asemejándose al resultado del algoritmo C4.5. Estos porcentajes de confiabilidad señalan a estas dos técnicas como las más adecuadas para el desarrollo de la presente investigación.

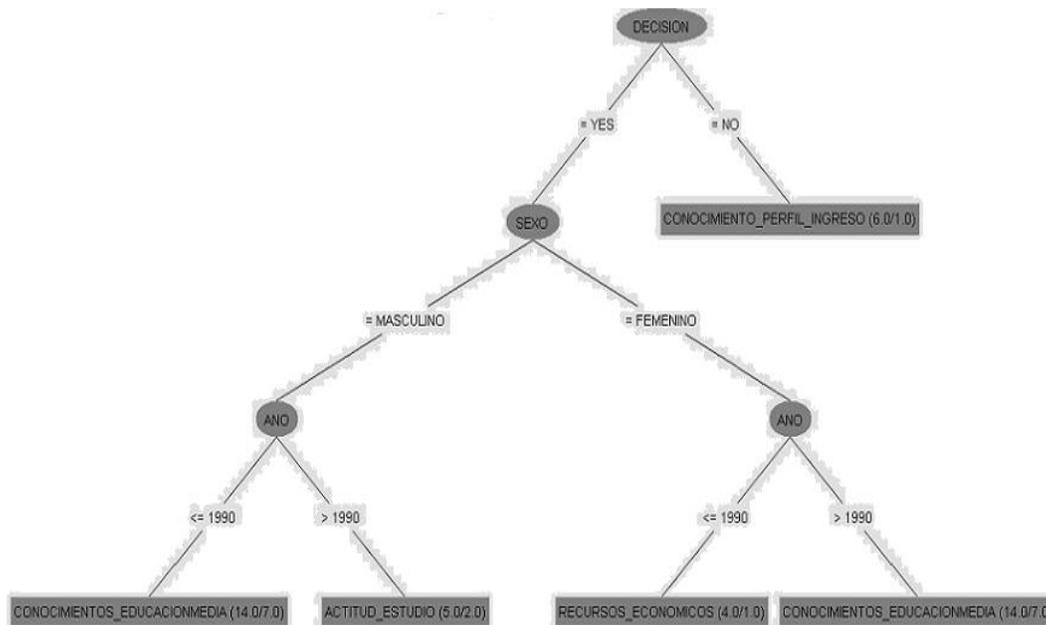


Figura 2. **Resultados del árbol de decisión.** Las autoras, 2013.

## **CONCLUSIONES Y RECOMENDACIONES**

La investigación mostró que los estudiantes de la Licenciatura en Computación desertan principalmente por las siguientes tres causas:

- Según los datos aportados por el estudio, se puede inferir que los conocimientos adquiridos por los estudiantes durante su permanencia en educación media no son suficientes para cursar estudios en la Licenciatura en Computación; principalmente por escasas competencias en matemática y lógica formal; aspectos importantes para los primeros semestres de la carrera.
- La falta de recursos económicos se constituye en un factor que incide en la deserción debido a que puede afectar el poder contar con equipos de computación para practicar fuera del recinto universitario.
- Falta de concentración en los estudios y las pocas horas dedicadas al mismo, se convierten en elementos que menoscaban el desempeño académico del estudiante y por ende se convierte en una causa de deserción.

Es importante destacar que la opinión de los profesores, coincide en que existe un factor de deserción estudiantil a parte de los ya mencionados, entre estos se destacó: El nivel de exigencia de las unidades curriculares que están en los semestres I, II y III.

Además el estudio permite afirmar, que las técnicas de Minería de Datos proporcionan una herramienta que permite establecer una aproximación a las causas que generan la deserción estudiantil. En tal sentido, el modelo predictivo planteado permitió arrojar resultados en cuanto a cuáles son las causas de deserción de los estudiantes de la Licenciatura en Computación, una vez que se reunieron, integraron, seleccionaron y limpiaron los datos y se creó un pequeño repositorio de datos para trabajar en la construcción y ejecución del modelo predictivo; resultados que coinciden con los arrojados luego del procesamiento de los cuestionarios aplicados a las unidades informantes.

Basándonos en la revisión documental realizada y en los resultados derivados de la investigación se invitó a la Coordinación del Programa de Ciencia y Tecnología de LUZ-NPF, a considerar los resultados presentados en esta investigación, debido a que no existen antecedentes de estudios previos, en cuanto a las causas de deserción de los estudiantes que

ingresan a la carrera, ya sea con el uso de herramientas como la desarrollada en la presente investigación, ni uso de métodos tradicionales. De este modo, el tipo de herramienta que aquí se emplea se convierte única en su área, para el contexto específico que aborda la investigación.

## REFERENCIAS

- Anuies (2007). *Retención y deserción en un grupo de Instituciones Mexicanas de Educación Superior*. Biblioteca de la educación superior. México.
- Britos P., Hossian A. (2005). *Minería de Datos*. Nueva Librería, Argentina, ISBN: 9871104308.
- Cabral, M. (2005). *Deserción en la educación superior pública en República Dominicana*. [Material en línea]. Disponible en: <http://www.psicologiacientifica.com/bv/psicologia-218-1-desercion-en-la-educacion-superior-publica-en-republica-domi.html> [Consulta: 2012, Diciembre 12].
- Chalabe, T; Pérez, L; Truninger, E. (2001). *Reflexiones sobre la deserción universitaria*. [Material en línea]. Disponible en: [http://rapes.unsl.edu.ar/Congresos\\_realizados/Congresos/V%20Encuentro%20-20Oct-2004/eje7/14.htm](http://rapes.unsl.edu.ar/Congresos_realizados/Congresos/V%20Encuentro%20-20Oct-2004/eje7/14.htm). [Consulta: 2013, Enero 15].
- Castellano, M. (2002). *Institutos Universitarios Tecnológicos y Colegios Universitarios en Venezuela*. [Material en línea]. Disponible: <http://www.uniparisdiderot.fr/comm/infodoc/Cdrom/Conferencias/>. [Consulta: 2012, Noviembre 28].
- Fuenmayor B., Perozo, S. y Rietveldt F. (2010). Percepción de los alumnos acerca de la Deserción Estudiantil en el Instituto Universitario de Tecnología de Cabimas. *Revista del Centro Regional de Investigación Humanística, Económica y Social* Año 13- N° 25. ISSN 1316-7790.
- González (2006). *Repetición y deserción universitaria en América Latina*. [Material en línea]. Disponible: <http://www.lpp-uerj.net/olped/documentos/1761.pdf>. [Consulta: 2012, Junio 10].

**Minería de datos aplicada a la deserción estudiantil. Caso: Licenciatura en Computación de la Universidad del Zulia-NPF.**  
**Yelitza J. Marcano, Rodney H. Rodríguez ( Pp. 31-51)**

---

Kamber, Han. (2006). *Data mining: concepts and techniques*. The Morgan Kaufmann Publishers, USA, ISBN: 1558609016.

Hernández J., Ferrari C. y Ramírez M. (2004). *Introducción a la minería de datos*. España. Pearson Educación.

Morales, E. (2009). *Descubrimiento de conocimiento en bases de datos*. [Material en línea] Disponible en: <http://ccc.inaoep.mx/~emorales/Cursos/KDD/principal.html> [Consulta: 2011, Diciembre 19].

Morazán F. (2011). *Deserción Escolar en Bachillerato Técnico en Computación*. Universidad Pedagógica Nacional. Caracas, Venezuela.

Sampieri, R.; Fernández, C.; Baptista, P. (2003). *Metodología de la Investigación*. Tercera Edición. Editorial Mc Graw Hill/Interamericana Editores, S.A. de C.V. México.

Sánchez G., Navarro W. y García A. (2005). *Factores de deserción estudiantil en la Universidad Surcolombiana*. Sede Nerva 2002-2005. Facultad de Educación.

Storti (2007). *Tecnología de la Información y la Comunicación*. Universidad Central de Venezuela. Venezuela.

Valero S. (2009). *Desarrollo de una herramienta de análisis de datos para predecir deserción escolar en la Universidad Tecnológica de Izúcar de Matamoros*. Tesis no publicada. Universidad Tecnológica de Izúcar de Matamoros. México.

Witten I. y Frank E. (2005). *Data mining, practical machine learning, tools and learning*. EUA: The Morgan Kaufmann Publishers.