

## Aplicación de Técnicas de Minería de Datos para Determinar las Interacciones de los Estudiantes en un Entorno Virtual de Aprendizaje

Angélica Jaramillo<sup>a</sup>, Henry Paz-Arias<sup>b</sup>

<sup>a</sup> Área de la Energía, Universidad Nacional de Loja, Ciudadela Universitaria Guillermo Falconí Espinosa La Argelia. Loja, Ecuador  
aejaramilloz@unl.edu.ec

<sup>b</sup> Facultad de Ingeniería en Sistemas, Escuela Politécnica Nacional, Ladrón de Guevara E11-25 y Andalucía.  
henry.paz@epn.edu.ec

**Resumen.** El presente artículo está enfocado en determinar las interacciones de los estudiantes del curso virtual de inglés de la Modalidad de Estudios a Distancia (MED) de la Universidad Nacional de Loja, para ello se realizó un análisis a la base de datos donde se encontraban los datos de los estudiantes correspondiente al periodo académico 2013 - 2014, para seleccionar los atributos necesarios que permitieron generar un modelo de minería de datos. De igual forma se consideró realizar un estudio de las diferentes técnicas de minería de datos donde se seleccionó la que más se adapta al proyecto, eligiendo la técnica de clasificación para generar los modelos a través de los algoritmos pertenecientes a la misma, posteriormente se efectuó un análisis de las metodologías de minería de datos comparando cada una de ellas con el fin de seleccionar la que ayude al desarrollo del proyecto eligiendo la metodología Crisp-dm porque contiene múltiples fases indicando cada una de las actividades que se deben cumplir, convirtiéndose de esta forma en una guía práctica. Además se desarrolló un análisis comparativo tomando en cuenta características de las herramientas de minería de datos donde se seleccionó RapidMiner para realizar los procesos mediante algunos algoritmos conjuntamente con los datos de los estudiantes los mismos que se dividieron en dos conjuntos, para entrenamiento y validación, obteniendo como resultado que el mejor algoritmo fue el decisión tree, ya que clasificó las instancias correctamente con un margen de error mínimo, así mismo presentó un árbol con los diferentes atributos dando las mejores reglas de las interacciones de los estudiantes, de tal forma se pudo generar el modelo mediante el cual se determinó que en gran mayoría los estudiantes tienen un nivel de interacción medio en el curso virtual de inglés, donde los factores que más influyen son las interacciones en las tareas, exámenes, recursos, situación laboral y estado civil del estudiante.

**Palabras Clave:** minería de datos, técnica de clasificación, modelo, algoritmo, metodología.

### 1 Introducción

La educación es la base del progreso de cualquier país, por ello en la actualidad los sistemas educativos de todo el mundo se enfrentan al desafío de utilizar las tecnologías de la información, teniendo un papel importante porque facilitan el aprendizaje en entornos virtuales prestando a los estudiantes la adquisición del

conocimiento en forma inmediata y amplia, sin que la distancia ni el tiempo sea un inconveniente en su formación académica [1].

Al respecto la Universidad Nacional de Loja cuenta con sistemas de información para brindar la facilidad de estudios a distancia ya sea en distintas carreras o cursos, estos sistemas almacenan grandes cantidades de información de los estudiantes como es el caso del curso virtual de inglés de la modalidad de estudios a distancia, el mismo que se ha tomado como objeto de estudio, pero tener numerosa información a disposición y no saber qué hacer con ella es un gran problema, es aquí donde interviene la minería de datos que contiene un conjunto de técnicas que se aplican para extraer conocimiento útil y comprensible, previamente desconocido, así mismo descubrir patrones para generar un modelo a través del análisis de la información de las interacciones en el curso, datos personales, institucionales y socioeconómicos del estudiante, que permitió determinar las interacciones de los estudiantes en el curso virtual, para que de esta manera ayude a la toma de decisiones, y por tanto un beneficio a la institución [2].

Crisp-dm es una metodología de las más usadas en la actualidad para la generación de proyectos de minería de datos, se obtuvo un modelo de análisis de datos, conjuntamente con la implementación de algoritmos de inteligencia artificial, ya incorporados en la herramienta de pre-procesamiento de datos RapidMiner [3].

El artículo está estructurado de la siguiente forma resumen, introducción, materiales y métodos, estado del arte, resultados alcanzados, conclusiones y referencias bibliográficas.

## 2 Materiales y Métodos

Para el desarrollo del artículo fue necesario el empleo de algunos métodos y metodología para ello se realizó una búsqueda de información bibliográfica y algunos casos de éxito.

En la recolección y organización de la información que se obtuvo se utilizó los siguientes métodos y técnicas:

**Estudio de Casos:** Sirvió para obtener un conocimiento más amplio de los casos reales actuales, los cuales ayudaron para tener una idea clara del problema, así mismo permitió realizar una exploración e investigación en profundidad de problemas específicos.

**Revisión bibliográfica:** Se sustentó la base teórica de la realización del proyecto, mediante consultas en fuentes bibliográficas, textos, artículos científicos, libros, tesis de grado y casos de éxito.

Dando como resultado el desarrollo del estado del arte en donde se menciona en que consiste la minería de datos, técnicas de minería de datos, herramientas que son utilizables dentro de la minería de datos.

**Metodología:** Se utilizó la metodología Crisp-dm que es muy importante porque contiene etapas las cuales están compuestas por actividades o una secuencia de pasos ordenados.

### 3 Estado del Arte

Dentro del estado del arte se han considerado temas relevantes concernientes acerca de la minería de datos.

#### 3.1 Minería de Datos

Proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos, para encontrar modelos a partir de los datos, para que este proceso sea efectivo, debería ser automático y el uso de los patrones descubiertos debería ayudar a la toma de decisiones, y por tanto, un beneficio a la organización, convertir datos en conocimiento [2], [3].

#### 3.2 Técnicas de Minería de Datos

Las tareas de la minería de datos son [2]:

**Técnicas supervisadas o predictivas:** Utilizar algunas variables o campos en una base de datos para predecir valores desconocidos o futuros de tal manera que especifican el modelo para los datos en base a un conocimiento previo [2].

**Técnicas no supervisadas o descriptivas:** Encontrar patrones que describan la información que puedan ser interpretadas.

Las técnicas de minería de datos serán utilizadas con el objetivo de obtener la información oculta en grandes cantidades de datos, las cuales son descritas a continuación [2].

**Agrupamiento:** Se agrupan datos dentro de un número de clases, se puede realizar mediante criterios de distancia o similitud, de forma que si las clases son similares entre sí estén agrupadas, la agrupación o clustering consiste en agrupar un conjunto de datos basándose en la similitud de los valores de sus atributos [6].

**Clasificación:** Árboles de decisión: Son estructuras que representan conjuntos de decisiones que generan reglas para la clasificación de un conjunto de datos. Entre los algoritmos que aplica es el J48, ID3, entre otros [7].

Son útiles para explorar un conjunto de datos y entender cómo ciertas variables de las interacciones de los estudiantes con el entorno virtual de aprendizaje inciden sobre otra [8].

Permite una organización eficiente de un conjunto de datos, debido a que los árboles son construidos a partir de la evaluación del primer nodo raíz y de acuerdo a su evaluación o valor tomado se va descendiendo en las ramas hasta llegar al final del camino u hojas del árbol [8].

**Reglas de asociación:** En minería de datos las reglas de asociación en base de datos se evalúan de acuerdo al soporte y a la confianza de las mismas, se utilizan para encontrar hechos que ocurren en común dentro de un conjunto de datos. Dicho de otra manera deben ocurrir ciertas condiciones para que se produzca cierta condición, también para buscar por medio de conjunto de datos reglas que revelan la naturaleza de las relaciones o asociaciones entre datos de las entidades [9], [10].

Se aplican en el análisis de la canasta de mercado, marketing cruzado con correo, diseño de catálogos, segmentación de clientes respecto a las compras y el soporte para la toma de decisiones [9].

### 3.3 Herramientas de Minería de Datos

**SAS Enterprise Miner:** Es una herramienta de minería de datos comercializada, crea modelos predictivos y descriptivos precisos sobre grandes volúmenes de datos a través de diferentes fuentes mediante un proceso transparente, lo que permite colaborar de manera más eficiente, incluye una interfaz de usuario intuitiva que incorpora los principios de diseño comunes establecidos para el software de SAS y herramientas de navegación adicionales para mover fácilmente alrededor del área de trabajo [11],[12].

**RapidMiner:** Es una herramienta de minería de datos desarrollado en Java, permite el desarrollo de procesos de análisis de datos mediante el encadenamiento de 500 operadores a través de un entorno gráfico, permite utilizar los algoritmos incluidos en weka, contiene técnicas de pre-procesamiento de datos, modelación predictiva y descriptiva, métodos de entrenamiento y prueba de modelos, visualización de datos, aprendizaje automático [9],[13].

**Weka:** Es una herramienta para el aprendizaje automático y minería de datos diseñado en Java, es de distribución de licencia GNU-GLP, contiene una colección de algoritmos para el análisis de datos y modelado predictivo, permite la visualización de datos, provee una interfaz gráfica [14].

Este programa se desarrolló en Java y dispone de tres entornos de trabajo gráficos y un entorno en modo consola, permitiendo la implementación de algoritmos para pre procesamiento de datos, clasificación, clustering, selección de atributos, reglas de asociación [15].

**Knime:** Está desarrollado sobre la plataforma Eclipse y programado en Java, su uso se basa en el diseño de un flujo de ejecución que plasme las distintas etapas de un proyecto de minería de datos y predecir posibles resultados [11], [16].

Es una plataforma de código abierto de fácil uso y comprensible para integración de datos, procesamiento, análisis, y exploración. Ofrece a los usuarios la capacidad de crear de forma visual flujos de datos, ejecutar selectivamente algunos o todos los pasos de análisis, y luego estudiar los resultados, modelos y vistas interactivas [16].

### 3.4 Metodología de Minería de Datos

Se ha considera el estudio de dos metodologías de minería de datos como lo es Semma, Crisp-dm, ya que estas son las más utilizadas en distintos proyectos.

#### 3.4.1 SEMMA

Creada por el SAS Institute, se define como “el proceso de selección, exploración y modelado de grandes volúmenes de datos para descubrir patrones de negocio desconocidos”. El nombre de esta terminología corresponde a las cinco fases básicas del proceso: sample (muestreo), explore (exploración), modify (modificación), model (modelado), assess (valoración) [17].

Se encuentra enfocada especialmente en aspectos técnicos, excluyendo actividades de análisis y comprensión del problema que se está abordando evidenciando que el modelo está orientado especialmente a aspectos técnicos [18].

#### 3.4.2 Crisp-dm

Es una metodología estándar para la construcción de proyectos de minería de datos con sus fases no necesariamente rígidas [19].

Puede ser integrada con una metodología de gestión de proyectos específica que complemente las tareas administrativas y técnicas, además es de libre distribución [19].

Organiza el desarrollo de un proyecto de minería de datos en una serie de fases o etapas que funcionan de manera cíclica e iterativa, cada una cuenta con tareas generales y específicas que permitan cumplir con los objetivos del proyecto [20], [21], [22].

##### 3.4.2.1 Fases de Crisp-dm

**Comprensión del negocio:** Comprender o definir el problema del negocio, lo cual es quizás el paso más importante de la metodología, permite entender los objetivos y requisitos que tendrá el proyecto [19].

Las tareas de esta fase es el establecimiento de los objetivos de negocio, evaluación de la situación mediante el inventario de recursos, requerimientos, suposiciones, restricciones, riesgos, contingencias, terminología, costes y beneficios,

establecimiento de los objetivos de minería de datos, generación del plan del proyecto y evaluación inicial de herramientas y técnicas [19], [20], [21].

**Comprensión de los datos:** Comprende la búsqueda de la información y de las variables que se utilizarán para la generación de los indicadores del proceso a los cuales se aplicara minería de datos, contiene algunas tareas como es la recolección de datos, teniendo claro desde qué lugar fueron obtenidos. Descripción de los datos, estableciendo los volúmenes de información con que se trabajará, la cantidad de registros, y los significados de cada campo o variable y los formatos en los que se encuentran. Exploración de los datos, indicando una estructura general de la información, comprobar frecuencia y distribución de los datos, verificación de la calidad de los datos, determinando la consistencia de los valores, comprobando la existencia de datos nulos y fuera de rango, identificando irregularidades para asegurar la completitud y exactitud de los datos [20],[21].

**Preparación de los datos:** Preparación de los datos para adaptarlos a las técnicas de minería de datos que se utilicen posteriormente, consta de algunas tareas como la selección de datos escogiendo un subconjunto de los datos recopilados en la etapa anterior. Limpieza de los datos, preparándolos para la fase de modelación, ya sea aplicando técnicas de normalización, discretización de campos numéricos, tratamiento de valores nulos, entre otros [20].

Estructuración de los datos con lo cual se pueden generar nuevos atributos a partir de los existentes o transformar valores de los atributos con que se cuenta. Integración de los datos, agrupar tablas o campos que se encuentren relacionadas, definiendo una estructura que las pueda contener. Formateo de los datos, transformar los datos sin modificar su significado, para que se puedan ajustar a las técnicas de minería de datos que se utilice [20], [22].

**Modelado:** Se elige las técnicas de modelado que sean más apropiadas para resolver el problema, aplica algunas tareas que son en base al objetivo principal del proyecto [19].

Generación del plan de prueba, diseñando un procedimiento para probar y validar el modelo. En general, se separa el conjunto de datos en dos: una parte de los datos destinada a entrenamiento del modelo y otra parte que será utilizada para las pruebas [21].

Construcción del modelo a partir de la técnica de modelado seleccionada, se aplica sobre el conjunto de datos para generar uno o más modelos. En este punto se van ajustando los parámetros de la técnica seleccionada de forma iterativa para obtener mejores resultados. Evaluación del modelo, interpretando los modelos en base al conocimiento existente y los criterios de éxito ya establecidos [19].

**Evaluación:** Se evalúa el modelo en base al cumplimiento de los criterios de éxito del problema, revisar el proceso seguido teniendo en cuenta los resultados obtenidos, para poder repetir algún proceso en el que a la vista del desarrollo posterior del proceso, se hayan podido cometer errores. Si el modelo generado es válido en función de los criterios de éxito establecidos en la primera fase [21].

Se evalúa el grado en el cual el modelo satisface los objetivos del negocio y busca determinar si hay alguna razón del negocio del porque el modelo sería deficiente [22].

## 4 Resultados Alcanzados

A lo largo del desarrollo del proyecto se ha considerado cada una de las etapas de la metodología llegado a los siguientes resultados.

### 4.1 Comparación de las Metodologías de Minería de Datos

Algunos modelos profundizan en mayor detalle sobre las tareas y actividades a ejecutar en cada etapa del proceso de minería de datos (como Crisp-dm), mientras que otros proveen sólo una guía general del trabajo a realizar en cada fase (como el proceso KDD o SEMMA) [16], [19].

SEMMA inicia el proyecto de minería a partir del conjunto de datos (la primera fase es el muestreo de los datos). Crisp-dm y KDD comienzan con un análisis del negocio y del problema organizacional. Catalyst considera cinco escenarios posibles como punto de partida, entre los cuales se encuentra el inicio desde un problema u oportunidad de negocio [16], [20].

KDD, Crisp-dm y Catalyst contemplan el análisis y comprensión del problema antes de comenzar el proceso de minería. SEMMA excluye esta actividad del modelo [17], [21].

En todos los modelos se contempla la selección y preparación de los datos esta situación se repite para la fase de modelado, donde se aplican las técnicas de minería para obtener los nuevos patrones [17].

La implementación de los resultados obtenidos es una fase que no está incluida en el modelo SEMMA. En Crisp-dm, se propone además una planificación para el control futuro y un análisis de cierre del proyecto [19], [20].

Los modelos Crisp-dm y Catalyst cuentan con un nivel de detalle con el que describen las tareas en cada fase del proceso, y porque incorporan actividades para la gestión del proyecto [19], [21].

La metodologías para la gestión de un proyecto de minería de datos, el modelo a tener en cuenta debería ser Crisp-dm ya que posee importantes características de las ya se hablado anteriormente [20], [22].

SEMMA y Crisp-dm comparten la misma esencia, estructurando el proyecto de explotación de datos en fases que se encuentran interrelacionadas entre sí [19].

SEMMA sólo es abierta en sus aspectos generales ya que está muy ligada a los productos SAS donde se encuentra implementada. Crisp-dm ha sido diseñada como una metodología neutra respecto a la herramienta que se utilice para el desarrollo del proyecto de explotación de datos siendo su distribución libre y gratuita [22].

#### **4.1.1 Elección de la Metodología**

La metodología a utilizar es Crisp-dm ya que cada una de sus fases se encuentra claramente estructurada definiendo de tal forma las actividades y tareas que se requieren para lograr el objetivo planteado es decir es la más completa entre las metodologías comparadas, es flexible por ende se puede hacer usos de cualquier herramienta de minería de datos.

#### **4.2 Desarrollo de la Metodología CRISP-DM**

En el presente artículo se describe la aplicación de técnicas de minería de datos conjuntamente con la herramienta RapidMiner para determinar las interacciones de los estudiantes en el curso virtual de inglés, para lo cual fue necesario aplicarlo en un escenario real con datos personales, institucionales, socioeconómicos y las interacciones de los estudiantes de la modalidad de estudios a distancia, para ello se empleó la metodología Crisp-dm como una guía que permita desarrollar el proyecto, a continuación cada una de sus fases:

##### **4.2.1 Compresión del Negocio**

Esta fase se centra en comprender los objetivos y los requerimientos del proyecto desde una perspectiva del negocio, y luego en convertir este conocimiento en la definición de un problema de minería de datos y en un plan preliminar designado para alcanzar los objetivos.

###### **4.2.1.1 Objetivos del Negocio**

Investigar sobre las diversas técnicas de minería de datos que permitan determinar la interacción de los estudiantes en un entorno virtual de aprendizaje.

Diseñar un modelo computacional aplicando técnicas de minería de datos para determinar la interacción de los estudiantes en un entorno virtual de aprendizaje.

Evaluar el modelo computacional en un escenario real a través de los datos de interacción de los estudiantes en un entorno virtual de aprendizaje.

###### **4.2.1.2 Requerimientos**

Tener la información suficiente de las interacciones de los estudiantes para la obtención del modelo.

Seleccionar técnicas de minería de datos adecuadas al problema a resolver.

Disponer de herramientas de minería de datos para la realización del modelo.

Contar con la colaboración continua por parte del director de tesis.

Tener apoyo del Director de la MED para obtener información exacta de los procesos que se utilizan en el curso virtual.

## 4.2.2 Comprensión de los Datos

En esta etapa se recolectó los datos relacionados con las interacciones de los estudiantes para una mejor comprensión de los mismos, de manera que es el primer acercamiento que se tiene para posteriormente realizar el análisis y de esta manera identificar algún inconveniente que exista, de tal forma que se analizó la estructura de los datos mediante consultas ejecutadas en la base de datos.

### 4.2.2.1 Recolección de Datos Iniciales

Los datos recolectados pertenecen a las interacciones de los estudiantes del curso virtual de inglés de la Modalidad de Estudios a Distancia perteneciente a la Universidad Nacional de Loja del periodo académico 2013 - 2014, para ello se descargó la información de la plataforma moodle perteneciente a la MED, entre la información con la que se trabajó se tiene las interacciones de los estudiantes según las actividades que desarrollaron en el curso, las mismas que se describen a continuación:

Interacción con los archivos compartidos con las temáticas del curso.

Realizaron tareas y evaluaciones, para la aprobación del curso.

Leer o imprimir los contenidos y actividades del curso.

Enviar las actividades al docente para su corrección y recibir sus calificaciones.

Evaluaciones On-Line y calificaciones.

De esta manera se ofrece a los estudiantes en formación la oportunidad de reforzar el aprendizaje brindando a través de contenidos y evaluaciones, para posibilitar la interacción estudiante-profesor y estudiante-herramientas.

Los datos se encuentran estructurados en archivos XML que consta de las interacciones, datos personales, institucionales y socioeconómicos de los estudiantes del curso, como son: número de accesos al curso, número de accesos a las tareas, número de veces que accede a un recurso, número de accesos a exámenes, descripción de los módulos, datos personales de los estudiantes, datos socioeconómicos e institucionales de los estudiantes.

En la siguiente figura (ver Figura 1) se puede observar las tablas que conforman la base de datos donde esta almacenada la información de las interacciones de cada uno de los estudiantes de la MED:

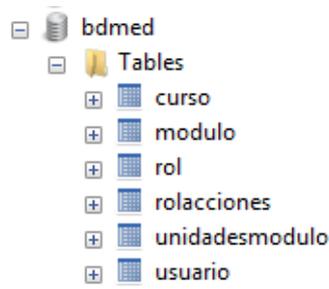


Fig. 1. Base de datos

### 4.2.3 Preparación de los Datos

La preparación de los datos se realizó para trabajar con las técnicas de minería de datos seleccionada, consta de algunas tareas como la selección de datos donde se eligió una estructura de datos recopilados en la etapa anterior, por otra parte se realizó la limpieza de los datos para poder generar el modelo de minería de datos, de tal manera que no debe contener valores nulos que permitieron obtener mejores resultados.

Además se realizó la estructura de los datos con lo cual se generaron nuevos atributos a partir de los existentes, así mismo se desarrolló la integración de los datos que consistió en agrupar tablas o campos que se encuentren relacionadas, finalmente se hizo el formateo de los datos que consiste en transformar los datos sin modificar su significado, para que se puedan ajustar a las técnicas de minería de datos.

Por lo tanto el objetivo de la presente etapa fue generar la estructura de datos final, con el propósito de obtener el modelo a través de técnicas de minería de datos.

#### 4.2.3.1 Estructura de los Datos

En esta tarea se realizó la construcción del data set final o estructura de datos que es útil para poder desarrollar el modelo computacional, donde se ha tomado en cuenta información personal, socioeconómicas y las interacciones de los estudiantes, para lo cual se trabajó con los siguientes campos para cumplir con el objetivo planteado, los mismos se pueden observar en la siguiente tabla (ver Tabla 1):

**Tabla 1.** Atributos de minería de datos para determinar las interacciones de los estudiantes.

Atributo	Tipo de Datos	Categorización
cedula	Nominal	
interaccionesrecurso	Int	- IRB - IRM - IRA
interaccionesexamen	Int	- IEB - IEM - IEA
interaccionestareas	Int	- ITB - ITM - ITA
numerointeracciones	Int	- Bajo - medio - alto
servicios	Nominal	- 1 - 2 - 3

ciudad	Nominal	- L - O
edad	Int	- a - b - c
genero	Nominal	- 0 - 1
estadocivil	Nominal	- S - C - D - V
carrera	Nominal	- Administración Pública - Artes Plásticas - Banca Y Finanzas - Bibliotecología e Información Científico - Comunicación Social - Contabilidad Y Auditoría - Cultura Física Y Deportes - Derecho - Diseño de Interiores Y Decoración - Economía - Educación Básica - Educación Musical - Enfermería - Físico Matemáticas - Informática Educativa - Ingeniería Agrícola - Ingeniería Agronómica - Ingeniería en Administración Y Producción - Ingeniería en Electrónica Y Telecomunicaciones - Ingeniería en Electromecánica - Ingeniería en Geología Ambiental - Ingeniería en Manejo Y Conservación - Ingeniería en Sistemas - Ingeniería Forestal - Laboratorio Clínico - Lengua Castellana Y Literatura - Medicina Humana - Medicina Veterinaria Y Zootecnia - Odontología - Psicología Clínica

		<ul style="list-style-type: none"> <li>- Psicología Educativa Y Orientación</li> <li>- Psicología Infantil Y Educación Parvularia</li> <li>- Psicorrehabilitacion Y Educación Especial</li> <li>- Químico Biológicas</li> <li>- Radiología e Imagen Diagnóstica</li> <li>- Trabajo Social</li> </ul>
numeroHijos	Nominal	<ul style="list-style-type: none"> <li>- Si</li> <li>- No</li> </ul>
trabajo	Nominal	<ul style="list-style-type: none"> <li>- Si</li> <li>- No</li> </ul>

#### 4.2.4 Modelado

En la presente etapa se realizó el modelado de minería de datos para ello se empleó el data set final de la etapa anterior (ver Tabla 1) conjuntamente con la técnica seleccionada los algoritmos que pertenecen a dicha técnica y finalmente la herramienta elegida de minería de datos para realizar los distintos procedimientos con la finalidad de obtener los resultados del modelo.

##### 4.2.4.1 Análisis de las Diversas Técnicas de Minería de Datos

Para el análisis de cada una de las técnicas de minería de datos se ha estudiado algunos casos de éxito donde estas han sido empleadas, obteniendo resultados confiables, además se realizó un análisis comparativo el cual se puede observar en la siguiente tabla (ver Tabla 2), cabe mencionar que dentro de cada una de las técnicas de minería de datos, estas constan con diferentes algoritmos que se pueden utilizar.

**Tabla 2.** Técnicas de minería de datos para determinar las interacciones de los estudiantes.

Técnica	Algoritmo	Casos de éxito
Agrupamiento O clustering	Simple-Kmeans	<ul style="list-style-type: none"> <li>- Aplicación de métodos de diseño centrado en el usuario y minería de datos para definir recomendaciones que promuevan el uso del foro en una experiencia virtual de aprendizaje [23].</li> </ul> <p>El estudio fue desarrollado con 30 estudiantes del primer ciclo de la carrera de Ingeniería en Sistemas Informáticos y Computación de la UTPL, con la técnica del clustering se procedió a agrupar a los</p>

		estudiantes del curso en diferentes grupos de acuerdo a su nivel de participación y semejanza de acceso a la plataforma, para descubrir patrones que reflejen comportamientos análogos en los estudiantes, con la ayuda del algoritmo SimpleKMeans se realizaron tres experimentaciones y también la construcción del modelo [23].
Clasificación	J48 REPTree ID3/C4.5	- Aplicación de técnicas de minería de datos para identificar patrones de comportamientos relacionados con las acciones del estudiante con el EVA de la UTPL [7].  Una de las técnicas que se empleará en la minería para determinar el comportamiento de los estudiantes en base a las acciones que éste realiza sobre el EVA, es la de clasificación, como son los árboles de decisión, reglas de decisión, éstos se utilizan para el indicador de la participación del estudiante en el curso ya que según lo analizado ayudan a predecir una o más variables discretas, basándose en otros atributos del conjunto de datos, el algoritmo hace predicciones. Los algoritmos utilizados fueron C4.5 o J48 y JRip para la experimentación [7].

#### 4.2.4.2 Determinar la Técnica de Minería de Datos

Para la selección de la técnica de minería de datos a utilizarla posteriormente se lo ha realizado mediante un cuadro comparativo que se lo puede observar en la Tabla 2, en donde se describe cada caso de éxito donde han alcanzado resultados idóneos, así mismo se ha podido identificar la técnica apropiada para el proyecto que es la clasificación para el análisis de las interacciones de los estudiantes en el entorno virtual de aprendizaje ya que es apropiada al problema a resolver y se tiene una comprensión de la misma.

Por otra parte se ha podido identificar una gran mayoría de casos de éxito relacionados con el estudio de entornos virtuales con diferentes fines en donde la técnica más aplicada es la antes mencionada, siendo esta eficiente en el momento de analizar grandes cantidades de datos y posteriormente el desarrollo de un modelo para que mediante este se pueda tomar decisiones y mejorar el uso de los entornos virtuales de aprendizaje.

#### 4.2.4.3 Cuadro Comparativo de Herramientas de Minería de Datos

Se realizó una recolección de información referente a las herramientas de minería de datos en los cuales sobresalen cada una de las características de las herramientas más utilizadas en proyectos similares según el estudio de algunos casos de éxito, entre las que se ha seleccionado se tiene SAS Enterprise Miner, RapidMiner, Weka, Knime, mismas que se puede observar a continuación:

**Tabla 3.** Herramientas de minería de datos [12], [13], [14], [15], [16], [17], [18].

CARACTERÍSTICAS	HERRAMIENTAS			
	SAS Enterprise Miner	RapidMiner	Weka	Knime
Licencia libre	X	✓	✓	✓
Multiplataforma	✓	✓	✓	✓
Puede combinar modelos	✓	✓	✓	X
Técnicas Descriptivas(agrupación)	✓	✓	✓	✓
Técnicas Predictivas(clasificación, otras)	✓	✓	✓	X
Interfaz amigable	✓	✓	X	X
Permite visualización de datos	✓	✓	✓	✓
Flexibilidad	X	✓	✓	✓
Fácil de Configurar	X	✓	✓	X
Fácil de Instalar	✓	✓	✓	✓
Conversión de datos	✓	✓	✓	X
Filtros	✓	✓	✓	X
Dispone de un módulo de integración con Weka y R	X	✓	X	X
Procesamiento de datos	✓	✓	✓	✓
Validación del modelo	✓	✓	✓	X

En la tabla anterior (ver Tabla 3.) se han mencionado algunas características que han sido tomadas en cuenta para la comparación de las mismas y posteriormente seleccionar según cumpla cada herramienta con dicha característica, donde se eligió la que más se adaptó al trabajo realizado.

#### 4.2.4.4 Selección de la Herramienta de Minería de Datos

Al realizar el análisis de las herramientas seleccionadas anteriormente (ver Tabla 3.) en base a algunas de sus características se pudo seleccionar a la herramienta RapidMiner para llevar a cabo cada una de las actividades acerca del modelado del presente proyecto, ya que se adapta al trabajo de titulación, posee una licencia libre, combinación de modelos, interfaz amigable, multiplataforma, empleo de técnicas, además permite aplicar varios algoritmos de minería de datos, con la integración de algunos complementos se puede utilizar los algoritmos incluidos en Weka, para poder generar el modelo y validarlo.

Así mismo se puede mencionar que no existe una herramienta que contenga todas las funcionalidades pero RapidMiner es ampliamente usada y probada a nivel internacional en aplicaciones empresariales, de gobierno y academia, posee gran cantidad de operadores que permiten generar el modelo para determinar las interacciones en entornos virtuales.

#### 4.2.4.5 Generar el Plan de Prueba

La generación del plan de pruebas consiste en probar la calidad y validez de los resultados obtenidos por el modelo, por ende es necesario generar un plan de pruebas mediante el cual se pueda probar la validez del modelo generado, para ello se trabajó con los datos pertenecientes a los estudiantes de la MED los cuales se los dividió en dos grupos uno para entrenamiento y el otro para emplearlo en la validación del modelo.

El conjunto de datos para entrenamiento es el 67% y el conjunto de datos restantes se los utilizó para realizar la validación de tal manera que da un resultado del 100% de datos utilizados para el modelado.

A continuación, se describe el plan de pruebas realizado con los diferentes algoritmos clasificados de la siguiente forma:

**Algoritmos de Reglas de decisión:** Los algoritmos utilizados dentro de esta clasificación corresponden a JRip, Ridor, Prism, K-NN, en donde se utilizó el 67% del conjunto de datos para entrenamiento y el 33% para la validación.

**Algoritmos de Árboles de decisión:** Los algoritmos utilizados dentro de esta clasificación corresponden a CHAID, Decision Tree, ID3, J48, en donde se utilizó el 67% del conjunto de datos para entrenamiento (E) y el 33% para la validación (V).

Entre los parámetros que se tomó en cuenta para evaluar los modelos generados son los siguientes: instancias clasificadas correctamente (accuracy), instancias clasificadas incorrectamente (classification\_error), estadística de Kappa que mide la coincidencia de la predicción con la clase real (Kappa), error cuadrático (squared\_error), error relativo (relative\_error), error absoluto (absolute\_error), presentando los resultados obtenidos en la siguiente tabla (ver Tabla 4 y Tabla 5):

**Tabla 4.** Resultados de Algoritmos.

Algoritmo	Datos	Instancias correctamente clasificadas (%)	Instancias incorrectamente clasificadas (%)	Índice de Kappa
DECISION TREE	E	87.71	12.29	0.73
	V	92.90	7.10	0.82
JRip	E	94.41	5.59	0.88
	V	92.63	7.37	0.82
RIDOR	E	89.66	10.34	0.77
	V	88.66	11.34	0.74
K-NN	E	98.74	1.26	0.97
	V	84.15	15.85	0.64
PRISM	E	98.46	1.54	0.97
	V	77.06	22.94	0.58
CHAID	E	91.06	8.94	0.81
	V	73.58	26.42	0.41
ID3	E	98.32	1.68	0.96
	V	82.10	17.90	0.59
J48	E	91.06	8.94	0.80
	V	91.49	8.51	0.15

**Tabla 5.** Resultados de Algoritmos.

Algoritmo	Datos	Error Cuadrático	Error Relativo (%)	Error Cuadrático Medio	Error Cuadrático Relativo
DECISION TREE	E	0.11	21.44	0.33	1.18
	V	0.07	13.33	0.25	4.29
JRip	E	0.05	10.18	0.23	0.80
	V	0.07	13.17	0.25	0.98
RIDOR	E	0.10	10.34	0.32	1.14
	V	0.11	11.34	0.31	1.23
K-NN	E	0.01	1.26	0.11	0.39
	V	0.16	15.85	0.39	1.51
PRISM	E	0.02	1.54	0.12	0.46
	V	0.26	26.06	0.50	2.37
CHAID	E	0.06	12.41	0.25	0.89

	V	0.17	27.35	0.41	5.46
ID3	E	0.01	1.97	0.09	0.39
	V	0.13	13.03	0.36	0.35
J48	E	0.08	16.07	0.29	1.01
	V	0.08	14.50	0.27	1.06

En las tablas anteriores (ver Tabla 4 y Tabla 5) se puede observar el resultado de cada algoritmo obtenido mediante la utilización de la herramienta RapidMiner conjuntamente con los datos de los estudiantes del curso virtual inglés de la MED, donde existe un porcentaje mínimo de error de clasificación en cada uno de los algoritmos, además se puede indicar que con el conjunto de entrenamiento de los datos la mayoría de los resultados obtenidos de los algoritmos son favorables es decir que sobrepasan el 90% de los datos han sido clasificados correctamente, los algoritmos que presentan mejores resultados se tiene el JRip 94.41%, K-NN 98.74%, Prism 98.46%, Chaid 91.06%, ID3 98.32 y el J48 91.06% . Así mismo con el conjunto de datos de validación los algoritmos que presentan los mejores resultados de datos clasificados correctamente se tiene el Decision Tree 92.90%, Jrip 92.63% y J48 91.49%.

Con el conjunto de datos utilizados en el entrenamiento el mejor algoritmo que presenta es el K-NN con un porcentaje de 98.74% es decir las instancias se han clasificado correctamente y un 1.26% de las instancias clasificadas incorrectamente, el índice de kappa 0.973, el error cuadrático 0.013, error relativo 1.26, error cuadrático medio 0.11 y finalmente el error cuadrático relativo con el 0.397; y con el conjunto de datos utilizada en la validación el algoritmo que ha arrojado el mejor resultado es el Decision Tree con el 92.90% que significa que dicho porcentaje es el de las instancias clasificadas correctamente, el 7.10% es el de las instancias clasificadas incorrectamente, el índice de kappa 0.823, el error cuadrático 0.068, error relativo 13.33, error cuadrático medio 0.253 y finalmente el error cuadrático relativo con el 4.285.

#### 4.2.5 Evaluación

En la presente etapa se realizó la evaluación del modelo para determinar si los datos obtenidos cumplen con el problema planteado, de tal forma que mediante la generación del modelo a través de los atributos (ver Tabla 1) se puede determinar las interacciones de los estudiantes del curso virtual de inglés de la modalidad de estudios a distancia.

En la siguiente figura se indican los resultados obtenidos de cada uno de los algoritmos ya sea tanto en el entrenamiento como en la validación en donde se muestra las instancias clasificadas correctamente (ver Fig. 2).

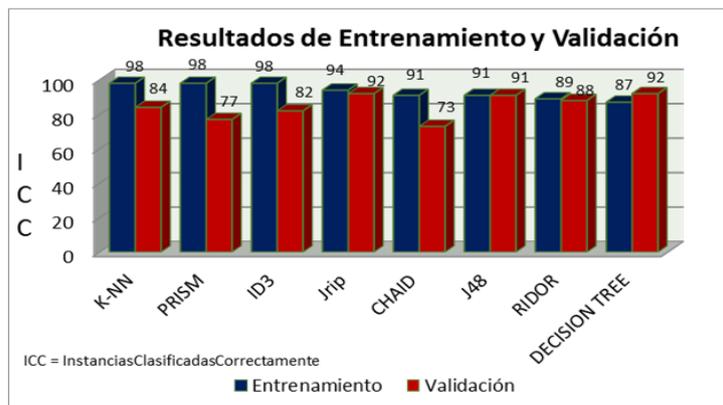


Fig. 2. Resultados por cada algoritmo

Así mismo en la siguiente figura se indican los resultados obtenidos de cada uno de los algoritmos en lo que respecta en la evaluación en donde se muestra las instancias clasificadas correctamente y las instancias clasificadas incorrectamente, donde se puede mencionar que el algoritmo que tiene mayor porcentaje es el Decision Tree con el 92,90% con un margen de error del 7,10%, el mismo se muestra a continuación (ver Fig. 3).



Fig. 3. Resultados de algoritmos de instancias clasificadas correcta e incorrectamente

Se realizó un análisis de los resultados obtenidos en la Minería de Datos, mediante una evaluación de los modelos obtenidos a través de la implementación de los algoritmos de la técnica de clasificación, los cuales fueron analizados en la fase anterior, para ello se utilizó datos de los estudiantes del curso virtual de inglés de la modalidad de estudios a distancia como son el número de accesos al curso, número de accesos a las tareas, número de veces que accede a un recurso, número de accesos a exámenes, datos personales de los estudiantes, datos socioeconómicos, datos institucionales.

#### 4.2.5.1 Determinar las Interacciones de los Estudiantes del Curso Virtual de Inglés Mediante Técnicas de Minería de Datos

Para poder determinar las interacciones de los estudiantes, se tomó en cuenta el mejor resultado de los algoritmos utilizados, que fueron analizados en la evaluación del modelo (Tabla 4 y Tabla 5 y Fig. 3) obteniendo el mejor resultado el algoritmo Decision Tree el cual presenta una buena clasificación con un 92.9% y menor margen de error del 7.1% en la validación del modelo.

Mediante el algoritmo Decision Tree se pudo determinar el nivel de interacción de los estudiantes conjuntamente con la utilización de cada uno de los atributos seleccionados que conforman el data set final (ver Tabla 1) de tal manera se obtuvieron los siguientes resultados durante la fase de entrenamiento del algoritmo el nivel de interacción es 19 altas, 438 medias y 190 bajas, así mismo en la fase de validación 7 altas, 239 medias y 86 bajas, los resultados finales del nivel de interacciones es 26 altas, 677 medias y 276 bajas (ver Fig. 4).

Según con los resultados obtenidos se pudo determinar que 677 corresponden al nivel de interacción medias de los estudiantes en el curso virtual de inglés equivalente a un porcentaje del 69%, el nivel de interacción media significa que el estudiante utiliza los recursos conscientemente teniendo definido sus intereses en cuanto al material.

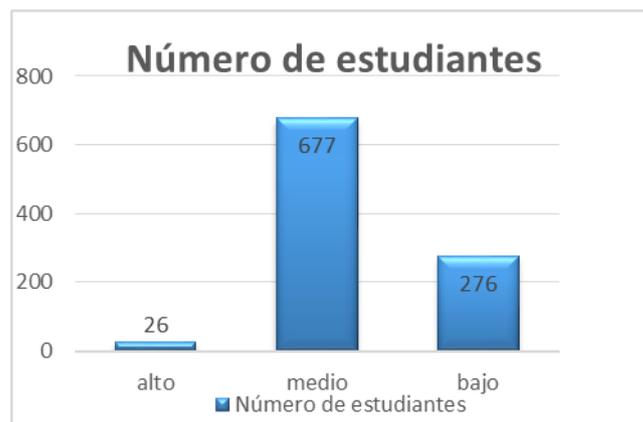


Fig. 4. Interacciones de los estudiantes

Además de obtener los resultados mediante el algoritmo Decision Tree se pudo evidenciar que de los 1069 datos de los estudiantes obtenidos para generar el modelo,

90 estudiantes es decir el 9% de los datos no han sido clasificados en ningún tipo del nivel de interacción (ver Fig. 5).



**Fig. 5.** Clasificación de los datos

#### 4.2.5.2 Reglas Obtenidas Mediante los Algoritmos de Minería de Datos

##### Nivel alto de interacción en el curso virtual

Cuando los estudiantes tienen interacciones altas con los exámenes y los recursos y tiene entre 25 y 29 años y es soltero, entonces las interacciones en el curso virtual son altas.

Si las interacciones con los recursos es medio y tiene entre 25 y 29 años y el género es femenino, entonces las interacciones en el curso virtual es alto.

Si las interacciones con los exámenes es medio y es mayor a 29 años y no trabaja, entonces las interacciones en el curso virtual es alto.

Si las interacciones con los exámenes y los recursos es medio y pertenece a otra ciudad y no tiene hijos, entonces las interacciones en el curso virtual es alto.

Si las interacciones con los exámenes es medio y las interacciones con los recursos y las tareas es alto y el género es femenino y no trabaja y pertenece a otra ciudad y es mayor a 29 años, entonces las interacciones en el curso virtual es alto.

Las interacciones con los exámenes y con los recursos es alto y no posee ningún tipo de servicio y no tiene hijos y tiene entre 25 y 29 años, entonces las interacciones en el curso virtual es alto.

##### Nivel medio de interacción en el curso virtual

Las interacciones en los exámenes es media y pertenece a otra ciudad y no tiene hijos y es casado, entonces las interacciones en el curso virtual es media.

Si las interacciones con los exámenes es medio y las interacciones con los recursos y las tareas es alto y tiene todos los servicios, no trabaja, el género es masculino, pertenece a otra ciudad y es mayor a 25 años, entonces las interacciones en el curso virtual es medio.

Si las interacciones con los recursos es alto y las interacciones con las tareas es medio y el género es masculino y posee todos los servicios y es mayor a 29 años, entonces las interacciones en el curso virtual es medio.

El estudiante no trabaja y el género es masculino y pertenece a otra ciudad y es soltero y solo posee un servicio que es número de celular, entonces las interacciones en el curso virtual es medio.

El estudiante no trabaja y el género es femenino y posee todos los servicios y pertenece a la ciudad de Loja y es soltero y es menor a 25 años y las interacciones con los recursos es medio, entonces las interacciones en el curso virtual es medio.

Las interacciones con los exámenes es bajo y las interacciones con las tareas es bajo y las interacciones con los recursos es medio y es casado, entonces las interacciones en el curso virtual es medio.

Las interacciones con los exámenes y con las tareas y con los recursos es bajo y posee los servicios y tiene hijos y es mayor a 29 años y trabaja y el género es masculino y pertenece a otra ciudad y es casado, entonces las interacciones en el curso virtual es medio.

Las interacciones con los exámenes es medio y las interacciones con las tareas y recursos es baja y tiene todos los servicios y es soltero, entonces las interacciones en el curso virtual es medio.

#### **Nivel bajo de interacción en el curso virtual**

Si las interacciones con los exámenes y los recursos o tareas son bajas y trabaja, entonces las interacciones en el curso virtual es bajo.

Si las interacciones con los exámenes y los recursos son bajas y el género es femenino y tiene hijos, entonces las interacciones en el curso virtual es bajo.

Si las interacciones con los exámenes y los recursos son bajas y tiene hijos y es menor a 25 años, entonces las interacciones en el curso virtual es bajo.

Si las interacciones con los exámenes, las tareas y los recursos son bajas y pertenece a la ciudad de Loja y no trabaja y posee todos los servicios y el género es masculino y es mayor a 29 años y tiene hijos, entonces las interacciones en el curso virtual es bajo.

El estudiante no trabaja y el género es masculino y pertenece a la ciudad de Loja y posee un servicio y es menor a 25 años y las interacciones con los recursos y las tareas es baja, entonces las interacciones en el curso virtual es bajo.

El estudiante trabaja y el género es masculino y pertenece a la ciudad de Loja y posee todos los servicios y es soltero y es menor a 25 años y las interacciones con los recursos y las tareas es baja y tiene hijos, entonces las interacciones en el curso virtual es bajo.

El estudiante trabaja y el género es femenino y posee todos los servicios y pertenece a otra ciudad y es casado y las interacciones con los exámenes es bajo, entonces las interacciones en el curso virtual es bajo.

Las interacciones con los exámenes y con las tareas y con los recursos es bajo y posee los servicios y tiene hijos y es mayor a 29 años y trabaja y el género es femenino y pertenece a la ciudad de Loja y es viudo, entonces las interacciones en el curso virtual es bajo.

#### 4.2.5.3 Factores Para Determinar las Interacciones de los Estudiantes

Los factores que influyeron en la realización del modelo se encuentran asociados entre sí, los cuales son datos: personales, institucionales, socioeconómicos e interacciones del estudiante los cuales se detallan a continuación:

**Interacciones en el curso:** interaccionestareas (número de accesos a las tareas), interaccionesrecurso (número de veces que accede a un recurso), interaccionesexamenes (número de accesos a exámenes).

**Datos personales de los estudiantes:** género, estadocivil, edad, servicios (teléfono, celular), ciudad (estudiantes que residen en Loja o en otra ciudad del país).

**Datos socioeconómicos de los estudiantes:** numeroHijos, trabajo (si el estudiante trabaja o no).

**Datos institucionales de los estudiantes:** carrera (a que carrera pertenece el estudiante).

A continuación se presenta cada uno de los atributos con sus respectivos pesos según los resultados obtenidos mediante el algoritmo Decision Tree de tal forma que se pueda determinar el que más influye en el modelo (ver Tabla 4).

**Tabla 6.** Porcentaje de los factores, atributos con sus respectivos pesos.

Atributo	Porcentaje del atributo (%)
Interacciones tareas	12.196
Interacciones recurso	10.946
Interacciones exámenes	13.299
Genero	4.562
Estado civil	9.509
Edad	8.671
Servicios	8.299
Carrera	9.137
Número hijos	5.346
Trabajo	8.126
Ciudad	9.908

Luego de obtener el peso de cada uno de los atributos pertenecientes a los datos de los estudiantes, los que más inciden en el objetivo principal del presente proyecto es determinar el nivel de interacción de los estudiantes del curso virtual inglés de la

MED son las interacciones en las tareas con un 12%, en los recursos con el 11% y en los exámenes el 13% como se puede observar en la siguiente figura (ver Fig. 6).

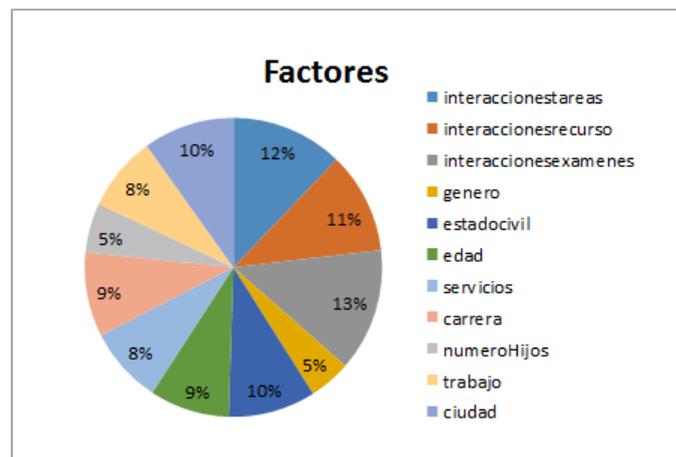


Fig. 6. Factores

#### 4.2.5.4 Análisis de los Resultados

La Universidad Nacional de Loja a través de la MED brinda cursos virtuales de aprendizaje entre estos se tiene el curso de inglés, el mismo que fue tomado como objeto de estudio para el presente proyecto, ya que este idioma es fundamental porque se lo utiliza como un medio de comunicación, así mismo para poder obtener becas e ir a otro país a especializarse, para laborar, además la mayoría de la información que se encuentra disponible ya sea en digital o impresa está en inglés, entre otras utilidades, es por ello que se realizó el análisis para determinar la interacción de los estudiantes en este curso virtual, tomando dos conjuntos de datos los cuales estuvieron conformados por 147 alumnos obteniendo un total de 32029 interacciones que se encuentran distribuidas en tareas, recursos y exámenes, también se obtuvo datos personales como edad, estado civil, servicios, género, dirección, de cada uno de los estudiantes, los cuales conformaron el data set.

Con estos datos se realizó el análisis mediante la minería de datos donde se obtuvo algunos inconvenientes que los resultados obtenidos de los modelos fueron muy bajos es decir no cumplían con un rango de aceptación, motivo por el cual fue necesario la incrementación de nuevos datos, mismos que fueron proporcionados por la unidad de telecomunicaciones, ya una vez que se contaba con 1069 datos se procedió a integrar nuevos atributos de estos estudiantes como son datos socioeconómicos (tienen hijos o no, situación laboral), datos institucionales (carrera a la que pertenecen), con estos parámetros se realizó nuevamente el análisis conjuntamente con la técnica de minería de datos y los algoritmos seleccionados obteniendo un mejoramiento en los resultados, porque el mayor porcentaje que arrojaron los algoritmo con los datos

anteriores fue del 77%, mientras que con los nuevos datos y atributos incorporados los resultados mejoraron obteniendo los porcentajes mayores al 90%, por ende mediante el modelo obtenido los resultados son confiables.

Consecuentemente se pudo determinar que 26 estudiantes pertenecen a la clasificación de la interacción alta, 677 estudiantes dentro de las interacciones medias y 276 estudiantes en interacciones bajas, estos resultados se obtuvieron a través de las reglas generadas por los algoritmos que conforman el modelo, presentando diferentes situaciones para cada nivel de interacción, las cuales se presentan a continuación:

El nivel de interacción de los estudiantes es alta en el curso virtual de inglés, cuando las interacciones en los exámenes y los recursos son altas, su edad es mayor a 25 años, su estado civil es soltero, no trabaja, no tienen hijos, pertenecen a otra ciudad y su género es femenino.

El nivel de interacción de los estudiantes es media en el curso virtual de inglés, cuando las interacciones en los exámenes, tareas y los recursos son medias, pertenecen a cualquier ciudad, puede tener cualquier edad, su estado civil es soltero o casado, puede poseer un trabajo o no, puede poseer todos los servicios o uno de ellos, puede tener hijos o no, puede ser masculino o femenino.

El nivel de interacción de los estudiantes es bajo en el curso virtual de inglés, cuando las interacciones en los exámenes, tareas y los recursos son bajas, pertenecen a la ciudad de Loja, es mayor a 29 años, su estado civil es casado, posee un trabajo, puede poseer todos los servicios o uno de ellos, tiene hijos, puede ser masculino o femenino.

Cabe indicar que las interacciones en las tareas se refiere a la revisión y envío de tareas, consulta de las sugerencias de las actividades que contiene las tareas; en las interacciones de los exámenes contiene revisión de los horarios de los que se puede rendir el examen, además saber la calificación y soluciones a las preguntas incorrectas que se han cometido y las interacciones de recursos se refiere cuando el estudiante descarga el material de apoyo que contiene cada unidad que se va a revisar en el curso.

Por lo tanto se puede indicar que las interacciones de los estudiantes que mayor prevalecen es el nivel medio con un porcentaje del 69% y el nivel bajo con el porcentaje del 25% en el curso virtual de inglés, significando que el estudiante utiliza los recursos conscientemente teniendo definido sus intereses en cuanto al material, es decir que no acceden con frecuencia al curso, por lo tanto realizan pocas consultas a las tareas, recursos y exámenes que deben cumplir.

## 5 Conclusiones

A través de los resultados obtenidos se ha podido concluir lo siguiente:

La minería de datos es muy importante dentro del campo de la educación ya que ayudó a extraer información que se encuentra oculta en los datos de tal forma permitió el análisis y la generación de nuevo conocimiento para poder determinar en nivel de interacción de los estudiantes.

RapidMiner es una herramienta de minería de datos potente ya que contiene complementos que permite hacer uso de diferentes algoritmos tanto de esta

herramienta como de otras herramientas, además tiene operadores que ayudan a facilitar el desarrollo de los procesos para crear los modelos aplicables para el análisis de los datos.

Para determinar el nivel de interacción en el curso de inglés se aplicó diferentes algoritmos de clasificación, presentando los mejores resultados el Decision Tree, ya que este algoritmo obtuvo el menor margen de error durante la clasificación de los datos de las interacciones en el curso (tareas, exámenes, recursos), datos personales, institucionales y socioeconómicos.

Mediante el modelo de minería de datos obtenido se pudo determinar que las interacciones de los estudiantes en el curso virtual de inglés que mayor prevalece es el nivel medio con un porcentaje del 69% y los factores que más influyeron en el modelo fueron las interacciones en los exámenes, tareas, recursos, el estado civil y la situación laboral del estudiante.

## Agradecimientos

Este trabajo ha sido auto-financiado por la autora Angélica Elizabeth Jaramillo Zhingre y contando con el apoyo de la carrera de Ingeniería en Sistemas de la Universidad Nacional de Loja. El presente artículo forma parte del trabajo de titulación denominado “Aplicación de técnicas de minería de datos para determinar las interacciones de los estudiantes en un entorno virtual de aprendizaje”.

La autora desea expresar su agradecimiento particular a la carrera de Ingeniería en Sistemas de la Universidad Nacional de Loja, y al Ing. Henry Paz por la tutoría prestada a lo largo del desarrollo del trabajo de titulación.

## Referencias

1. Gómez, LM, Macedo, JC: Importancia de las TIC en la Educación Básica Regular, Universidad Nacional Mayor de San Marcos – Facultad de Educación.
2. Minería de Datos, Universidad de Extremadura - Campus Libre y Abierto, [http://cala.unex.es/cala/epistemowikia/index.php?title=Miner%C3%ADa\\_de\\_Datos](http://cala.unex.es/cala/epistemowikia/index.php?title=Miner%C3%ADa_de_Datos)
3. Ordoñez, KF.: Aplicación de técnicas de minería de datos para predecir la deserción de los estudiantes de primer ciclo de la Modalidad Abierta y a Distancia de la UTPL, Universidad Técnica Particular de Loja – Área Técnica, <http://dspace.utpl.edu.ec/bitstream/123456789/7897/1/Ordonez%20Brice%C3%B1o%20Karla-%20Informatica.pdf>
4. Carmen, M., Galán, SJ: Definición de Minería de Datos, Universidad Carlos III de Madrid, En línea: [http://www.oocities.org/es/mineria.datos/definicion\\_tecnicas\\_mineria\\_datos.pdf](http://www.oocities.org/es/mineria.datos/definicion_tecnicas_mineria_datos.pdf)
5. Instituto de Investigación en Inteligencia Artificial, Minería de Datos o Data Mining, Consejo Superior de Investigaciones Científicas - Instituto de Investigación en Inteligencia Artificial, <http://www.iiia.csic.es/udt/files/DataMining.pdf>
6. Romero, C., Ventura, S., Hervás, C.: Escuela Politécnica Superior Universidad de Córdoba, Estado actual de la aplicación de la minería de datos a los sistemas de enseñanza basada en web, [http://www.investigacion.frc.utn.edu.ar/labsis/Publicaciones/congresos\\_labsis/cynthia/CICA\\_2009\\_Aplicacion\\_Mineria\\_Datos\\_basada\\_ense%F1anza\\_web.pdf](http://www.investigacion.frc.utn.edu.ar/labsis/Publicaciones/congresos_labsis/cynthia/CICA_2009_Aplicacion_Mineria_Datos_basada_ense%F1anza_web.pdf)

7. Sarango, MY.; Aplicación de técnicas de minería de datos para identificar patrones de comportamientos relacionados con las acciones del estudiante con el EVA de la UTPL, Universidad Técnica Particular de Loja – Escuela de Ciencias de la Computación, <http://dspace.utpl.edu.ec/bitstream/123456789/2387/1/MarciaSarangoTsis.pdf>
8. Solarte, GR, Soto, JA: Árboles de decisiones en el diagnóstico de enfermedades cardiovasculares, Universidad Tecnológica de Pereira, <http://revistas.utp.edu.co/index.php/revistaciencia/article/viewFile/1487/947>
9. García, JA, Acevedo AM: Análisis para predicción de ventas utilizando minería de datos en almacenes de ventas de grandes superficies, Universidad Tecnológica de Pereira - Facultad de ingenierías: eléctrica, electrónica, física y Ciencias de la computación - Ingeniería de sistemas y computación, <http://repositorio.utp.edu.co/dspace/bitstream/11059/1339/1/006312G216.pdf>
10. Velandia, RA, Hernández, FL: Evaluación de Algoritmos de extracción de reglas de decisión para el diagnóstico de huecos de tensión, Universidad Industrial de Santander, <http://tangara.uis.edu.co/biblioweb/tesis/2010/134742.pdf>
11. Cubero, JC, Berzal, F.: Herramientas de Minería de Datos, Universidad de Granada – Departamento de Ciencias de la Computación e Inteligencia Artificial, <http://elvex.ugr.es/decsai/intelligent/workbook/D1%20KNIME.pdf>
12. Qualex Consulting Services, SAS Enterprise Miner, [http://www.qlx.com/Software\\_Sales/enterprise\\_miner.html](http://www.qlx.com/Software_Sales/enterprise_miner.html)
13. Beltrán, D. Poveda, D.: RAPIDMINER, Universidad Nacional de Colombia- Facultad de Ciencias Económicas - Unidad de Informática y Comunicaciones, [http://www.fce.unal.edu.co/uifce/pdf/Rapid\\_Miner.pdf](http://www.fce.unal.edu.co/uifce/pdf/Rapid_Miner.pdf)
14. WEKA: Software de minería de datos en JAVA, Universidad de Waikato, <http://www.cs.waikato.ac.nz/ml/weka/>
15. García FJ: Aplicación de Técnicas de minería de datos a datos obtenidos por el centro de Andalucía de Medio Ambiente, Universidad de Granada – Master Universitario en Estadística Aplicada, [http://masteres.ugr.es/moea/pages/tfm-1213/tfm\\_garciagonzalezfrancisco\\_1/](http://masteres.ugr.es/moea/pages/tfm-1213/tfm_garciagonzalezfrancisco_1/)
16. KNIME, KNIME, <http://www.knime.org/>
17. Moine, JM, Haedo, AS, Gordillo, S.: Estudio comparativo de metodologías para minería de datos, Universidad Nacional de La Plata - Facultad de Informática, [http://sedici.unlp.edu.ar/bitstream/handle/10915/20034/Documento\\_completo.pdf%3Fsequence%3D1](http://sedici.unlp.edu.ar/bitstream/handle/10915/20034/Documento_completo.pdf%3Fsequence%3D1)
18. Moine, JM, Haedo, AS, Gordillo, S.: Análisis comparativo de metodologías para la gestión de proyectos de minería de datos, Universidad Nacional de La Plata - Facultad de Informática, [http://sedici.unlp.edu.ar/bitstream/handle/10915/18749/Documento\\_completo.pdf%3Fsequence%3D1](http://sedici.unlp.edu.ar/bitstream/handle/10915/18749/Documento_completo.pdf%3Fsequence%3D1)
19. Gironés, J.: Metodologías y estándares, Universidad Abierta de Cataluña, [http://www.exabyteinformatica.com/uoc/Administracio\\_i\\_direccio\\_dempreses/Business\\_analytics/Business\\_analytics\\_\(Modulo\\_3\).pdf](http://www.exabyteinformatica.com/uoc/Administracio_i_direccio_dempreses/Business_analytics/Business_analytics_(Modulo_3).pdf)
20. Herrera, MA, Acosta, JD: Estudio sobre el estado de las soluciones ict y de los casos prácticos de aplicación de la minería de datos a nivel mundial en al menos 5 casos representativos, Universidad EAFIT, [https://repository.eafit.edu.co/bitstream/handle/10784/2457/AcostaVasquez\\_JuanDavid\\_2006.pdf?sequence=1&isAllowed=y](https://repository.eafit.edu.co/bitstream/handle/10784/2457/AcostaVasquez_JuanDavid_2006.pdf?sequence=1&isAllowed=y)
21. Martínez, CÁ: Aplicación de técnicas de minería de datos para mejorar el proceso de Control de Gestión en Entel, universidad de chile - Facultad de Ciencias Físicas y Matemáticas, [http://www.tesis.uchile.cl/bitstream/handle/2250/112065/cf-martinez\\_ca.pdf?sequence=1](http://www.tesis.uchile.cl/bitstream/handle/2250/112065/cf-martinez_ca.pdf?sequence=1)
22. Fischer, ES: Modelo para la Automatización del Proceso de Determinación de Riesgo de Deserción de Estudiantes Universitarios, Universidad de Chile - Facultad de Ciencias

Físicas y Matemáticas,  
<http://preu.unillanos.edu.co/sites/default/files/fields/documentos/PREDICION%20DESERC I.pdf>

23. Valdivieso, PM: Aplicación de métodos de diseño centrado en el usuario y minería de datos para definir recomendaciones que promuevan el uso del foro en una experiencia virtual de aprendizaje, <http://www.redalyc.org/pdf/3314/331427213010.pdf>