# UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA

*La Universidad Católica de Loja*

## ÁREA TÉCNICA

TITULO DE INGENIERO EN SISTEMAS INFORMÁTICOS Y COMPUTACIÓN

Aplicación de Técnicas de Minería de datos para determinar patrones de comportamiento en las actividades realizadas en el EVA por parte de los estudiantes de Modalidad a Distancia.

TRABAJO DE TITULACIÓN.

**AUTORA**: Betancourt Granillo, Lizzette Gabriela

**DIRECTOR**: Riofrío Calderón, Guido Eduardo, Mg.

LOJA –ECUADOR

2016



*Septiembre, 2016*

## APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN

Magister.
Guido Eduardo Riofrío Calderón
**DOCENTE DE LA TITULACIÓN**

De mi consideración:

En el presente trabajo de titulación: Aplicación de Técnicas de Minería de datos para determinar patrones de comportamiento en las actividades realizadas en el EVA por parte de los estudiantes de Modalidad a Distancia realizado por Betancourt Granillo Lizzette Gabriela, ha sido orientado y revisado durante su ejecución, por cuanto se aprueba la presentación del mismo.

Loja, Noviembre de 2016

f) . . . . . .

## DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

"Yo Betancourt Granillo Lizzette Gabriela declaro ser autor (a) del presente trabajo de titulación: Aplicación de Técnicas de Minería de datos para determinar patrones de comportamiento en las actividades realizadas en el EVA por parte de los estudiantes de Modalidad a Distancia, de la Titulación Sistemas Informáticos y Computación, siendo Guido Eduardo Riofrío Calderón director (a) del presente trabajo; y eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones legales. Además, certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

Adicionalmente declaro conocer y aceptar la disposición del Art. 88 del Estatuto Orgánico de la Universidad Técnica Particular de Loja que en su parte pertinente textualmente dice: "Forman parte del patrimonio de la Universidad la propiedad intelectual de investigaciones, trabajos científicos o técnicos y tesis de grado que se realicen a través, o con el apoyo financiero, académico o institucional (operativo) de la Universidad".

f.....

Autora: Betancourt Granillo Lizzette Gabriela

Cédula: 1104968415

## DEDICATORIA

A Dios y a la Virgencita del Cisne, quienes me dieron la sabiduría y la fe para continuar y poder superar los problemas que se me presentaron, por bendecirme y permitirme llegar hasta estas instancias.

A mis padres, Luis y Etelvira, quienes son el pilar fundamental de mi vida, que con sus consejos, comprensión, ánimos y apoyo incondicional hicieron que me motive aún más para cumplir esta meta, porque sin su apoyo no lo hubiera logrado.

A mis hermanos Bladimir, Joffre y Jhuliana, por estar en todo momento apoyándome.

A mis sobrinos, Mili y Jericito, por sacarme una sonrisa, en esos momentos que no daba más.

A Robert, por su apoyo incondicional durante mi vida universitaria.

A mis amigos con los que compartí muchas experiencias.

Lizzette Gabriela Betancourt Granillo

## AGRADECIMIENTO

A Dios y la Virgencita del Cisne, por darme salud y la sabiduría para seguir adelante.

A mis padres y hermanos, parte fundamental de mi vida, quienes nunca me dejaron caer, por siempre motivarme para que cumpla esta meta.

A mi director de trabajo de Titulacion, Mgs Guido Riofrío, porque su ayuda, orientación, motivación y confianza, impartiéndome sus conocimientos y brindarme su apoyo en el transcurso de este trabajo.

Muchas gracias a todos, a quienes de una u otra manera me apoyaron con un granito de arena para poder culminar este proyecto.

Lizzette Gabriela Betancourt Granillo

**INDICE DE CONTENIDOS**

## INDICE DE TABLAS

## INDICE DE FIGURAS

## INDICE DE ECUACIONES

## RESUMEN

La minería de datos es el proceso de descubrir relaciones, patrones o tendencias al examinar grandes cantidades de datos. En la Universidad Técnica Particular de Loja desde varios ciclos se han implementado actividades en línea (foro, chat y video colaboración) en el proceso de enseñanza aprendizaje de la modalidad a distancia, luego de varios ciclos de su implementación surge la necesidad de evaluar el impacto que han tenido estas actividades, además de encontrar y evaluar patrones de comportamiento. En el presente trabajo de titulación se aplica técnicas de minería de datos como las reglas de asociación y los mapas autoorganizados de Kohonen a cada una de las actividades en línea en el periodo académico Abril- Agosto 2015 de las titulaciones de modalidad a distancia, con el fin de analizar y evaluar el comportamiento de los estudiantes con dichas actividades.



**PALABRAS CLAVES:**

Minería de datos, patrones de comportamiento, reglas de asociación, algoritmo a priori, mapas autoorganizados de Kohonen.

## ABSTRACT

Data mining is the process of discovering relationships, patterns or trends in examining large amounts of data. In the Universidad Técnica Particular de Loja makes several cycles have been implemented activities online (Forum, chat, and video collaboration) in the process of teaching-learning of the mode to distance. after several cycles of its implementation arises the need to evaluate the impact these activities have had, as well as find and evaluate behavioral patterns.

This work applies techniques of data mining Association rules and self-organized maps of Kohonen to each of the online activities for the academic period April - August 2015 of the degrees of modality to distance, in order to analyze and evaluate the behavior of students with these activities.



**KEYWORDS:**

Data mining, behavior patterns, association rules, Apriori algorithm, Kohonen's Self Organizing Maps

# CAPÍTULO 1. ESTADO DEL ARTE

## 1.1 **Introducción**

Actualmente en las instituciones educativas se han adaptado entornos de aprendizaje virtual, los cuales pretenden ser beneficiosos para el aprendizaje de los diferentes estudiantes ya que con este se pretende cubrir o satisfacer algunas dudas mediante la interacción estudiante-docente y viceversa con actividades en línea.

Es por esto que mediante la aplicación de técnicas de minería de datos se pretende determinar y evaluar patrones de comportamiento de los estudiantes de la modalidad a Distancia de la UTPL, además de evaluar el impacto que ha tenido la realización de actividades como foros, chats y video colaboración que el Entorno Virtual de Aprendizaje (EVA) ofrece.

Uno de los entornos virtuales de aprendizaje más populares en Latinoamérica es el Moodle, este permite la gestión y presentación de materiales educativos a los estudiantes, sin embargo también los docentes pueden realizar cursos en línea, debates, foros, chats, etc., su objetivo principal es el de permitir el aprendizaje en cualquier parte y cualquier momento con el único requerimiento de poseer un ordenador y conexión a internet.

## 1.2 **Entorno virtual de aprendizaje**

Los entornos virtuales de aprendizaje (EVA) o Virtual Learning Environment (VLE), plataformas e-learning o plataformas educativas son sistemas de software en línea que están diseñados para ayudar a los estudiantes con las diversas tareas de aprendizaje, como gestionar cada uno de los contenidos y recursos necesarios para que sus actividades sean exitosas, además de ser una interacción didáctica en la web. (heatherwilliamson, n.d.)

Según Boneu (2007) hay cuatro características básicas que las plataformas de e-learning deben tener:

- **Interactividad:** la persona que usa la plataforma debe saber que él es el protagonista de su formación.
- **Flexibilidad:** conjunto de funcionalidades que permiten al sistema adaptarse con facilidad en la organización en donde se implementará.
- **Escalabilidad:** capacidad que tiene la plataforma e-learning para trabajar con poco o gran número de usuarios.
- **Estandarización:** capacidad de utilizar cursos realizados por terceros, es decir que los cursos están disponibles para los creadores del mismo o para otros que cumplen los estándares requeridos.

Existen dos tipos de plataformas virtuales de aprendizaje: plataformas comerciales y plataformas open source, esta última permite el acceso directo al código fuente para personalizarlo si se requiere. (Belloch, 2012)

Los sistemas de software más utilizados son los sistemas de gestión del aprendizaje (Learning Management Systems) o LMS, algunos ejemplos de LMS de código abierto son.

- Moodle (Entorno modular de aprendizaje dinámico orientado a objetos), es una plataforma de aprendizaje que está en un servidor web, se lo define como un sistema de gestión de contenidos (CMS), es un paquete de software diseñado para ayudar a los docentes a crear fácilmente cursos en línea de calidad, (Miratía, 2008) además ofrece a los estudiantes un servicio automatizado y personalizado a sus necesidades e intereses. (Martinez, 2008)

- Dokeos es el primer sistema de gestión de aprendizaje que integra auditoría en línea, interacción, seguimiento y videoconferencia en un mismo software libre. Proporciona información exacta de los alumnos como número de accesos a las herramientas, tiempo de conexión, etc. (Miratía, 2008). Permite la interacción entre alumnos y maestros, realización de evaluaciones y actividades didácticas. Es muy fácil de usar aunque con algunas limitantes de funcionalidad y seguridad además de presentar algunos problemas con el manejo de ciertos paquetes SCORM[1]. (Valle, Arévalo, & Muñoz, 2014)

Los LMS permiten realizar cinco funciones principales:

- **Administración**: facilita la gestión de usuarios y gestión de entornos de aprendizaje.
- **Comunicación**: permite interacción de alumnos con alumnos, profesor a alumnos o viceversa, etc., por medio de chats, foros, correo electrónico en una o doble vía, asincrónicamente o sincrónica. (Ortiz F, 2007)
- **Gestión de contenidos**: disponen de un sistema de almacenamiento y gestión de archivos que permiten realizar funciones básicas, como visualizar, organizar en carpetas, descargar o cargar archivos al entorno de aprendizaje.

[1] Conjunto de estándares que permite crear objetos pedagógicos estructurados.

- **Gestión de grupos:** realizan operaciones de modificación o borrado de grupos de alumnos, además crea “escenarios virtuales” para trabajos cooperativos de los miembros de un grupo.
- **Evaluación:** permiten la creación, edición y realización de ciertos tipos de test, además la retroalimentación, publicación de calificaciones, estadísticas sobre resultados y progreso de cada alumno. (Fernández, 2010)

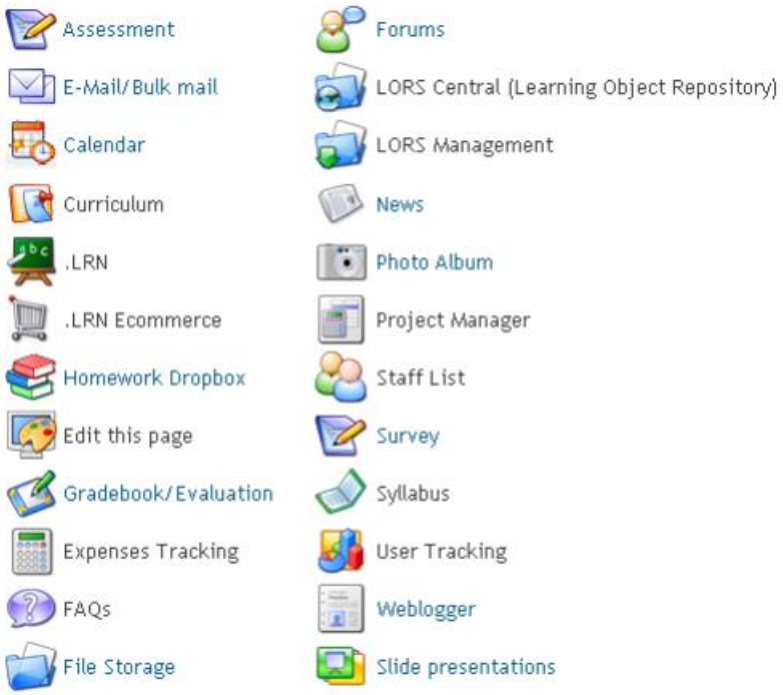



Figura 1. Conjunto de herramientas LMS

Fuente: (Fernández, 2010)

## 1.3 **E-learning**

La educación a distancia es la base fundamental para el desarrollo del e-learning, ya que permite resolver problemas como: índole geográfica, porque no es necesario desplazarse de un lugar a otro; tiempo, ya que el estudiante escoge su propio horario. (Gallego & Martinez, 2003)

El término e-learning hace referencia al aprendizaje electrónico es decir, es un proceso enseñanza-aprendizaje en donde hay interacción entre docente y estudiante o viceversa, por medio de dispositivos electrónicos conectados a Internet para proporcionar a contenidos educativos. (Moreira & Segura, 2009)

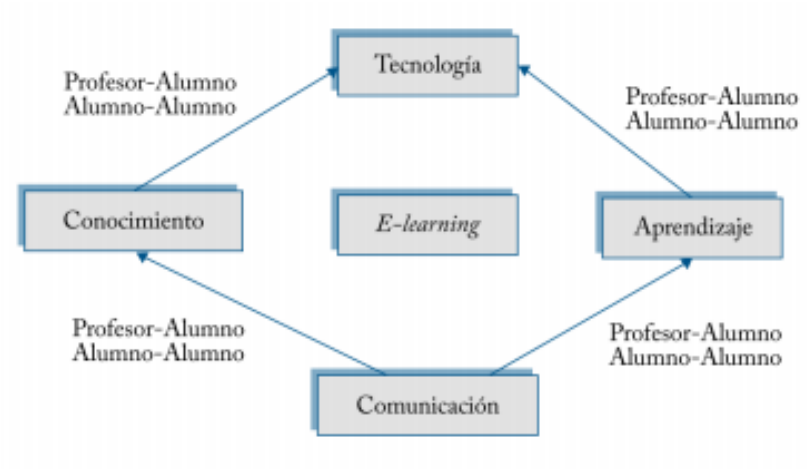



Figura 2. Interacción en el E-learning

Fuente: (Cabero & Gisbert, 2005)

El e-learning es un sistema de impartición de información a distancia, apoyado en las TIC, ya que combina distintos elementos pedagógicos: la instrucción directa clásica (presencial o de autoestudio), las prácticas, los contactos en tiempo real (presenciales, videoconferencia o chats) y los contactos diferidos (tutores, foros de debate, correo electrónico). (García, 2002)

“e-learning se refiere al proceso de aprendizaje a distancia que se facilita mediante el uso de las tecnologías de la información y comunicación” (Barberá, 2008)

“De acuerdo a (IEEE LTSC, 2001), los sistemas de educación basados en ordenadores se denominan LTS (Learning Technology Systems) y son definidos como sistemas de aprendizaje, educación y entrenamiento que son soportados por las Tecnologías de la información y de la comunicación.“.(Morales, 2010)

Boneu (2007) menciona que los sistemas e-learning pueden clasificarse según su tecnología:

- El CBT (*Computerbased training*) es un aprendizaje basado en el computador con mecanismos de retroalimentación, en donde el alumno es un ente más activo dentro de su proceso formativo.
- El WBL (*Web Based Learning*), es el aprendizaje haciendo uso de la web, su propósito es realizar la entrega de contenido a través de una red que se complementa con la participación de tutores. (Morales, 2010)
- El IBT (*Internet Based training*), es la evolución del CBT, no se requiere estrictamente la web, se puede hacer uso de servicio del protocolo TCP/IP.

Tabla 1. Ventajas y desventajas de la educación virtual

| Ventajas | Desventajas / inconveniente |
|---|---|
| Ofrece a los alumnos amplio volumen de información. | Requiere más tiempo por parte del docente. |
| Ahorra costos y desplazamientos | Habilidades para aprendizaje autónomo por parte de los estudiantes. |
| Facilita el uso de materiales en diferentes cursos. | Supone baja calidad de cursos y contenido actuales. |
| Interactividad con profesores, estudiantes e información | Requiere más trabajo que lo convencional. |
| Propicia una formación just in time y just for me | Depende de una conexión a internet y que sea rápida |

Fuente: (Cabero & Gisbert, 2005)

En las plataformas e-learning en la actualidad existe un elemento que ha tenido mucho éxito, como es el campus virtual, este se lo puede definir como un lugar virtual para enseñanza, aprendizaje e investigación que integran herramientas TIC (Tecnologías de la información y las comunicaciones), además es un espacio para la administración y organización de todas las actividades y procesos de una universidad.

Un modelo pedagógico, es una muestra de alternativas posibles de enseñanza-aprendizaje, estos son un conjunto de conceptos, principios y esquemas que pretenden dar un fundamento a problemas relacionados con fines educativos. (Cabero, 2006)

La tarea pedagógica para un tutor es asumir la responsabilidad y compromiso ético de decidir cuáles son los contenidos más valiosos para la enseñanza, además justificar los conocimientos que tengan relación con el contexto social y profesional. (Escobar, 2009)

Desde el punto de vista de (Cabero, 2006) se debería abrir una etapa en la que se tome en cuenta que los procesos de enseñanza-aprendizaje son sistémicos, a las características de los estudiantes para garantizar el éxito de las acciones formativas se deben adaptar variables como se muestra a continuación.

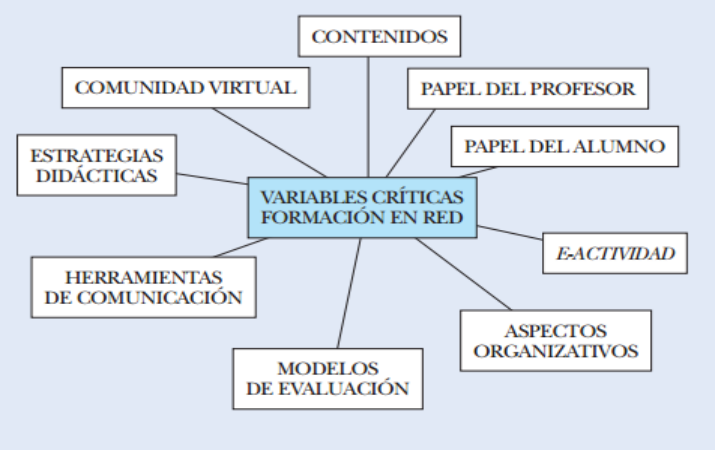



Figura 3. Variables críticas en la formación en red

Fuente: (Cabero, 2006)

Destacando algunas variables críticas de la formación en red:

- **Papel del profesor**, su modelo pedagógico es centrado en la enseñanza, es muy importante ya que este no actúa igual como lo hace en la formación tradicional, pasó de ser un transmisor de información a ser tutor y orientador virtual, cumple diversas funciones de las académicas (técnica, orientadora, social, etc). (Cabero, 2006) El cual se define como un agente que guía, orienta y evalúa el aprendizaje, siempre trata de proporcionar la mejor metodología de enseñanza dirigida al alumno. (Rivera, 2004)

- **Papel del alumno** también es importante porque es el receptor de la nueva información, si este mantiene su recepción pasiva el proceso de educación virtual fracasar. Es por esto que las e-actividades ayudarán que los alumnos dejen de ser pasivos llevando a cabo acciones de e-learning y no de e-reading. El alumno e-learning debe tener destrezas para identificar una necesidad, conocer cuando hay necesidad de nueva información, saber cómo trabajar con diversas fuentes, dominar sobrecarga de información, evaluarla y dominarla al punto de resolver su necesidad y comunicar la información a otros. (Cabero, 2006).

### 1.3.1 **Estrategias de aprendizaje.**

Para Benito (2009) el termino estrategia hace referencia a que los procedimientos usados para aprender son una parte fundamental y decisiva para el resultado final de este proceso.

Las estrategias se pueden clasificar dependiendo de las funciones cognitivas a realizar:

- Asociativas: implican operaciones básicas y elementales, incrementan la posibilidad de recordar la información sin hacer cambios estructurales en ella.
- De elaboración: es un paso intermedio entre las estrategias asociativas y de organización, se pueden originar operaciones más simples, en donde se estableces algunas relaciones.
- De organización: establece relaciones internas entre elementos que componen los materiales de aprendizaje y los conocimientos previos que posea.

Algunas funciones cognitivas depende de algunos aspectos de la personalidad es por esto que se incluye las estrategias de apoyo, son aquellas que, en lugar de dirigirse directamente al aprendizaje de los materiales, tienen como misión incrementar la eficacia de ese aprendizaje y mejorar las condiciones en las que se produce.

### 1.3.2 Estilos de aprendizaje.

El término estilo de aprendizaje se refiere a que cada persona puede hacer uso de su propio estrategia de estudio a la hora de aprender. Las estrategias pueden cambiar debido al ámbito que desea aprender dependiendo de sus preferencias; así mismo utilizan diversas velocidades con mayor o menor eficacia para aprender incluso cuando tienen las mismas motivaciones. (Cazau, 2010)

Según Alonso, Gallego yHoney (como se citó en Cardozo, 2012) se define como estilos de aprendizaje a los diversos rasgos cognitivos, afectivos y fisiológicos que sirven como indicadores estables, de cómo las personas perciben, interaccionan y responden a sus ambientes de aprendizaje.

Aunque no existe una definición única de estilos de aprendizaje, una forma de describirlo es que son las diversas maneras en el que una persona puede aprender algo de su agrado basándose en la interacción, aceptación y procesamiento de la información (Franco, Yamasaki, & Domínguez, 2010). Algunos estilos de aprendizajes se desglosan a continuación en la tabla 2.

Tabla 2. Clasificación de estilos de aprendizaje

| Teorias | Estilos de aprendizaje |
|---|---|
| EL MODELO DE LOS CUADRANTES CEREBRALES DE NED HERRMANN | ❖ Cortical izquierdo<br>❖ Límbico izquierdo<br>❖ Límbico derecho |

| Teorias | Estilos de aprendizaje |
|---|---|
| | ❖ Cortical derecho |
| HONEY Y MUMFORD | ❖ Activo<br>❖ Reflexivo<br>❖ Pragmático<br>❖ Teórico |
| MODELO DE PROCESAR LA INFORMACIÓN ( DAVID KOLB) | ❖ Convergente<br>❖ Acomodador<br>❖ Asimilador<br>❖ Divergente. |
| CATEGORÍA BIPOLAR (FELDER Y SILVERMAN) | ❖ Activo / reflexivo<br>❖ Sensorial / intuitivo<br>❖ Visual / verbal<br>❖ Secuencial / global |
| MODELO VARK DE NEIL FLEMING | ❖ Visual<br>❖ Auditivo<br>❖ Lectura/Escritura<br>❖ Kinestésico |

Fuente: (Franco et al., 2010)

Según Gallego& Martinez(2003) existen varios los estudios que confirman que los estilos de aprendizaje están relacionados con el éxito académico.

Luego de varias investigaciones (Alonso, Gallego y Honey, 1999) concluyen que los estudiantes aprenden con más efectividad cuando se enseña con los estilos de aprendizaje predominantes. Pero también se debe tener en cuenta que el éxito de los estudiantes no solo depende de los estilos de aprendizaje propuestos, sino dependen también por los métodos de enseñanza de los docentes.

### 1.4 Minería de datos

La revolución digital ha sido el causante que la información digitalizada sea más fácil de capturar, procesar, almacenar, distribuir y transmitir. Con el gran avance en la tecnología e informática se continua recogiendo y almacenando gran cantidad de información en las base de datos. Con la minería de datos (MD) se pretende buscarle algún sentido a toda esta

información almacenada.

La minería de datos se puede definir como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos.

Varios autores mencionan que minería de datos es sinónimo de KDD, esta es una definición errónea. La minería de datos es solo una etapa del proceso de extracción de conocimiento a partir de datos (KDD) (Pérez & Santín, 2007)este proceso consta de varios pasos, éstos son aplicados de una manera iterativa e interactiva aseguran que un conocimiento útil se extraiga de los datos.

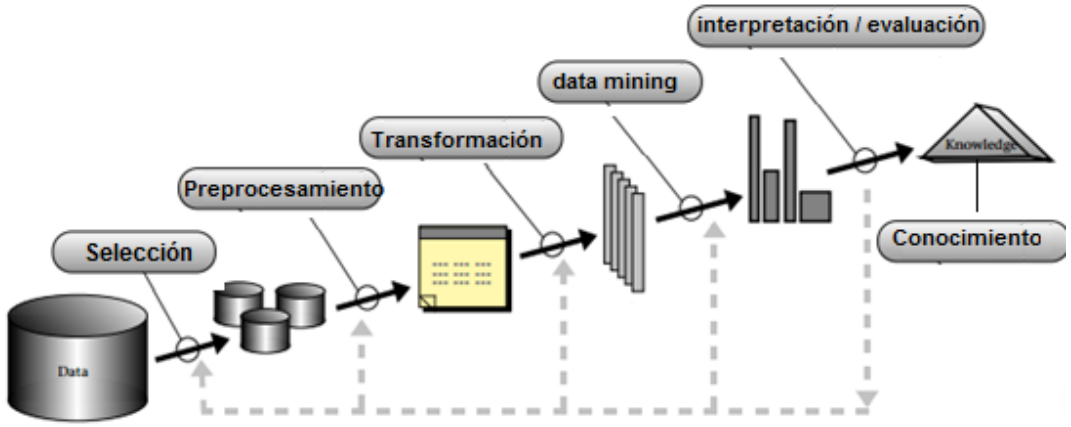



Figura 4. Etapas de KDD
Fuente:(Fayyad, Piatetsky-Shapiro, & Smyth, 1996)

En las fases de selección, limpieza y transformación se eliminan o corrigen datos incorrectos y toma decisiones sobre que variables o atributos son relevantes. Estas etapas se las considera como preparación de datos. Mientras que la fase de data mining o minería de datos se decide cual es la tarea a realizar y el método con el que se va a trabajar. En la siguiente etapa que es la fase de evaluación o interpretación se evalúan los patrones y si es necesario se vuelven a fases anteriores, para resolver posibles conflictos con el conocimiento.(Hernández, Ramírez, & Ferri, 2004)

Como menciona Pérez y Santín (2007) existen varias interpretaciones del concepto de minería de datos, entre una de ellas SAS[2] Institute define el concepto de Data mining como el proceso de seleccionar, explorar, modificar, modelizar y valorar las grandes cantidades de datos con el objetivo de descubrir patrones desconocidos. En la siguiente figura se ilustra las fases de minería de datos según SAS Institute.

[2] Es uno de los principales fabricantes de business intelligence software

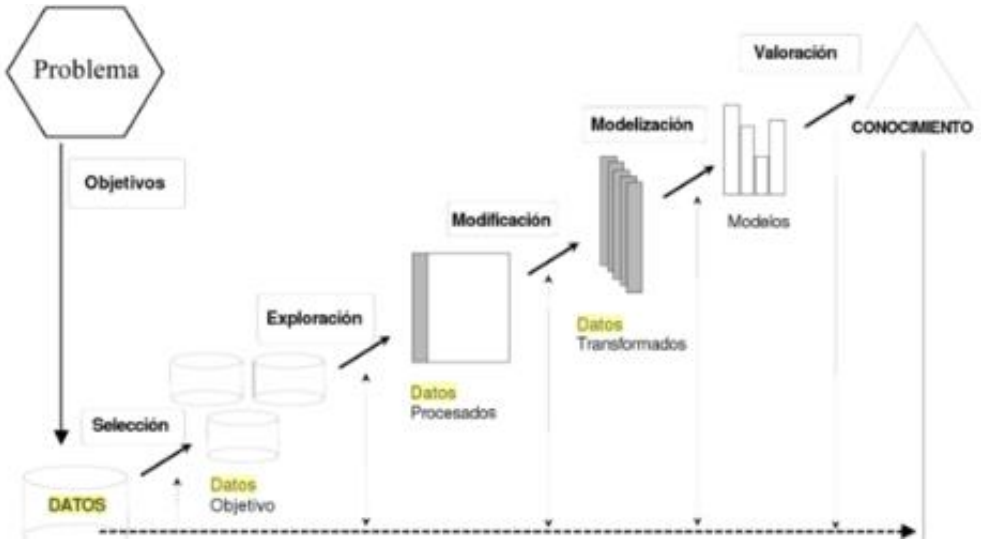



Figura 5. Fases de minería de datos según SAS Institute

Fuente: (Pérez & Santín, 2007)

### 1.4.1 **Tipos de modelo.**

La minería de datos tiene como objetivo extraer el conocimiento a partir de un conjunto de datos. Dicho conocimiento puede estar en forma de relaciones, patrones o reglas inferidos de los datos o de una descripción más concisa, estas relaciones constituyen un modelo de datos analizados, que se pueden representar de diversas formas y cada una determinar el tipo de técnica que puede usarse.

Los modelos pueden clasificarse en predictivo y descriptivo. Los modelos predictivos son aquellos que intentan predecir o estimar valores futuros o desconocidos de las variables de interés. Mientras que los modelos descriptivos identifican patrones que explican o describen los datos, es decir, exploran el comportamiento de los datos examinados no para predecir nuevos datos.(Hernández et al., 2004)

### 1.4.2 **Técnicas de Minería de datos**

Las técnicas de minería de datos son una etapa del proceso de KDD, estas intentan obtener un conjunto de reglas o patrones partiendo de los datos coleccionados, extraen información de los datos partiendo generalmente de algoritmos. Dichas técnicas se clasifican dentro de dos grupos: no supervisadas o descriptivas y supervisadas o predictivas.

#### *1.4.2.1 **No supervisadas o descriptivas.***

##### *1.4.2.1.1 Clustering.*

También llamado agrupamiento, permite la identificación de grupos en donde los datos tienen gran similitud entre si y disimilitud con otros grupos de datos. (Molina & García, n.d.)Por ejemplo, una biblioteca en donde los libros tienen gran variedad de temas disponibles.

El desafío es mantener los libros juntos, de manera que los lectores puedan tomar varios libros sin necesidad de molestarse. Mediante la técnica de agrupamiento, podemos mantener los libros de contenido similar en un clúster. Es decir cuando los lectores requieran de un tema en especial, solo necesitarían ir al estante o lugar en donde se encuentre, sin necesidad de recorrer toda la biblioteca para buscar. (Diwate & Sahu, 2014)

- **Redes neuronales:** Inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema que permite interconectar las neuronas en una red que presta colaboración para la producción de estímulos.

- **Algoritmo Kmeans**: es uno de los algoritmos más conocidos de agrupamiento en donde divide la data en k grupos, su principal idea es la de definir el k centroide, toma los objetos y ubica en su centroide más cercano. El siguiente paso es recalcular el centroide de cada grupo y se vuelve a distribuir los según el centroide más cercano. Los pasos se repiten hasta cuando ya no hay cambios en los grupos de un paso a otro. (Pascual & Sánchez, n.d.)

- **Cobweb**.- es un algoritmo de Clustering jerárquico, utiliza aprendizaje incremental, es decir, realiza agrupaciones de instancia a instancia. Mientras se ejecuta el algoritmo se genera un árbol de clasificación donde las hojas representan los segmentos y el nodo raíz contiene el conjunto de datos de entrada. Las instancias se van añadiendo una a una de manera que se va actualizando, dicho actualización consiste en encontrar el mejor sitio donde incluir la nueva instancia, con esta operación se puede reestructurar el árbol. Para saber cómo y dónde actualizar el árbol, es tener una medida llamada utilidad de categoría, que mide la calidad general de una partición de instancias en un segmento. Este algoritmo es sensible a dos parámetros:

    - **Acuity**: la utilidad de la categoría se basa en una estimación de la media y desviación estándar del valor de los atributos.

- **Cut-off:** se usa para evitar el crecimiento del número de segmentos (Garre, Cuadrado, Silicia, Rodriguez, & Rejas, n.d.)

- **Algoritmos genéticos y evolutivos**: es un método de búsqueda que imita la evolución de la teoría de evolución biológica de Charles Darwin para la resolución de problemas realizando la optimización de procesos. (Arranz de la Peña & Parra Truyol, n.d.)

- **Máquinas de vectores de soporte (Support vector machines):** se aplican a problemas de clasificación y regresión. Las SVM están basadas en aprendizaje estadístico, funcionan con datos dimensionales, este algoritmo pretende encontrar hiperplanos óptimos que separen de la mejor manera un conjunto de datos en dos o más clases. Cuando los datos no son separables linealmente se procede a realizar estrategias de cambio de dimensionalidad. (Ramírez, 2007)

- **Redes de Kohonen:** También llamado mapa auto-organizado o SOM (Self-Organizing Maps), su comportamiento es similar al cerebro, ya que se encarga de descubrir rasgos comunes, regularidades en los datos entrantes para auto organizarlos e incorporarlos a su estructura de conexiones en función de los datos procedentes del exterior.

*1.4.2.1.2 Asociación.*

(Hernández et al., 2004) dice que las reglas de asociación son una manera muy popular de expresar patrones de datos de una base de datos. Estos patrones pueden servir para conocer el comportamiento general del problema que genera la base de datos, de esta manera se tenga más información que pueda asistir en la toma de decisiones.

Son utilizadas cuando el objetivo es realizar un análisis buscando relaciones dentro de un conjunto de datos, se usan para predecir comportamientos. (Molina & García, n.d.) Uno de los algoritmos más usados es el algoritmo A priori.
Frecuentemente dadas las reglas de asociación se trabaja con dos medidas para conocer la calidad de la regla, cobertura (support) y confianza (confidence).

La cobertura de una regla es el número de instancias que la regla predice correctamente, mientras que la confianza (también llamada precisión) mide el porcentaje de veces que la regla se cumple cuando se puede aplicar.(Hernández et al., 2004)

- **Algoritmo A priori:** Este algoritmo se basa en el conocimiento previo o “a priori” de los conjuntos frecuentes, sirve para reducir el espacio en búsqueda y aumentar la eficiencia.

  “Se basa en la búsqueda de los conjuntos de ítems con determinada cobertura. Para ello se construyen los conjuntos formados por un solo ítem que supera la cobertura mínima. Este conjunto de datos se utiliza para construir el conjunto de conjuntos de dos ítems y así sucesivamente hasta que llegue a un tamaño en el cual no existan conjuntos de ítems con la cobertura requerida” (Hernández et al., 2004, p. 240).

- **Regresión Logística:** Este es un modelo de la regresión lineal simple, el cual predice la probabilidad que un resultado pueda tener dos valores (binaria o dicotómica) en función de variables independientes, que puede ser cualitativas o cuantitativas, adopta dos posibles valores: 1 y 0, éxito y fracaso, positivo y negativo, etc.

- **CN2:** el algoritmo cn2 fue propuesto en base a dos algoritmos, el AQ que encuentra el mejor conjunto de reglas por medio de la búsqueda en estrella y el algoritmo TDIDT (Top Down Induction of DecisionTrees), por finalizar la búsqueda cuando las reglas no tengan el nivel de significancia estadística. El algoritmo funciona encontrando la mejor regla a partir de un conjunto de datos y eliminando los restantes, este finaliza cuando el conjunto de datos queda vacío o no se encuentren reglas que tengan un minino nivel de significancia exigido (Muñoz & Moreno, 2010).

#### *1.4.2.2* ***Supervisadas o predictivas:***

(Hernández et al., 2004) afirma: “Se trata de problemas y tareas en las que hay que predecir uno o más valores para uno o más problemas” (p.139). Estas tareas se desglosan en Técnicas de predicción y de asociación.

A continuación algunas de las técnicas más representativas son:

- **Arboles de decisión:** son estructuras en formas de árbol que representan un conjunto de decisiones, dichas decisiones contienen reglas para la clasificación de un conjunto de datos.

  “Los sistemas de aprendizaje basados en arboles de decisión son quizás el método

más fácil de utilizar y entender"(Hernández et al., 2004, p.281). Un árbol de decisión es una estructura jerárquica que está formada por un conjunto de nodos, en donde cada nodo hace referencia a una condición o regla, que puede tener valores de verdadero o falso. De tal modo que en la decisión final se pueda determinar si siguiendo las condiciones desde el nodo raíz del árbol hasta los nodos hijos se cumple lo esperado.

- **Algoritmo J48:** También conocido como el algoritmo C45. Permite la predicción y clasificación basada en la teoría de la clasificación de datos., además permite trabajar con valores continuos para los atributos, separando los posibles resultados y dos ramas y poder escoger un rango de medida apropiada.(Haro & Perez, 2014)

- **Algoritmo REPTree:** método de aprendizaje en el cual se construye un árbol de decisión usando la información de varianza y lo poda usando como criterio la reducción del error. Solo clasifica valores para atributos numéricos una vez.(Hernández& Abilowo, n.d.)

- **Métodos Bayesianos**: su principal característica es el uso de distribuciones de probabilidad para cuantificar la incertidumbre de los datos que se desea modelar.

  Hernández (2004), es una de las más usadas en problemas de inteligencia artificial, aprendizaje automático y minería de datos; Es un método práctico para realizar inferencias a partir de los datos, la misma que se basa en estimar la probabilidad de pertenencia mediante la estimación de probabilidades, utilizando el algoritmo de Bayes.

  Para usar los métodos bayesianos es recomendable trabajar con bastantes datos, ya que si se tiene pocos datos no se puede realizar predicciones y no podría proporcionar un modelo de datos correcto.

  Estos métodos utilizan la tarea de clasificación para extraer los patrones de comportamiento, algunos algoritmos utilizan este tipo técnica como el clasificador:

  - Algoritmo Naïve Bayes: es un clasificador probabilístico, se basa en modelo de probabilidades que integra suposiciones de independencia (no tienen efecto sobre la realidad). ("IBM Knowledge Center," 2013)

- Redes bayesianas (RBs): “es un formalismo que ha demostrado su potencialidad como modelo de representación del conocimiento con incertidumbre,… Es una herramienta muy atractiva en su uso como representación del conocimiento, aspecto muy importante de la minería de datos” (Hernández et al., 2004, p. 263).

- **Regresión lineal:** Es el modelo más usado de regresión, sirve para formar relaciones entre datos, es rápida y eficaz, excepto en espacios multidensionales en donde es insuficiente y no puede relacionarse con más de dos variables, estos se modelan en una línea recta. Según Molina y Gracia, (2006) citado en (Toro, F. 2015) la RL es la forma más simple de regresión, se caracteriza por el uso de dos variables una aleatoria Y (llamada variable respuesta) y una aleatoria X (variable predictora).

Tabla 3. Algunas técnicas de Minería de datos

| | PREDICTIVO | | DESCRIPTIVO | | |
|---|---|---|---|---|---|
| NOMBRE | **Clasificación** | **Regresión** | **Agrupamiento** | **Reglas de Asociación** | **Correlaciones/ Factorizaciones** |
| **Redes neuronales** | ✓ | ✓ | ✓ | | |
| **Árboles de decisión ID3, C5.0** | ✓ | | | | |
| **Árboles de decisiones CART** | ✓ | ✓ | | | |
| **Otros árboles de decisión** | ✓ | ✓ | ✓ | ✓ | |
| **Redes de Kohonen** | | | ✓ | | |
| **Regresión lineal y logarítmica** | | ✓ | | | |
| **Regresión logística** | ✓ | | | ✓ | |
| **Kmeans** | | | ✓ | | |
| **Apriori** | | | | ✓ | |

| | | | | | |
|---|---|---|---|---|---|
| **NaiveBayes** | ✓ | | | | |
| **Vecinos más próximos** | ✓ | ✓ | ✓ | | |
| **Análisis factorial y de componentes principales** | | | | | ✓ |
| **Twostep, Cobweb** | | | ✓ | | |
| **Algoritmos genéticos y evolutivos** | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Máquinas de vectores de soporte** | ✓ | ✓ | ✓ | | |
| **CN2 rules (cobertura)** | ✓ | | | ✓ | |
| **Análisis discriminante multivariante** | ✓ | | | | |

Fuente. (Hernández et al., 2004)

## 1.5 **Herramientas de Minería de datos**

Las herramientas de minería de datos permiten la extracción de patrones para describir y entender mejor los datos y poder predecir comportamientos a futuro. (Pérez & Santín, 2007)

Existen varias herramientas para minería de datos, estas se clasifican en herramientas comerciales y de Open Source (código libre), en este caso veremos más sobre herramientas Open Source.

### 1.5.1 **Weka[3] (Waikato Environment for Knowledge Analysis)**

Es una herramienta visual bajo la licencia general pública GNU desarrollada en Java, creada por los investigadores de la Universidad de Waikato de Nueva Zelanda, contiene un conjunto de algoritmos de aprendizaje automático para tareas de minería de datos.

[3] Weka: http://www.cs.waikato.ac.nz/ml/weka/

Con el objetivo de facilitar su uso Weka contiene interfaz gráfica para el usuario, con la finalidad de que pueda configurar las herramientas necesarias. ("Weka 3: Software de minería de datos en Java," n.d.)

Este entorno sirve para la experimentación de análisis de datos que permite aplicar, analizar y evaluar técnicas. También proporciona acceso a base de datos SQL gracias a la conexión JDBC[4], que mediante una consulta permite procesar el resultado.

**Características**

- La carga de datos se la puede realizar con archivos de formato arff, csv, c4.5.
- Se puede integrar en otros paquetes Java.
- Está constituido por paquetes de código libre como técnicas de pre-procesamiento, clasificación, agrupamiento, asociación y visualización.
- Es débil en casos de estadística clásica
- No guarda parámetros de escala para aplicar a datos futuros.
- No cuenta con servicio de automatización de parámetros

### 1.5.2 **RapidMiner[5]**

RapidMiner anteriormente YALE (Yet Another Learning Environment), es una plataforma de software bajo la licencia general pública GNU, que proporciona un entorno para el aprendizaje automático, minería de datos, minería de texto, análisis predictivo y análisis de negocios.

**Características**

- Multiplataforma
- Incluye muchos algoritmos de aprendizaje de Weka
- Puede usarse de varias maneras: por línea de comandos, Bash, desde otros programas usando a sus librerías
- Proporciona más de 500 operadores orientados al análisis, procesamiento y visualización de datos. (Cardona, 2011)

### 1.5.3 **R Project[6]**

Es un lenguaje y entorno de computación y gráficos estadísticos, con un conjunto de servicios para la manipulación de datos, cálculo y representación gráfica.

---

[4] Java DatabaseConnectivity

[5] RapidMiner: https://rapidminer.com/

[6] R Project: https://www.r-project.org/

R ofrece una amplia variedad de técnicas estadísticas y gráficos, incluyendo modelos lineales y no lineales, pruebas estadísticas clásicas, clasificación, agrupación, entre otros.

Es uno de las principales herramientas usadas en la investigación estadística y se ejecuta en una amplia variedad de las plataformas UNIX, Windows y Mac OS. Es mantenido y distribuido por un equipo de estadísticos e informáticos que trabajan en la industria. (Kosorus, Honigl, & Kung, 2011)

### 1.5.4 **Knime[7] (Konstanz Information Miner)**

Es una plataforma código abierto de análisis de datos y presentación de información, basada en la plataforma Eclipse. Posee 100 nodos de procesamiento de datos, pre-procesamiento de datos, limpieza, modelado, análisis y minería de datos, además vistas iterativas, como gráficos de dispersión, paralelo, coordenadas, entre otros.

En esta herramienta se puede realizar el pre-procesamiento de datos, además integra componentes para aprendizaje automático y minería de datos a través de su concepto de segmentación de datos, es por esto que ha llamado la atención de la inteligencia de negocios y análisis de datos financieros. (Goopta, 2014)

En la encuesta anual del año 2015, el portal internacional de Minería de Datos KDnuggets[8], obtuvo el siguiente ranking de las herramientas más usadas durante ese año.

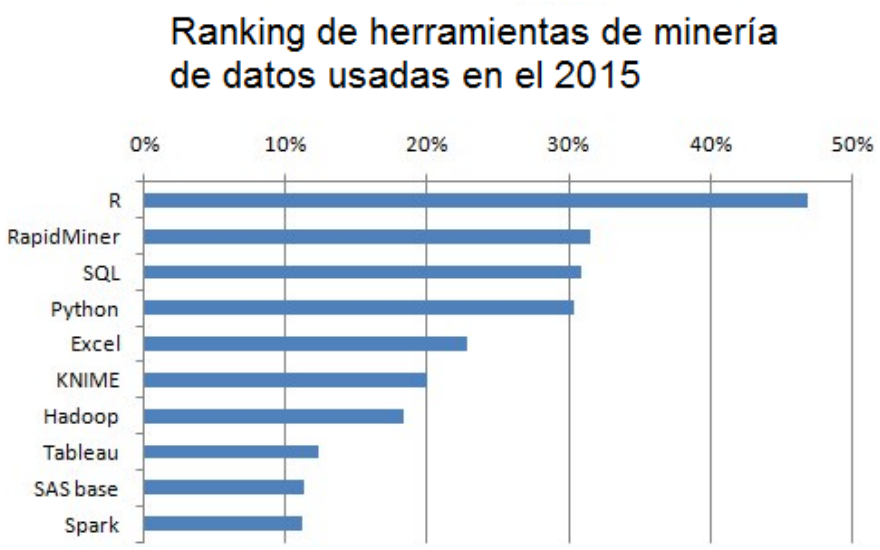



Figura 6. Ranking de las herramientas más usadas del 2015

Fuente: (KDnuggets, 2015)

[7] Knime: https://www.knime.org/

[8] KDnuggets: http://www.kdnuggets.com/

R ha conseguido destronar del primer lugar del ranking del 2014 a RapidMiner con un 49% de participación. Mientas que RapidMiner con respecto a años anteriores ha disminuido y tiene una participación de 33% al igual que Python.

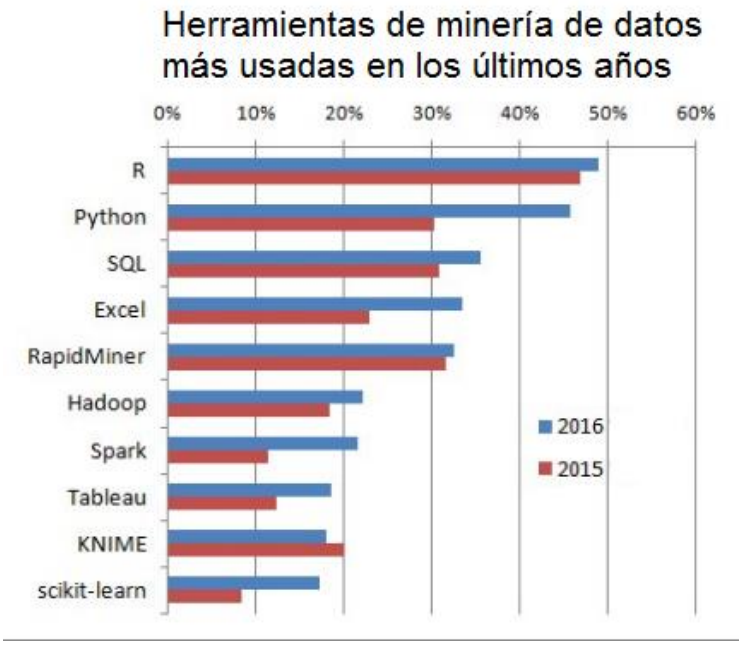



Figura 7. Ranking de las herramientas más usadas del 2015 y 2016
Fuente:(KDnuggets, 2016)

Comparando el top 10 de KDnuggets del año 2015 con el primer trimestre del 2016, el uso de la herramienta R se va posicionando entre las principales herramientas, seguidamente se encuentra Python.

Es por esto que para el presente trabajo de fin de Titulacion se ha decidido trabajar con la herramienta R. Ya que es la herramienta que últimamente está sobresaliendo por su sin número de funciones como por ejemplo, la rapidez que posee al manipular los datos, la facilidad que tiene para resolver datos estadísticos por medio de varias funciones, etc.

## 1.6 Metodología para el proyecto de minería de datos

Las metodologías permiten llevar a cabo el proceso de minería de datos en forma sistemática y no trivial, especifica cómo hacerlo, no solo define las fases del proceso sino también como deberían realizarse las tareas.

Como lo menciona Moine, Haedo y Gordillo (2013), existen algunas metodologías de minería de datos, como SEMMA, Catalyst y CRISP-DM. Estos autores concluyen que las

metodologías Catalyst y CRISP-DM son considerados como metodologías de minería de datos, ya que describen cada una de las actividades de cada fase, guiando así al desarrollo de la metodología.

Según la encuesta realizada por KDnuggets en el año 2007 y 2014, los resultados son sorprendentes, debido a que CRISP-DM es la metodología más usada.

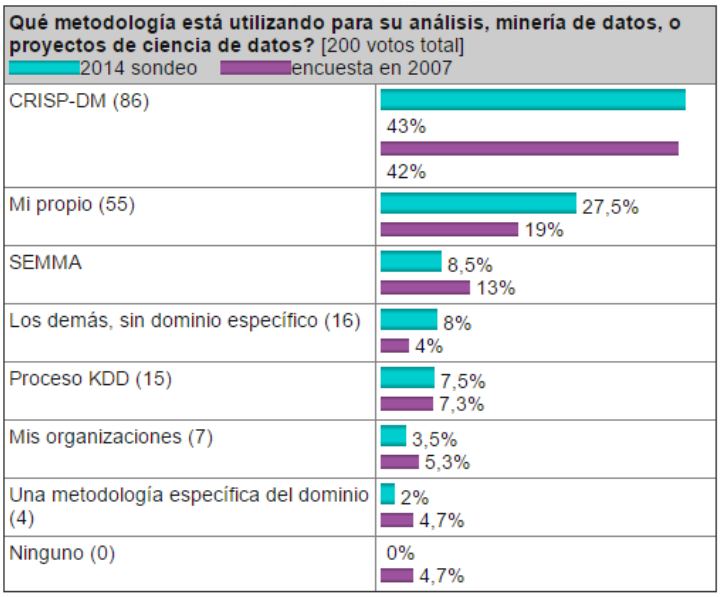



Figura 8. Encuesta realizada por la KDnuggets sobre metodologías de Data mining
Fuente: (Kdnuggets, 2014)

### 1.6.1 Metodología CRISP-DM(Chapman et al., 2000)

La metodología Crisp–Dm (Cross Industry Standard Processfor Data Mining) es actualmente la guía de referencia más usada en proyectos de data mining, consta de cuatro niveles de abstracción, organizados de manera jerárquica en tareas que van desde un nivel general hasta los casos más específicos.

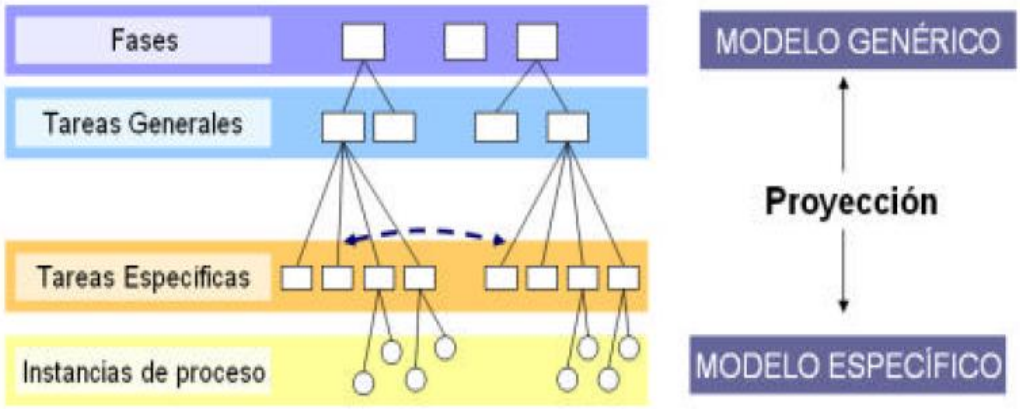



Figura 9. Modelo de la metodología CRISP-DM
Fuente: (Chapman et al., 2000)

El primer nivel son las fases que son 6 (figura 10), en donde cada una está estructurada en varias tareas generales de segundo nivel o subfases. De las tareas generales se distribuyen tareas específicas y de estas sale el cuarto nivel denominado instancias de procesos que devuelve un conjunto de acciones, decisiones o resultados del proyecto.

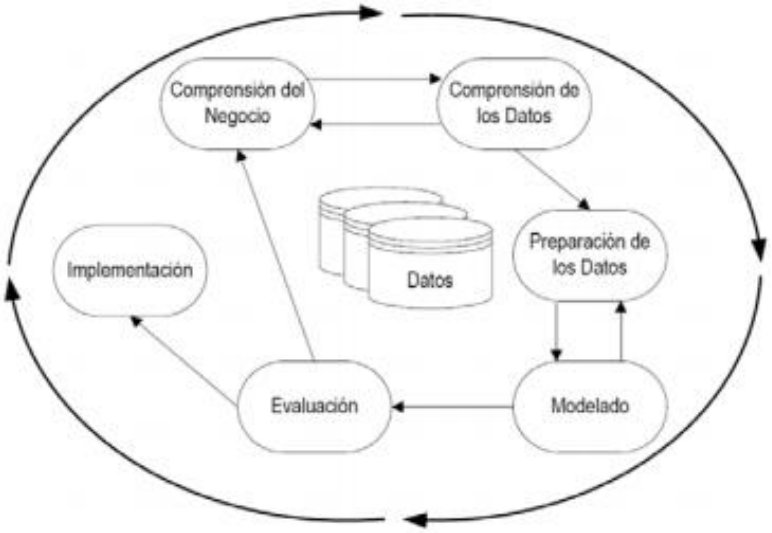



Figura 10. Fases de la metodología CRISP-DM
Fuente: (Chapman, y otros, 2000)

### *1.6.1.1* ***Descripción de las fases de la metodología CRISP-DM***

#### *1.6.1.1.1 Comprensión del negocio*

Esta fase es probablemente la más importante porque agrupa las tareas de comprensión de objetivos y requisitos del proyecto. Es importante entender completamente el problema que se desea resolver, esto permitirá recolectar datos correctos e interpretar de mejor manera los resultados.

- Determinar los objetivos del negocio
- Evaluación de la situación
- Determinar objetivos de DM

#### *1.6.1.1.2 Comprensión de los datos*

Esta fase comprende la recolección inicial de los datos, con el objetivo de tener contacto con el problema, familiarizándose, identificar la calidad y establecer relaciones que permitan definir las primeras hipótesis. Las principales tareas de esta fase son:

- Recolección de datos iniciales
- Descripción de datos
- Exploración de datos
- Verificar la calidad de los datos

#### *1.6.1.1.3 Preparación de los datos*

Una vez ejecutada la fase de comprensión de datos, se procede a la preparación para poder adaptarlos a las técnicas de minería de datos. La preparación de datos incluye a tareas generales como selección de datos, a los que se los que se les va a aplicar la limpieza. Sus principales tareas son:

- Selección de datos
- Limpieza de datos
- Estructurar los datos
- Integración de datos
- Formateo de datos

#### *1.6.1.1.4 Modelado*

En esta fase se selecciona que técnica es la adecuada para el proceso minería de datos, las técnicas se eligen de acuerdo a los siguientes parámetros:

- ✓ Ser apropiada para el problema
- ✓ Disponer de datos adecuados
- ✓ Cumplir con los requisitos del problema
- ✓ Tiempo adecuado para obtener un modelo
- ✓ Conocimiento de la técnica

Antes del modelamiento de datos, se debe elegir un método de evaluación del modelo, teniendo presente que el modelo dependerá de las características de los datos y la precisión que se quieran lograr con el modelo.

Algunas de las tareas de esta fase son:

- Selección de la técnica de modelado
- Generación del plan de prueba
- Construcción del modelo
- Evaluación del modelo

#### *1.6.1.1.5 Evaluación*

En esta fase se evalúa el modelo, teniendo en cuenta los objetivos del proyecto, si el modelado generado es válido en función de los objetivos se procede a la explotación del modelo.

- Las tareas dentro de esta fase son:
- Evaluación de resultados
- Determinación de futuras fases

## 1.7 Trabajos Relacionados

### 1.7.1 Algoritmos de Data Mining aplicados en la enseñanza basada en la Web

"El objetivo de este trabajo es presentar una recopilación de la investigación de técnicas de Data Mining aplicadas al ámbito educativo y en la enseñanza basada en la web." (Corso & Alfaro, 2011), con la finalidad de brindar a los docentes que administran y utilizan estas plataformas educativas, conocimiento relacionado con el uso por parte de los estudiantes, favoreciendo así el proceso de aprendizaje.

Las técnicas que se implementaron son algoritmos de clasificación y agrupamiento, descubrimiento de reglas de asociación y patrones de secuencia en la herramienta Weka.

Los datos obtenidos de los archivos "web log" generados por los estudiantes, han sido limpiados, de modo que se han eliminado registros como los log de las imágenes que son parte de la página HTML, de ha identificado las acciones del usuario en un determinado sitio web, con la finalidad de analizar el comportamiento.

La extracción de reglas de asociación se ha aplicado con éxito en sistemas de enseñanza web para descubrir relaciones o asociaciones entre distintas páginas web visitadas, actividades realizadas, calificaciones obtenidas, etc.

### 1.7.2 Reglas de Asociación con los datos de una biblioteca universitaria (Malberti & Elida, 2015)

La Biblioteca de la FCEFN-UNSJ, dispone del Sistema MicroISIS, en el cual se registra diariamente la circulación de sus colecciones. Los bibliotecarios además de: facilitar cada recurso librario y registrar su movimiento en el sistema de biblioteca –préstamo, renovación,

consulta o devolución, colaboran en las tareas vinculadas con el desarrollo y acceso a la colección, y participan en las actividades relacionadas con la asignación y organización del material librario en las estanterías.

Con el objetivo de promover e impulsar el uso de los recursos de la biblioteca y a la vez facilitar la labor por parte de su personal. Se aplicó el descubrimiento de reglas de asociación en una biblioteca universitaria, para proyectar la conveniencia en la disposición física del material librario. Se han planteado dos escenarios "Espacios Físicos Cerrados" y "Espacios Físicos Abiertos", para la toma de decisiones en cuanto a la disposición física del material librario, en cada uno de los escenarios mencionados.

La minería de datos si bien es cierto es muy utilizada, pero en el campo de la bibliotecología no se ha explotado su material, es por esto que con este trabajo se pretende ayudar a impulsar al personal a incursionar en las nuevas tecnologías, además de lograr un mayor aprovechamiento de los datos recolectados.

### 1.7.3 Detección de Patrones de Deserción Estudiantil en Programas de Pregrado de Instituciones de Educación Superior con CRISP-DM (Timarán & Jiménez, 2009)

En la Universidad de Nariño e Institución Universitaria CESMAG, se pretende detectar patrones de deserción estudiantil a partir de los datos socioeconómicos y académicos de los estudiantes de los programas de pregrado.
Se construyó un repositorio de datos con la información de los estudiantes que ingresaron a estas universidades entre el primer semestre de 2004 y segundo semestre de 2006, con una ventana de observación hasta el 2011.

Se seleccionaron las tareas de minería de datos clasificación, asociación y agrupamiento para descubrir conocimiento sobre deserción estudiantil. Para la tarea de asociación se utilizó el algoritmo a priori implementado en la herramienta Weka, para evaluar las reglas de asociación resultantes se utilizaron parámetros de soporte y confianza, métricas que permiten conocer la calidad de la regla.

Con el fin de generar reglas de asociación fuertes, se estableció el soporte mínimo en 3% y la confianza en 80%, se generaron 1957 reglas.

De acuerdo a los resultados, dentro de los factores asociados a la deserción estudiantil

están el ser soltero, tener un promedio bajo, haber perdido materias en los primeros semestres y provenir de un colegio público.

## CAPÍTULO 2. ANÁLISIS DE LA BASE DE DATOS

## 2.1 Modelo de base de datos – Moodle

Moodle usa una base de datos con tablas definidas, con un SQL simple, que permite el funcionamiento con una amplia variedad de motores de base de datos, como Mysql y PostgreSql (PARRA & RODRIGUEZ, 2007). En los Moodle hay alrededor de 200 tablas en su base de datos.

Los nombres de las tablas van precedidos del prefijo mdl_ más el nombre de la tabla, estos deben ser de nombre corto, sencillo y en minúscula.
A continuación se muestra aquellas tablas que están relacionadas con las acciones que realizan los estudiantes en el Moodle.

### 2.1.1 Users

Esta tabla contiene todos los datos personales de los usuarios, además contiene datos como nombre de usuario, contraseña, hora del primer y último acceso, el tiempo que duró la sesión.

### 2.1.2 Roles

Moodle para poder diferenciar entre los tipos de usuarios, tiene un sistema de roles en función de permisos que se le puede otorgar al usuario, de acuerdo a este rol el usuario tendrá la capacidad de realizar acciones en el sitio.
La siguiente tabla contiene toda la información sobre los diferentes roles existentes

Tabla 4. Descripción de tabla mdl_role

| Nombre de la tabla: | | **mdl_role** |
|---|---|---|
| Atributo | **Tipo de dato** | **Descripción** |
| **Id** | Bigint(10) | Identificador del rol |
| **Name** | Varchar(100) | Nombre completo del rol |
| **Shortname** | Varchar(100) | Nombre corto del rol |
| **Description** | Text | Descripción de cada rol |

Fuente. (Garg & Neilsen, 2012)

La tabla mdl_role tiene relación con la tabla mdl_role_assignments y mdl_context ya que en estas se almacena información sobre que contexto tiene asignación el usuario; los contextos

establecen los distintos niveles de capacidades de un usuario en orden jerárquico, es decir, si un usuario tiene la capacidad en el nivel más alto de jerarquía, heredará los permisos de niveles inferiores.(Garg & Neilsen, 2012).

Tabla 5. Descripción de la tabla md_role_assignments y mdl_context

| Nombre de la tabla: | | **mdl_role_assignments** |
|---|---|---|
| **Atributo** | **Tipo de dato** | **Descripción** |
| **Userid** | int(10) | Identificador del usuario |
| **Roleid** | Int(10) | Identificar del rol |
| **Nombre de la tabla:** | | **mdl_context** |
| **Instanceid** | Int(10) | Identificador relacionado con el id de un curso |
| **contextlevel** | Int(10) | Nivel de contexto (50) |

Fuente. (Garg & Neilsen, 2012)

### 2.1.3 Course

En esta sección van alojados los diversos cursos que están compuestos por las categorías de los cursos, es decir la tabla mdl_course contiene esta información, tiene 32 campos de los cuales solo se usan pocos, los campos que no se utilizan son para su uso futuro.
Esta tabla es solamente de lectura y en esta se puede recuperar los cursos en los que un estudiante está matriculado. Los camposutilizados son: id, category, fullname, shortname and summary. (Ramagiri & Jeffery, n.d.)

Tabla 6. Descripción de la tabla mdl_course

| Nombre de la tabla: | | **Mdl_course** |
|---|---|---|
| **Atributo** | **Tipo de dato** | **Descripción** |
| **Id** | Int(10) | Es un campo de incremento automático y se utiliza como clave principal, Además es el identificador del curso |
| **Category** | Int(20) | Identificador de la categoría del curso |
| **Fullname** | varchar(254) | Es el nombre del curso |
| **Shortname** | varchar(15) | Es el número de llamada del curso |
| **Summary** | text(50) | El resumen es la descripción del curso |

Fuente. (Ramagiri & Jeffery, n.d.)

Las categorías están compuestas por información de cursos, sirven para organizar de

manera que sea más fiable la búsqueda de dichos cursos. (Gonzáles, n.d.)

### 2.1.4 Logs

Los logs usualmente guardan las acciones que realizan los usuarios dentro del Moodle, estas acciones se consiguen de acuerdo al identificador del usuario, el curso al que pertenece y una actividad o recurso específico. (Sael, Marzak, & Behja, 2013)
Estos ficheros guardan información de las visitas que realiza el usuario al sitio.

Tabla 7. Descripción de la tabla mdl_log

| Nombre de la tabla: | | **Mdl_log** |
|---|---|---|
| **Atributo** | **Tipo de dato** | **Descripción** |
| **Userid** | Begint(10) | Identificador del usuario |
| **Course** | Begint(10) | Número que identifique el curso |
| **Module** | Varchar(20) | Módulo o sitio que el usuario realiza la acción |
| **Action** | Varchar(20) | Acción que realiza el usuario en un módulo |

Fuente. (Ramagiri & Jeffery, n.d.)

### 2.1.5 Foro.

Es una herramienta de comunicación asíncrona en los Moodle, funciona como una pizarra virtual en donde los estudiantes y docentes pueden interactuar por medio de mensajes y respondiendo estos, creando hilos de conversación. Para mantener la interacción no es necesario estar online al igual que la otra persona, se puede adjuntar archivos externos, y ofrece la posibilidad de suscribirse a foros de su interés, recibiendo por correo electrónico cada nueva interacción.

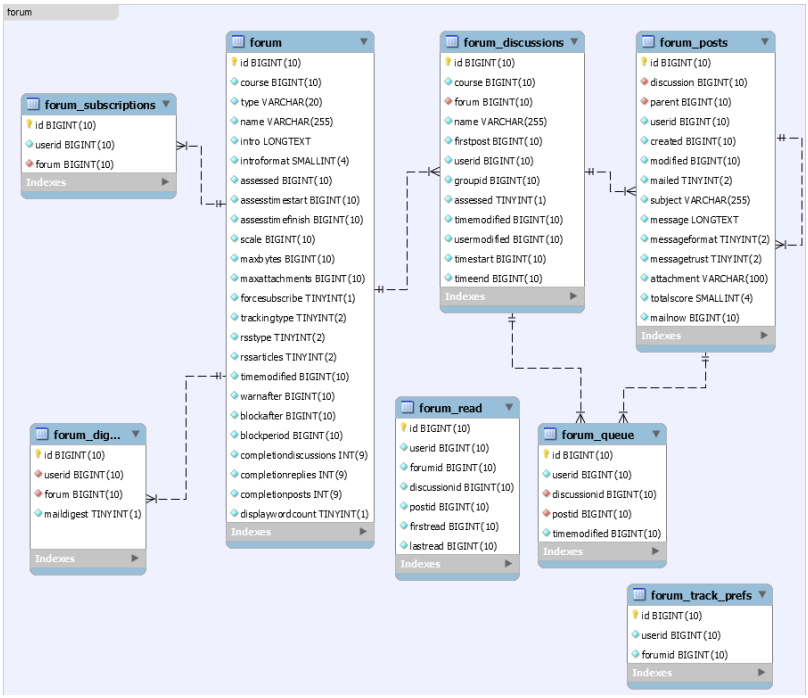



Figura 11. Diagrama E/R Foro

Fuente: (PARRA & RODRIGUEZ, 2007)

Si se desea tener información acerca del módulo foros hay 8 tablas que poseen información que se necesita.

- Mdl_forum
- Mdl_forum_discussions
- Mdl_forum_post
- Mdl_forum_read
- Mdl_forum_queue
- Mdl_forum_track_prefs
- Mdl_forum_suscriptions
- Mdl_forum_ratings

Tabla 8. Descripción de la tabla mdl_forum

| Nombre de la tabla | **mdl_forum** | |
|---|---|---|
| Atributo | **Tipo de dato** | **Descripción** |
| **id** | int(10) | Identificador de la tabla(primary key) |
| **course** | int(10) | Índice que se refiere a la tabla mdl_course |
| **type** | Enum(‘single’, ’news’,’general’, ‘social’,’eachuser’, | Tipo de foro |

| Nombre de la tabla | | **mdl_forum** |
|---|---|---|
| | 'teacher') | |
| **name** | varchar(255) | Titulo o nombre del foro |
| **intro** | text | Texto introductorio del foro |
| **open** | tinyint(2) | Permitir que los estudiantes abran nuevos temas |
| **assessed** | int(10) | Acciones usuario, modo calificación |
| **assesspublic** | int(4) | Modo de vista de calificaciones |
| **assesstimestart** | int(10) | Fecha inicial restricción de calificaciones |
| **assesstimefinish** | int(10) | Fecha final restricción de calificaciones |
| **scale** | int(10) | Escala de calificaciones |
| **maxbytes** | int(10) | Tamaño máximo del archivo adjunto |
| **forcesuscribe** | tinyint(1) | Forzar suscripción al foro |
| **trackingtype** | tinyint(2) | Leer rastreo de foro |
| **rsstype** | tinyint(2) | Tipos de registros |
| **rssarticles** | tinyint(2) | Registro de artículos |
| **timemodified** | int(10) | Ultima fecha de modificación del foro |

Fuente. (PARRA & RODRIGUEZ, 2007)

### 2.1.6 Chat.

Esa es una herramienta de comunicación síncrona que posee el Moodle, permite a los usuarios interactuar en tiempo real, para poder comunicarse es necesario tener una sala de chat que es creada por el docente o encargado.

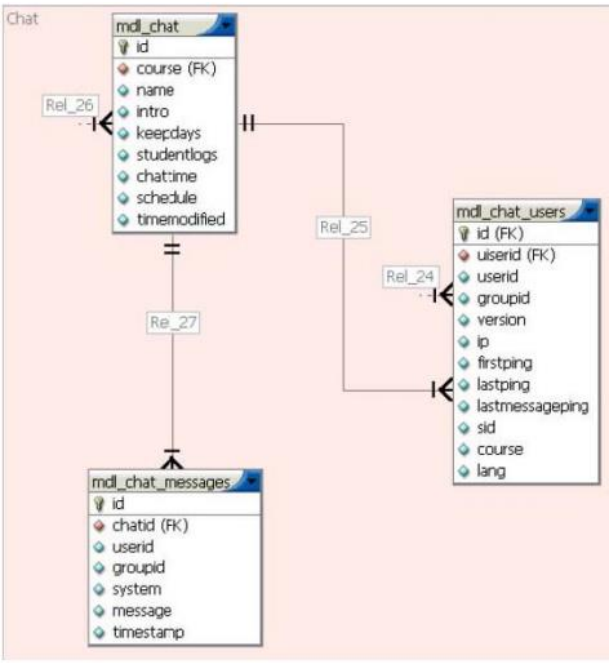



Figura 12. Diagrama E/R del módulo chat

Fuente: (PARRA & RODRIGUEZ, 2007)

En el módulo chat existen 4 tablas tales como:

- Mdl_chat: cada una es una diferente sala de chat.
- mdl_chat_messages: todos los mensajes de chat reales.
- mdl_chat_messages_current
- mdl_chat_users: mantiene un registro de cada uno de los participantes que están en las salas de chat.

A continuación se detalla atributos de la tabla mdl_chat que es una de las principales.

Tabla 9. Descripción de tabla mdl_chat

| Nombre de la tabla: | **mdl_chat** | |
|---|---|---|
| Atributo | **Tipo de dato** | **Descripción** |
| **id** | int(10) | Identificador de la tabla PRIMARY KEY |
| **course** | int(10) | Índice que refiere a la tabla mdl_course |
| **name** | varchar(255) | Titulo o nombre de la sala |
| **intro** | text | Introducción a la conversación |
| **keepdays** | int(11) | Guardar sesiones pasadas |
| **studentlogs** | int(4) | Todos pueden ver sesiones pasadas |
| **chattime** | int(10) | Próxima cita |

| | | |
|---|---|---|
| **shedule** | int(4) | Repetidor de sesiones |
| **timemodified** | int(10) | Ultima fecha de modificación |

Fuente. (PARRA & RODRIGUEZ, 2007)

## 2.2 Modelo de la Base de Datos - UTPL

### 2.2.1 Tablas relacionadas con las acciones del estudiante

En esta sección se detallan aquellas tablas que están relacionadas con las acciones o comportamientos que realizan los estudiantes de la modalidad a distancia de la Universidad Técnica Particular de Loja cuando interactúan con actividades como foro, chat y video-colaboración que están alojadas en el Entorno Virtual de aprendizaje.

Mediante el análisis a la Base de datos de la UTPL, se ha determinado que las tablas que almacenan datos con respecto a las acciones en el Eva son las siguientes.

Tabla 10. Tablas del EVA para adquisición de datos

| Nombre de la Tabla | Descripción |
|---|---|
| mdl_syllabus_utpl | Información del sistema académico |
| mdl_syllabus_pdo | Información del periodo académico |
| mdl_log | Registra cada una de las interacciones del usuario con el EVA |
| mdl_enrol_utpl | Información sobre la matrícula del estudiante |
| mdl_role_assignments | Información referente a los tipos de roles que existe en el sistema |
| mdl_user_utpl | Información sobre los estudiantes |
| mdl_context | Muestra los diferentes contextos que puede tener el usuario |
| mdl_course | Información de cada curso existente |
| mdl_course_categories | Muestra la categoría que tienen los cursos |

Fuente. Base de datos- UTPL

Elaboración: propia

Una vez identificadas las relaciones entre las diversas tablas, se procedió a la extracción de datos, ya que nos permiten encontrar patrones de comportamiento, que nos ayudan a saber cuáles son las acciones que hacen los estudiantes de modalidad a distancia cuando navegan en el Eva en actividades como: foro, chat, video-colaboración.

### 2.2.2 Atributos relaciones con las acciones del estudiante en el EVA

Cada tabla contiene atributos importantes y relevantes que reflejan las actividades que realizan los estudiantes de modalidad a distancia, en la tabla 6 se muestran dichos atributos que son requeridos para el proceso de minería de datos. Una vez definido los atributos, se podrá obtener mediante consultas SQL toda la información relacionada con las acciones del usuario en actividades (foro, chat, video-colaboración) del EVA durante su interacción.

Tabla 11. Tablas y atributos con acciones sobre el Eva

| Tabla | Atributo | Descripción |
|---|---|---|
| **mdl_user_utpl** | ✓ Userid<br>✓ Sexo | Esta tabla contiene datos como el género del usuario |
| **Mdl_syllabus_utpl** | ✓ courseid<br>✓ componente<br>✓ paralelo<br>✓ pdoid | Contiene atributos de los cursos dictados de los diversos periodos académicos. |
| **mdl_syllabus_pdo** | ✓ id<br>✓ name | Almacena los datos de cada periodo académico |
| **mdl_course** | ✓ id<br>✓ category | Almacena los cursos de las diversas categorías existentes |
| **mdl_course_categories** | ✓ id | Diversas categorías de cursos que existen dictados en cierta modalidad de estudios |
| **mdl_log** | ✓ userid<br>✓ course<br>✓ module<br>✓ action | Datos del usuario dentro del Eva, en donde se puede identificar las diversas acciones que se realiza. |
| **mdl_log_display** | ✓ module<br>✓ action | Se encuentran los diferentes tipos de acciones por cada uno de los módulos. |
| **Mdl_enrol_utpl** | ✓ courseid<br>✓ pdoid | Se encuentran datos referentes al curso, periodo correspondiente y rol del estudiante |
| **mdl_role_assignments** | ✓ userid<br>✓ contextid<br>✓ roleid | Indica que el usuario tenga rol de estudiante |
| **mdl_context** | ✓ id<br>✓ instanceid | Indica en qué contexto se encuentra el usuario |

| | | |
|---|---|---|
| | ✓ contextlevel | |

Fuente. Base de datos- UTPL

Elaboración: propia

En la tabla mdl_user_utpl se ha creído conveniente extraer solamente el identificador y el género, ya que por cuestiones de seguridad no se tiene acceso a los demás datos.

Con lo visto en la tabla 6, lo que se pretende es la extracción de datos más relevantes dentro del sistema, esto permitirá saber:

- El número de materias que cursan los estudiantes
- Número de acciones que realizan los estudiantes en los módulos de chat, foro, video colaboración
- Nombres de cursos y paralelos
- Carrera a la cual pertenecen los cursos
- Tipo de materia que cursan
- Centro UTPL al que pertenecen.

Una vez que se posean estos datos, mediante la tabla mdl_log, que es en donde se aloja toda la información referente a acciones de usuarios, se procederá al respectivo análisis de las acciones realizadas por los estudiantes de modalidad a distancia. Únicamente se extraerán los datos con los que se pueda trabajar de manera eficiente en la minería de datos.

### 2.2.3 Tipos de acciones en el EVA

Dentro de las actividades en línea como lo son foros, chat y video colaboraciones existen diversos tipos de acciones que hacen referencia a lo que los estudiantes realizan cuando interactúan con el Eva. En la siguiente tabla se las detalla.

Tabla 12. Tipos de acciones según actividades

| Modulo | Acción | Descripción |
|---|---|---|
| **Chat** | historial | Veces que el usuario revisa el historial de un determinado chat. |
| | Talk | Veces que el usuario tiene una conversación |
| | Report | Veces que el Usuario ve los reportes de todos |

| | | |
|---|---|---|
| | | los chats |
| | view | Veces que el usuario revisa una actividad planeada |
| | viewall | Veces que el usuario revisa los chat de todas las asignaturas |
| **Elluminate (Videocolaboración)** | view | Ver la sala en la que se da la video colaboración |
| | viewall | Ver todos los usuarios que participan en las video colaboraciones |
| | view meeting | Veces que el usuario ingresa a una video colaboración |
| | viewrecording | Ver grabaciones de las video colaboraciones |
| **Fórum** | add post | El usuario da una respuesta a algo preguntado por el docente |
| | Delete Post | Veces que el usuario elimina un post que ha realizado en el foro |
| | Search | Veces que el usuario busca foros |
| | subscribe | Se suscribe a un nuevo foro o debate. |
| | Subscribe all | Veces que el usuario se subscribe a todos los foros existentes |
| | unsubscribe | El usuario de suscribe de cierto foro |
| | update post | Veces que el usuario actualizó su respuesta en el foro |
| | viewdiscussion | Veces que el usuario revisa las discusiones dentro de un foro |
| | viewforum | Veces que el usuario ingresa a revisar un foro |
| | viewforums | Veces que el usuario ingresa a ver todos los foros |
| | viewsubscribers | Ver suscripciones que posee el usuario |

Únicamente de la tabla mdl_log se extrae la interacción que realiza el estudiante con el EVA. Las demás tablas se las usa para obtener información relevante del estudiante.

# CAPÍTULO 3. MINERIA DE DATOS

**Introducción**

Para realizar el proceso de minería de datos se ha creído conveniente usar la metodología CRISP-DM, como se vio en el capítulo 1.6 es una de las metodologías más usadas según la encuesta aplicada por Kdnuggest a sus usuarios en el año 2014 y la que mejor se adapta a nuestro problema.

Al proceso de minería se lo llevará a cabo usando el lenguaje de programación la R Project con la herramienta de programación R Studio[9] en donde se ejecutará la técnica de minería de datos, obteniendo así un modelo descriptivo el que nos permitirá encontrar los patrones de comportamiento de los estudiantes.

## 3.1 **Problemática**

Actualmente en las instituciones educativas se han adaptado entornos de aprendizaje virtual, los cuales pretenden ser beneficiosos para el aprendizaje de los diferentes estudiantes ya que con este se pretende cubrir o satisfacer algunas dudas mediante la interacción estudiante-docente y viceversa con actividades en línea.

En la UTPL, se ha implementado desde hace varios ciclos la realización de actividades en línea (foro, chat y video colaboración), en el proceso de enseñanza aprendizaje de Modalidad a Distancia y una vez que han transcurrido varios semestres de la aplicación de esta práctica, surge la necesidad de evaluar el impacto que han tenido estas actividades en el rendimiento académico de los estudiantes además de evaluar ciertos comportamientos, en general establecer si hay relaciones entre las diferentes variables involucradas y sobre todo como se puede mejorar los procesos de aprendizaje.

Es por esto que mediante la aplicación de técnicas de minería de datos se pretende determinar y evaluar patrones de comportamiento de los estudiantes de la modalidad a Distancia de la UTPL, además de evaluar el impacto que ha tenido la realización de actividades como foros, chats y video colaboración que el Entorno Virtual de Aprendizaje (EVA) ofrece.

Luego de analizar la problemática, se han definido todas las interacciones de las actividades que realizan los estudiantes en el EVA. Además de estas variables se obtiene información de los estudiantes como son:

- Genero
- Número de materias cursadas
- Centro UTPL

[9] R Studio: https://www.rstudio.com/

También existen variables que definen en que componentes se efectuaron las interacciones, por ejemplo.

- Nombre de la materia
- Paralelo
- Titulación a la que pertenece
- Tipo de materia
- Área a la que pertenece la Titulación

En cuanto a las variables que reflejan las interacciones de los estudiantes en los módulos de: chats, foros y videocolaboraciones son las siguientes.

- historial
- Talk
- Report
- view
- viewall
- view
- viewall
- view meeting
- viewrecording
- add post
- Delete Post
- Search
- Subscribe
- Subscribe all
- unsubscribe
- update post
- viewdiscussion
- viewforum
- viewforums
- viewsubscribers

## 3.2 Fase I: Comprensión del negocio

### 3.2.1 Determinar los objetivos del negocio

#### *3.2.1.1 Contexto*

Para la solución del proyecto, como población se ha elegido en primera instancia la modalidad a distancia de la UTPL, con materias troncales y genéricas de primer ciclo de la titulación del área biológica Gestión ambiental en periodo académico Abril 2015- Agosto 2015.

En segunda instancia se decidió cambiar la muestra, igualmente se tomaron en cuenta las materias troncales y genéricas de primer ciclo de modalidad a distancia en el mismo periodo, pero se ha considerado utilizar registros de todas las titulaciones:

- Derecho
- Comunicación Social

- Psicología
- Ciencias de la educación mención en:
  - Educación básica
  - Educación infantil
  - Física y Matemática
  - Inglés
  - Lengua y Literatura
  - Química y Biología
- Administración de Empresas
- Administración en Gestión Pública
- Administración en Banca y Finanzas
- Administración de Empresas Turísticas y Hoteleras
- Contabilidad y Auditoría
- Economía
- Asistencia Gerencial y Relaciones Públicas
- Informática
- Gestión Ambiental

#### *3.2.1.2 Objetivos de negocio*

La UTPL proporciona a sus estudiantes el entorno virtual de aprendizaje (EVA) en el cual existen actividades online como foros, chats y videocolaboraciones, los cuales permiten a los estudiantes:

- Complementar temas de estudio
- Permite interacción de alumnos con alumnos y de alumnos con docentes.

### 3.2.2 Evaluación de la situación

Una vez analizada la problemática se ha podido observar cual es la situación actual del problema, permitiendo encontrar los requerimientos para la ejecución del mismo.

#### *3.2.2.1 Requisitos*

Para la ejecución del proyecto lo que primeramente se necesita tener acceso a la base de datos de la UTPL, para por medio de esta continuar con las fases siguientes.

#### *3.2.2.2 Restricciones*

No se puede tener acceso a los datos personales de los estudiantes debido a que la

información almacenada en la base de datos de la universidad es confidencial.

#### *3.2.2.3 Inventario de recursos*

Los recursos a utilizar para la ejecución del proyecto, son de tipo software.

Tabla 13. Software a utilizar

| Herramienta | Uso |
|---|---|
| **XAMPP** | Servidor para el levantamiento del gestor de base de datos. |
| **MySQL WORKBECH** | Usada para consulta y extracción de información de la base de datos |
| **OPEN REFINE** | Tratamiento, limpieza y transformación de datos |
| **R STUDIO** | Transformación de datos y aplicación de técnicas de DM |

Elaboración: propia

#### *3.2.2.4 Terminología*

**EVA**: entorno virtual de aprendizaje.

**UTPL**: Universidad Técnica Particular de Loja.

**Titulaciones**: Carreras ofertadas por la universidad

### 3.2.3 Determinar los objetivos de minería

- Diseñar un modelo descriptivo de las actividades que se efectúan en el EVA: Chat, Foro, Video colaboración.
- Identificar los patrones de comportamiento relevantes y que expliquen de manera significativa las relaciones implícitas en los procesos de interactividad en el EVA

### 3.2.4 Plan del proyecto

Para la ejecución del proyecto se estimó 12 meses, distribuidos en varias componentes, a continuación la descripción.

Tabla 14. Plan de proyecto

| Componentes | Plazo(Meses) |
|---|---|
| **Estado del arte** | 2.5 |
| **Adquisición de datos** | 1.5 |
| **Aplicar técnicas de Pre-procesamiento, limpieza y selección de datos** | 2 |

| Componentes | Plazo(Meses) |
|---|---|
| **Selección de herramientas de Data Mining a utilizar** | 2 |
| **Ejecutar las experimentaciones planificadas** | 2 |
| **Evaluación de resultados** | 2 |

Fuente: propia

## 3.3 Fase II: Comprensión de los datos

### 3.3.1 Recolección de datos

Para la presente tarea se usará información referente a las interacciones en el EVA de los estudiantes de modalidad a distancia, la misma que se encuentra almacenada en la base de datos de la universidad.

Una vez obtenido el acceso a la base de datos UTPL, se procedió a analizar y buscar la información necesaria para el proceso de minería según nuestro problema.

### 3.3.2 Descripción de datos

Los datos necesarios se los encontró mediante consultas SQL en varias tablas relacionadas de la base de datos.

En total se usaron 9 tablas relacionados, como:

- mdl_syllabus_utpl
- mdl_syllabus_pdo
- mdl_log
- mdl_enrol_utpl
- mdl_role_assignments
- mdl_user_utpl
- mdl_context
- mdl_course
- mdl_course_categories

De cada una de las tablas se usaron varios atributos, los cuales se detallan claramente en el Capítulo 2.2.2.

Algunos atributos se los uso netamente para la relación entre las diversas tablas, más no para el dataset final.

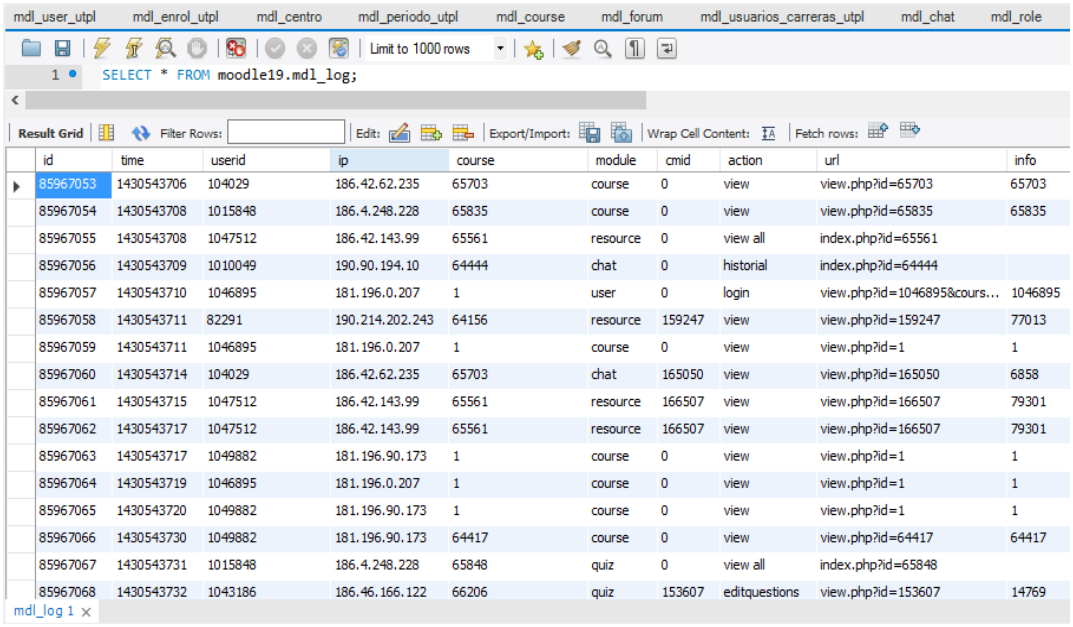
## 3.4 Fase III: Preparación de datos

### 3.4.1 Selección de datos

El primer paso para la extracción de datos, consiste en seleccionar los datos necesarios para realizar la minería, para la ejecución de este análisis, la información se encuentra en la

tabla mdl_log y demás tablas relacionadas. Dentro de esta tabla, se encuentran todas las interacciones que realizan los estudiantes cuando ingresan al EVA, es decir se registra cada acción realizada por el estudiante en la plataforma.

Para la extracción de datos se utilizó MySQL WORKBENCH, que permitió ver y analizar cada una de las tablas que contiene la información necesaria, los principales datos que fueron analizados, son los que están alojados en la tabla mdl_log como lo muestra la siguiente figura.


mdl_user_utpl mdl_enrol_utpl mdl_centro mdl_periodo_utpl mdl_course mdl_forum mdl_usuarios_carreras_utpl mdl_chat mdl_role

Limit to 1000 rows

```
1  SELECT * FROM moodle19.mdl_log;
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content: | Fetch rows:

| id | time | userid | ip | course | module | cmid | action | url | info |
|---|---|---|---|---|---|---|---|---|---|
| 85967053 | 1430543706 | 104029 | 186.42.62.235 | 65703 | course | 0 | view | view.php?id=65703 | 65703 |
| 85967054 | 1430543708 | 1015848 | 186.4.248.228 | 65835 | course | 0 | view | view.php?id=65835 | 65835 |
| 85967055 | 1430543708 | 1047512 | 186.42.143.99 | 65561 | resource | 0 | view all | index.php?id=65561 | |
| 85967056 | 1430543709 | 1010049 | 190.90.194.10 | 64444 | chat | 0 | historial | index.php?id=64444 | |
| 85967057 | 1430543710 | 1046895 | 181.196.0.207 | 1 | user | 0 | login | view.php?id=1046895&cours... | 1046895 |
| 85967058 | 1430543711 | 82291 | 190.214.202.243 | 64156 | resource | 159247 | view | view.php?id=159247 | 77013 |
| 85967059 | 1430543711 | 1046895 | 181.196.0.207 | 1 | course | 0 | view | view.php?id=1 | 1 |
| 85967060 | 1430543714 | 104029 | 186.42.62.235 | 65703 | chat | 165050 | view | view.php?id=165050 | 6858 |
| 85967061 | 1430543715 | 1047512 | 186.42.143.99 | 65561 | resource | 166507 | view | view.php?id=166507 | 79301 |
| 85967062 | 1430543717 | 1047512 | 186.42.143.99 | 65561 | resource | 166507 | view | view.php?id=166507 | 79301 |
| 85967063 | 1430543717 | 1049882 | 181.196.90.173 | 1 | course | 0 | view | view.php?id=1 | 1 |
| 85967064 | 1430543719 | 1046895 | 181.196.0.207 | 1 | course | 0 | view | view.php?id=1 | 1 |
| 85967065 | 1430543720 | 1049882 | 181.196.90.173 | 1 | course | 0 | view | view.php?id=1 | 1 |
| 85967066 | 1430543730 | 1049882 | 181.196.90.173 | 64417 | course | 0 | view | view.php?id=64417 | 64417 |
| 85967067 | 1430543731 | 1015848 | 186.4.248.228 | 65848 | quiz | 0 | view all | index.php?id=65848 | |
| 85967068 | 1430543732 | 1043186 | 186.46.166.122 | 66206 | quiz | 153607 | editquestions | view.php?id=153607 | 14769 |

mdl_log 1

Figura 13. Tabla mdl_log
Elaboración: propia

Debido a la gran cantidad de información que posee la base de datos de la Universidad Técnica Particular de Loja, como se mencionó en la fase I, se ha creído conveniente extraer una muestra de las interacciones que realizan los usuarios dentro del EVA, en la carrera de Gestión Ambiental en las materias de primer ciclo en el periodo académico Abril 2015-Agosto 2015.

| idUsuario | Genero | MateriasMatri | Carrera | idCurso | Centro | materia | Modulo | Actividad | NumAcciones |
|---|---|---|---|---|---|---|---|---|---|
| 6892 | M | 2 | Gestion Ambiental | 64993 | QUITO | BIOLOGÍA GENERAL | forum | view forum | 2 |
| 49984 | F | 2 | Gestion Ambiental | 64993 | QUITO | BIOLOGÍA GENERAL | chat | historial | 1 |
| 49984 | F | 2 | Gestion Ambiental | 64993 | QUITO | BIOLOGÍA GENERAL | chat | talk | 13 |
| 49984 | F | 2 | Gestion Ambiental | 64993 | QUITO | BIOLOGÍA GENERAL | chat | view | 2 |
| 49984 | F | 2 | Gestion Ambiental | 64993 | QUITO | BIOLOGÍA GENERAL | chat | view all | 2 |
| 49984 | F | 2 | Gestion Ambiental | 64993 | QUITO | BIOLOGÍA GENERAL | elluminate | view | 23 |
| 49984 | F | 2 | Gestion Ambiental | 64993 | QUITO | BIOLOGÍA GENERAL | elluminate | view all | 18 |
| 49984 | F | 2 | Gestion Ambiental | 64993 | QUITO | BIOLOGÍA GENERAL | elluminate | view meeting | 25 |
| 49984 | F | 2 | Gestion Ambiental | 64993 | QUITO | BIOLOGÍA GENERAL | forum | add post | 1 |
| 49984 | F | 2 | Gestion Ambiental | 64993 | QUITO | BIOLOGÍA GENERAL | forum | update post | 1 |
| 49984 | F | 2 | Gestion Ambiental | 64993 | QUITO | BIOLOGÍA GENERAL | forum | view forum | 10 |
| 49984 | F | 2 | Gestion Ambiental | 64993 | QUITO | BIOLOGÍA GENERAL | forum | view forums | 13 |
| 54432 | F | 2 | Gestion Ambiental | 64993 | QUITO-SAN RAFAEL | BIOLOGÍA GENERAL | chat | talk | 26 |
| 54432 | F | 2 | Gestion Ambiental | 64993 | QUITO-SAN RAFAEL | BIOLOGÍA GENERAL | chat | view | 4 |
| 55260 | M | 2 | Gestion Ambiental | 64993 | QUITO-TURUBAMBA | BIOLOGÍA GENERAL | chat | view | 1 |
| 55260 | M | 2 | Gestion Ambiental | 64993 | QUITO-TURUBAMBA | BIOLOGÍA GENERAL | elluminate | view | 1 |
| 94311 | F | 2 | Gestion Ambiental | 64993 | MACHALA | BIOLOGÍA GENERAL | chat | historial | 1 |
| 94311 | F | 2 | Gestion Ambiental | 64993 | MACHALA | BIOLOGÍA GENERAL | chat | view | 3 |

Figura 14. Muestra #1 de la extracción de datos

Fuente: Base de datos UTPL
Elaboración: propia

Luego de realizar las siguientes fases de data mining se optó por la extracción de más datos, por el motivo de que la muestra obtenida, fue muy pequeña para el total de la población.

Para la nueva muestra, como se mencionó anteriormente se realizó la adquisición de datos de las materias troncales y genéricas de primer ciclo de todas las titulaciones que brinda dicha universidad en su modalidad a distancia, en el periodo académico Abril 2015- Agosto 2015.

| idUsuario | Genero | MateriasMatricula | Carrera | Area | TipoMater | Centro | Materia | paralelo | Herramienta | Actividad | NumAccione | periodo |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1046348 | F | 2 | Economia | Administrativa | Troncal | ALAMOR | HISTORIA DEL PENSAMIENTO ECONÓMICO | A1 | chat | talk | 3 | Abr/2015 - Ago/2015 |
| 1046348 | F | 2 | Economia | Administrativa | Troncal | ALAMOR | HISTORIA DEL PENSAMIENTO ECONÓMICO | A1 | chat | view | 1 | Abr/2015 - Ago/2015 |
| 1046348 | F | 2 | Economia | Administrativa | Troncal | ALAMOR | HISTORIA DEL PENSAMIENTO ECONÓMICO | A1 | chat | view all | 1 | Abr/2015 - Ago/2015 |
| 1046348 | F | 2 | Economia | Administrativa | Troncal | ALAMOR | HISTORIA DEL PENSAMIENTO ECONÓMICO | A1 | elluminate | view | 2 | Abr/2015 - Ago/2015 |
| 1046348 | F | 2 | Economia | Administrativa | Troncal | ALAMOR | HISTORIA DEL PENSAMIENTO ECONÓMICO | A1 | elluminate | view all | 1 | Abr/2015 - Ago/2015 |
| 1046348 | F | 2 | Economia | Administrativa | Troncal | ALAMOR | HISTORIA DEL PENSAMIENTO ECONÓMICO | A1 | elluminate | view recording | 1 | Abr/2015 - Ago/2015 |
| 1046348 | F | 2 | Economia | Administrativa | Troncal | ALAMOR | HISTORIA DEL PENSAMIENTO ECONÓMICO | A1 | forum | view forum | 2 | Abr/2015 - Ago/2015 |
| 1046348 | F | 2 | Economia | Administrativa | Troncal | ALAMOR | HISTORIA DEL PENSAMIENTO ECONÓMICO | A1 | forum | view forums | 3 | Abr/2015 - Ago/2015 |
| 1046348 | F | 2 | Economia | Administrativa | Generica | ALAMOR | MATEMÁTICA | A1 | chat | historial | 1 | Abr/2015 - Ago/2015 |
| 1046348 | F | 2 | Economia | Administrativa | Generica | ALAMOR | MATEMÁTICA | A1 | chat | talk | 2 | Abr/2015 - Ago/2015 |
| 1046348 | F | 2 | Economia | Administrativa | Generica | ALAMOR | MATEMÁTICA | A1 | chat | view | 2 | Abr/2015 - Ago/2015 |
| 1046348 | F | 2 | Economia | Administrativa | Generica | ALAMOR | MATEMÁTICA | A1 | chat | view all | 1 | Abr/2015 - Ago/2015 |
| 1046348 | F | 2 | Economia | Administrativa | Generica | ALAMOR | MATEMÁTICA | A1 | forum | add post | 1 | Abr/2015 - Ago/2015 |
| 1046348 | F | 2 | Economia | Administrativa | Generica | ALAMOR | MATEMÁTICA | A1 | forum | view forum | 9 | Abr/2015 - Ago/2015 |

Figura 15. Muestra #2 de la extracción de datos

Fuente: Base de datos UTPL

Elaboración: propia

Con titulaciones como:

- Asistencia Gerencial y Relaciones Públicas
- Administración de Empresas
- Administración en Gestión Pública
- Administración en Banca y Finanzas
- Administración de Empresas Turísticas y Hoteleras
- Contabilidad y Auditoría
- Ciencias de la educación mención Física y Matemática
- Ciencias de la educación mención Inglés
- Ciencias de la educación mención Lengua y Literatura

, no se obtuvieron registros con las herramientas de foros, chats y video colaboración.

Cabe destacar que en la tabla mdl_log posee diversas actividades de cada uno de los módulos (foros, chats, Videocolaboración), para este proceso no se usarán algunas de ellas, ya que son acciones que el estudiante pocamente usa. (Samaniego, 2016) En la siguiente tabla, se puede observar las variables que no se las usará en el proceso.

Tabla 15. Interacciones no usadas

| Modulo | Acción |
|---|---|
| Chat | Report |
| | view |
| | viewall |
| Fórum | Delete Post |
| | Search |
| | Subscribe all |
| | viewforums |
| | unsubscribe |

Elaboración: propia

En el módulo chat, no se usará las variables report, view, viewall, ya que no brinda información de cuantas veces el estudiante interactúa en el chat, las únicas variables que obtiene esta información es **chatTalk e historial.**

Mientras que en el módulo foro, una de las variables que demuestra la interacción que realiza el estudiante es **addpost, updatepost, viewforum,** las demás variables Delete Post, Search, viewforums, Subscribe all y unsubscribe, no aportan con información valiosa. En el anexo 2, se muestra la consulta SQL para la selección de los datos.

En el módulo de video colaboración se utilizará las cuatro variables: view, viewall, view meeting y viewrecording.

### 3.4.2 Limpieza de datos

Luego de la extracción de los datos se procedió a preparar los mismos, ya que si estos no son de calidad pueden entorpecer el descubrimiento de patrones de comportamiento, siendo así el proceso inútil.

El objetivo de la preparación de datos es obtener una vista minable, es decir que los datos incluyan variables de interés para el problema en concreto en el formato adecuado.
Para tener el conjunto de datos minable, se realizó la conversión de filas por columnas, es decir que un determinado estudiante, presente en una sola fila cada una de las variables con su respectivo valor. (Ver anexo 3)

| idUsuario | Genero | Carrera | TipoMateria | Centro | Materia | chat_historial | chat_report | chat_talk | chat_viewAllChat | chat_viewChat |
|---|---|---|---|---|---|---|---|---|---|---|
| 1046348 | F | Economia | Troncal | ALAMOR | HISTORIA DEL PENSAMIENTO ECONÓMICO | 0 | 0 | 3 | 1 | 1 |
| 1046348 | F | Economia | Generica | ALAMOR | MATEMÁTICA | 1 | 0 | 2 | 1 | 2 |

Figura 16. Vista Minable de los datos

Elaboración: Propia

En esta fase de limpieza de datos, el objetivo es corregir datos inconsistentes, identificar valores erróneos, eliminar registros duplicados, adecuar valores faltantes, eliminación de espacios o caracteres especiales. Para asegurar la calidad de los datos es necesario verificar cada uno de los datos antes mencionados.

Es por esto que se ha utilizado la herramienta **Open Refine**[10]**,** que sirve para la limpieza y transformación de datos. Antes de empezar con la limpieza y transformación de los datos se tenía un total de 11030 registros.

[10] Open Refine: http://openrefine.org/

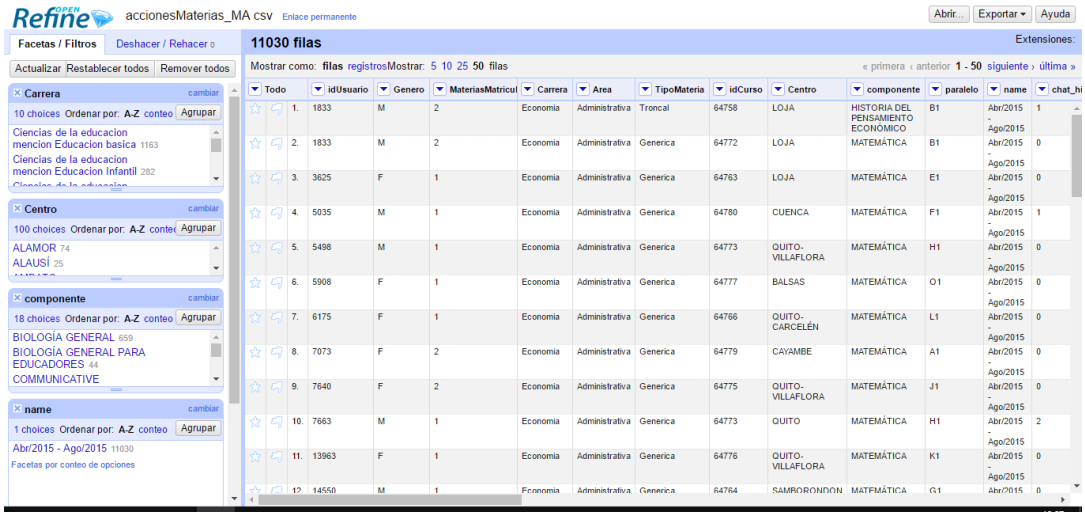



Figura 17. Datos adquiridos

Elaboración: propia

Una vez que los datos han sido cargados a la herramienta antes mencionada, se realizó la limpieza en el que se utilizaron opciones como:

**Text facet:** permite analizar y filtrar los datos, de manera que filtre los datos pertenecientes a una columna, y permita la edición de los mismos,

- Se eliminaron filas de aquellos datos que no poseían valores en sus atributos.
- Se completaron datos faltantes

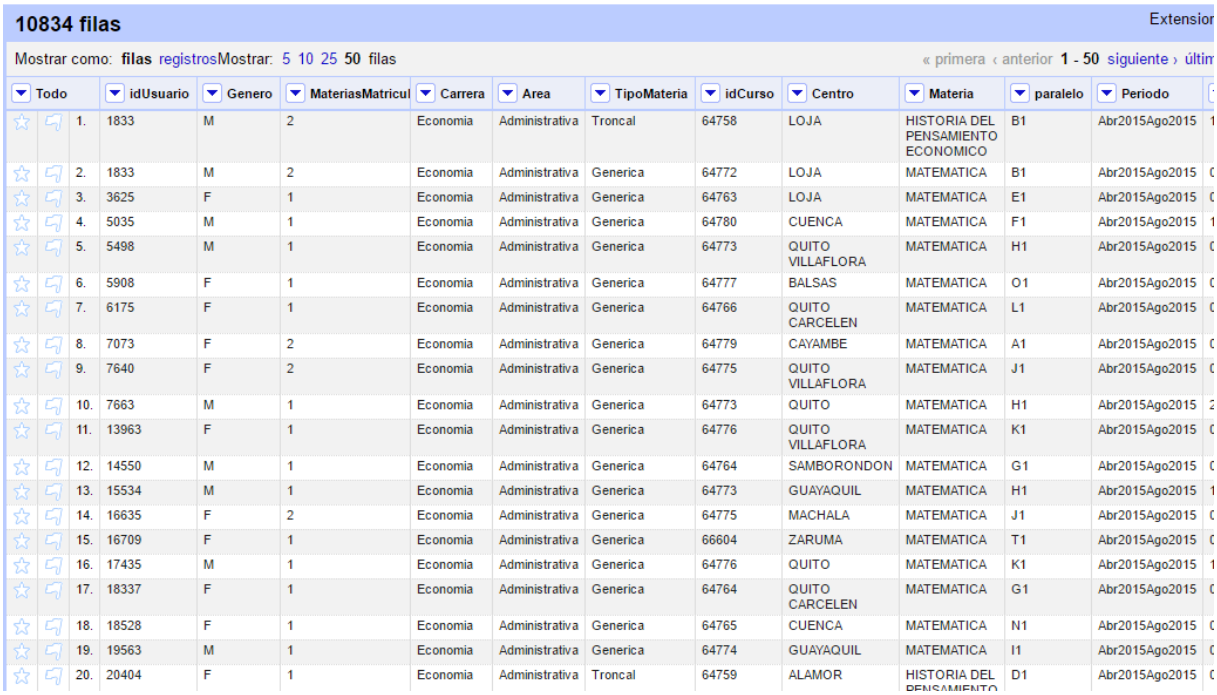
**Transform:** esta opción permite la transformación de los datos, es decir permite:

- Reemplazar valores
- Añadir caracteres
- Eliminar espacios es blanco del inicio y final de los valores
- Fueron removidos caracteres especiales.

Al analizar el atributo Género con la opción **textfacet,** se encontraron datos vacíos y otros valores que no pertenecían al género como s/n**,** a estos se procedió a eliminarlos.

Finalizado el proceso la data quedó con un total de 10834 datos, que muestra la interacción de 7413 estudiantes, esta muestra para la investigación se la tiene en un archivo de formato .csv.

En el anexo 4 se puede visualizar como se realizó la limpieza.



10834 filas

Mostrar como: filas registrosMostrar: 5 10 25 50 filas « primera ‹ anterior 1 - 50 siguiente › últim

| Todo | | idUsuario | Genero | MateriasMatricul | Carrera | Area | TipoMateria | idCurso | Centro | Materia | paralelo | Periodo |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. | | 1833 | M | 2 | Economia | Administrativa | Troncal | 64758 | LOJA | HISTORIA DEL PENSAMIENTO ECONOMICO | B1 | Abr2015Ago2015 |
| 2. | | 1833 | M | 2 | Economia | Administrativa | Generica | 64772 | LOJA | MATEMATICA | B1 | Abr2015Ago2015 |
| 3. | | 3625 | F | 1 | Economia | Administrativa | Generica | 64763 | LOJA | MATEMATICA | E1 | Abr2015Ago2015 |
| 4. | | 5035 | M | 1 | Economia | Administrativa | Generica | 64780 | CUENCA | MATEMATICA | F1 | Abr2015Ago2015 |
| 5. | | 5498 | M | 1 | Economia | Administrativa | Generica | 64773 | QUITO VILLAFLORA | MATEMATICA | H1 | Abr2015Ago2015 |
| 6. | | 5908 | F | 1 | Economia | Administrativa | Generica | 64777 | BALSAS | MATEMATICA | O1 | Abr2015Ago2015 |
| 7. | | 6175 | F | 1 | Economia | Administrativa | Generica | 64766 | QUITO CARCELEN | MATEMATICA | L1 | Abr2015Ago2015 |
| 8. | | 7073 | F | 2 | Economia | Administrativa | Generica | 64779 | CAYAMBE | MATEMATICA | A1 | Abr2015Ago2015 |
| 9. | | 7640 | F | 2 | Economia | Administrativa | Generica | 64775 | QUITO VILLAFLORA | MATEMATICA | J1 | Abr2015Ago2015 |
| 10. | | 7663 | M | 1 | Economia | Administrativa | Generica | 64773 | QUITO | MATEMATICA | H1 | Abr2015Ago2015 |
| 11. | | 13963 | F | 1 | Economia | Administrativa | Generica | 64776 | QUITO VILLAFLORA | MATEMATICA | K1 | Abr2015Ago2015 |
| 12. | | 14550 | M | 1 | Economia | Administrativa | Generica | 64764 | SAMBORONDON | MATEMATICA | G1 | Abr2015Ago2015 |
| 13. | | 15534 | M | 1 | Economia | Administrativa | Generica | 64773 | GUAYAQUIL | MATEMATICA | H1 | Abr2015Ago2015 |
| 14. | | 16635 | F | 2 | Economia | Administrativa | Generica | 64775 | MACHALA | MATEMATICA | J1 | Abr2015Ago2015 |
| 15. | | 16709 | F | 1 | Economia | Administrativa | Generica | 66604 | ZARUMA | MATEMATICA | T1 | Abr2015Ago2015 |
| 16. | | 17435 | M | 1 | Economia | Administrativa | Generica | 64776 | QUITO | MATEMATICA | K1 | Abr2015Ago2015 |
| 17. | | 18337 | F | 1 | Economia | Administrativa | Generica | 64764 | QUITO CARCELEN | MATEMATICA | G1 | Abr2015Ago2015 |
| 18. | | 18528 | F | 1 | Economia | Administrativa | Generica | 64765 | CUENCA | MATEMATICA | N1 | Abr2015Ago2015 |
| 19. | | 19563 | M | 1 | Economia | Administrativa | Generica | 64774 | GUAYAQUIL | MATEMATICA | I1 | Abr2015Ago2015 |
| 20. | | 20404 | F | 1 | Economia | Administrativa | Troncal | 64759 | ALAMOR | HISTORIA DEL PENSAMIENTO | D1 | Abr2015Ago2015 |

Figura 18. Datos limpios

Elaboración: propia

Para el mejor entendimiento de los atributos se optó modificar el nombre de cada uno de ellos.

Tabla 16. Modificación de atributos

| Modulo | Atributo | Modificación |
|---|---|---|
| **Chat** | Talk | veces_conversacion_chat |
| | historial | ve_historial_chat |
| **Elluminate (Videocolaboración)** | view | verSala_videoColaboracion |
| | viewall | Ve_todas_VideoColaboraciones |
| | view meeting | ingreso_VideoColaboracion |
| | viewrecording | ve_grabacionesde_VideoColaboracion |
| **Fórum** | add post | AniadeForo |
| | viewforum | ve_foro |
| | Updatepost | ActualizaPost |

Elaboración: propia

Analizando los datos limpios se mostraron varias inconsistencias en los mismos, como por ejemplo: un determinado estudiante, ve un solo foro 84 veces; es por esto que se procedió a eliminar los valores de estos datos, dejándolos con valores creíbles y al resto ponerlos en cero, este proceso se lo desarrolló en la herramienta R Studio.

Es decir que en la actividad ve foro, hay 2110 estudiantes que no realizan la acción su interacción es de 0, adicional a esto se incrementa los valores inconsistes que dan un total son 2209.

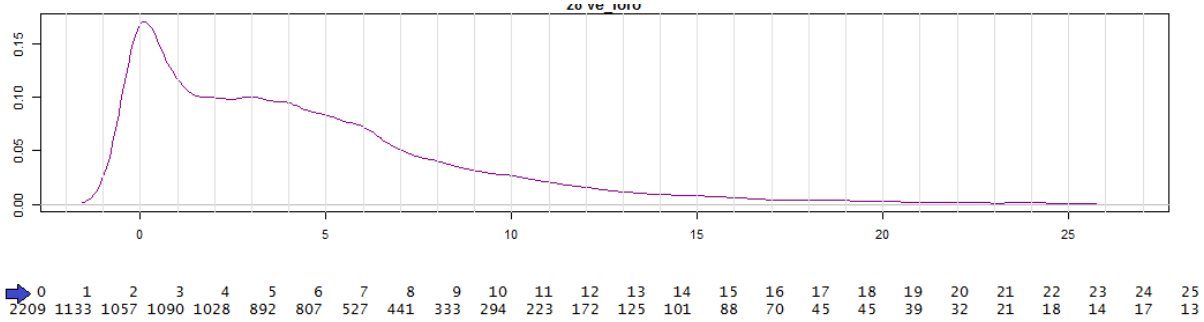



Figura 19. Eliminación de datos inconsistentes
Elaboración: propia

Para la generación de graficas se usa la función plot(density(x)), donde x es un vector numérico, esta función muestra las estadísticas básicas para la estimación de la densidad, por lo general se usa porque es una manera más eficaz para ver la distribución de una variable.
El usuario puede usar otros kernel, no solo el que viene por defecto el de Gauss.

Se usan valores como media, mediana, valor máximo y mínimo, cuartiles, para los diferentes ejes de la gráfica se utilizan los valores mínimos y máximos.

Este proceso se hizo con los 9 atributos numéricos obtenidos. (Ver anexo 4)

Finalmente como se tiene atributos con valores numéricos, se realizó la discretización de datos, el cual consiste con la búsqueda de intervalos adecuados; para esto se desarrolló un algoritmo en la herramienta R Studio.

#### *3.4.2.1 Discretización.*

Su objetivo es la conversión de un valor numérico a un valor nominal el cual representa un intervalo o bins. (KrzysztofJ, Witold, Roman, & Kurgan, 2007)

Este proceso es una tarea fundamental del pre-procesamiento de datos, no porque algunos métodos de aprendizaje no usen atributos continuos sino porque los datos transformados en

intervalos son cognitivamente más relevantes para la interpretación.

Para la discretización se usó indicadores, como lo definen CIDA (1996) son "herramientas para clarificar y definir con mayor precisión los objetivos y el impacto... son medidas verificables de los cambios o resultados... están diseñados para proporcionar un estándar contra el que medir, estimar, o demostrar el progreso... contra objetivos establecidos, hacia la entrega de... entradas, salidas y producir... ... el logro de objetivos"

Para la discretización se utilizaron 3 intervalos, los registros que tengan un buen porcentaje de interacción en cada actividad se denominaran alto (A), los que tengan poca interacción serán bajo (B), mientras que los que la interacción es moderada se denominará medio (M).

Tabla 17. Discretización de atributos numéricos

| Atributo | Intervalos | | |
|---|---|---|---|
| | Bajo (B) | Medio (M) | Alto(A) |
| **ve_historial_chat** | 0.00 - 3.33 | 3.33 - 6.67 | 6.67 - 10.00 |
| **veces_conversacion_chat** | 0.00 - 3.33 | 3.33 - 6.67 | 6.67 - 10.00 |
| **verSala_videoColaboracion** | 0 - 4 | 4 - 8 | 8 - 12 |
| **Ve_todas_VideoColaboraciones** | 0 - 5 | 5 - 10 | 10 - 15 |
| **ingreso_VideoColaboracion** | 0 - 5 | 5 - 10 | 10 - 15 |
| **ve_grabacionesde_VideoColaboracion** | 0.00 - 3.67 | 3.67 - 7.33 | 7.33 - 11.00 |
| **AniadeForo** | 0.00 - 2.33 | 2.33 - 4.67 | 4.67 - 7.00 |
| **ActualizaPost** | 0.00 - 3.67 | 3.67 - 7.33 | 7.33 - 11.00 |
| **ve_foro** | 0.00 - 8.33 | 8.33 -16.67 | 16.67 - 25.00 |

Elaboración: propia

## 3.5 Fase IV: Modelado

Una vez que se tiene el dataset final, se procede a la creación del modelado, tomando en cuenta la técnica de minería seleccionada.

### 3.5.1 Selección de la técnica de modelado

#### *3.5.1.1 Muestra #1.*

Como se mencionó en el ítem 3.4.1 con la primera muestra extraída de la base de datos UTPL que corresponde a las interacciones en chats, viodeocolaboraciones y foros de la Titulación de Gestión Ambiental, se utilizó la técnica de Clustering, Mapas de Kohonen, que como se lo indicó en el capítulo 1 (1.4.2.1.1) permiten descubrir rasgos comunes dentro de

un conjunto de datos, para auto organizarlos en función de los datos procedentes del exterior.

Estos mapas simulan el funcionamiento del cerebro, es decir, esta red consta de dos capas con N neuronas de entrada y M de salida, cuando entra información, cada una de las M salidas la recibe a través de conexiones con pesos, una vez que la red evoluciona sólo una neurona de salida se activará quedando como la neurona vencedora.

Para entrenar, primero se debe tener un conjunto de datos entrantes, que se dividirá en tres grupos, entrenamiento, prueba, validación. (Merelo, 2004)

A continuación se detalla cada uno de los paso a seguir:

1) Primeramente se inicializan los pesos, con valores aleatorios pequeños, Wji.
2) Luego se presenta la información de entrada en forma de un vector, $E_k = (e_1 \dots \dots e_n)$, en donde e, son valores continuos.
3) Se realiza la determinación de la neurona vencedora de la capa de salida, en donde el vector de pesos Wj, sea el más aproximado a la entrada Ek. Para ello se realiza el cálculo de las distancias entre los vectores antes mencionados. Uno de los más empleados es la distancia euclidiana.

$$d^2(w_{ij}, x) = \sum_{k=1}^{n} (w_{ijk} - x_k)^2$$

Ecuación 1. Fórmula para calcular la distancia euclidiana
Fuente: (Flórez & Fernández, 2008)

4) La determinación de la neurona ganadora es la que cuya distancia es la menor de todas.
5) Cuando se tiene la neurona ganadora se realiza la actualización de los pesos de esta neurona y de sus vecinas, la más usada para esta actualización es la compatible con la distancia euclidiana que se muestra a continuación:

$$w_{ij}(t+1) = w_{ij}t + \beta(t)({e_i}^k - w_{j*i}(t))$$

Ecuación 2. Actualización de pesos, compatible con la distancia euclidiana
Fuente: (Flórez & Fernández, 2008)

, donde $\beta(t)$ es un parámetro denominado ritmo de aprendizaje, puede usarse con la expresión $\beta(t) = \frac{1}{t}$.

6) Cuando se realicen todas las iteraciones, el proceso termina, de lo contrario vuelve al paso 2. (Flórez & Fernández, 2008)

#### *3.5.1.2 Muestra #2.*

Para la selección de la técnica del modelado de la segunda muestra (ver figura 15), se procedió a ver las características de los datos.

Una vez analizados las técnicas de minería presentados en el estado del arte, se usará la técnica de asociación con el algoritmo Apriori.

Este permite expresar asociaciones entre varias variables. Por ejemplo: cuando se compra un artículo por Amazon.com, asocian su compra con algún otro artículo similar que le pueda interesar al cliente y le sugiere.
Las reglas de asociación generadas mediante el algoritmo A priori, permiten expresar patrones de comportamiento entre los datos disponibles, las reglas son de tipo:

$$Si: X \rightarrow Y$$

Ecuación 3. Fórmula reglas de asociación.
Fuente: (Malberti & Elida, 2015)

Donde X e Y se denominan antecedente (LHS, lado izquierdo) y consecuente (RHS, lado derecho), a menudo las reglas se restringen a un solo elemento en el consecuente.

Las reglas de asociación usan varias métricas para verificar la calidad de la regla, a continuación se detallan:

- Soporte (Support): número de veces con que X aparece dentro de un conjunto de transacciones.

$$supp\ (X \rightarrow Y) = \frac{cantidad\ de\ repeticiones\ de\ X \rightarrow Y}{total\ de\ transacciones} = supp(X \cup Y)$$

Ecuación 4. Fórmula para calcular el soporte
Fuente: (Malberti & Elida, 2015)

- Confianza (Confidence): Malberti & Elida (2015) mencionan que es “probabilidad de que las transacciones que contienen el antecedente de la regla, también tenga el consecuente” (p. 34), es decir, la confianza mide con qué frecuencia aparece Y en las transacciones que incluyen X.

$$conf(X \rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)}$$

Ecuación 5. Fórmula para calcular la métrica confianza

Fuente: (Malberti & Elida, 2015)

- Lift: cuantifica la relación entre $X \rightarrow Y$:
  - lift > 1 : aparece una cantidad de veces superior a lo esperado
  - lift = 1: conjunto de datos aparece una cantidad de veces acorde a lo esperado.
  - lift < 1: aparece una cantidad de veces inferior a lo esperado

$$lift(X \rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)supp(Y)}$$

Ecuación 6. Fórmula para calcular la métrica Lift

Fuente: (Malberti & Elida, 2015)

A continuación se realiza un ejemplo para el mejor entendimiento del proceso para generar reglas de asociación.

Dada la siguiente tabla de transacciones;

Tabla 18 . Transacciones

| Id | Transacción |
|---|---|
| **1** | Pan, leche, pañales |
| **2** | Pan, pañales, cerveza, huevos |
| **3** | Leche, pañales, cerveza, refresco, café. |
| **4** | Pan, leche, pañales, cerveza |
| **5** | Pan, refresco, leche, pañales |

Calcular el soporte de cada uno de los ítems.

1. Primeramente se saca la frecuencia de cada variable, es decir el número de repeticiones de variable dentro de la tabla de transacciones.

Tabla 19. Frecuencia de los variables para generar reglas de asociación

| Ítem | Frecuencia |
|---|---|
| **Cerveza** | 3 |

| | |
|---|---|
| **Pan** | 4 |
| **Refresco** | 2 |
| **Pañales** | 5 |
| **Leche** | 4 |
| **Huevos** | 1 |
| **Café** | 1 |

Fuente: (Pitol, 2014)

2. Calcular el soporte de las variables (ver ecuación 4), para este ejemplo se usará los variables que sean mayor del 50%.

Tabla 20. Cálculo de la métrica soporte de las variables

| Variable | Soporte |
|---|---|
| **Cerveza** | $supp\ (cerveza) = \frac{3}{5} = 0.6 = 60\%$ |
| **Pan** | $supp\ (pan) = \frac{4}{5} = 0.8 = 80\%$ |
| **Refresco** | $supp\ (refresco) = \frac{2}{5} = 0.4 = 40\%$ |
| **Pañales** | $supp\ (Pañales) = \frac{5}{5} = 1 = 100\%$ |
| **Leche** | $supp\ (leche) = \frac{4}{5} = 0.8 = 80\%$ |
| **Huevos** | $supp\ (huevos) = \frac{1}{5} = 0.2 = 20\%$ |
| **Café** | $supp\ (Café) = \frac{1}{5} = 0.2 = 20\%$ |

Fuente: (Pitol, 2014)

Una vez obtenido el soporte de las variables, se procede a realizar combinaciones con 2 variables, para esto se sigue el mismo proceso antes visto.

Tabla 21. Cálculo de la métrica soporte para la combinación de 2 variables

| Conjunto | Frecuencia | Soporte |
|---|---|---|
| **Cerveza, pan** | 2 | $supp\ (cerveza, pan) = \frac{2}{5} = 0.4 = 40\%$ |
| **Cerveza, pañales** | 3 | $supp\ (cerveza, pañales) = \frac{3}{5} = 0.6 = 60\%$ |
| **Cerveza, leche** | 2 | $supp\ (cerveza, leche) = \frac{2}{5} = 0.4 = 40\%$ |

| | | |
|---|---|---|
| **Pan, pañales** | 4 | $supp\ (pan, pañales) = \frac{4}{5} = 0.8 = 80\%$ |
| **Pan, leche,** | 3 | $supp\ (pan, leche) = \frac{3}{5} = 0.6 = 60\%$ |
| **Pañales , leche** | 4 | $supp\ (\ pañales, leche) = \frac{4}{5} = 0.8 = 80\%$ |

Fuente: (Pitol, 2014)

Seguidamente se hace la combinación con 3 conjuntos, usando el conjunto de variables generado anteriormente. Se calcula el soporte para el nuevo conjunto de datos.

Tabla 22. Cálculo de la métrica soporte para la combinación de 3 variables

| Conjunto | Frecuencia | Soporte |
|---|---|---|
| **Cerveza, pañales, pan** | 2 | $supp\ (cerveza, pañales \rightarrow\ pan) = \frac{2}{5} = 0.4 = 40\%$ |
| **Cerveza, pañales, leche** | 2 | $supp\ (cerveza, leche \rightarrow\ pañales) = \frac{2}{5} = 0.4 = 40\%$ |
| **Pan, pañales, leche** | 3 | $supp\ (pan, pañales \rightarrow leche) = \frac{3}{5} = 0.6 = 60\%$ |
| **Pan, leche, cerveza** | 1 | $supp\ (pan, leche \rightarrow cerveza) = \frac{1}{5} = 0.2 = 20\%$ |

Fuente: (Pitol, 2014)

El conjunto de ítems que tiene más del 50% es **pan, pañales, leche,** las posibles reglas obtenidas serán:

- Pan => Pañales, Leche
- Pañales => Pan , Leche
- Leche = > Pan, Pañales
- Pan, Pañales => Leche
- Pan, Leche => Pañales
- Leche, Pañales => Pan

De la regla Pan, Pañales => Leche se obtendrá la confianza para lo cual se usará la ecuación 5

$$conf\,(pan, pañales \rightarrow leche) = \frac{3}{4} = 0.75 = 75\%$$

Ecuación 7. Cálculo de la métrica confianza para la regla pan, pañales →Leche

Fuente: (Pitol, 2014)

Para obtener la métrica lift (ver ecuación 6) de esta regla se realiza la siguiente ecuación:

$$lift(pan, pañales \rightarrow leche) = \frac{3}{4 * 4} = 0{,}19$$

Ecuación 8. Cálculo de la métrica Lift para la regla pan, pañales →Leche
Fuente: (Pitol, 2014)

### 3.5.2 Construcción del modelo

En esta fase de seleccionan y se aplican diversas técnicas, por tal razón lo más importante es elegir la técnica correcta, tomando en cuenta ciertos criterios como el problema a resolver, los datos a trabajar, etc. (Chapman et al., 2000)

Como se mencionó anteriormente para la construcción del modelo en la primera muestra se eligió la técnica de Clustering, mapas de Kohonen, mientras que para la segunda muestra se eligió la técnica de asociación, algoritmo a priori, para la obtención de patrones de comportamiento en las actividades realizadas en el EVA por parte de los estudiantes de modalidad a distancia.

#### *3.5.2.1 Aplicación de la técnica de Clustering: Mapas de Kohonen*

Para aplicar mapas de Kohonen en la herramienta R Studio, primeramente se procedió a instalar el paquete Kohonen, este paquete tiene énfasis en la visualización. Antes de la fase de entrenamiento, se procedió a eliminar las variables de tipo clase, pero estas serán utilizadas luego para la visualización.

Una vez realizado el entrenamiento se realizó la topología o estructura del mapa, las más frecuentes son la rectangular y la hexagonal. Esta topología es de 5 columnas x 4 filas de tipo hexagonal.

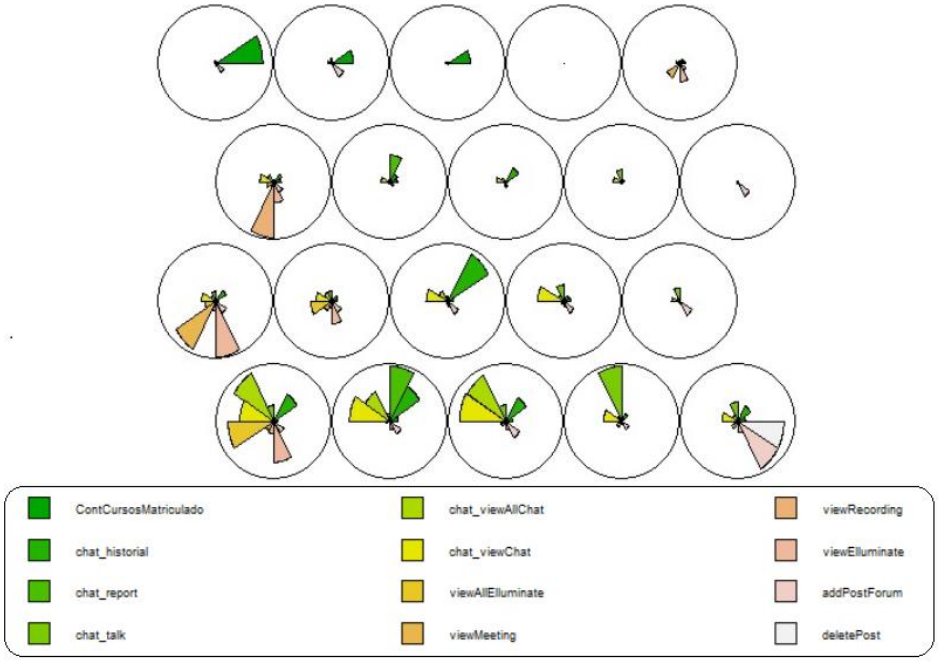



Figura 20. Clustering con todas las variables

Elaboración: propia

En la figura anterior, se muestran 20 nodos, que surgen luego de realizar el modelo SOM, en el cual se puede observar las agrupaciones dependiendo del nivel de interacción que posee cada cual, es decir, en el nodo o neurona 20 se puede ver que las variables más usadas son addPostForum y deletePost, mientras que en la neurona 16 las variables que contienen más interacción son: chat_viewAllChat, viewAllElluminate y viewElluminate, la variable chat_viewChat tienen interacción menor con respecto a las antes mencionadas.

Cabe mencionar que para esta experimentación se usaron todas las variables a excepción de algunas que tenían valor de 0

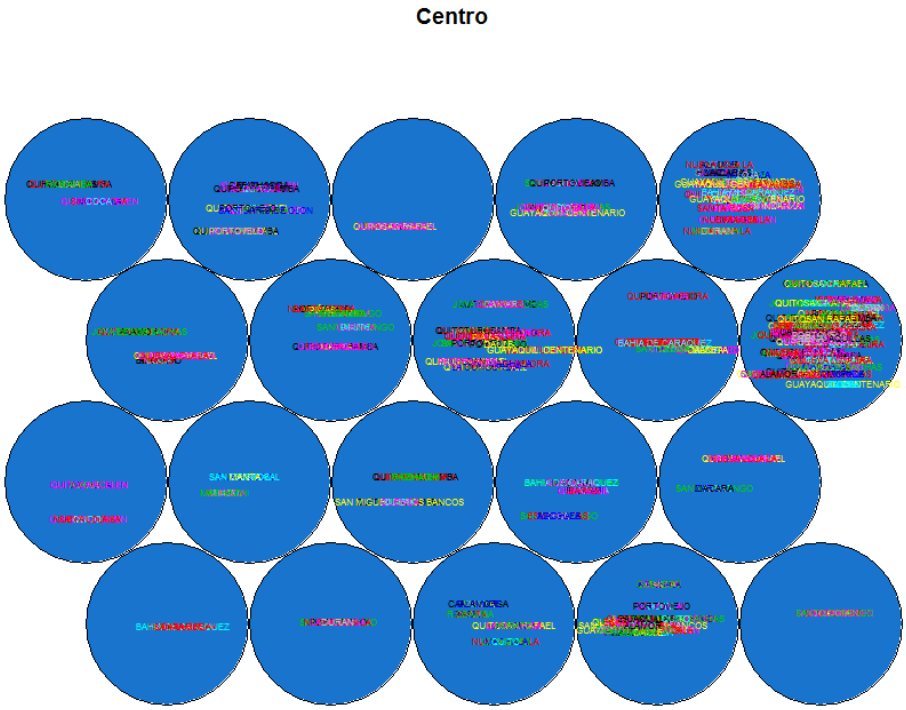



Figura 21. Clustering con la variable Centro
Elaboración: propia

También se realizó Clustering, con el variable Centro, en la cual se muestra que dependiendo el número de acciones en cada una de las variables se realiza la agrupación, como se observa en la figura 21, neurona 10, que posee un conglomerado bastante lleno en comparación con las demás neuronas o nodos.

#### *3.5.2.2 Aplicación de la técnica de Asociación: algoritmo A priori*

Con la implementación del algoritmo a priori en la herramienta R Studio, se hicieron varias configuraciones:
En el parámetro Maxlen, que se refiere al número máximo de elementos en el conjunto de reglas se colocó el valor de 3, es decir que este conjunto, estará conformadas por 3 elementos. Mientras que en el parámetro minlen se deja el valor por defecto que es 1.

Para la generación de combinaciones es decir el soporte se ha decidido trabajar con el 50%, mientras que para el parámetro confianza se trabaja con el 80%.

Es importante antes de seguir, conocer los rangos establecidos para cada actividad:

- **Interacción Baja:** Se refiere a los estudiantes que no tienen una participación o interacción buena con la herramienta.
- **Interacción media:** Se refiere a los estudiantes que no tienen una interacción baja ni alta, se mantienen en una interacción moderada con la herramienta.
- **Interacción Alta:** Son estudiantes que mayormente presentan interacción en cada una de las herramientas.

Una vez realizado el algoritmo a priori, con los valores antes mencionados se obtuvieron las siguientes reglas:

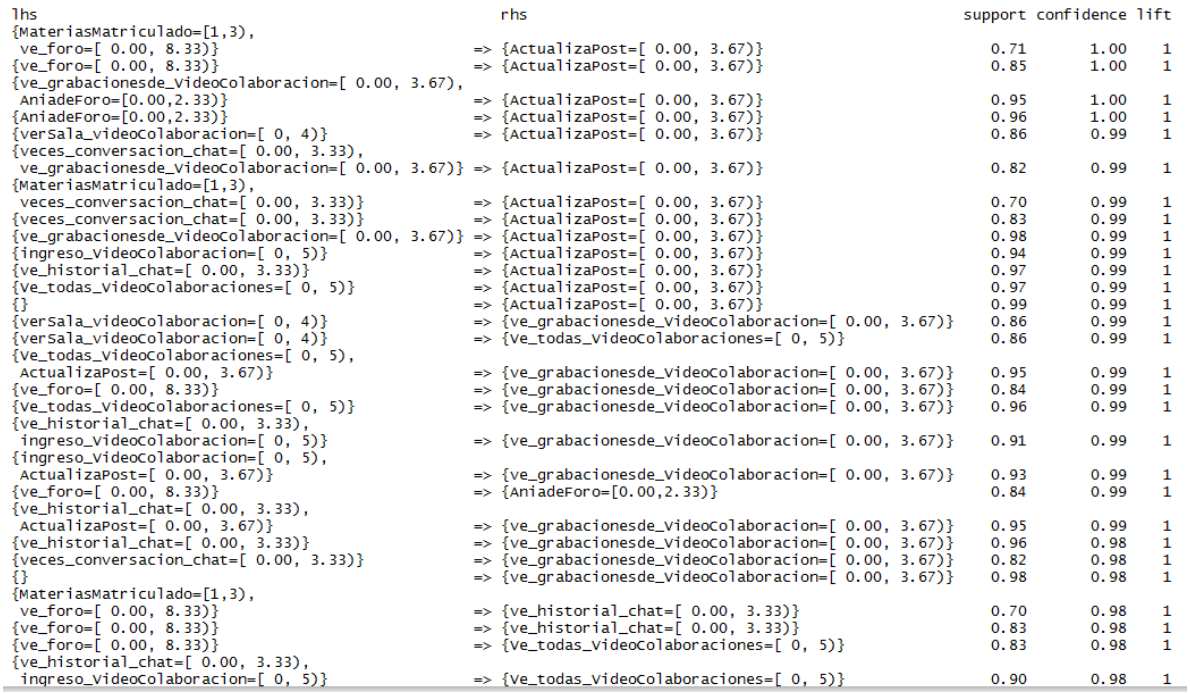
| | lhs | | rhs | support | confidence | lift |
|---|---|---|---|---|---|---|
| 1 | {ingreso_videoColaboracion=[ 0, 5), ve_foro=[ 0.00, 8.33)} | => | {verSala_videoColaboracion=[ 0, 4)} | 0.74 | 0.91 | 1.1 |
| 2 | {Ve_todas_videoColaboraciones=[ 0, 5), ingreso_videoColaboracion=[ 0, 5)} | => | {verSala_videoColaboracion=[ 0, 4)} | 0.84 | 0.91 | 1.0 |
| 3 | {veces_conversacion_chat=[ 0.00, 3.33), ingreso_videoColaboracion=[ 0, 5)} | => | {verSala_videoColaboracion=[ 0, 4)} | 0.72 | 0.91 | 1.0 |
| 4 | {ingreso_videoColaboracion=[ 0, 5), ve_grabacionesde_videoColaboracion=[ 0.00, 3.67)} | => | {verSala_videoColaboracion=[ 0, 4)} | 0.84 | 0.91 | 1.0 |
| 5 | {verSala_videoColaboracion=[ 0, 4), AniadeForo=[0.00,2.33)} | => | {ve_foro=[ 0.00, 8.33)} | 0.74 | 0.88 | 1.0 |
| 6 | {ingreso_videoColaboracion=[ 0, 5), AniadeForo=[0.00,2.33)} | => | {verSala_videoColaboracion=[ 0, 4)} | 0.82 | 0.90 | 1.0 |
| 7 | {ve_historial_chat=[ 0.00, 3.33), ingreso_videoColaboracion=[ 0, 5)} | => | {verSala_videoColaboracion=[ 0, 4)} | 0.83 | 0.90 | 1.0 |
| 8 | {ingreso_videoColaboracion=[ 0, 5), ActualizaPost=[ 0.00, 3.67)} | => | {verSala_videoColaboracion=[ 0, 4)} | 0.84 | 0.90 | 1.0 |
| 9 | {veces_conversacion_chat=[ 0.00, 3.33), AniadeForo=[0.00,2.33)} | => | {ve_foro=[ 0.00, 8.33)} | 0.71 | 0.88 | 1.0 |
| 10 | {verSala_videoColaboracion=[ 0, 4)} | => | {ingreso_videoColaboracion=[ 0, 5)} | 0.85 | 0.98 | 1.0 |
| 11 | {Ve_todas_videoColaboraciones=[ 0, 5), ve_foro=[ 0.00, 8.33)} | => | {verSala_videoColaboracion=[ 0, 4)} | 0.75 | 0.90 | 1.0 |
| 12 | {veces_conversacion_chat=[ 0.00, 3.33), verSala_videoColaboracion=[ 0, 4)} | => | {ve_foro=[ 0.00, 8.33)} | 0.64 | 0.88 | 1.0 |
| 13 | {Ve_todas_videoColaboraciones=[ 0, 5), AniadeForo=[0.00,2.33)} | => | {ve_foro=[ 0.00, 8.33)} | 0.82 | 0.87 | 1.0 |
| 14 | {ve_grabacionesde_videoColaboracion=[ 0.00, 3.67), ve_foro=[ 0.00, 8.33)} | => | {verSala_videoColaboracion=[ 0, 4)} | 0.75 | 0.89 | 1.0 |
| 15 | {veces_conversacion_chat=[ 0.00, 3.33), Ve_todas_videoColaboraciones=[ 0, 5)} | => | {verSala_videoColaboracion=[ 0, 4)} | 0.72 | 0.89 | 1.0 |

Figura 22. Reglas de asociación obtenidas

Elaboración: propia

En total se obtuvieron 134 reglas las cuales se las ordenó dependiendo de la métrica, por ejemplo en la figura 22 se lo realizó por la métrica lift.

A continuación se puede observar algunas de las reglas de asociación ordenadas por la métrica confidence (confianza), en las cuales se puede percibir que el consecuente (rhs) en la mayoría son las variables mencionadas en la tabla 18.



| lhs | rhs | support | confidence | lift |
|---|---|---|---|---|
| {MateriasMatriculado=[1,3), ve_foro=[ 0.00, 8.33)} | => {ActualizaPost=[ 0.00, 3.67)} | 0.71 | 1.00 | 1 |
| {ve_foro=[ 0.00, 8.33)} | => {ActualizaPost=[ 0.00, 3.67)} | 0.85 | 1.00 | 1 |
| {ve_grabacionesde_videoColaboracion=[ 0.00, 3.67), AniadeForo=[0.00,2.33)} | => {ActualizaPost=[ 0.00, 3.67)} | 0.95 | 1.00 | 1 |
| {AniadeForo=[0.00,2.33)} | => {ActualizaPost=[ 0.00, 3.67)} | 0.96 | 1.00 | 1 |
| {verSala_videoColaboracion=[ 0, 4)} | => {ActualizaPost=[ 0.00, 3.67)} | 0.86 | 0.99 | 1 |
| {veces_conversacion_chat=[ 0.00, 3.33), ve_grabacionesde_videoColaboracion=[ 0.00, 3.67)} | => {ActualizaPost=[ 0.00, 3.67)} | 0.82 | 0.99 | 1 |
| {MateriasMatriculado=[1,3), veces_conversacion_chat=[ 0.00, 3.33)} | => {ActualizaPost=[ 0.00, 3.67)} | 0.70 | 0.99 | 1 |
| {veces_conversacion_chat=[ 0.00, 3.33)} | => {ActualizaPost=[ 0.00, 3.67)} | 0.83 | 0.99 | 1 |
| {ve_grabacionesde_videoColaboracion=[ 0.00, 3.67)} | => {ActualizaPost=[ 0.00, 3.67)} | 0.98 | 0.99 | 1 |
| {ingreso_videoColaboracion=[ 0, 5)} | => {ActualizaPost=[ 0.00, 3.67)} | 0.94 | 0.99 | 1 |
| {ve_historial_chat=[ 0.00, 3.33)} | => {ActualizaPost=[ 0.00, 3.67)} | 0.97 | 0.99 | 1 |
| {ve_todas_videoColaboraciones=[ 0, 5)} | => {ActualizaPost=[ 0.00, 3.67)} | 0.97 | 0.99 | 1 |
| {} | => {ActualizaPost=[ 0.00, 3.67)} | 0.99 | 0.99 | 1 |
| {verSala_videoColaboracion=[ 0, 4)} | => {ve_grabacionesde_videoColaboracion=[ 0.00, 3.67)} | 0.86 | 0.99 | 1 |
| {verSala_videoColaboracion=[ 0, 4)} | => {ve_todas_videoColaboraciones=[ 0, 5)} | 0.86 | 0.99 | 1 |
| {ve_todas_videoColaboraciones=[ 0, 5), ActualizaPost=[ 0.00, 3.67)} | => {ve_grabacionesde_videoColaboracion=[ 0.00, 3.67)} | 0.95 | 0.99 | 1 |
| {ve_foro=[ 0.00, 8.33)} | => {ve_grabacionesde_videoColaboracion=[ 0.00, 3.67)} | 0.84 | 0.99 | 1 |
| {ve_todas_videoColaboraciones=[ 0, 5)} | => {ve_grabacionesde_videoColaboracion=[ 0.00, 3.67)} | 0.96 | 0.99 | 1 |
| {ve_historial_chat=[ 0.00, 3.33), ingreso_videoColaboracion=[ 0, 5)} | => {ve_grabacionesde_videoColaboracion=[ 0.00, 3.67)} | 0.91 | 0.99 | 1 |
| {ingreso_videoColaboracion=[ 0, 5), ActualizaPost=[ 0.00, 3.67)} | => {ve_grabacionesde_videoColaboracion=[ 0.00, 3.67)} | 0.93 | 0.99 | 1 |
| {ve_foro=[ 0.00, 8.33)} | => {AniadeForo=[0.00,2.33)} | 0.84 | 0.99 | 1 |
| {ve_historial_chat=[ 0.00, 3.33), ActualizaPost=[ 0.00, 3.67)} | => {ve_grabacionesde_videoColaboracion=[ 0.00, 3.67)} | 0.95 | 0.99 | 1 |
| {ve_historial_chat=[ 0.00, 3.33)} | => {ve_grabacionesde_videoColaboracion=[ 0.00, 3.67)} | 0.96 | 0.98 | 1 |
| {veces_conversacion_chat=[ 0.00, 3.33)} | => {ve_grabacionesde_videoColaboracion=[ 0.00, 3.67)} | 0.82 | 0.98 | 1 |
| {} | => {ve_grabacionesde_videoColaboracion=[ 0.00, 3.67)} | 0.98 | 0.98 | 1 |
| {MateriasMatriculado=[1,3), ve_foro=[ 0.00, 8.33)} | => {ve_historial_chat=[ 0.00, 3.33)} | 0.70 | 0.98 | 1 |
| {ve_foro=[ 0.00, 8.33)} | => {ve_historial_chat=[ 0.00, 3.33)} | 0.83 | 0.98 | 1 |
| {ve_foro=[ 0.00, 8.33)} | => {ve_todas_videoColaboraciones=[ 0, 5)} | 0.83 | 0.98 | 1 |
| {ve_historial_chat=[ 0.00, 3.33), ingreso_videoColaboracion=[ 0, 5)} | => {ve_todas_videoColaboraciones=[ 0, 5)} | 0.90 | 0.98 | 1 |

Figura 23. Reglas de asociación ordenadas por la métrica Confidence

Elaboración: propia

En el anexo 5, se puede observar el grafo generado por las reglas de asociación ordenadas por la métrica confianza.

## 3.6 **Fase V: Evaluación**

Una vez analizadas cada una de las 134 reglas de asociación obtenidas podemos decir que, las reglas con mayor confianza se concentran en las variables más frecuentes que son:

Tabla 23. Variables más frecuentes según reglas de asociación

| Variables más frecuentes |
|---|
| ActualizaPost=[ 0.00, 3.67) |
| ve_grabacionesde_VideoColaboracion=[ 0.00, 3.67) |
| Ve_todas_VideoColaboraciones=[ 0, 5) |
| ve_historial_chat=[ 0.00, 3.33) |
| AniadeForo=[0.00,2.33 |

Elaboración: propia

Seguidamente se realiza un análisis estadístico, para comprobar si las variables indicadas anteriormente son las más utilizadas por los estudiantes.

En la actividad Chat se puede observar que:

- Ver historial de chat: el 97% de la población es decir que 10540 registros, pertenecen a una interacción baja, mientras que el 2% a interacción media y tan sólo el 1% tiene interacción alta.


Figura 24. Interacción chat - ver historial
Elaboración: propia

- Veces que tiene conversación: en esta actividad los estudiantes tienen el 83% de interacción baja es decir 9021 registros, la interacción media el 6% con 636 registros e interacción alta el 11% con 1177 registros.

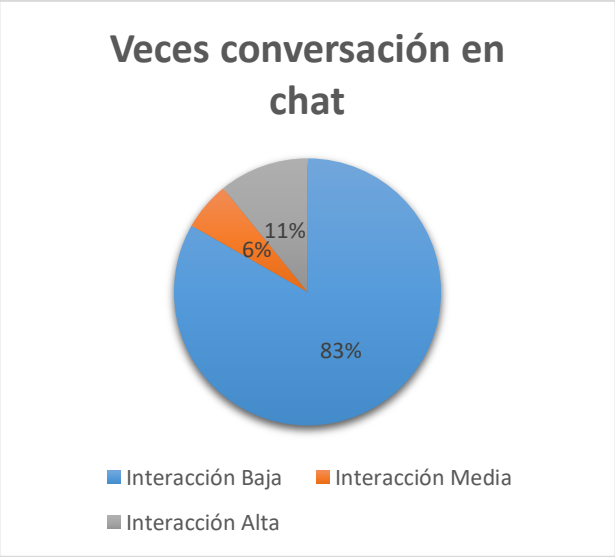



Figura 25. Interacción chat - veces_conversacion_chat
Elaboración: Propia

En la actividad Video Colaboración se obtuvieron los siguientes resultados:

- Ingreso a Video Colaboración: la interacción que tienen los estudiantes es baja con el 95%, mientas que la interacción media es de 4% y la interacción alta es tan solo el 1%.

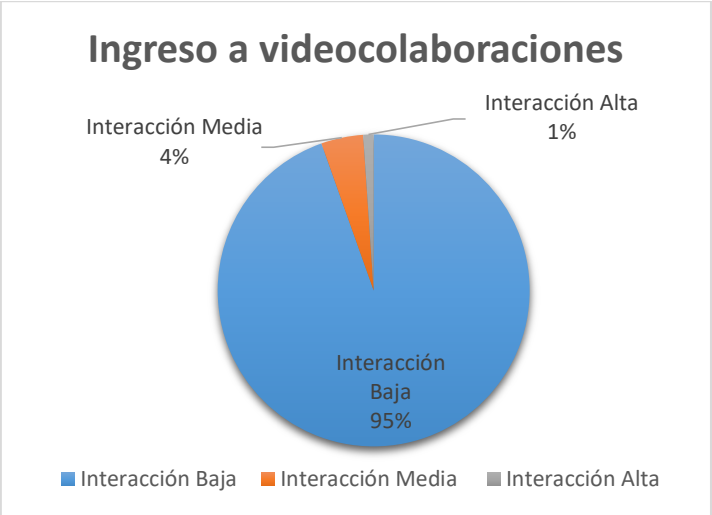



Figura 26. Interacción Videocolaboración - ingreso_VideoColaboracion

Elaboración: Propia

- Ver todas las Video colaboraciones: Los estudiantes que interactúan en esta actividad lo hacen de manera baja con el 97%, por lo tanto solo el 2% lo hace de manera media y el 1% de manera alta.


Figura 27. Interacción Videocolaboración - Ve_todas_VideoColaboraciones

Elaboración: Propia

- Ver grabaciones de Video Colaboración: en esta actividad como se puede observar la interacción es baja con el 99% que corresponden a 10662 registros, el 1% restante

pertenece a la interacción media.


Figura 28. Interacción Videocolaboración - ve_grabacionesde_VideoColaboracion

Elaboración: Propia

- Ver sala de video colaboración: la interacción del estudiante en esta actividad al igual que las demás es baja teniendo el 87%, mientras que la interacción media pose el 10% y el 3% la interacción alta

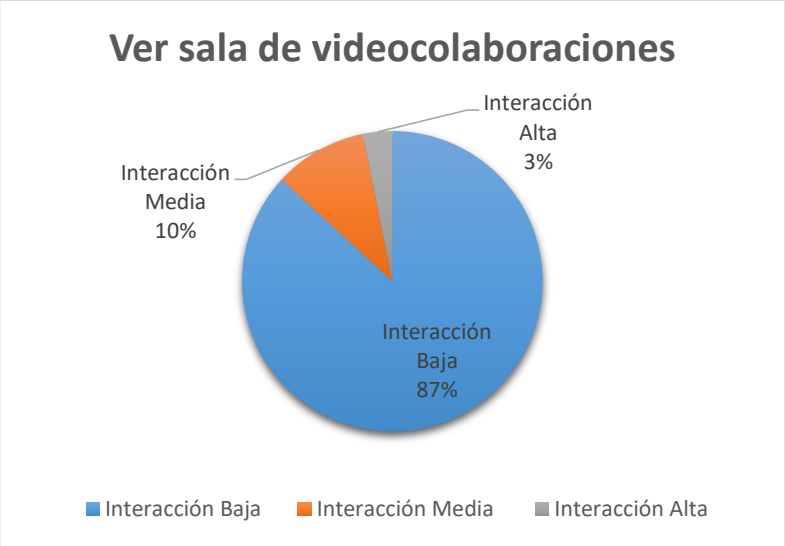



Figura 29. Interacción Videocolaboración - verSala_videoColaboracion

Elaboración: Propia

En la actividad foros se observa que:

- Añadir Foro: en esta actividad la interacción de los estudiantes es baja con el 97% y el 3% pertenece a la interacción media, por lo tanto no existe una interacción alta.


Figura 30. Interacción Foro - AniadeForo

Elaboración: Propia

- Actualiza post: la interacción de los estudiantes actualizando post en foros es baja con el 99% es decir 10767 registros, tan solo el 1% con 61 registros representa una interacción media.

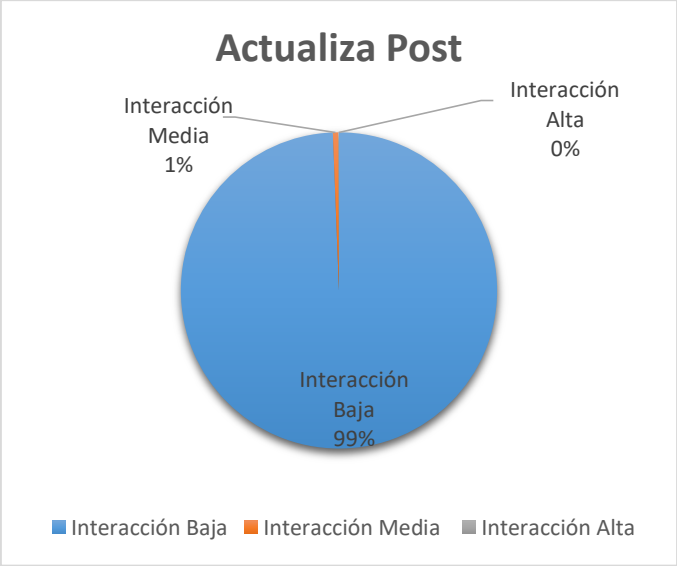



Figura 31. Interacción Foro – ActualizaPost

Elaboración: Propia

- Ver Foro: la presente actividad varia en cuanto a los resultados de foros, ya que el 85% de la interacción es baja mientras que, el 13% interactúa de manera media y apenas el 2% presenta interacción alta.

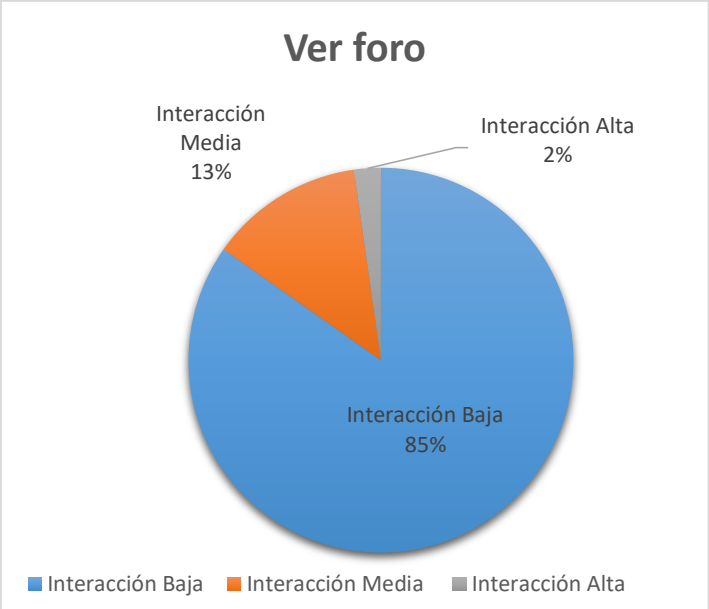



Figura 32. Interacción Foro - ve_foro

Elaboración: Propia

En cada una de las actividades, la interacción baja es predominante, los resultados si varían pocamente con la interacción media y alta.

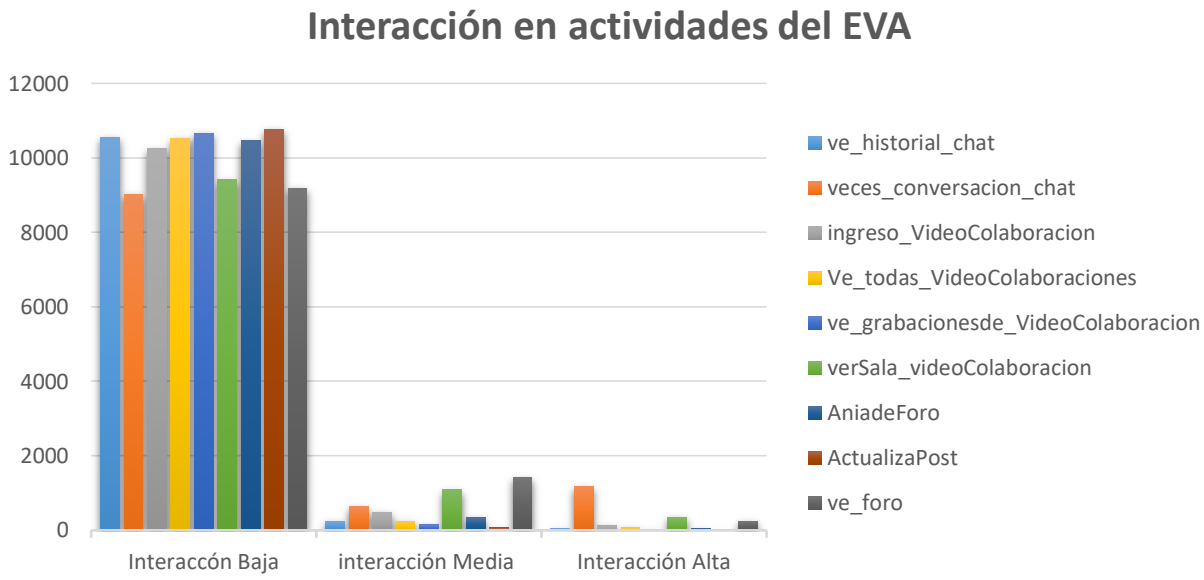



Figura 33. Interacción en actividades del EVA

Elaboración: propia

Por tanto dentro del conjunto de reglas encontradas, se puede visualizar que en todas las interacciones existe mayor interacción con los estudiantes de género femenino y menor con el género masculino.

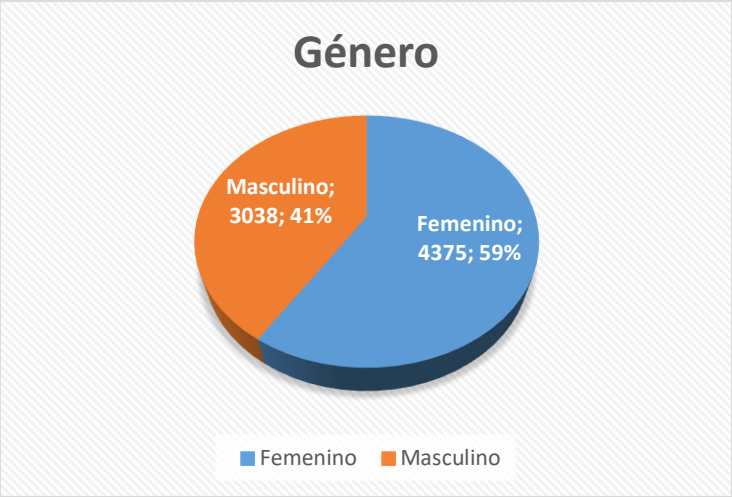



Figura 34. Interacción en actividades referente a género

Elaboración: propia

Como se puede observar en la siguiente gráfica, se tiene reglas que poseen un buen nivel de confianza y de soporte, pero, son estás que denotan que el conjunto de elementos aparece una cantidad de veces a lo esperado, es decir, aquellas que tienen baja interacción son las que predominan.

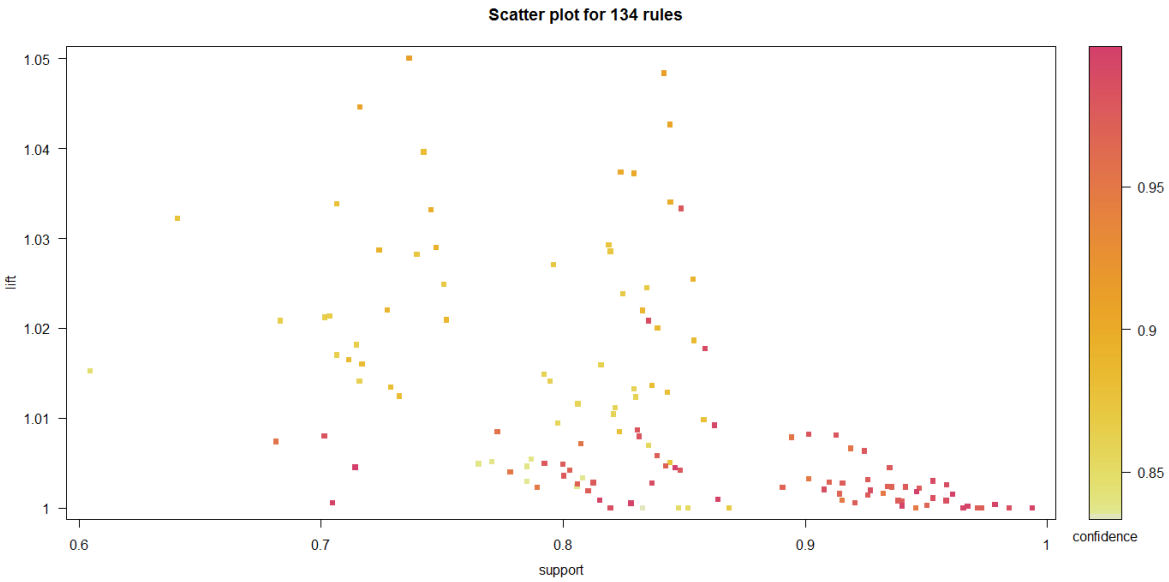



Figura 35. Resultados de reglas de asociación

Elaboración: propia

Mediante el análisis estadístico se encontró que las variables más frecuentes indicadas en la tabla 23, son las que mayor nivel de interacción baja tienen (ver tabla 24) con respecto a las demás variables.

Tabla 24. Resultado del análisis estadístico

| Variable | Interacción Baja | Interacción Media | Interacción Alta |
|---|---|---|---|
| **ActualizaPost** | 99% | 1% | 0% |
| **ve_grabacionesde_VideoColaboracion** | 99% | 1% | 0% |
| **Ve_todas_VideoColaboraciones** | 97% | 2% | 1% |
| **ve_historial_chat** | 97% | 2% | 1% |
| **AniadeForo** | 97% | 3% | 0% |

Elaboración: propia

Una vez que se ha analizado cada uno de los resultados de las actividades en el EVA (foro, chat, Videocolaboración), podemos decir, que con los estudiantes de modalidad a distancia en el periodo Abril - Agosto 2015 matriculados en materias de Primer Ciclo no se encuentran patrones de comportamiento relevantes, que reflejen la relación de los estudiantes con dichas actividades.

## CONCLUSIONES

Una vez realizada la experimentación con los datos obtenidos se puede concluir lo siguiente:

- En la realización de la limpieza de los datos se pudo observar que de toda la muestra recogida, solamente redujo un 2% debido a las diferentes inconsistencias presentadas, además se determinó que la mayor población es de género Femenino.

- La utilización de la técnica de Clustering, mediante el algoritmo Mapas o red de Kohonen utilizado en la primera experimentación del presente trabajo permitió la agrupación de los estudiantes de acuerdo a las diferentes actividades del EVA más utilizadas de la titulación Gestión Ambiental.

- La aplicación de la técnica de asociación mediante el algoritmo a priori, permitió la obtención de reglas de asociación entre los datos, no se pudieron encontrar patrones de comportamiento relevantes que puedan mostrar como los estudiantes se comportan en actividades del Eva, como chats, foros y videocolaboraciones.

- Dentro del análisis de los resultados de las reglas de asociación obtenidas, se puede decir que, los estudiantes de modalidad a distancia tienen mayor uso de la interacción baja en actividades como: Actualizar post, añadir foro, Ver historial de chat, Ver grabaciones y ver todas las videocolaboraciones

- Dentro de la interacción media las actividades usadas dentro del rango establecido son: ver foro y ver sala de Videocolaboración, mientras que en la interacción alta la única actividad que predomina es el número de veces que el estudiante tiene conversaciones.

## RECOMENDACIONES

A partir de los resultados encontrados en el presente trabajo de Titulación, y con el fin de dar continuidad y mejoras a trabajo futuros, se presentan las siguientes recomendaciones:

- Que el departamento encargado del manejo de la base de datos de la universidad, posea un diccionario de datos, el cual permita a los estudiantes guiarse para poder realizar la selección y extracción de datos.

- Realizar la investigación con datos de dos ciclos académicos para comparar si influye en algo que los estudiantes sean de primer ciclo o no, además de implementar datos como notas académicas generadas por el uso de las herramientas (foro, videocolaboración), para poder determinar si influye en el rendimiento académico.

- Para la recopilación y selección de datos, se debe tener muy claro el escenario en el cual se va a trabajar, para poder extraer los datos necesarios.

- Obtener e incluir datos personales y socioeconómicos, para ver si esto incluye o no, en el comportamiento de los estudiantes y poder encontrar patrones de comportamiento relevantes.

## BIBLIOGRAFÍA


- Arranz de la Peña, J., & Parra Truyol, A. (n.d.). ALGORITMOS GENÉTICOS.
- Belloch, C. (2012). Entornos Virtuales de Aprendizaje, 1–9.
- Cabero, J. (2006). Bases pedagógicas del e-learning. *DIM: Didáctica, Innovación y Multimedia*, (6).
- Cabero, J., & Gisbert, M. (2005). *La formación en Internet: guía para el diseño de materiales didácticos*. MAD-Eduforma. Retrieved from https://books.google.com/books?id=-sJrbH58xj0C&pgis=1
- Cardona, J. (2011). SISTEMA DE APOYO PARA LA ACREDITACIÓN DE LA CALIDAD DE PROGRAMAS ACADEMICOS DE LA UNIVERSIDAD DE CALDAS, APLICANDO TÉCNICAS EN MINERÍA DE DATOS.
- Cazau, P. (2010). Estilos de aprendizaje: Generalidades.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., & Shearer, C. (2000). Metodología CRISP-DM para minería de datos.
- Diwate, R. B., & Sahu, A. (2014). Data Mining Techniques in Association Rule : A Review.
- Escobar, P. (2009). CURSOS DE ESPECIALIZACIÓN EN : FORMACIÓN VIRTUAL.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, *17*(3), 37.
- Fernández, A. (2010). Las Plataformas E-Learning Para La Enseñanza Y El Aprendizaje Universitario en Internet. *Las Plataformas De Aprendizaje. Del Mito a La Realidad*, 45–73. Retrieved from http://eprints.ucm.es/10682/1/capituloE_learning.pdf
- Flórez, R., & Fernández, J. M. (2008). *Las redes neuronales artificiales : fundamentos teóricos y aplicaciones prácticas*. Netbiblo.
- Franco, C., Yamasaki, L., & Domínguez, A. (2010). Desarrollando un modelo adaptativo jerárquico basado en preferencias de aprendizaje, para entornos e-learning y b-learning para MOODLE. *Crisis analógica, futuro digital: actas del IV Congreso Online del Observatorio para la Cibersociedad, celebrado del 12 al 29 de noviembre de 2009*. Retrieved from http://dialnet.unirioja.es/servlet/citart?info=link&codigo=3324116&orden=272866
- Gallego, A., & Martinez, E. (2003). Estilos de aprendizaje y e-learning. Hacia un mayor rendimiento académico. *RED: Revista de Educación a Distancia*, (1), 1–10. http://doi.org/http://hdl.handle.net/10317/982
- García, C. (2002). E-learning-Teleformación . Diseño , desarrollo y evaluación de la formación a través de internet.
- Garg, A., & Neilsen, M. (2012). SCORM BASED LEARNING MANAGEMENT

SYSTEM FOR ONLINE TRAINING.

- Garre, M., Cuadrado, J., Silicia, M., Rodriguez, D., & Rejas, R. (n.d.). Comparación de diferentes algoritmos de clustering en la estimación de coste en el desarrollo de software.
- Gonzáles, A. (n.d.). Guía de apoyo para el uso de Moodle. Retrieved from https://download.moodle.org/docs/es/1.9.4_usuario_administrador.pdf
- Goopta, C. (2014). Six of the Best Open Source Data Mining Tools. Retrieved from http://thenewstack.io/six-of-the-best-open-source-data-mining-tools/
- Haro, V., & Perez, W. (2014). Data Warehouse para el Centro de Documentacion Regional "Juan Bautista Vazquez." Retrieved January 25, 2016, from http://dspace.ucuenca.edu.ec/bitstream/123456789/19878/1/tesis.pdf
- heatherwilliamson. (n.d.). Definitions: Technology enhanced learning environments areas. Retrieved from http://www.jisc.ac.uk/whatwedo/programmes/elearning/tele/definitions.aspx
- Hernandez, H., & Abilowo, R. (n.d.). Evaluación de modelos para la predicción de la Bolsa. Retrieved January 25, 2016, from http://www.it.uc3m.es/jvillena/irc/practicas/06-07/25.pdf
- Hernández, J., Ramírez, M. J., & Ferri, C. (2004). *Introducción a la minería de datos*. Editorial Alhambra S. A. (SP). Retrieved from https://books.google.com/books?id=x3LuAAAACAAJ&pgis=1
- IBM Knowledge Center. (2013, January 1). Retrieved from http://www-01.ibm.com/support/knowledgecenter/SSEPGG_9.7.0/com.ibm.im.overview.doc/c_naive_bayes_classification.html?lang=es
- Kdnuggets. (2014). What main methodology are you using for your analytics, data mining, or data science projects ? Retrieved from http://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html
- KDnuggets. (2015). Poll: What Analytics, Data Mining, Data Science software/tools you used in the past 12 months? Retrieved from http://www.kdnuggets.com/polls/2015/analytics-data-mining-data-science-software-used.html
- KDnuggets. (2016). KDnuggets Analytics/Data Science 2016 Software Poll: top 10 most popular tools in 2016. Retrieved from http://www.kdnuggets.com/2016/06/r-python-top-analytics-data-mining-data-science-software.html
- Kosorus, H., Honigl, J., & Kung, J. (2011). Using R, WEKA and RapidMiner in Time Series Analysis of Sensor Data for Structural Health Monitoring. In *2011 22nd International Workshop on Database and Expert Systems Applications* (pp. 306–310).

IEEE. http://doi.org/10.1109/DEXA.2011.88

- Malberti, M., & Elida, G. (2015). Reglas de Asociación con los datos de una biblioteca universitaria. *Revista Cubana de Ciencias Informáticas*, *9*(4), 30–45.
- Martinez, I. (2008). Moodle, la plataforma para la enseñanza y organización escolar, 1–12. Retrieved from https://addi.ehu.es/handle/10810/6876
- Merelo, J. J. (2004). Tutorial: mapas organizativos de Kohonen Mapa autoorganizativo de Kohonen. Retrieved from http://geneura.ugr.es/~jmerelo/tutoriales/bioinfo/Kohonen.html
- Miratía, O. (2008). Moodle y Dokeos. Dos plataformas de software libre para la educación a distancia. *VI Congreso Internacional de Educación Superior "Universidad 2008,"* (November), 1–9. Retrieved from http://www.researchgate.net/publication/229010433
- Molina, J., & García, J. (n.d.). Técnicas de Minería de Datos basadas en Aprendizaje Automático. Retrieved February 19, 2016, from https://santiagozapatakdd.files.wordpress.com/2011/03/curso-kdd-full-cap-3.pdf
- Morales, E. (2010). *Gestión del conocimiento en sistemas «e-learning», basado en objetos de aprendizaje, cualitativa y pedagógicamente definidos*. Universidad de Salamanca. Retrieved from https://books.google.com/books?id=Z9y6-5fKOGkC&pgis=1
- Moreira, M., & Segura, J. (2009). E-learning: enseñar y aprender en espacios virtuales, 1–29.
- Ortiz F., L. F. (2007). Campus Virtual: la educación más allá del LMS. *RUSC. Universities and Knowledge Society Journal*, *4*(1), 3. http://doi.org/10.4067/S0718-50062011000600005
- PARRA, W., & RODRIGUEZ, A. (2007). SOFTWARE DE APOYO AL DIAGNÓSTICO Y CLASIFICACIÓN DE ESTUDIANTES POR ESTILO DE APRENDIZAJE EN EL SISTEMA DE GESTIÓN DE APRENDIZAJE MOODLE. Retrieved December 22, 2015, from http://repositorio.uis.edu.co/jspui/bitstream/123456789/2487/2/122725.pdf
- Pascual, D., & Sánchez, S. (n.d.). Algoritmos de agrupamiento. Retrieved from http://marmota.dlsi.uji.es/WebBIB/papers/2007/1_Pascual-MIA-2007.pdf
- Pérez, C., & Santín, D. (2007). *Minería de datos: técnicas y herramientas*. Retrieved from https://books.google.com/books?hl=es&lr=&id=wz-D_8uPFCEC&pgis=1
- Pitol, F. (2014). Reglas de asociación, algoritmo apriori. Retrieved from http://ferminpitol.blogspot.com/2014/05/reglas-de-asociacion-algoritmo-apriori.html
- Ramagiri, R., & Jeffery, C. (n.d.). Integrating the Moodle Course Management System into a Collaborative Virtual Environment Project Report Submitted in Partial Fulfillment for the Degree of Master's in Computer Science.

- Ramírez, A. (2007). Técnicas de minería de datos aplicadas a la construcción de modelos de score crediticio : Estado del arte. *Universidad Nacional de Colombia*, 1–10.
- Rivera, K. (2004). *Índice*.
- Sael, N., Marzak, A., & Behja, H. (2013). Web Usage Mining Data Preprocessing and Multi Level Analysis on Moodle.
- Samaniego, J. B. (2016). Análisis de patrones de comportamiento de los estudiantes que utilizan dispositivos móviles en su proceso de enseñanza aprendizaje, a través de técnicas de minería de datos.
- Timarán, R., & Jiménez, J. (2009). *Detección de Patrones de Deserción Estudiantil en Programas de Pregrado de Instituciones de Educación Superior con CRISP-DM*. Fundación Iberoamericana para la Educación, la Ciencia y la Cultura. Retrieved from www.oei.es/historico/congreso2014/memoriactei/758.pdf
- Valle, Y., Arévalo, C., & Muñoz, J. (2014). Conversión de Mapas Conceptuales en Objetos de Aprendizaje bajo el estándar SCORM, 2–7.
- Weka 3: Software de minería de datos en Java. (n.d.). Retrieved from http://www.cs.waikato.ac.nz/ml/weka/

# ANEXOS

## ANEXO 1

### Carrera y materias de modalidad distancia extraídas.

| Categoria | Carrera | Area | Materia | Tipo Materia | Total paralelos | Total alumnos |
|---|---|---|---|---|---|---|
| 5524 | Gestion Ambiental | Biologica | Introduccion a las ciencias ambientales | Troncal | 16 | 562 |
| | | | Biologia General | Generica | 14 | 659 |
| 5510 | Informatica | Técnica | Fundamentos Informáticos | Troncal | 9 | 387 |
| | | | Logica de la programacion | Troncal | 8 | 233 |
| 5504 | Economía etc | Administrativa | Historia del pensamiento económico | Troncal | 6 | 221 |
| | | | Matemáticas | Generica | 20 | 1261 |
| 5512 | Derecho | SocioHumanistica | Derecho constitucional | Troncal | 26 | 1603 |
| | | | Introduccion al derecho | Generica | 24 | 1379 |
| 5520 | Comunicación Social | SocioHumanistica | Introduccion a la comunicació | Troncal | 6 | 274 |
| | | | Literatura | Generica | 6 | 259 |
| 5509 | Psicología | SocioHumanistica | Psicología social | Troncal | 14 | 858 |
| | | | Psicologia general | Generica | 23 | 1328 |
| 5516 | Ciencias de la educacion mención Educación Básica | SocioHumanistica | Sociología de la educación | Troncal | 8 | 349 |
| | | | Pedagogía general | Generica | 16 | 814 |
| 5517 | Ciencias de la educacion mención Educación Infantil | SocioHumanistica | Introduccion a la educación pr | Troncal | 6 | 282 |
| | | | Pedagogía general | Generica | | |
| 5527 | Ciencias de la educacion mención Inglés | SocioHumanistica | Reading and writting I | Generica | 6 | 242 |
| | | | Communicative grammar | Generica | 7 | 275 |
| 5513 | Químico Biológicas | SocioHumanistica | Biología general para educad | Troncal | 2 | 44 |

## Anexo 2

### Consultas SQL para la extracción de datos

De las consultas que se muestran a continuación se extraen las acciones que realizan los estudiantes de modalidad abierta y a distancia en el EVA, con sus respectivos atributos y la carrera que cursan.

```
SELECT DISTINCT
    user.userid as idUsuario, user.sexo as Genero, Ma.MateriasMatriculado ,
    c.id as idCurso, UPPER(enRol.centro) AS Centro, su.componente, su.paralelo,
    Interaccion.module ,Interaccion.Actividad, Interaccion.NumAcciones, pdo.name
FROM  mdl_user_utpl user,
(       select  user.userid, count(distinct c.id) as MateriasMatriculado, user.sexo as GENERO, rol.roleid
        from mdl_user_utpl as user, mdl_role_assignments as rol, mdl_context as context, mdl_course as c,
            mdl_syllabus_pdo as pdo, mdl_course_categories as cc, mdl_course_sections cs
        where   user.userid = rol.userid    and rol.roleid =5 and context.contextlevel=50 and context.id=rol.contextid
                and context.instanceid=c.id
                and cs.course=c.id  and pdo.pdoid ='08d494b3-9baf-0098-e053-ac10360d0098'
                and cc.id in (5524) and c.category=cc.id
        group by user.id
) Ma,
(    SELECT DISTINCT log.action AS Actividad, COUNT(log.action) AS NumAcciones,log.module,log.course,log.userid
    FROM  mdl_log AS log
    WHERE log.course IN (65005,65006,65007,65008,65009,65010,65011,65012,65013,65014,65015,65016,65017,65018,65019,65020
                        ,64990 , 64991, 64992, 64993, 64994, 64995, 64996, 64997, 64998, 64999, 65000, 65001, 65002, 65003 )
            AND (log.module = 'chat' || log.module = 'forum' || log.module = 'elluminate')
    GROUP BY log.course , log.userid , log.module , log.action
) Interaccion
    inner join mdl_role_assignments assg on assg.userid = Interaccion.userid
    inner join mdl_enrol_utpl enRol on enRol.userid =  Interaccion.userid
    inner join mdl_course  c on c.id= enRol.courseid
    inner join mdl_syllabus_utpl su on su.courseid= c.id
    inner join mdl_syllabus_pdo pdo on pdo.id= su.pdoid
    inner join mdl_context contex on contex.id= assg.contextid and contex.instanceid = c.id  and contex.contextlevel=50

WHERE assg.roleid = 5 AND  Interaccion.userid = user.userid and Interaccion.userid = Ma.userid
        AND c.id IN (65005,65006,65007,65008,65009,65010,65011,65012,65013,65014,65015,65016,65017,65018,65019,65020
                    ,64990 , 64991, 64992, 64993, 64994, 64995, 64996, 64997, 64998, 64999, 65000, 65001, 65002, 65003)
ORDER BY su.componente,su.paralelo , user.userid , Interaccion.module , Interaccion.Actividad
```

## Anexo 3

### Generación de Vistas para datos minables

Vista #1

```
CREATE VIEW `economia_V1` AS(
    select economia.idUsuario, economia.Genero, economia.MateriasMatriculado,
           economia.Carrera, economia.Area, economia.TipoMateria,
           economia.idCurso, economia.Centro, economia.componente,
           economia.paralelo, economia.name,
        #Chat  5
        case when Actividad='historial' and module='chat' then NumAcciones end as chat_historial ,
        case when Actividad='report' and module='chat' then NumAcciones end as chat_report,
        case when Actividad='talk' and module='chat' then NumAcciones end as chat_talk ,
        case when Actividad='view'  and module='chat' then NumAcciones end as chat_viewChat ,
        case when Actividad='view all' and module='chat' then NumAcciones end as chat_viewAllChat ,
        #VideoColaboracion    4
        case when Actividad='view all' and module='elluminate'then NumAcciones end as viewAllElluminate ,
        case when Actividad='view meeting' and module='elluminate' then NumAcciones end as viewMeeting ,
        case when Actividad='view recording' and module='elluminate' then NumAcciones end as viewRecording ,
        case when Actividad='view' and module='elluminate' then NumAcciones end as viewElluminate ,
        #forum 13
        case when Actividad='add post' and module='forum' then NumAcciones end as addPostForum ,
        case when Actividad='delete post' and module='forum' then NumAcciones end as deletePost ,
        case when Actividad='mark read' and module='forum' then NumAcciones end as markRead,
        case when Actividad='search' and module='forum' then NumAcciones end as search ,
        case when Actividad='subscribe' and module='forum' then NumAcciones end as subscribe ,
        case when Actividad='subscribeall' and module='forum' then NumAcciones end as subscribeall ,
        case when Actividad='unsubscribe' and module='forum' then NumAcciones end as unsubscribe,
        case when Actividad='unsubscribe all' and module='forum' then NumAcciones end as unsubscribeAll,
        case when Actividad='update post' and module='forum' then NumAcciones end as  updatePost,
        case when Actividad='view discussion' and module='forum' then NumAcciones end as viewDiscussion ,
        case when Actividad='view forum' and module='forum' then NumAcciones end as viewForum ,
        case when Actividad='view forums'  and module='forum' then NumAcciones end as viewForums ,
        case when Actividad='view subscribers' and module='forum' then NumAcciones end as viewSubscribers
    from economia
);
```

Vista #2

```
CREATE VIEW `economia_V2` AS(
        select idUsuario, Genero, MateriasMatriculado, Carrera, Area,
        TipoMateria, idCurso, Centro, componente, paralelo,name,
        sum(chat_historial) as chat_historial,
        sum(chat_report) as chat_report,
        sum(chat_talk) as chat_talk,
        sum(chat_viewAllChat) as chat_viewAllChat,
        sum(chat_viewChat) as chat_viewChat,
        sum(viewAllElluminate) as viewAllElluminate,
        sum(viewMeeting) as viewMeeting ,
        sum(viewRecording) as viewRecording,
        sum(viewElluminate) as viewElluminate ,
        sum(addPostForum) as addPostForum ,
        sum(deletePost) as deletePost,
        sum(markRead) as markRead,
        sum(search) as search,
        sum(subscribe) as subscribe,
        sum(subscribeall) as subscribeall,
        sum(unsubscribe) as unsubscribe ,
        sum(unsubscribeAll) as unsubscribeAll ,
        sum(updatePost) as updatePost ,
        sum(viewDiscussion) as viewDiscussion,
        sum(viewForum) as viewForum,
        sum(viewForums) as viewForums,
        sum(viewSubscribers) as viewSubscribers
    from economia_V1
    group by idUsuario,componente
);
```

Vista #3

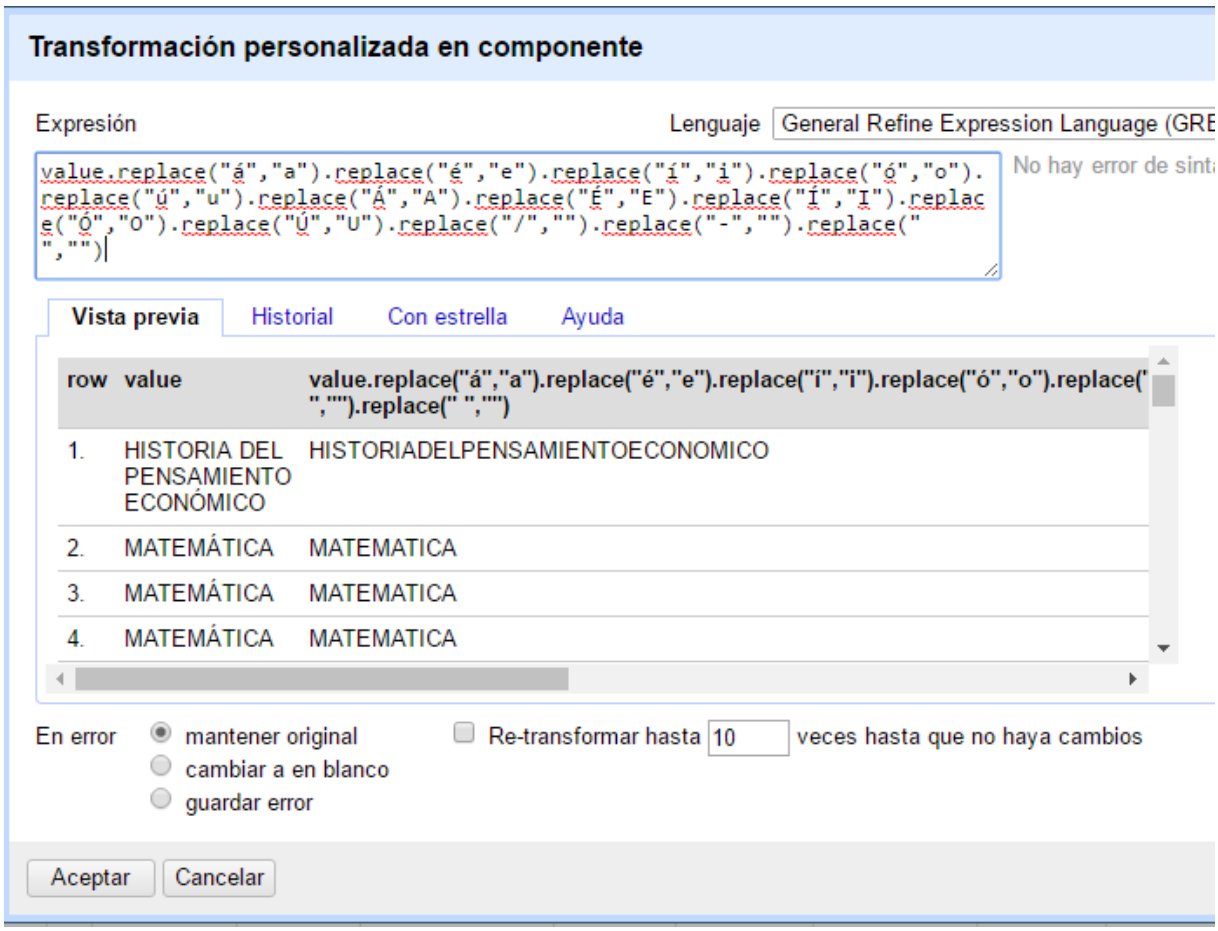
```sql
create view economia_V3 as (
    select idUsuario, Genero, MateriasMatriculado, Carrera, Area,
    TipoMateria, idCurso, Centro, componente, paralelo,name,
    coalesce(chat_historial,0)as chat_historial,
    coalesce(chat_report,0)as chat_report,
    coalesce(chat_talk ,0)as chat_talk,
    coalesce(chat_viewAllChat ,0)as chat_viewAllChat,
    coalesce(chat_viewChat ,0)as chat_viewChat,

    coalesce(viewAllElluminate,0)as viewAllElluminate,
    coalesce(viewMeeting,0)as viewMeeting,
    coalesce(viewRecording ,0)as viewRecording,
    coalesce(viewElluminate ,0)as viewElluminate,

    coalesce(addPostForum,0)as addPostForum,
    coalesce(deletePost,0)as deletePost,
    coalesce(markRead,0)as markRead, #ueva
    coalesce(search,0)as search,
    coalesce(subscribe,0)as subscribe,
    coalesce(subscribeall,0) as subscribeall,
    coalesce(unsubscribe,0) as unsubscribe,
    coalesce(unsubscribeAll,0) as unsubscribeAll,
    coalesce(updatePost,0) as updatePost,
    coalesce(viewDiscussion,0)as viewDiscussion,
    coalesce(viewForum,0) as viewForum,
    coalesce(viewForums,0) as viewForums,
    coalesce(viewSubscribers,0) as viewSubscribers
    from economia_V2
    order by  idUsuario
);
```

## Anexo 4

### Limpieza y transformación de datos en OpenRefine



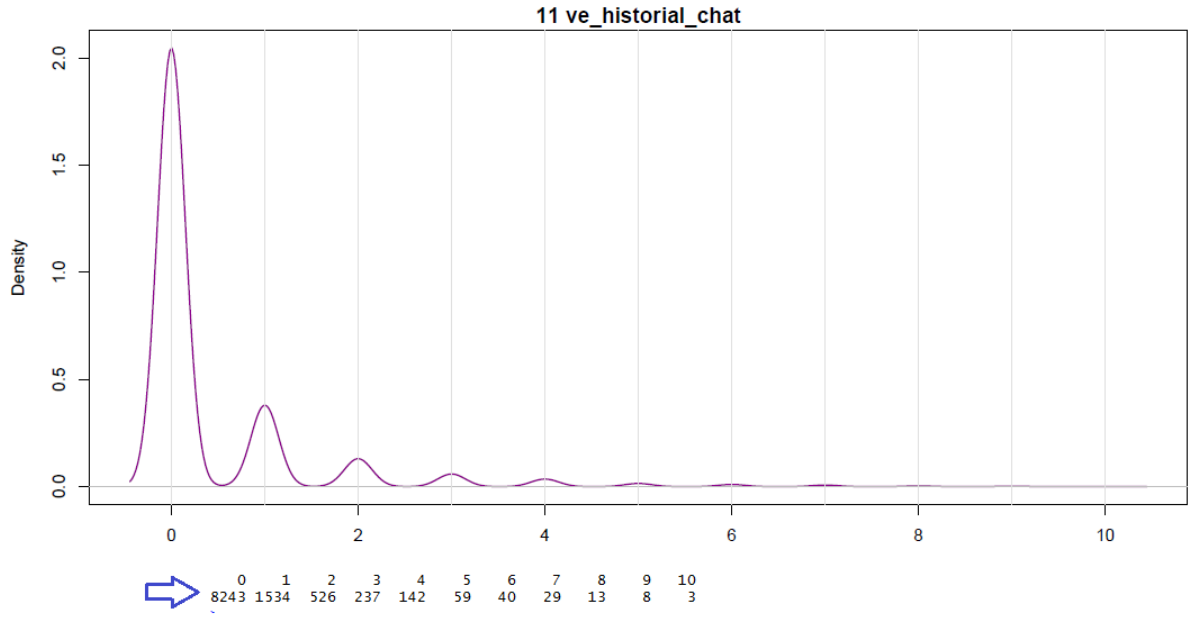
## Limpieza y transformación de datos en R Studio

**Variable:** Ve_historial_chat

Sin limpieza

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 17 | 18 | 19 | 21 | 23 | 27 | 29 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8217 | 1534 | 526 | 237 | 142 | 59 | 40 | 29 | 13 | 8 | 3 | 6 | 5 | 1 | 4 | 1 | 1 | 3 | 1 | 1 | 1 | 1 | 1 |

Con limpieza



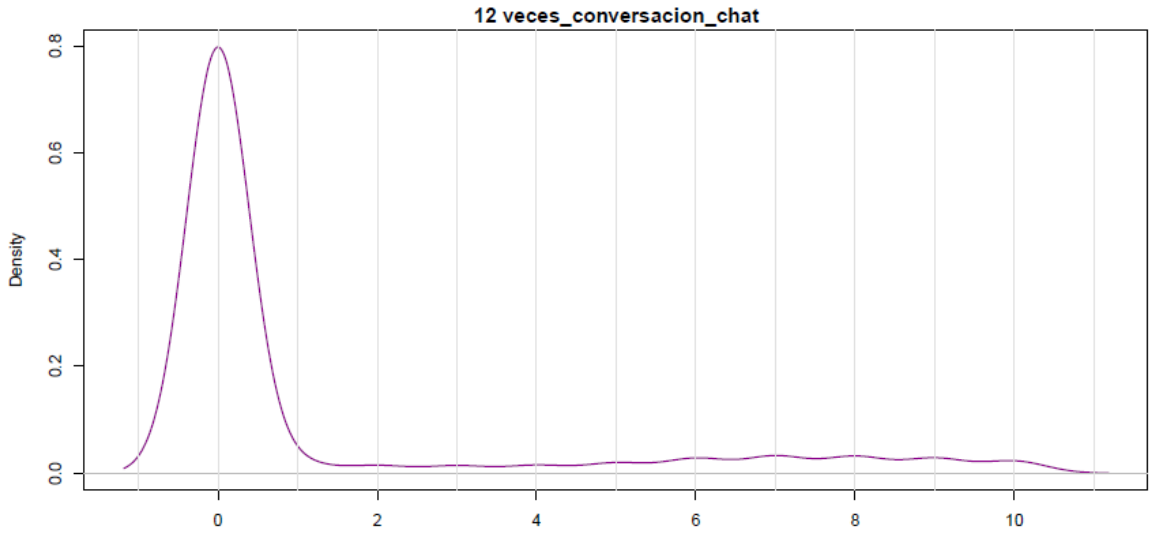

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 8243 | 1534 | 526 | 237 | 142 | 59 | 40 | 29 | 13 | 8 | 3 |

**Variable:** veces_conversacion_chat

Sin limpieza

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6721 | 195 | 147 | 144 | 152 | 198 | 286 | 330 | 321 | 288 | 238 | 246 | 185 | 163 | 147 | 129 | 107 | 110 | 79 | 69 | 72 | 48 | 47 | 40 | 37 | 33 | 29 | 27 | 20 |
| 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 56 | 59 | 62 | 63 | 68 | 70 |
| 20 | 22 | 22 | 19 | 16 | 12 | 10 | 13 | 10 | 9 | 6 | 4 | 7 | 7 | 5 | 2 | 2 | 3 | 2 | 3 | 3 | 7 | 3 | 1 | 2 | 1 | 1 | 4 | 1 |
| 73 | 75 | 76 | 77 | 78 | 116 | 132 | 134 | | | | | | | | | | | | | | | | | | | | | |
| 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | | | | | | | | | | | | | | | | | | | | | |

Con limpieza



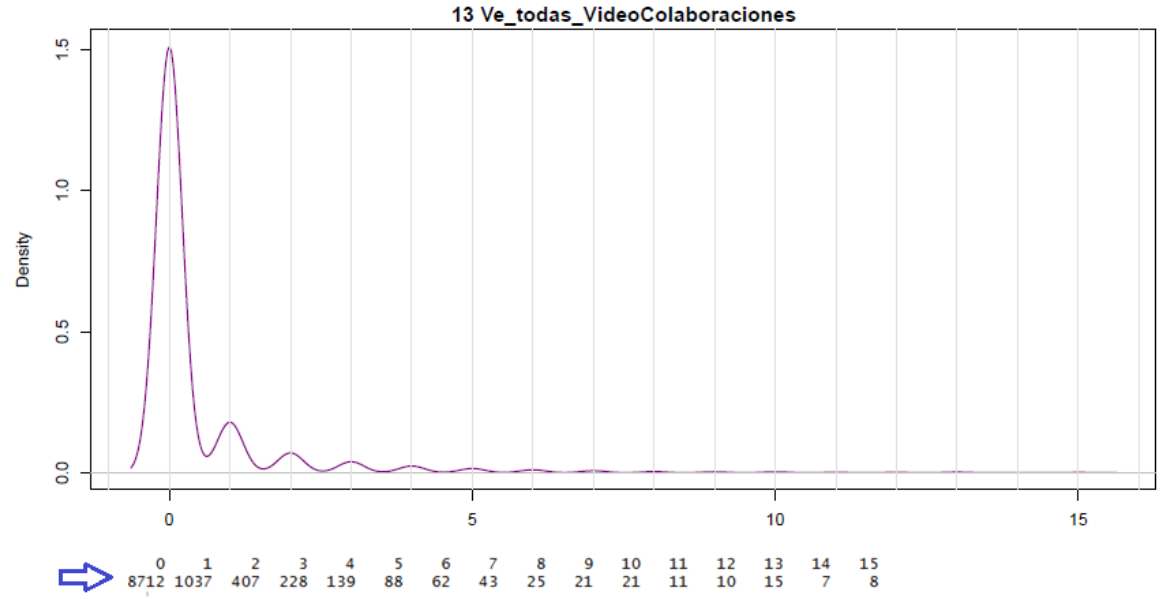

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 8535 | 195 | 147 | 144 | 152 | 198 | 286 | 330 | 321 | 288 | 238 |

**Variable:** Ve_todas_VideoColaboraciones

Sin limpieza

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8683 | 1037 | 407 | 228 | 139 | 88 | 62 | 43 | 25 | 21 | 21 | 11 | 10 | 15 | 7 | 8 | 1 | 1 | 7 | 1 | 1 | 3 | 3 | 2 | 1 |
| 25 | 27 | 28 | 31 | 32 | 39 | 50 | | | | | | | | | | | | | | | | | | |
| 1 | 1 | 1 | 2 | 2 | 1 | 1 | | | | | | | | | | | | | | | | | | |

Con limpieza



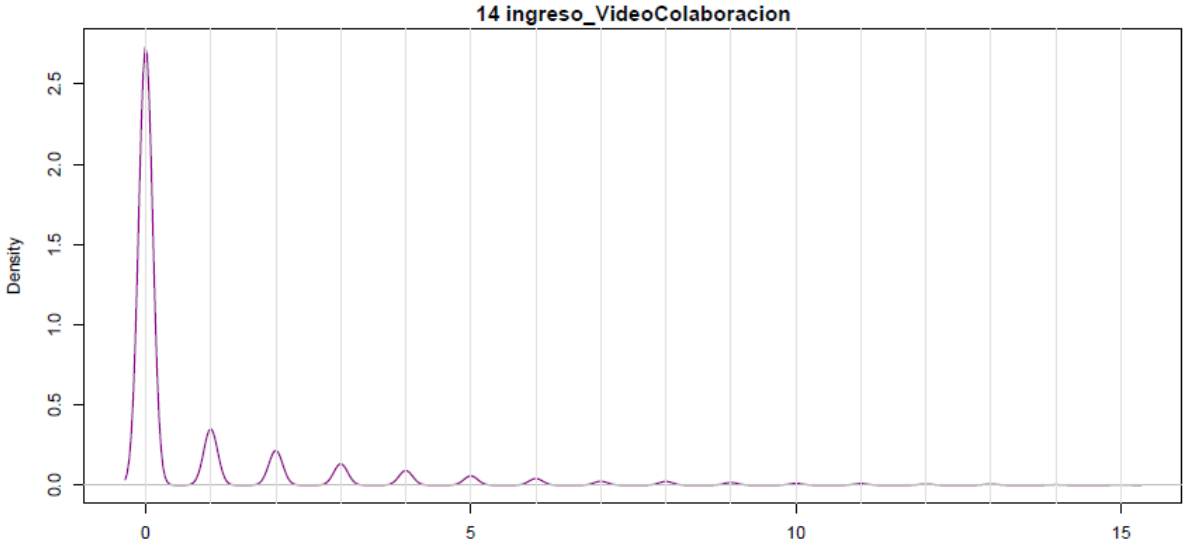

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8712 | 1037 | 407 | 228 | 139 | 88 | 62 | 43 | 25 | 21 | 21 | 11 | 10 | 15 | 7 | 8 |

**Variable:** ingreso_VideoColaboracion

Sin limpieza

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7880 | 1017 | 626 | 386 | 263 | 167 | 120 | 70 | 67 | 48 | 32 | 29 | 20 | 19 | 9 | 8 | 14 | 4 | 6 | 5 | 4 | 4 | 3 | 3 | 5 | 4 | 4 | 1 | 2 | 1 |
| 31 | 32 | 34 | 37 | 38 | 50 | 54 | 73 | 76 | 243 | | | | | | | | | | | | | | | | | | | | |
| 1 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | | | | | | | | | | | | | | | | | | | | |

Con limpieza



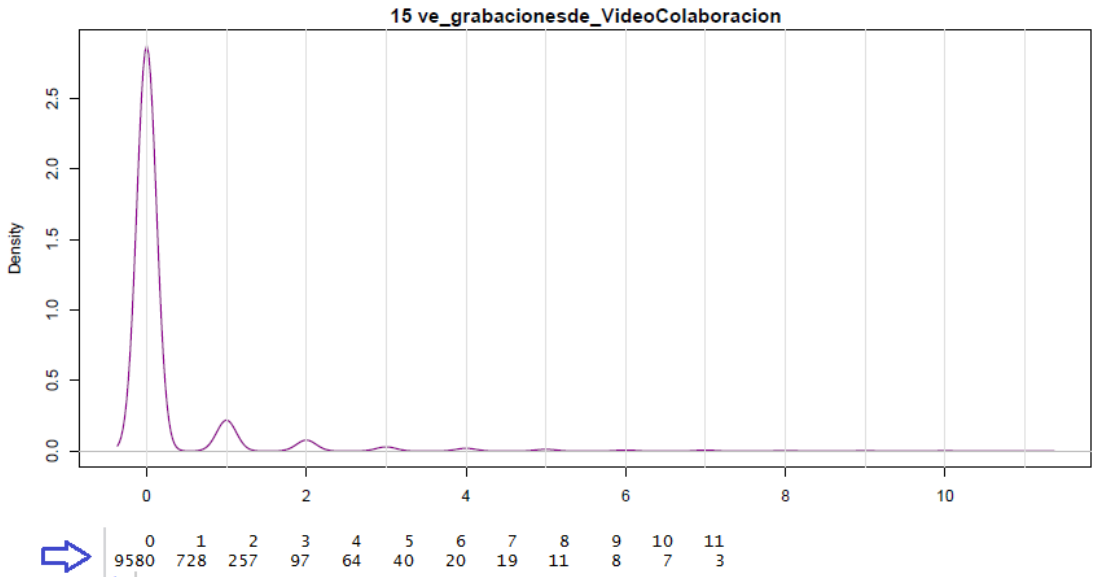

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7953 | 1017 | 626 | 386 | 263 | 167 | 120 | 70 | 67 | 48 | 32 | 29 | 20 | 19 | 9 | 8 |

**Variable:** ve_grabacionesde_VideoColaboracion

Sin limpieza

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 13 | 14 | 15 | 20 | 21 | 23 | 24 | 41 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 9571 | 728 | 257 | 97 | 64 | 40 | 20 | 19 | 11 | 8 | 7 | 3 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Con limpieza



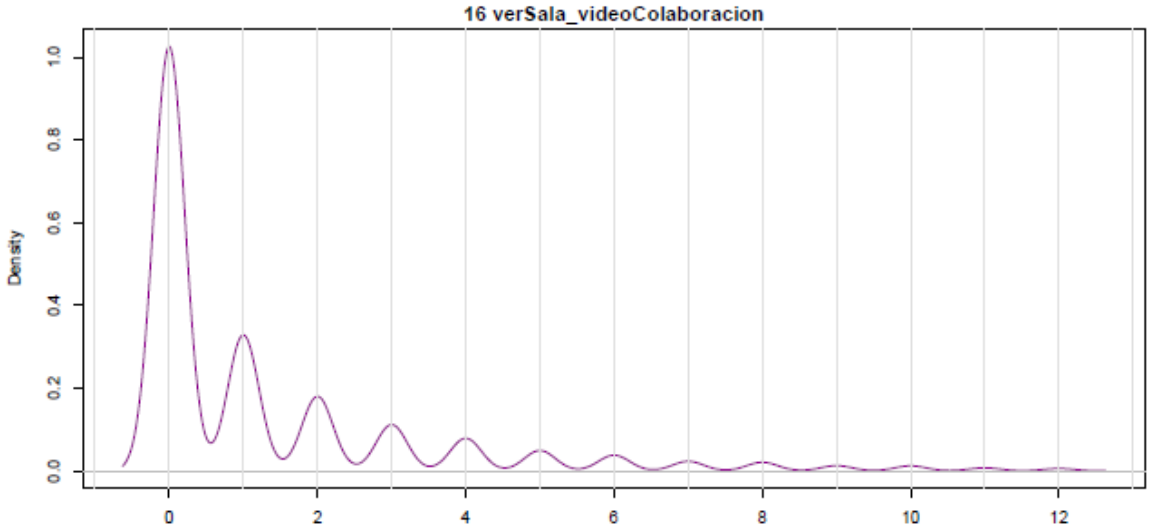

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 9580 | 728 | 257 | 97 | 64 | 40 | 20 | 19 | 11 | 8 | 7 | 3 |

**Variable:** verSala_videoColaboracion

Sin limpieza

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5668 | 1881 | 1031 | 640 | 451 | 282 | 218 | 133 | 120 | 73 | 71 | 43 | 34 | 31 | 30 | 24 | 17 | 14 | 12 | 9 | 7 | 6 | 1 | 7 | 5 | 3 | 3 | 1 | 3 | 2 |
| 30 | 31 | 32 | 33 | 34 | 37 | 39 | 43 | 44 | 48 | | | | | | | | | | | | | | | | | | | | |
| 1 | 2 | 1 | 2 | 1 | 1 | 3 | 1 | 1 | 1 | | | | | | | | | | | | | | | | | | | | |

Con limpieza



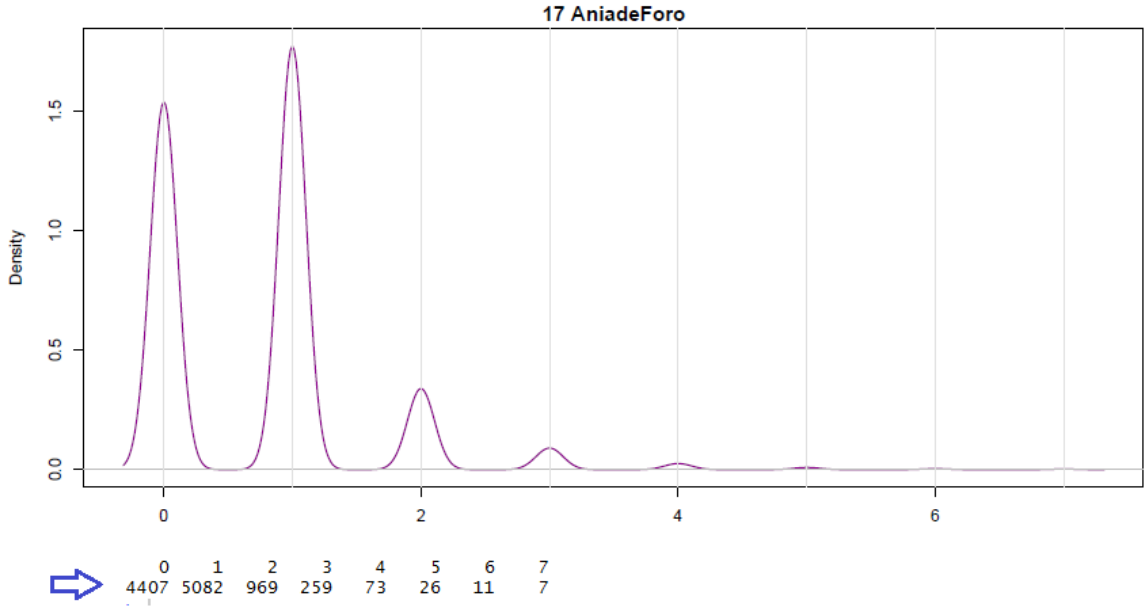

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5857 | 1881 | 1031 | 640 | 451 | 282 | 218 | 133 | 120 | 73 | 71 | 43 | 34 |

**Variable:** AniadeForo

Sin limpieza

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 14 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4391 | 5082 | 969 | 259 | 73 | 26 | 11 | 7 | 7 | 2 | 4 | 1 | 1 | 1 |

Con limpieza

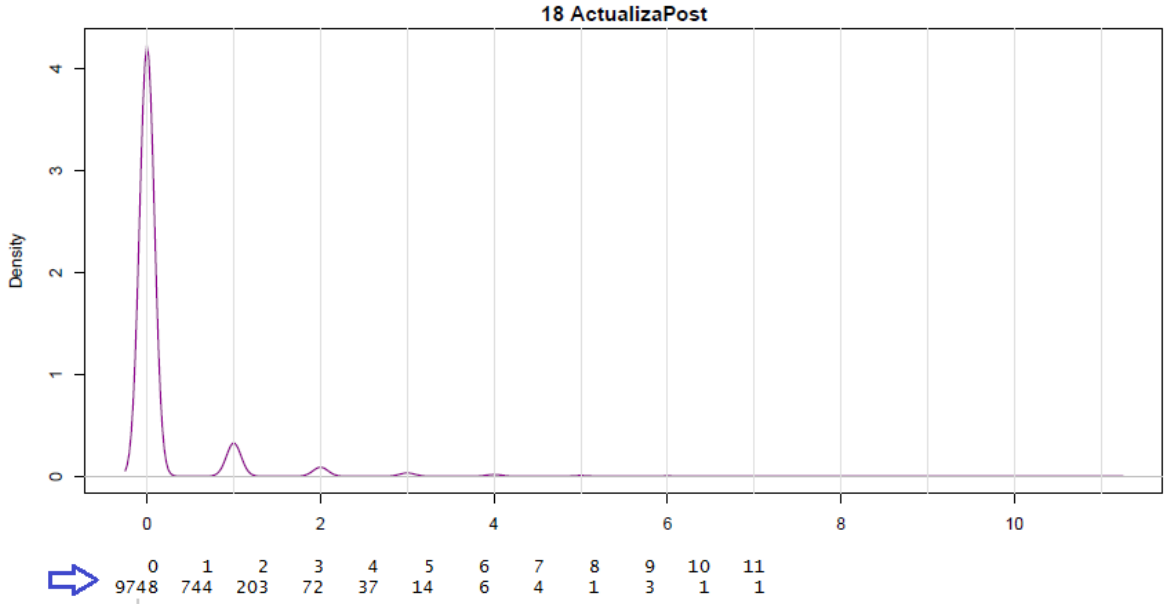



**Variable:** ActualizaPost

Sin limpieza

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 9748 | 744 | 203 | 72 | 37 | 14 | 6 | 4 | 1 | 3 | 1 | 1 |

Con limpieza

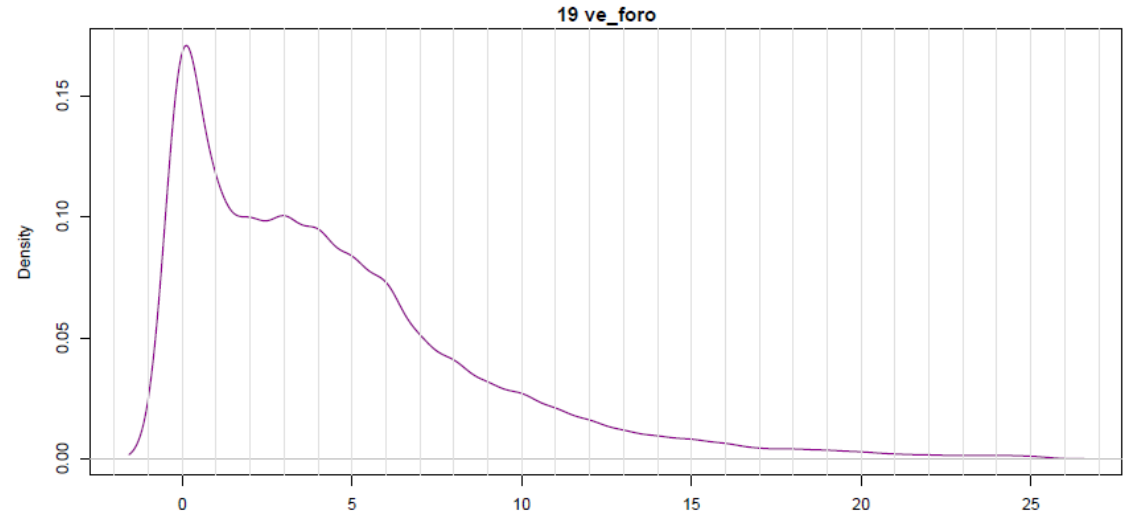



**Variable:** ve_foro

Sin limpieza

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2110 | 1133 | 1057 | 1090 | 1028 | 892 | 807 | 527 | 441 | 333 | 294 | 223 | 172 | 125 | 101 | 88 | 70 | 45 | 45 | 39 | 32 | 21 | 18 | 14 | 17 | 13 | 10 | 13 | 6 | 7 |
| 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 49 | 50 | 51 | 54 | 55 | 56 | 58 | 60 | 62 | 64 | 72 | 78 |
| 8 | 6 | 5 | 2 | 4 | 3 | 2 | 4 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 84 | 87 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Con limpieza




| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2209 | 1133 | 1057 | 1090 | 1028 | 892 | 807 | 527 | 441 | 333 | 294 | 223 | 172 | 125 | 101 | 88 | 70 | 45 | 45 | 39 | 32 | 21 | 18 | 14 | 17 | 13 |

## Anexo 5

### Grafo generado de las reglas de asociación según métrica

En el presente grafo se muestra cada una de las reglas de asociación generadas en la herramienta R Studio, cabe destacar que los nodos más pequeños hacen referencia a un soporte menor