



*Minería y almacenamiento de archivos de las aulas virtuales de la Universidad  
Técnica de Machala*

*Mining and file storage of virtual classrooms of the Technical University of  
Machala*

*Mineração e armazenamento de arquivos de salas de aula virtuais da  
Universidade Técnica de Machala*

Karina Elizabeth García-Galarza <sup>I</sup>  
[gkarinae@psg.ucacue.edu.ec](mailto:gkarinae@psg.ucacue.edu.ec)  
<https://orcid.org/0000-0001-5434-2143>

Javier Bernardo Cabrera-Mejía <sup>II</sup>  
[jcabreram@ucacue.edu.ec](mailto:jcabreram@ucacue.edu.ec)  
<https://orcid.org/0000-0003-2027-0211>

**Correspondencia:** [gkarinae@psg.ucacue.edu.ec](mailto:gkarinae@psg.ucacue.edu.ec)

Ciencias de las ingenierías  
Artículo de investigación

**\*Recibido:** 25 de noviembre de 2019 **\*Aceptado:** 29 diciembre de 2019 **\* Publicado:** 17 de enero 2020

- I. Ingeniero de Sistemas, Jefatura de Posgrados Universidad Católica de Cuenca, Cuenca, Ecuador.
- II. Ingeniero en Electrónica, Docente de la Unidad Académica de Ingeniería, Industria y Construcción, Jefatura de Posgrados Universidad Católica de Cuenca, Cuenca, Ecuador.

## Resumen

La minería de datos es un campo emergente, que puede ayudar a la obtención de data de una organización determinada. El presente artículo busca determinar la capacidad de almacenamiento y la cantidad de datos creados en la plataforma virtual de la Universidad Técnica de Machala, que fueron generados en los años 2016, 2017 y 2018. Así también se propone determinar qué tipo de documentos se generan y en qué formato. La investigación se basa en cuatro pasos: selección de datos, pre-procesamiento de datos, minería de datos y análisis de resultados. La selección de datos consistió en recuperar la información de los respaldos de la base de datos Postgres mediante Navicat, el pre-procesamiento se ejecutó mediante el software Open Refine, realizando una depuración de los datos, para la aplicación de la minería se utilizó el software Weka empleando técnicas de clasificación y clúster: árbol de decisiones y K-means. Finalmente, en el análisis de resultados se derivó como principal conclusión lo que sigue a continuación: mediante el uso de un modelo matemático podemos obtener una mejor precisión en la predicción del almacenamiento de las aulas virtuales de la Universidad Técnica de Machala.

**Palabras clave:** Aula virtual; minería de datos; base de datos; plataforma virtual; universidad.

## Abstract

Data mining is an emerging field, which can help to obtain data from a particular organization. This article seeks to determine the storage capacity and the amount of data created in the virtual platform of the Technical University of Machala, which were generated in 2016, 2017 and 2018. Thus, it is also proposed to determine what type of documents are generated and in what format. The research is based on four steps: data selection, data pre-processing, data mining and analysis of results. The data selection consisted of recovering the information of the backups of the Postgres database through Navicat, the pre-processing was executed through the Open Refine software, performing a data purification, for the mining application the software was used Weka using classification and cluster techniques: decision tree and K-means. Finally, the following conclusion was derived in the analysis of results as follows: through the use of a mathematical model we can obtain a better precision in the prediction of the storage of virtual classrooms of the Technical University of Machala.

**Keywords:** Virtual classroom; data mining; database; virtual platform; college

## Resumo

A mineração de dados é um campo emergente, que pode ajudar a obter dados de uma organização específica. Este artigo procura determinar a capacidade de armazenamento e a quantidade de dados criados na plataforma virtual da Universidade Técnica de Machala, gerados em 2016, 2017 e 2018. Assim, propõe-se também determinar que tipo de documentos são gerados e em que formato. A pesquisa se baseia em quatro etapas: seleção de dados, pré-processamento de dados, mineração de dados e análise de resultados. A seleção dos dados consistiu em recuperar as informações dos backups do banco de dados Postgres através do Navicat, o pré-processamento foi executado através do software Open Refine, realizando uma purificação de dados, para o aplicativo de mineração o software foi utilizado Weka usando técnicas de classificação e cluster: árvore de decisão e meios K. Finalmente, a seguinte conclusão foi derivada na análise dos resultados da seguinte forma: através do uso de um modelo matemático, podemos obter uma melhor precisão na previsão do armazenamento de salas de aula virtuais da Universidade Técnica de Machala.

**Palavras-chave:** Sala de aula virtual; mineração de dados; banco de dados; plataforma virtual; faculdade

## Introducción

La incorporación y el uso de la web en el proceso enseñanza – aprendizaje comporta hoy por hoy un complemento para el desarrollo de las actividades académicas, en cuyo contexto la interacción del profesor y el alumno se establece mediante una comunicación sincrónica y asincrónica respectivamente. La manera en que los estudiantes y profesores se comunican entre sí, en la enseñanza virtual o e-learning, favorece un intercambio de conocimientos y permite que puedan aprovechar diferentes recursos y medios, para involucrarse en una combinación de actividades, y al mismo tiempo, propiciar la construcción de aprendizajes colaborativos de manera rápida, efectiva, actualizada y oportuna. (Chen, Xu, Tang, & Chen, 2017).

Independientemente del lugar o coincidir en el tiempo, un alumno puede interactuar y requerir un recurso compartido con otros compañeros, revisar un material propuesto por el profesor e incluso hallar nuevas formas de establecer relaciones sociales. En ese orden, la Universidad Técnica de Machala (UTMACH) ha implementado un entorno virtual de aprendizaje (EVA), usando la plataforma Moodle como mecanismo de ayuda para la docencia en modalidad presencial, cuya

característica primordial sea la cercanía y cohesión del grupo, que genere un entorno socio-afectivo propicio para el proceso de aprendizaje.

El EVA es un sistema de enseñanza aprendizaje virtual que está clasificado por categorías, las cuales corresponden en primer orden a las facultades, luego a las carreras, a su vez en los periodos académicos y finalmente en las cátedras. Cada aula virtual puede restringir el uso público, es decir sólo tendrán acceso estudiantes y profesores de la universidad, en correspondencia con el manejo de usuarios y claves en cada una de las cuentas de los profesores y estudiantes.

Desde ese escenario, cabe resaltar que grandes sistemas informáticos generan volúmenes de información, que con un buen tratamiento se pueden convertir en conocimiento para la ayuda en la toma de decisiones tanto académicas como organizativas y gerenciales (Ochoa, L. L., Paredes, K. R., Araya, C. B., Santa, U. C. De, Ochoa, L. L., Paredes, K. R., & Araya, C. B. (2017). Es por eso, que, al utilizar la herramienta que permite compilar una minería de data, se trata de transformar esos datos crudos en información válida. En ese orden y según, (Martínez-abad & Hernández-Ramos, 2018), la minería de datos, se configura en un conjunto de procedimientos estadísticos, que intenta descubrir patrones en grandes volúmenes de datos para la toma de decisiones, mediante la analítica predictiva. Asimismo, utiliza un sistema de base de datos, inteligencia artificial y aprendizaje automático.

Las técnicas de minería de datos provienen de una suerte de fusión realizada entre la denominada inteligencia artificial y la estadística, ésta se basa en el uso de algoritmos que devuelven resultados de un conjunto de datos, ( Emilio, H., Ludovico, J., Martín, H., Romina, L., Victoria, C., Mirta, E., Hernán, F. (2016).). En el marco de lo cual, existe muchas técnicas que se pueden utilizar, entre ellas la clasificación, asociación, y el agrupamiento, entre otras.

Por su parte, el algoritmo de árbol de decisión corresponde a una técnica de la minería de datos utilizada para la clasificación y predicción de una variable objetivo, si esta variable es discreta se elabora un árbol de clasificación y el resultado a predecir es una clase discreta, por el contrario, si es continua se obtiene un árbol de regresión y el resultado a predecir es un valor numérico (Oña, R. De, Garrido, C., De, J., Oña, R. De, & Garrido, C. (2017)).

Un ejemplo de ello lo comporta el algoritmo Random Forest (Luis & Ortiz, 2017), es un algoritmo de clasificación y regresión que consiste en una combinación de árboles predictores, de esta

manera, cada árbol depende de los valores de un vector aleatorio de un número de casos. Esta muestra será el conjunto de entrenamiento para construir el árbol. El diseño permite el entrenamiento de los datos usando una mezcla de variables. Cada árbol crece de forma extensa sin ningún tipo de poda.

El algoritmo M5P, según (Sihag, Tiwari, & Ranjan, 2018), es un modelo de árbol de decisión binario que construye segmentos (nodos) basados en la medida de desviación estándar de los valores de clase de la base de datos, que puede predecir valores continuos, minimizando la variación de los elementos agrupados bajo cada nuevo nodo generado. El M5P genera modelos que son compactos, así como fáciles de comprender.

Existen estudios de minería de datos que permiten transformar los datos no relacionados en conocimiento, y así, realizar predicciones más efectivas. Una investigación realizada en el año 2019 sobre minería de datos educativos, para identificar la relación entre cociente intelectual, estilos de aprendizaje, inteligencia emocional e inteligencias múltiples de estudiantes de ingeniería, fue aplicada a 282 estudiantes utilizando técnicas de asociación y agrupamiento, cuyos resultados permitieron establecer la relación entre estas variables Arevalo-Marin, P., Cabrera-Piedra, Y., Cabrera-Sarango, M., Caicedo-Rodríguez, J., Camacho-Veliz, A., Cartuche-Morocho, S., ... others. (2019).

Otra investigación realizada en el año 2017, se trae a colación en este estudio, dado que la misma dio cuenta de la realización de una implementación de minería de datos en el proceso educativo, en la gestión académica de las instituciones de educación superior, cuyo objetivo general radicó en mejorar la calidad en el desempeño educativo de manera inclusiva. (Terán, Saltos, de la Plata, & Terán, 2017).

Poseer datos y desconocer su utilidad, puede significar la diferencia que existe entre las empresas. La UTMACH por su parte, posee una deficiente capacidad de almacenamiento de archivos de las aulas virtuales y cada periodo se elimina información de aulas que reposan en los servidores. Como consecuencia, la plataforma colapsa y restringe la capacidad de almacenamiento de información.

El análisis predictivo es una herramienta que forma parte de un conjunto de técnicas, capaz de detectar patrones, donde la entrada de datos y una situación permiten que se comparen con un modelo, que incluye combinar los datos requeridos con la información de antecedentes, para que

el sistema realice una apreciación. La importancia de un modelo predictivo de almacenamiento de información en las aulas virtuales permitirá optimizar los recursos tecnológicos. Por lo tanto, el objetivo de esta investigación es establecer un modelo que permita predecir la capacidad de almacenamiento en las aulas virtuales de la UTMACH.

El presente artículo, se encuentra organizado de la siguiente manera: se presentan y comentan trabajos relacionados con el estudio, la metodología utilizada, los resultados obtenidos, las conclusiones y finalmente el aporte de la investigación.

## **Materiales y Métodos**

### **Reseña referencial de algunos estudios previos como antecedentes de este trabajo**

Se presenta en primera instancia, el trabajo de Mejía, (2019) en el cual se indica que la técnica de regresión logística de la minería de datos en el proceso de descubrimiento de conocimiento (KDD), de las bases de datos operativas o transaccionales permiten encontrar información relevante a partir de datos “ocultos”. Su finalidad, es mostrar el proceso de obtención de conocimiento a partir de bases de datos para el soporte de toma de decisiones en las empresas.

Por su parte, también se presenta la investigación de Retamar, S., De Battista, A., Ramos, L., Nuñez, J. P., Savoy, F., & De Gracia, L. (2019.) sobre minería de datos, para la detección de factores de influencia en el test Apgar en los nacimientos ocurridos entre los años 2009 y 2017, el cual muestra, que estos autores usaron técnicas de selección de atributos y se aplicaron tres modelos de clasificación de árboles de decisión: J48, REP Tree y RamdonTree. Los autores concluyeron que la técnica de árbol de decisión permite analizar decisiones para que los modelos generados sean comprensibles para los usuarios finales.

(Retamar et al., 2019) presentan un estudio sobre innovación en minería de datos para el tratamiento de imágenes, utilizando el método de agrupación K-media para conjuntos de datos aplicados en la agroindustria. Los autores presentan una modificación del método utilizado, que permite mayor flexibilidad en casos de agricultura, a través de ejemplos numéricos para el agrupamiento o clustering sobre imágenes.

Asimismo, en el año 2018 se presentó una investigación sobre la aplicación de minería de datos para la clasificación de programas universitarios de ingeniería industrial, acreditados en alta

calidad en Colombia, utilizando análisis de componentes principales (PCA) y clúster. La investigación les permitió clasificar tres grandes componentes principales y perfiles de programas acreditados (Fontalvo-Herrera, Delahoz, & Mendoza-Mendoza, 2018).

Aunado a los anteriores, también se toma el trabajo realizado en el año 2018 sobre análisis del desempeño del profesorado universitario en el uso de Moodle, a través de técnicas de minería de datos: propuestas de necesidades formativas, donde se analizó el desempeño de 484 profesores de la Universidad Autónoma de Yucatán para identificar patrones de comportamiento. Los autores concluyen que el modelo obtenido tuvo una efectividad del 90% (Herrera, Domínguez, González, & Chacón, 2018).

Además de los anteriores, se reconoce en estudio realizado en el año 2017 sobre el análisis de la deserción de estudiantes universitarios mediante técnicas de minería de datos, utilizando árboles de decisiones, aplicados a la Universidad Católica del Norte en Antofagasta y Coquimbo, donde se obtuvo como resultado un 78,3% de deserción. Los autores concluyen la curva ROC de 76%, 75% y 83% de acierto para los clasificadores de red bayesiana, árbol de decisión y red neuronal respectivamente. (Miranda & Guzmán, 2017).

Así también, una investigación presentada en el año 2017 sobre minería de datos educativos para la detección de recursos claves en la plataforma Moodle, obtuvo modelos predictivos basados en árboles de decisión. Los autores concluyen que se puede obtener información relevante en la base de datos de Moodle, utilizando algoritmos de clasificación y mejorar el rendimiento académico de cursos sucesivos. (Galindo, E. G., Gómez, A. Z., Rodríguez, M. L., Azofra, A. A., Quesada, A. R., & Ortiz, J. L. O. (2017).

Finalmente, en el año 2016 se presentó una investigación que permite aplicar minería de datos para descubrir rutas de aprendizaje frecuente en Moodle. En este estudio participaron 84 estudiantes universitarios, mediante el uso de la técnica de clúster relacionando las calificaciones de los estudiantes y los Blog del aula virtual Moodle para determinar modelos de comportamiento de los estudiantes (Bogarín Vega, Romero Morales, & Cerezo Menéndez, 2016).

## **Metodología**

En este trabajo de investigación se realiza el siguiente procedimiento:

- 1) Selección de datos, en esta fase definimos las fuentes de datos a utilizar.
- 2) Pre-procesamiento de datos, se corrige errores en los datos, duplicación de registros, registros nulos, eliminación de datos innecesarios y transformación de atributos.
- 3) Minería de datos, se determina el modelo que se va a emplear.
- 4) Análisis de resultados, se analiza la información obtenida para que sean aporte a la toma de decisiones.

## Análisis y discusión de los Resultados

### Selección de datos.

La información se obtuvo desde respaldos de Posgresql, que es utilizada por la plataforma virtual Moodle y que corresponden a los años 2016, 2017 y 2018, con un total de 3804 cursos virtuales. Los datos se extrajeron mediante el software Navicat a un archivo de Valores Separados por Coma (CSV). En la tabla 1 se observa la cantidad de aulas virtuales con su respectiva dependencia.

Tabla 1. Cursos virtuales UTMACH.

Nombre	Cursos 2016	Cursos 2017	Cursos 2018	Total
Centro de Educación Continua	0	1	46	47
Centro de Posgrado	0	0	1	1
Curso de Nivelación	1	0	1	2
Facultad de Ciencias Agropecuarias	20	26	100	146
Facultad de Ciencias Empresariales	373	447	768	1588
Facultad de Ciencias Químicas y de la Salud	202	206	576	984
Facultad de Ciencias Sociales	129	164	460	753
Facultad de Ingeniería Civil	86	97	100	283
Total	811	941	2052	3804

Fuente: Elaboración propia 2020

### Pre-procesamiento de datos

Esta fase, es la que mayor tiempo demanda. Se inicia, con la obtención de datos sobre el uso del aula virtual Moodle en relación con la utilización de su capacidad de almacenamiento. Se determinan las actividades y recursos más utilizados, el tipo de archivos que almacena un estudiante o profesor, y el tamaño de almacenamiento del aula.



Como los datos que se encuentran en los archivos de las bases de datos correspondientes a los años 2016, 2017 y 2018, y las diferentes versiones de la plataforma Moodle que han actualizado la estructura de sus tablas, con la ayuda de las herramientas Navicat y OpenRefine se procede a unir los datos en un archivo CSV.

A continuación, se corrigen valores duplicados, eliminación de datos incorrectos e innecesarios con la finalidad de poseer información válida. Además, empleamos el software OpenRefine para mejorar la estructura de los datos recabados. Los parámetros identificados son: cantidad de estudiantes, tareas, recursos, carpetas, foros, cuestionarios, URL, tamaño de almacenamiento, archivos, tipos de archivos y cantidad de aulas virtuales.

### **Minería de datos**

Con respecto a esta fase, se determina la técnica a utilizar, se ejecutan varios experimentos y seleccionamos el que mejor resultado nos proporcionó. Utilizamos el software Weka 3.8 (librería java), que puede resolver una amplia variedad de tareas de machine learning para la clasificación, regresión y clustering. El uso de los algoritmos de M5P, es un método de aprendizaje mediante árboles de decisión, donde cada hoja tiene asociada una clase que permite calcular el valor estimado de la instancia mediante regresión lineal y K-means como método de agrupamiento o clustering para establecer si los datos tienen características similares o diferentes, buscando patrones. Estos algoritmos otorgan alta eficiencia de predicción de modelos y refinamiento iterativo.

### **Análisis de resultados**

Algunos puntos importantes que influyen en el crecimiento del aula virtual y su almacenamiento, mediante el algoritmo M5P permite determinar el tamaño de almacenamiento de un aula virtual con base en su estructura, cuyo resultado se muestra en la figura 1, que a continuación se interpone:

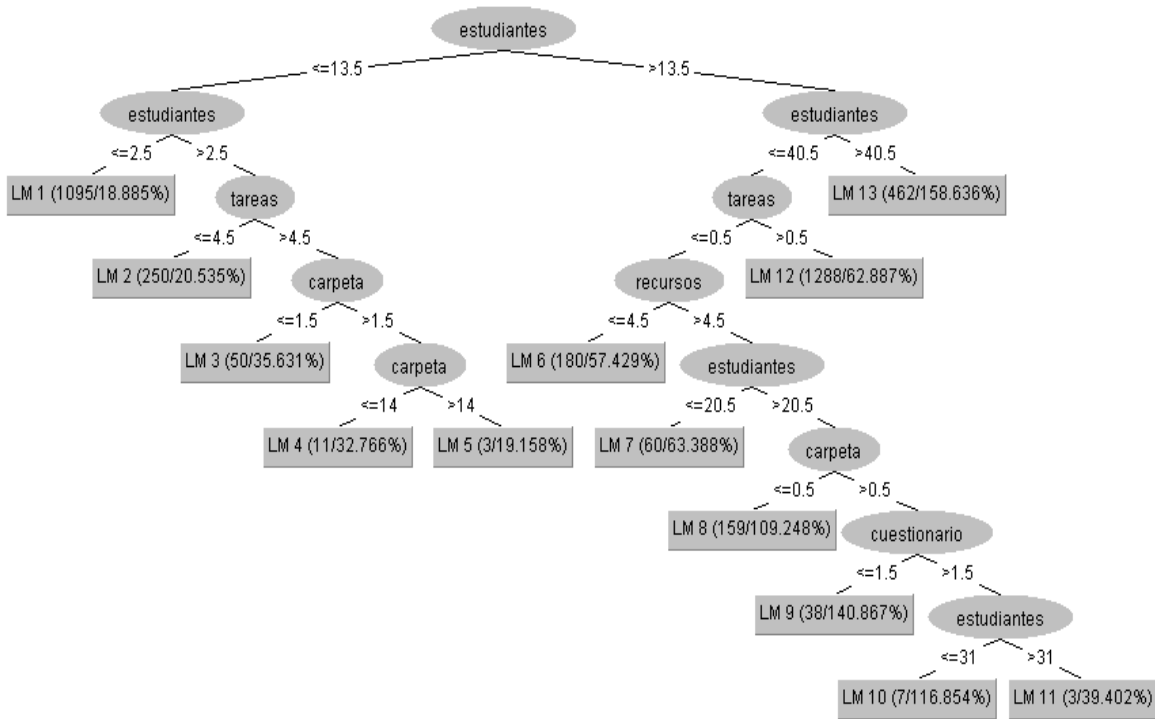


Figura 1. Reglas de clasificación con base en la estructura del aula virtual.

La figura 1, presenta la técnica de clasificación realizada y obtenida con el algoritmo M5P, el cual permite mostrar los atributos que influyen en el tamaño del aula virtual. El resultado proporciona 13 reglas, determinadas por los valores que se muestran en la tabla 2.

Tabla 2. Modelo de valoración del tamaño de un curso virtual.

	<b>Vest</b>	<b>Vtar</b>	<b>vrec</b>	<b>vcar</b>	<b>vfor</b>	<b>vcue</b>	<b>vURL</b>	<b>vval</b>
<b>LM1</b>	0.37788	-0.3707	9.0861	0.8622	-0.6198	-0.0129	0	40.639
<b>LM2</b>	17.0405	0.4734	11.7074	4.6558	-22.0171	- 2.5271	0	27.7535
<b>LM3</b>	3.8395	1.0437	2.4044	19.4341	- 0.6198	- 7.2275	0	296.5278
<b>LM4</b>	3.8395	1.0437	2.4044	41.4617	- 0.6198	- 7.2275	0	264.7167
<b>LM5</b>	3.8395	1.0437	2.4044	46.5248	- 0.6198	- 7.2275	0	368.9685
<b>LM6</b>	0.4919	0.345	2.1482	1.4799	-14.9372	106.1797	4.4343	366.2707
<b>LM7</b>	0.4919	0.345	1.7064	1.4799	-37.7348	64.4172	104.6354	661.2308
<b>LM8</b>	0.4919	0.345	22.3547	1.4799	-141.6438	904.2273	7.3344	598.4152
<b>LM9</b>	0.4919	0.345	-23.5761	1.4799	-190.6302	214.7722	7.3344	1621.7443
<b>LM10</b>	-90.1426	0.345	-76.0614	1.4799	381.6545	314.6853	7.3344	6223.5014
<b>LM11</b>	6223.5014	0.345	-51.8926	1.4799	-381.6545	314.6853	7.3344	6167.9169
<b>LM12</b>	10.1504	11.4454	8.3305	30.8221	-1.4724	-15.9169	0	341.026
<b>LM13</b>	12.4877	15.7795	23.4939	70.9225	-115.2747	0.884	0	330.059

Fuente: Elaboración propia

Estos valores del modelo establecido por el algoritmo permiten predecir el tamaño de un aula virtual, basado en su estructura y se expresa en la siguiente ecuación, a saber:

$$\begin{aligned}
 \text{tamañoMb} = & (n\text{Estudiantes} * \text{vest}) + (n\text{Tareas} * \text{vtar}) + (n\text{Recursos} * \text{vrec}) \\
 & + (n\text{Carpetas} * \text{vcar}) + (n\text{Foro} * \text{vfor}) + (n\text{Cuestionario} * \text{vcue}) + (n\text{Url} * \text{vurl}) \\
 & + \text{vval}
 \end{aligned}$$

Así como también, en las aulas virtuales se determinó el siguiente modelo utilizando el algoritmo M4P, como se muestra en la figura 2.

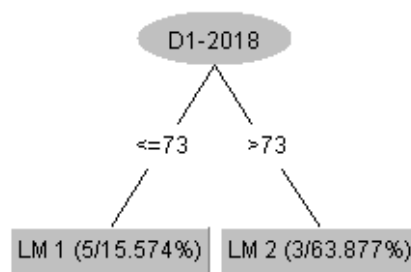


Figura 2. Reglas de clasificación para la cantidad de aulas virtuales.

En la figura 2, se muestran las reglas de clasificación para la determinación de las aulas virtuales por departamentos, lo cual consiste en utilizar dos reglas analizando la cantidad de aulas del último periodo por departamento que permite la obtención de la data, de ello se presenta la valoración de la cantidad de aulas por departamento, como se presenta en la Tabla 3, inserta a renglón seguido:

Tabla 3. Valoración de la cantidad de aulas por departamento.

	<b>vPeriodo</b>	<b>Valor</b>
<b>LM1</b>	1.0405	56.4404
<b>LM2</b>	1.1561	22.7116

Fuente: Elaboración propia

En la tabla 3, se muestra la valoración resultante del modelo, que permite predecir la cantidad de aulas virtuales por periodo dado por los datos históricos.

$$cAulas=(vPeriodo*nAulas)+valor$$

Los atributos establecidos determinan el espacio de almacenamiento que puede ocupar un usuario en la plataforma virtual. En total se obtuvo la siguiente cifra a saber: 28769 registros entre los años 2016, 2017 y 2018 como se muestra en la tabla 4, que a continuación se inserta:

Tabla 4. Atributos de espacio de almacenamiento por usuario.

<b>Atributo</b>	<b>Descripción</b>
Id_usuario	Código único en el aula virtual que se asigna a un usuario.
Espacio_utilizado	Megabytes utilizados en la plataforma.
Cantidad_archivos	Cantidad de archivos generados en la plataforma.
Aulas_asignadas	Número de aulas matriculado en usuario.

Fuente: Elaboración propia

En la tabla 4, se pueden observar los atributos que permiten determinar la capacidad de almacenamiento que utiliza un usuario determinado en la plataforma. Para la obtención de este resultado se realizaron 4 clústeres, a los fines de determinar los datos como se muestra en la tabla 5, que se presenta a renglón seguido:

Tabla 5. Clúster de almacenamiento por usuarios.

	<b>Full Data</b>	<b>Clúster 0</b>	<b>Clúster 1</b>	<b>Clúster 2</b>	<b>Clúster 3</b>
	<b>(28769)</b>	<b>(10468)</b>	<b>(3847.0)</b>	<b>(8288)</b>	<b>(6166)</b>
Espacio_utilizado(MB)	133.6389	12.1526	32.583	95.5416	454.1435
Cantidad_archivos	64.0577	35.7888	48.2636	65.9059	119.4197
Aulas_asignadas	4.2397	1.4514	3.0	4.8352	8.9465

Fuente: Elaboracion propia

En la tabla 5, se muestra el resultado de 4 clústeres que son determinados por el espacio utilizado, así como la cantidad de archivos y las aulas asignadas a un usuario. Se observa que en una media de 133.64 MB es consumido por un usuario, con un total de 4 aulas virtuales asignadas y 64 archivos.

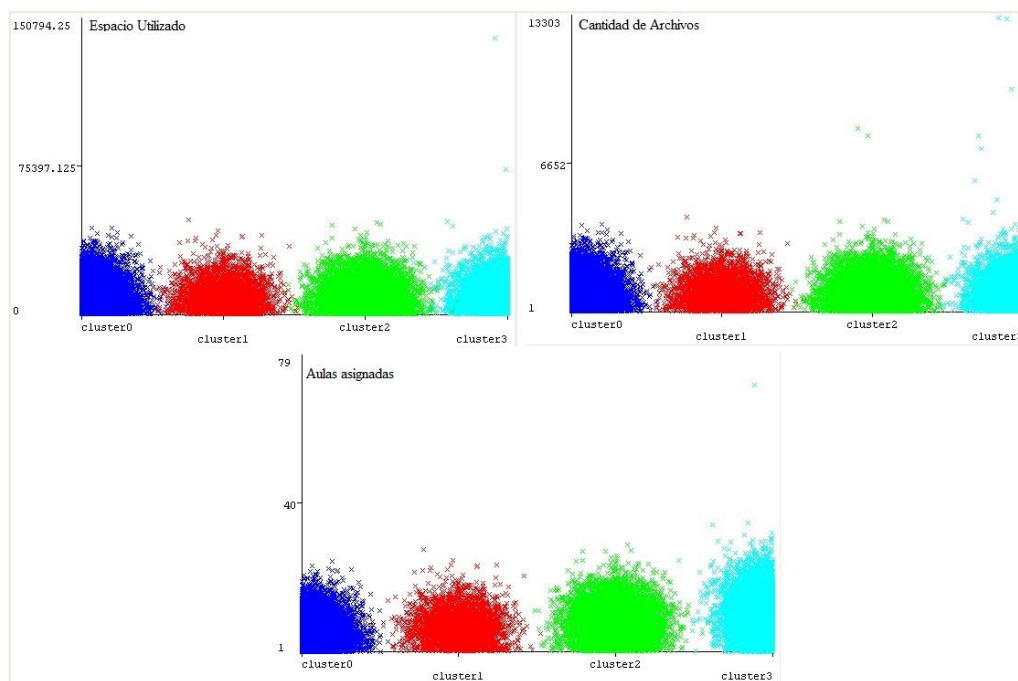


Figura 3. Clúster de datos de almacenamiento por usuario.

Además de lo anteriormente expuesto, se insertan acá los resultados de la realización del análisis de clúster para determinar los grupos de archivos más utilizados dentro de las aulas virtuales de la UTMACH, de ello se obtuvo que existe: un total de 77009 archivos generados durante los años 2016, 2017 y 2018, de lo que se determina lo siguiente:

Tabla 6. Clasificación de archivos en el Moodle UTMACH.

	<b>Full Data</b>	<b>Clúster 0</b>	<b>Clúster 1</b>	<b>Clúster2</b>	<b>Clúster 3</b>
	<b>(77009)</b>	<b>(8604)</b>	<b>(341)</b>	<b>(16273)</b>	<b>(51791)</b>
Tipo	Docx	pptx	moodle.backup	pdf	docx
Megabytes	5.40	4.64	269.06	10.46	2.18

Fuente: Elaboracion propia

Se aplica el algoritmo K-means con la información de los archivos generados en la plataforma virtual de la UTMACH. En la tabla 6 se muestra cuatro clústeres con los atributos: tipo y megabytes (espacio utilizado en la plataforma) y se destacan los documentos docx, pptx, backup y pdf como los principales elementos manejados en el aula virtual, así como también la relación con el espacio de almacenamiento.

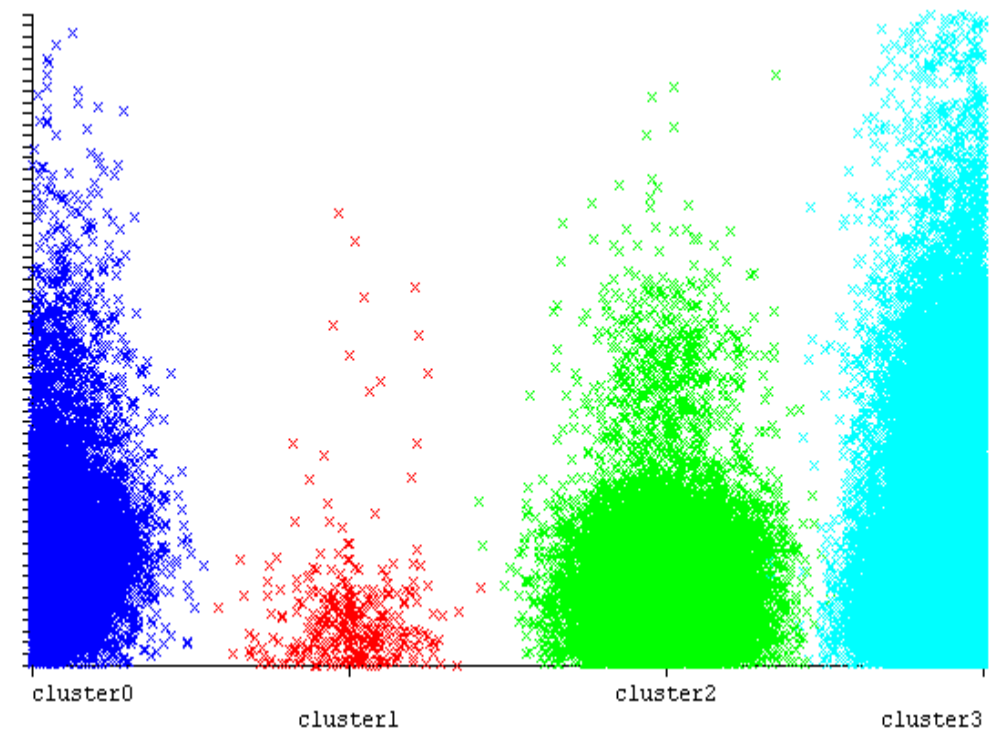


Figura 4. Clustering de los archivos del Moodle UTMACH.

En esta figura 4, como se puede observar, se muestra la relación de los datos entre el tipo de archivo y los clústeres, de ello se obtuvo que los documentos de texto, formato office, tales como Word, también se emplea información presentada en formato PDF, los cuales son los más utilizados en la plataforma Moodle de la Universidad Técnica de Machala.

## Conclusiones

Luego de un proceso de revisión bibliográfica, de vinculación de estudios previos referenciales para este artículo y la aplicación de la técnica de minería de datos, se obtuvo una serie de resultados que permiten arribar a las siguientes conclusiones, a saber:

Los principales archivos que son utilizados en la plataforma virtual son los documentos de texto, presentación, PDF y copias de seguridad de Moodle.

Mediante el uso de un modelo matemático podemos obtener una mejor precisión en la predicción del almacenamiento de las aulas virtuales de la Universidad Técnica de Machala.

La Universidad Técnica de Machala dispone de un insuficiente espacio de almacenamiento para satisfacer todas las necesidades de las aulas virtuales. Aunado a ello, los profesores y estudiantes no utilizan todas las actividades y recursos que proporciona la plataforma.

La capacidad de almacenamiento de la información de los usuarios oscila entre 12 MB y 500 MB. A su vez, se detecta un importante crecimiento de aulas virtuales cada semestre.

El tamaño del aula virtual está determinado principalmente por el número de estudiantes matriculados y las principales actividades de utilización son las tareas, archivos, carpetas, foros, cuestionarios y URL.

Existe un gran consumo de almacenamiento en copias de seguridad de las aulas virtuales influyendo considerablemente en el crecimiento estructural de las mismas.

## Referencias

- 1) Bogarín Vega, A., Romero Morales, C., & Cerezo Menéndez, R. (2016). Aplicando minería de datos para descubrir rutas de aprendizaje frecuentes en Moodle. EDMETIC. <https://doi.org/10.21071/edmetic.v5i1.4017>
- 2) Chen, J., Xu, J., Tang, T., & Chen, R. (2017). WebIntera-classroom : an interaction-aware virtual learning environment for augmenting learning interactions, 4820. <https://doi.org/10.1080/10494820.2016.1188829>
- 3) Emilio, H., Ludovico, J., Martín, H., Romina, L., Victoria, C., Mirta, E., ... Hernán, F. (2016). Resumen Extendido correspondiente al Informe de Avance Título del

Proyecto : Análisis Comparativo de Modelos de Clasificación de Minería de Datos ( Data Mining ). Su aplicación en la predicción de perfiles de alumnos en riesgo de deserción .

- 4) Fontalvo-Herrera, T. J., Delahoz, E. J., & Mendoza-Mendoza, A. A. (2018). Aplicación de Minería de Datos para la Clasificación de Programas Universitarios de Ingeniería Industrial Acreditados en Alta Calidad en Colombia. *Información Tecnológica*. <https://doi.org/10.4067/s0718-07642018000300089>
- 5) Galindo, E. G., Gómez, A. Z., Rodríguez, M. L., Azofra, A. A., Quesada, A. R., & Ortiz, J. L. O. (2017). Minería de datos educativos para la detección de recursos clave. *Revista de Innovación y Buenas Prácticas Docentes*, 3, 18–28.
- 6) Herrera, P. J. C., Domínguez, V. H. M., González, A. Z., & Chacón, P. E. C. (2018). Análisis del desempeño del profesorado universitario en el uso de MOODLE a través de técnicas de minería de datos: propuestas de necesidades formativas. *RED: Revista de Educación a Distancia*, (58), 10.
- 7) Luis, J., & Ortiz, O. (2017). Minería de datos educativos para la detección de recursos clave, (February 2018). <https://doi.org/10.21071/ripadoc.v3i0.9960>
- 8) Martínez-abad, F., & Hernández-ramos, P. (2018). Técnicas de minería de datos con software libre para la detección de factores asociados al rendimiento, 135–145. <https://doi.org/10.21703/rexe.Especial3>
- 9) Mejía, J. C. G. (2019). Aplicación de la técnica regresión logística de la minería de datos en el proceso de descubrimiento de conocimiento (KDD) en bases de datos operativas o transaccionales. *Perspectiv@ S*, 14(13), 51–55.
- 10) Miranda, M. A., & Guzmán, J. (2017). Análisis de la deserción de estudiantes universitarios usando técnicas de minería de datos. *Formacion Universitaria*. <https://doi.org/10.4067/S0718-50062017000300007>
- 11) Ochoa, L. L., Paredes, K. R., Araya, C. B., Santa, U. C. De, Ochoa, L. L., Paredes, K. R., & Araya, C. B. (2017). Survey of academic data using data mining tools *Exploración de Datos Académicos Utilizando Herramientas de Minería de Datos*, (July 2017), 19–21.
- 12) Oña, J. De, Oña, R. De, Garrido, C., De, J., Oña, R. De, & Garrido, C. (2017). Extraction of attribute importance from satisfaction surveys with data mining



- techniques : a comparison between neural networks and Extraction of attribute importance from satisfaction surveys with data mining techniques : a comparison between neural netw, 7867. <https://doi.org/10.1080/19427867.2015.1136917>
- 13) Retamar, S., De Battista, A., Ramos, L., Nuñez, J. P., Savoy, F., & De Gracia, L. (2019). Minería de datos para la detección de factores de influencia en el test Apgar. *Revista Digital Del Departamento de Ingeniería e Investigaciones Tecnológicas*, 4(1).
  - 14) Sihag, P., Tiwari, N. K., & Ranjan, S. (2018). Prediction of cumulative infiltration of sandy soil using random forest approach. *Journal of Applied Water Engineering and Research*, 0(0), 1–25. <https://doi.org/10.1080/23249676.2018.1497557>
  - 15) Terán, H. E. E., Saltos, M. A., de la Plata, C. M., & Terán, C. E. E. (2017). Implementación de minería de datos en la gestión académica de las instituciones de educación superior. *Didasc@ Lia: Didáctica y Educación*, 8(3), 203–212.

## References

- 1) Bogarín Vega, A., Romero Morales, C., & Cerezo Menéndez, R. (2016). Applying data mining to discover frequent learning paths in Moodle. *EDMETIC* <https://doi.org/10.21071/edmetic.v5i1.4017>
- 2) Chen, J., Xu, J., Tang, T., & Chen, R. (2017). WebIntera-classroom: an interaction-aware virtual learning environment for augmenting learning interactions, 4820. <https://doi.org/10.1080/10494820.2016.1188829>
- 3) Emilio, H., Ludovico, J., Martín, H., Romina, L., Victoria, C., Mirta, E.,... Hernán, F. (2016). Extended Summary corresponding to the Progress Report Project Title: Comparative Analysis of Data Mining Classification Models. Its application in the prediction of profiles of students at risk of dropping out.
- 4) Fontalvo-Herrera, T. J., Delahoz, E. J., & Mendoza-Mendoza, A. A. (2018). Data Mining Application for the Classification of University Programs of Industrial Engineering Accredited in High Quality in ColoMBia. *Technological information*. <https://doi.org/10.4067/s0718-07642018000300089>
- 5) Galindo, E. G., Gómez, A. Z., Rodríguez, M. L., Azofra, A. A., Quesada, A. R., & Ortiz, J. L. O. (2017). Educational data mining for the detection of key resources. *Journal of Innovation and Good Teaching Practices*, 3, 18–28.

- 6) Herrera, P. J. C., Domínguez, V. H. M., González, A. Z., & Chacón, P. E. C. (2018). Analysis of the performance of university teachers in the use of MOODLE through data mining techniques: proposals for training needs. *NETWORK: Distance Education Magazine*, (58), 10.
- 7) Luis, J., & Ortiz, O. (2017). Educational data mining for the detection of key resources, (February 2018). <https://doi.org/10.21071/ripadoc.v3i0.9960>
- 8) Martínez-abad, F., & Hernández-ramos, P. (2018). Data mining techniques with free software for the detection of factors associated with performance, 135–145. <https://doi.org/10.21703/rexe.Especial3>
- 9) Mejía, J. C. G. (2019). Application of the logistic regression technique of data mining in the process of knowledge discovery (KDD) in operational or transactional databases. *Perspective @ S*, 14 (13), 51–55.
- 10) Miranda, M. A., & Guzmán, J. (2017). Dropout analysis of university students using data mining techniques. *University education*. <https://doi.org/10.4067/S0718-50062017000300007>
- 11) Ochoa, L. L., Paredes, K. R., Araya, C. B., Santa, U. C. De, Ochoa, L. L., Paredes, K. R., & Araya, C. B. (2017). Survey of academic data using data mining tools Exploration of Academic Data Using Data Mining Tools, (July 2017), 19–21.
- 12) Oña, J. De, Oña, R. De, Garrido, C., De, J., Oña, R. De, & Garrido, C. (2017). Extraction of attribute importance from satisfaction surveys with data mining techniques: a comparison between neural networks and Extraction of attribute importance from satisfaction surveys with data mining techniques: a comparison between neural netw, 7867. <https://doi.org/10.1080/19427867.2015.1136917>
- 13) Retamar, S., De Battista, A., Ramos, L., Nuñez, J. P., Savoy, F., & De Gracia, L. (2019). Data mining for the detection of influence factors in the Apgar test. *Digital Magazine of the Department of Engineering and Technological Research*, 4 (1).
- 14) Sihag, P., Tiwari, N. K., & Ranjan, S. (2018). Prediction of cumulative infiltration of sandy soil using random forest approach. *Journal of Applied Water Engineering and Research*, 0 (0), 1–25. <https://doi.org/10.1080/23249676.2018.1497557>

- 15) Terán, H. E. E., Saltos, M. A., de la Plata, C. M., & Terán, C. E. E. (2017). Data mining implementation in the academic management of higher education institutions. *Didasc @ Lia: Didactics and Education*, 8 (3), 203-212.

## Referências

1. Bogarín Vega, A., Romero Morales, C. e Cerezo Menéndez, R. (2016). Aplicação da mineração de dados para descobrir caminhos de aprendizado frequentes no Moodle. *EDMETIC* <https://doi.org/10.21071/edmetic.v5i1.4017>
2. Chen, J., Xu, J., Tang, T. e Chen, R. (2017). WebIntera-classroom: um ambiente de aprendizado virtual sensível à interação para aumentar as interações de aprendizado, 4820. <https://doi.org/10.1080/10494820.2016.1188829>
3. Emilio, H., Ludovico, J., Martín, H., Romina, L., Victoria, C., Mirta, E., ... Hernán, F. (2016). Resumo estendido correspondente ao Título do Projeto do Relatório de Progresso: Análise Comparativa de Modelos de Classificação de Mineração de Dados. Sua aplicação na previsão de perfis de alunos em risco de abandono escolar.
4. Fontalvo-Herrera, T.J., Delahoz, E.J. & Mendoza-Mendoza, A. A. (2018). Aplicativo de mineração de dados para a classificação de programas universitários de engenharia industrial credenciados em alta qualidade no ColoMBia. *Informação tecnológica* <https://doi.org/10.4067/s0718-07642018000300089>
5. Galindo, E.G., Gómez, A.Z., Rodríguez, M.L., Azofra, A. A., Quesada, A.R. & Ortiz, J. L. O. (2017). Mineração de dados educacionais para a detecção dos principais recursos. *Jornal de Inovação e Boas Práticas de Ensino*, 3, 18–28.
6. Herrera, P.J. C., Domínguez, V.H.M., González, A.Z., & Chacón, P.E.C. (2018). Análise do desempenho de professores universitários no uso do MOODLE por meio de técnicas de mineração de dados: propostas para necessidades de treinamento. *REDE: Revista de Educação a Distância*, (58), 10.
7. Luis, J. & Ortiz, O. (2017). Mineração de dados educacionais para a detecção dos principais recursos (fevereiro de 2018). <https://doi.org/10.21071/ripadoc.v3i0.9960>
8. Martínez-abad, F., e Hernández-ramos, P. (2018). Técnicas de mineração de dados com software livre para a detecção de fatores associados ao desempenho, 135–145. <https://doi.org/10.21703/rexe.Especial>.

9. Mejía, J. C. G. (2019). Aplicação da técnica de regressão logística da mineração de dados no processo de descoberta de conhecimento (KDD) em bancos de dados operacionais ou transacionais. *Perspectiva @ S*, 14 (13), 51–55.
10. Miranda, M. A. e Guzmán, J. (2017). Análise de evasão de estudantes universitários utilizando técnicas de mineração de dados. *Formação universitaria*. <https://doi.org/10.4067/S0718-50062017000300007>
11. Ochoa, L.L., Paredes, K.R., Araya, C.B., Santa, U. C. De, Ochoa, L.L., Paredes, K.R., & Araya, C.B. (2017). Pesquisa de dados acadêmicos usando ferramentas de mineração de dados Exploração de dados acadêmicos usando ferramentas de mineração de dados, (julho de 2017), 19–21.
12. Oña, J. De, Oña, R. De, Garrido, C., De, J., Oña, R. De e Garrido, C. (2017). Extração da importância dos atributos de pesquisas de satisfação com técnicas de mineração de dados: uma comparação entre redes neurais e Extração da importância dos atributos de pesquisas de satisfação com técnicas de mineração de dados: uma comparação entre redes neurais, 7867. <https://doi.org/10.1080/19427867.2015.1136917>
13. Retamar, S., De Battista, A., Ramos, L., Nuñez, J.P., Savoy, F., & De Gracia, L. (2019). Mineração de dados para a detecção de fatores de influência no teste de Apgar. *Revista Digital do Departamento de Engenharia e Pesquisa Tecnológica*, 4 (1).
14. Sihag, P., Tiwari, N. K. & Ranjan, S. (2018). Previsão da infiltração cumulativa de solo arenoso utilizando abordagem florestal aleatória. *Jornal de Engenharia e Pesquisa de Água Aplicada*, 0 (0), 1–25. <https://doi.org/10.1080/23249676.2018.1497557>
15. Terán, H.E.E., Saltos, M.A. de la Plata, C.M. & Terán, C.E. E. (2017). Implementação de data mining na gestão acadêmica de instituições de ensino superior. *Didasc @ Lia: Didática e Educação*, 8 (3), 203-212.

©2019 por los autores. Este artículo es de acceso abierto y distribuido según los términos y condiciones de la licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0) (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).