

# Minería de datos en los Sistemas de Gestión de Aprendizaje en la Educación Universitaria

Data mining in Learning Management Systems in University Education

Byron G. Hidalgo Cajo<sup>1,2</sup>

<sup>1</sup> Universidad Rovira i Virgili, España

<sup>2</sup> Universidad Nacional de Chimborazo, Ecuador

bhidalgo@unach.edu.ec

**RESUMEN.** El presente estudio describe la minería de datos educativos (MDE) aplicados a los sistemas de gestión de aprendizaje (LMS) con el fin de identificar patrones de comportamiento, que conducirán a la elección de recursos y actividades más adecuadas en Moodle. La investigación es de tipo transversal y descriptiva – exploratoria, se emplea tres herramientas informáticas como son: Excel 2016, Statistic Program for Social Sciences (SPSS) y Rapidmining 7.5, se utilizó datos de 125 estudiantes, monitoreados en un curso b-learning. Obteniendo los siguientes resultados, los estudiantes se conectan significativamente en el horario de la mañana, su actividad de interacción mensual es equitativa excepto los meses de inicio y fin del periodo que son menores, las actividades que más interactúa el estudiante es la revisión de tareas, se creó 3 grupos mediante la técnica de clustering y con el empleo del algoritmo de k medias se obtuvo el vector de desempeño promedio, obteniendo una distancia considerable entre estos grupos, concluyendo que existe una correlación entre el nivel de actividad y su rendimiento académico, como también recopila los perfiles de comportamiento y los compara con su nivel de actividad, con el objetivo de crear grupos con características similares garantizando que los recursos y actividades se adapten a las necesidades de los estudiantes.

**ABSTRACT.** This study describes the mining of educational data (MDE) applied to learning management systems (LMS) in order to identify patterns of behavior, which will lead to the choice of resources and more appropriate activities in Moodle. The research is transversal and descriptive - exploratory, using three computer tools such as: Excel 2016, Statistical Program for Social Sciences (SPSS) and Rapidmining 7.5, data from 125 students, monitored in a b-learning course. Obtaining the following results, the students connect significantly in the morning schedule, their monthly interaction activity is equitable except the months of beginning and end of the period that are minor, the activities that most interact the student is the revision of tasks, 3 groups were created using the clustering technique and with the use of the average k algorithm, the average performance vector was obtained, obtaining a considerable distance between these groups, concluding that there is a correlation between the level of activity and its academic performance, as It also collects behavioral profiles and compares them with their level of activity, with the aim of creating groups with similar characteristics ensuring that resources and activities are adapted to the needs of students.

**PALABRAS CLAVE:** B-learning, LMS, Moodle, Aprendizaje, Modelo educativo, Redes de información, Base de datos, Minería de datos educativos, Algoritmo.

**KEYWORDS:** B-learning, LMS, Moodle, Learning, educational model, Information networks, Database, Educational data mining, Algorithm.

## 1. Introducción

Los LMS almacena grandes cantidades de datos que usualmente no son explorados en las instituciones educativas en sus modalidad e-learning y b-learning, estas herramientas tecnológicas que es un complemento de la educación presencial facilitan la comunicación interacción y colaboración dentro del proceso enseñanza aprendizaje (Xhafa et al., 2010).

Moodle ha surgido bajo la necesidad del acompañamiento al estudiante de manera ubicua con el modelo b-learning o aprendizaje mixto que permite implementar las mejores prácticas del aprendizaje presencial con las funcionalidades más óptimas del aprendizaje electrónico, el rol del docente debe incluir el uso de herramientas tecnológicas, metodologías innovadoras y modelos de diseño instruccional que integren las herramientas que ofrecen las TIC.

Según Irasema et al. (2016), la inclusión de las TIC en la educación exige que el profesorado se integre en esta dinámica, cambie los tradicionales esquemas de enseñanza y sea capacitado en el uso y manejo de estas nuevas tecnologías, lo cual tendrá como consecuencia diversas oportunidades para hacer frente a estos cambios.

En la actualidad se hace imprescindible en el proceso de enseñanza aprendizaje a nivel universitario, incorporar las TICs (Núñez et al., 2011) no solo para descargar información, sino que se utilice este recurso a través del análisis crítico y reflexivo, del desarrollo de un aprendizaje autónomo y responsable con la ventaja de la libertad de horarios (Arráez Aybar, Millán Núñez, Carabantes Alarcón & Lozano Fernández, 2008; Martín García, García del Dujo & Muñoz Rodríguez, 2014).

La interacción, comunicación y el aprendizaje colaborativo estimula la autonomía y la motivación intrínseca de los estudiantes (Donnelly, 2010). Es así que los LMS, en forma de plataformas virtuales de enseñanza se han ido asociando de forma creciente al sistema de aprendizaje presencial, convirtiéndose en un elemento más que enlace los diferentes actores educativos.

Cabe señalar que los LMS son herramientas que ayudan en el proceso enseñanza aprendizaje y que sus características del aprendizaje virtual con lleva en su operatividad registros que son extraídos con el objetivo de torturar estos datos hasta que confiesen, por lo que la MDE son frecuentemente utilizados para encontrar este nuevo conocimiento. Los Sistemas de información almacenan todas sus actividades en ficheros o bases de datos que, procesados correctamente, puedan ofrecer información muy relevante para el profesor e investigadores (Bogarín Vega, Romero Morales & Rebeca, 2016).

La extracción y análisis de los datos ha cobrado una especial importancia en estos últimos años en múltiples ámbitos de su aplicación, al poder transformar estos en información y conocimiento de gran utilidad mediante la aplicación de diversas técnicas (Martín Galán & Rodríguez Mateos, 2012). El data mining es una área o campo multidisciplinar en el que confluyen múltiples técnicas y procedimientos tanto de las ciencias de la computación y de la matemática como de las ciencias de la información (Klößgen & Zytkow, 2002).

La minería de datos construye modelos analíticos que descubren patrones y tendencias interesantes en la información de uso del estudiante (Romero, Ventura & García, 2008).

Sin duda alguna la sociedad actual necesita de personas que sean capaces de comprender conceptos complejos, y desarrollar la capacidad para trabajar con ellos de forma creativa y generar nuevas ideas, nuevas teorías, nuevos productos y nuevos conocimientos (MECD, 2015).

Bajo este esquema, el modelo pedagógico de la educación presencial, complementado con la modalidad virtual o b-learning supervisado por el docente y desarrollado con el uso del LMS Moodle ofrece a los estudiantes actividades y recursos en línea, estas interacciones son datos que son obtenidos mediante el uso de

MDE que conduce a la posibilidad de elegir los recursos y actividades más adecuadas, de acuerdo a los perfiles de comportamiento y adaptarlos a las características de los estudiantes en el contexto educativo.

## 2. Materiales y métodos o Metodología computacional

Son muchas las instituciones académicas universitarias de todo el mundo que han venido utilizando la información contenida en los ficheros de actividad de los LMS para medir el grado de calidad de la enseñanza, el presente caso de estudio da a conocer el nivel de interacción y comportamiento más frecuentes de los recursos que los estudiantes utilizan en la actividad académica mediada por Moodle.

La Universidad Nacional de Chimborazo (UNACH), es una institución de educación superior pública, imparte una educación formal presencial y como complemento a este tipo de instrucción utiliza ambientes b-learning. Se cuenta con el Centro de Trasferencia Educativa (CTE), el que brinda múltiples servicios al personal docente y estudiantil, uno de estos servicios es el sistema Informático para el control académico (SICOA), el que incluye: la matrícula de estudiantes, sistema de actividades académicas, portafolio docentes - estudiantes, creación de aulas virtuales entre otros.

Se utiliza la plataforma Moodle para la creación y gestión del aprendizaje virtual, lugar en que cada uno de los docentes organiza sus actividades y recursos de acuerdo a sus objetivos de aprendizaje.

### Método

La presente investigación es de tipo transversal y descriptiva – exploratoria, basado en las observaciones del curso ya que se busca especificar las propiedades importantes de un grupo de estudiantes para su análisis.

La aplicación de la minería de datos educativos requiere de la realización de una serie de actividades previas encaminadas a preparar los datos de entrada, es así que se utilizaron técnicas cuantitativas, haciendo uso de herramientas de recogida de información y análisis del campo de MDE.

### Muestra de datos

El estudio se basó en un universo compuesto de 125 estudiantes con edades comprendidas entre 22-28 años de la Carrera de Medicina, en el periodo académico Octubre 2016 - Marzo 2017, siendo los logs de Moodle la fuente de datos principal y altamente útiles para la generación de estadísticas de uso, que darán lugar a la realización del estudio,

El acceso al LMS se da a través de la identificación de un nombre de usuario y una contraseña, cada vez que un estudiante accede al campus virtual, se genera un número de sesión que identifica al usuario según su perfil y le permite el acceso al contenido del LMS.

A efectos de la investigación, el comportamiento de los estudiantes se describe como un proceso de actividades llevadas a cabo en un sistema LMS con el fin de hallar patrones de comportamiento en la utilización de los recursos del sistema, ya que estos registros son almacenados en base de datos como MySQL. Por lo tanto, el análisis del comportamiento de los estudiantes se llevó a cabo utilizando los principios del proceso de minería de datos (Aalst & Wil, 2011).

Se sigue el siguiente proceso de datos (Figura 1):



Figura 1. Proceso de la Investigación en EDM. Fuente: Elaboración propia.

## Procedimiento

La recolección de los datos se extrajo de los registros de navegación de la plataforma Moodle a manera de informes, accediendo a los registros del servidor, seleccionando el curso, los estudiantes y las actividades realizadas para su respectivo análisis.

Para garantizar el proceso de análisis se realizó un pre procesamiento de los datos, con el objetivo de contribuir significativamente en la creación de patrones de calidad de la información, para obtener datos minables y que puedan ser usados en técnicas de limpieza y transformación, ya que los logs de acceso a los servidores web presentan frecuentemente datos duplicados, acceso realizados por robots, registros basados en el código de estado, que son datos no válidos, identificados en el ambiente b-learning de la plataforma universitaria.

En el estudio se empleó tres herramientas informáticas como son: Excel 2016 utilizada para el filtro de los datos por su excelente visibilidad y facilidad de trabajo, Statistic Program for Social Sciences (SPSS) para Windows versión 23, para el análisis estadístico y Rapidmining 7.5 para la modelización estadística y visualización de los datos.

Con respecto a la instrumentación y análisis se puede exponer que la MDE consiste en la aplicación de modelos estadísticos exploratorios y predictivos orientados a la extracción de patrones de comportamiento significativos de las diferentes tablas que registran las actividades en las plataformas educativas (Chen, Ma, Jong, Hsia & Lin, 2008).

De todos los métodos con los que se cuenta en la MDE, el que se emplea en el estudio es el modelo K-Means para la obtención de conglomerados (Cluster) o grupos de actividad de características similares que tiene como objetivo encontrar patrones o grupos interesantes en un conjunto de datos, y que dichos patrones sirvan para la clasificación, diseño de estrategias, soporte de decisiones, organización de la información entre otras (Azoumana, 2013) (Figura 2).

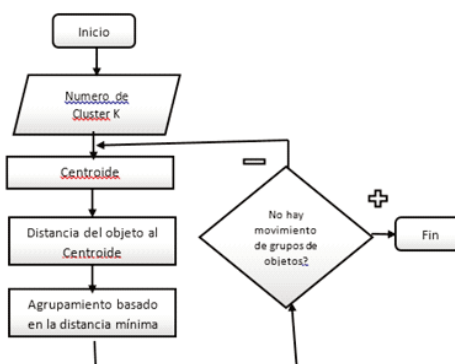


Figura 2. Algoritmo de Clustering K-Means. Fuente: Elaboración propia.

### 3. Resultados y discusión

Para el proceso de obtención de datos en el estudio se identificó la fuente de datos de los servidores donde se encuentra instalado el LMS Moodle para realizar el análisis. La primera fuente es el log de conexión del servidor apache.

El formato de una entrada en el log se divide en (Tabla 1):

ATRIBUTO	DESCRIPCION DEL EVENTO
Fecha/ Hora	Fecha y hora en que el estudiantes realizó el acceso
Identificación del estudiante	Cada estudiante es identificado con un usuario y contraseña para su ingreso mientras que para la plataforma tiene su identificación
Identificación del Curso	Cada curso de estudio tiene su identificación única
Dirección IP	La IP del dispositivo usado para acceder
Actividad	Nombre de la actividad o acción realizada en la plataforma por parte del estudiante
Información	Información sobre la actividad o acción

Tabla 1. Variables del registro de eventos (LOG) de Moodle. Fuente: Elaboración propia.

Se muestra en la tabla 1, las variables que determinaran la interacción que tiene cada estudiante en la plataforma Moodle. Estas variables se calculan a partir del registro de ingreso en la plataforma y la generación de las diferentes tablas de bases de datos (Figura 3).



Figura 3. Selección de datos logs Moodle. Fuente: Elaboración propia.

Los datos obtenidos de Moodle son procesados y convertidos en ficheros Excel con un total de 39850 registros lo que se procede a su filtrado con el objetivo de visualizar la información válida mediante el proceso de: eliminación de registros ya sean vacíos, repetidos, etc. y crear un nuevo fichero con datos validados y útiles para el estudio como son: Fecha, Dirección IP, Usuario, Interacción y Acción con un total de registros validos de 31192. (Figura 4)

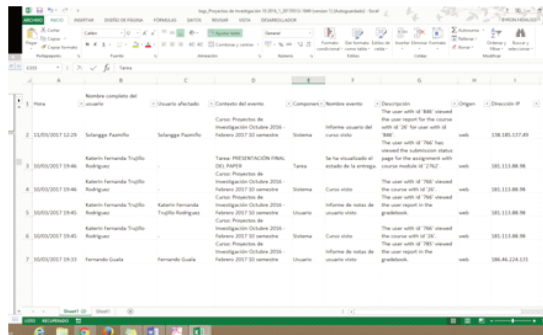


Figura 4. Datos de Moodle en Excel pasa su filtrado. Fuente: Elaboración propia.

Los registros obtenidos luego del filtrado son de 31192, luego de este proceso el resultado es de 28330 registros de estudiantes, 2862 registros del docente, este proceso se analizó por medio del SPSS, por fines de estudio se tomara en cuenta solamente los registros de los estudiantes. (Tabla 2).

		Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje acumulado
Validos	Docente	2862	9,2	9,2	9,2
	Estudiantes	28330	90,8	90,8	100,0
	Total	31192	100,0	100,0	

Tabla 2. Actividad de la plataforma educativa. Fuente: Elaboración propia.

Tal y como se ha comprobado en los análisis descriptivos, el docente acumula un 9,2 % de actividad registrada en la plataforma frente a un 90.8 % de los estudiantes.

Finalmente las variables resultantes para el análisis de los registros validos del LMS Moodle son (Tabla 3):

Nombre	Descripción	Método de extracción
Lugar de trabajo	Se recoge la información de acuerdo a la dirección IP de donde accedieron al LMS	Su extracción se realiza mediante la variable IP
Horario Habitual de trabajo	La hora promedio en que suele conectarse al LMS	Su extracción se realiza mediante la variable fecha extrayendo del mismo la hora que servirá para conocer el horario habitual de trabajo
Conexión	Conexiones por parte de estudiante a la plataforma de manera mensual	Su extracción se realiza mediante la variable fecha extrayendo del mismo el mes que servirá para conocer la conexión mensual del estudiante
Interacción con los Recursos y actividades utilizadas	Actividades y recursos que el estudiante utiliza en la plataforma	Su extracción se realiza mediante la variable Componente para conocer la interacción o utilización de los recursos

Tabla 3. Variables que muestran la interacción de los estudiantes en Moodle. Fuente: Elaboración propia.

#### Lugar Habitual de la Actividad Académica por parte de los estudiantes

Se analiza el lugar más frecuente que los estudiantes realizan sus diferentes actividades académicas, con la localización de la dirección IP de conexión del dispositivo.

Como primer filtro se conoce que las direcciones IP que estén contemplados en los rangos (190.xxx.xxx.xxx) se considera que los estudiantes realizan sus actividades en la UNACH y con direcciones IP diferente son realizadas fuera la institución lo que se considerara como casa. (Tabla 4 y Figura 5).

		Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje acumulado
Validos	Universidad	14272	50,4	50,4	50,4
	Casa	14058	49,6	49,6	100,0
	Total	28330	100,0	100,0	

Tabla 4. Tendencia del Lugar habitual de la actividad académico por parte de los estudiantes. Fuente: Elaboración propia.

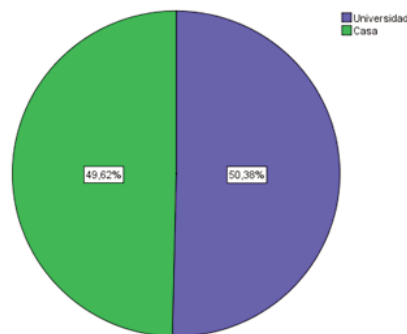


Figura 5. Tendencia del Lugar habitual de trabajo académico por parte de los estudiantes. Fuente: Elaboración propia.

Se describe los resultados del análisis de los datos extraídos, mediante la interpretación basada en un análisis estadístico y cuantitativo, dando a conocer que el 50.4 % de los estudiantes realizan sus actividades en la universidad frente al 49.6 % que lo realizan en la casa, si bien esta información es muy relativa, dado que el docente mantiene clases presenciales 2 horas a la semana utilizando la plataforma en los laboratorios de la institución.

#### Horario de conexión habitual

El horario habitual de las actividades académicas en el LMS Moodle se estableció en la variable fecha de los registros de la base de datos, y mediante SPSS se extrae la hora de conexión de los estudiantes.

Se estableció la siguiente franja horaria (Tabla 5 y 6 y Figura 6):

1. Mañana	06:00 a 11:59
2. Tarde	12:00 a 18:59
3. Noche	19:00 a 24:59
4. Madrugada	01:00 a 05:59

Tabla 5. Franja Horaria de conexión a la plataforma por parte de los estudiantes. Fuente: Elaboración propia.

	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje acumulado
Validos Mañana	12604	45,5	45,5	44,5
Tarde	6418	22,7	22,7	67,1
Noche	8771	31,0	31,0	98,1
Madrugada	534	1,9	1,9	100,0
Total	28330	100,0	100,0	

Tabla 6. Horario de conexión a la plataforma por parte de los estudiantes. Fuente: Elaboración propia.

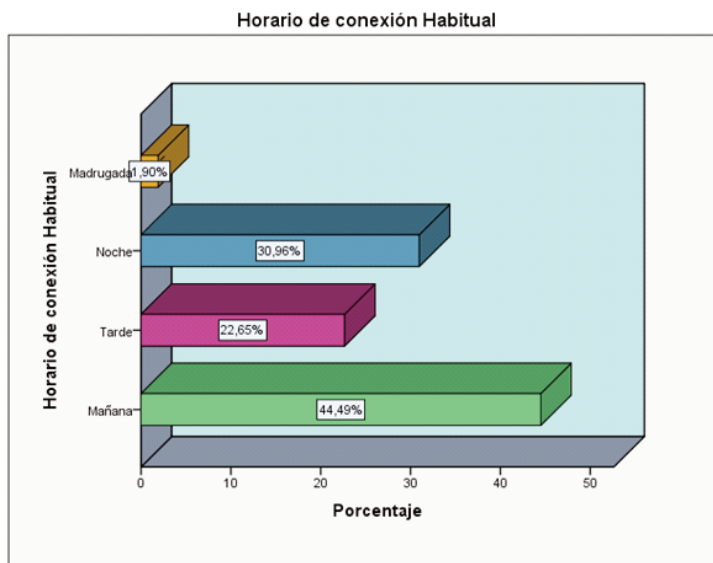


Figura 6. Horario de conexión a la plataforma por parte de los estudiantes. Fuente: Elaboración propia.

Se destaca la tendencia de la interacción estudiantil en el LMS por la mañana debido a que los estudiantes realizan su actividad académica con el docente en los laboratorios de la institución con la interacción del LMS en un 45 %, sin embargo el 31 % realiza sus actividades en la noche y un 22.7 % en la tarde y solamente el 1.9 % lo realiza en la madrugada, notándose que el horario de la noche es el más utilizado por los estudiantes para conectarse con la plataforma educativa.

#### Conexión a la actividad académica por parte del estudiante con periodicidad mensual

La conexión mensual, se establece con la variable fecha de la cual se extrae los datos en una nueva variable denominada mes, en donde se obtiene información de la conexión de los estudiantes a la plataforma educativa mensual (Tabla 7 y Figura 7).

		Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje acumulado
Validos	Oct	2350	8,3	8,3	8,3
	Nov	8209	29,0	29,0	37,3
	Dec	5326	18,8	18,8	56,1
	Jan	5440	19,2	19,2	75,3
	Feb	5173	18,3	18,3	93,6
Mar	1832	6,5	6,5	100,0	
	Total	28330	100,0	100,0	

Tabla 7. Tendencia del Lugar habitual de la actividad académico por parte de los estudiantes. Fuente: Elaboración propia.



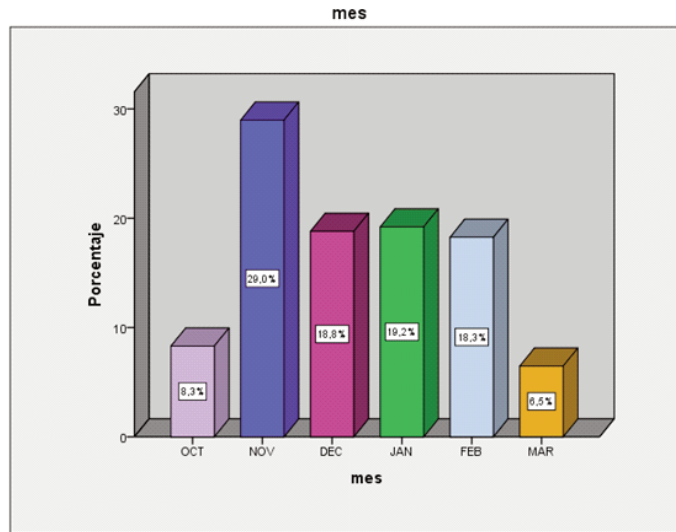


Figura 7. Frecuencia de conexión por mes a la plataforma por parte de los estudiantes. Fuente: Elaboración propia.

Una vez analizado la frecuencia de la variable mes (6 meses- periodo académico) donde se evidencia que el mes más frecuente de conexión es Noviembre 2016 con el 29 %, y notándose que los meses de Diciembre 2016, Enero 2017 y Febrero 2017 son meses que su conexión son muy equitativos esto se nota debido a que el proceso de tutoría y trabajo autónomo son equilibrados, no siendo así la diferencia con porcentajes del 8.3 % y 6.5 % al inicio y final del periodo.

#### Interacción con los Recursos y Actividades

Moodle almacena varios registros que corresponden a una acción (agregar, borrar, actualizar, ver) con algún recurso o actividad del LMS realizadas por los estudiantes (Tabla 8 y Figura 8).

	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje acumulado
Validos Tarea	10649	37,6	37,6	37,6
Sistema	10129	35,8	35,8	73,3
Archivos enviados	2514	8,9	8,9	82,2
URL	2073	7,3	7,3	89,5
Foro	1069	3,8	3,8	93,3
Recurso	1008	3,6	3,6	96,9
Carpeta	588	2,1	2,1	98,9
Usuario	129	,5	,5	99,4
Pagina	83	,3	,3	99,7
Chat	25	,1	,1	99,8
Taller	24	,1	,1	99,9
Libro	15	,1	,1	99,9
Comentarios de la entrega	13	,0	,0	100,0
Informe General	11	,0	,0	100,0
<b>Total</b>	<b>28330</b>	<b>100,0</b>	<b>100,0</b>	

Tabla 8. Frecuencia de la Interacción de los recursos y actividades utilizadas. Fuente: Elaboración propia.

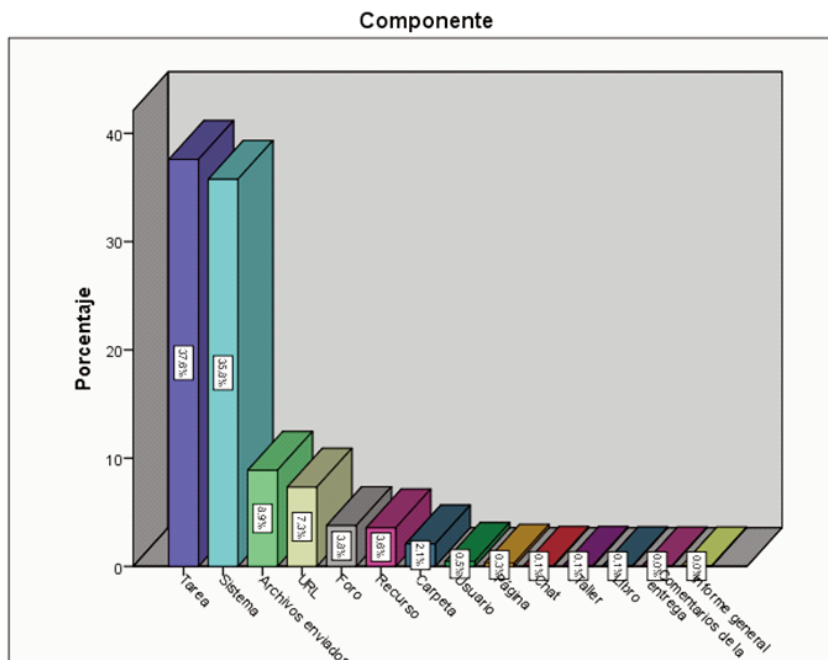


Figura 8. Frecuencia de la Interacción de los recursos y actividades. Fuente: Elaboración propia.

La información obtenida describe que el estudiante en un 35.8 % ingresa a la plataforma y utiliza con frecuencia la actividad Tarea en el 37.6 % debido a que el estudiante se encuentra muy pendiente de la actividad autónoma que debe realizar mediados por la plataforma, seguidamente utiliza el envío de los diferentes archivos resultantes de las tareas propuestas por el docente en un 8.9 %, para finalmente interactuar con los diferentes recursos que el docente ha facilitado en la plataforma como medio de consulta constante y ayuda en el desarrollo de sus actividades como son: los enlaces URL en un 7.3 %, foro 3.8 %, Recurso y Material 3.6%, archivos en carpeta 2.1 % y demás recursos disponibles para el estudiante.

#### Clustering K-means

Los resultados de la agrupación en base al algoritmo de k-Means (Agrawal & Srikant, 1994) se procedió a agrupar a los estudiantes en tres clúster de acuerdo al nivel de participación y recursos utilizados en la plataforma, para descubrir patrones que refleje comportamientos análogos en los estudiantes por medio de la herramienta Rapidminer (Figura 9).

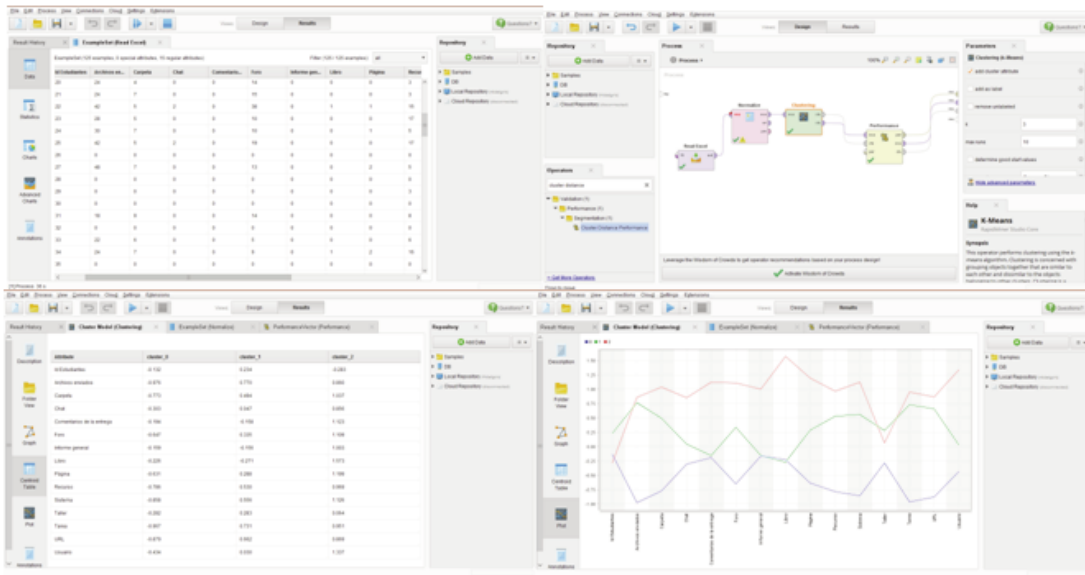


Figura 9. Proceso de Clustering K-means con Rapidminer. Fuente: Elaboración propia.

Como se demuestra en la ilustración el proceso es determinado de acuerdo a la cantidad de conexiones que ha realizado el estudiante y a la interacción con la actividad o recurso, para luego realizar la diagramación del proceso de agrupación (clúster) basadas en distancias compitiendo las variables en los mismos términos y que no sean dominados por las variables más grandes, consiguiendo la media y la desviación estándar para luego utilizar el algoritmo k-means donde se especifica el modelo de clúster, en el caso de estudio tenemos 3 grupos y la hoja de datos con los clúster, obteniendo los centroides de los mismos

#### PerformanceVector:

```
Avg. within centroid distance: -8.906
Avg. within centroid distance_cluster_0: -2.449
Avg. within centroid distance_cluster_1: -8.635
Avg. within centroid distance_cluster_2: -31.003
Davies Bouldin: -1.935
```

Es así que los tres clúster se diferencia notablemente como se visualiza en la ilustración de paralelas coordenadas y que se encuentra agrupadas por el clúster\_0 denominado como bajo conformado con 56 estudiantes, clúster\_1 denominado medio con 52 estudiantes y finalmente clúster 2 denominado alto con 17 estudiantes, sobresaliendo el clúster\_2 que es el de mayor conexión e interacción que ha tenido con los recursos que brinda el LMS, sin embargo posee una baja considerable en la conexión con el recurso taller, en lo referente a los clúster 0 y 1 la interacción que se realizan son muy paralelas en el uso de los recursos mas no en la cantidad de conexiones realizadas.

Se analiza con un vector de desempeño el cálculo de la distancia promedio de cada uno de los puntos de la base de datos a la del centrirodede los clúster que están asignados, en el caso de estudio  $k = 3$ , Cluster\_0 = 2.449, Cluster\_1=8.635, Cluster\_2=31.00.3, y finalmentese obtiene un promedio de distancia de los 3 grupos correspondiente a 8.906.

## 4. Discusión

Se utiliza en el estudio un enfoque puramente basado en MDE para analizar datos que reflejen el comportamiento de los estudiantes y detectar patrones de interacción con las diferentes actividades y recursos

que Moodle ofrece, lo que coincide con (POULOVÁ, 2010) que sugieren que emplear MDE para analizar datos sobre el comportamiento de los estudiantes puede ser beneficioso especialmente para revelar, mapear y monitorear el aprendizaje.

Diversos estudios se han realizado en el área de MDE entre los que podemos destacar (Senechal, 2013), (Marques, 2014), (Conti, 2011), según Senechal (2013) en su estudio Análisis y Pre procesamiento de datos utilizando técnicas de MDE para Moodle analiza la técnica de agrupamiento para generar los perfiles de aprendizaje, con la técnica de redes neuronales y K means, Marques (2014) realizó un estudio con el objetivo de identificar estudiantes con riesgo de reprobación por medio de MDE, con la finalidad de caracterizar los perfiles de los estudiantes con la técnicas de agrupamiento y predicción, según Conti (2011) en su investigación analiza las actividades de Moodle para identificar patrones referente a las tareas con un análisis de plazo y cumplimiento de las mismas con el fin de evitar la reprobación del estudiante con la integración de procesos KDD en Moodle utilizandola técnica de agrupamiento y clasificación de MDE.

Según Pinazo Espigares y García Perez (2011) en su investigación minería de datos educativos en plataformas virtuales evalúa el uso de la plataforma mediante técnicas de MDE empleando el análisis de conglomerados, elaboración de graficas de mallas y reglas predictivas a partir de los datos obtenidos se concluye que los debates en los foros y chats es donde se refleja el aprendizaje. Sosa & Sosa Bruchmann (2014) afirman que los datos extraídos de los LMS generan conocimiento potencialmente útil que posteriormente puede ser utilizado para mejorar el diseño, tanto de materiales, la organización de actividades y evaluaciones entre otros aspectos, para la obtención de este conocimiento es necesario la utilización de técnicas de MDE que permitan el procesamiento de grandes volúmenes de datos.

Lo importante de estos estudios que hablan de los mecanismos para la aplicación de MDE sirven para adaptarlos a la realidad de cada institución educativa que ha decidido realizar un análisis de sus datos siendo necesario tener en cuenta a los diferentes actores dentro de un ambiente educativo que maneja el LMS, estos parámetros ayudarán a la institución a obtener resultado centralizados que consisten en información precisa que ayuden a señalar los problemas que la institución desea corregir ya sea en las plataformas LMS o en lo pedagógico (Villegas-Ch, Luján-Mora & Buenaño-Fernandez, n.d.).

## 5. Conclusiones

Se demuestra la utilidad de la aplicación de MDE en LMS en la plataforma Moodle, permitiendo aplicar la extracción de datos con respecto a la interacción entre el estudiante y la plataforma en términos de actividades y recursos utilizados.

Con la MDE se logró la extracción del conocimiento de la base de datos que permitió analizar automáticamente una gran cantidad de datos y decidir qué información es más relevante para la toma de decisiones, que sellévó a cabo mediante evaluaciones de las diferentes interacciones que se ha considerado como principales, de la misma manera se pudo detectar información no visible de Moodle, lo que ayudo a determinar patrones de conducta, establecer grupos con interés comunes, si bien los estudiantes interaccionan significativamente mucho dependerá del docente en el proceso enseñanza aprendizaje mediados por el LMS, ya que necesita de conocimiento apropiado para la gestión y uso de plataformas educativas aprovechando todas las bondades que brinda Moodle.

El estudio demuestra claramente que existe una correlación entre el nivel de actividad de los estudiantes en la plataforma Moodle y su rendimiento académico, como también recopila los perfiles de comportamiento y los compara con su nivel de actividad con el fin de analizar cómo estos atributos afectan el nivel de actividad de los estudiantes, con el objetivo de crear grupos con características similares garantizando que los recursos y actividades se adapten a las necesidades de los estudiantes.

Se da a conocer que el LMS es utilizado como repositorio de información y en menor grado como

herramientas de evaluación y comunicación efectiva, existiendo debilidades en el ámbito colaborativo, así como la aplicación de metodologías didácticas en el LMS es débil, a pesar de que los docentes utilizan metodologías generalizadas para el diseño, creación de sus aulas virtuales.

La educación ha cambiado su enfoque centrándose en los estudiantes de lo que deben aprender y cómo hacerlo, la minería de datos es la solución adecuada para cubrir estas necesidades, teniendo en cuenta que el análisis de datos permite conocer las fortalezas y debilidades de los estudiantes, este conocimiento permite crear métodos educativos adaptativos establecidos en la mejora del proceso enseñanza aprendizaje.

Cómo citar este artículo / How to cite this paper

Hidalgo Cajo, B. G. (2018). Minería de datos en los Sistemas de Gestión de Aprendizaje en la Educación Universitaria. *Campus Virtuales*, 7(2), 115-128. ([www.revistacampusvirtuales.es](http://www.revistacampusvirtuales.es))

## Referencias

- Aalst, V. der; Wil, M. P. (2011). Process Mining\_ Discovery, Conformance and Enhancement of Business. Processes-Springer-Verlag Berlin Heidelberg. (15-03-2018) ([https://link.springer.com/chapter/10.1007/978-3-642-19345-3\\_4](https://link.springer.com/chapter/10.1007/978-3-642-19345-3_4))
- Agrawal, R.; Srikant, R. (1994). Fast Algorithms for Mining Association Rules. University of Wisconsin, Madison. (12-03-2018) ([https://www.it.uu.se/edu/course/homepage/infoutv/ht08/vldb94\\_rj.pdf](https://www.it.uu.se/edu/course/homepage/infoutv/ht08/vldb94_rj.pdf))
- Arráez Aybar, A. L.; Millán Núñez, C. J.; Carabantes Alarcón, D.; Lozano Fernández, R. (2008). Adquisición de competencias transversales en alumnos de pregrado de Ciencias de la Salud en la Universidad Complutense : una experiencia positiva, 11(3), 169-177. (15-03-2018) (<http://scielo.isciii.es/pdf/edu/v11n3/original3.pdf>)
- Azoumana, K. (2013). Análisis de la deserción estudiantil en la Universidad Simón Bolívar, facultad Ingeniería de Sistemas, con técnicas de minería de datos. *Pensam.am*, 6(10), 41-51. (09-03-2018) (<http://www.coruniamericana.edu.co/publicaciones/ojs/index.php/pensamientoamericano/article/view/137/152>)
- Bogarin Vega, A.; Romero Morales, C.; Rebeca, C. M. (2016). Aplicando minería de datos para descubrir rutas de aprendizaje frecuentes en Moodle Applying data mining to discover common learning routes in Moodle, 73, 5(1), 73-92.
- Chen, C.-M.; Ma, C.-H.; Jong, B.-S.; Hsia, Y.-T.; Lin, T.-W. (2008). Using data mining to discover the correlation between web learning portfolios and achievements. In *Frontiers in Education Conference, 2008. FIE 2008. 38th Annual* (p. F2F-9). IEEE.
- Conti, F. de. (2011). Mineração de dados no Moodle: análise de prazos de entrega de atividades. *TREGA DE ATIVIDADES*. (18-03-2018) (<http://repositorio.ufsm.br/handle/1/5389>)
- Donnelly, R. (2010). Interaction analysis in a "Learning by Doing" problem-based professional development context. *Computers & Education*, 55(3), 1357-1366. DOI: <https://doi.org/10.1016/j.compedu.2010.06.010>
- Irasema, T.; Monroy, S.; Luz, S.; Mendoza, H. (2016). Plataformas LMS, una alternativa TI de éxito en los sistemas educativos de nivel superior, 4(7). (15-03-2018) (<https://repository.uaeh.edu.mx/revistas/index.php/huejutla/article/view/1137/1137>)
- Klößgen, W.; Zytkow, J. M. (2002). *Handbook of data mining and knowledge discovery*. Oxford University Press, Inc.
- Marques, J. L. de Q. (2014). Mineração de dados educacionais: um estudo de caso utilizando o ambiente virtual do SENAI.
- Martín Galán, B.; Rodríguez Mateos, D. (2012). La Evaluación de la formación Universitaria semipresencial y en línea en el contexto del EEES mediante el uso de los informes de actividad de la plataforma Moodle. *RIED*. (16-03-2018) (<http://search.proquest.com/docview/1160567154/fulltextPDF/4FE4249D87CD41EEPC/2?accountid=12268>)
- Martín García, A. V.; García del Dujo, A.; Muñoz Rodríguez, J. M. (2014). Factores determinantes de adopción de Blende Learning en Educación Superior. *Adaptación del Modelo UTAT*. DOI: <https://doi.org/10.5944/educxx1.17.2.11489>
- MECD. (2015). Indicadores de la OCDE. (14-03-2018) (<http://www.mecd.gov.es/dctm/inee/internacional/panorama-de-la-educacion-2015.-informe-espanol.pdf?documentId=0901e72b81ee9fa3>)
- Núñez, J. C.; Cerezo, R.; Bernardo, A.; Rosário, P.; Valle, A.; Fernández, E. (2011). Implementation of training programs in self-regulated learning strategies in Moodle format: Results of a experience in higher education. (15-05-2018) (<http://www.unioviado.net/reunido/index.php/PST/article/view/9034/8898>)
- Pinazo Espigares, J. M.; García Perez, R. (2011). Minería de datos educativos en plataformas virtuales de aprendizaje musical Educational Data Mining with Learning Management Systems educativos más característicos de la Sociedad del Conocimiento . Esto es , el aprendizaje regulado esta sociedad en un pri. (18-05-2018) (<https://ojs.uv.es/index.php/LEEME/article/view/9822>)
- POULOVÁ, P. (2010). Uplatnění elearningu na českých univerzitách–desetiletá historie. In *Sborník příspěvků: 6. mezinárodní konference o distančním vzdělávání-DisCo 2010* (pp. 124-126).
- Romero, C.; Ventura, S.; García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers and*

- Education, 51(1), 368-384. DOI: <https://doi.org/10.1016/j.compedu.2007.05.016>
- Senechal, A. C. L. (2013). Análise e pré-processamento de dados utilizando técnicas de mineração de dados educacionais para o Moodle. Trabalho de Conclusão de Curso.
- Sosa, M. O.; Sosa Bruchmann, E. C. (2014). Estudio de tecnicas de Data Mining aplicadas al analisis de datos generados con la metodologia Blended Learning. 163-167. (13-05-2018)  
([http://sedici.unlp.edu.ar/handle/10915/41977%5Cnhttp://sedici.unlp.edu.ar/bitstream/handle/10915/41977/Documento\\_completo.pdf?sequence=1](http://sedici.unlp.edu.ar/handle/10915/41977%5Cnhttp://sedici.unlp.edu.ar/bitstream/handle/10915/41977/Documento_completo.pdf?sequence=1))
- Villegas-Ch, W.; Luján-Mora, S.; Buenaño-Fernandez, D. (n.d.). Data mining toolkit for extraction of knowledge from LMS. DOI: <https://doi.org/10.1145/3175536.3175553>
- Xhafa, F.; Caballé, S.; Abraham, A.; Daradoumis, T.; Alejandro, A.; Perez, J. (2010). Computational Intelligence for Technology Enhanced Learning Studies in Computational Intelligence. (15-05-2018) ([https://link.springer.com/chapter/10.1007/978-3-642-11224-9\\_9](https://link.springer.com/chapter/10.1007/978-3-642-11224-9_9))

