

# Minería de datos educativos para monitoreo y Mejorando el Rendimiento Académico en la Universidad Niveles

Ezequiel U Okike<sup>1</sup>, Merapelo Mogorosi<sup>2</sup>

Departamento de Ciencias de la Computación  
Universidad de Botswana, Gaborone, Botswana

**Resumen:** este estudio aplicó minería de datos educativos en 712 muestras de registros extraídos del sistema de gestión de aprendizaje Moodle (LMS) en una universidad africana para medir los patrones de uso de los recursos LMS de los estudiantes y el personal y, por lo tanto, determinar si la cantidad de participación medida en la cantidad de tiempo dedicado al uso de los recursos del LMS mejoró

rendimiento académico de los estudiantes. Los datos recopilados de Moodle LMS se preprocesaron y analizaron utilizando algoritmos de aprendizaje automático de agrupación, clasificación y visualización de las herramientas del sistema WEKA. El conjunto de datos constaba de las herramientas del curso (cuestionario, tarea, chat, foro, URL, carpeta y archivos), uso de las herramientas por parte del profesor y del estudiante. Además, se utilizó SPSS para obtener una matriz de coeficientes de correlación para las herramientas del curso, las pruebas y la calificación final. Se realizó un análisis de correlación para verificar si los estudiantes el uso de las herramientas del curso tuvo un impacto en el rendimiento académico de los estudiantes. Los hallazgos indicaron el patrón de uso para el curso 1 como Cuestionario (38358), Sistema (17910), Foro (8663), Archivo (8566), Asignación (1235), Carpeta (514), Envío de archivo (172) y Chat (37); Course2 como sistema (11920), prueba (8208), foro (4476), archivo (4394), tarea (257), chat (247), URL (125) y envío de archivos (38); Course3 como sistema (2622), Archivo (1022), Carpeta (570), Foro (258) y URL (2). En general, evaluar la correlación entre el uso de los recursos del LMS y los

rendimiento, los hallazgos indicaron que existe una relación significativa entre el uso de los recursos LMS y el rendimiento académico de los estudiantes en un nivel de 0.01 significativo. Los hallazgos son útiles para fines de planificación académica estratégica con datos LMS en la universidad.

**Palabras clave:** minería de datos educativos; gestión del aprendizaje sistemas; herramientas del sistema Weka; rendimiento académico mejorado

## I. INTRODUCCIÓN

La minería de datos (DM) y el aprendizaje automático (ML) como subdisciplinas de la informática proporcionan herramientas poderosas para el descubrimiento de conocimientos a partir de conjuntos de datos masivos [1,2]. Como proceso de descubrimiento de patrones en los datos, un proceso de DM debe ser automático o semiautomático. Los patrones descubiertos deben ser significativos en el sentido de que conducen a alguna ventaja, generalmente una ventaja económica [3]. El concepto de minería de datos educativos (EDM) se define en este documento como la aplicación de la minería de datos para derivar patrones significativos de los repositorios del sistema educativo que, a su vez, podrían usarse para mejorar las experiencias de enseñanza y aprendizaje. Uno de esos educativos

La herramienta del sistema de enseñanza y aprendizaje es el sistema de gestión del aprendizaje (LMS). Las herramientas LMS son plataformas computarizadas de enseñanza y aprendizaje para crear contenidos, impartir cursos a

alumnos y la gestión de cursos en entornos de enseñanza y aprendizaje [4,5].

### A. Planteamiento del Problema

Muchas universidades africanas han invertido en sistemas de gestión del aprendizaje. A lo largo de los años, los datos masivos también han acumulado a través del LMS, pero que no ha sido adecuadamente minado para brindar información útil para las decisiones estratégicas en los niveles universitarios. El objetivo de este artículo es demostrar con evidencia empírica la utilidad de EDM para descubrir patrones ocultos pero útiles a partir de datos de enseñanza y aprendizaje acumulados a través del LMS. Varios estudios, por ejemplo [4, 5, 6, 7, 8, 9], también han demostrado el lugar central de los LMS en la enseñanza y el aprendizaje, aunque solo unos pocos abordaron la necesidad de EDM a partir de registros de datos de LMS.

Los objetivos específicos de este estudio fueron:

Los objetivos específicos de este estudio fueron:

- 1) Para medir el uso de los estudiantes y el personal de los recursos LMS en enseñanza y aprendizaje a nivel universitario.
- 2) Evaluar la correlación entre el uso potencial de los recursos del LMS y el desempeño de los estudiantes.

### B. Preguntas de investigación

En el estudio se investigaron las siguientes preguntas de investigación:

- 1) ¿Cuáles fueron los patrones reconocibles en la enseñanza y aprendiendo de los registros de Moodle LMS para el conjunto de datos de muestra?
- 2) ¿Existe alguna correlación entre el uso potencial de LMS y desempeño predominante de los estudiantes en la universidad?

El resto de este artículo consta de tres secciones y una lista de referencias. La segunda sección presenta una revisión de la literatura relevante para el estudio. La sección tres presenta el estudio empírico y la metodología. La cuarta sección presenta los resultados y la discusión del estudio.

## II. REVISIÓN DE LA LITERATURA

La aplicación de los sistemas de gestión del aprendizaje y su eficacia en la educación superior ha sido ampliamente discutida en la literatura [6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17]. Un enfoque común para el modelado de EDM es una combinación de procesos secuenciales que incluye la recopilación de datos de los LMS, el preprocesamiento de datos, la extracción y el análisis de datos, el preprocesamiento, la generación de resultados y la aplicación (consulte la Fig. 1).

En [9] se informó sobre una exitosa minería de datos educativos (EDM) que produjo resultados satisfactorios en la Universidad de Córdoba, España. El enfoque EDM permitió el descubrimiento de nuevas reglas de asociación que se utilizaron para mejorar el diseño de cursos en línea en la universidad.

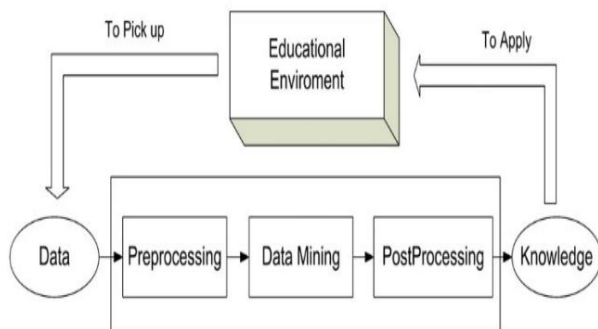
De manera similar, las estrategias para extraer datos de los registros de actividad que se encuentran en Moodle LMS se investigaron en [6]. Aplicando minería de datos y usando estadísticas simples para analizar los registros, el autor recomendó el uso de las aplicaciones Access Watch Analog y Web Start para inferir las actitudes de los estudiantes hacia el aprendizaje y para predecir los puntajes de los exámenes a través de regresiones múltiples. La Fig. 2 muestra un modelo de minería de datos usando LMS.

Siguiendo un enfoque de ciclo de datos, la figura 3 demuestra la utilidad de la minería de datos en el contexto de este documento. Todo proceso de toma de decisiones se basa en un dato transformado en información que culmina con la toma de una decisión. El ciclo comienza con la identificación del problema, la recopilación y el almacenamiento de datos utilizando las herramientas adecuadas, el preprocesamiento de datos, la extracción de datos y el descubrimiento de nuevos conocimientos a partir de datos que proporcionan la retroalimentación necesaria del sistema para actividades futuras.

La Fig. 3 representa el ciclo de datos en el descubrimiento de conocimiento mediante la minería de datos.

Una breve explicación de cada etapa del DCKD incluye:

- 1) *Definición del problema*: Una definición inicial del problema, o la misión, o el propósito, para el cual se requieren los datos.
- 2) *Identificar fuentes de datos*: entender qué son los datos pertinentes y dónde se pueden ubicar.
- 3) *Recopilación y almacenamiento de datos*: Recuperación de datos de diversas fuentes y almacenándolas en un lugar accesible.
- 4) *Minería de datos*: selección de datos relevantes fuera del Big Datos usando las herramientas apropiadas de DM y ML.
- 5) *Descubrimiento de conocimiento*: Nuevo conocimiento descubierto y presentado a los tomadores de decisiones (Clasificado, Agrupado y Visualizado).
- 6) *Aprendizaje y toma de decisiones*: La etapa final que es el propósito del ciclo de datos. Los resultados se muestran a los responsables de la toma de decisiones y se toman las decisiones.
- 7) *Retroalimentación para ciclos posteriores*: Esta etapa no siempre es necesaria. Sin embargo, muy a menudo, la necesidad de tomar una determinada decisión es repetitiva, por lo que el cliente (el que toma la decisión) puede afectar la utilidad y la eficacia del ciclo al enviar comentarios y cambios.



HIGO. 1. Procesos de EDM Fuente: (García et al., 2011).

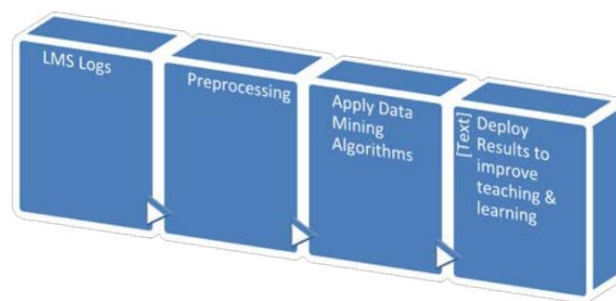


Fig. 2. Minería de datos LMS.

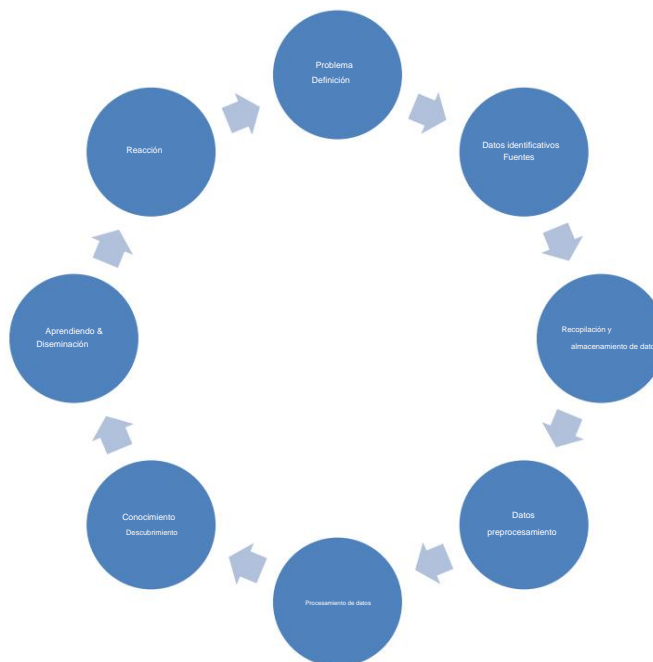


Fig. 3. Ciclo de Datos en el Descubrimiento del Conocimiento (DCKD) usando Minería de Datos.

#### tercero ESTUDIO EMPÍRICO

##### A. Metodología

Una muestra de 712 registros de datos de Moodle acumulados usando En el estudio se utilizó el sistema de gestión de aprendizaje Moodle. La muestra estuvo compuesta por tres cursos seleccionados al azar de tres facultades. Los registros compuestos por estudiantes, personal académico y datos del curso en el formato: nombre completo del usuario, descripción, hora, componentes, usuario afectado, contexto del evento, nombre del evento, origen y dirección IP. En la selección de la muestra se observaron los siguientes criterios:

- 1) Se obtuvo permiso para realizar el estudio de que se obtuvieron los datos.
- 2) La experiencia de un profesor que imparte un curso sobre Se consideró Moodle LMS.
- 3) Se ofreció un curso seleccionado en el primer semestre.
- 4) Un curso seleccionado fue impartido por un profesor que mostró entusiasmo en las actividades en línea.

### B. Instrumento de recopilación de datos

Los registros se descargaron como archivos .csv, se prepararon y preprocesaron utilizando una herramienta Waikato Environment for Knowledge Analysis (WEKA).

### C. Preparación y preprocesamiento de datos

La preparación y el preprocesamiento consistieron en la extracción, limpieza, agregación/integración, filtrado y transformación de datos. Los datos se extrajeron utilizando esquemas de aprendizaje automático seleccionados de la herramienta WEKA. Los esquemas seleccionados incluían herramientas de preprocesamiento, clasificación, agrupamiento y visualización. Los algoritmos se aplicaron directamente a cada conjunto de datos para cada etapa de preprocesamiento, clasificación, agrupación y visualización invocada desde el menú de esquemas de la siguiente manera:

- 1) Se selecciona un archivo de datos del menú de archivos.
- 2) Se seleccionan los atributos importantes de los datos.
- 3) Se crearon agregados de atributos existentes utilizando la hoja de cálculo.
- 4) Se seleccionó un esquema de aprendizaje automático en el menú Esquemas.
- 5) Los resultados se vieron como árboles, texto o gráficos tridimensionales.

- 6) Se revisaron las selecciones de atributos/agregados.
- 7) El esquema se volvió a ejecutar con los datos revisados.

Además, para mantener la independencia del formato, los datos se convirtieron en una representación intermedia:

Formato de archivo de relación de atributos (ARFF). Los registros de datos del semestre 1 del año académico 2017/18 se descargaron, extrajeron y utilizaron para el estudio. El proceso de extracción se muestra en la Fig. 4.

El filtrado de datos se realizó a través de la herramienta de filtrado del sistema WEKA y se centró en los atributos necesarios para su uso en los algoritmos de aprendizaje automático seleccionados ZeroR y J48.

### D. Minería y análisis de datos mediante aprendizaje automático

#### Algoritmos ZERO R y J48

Se aplicaron algoritmos de aprendizaje automático J48 y ZeroR en los conjuntos de datos. La justificación para la elección de los algoritmos J48 y ZeroR en este estudio se debió a su utilidad percibida en la clasificación, el agrupamiento y la visualización [18, 19] El algoritmo J48 usa aprendizaje no supervisado para formar grupos de un conjunto de datos, mientras que el algoritmo ZeroR usa aprendizaje supervisado. Ambos clasificadores diferencian cada caso según unos criterios establecidos.

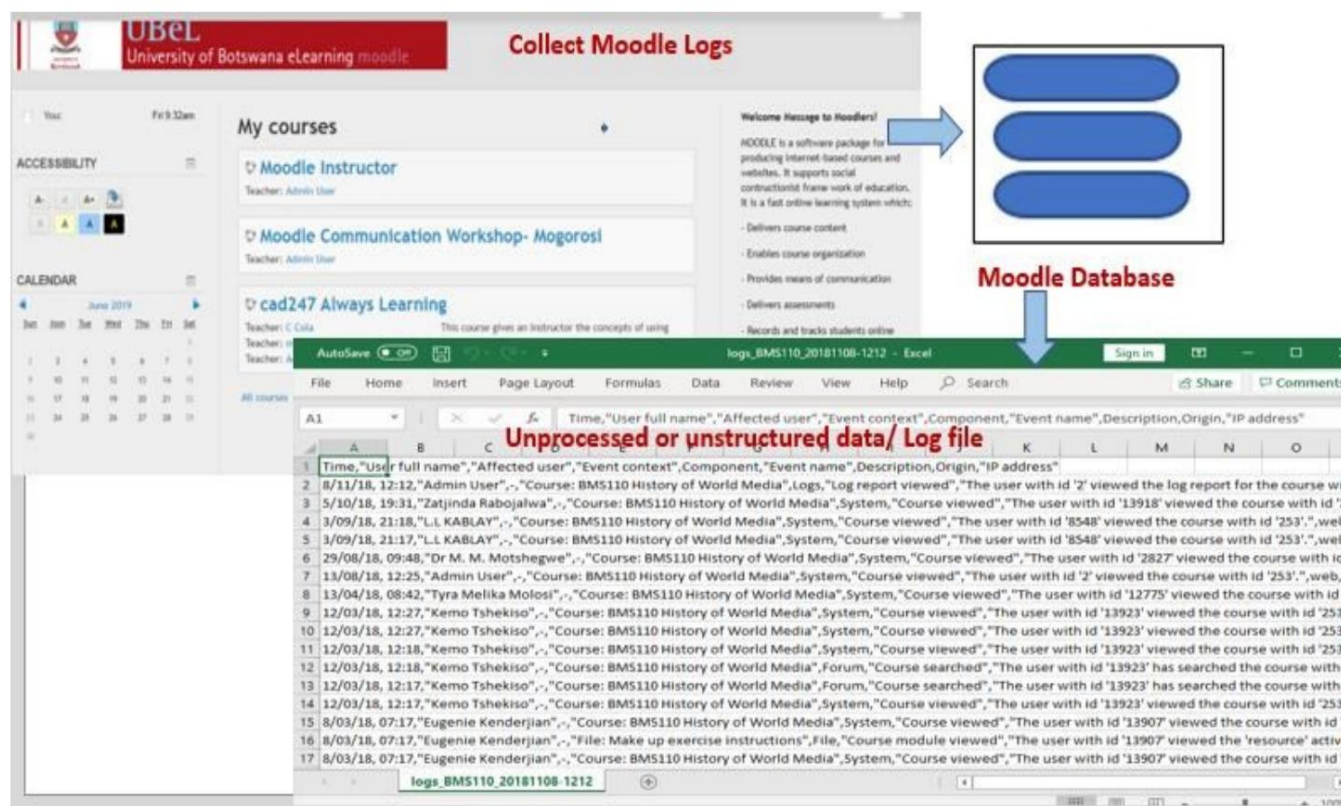


Fig. 4. Extracción de datos de Moodle LMS para minería de datos.

## V. RESULTADOS Y DISCUSIÓN

## A. Clasificación con Algoritmos ZeroR y J48

La Fig. 5 muestra los resultados del uso de ZeroR para predecir Course1. El valor de la clase de predicción fue Cuestionario, que tardó 0,09 segundos en construir el modelo. El número de instancias clasificadas correctamente fue de 38530 (50,553 %) mientras que el número de instancias clasificadas incorrectamente fue de 37687 (49,447 %).

El error absoluto medio y el error cuadrático medio de estas predicciones fueron 0,147 y 0,2711 respectivamente. Había cinco atributos (Nombre completo del usuario, Contexto del evento, Componentes, Nombre del evento y Origen) y el enfoque estaba en el atributo Componente que explica cómo se usaron las actividades y los recursos del curso durante el semestre. La accesibilidad de los registros del curso por parte de un profesor y los estudiantes fue de 76217, que se muestra como el número total de instancias. El número de instancias clasificadas correctamente fue 38530 (50,556%) y las instancias clasificadas incorrectamente fueron 37687 (49,4%). Cabe señalar que cuando un algoritmo se basa en la probabilidad, existe el riesgo de errores de tipo 1 (falsos positivos) y error de tipo 2 (falsos negativos). Esto explica el notable error absoluto medio y el error cuadrático medio anterior. La precisión detallada por clase se muestra a continuación en la matriz de confusión (Tabla I) que muestra las herramientas de actividad con sus valores de accesibilidad.

La Tabla II, la Tabla III y la Fig. 6 muestran los resultados de predicción de actividades en el curso1 utilizando el algoritmo J48.

De la Tabla II el resumen estadístico de los resultados muestra 75463 (99%) como número de instancias clasificadas correctamente y 754 (0,98%) como instancias clasificadas incorrectamente. El error absoluto medio es 0,0028, el error cuadrático medio es 0,039 y el error absoluto relativo es 1,91%. Los errores son mínimos. La precisión detallada por clase se muestra en la matriz de confusión (Tabla III).

Como algoritmo de aprendizaje automático, J48 utiliza la técnica del árbol de decisiones. Cuando se aplica a los registros de LMS en actividades y recursos del curso, el patrón de clasificación de actividades muestra que Cuestionario (columna c) tiene la actividad más alta en 38358, seguido por Sistema (columna b) en 17910, Foro (columna d) en 8663, Archivo (columna f) en 8566, y Asignación (columna e) en 1238, Carpeta (columna g) en 514, Envío de archivo (columna h) en 172 y chat (columna i) en 37.

La Tabla IV y la Tabla V muestran el resultado de la predicción del curso 2 y su matriz de confusión en la clasificación utilizando el algoritmo J48.

De la Tabla V el número de instancias clasificadas correctamente es 29765 (99%), mientras que el número de instancias clasificadas incorrectamente es 276 (0,91%). El error absoluto medio fue 0,0026, el error cuadrático medio fue 0,0387 y el error absoluto relativo fue 1,84%. Los errores son mínimos. La precisión detallada por clase se muestra a continuación en la Tabla VI.

La Tabla V anterior muestra la matriz de confusión de todas las clases. El patrón muestra que la columna b) del sistema tiene la actividad más alta en 11920, seguida de Cuestionario (columna c) en 8308, Foro (columna d) en 4476, Archivo (columna g) en 4398, Asignación (columna f) en 257, Chat (columna e) en 247, URL

(columna h) en 125 y Presentación de expediente (columna i) en 38.

De manera similar, para ZeroR, la predicción del curso 2 en la clasificación se muestra en la Fig. 6, Tabla VI y Tabla VII. Se mantiene la misma explicación ofrecida en J48, excepto que la predicción identificó al sistema como predictor, como se muestra en la Fig. 6 y la Tabla VI.

Los resultados del valor de clase predicho de ZeroR fueron Cuestionario (38530), Sistema (18108), Foro (8802), Archivo (8566), Asignación (1395), Carpeta (514), Envío de archivos (172) y Chat (122). El algoritmo ZeroR predijo que la herramienta Quiz sería la más alta. Esto explica el patrón de cómo el disertante enseñó la clase. El acceso al curso (Sistema: 18108) llevó a los alumnos a usar la herramienta de prueba (38530) con más frecuencia porque tenían que discutir (Foro 8802) el concepto del tema dado y leer notas (Archivo: 8566). Los alumnos también tenían que trabajar en las tareas asignadas (1395) para que el profesor verificara si habían entendido los conceptos del tema y enviar (172) sus tareas para corregirlas. La precisión detallada por clase se muestra en la Tabla VII, mientras que la matriz de confusión se muestra en la Tabla VIII.

El informe sobre la Matriz de Confusión anterior (Tabla VIII) muestra las herramientas de actividad con sus valores de accesibilidad. Los resultados de ZeroR predijeron el valor de clase como Sistema (12109), Cuestionario (8308), Foro (4485), Archivo (4394), Asignación (309), URL (125), Chat (125) y Envío de archivos (38). El algoritmo ZeroR predijo que la herramienta del sistema sería la más alta, seguida de la prueba, el foro, el archivo, la tarea, la URL y el chat. Esto explica el patrón de cómo el disertante enseñó la clase. Los alumnos accedieron al curso (Sistema: 12109) con más frecuencia, lo que los llevó a utilizar la herramienta de prueba. La Tabla XI anterior y la Tabla XII

muestra los resultados del uso de ZeroR para predecir Course3. La figura tiene cinco atributos (Nombre completo del usuario, Contexto del evento, Componentes, Nombre del evento y Origen) y se enfocó en el atributo Componente que explica cómo se usaron las actividades y los recursos del curso durante el semestre. El valor de la clase de predicción fue Sistema, que tardó 0 segundos en construir el modelo. El número de instancias clasificadas correctamente fue de 2622 (58,5007 %) mientras que el número de instancias clasificadas incorrectamente fue de 1860 (41,4993 %). La accesibilidad o los registros del curso por parte del profesor y los estudiantes fue de 4482, que se muestra como el número total de instancias. El algoritmo ZeroR se basa en la probabilidad, de ahí la probable presencia de errores. Los errores de media absoluta y raíz cuadrática media son mínimos en el rango de 0.1676 y 0.2894 respectivamente.

La Tabla XII presenta una precisión detallada por clase y Matriz de confusión en la clasificación para ZeroR

De la Tabla IX, los resultados anteriores tienen 5 atributos y se enfocan en el atributo del componente que explica cómo se usaron las actividades y los recursos del curso durante el semestre. Las actividades del curso fueron Sistema (2622),

Archivo (1022), Carpeta (570), Foro (258) y URL (2). Esto sugiere que el profesor usó el sistema para crear archivos, foros y carpetas para el curso. Las carpetas probablemente se usaron para agregar notas para los estudiantes. Los archivos podrían haberse utilizado para crear materiales didácticos como notas, mientras que el foro podría haberse utilizado para crear temas de debate.

```

Scheme:      weka.classifiers.rules.ZeroR
Relation:    bms110_seml_2015-16-weka.filters.unsupervised.attribute.Remove-R1,3,8
Instances:   76217
Attributes:  5
              User full name
              Event context
              Component
              Event name
              Origin
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

ZeroR predicts class value: Quiz

Time taken to build model: 0.09 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      38530          50.553 %
Incorrectly Classified Instances    37687          49.447 %
Kappa statistic                    0
Mean absolute error                 0.147
Root mean squared error             0.2711
Relative absolute error              100      %
Root relative squared error          100      %
Total Number of Instances          76217

```

Fig. 5. Predicción de Course1 en Clasificación – ZeroR.

TABLA I. MATRIZ DE CONFUSIÓN- CURSO1

A	B	C	D	Y	F	—	H	I	Clasificado como
0	0	8	0	0	0	0	0	0	A = Registros
0	0	18108	0	0	0	0	0	0	B=Sistema
0	0	38530	0	0	0	0	0	0	C=Questionario
0	0	8802	0	0	0	0	0	0	D=Foro
0	0	1395	0	0	0	0	0	0	E = Asignación
0	0	8566	0	0	0	0	0	0	F=Archivo
0	0	514	0	0	0	0	0	0	G=Carpetas
0	0	172	0	0	0	0	0	0	H=Envío de archivos
0	0	122	0	0	0	0	0	0	yo=chatear

TABLA II. PREDICCIÓN DEL CURSO 1 EN CLASIFICACIÓN MEDIANTE EL ALGORITMO J48

Artículo	Valor	
Instancias clasificadas correctamente	75463	99.0107%
Instancias clasificadas incorrectamente	754	0,9893 %
Estadística Kappa	0.9851	
Error absoluto medio	0.0028	
Error cuadrático medio de la raíz	0.039	
Error absoluto relativo	1.9191%	
Raíz Error cuadrático relativo	14.3847%	
Número total de instancias	76217	

CUADRO III. MATRIZ DE CONFUSIÓN CURSO 1 UTILIZANDO EL ALGORITMO J48

A	B	C	D	Y	F		H	I	Clasificado como
8	0	0	0	0	0	0	0	0	A = Registros
0	17910	47	111	23	15	2	0	0	B=Sistema
0	46	38358	100	26	0	0	0	0	C=Cuestionario
0	52	36	8663	35		0	0	0	D=Foro
0	30	42	88	1235	0	0	0	0	E = Asignación
0	0	0	0	0	8566	0	0	0	F=Archivo
0	0	0	0	0	0	514	0	0	G=Carpeta
0	0	0	0	0	0	0	172	0	H=Envío de archivos
0	10	0	2	1	72	0	0	37	yo=chatear

```
=== Run information ===
```

```

Scheme:      weka.classifiers.rules.ZeroR
Relation:    bns201_seml_2014-15-weka.filters.unsupervised.attribute.Remove-R1,3,8
Instances:   30041
Attributes:  5
              User full name
              Event context
              Component
              Event name
              Origin
Test mode:   10-fold cross-validation

```

```
=== Classifier model (full training set) ===
```

```
ZeroR predicts class value: System
```

```
Time taken to build model: 0.06 seconds
```

Fig. 6. Predicción del Curso 2 en Clasificación usando ZeroR.

CUADRO IV. PREDICCIÓN DEL CURSO 2 EN CLASIFICACIÓN MEDIANTE EL ALGORITMO J48

Artículo	Valor	
Instancias clasificadas correctamente	29765	98.0813%
Instancias clasificadas incorrectamente	276	0,9187 %
Estadística Kappa	0.9872	
Error absoluto medio	0.0026	
Error cuadrático medio de la raíz	0.0387	
Error absoluto relativo	1.8404%	
Raíz Error cuadrático relativo	14.46437%	
Número total de instancias	30041	



CUADRO VI. MATRIZ DE CONFUSIÓN CURSO 2 – J48

A	B	C	D	Y	F		H	I	j	Clasificado como
0	2	0	0	0	0	0	0	0	0	A = Registros
0	11920	115	41	6	4	22	1	0	0	B=Sistema
0	0	8308	0	0	0	0	0	0	0	C=Cuestionario
0	1	8	4476	0	0	0	0	0	0	D=Foro
0	0	4	0	247	0	0	0	0	0	E = Charla
0	0	0	0	0	257	0	0	0	0	F=Asignación
0	0	0	0	0	0	4394	0	0	0	G = Archivo
0	0	0	0	0	0	0	125	0	0	H=URL
0	0	0	0	0	0	0	0	38	0	I=Envío de archivos
0	1	0	0	0	0	0	0	0	0	J=Informe de actividad

TABLA VII. PREDICCIÓN DEL CURSO 2 EN CLASIFICACIÓN UTILIZANDO ALGORITMO CERO

Artículo	Valor	
Instancias clasificadas correctamente	12109	40.30807%
Instancias clasificadas incorrectamente	17932	59,6918 %
Estadística Kappa	0	
Error absoluto medio	0.1434	
Error cuadrático medio de la raíz	0.2678	
Error absoluto relativo	100%	
Raíz Error cuadrático relativo	100%	
Número total de instancias	30041	

TABLA VIII. MATRIZ DE CONFUSIÓN- CURSO2

A	B	C	D	Y	F		H	I	j	Clasificado como
0	2	0	0	0	0	0	0	0	0	A = Registros
0	12109	0	0	0	0	0	0	0	0	B=Sistema
0	8308	0	0	0	0	0	0	0	0	C=Cuestionario
0	4485	0	0	0	0	0	0	0	0	D=Foro
0	270	0	0	0	0	0	0	0	0	E = Charla
0	309	0	0	0	0	0	0	0	0	F=Asignación
0	4394	0	0	0	0	0	0	0	0	G = Archivo
0	125	0	0	0	0	0	0	0	0	H=URL
0	38	0	0	0	0	0	0	0	0	I=Envío de archivos
0	1	0	0	0	0	0	0	0	0	J=Informe de actividad

TABLA IX. PREDICCIÓN DEL CURSO 3 EN CLASIFICACIÓN UTILIZANDO ALGORITMO CERO

Artículo	Valor	
Instancias clasificadas correctamente	2622	58.5007%
Instancias clasificadas incorrectamente	1860	41,4993 %
Estadística Kappa	0	
Error absoluto medio	0.1676	
Error cuadrático medio de la raíz	0.2894	
Error absoluto relativo	100%	
Raíz Error cuadrático relativo	100%	
Número total de instancias	4482	

TABLA X. MATRIZ DE CONFUSIÓN- CURSO2

A	B	C	D	Y	F	.....	H	I	j	Clasificado como
0	6	0	0	0	0	0	0	0	0	A = Registros
0	2622	0	0	0	0	0	0	0	0	B=Sistema
0	1022	0	0	0	0	0	0	0	0	C=Archivo
0	258	0	0	0	0	0	0	0	0	D=Foro
0	570	0	0	0	0	0	0	0	0	E = Carpeta
0	2	0	0	0	0	0	0	0	0	F=Informe de actividad
0	2	0	0	0	0	0	0	0	0	G = dirección URL

De manera similar, la Fig. 7 muestra el árbol podado J48 y el contexto del evento es curso3. El resumen de clasificación (Tabla X) muestra instancias correctamente clasificadas como 4482 (100%). El error absoluto medio de la raíz es 0, el error cuadrático medio de la raíz es 0 y el error absoluto relativo también es 0. La Tabla XI muestra la matriz de confusión donde la diagonal principal muestra números que representan la interacción con el sistema y el total de estos números representa el comportamiento de registro de los estudiantes y el (los) profesor(es). Por lo tanto, el algoritmo J48 predice System como el clasificador principal.

La Fig. 7 muestra la predicción de las actividades del curso 3 en la clasificación usando J48 con 5 atributos. La Tabla X muestra el resumen de los resultados reportados como instancias clasificadas correctamente como 4482 (100%) y 0 (0%) como instancias clasificadas incorrectamente. El error absoluto medio fue 0, el error cuadrático medio fue 0 y el error absoluto relativo fue 0%. Los errores son mínimos.

La Tabla XI muestra 7 clases (Registros, Sistema, Archivo, Foro, Carpeta, Informe de actividad y URL) en la matriz de confusión para el Curso3. El patrón muestra que Sistema (columna b) tiene la actividad más alta en 2622, seguido de Archivo (columna c) en 1022, Carpeta (columna e) en 570, Foro (columna d) en 258 y URL (columna g) en 2.

#### B. Patrón de uso Discutir

1) *Pregunta de investigación 1:* Con referencia a la pregunta de investigación 1, el análisis de datos reveló una imagen mixta. Hay un uso sustancial de la herramienta de prueba (alrededor del 70%) con fines de evaluación, como medir el nivel de logro de los objetivos de instrucción. Vale la pena señalar que otra función que también se utilizó sustancialmente fueron las herramientas de Recursos (Archivo, Carpeta y URL), que se utilizaron principalmente para publicar notas y comunicarse entre el profesor y los estudiantes. Se pueden considerar como un punto de entrada para que los profesores digitalicen la enseñanza y el aprendizaje. En total, parece que el patrón

del uso identificado anteriormente se complementan con el aprendizaje mixto, en el que se mezclan los enfoques tradicional y basado en la tecnología según los objetivos de instrucción.

2) *Pregunta de investigación 2:* Con referencia a la pregunta de investigación 2, ya se revelaron patrones en los resultados presentados. Se observó claramente que el estudiante inicia sesión en su portal Moodle para verificar el nuevo contenido del curso (uso del sistema). Habitualmente, los alumnos habrían sido alertados por el profesor de la existencia de nuevos contenidos. Los estudiantes podían descargar el contenido del curso en sus dispositivos personales para leer y discutir entre ellos. Sin embargo, también podrían hacer discusiones limitadas utilizando foros creados por el disertante. Los estudiantes pueden pasar un tiempo mínimo en línea cuando los contenidos del curso no son atractivos. Además, los estudiantes podrían haber exhibido el "síndrome del estudiante", donde había prisa por hacer las tareas justo antes de la fecha límite. Estas actividades eran discernibles de los registros de actividad de los estudiantes y sus fechas de envío.

Por el lado de los disertantes, se sugirió el patrón típico de uso, publicación de notas utilizando la herramienta de recursos y creación de foros. Además, se observó que los profesores hicieron buen uso de cuestionarios, tareas, chats y URL.

En términos de correlación entre el uso potencial de LMS y rendimiento predominante de los estudiantes (pregunta de investigación 2).

La Fig. 8 y la Tabla VII, la Tabla XIII y la Tabla XIV muestran la interacción del estudiante con LMS y el desempeño en la final examen.

La Fig. 8 muestra que Archivos, Cuestionario, Prueba1, Prueba2 y Exámenes finales contribuyeron a la calificación final. La correlación se explica en las Tablas XII, XIII y XIV para el curso 1, curso 2 y curso 3 respectivamente. Las tablas solo reflejan componentes correlacionados con un nivel de significancia de 0,01.



```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    bns305_seml_2014-15-weka.filters.unsupervised.attribute.Remove-R1,3,8
Instances:   4482
Attributes:  5
              User full name
              Event context
              Component
              Event name
              Origin
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

```

Fig. 7. Predicción de Course3 en Clasificación – J48.

TABLA XI. RESUMEN DE CLASIFICACIÓN

Artículo	Valor	
Instancias clasificadas correctamente	4482	100%
Instancias clasificadas incorrectamente	0	0 %
Estadística Kappa	1	
Error absoluto medio	0	
Error cuadrático medio de la raíz	0	
Error absoluto relativo	0 %	
Raíz Error cuadrático relativo	0%	
Número total de instancias	4482	

TABLA XII. MATRIZ DE CONFUSIÓN UTILIZANDO J48- CURSO3

A	B	C	D	Y	F		Clasificado como
6	0	0	0	0	0	0	A = Registros
0	2622	0	0	0	0	0	B=Sistema
0	0	1022	0	0	0	0	C=Archivo
0	0	0	258	0	0	0	D=Foro
0	0	0	0	570	0	0	E = Carpeta
0	0	0	0	0	2	0	F=Informe de actividad
0	0	0	0	0	0	2	G = dirección URL

TABLA XIII. RESUMEN DE CORRELACIONES SIGNIFICATIVAS EN EL CURSO 1

Herramienta	archivos	Carpeta	Examen	Examen final	Prueba1	Prueba2	Nota final
archivos		.390	.402				
Carpeta	.390		.869	.180			.171
Examen	.402	.869		.207	.122	.133	.202
Examen final		.180	.207			.452	.798
Prueba1				.417		.334	.557
Prueba2				.452	.334		.627
Nota final		.171	.202	.798	.557	.627	

La correlación es significativa en el nivel 0.01 (2 colas)

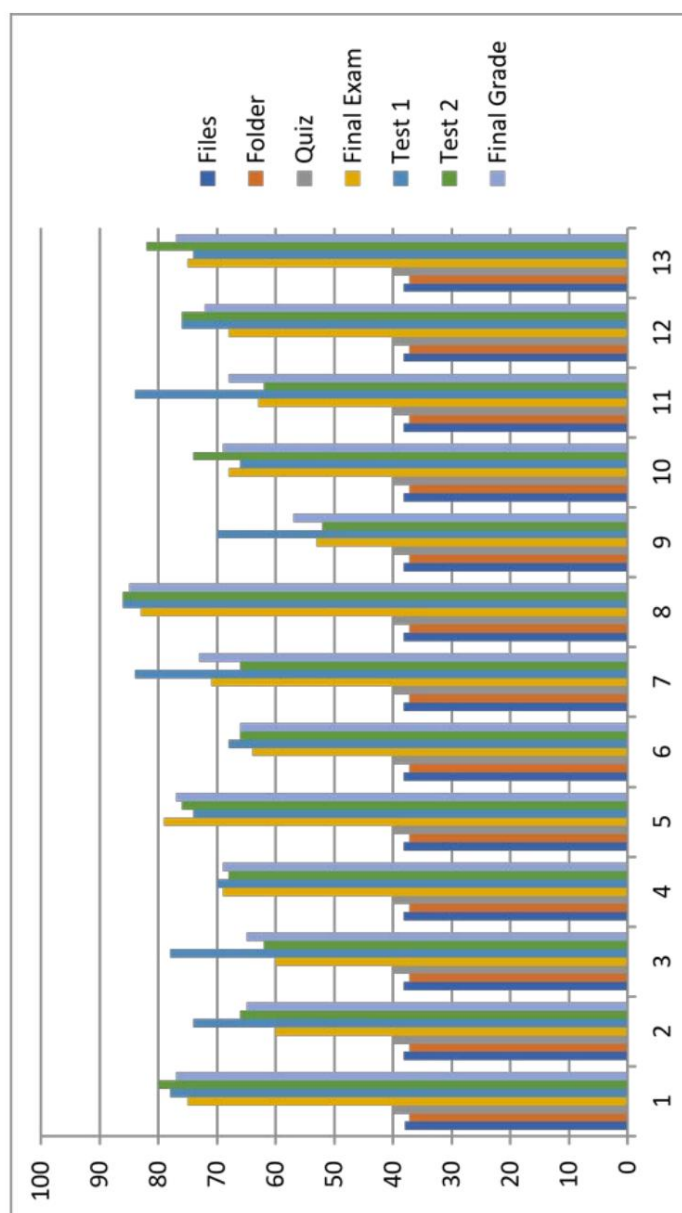


Fig. 8. Análisis de datos basado en la matriz de correlación de las herramientas del curso y el rendimiento académico de los estudiantes.

CUADRO XIV. RESUMEN DE CORRELACIONES SIGNIFICATIVAS EN EL CURSO 2

Herramienta	archivos	Carpeta	Examen	Asignación	Foro	Examen final	Prueba1	Prueba2	Nota final
archivos		0.342	0.607	0.675	0.540				0.232
Carpeta	0.342		0.859	0.816	-0.726	0.288		0.200	0.372
Examen	0.607	0.859		0.901	-0.786	0.254			0.388
Asignación	0.675	0.816	0.901		-0.885	0.302	0.212	0.241	0.438
Foro	-0.540	-0.726	-0.786	-0.885		-0.281	-0.226	-0.224	0.419
Examen final		0.288	0.254	-0.302	-0.281		0.465	0.563	0.761
Prueba1				0.212	-0.226	0.465		0.416	0.582
Prueba2		0.200		0.241	-0.224	0.563	0.416		0.564
Nota final	0.232	0.372	0.388	0.438	-0.419	0.731	0.582	0.564	

La correlación es significativa en el nivel 0.01 (2 colas)

TABLA XV. RESUMEN DE LA CORRELACIÓN SIGNIFICATIVA DEL RENDIMIENTO DE LOS ESTUDIANTES - CURSO 3

Herramienta	archivos	Carpeta	Examen	Foro de tareas		Examen final	Prueba1	Prueba2	Nota final
archivos		0.394	0.406	0.540	0.513				0.144
Carpeta	0.394		0.869	0.858	-0.671	0.184		0.145	0.276
Examen	0.406	0.869		0.847	-0.724	0.209	0.144	0.156	0.286
Asignación	0.540	0.858	0.847		-0.742	0.216	0.148	0.174	0.298
Foro	-0.513	-0.671	-0.727	-0.742		-0.232	-0.211	-0.202	-0.331
Examen final		0.184	0.209	0.216	-0.232		0.414	0.432	0.731

La correlación es significativa en el nivel 0.01 (2 colas)

### C. Análisis de Correlaciones

En el Curso 1 (Tabla XII) la correlación entre las interacciones de los estudiantes con LMS y su desempeño se observó de la siguiente manera:

Existe una relación significativa al nivel 0,01 entre Archivos y Carpeta (0,390), Archivos y Cuestionario (0,402); Cuestionario y Archivos (0,402), Cuestionario y Carpeta (0,869), Cuestionario y Examen Final (0,207), Cuestionario y Calificación Final (0,202); Examen Final y Carpeta (0,180), Examen Final y Cuestionario (0,207), Examen Final y Test1 (0,417); Examen Final y Prueba2 (0,452), Examen Final y Nota Final (0,709); Test1 y Examen Final (0,417), Test1 y Test2 (0,334), Test1 y Nota Final (0,557); Test2 y Examen Final (0,452), Test2 y Test1 (0,334), Test2 y Nota Final (0,627); Nota Final y Carpeta (0,171), Nota Final y Cuestionario (0,202); Nota Final y Examen Final (0,798), Nota Final y Prueba1 (0,557), Nota Final y Prueba2 (0,627); Por lo tanto, los diversos componentes influyeron en la calificación final. Por lo tanto, existe una relación significativa entre el uso de herramientas LMS y el desempeño de los estudiantes. Dado que el valor del Coeficiente de Pearson es superior a 0,01, aceptamos la hipótesis nula (pregunta de investigación 2).

En curso 2 (Tabla XIII). Existe una relación significativa en el nivel 0.01 entre Archivos y Carpeta (0.342), Archivos y Cuestionario (0.607), Archivos y Asignación (0.675), Archivos y Foro (-0.540) y Expedientes y Nota Final (0.232); Carpeta y Archivos (0.342), Carpeta y Cuestionario (0.859), Carpeta y Tarea (0.816), Carpeta y Foro (-0.726), Carpeta y Examen Final (0.288), Carpeta y Prueba2 (0.200) y Carpeta y Calificación Final (0.372); Cuestionario y archivos (0.607), Cuestionario y carpeta (0.859), Cuestionario y tarea (0.901), Cuestionario y foro (-0.786), Cuestionario y examen final (0.254) y Cuestionario y calificación final (0.388); Tareas y Archivos (0.675), Tareas y Carpeta (0.816), Tareas y Prueba (0.901), Tareas y Foro (-0.885), Tareas y Examen Final (0.302), Tareas y Prueba1 (0.212), Tareas y Prueba2 (0.241) y Trabajos y Nota final (0.438); Foro y Archivos (-0.540), Foro y Carpeta (-0.726), Foro y Cuestionario (-

0.786), Foro y Tarea (-0.885), Foro y Examen Final (-0.281), Foro y Prueba1 (-0.226), Foro y Prueba2 (-0.224), Foro y Nota Final (-0.419); Examen final y carpeta (0.288), Examen final y prueba (0.254), Examen final y tarea (0.302), Examen final y foro (-0.281), Examen final y prueba 1 (0.465), Examen final y prueba 2 (0.563), Examen final Examen y Nota Final (0.761); Prueba1 y Tarea (0.212), Prueba1 y Foro (-0.226), Prueba1 y Examen Final (0.465), Prueba1 y Prueba2 (0.416), Prueba1 y Calificación Final (0.582); Test2 y Carpeta (0.200), Test2 y Asignación (0.241), Test2 y

Foro (-0,224), Prueba2 y Examen Final (0,563), Prueba2 y Prueba1 (0,416), Prueba2 y Calificación Final (0,564); Calificación Final y Archivos (0.232), Calificación Final y Carpeta (0.372), Calificación Final y Cuestionario (0.388), Calificación Final y Tarea (0.438), Calificación Final y Foro (-0.419), Calificación Final y Calificación Final (0.761), Nota Final y Prueba1 (0,582), Nota Final y Prueba2 (0,564); Por lo tanto, los diversos componentes influyen en la calificación final. Dado que el valor del Coeficiente de Pearson es superior a 0,01, aceptamos la hipótesis nula y concluimos que existe una relación significativa entre el uso de LMS y el rendimiento predominante de los estudiantes. Los resultados muestran que los archivos, la carpeta, el cuestionario, la tarea, el foro, el examen final, la prueba 1 y la prueba 2 tuvieron un impacto significativo en el rendimiento académico. En este caso el profesor utilizó las herramientas (pregunta de investigación 2).

En el curso 3 (Tabla XIV) Existe relación significativa a nivel 0.01 entre Archivos y Carpeta (0.394), Archivos y Cuestionario (0.406), Archivos y Asignación (0.540), Archivos y Foro (-0.513) y Expedientes y Nota Final (0.144); Carpeta y Archivos (0.394), Carpeta y Cuestionario (0.869), Carpeta y Tarea (0.858), Carpeta y Foro (-0.671), Carpeta y Examen Final (0.184), Carpeta y Prueba2 (0.145) y Carpeta y Calificación Final (0.276); Cuestionario y archivos (0.406), Cuestionario y carpeta (0.869), Cuestionario y tarea (0.847), Cuestionario y foro (-0.727), Cuestionario y examen final (0.209), Cuestionario y prueba1 (0.144), Cuestionario y prueba2 (0.156) y Examen y Calificación Final (0.286); Tareas y Archivos (0.540), Tareas y Carpeta (0.858), Tareas y Prueba (0.847), Tareas y Foro (-0.742), Tareas y Examen Final (0.216), Tareas y Prueba1 (0.148), Tareas y Prueba2 (0.174) y Trabajos y Nota Final (0.298); Foro y Archivos (-0.513), Foro y Carpeta (0.671), Foro y Prueba (-0.727), Foro y Tarea (-0.742), Foro y Examen Final (-0.232), Foro y Prueba1 (-0.211), Foro y Test2 (-0.202), Foro y Nota Final (-0.331); Examen final y carpeta (0.184), Examen final y prueba (0.209), Examen final y tarea (0.216), Examen final y foro (-0.232), Examen final y prueba 1 (0.414), Examen final y prueba 2 (0.432), Examen final Examen y Nota Final (0.731). Por lo tanto, los diversos componentes influyen en la calificación final. Por lo tanto, existe una relación significativa entre el uso de herramientas LMS y el rendimiento de los estudiantes. Dado que el valor del Coeficiente de Pearson es superior a 0,01, aceptamos la hipótesis nula y concluimos que existe una relación significativa entre el uso de LMS y el rendimiento predominante de los estudiantes.

### NOSOTROS. CONCLUSIÓN

En conclusión, utilizando nuestro enfoque de minería de datos educativos y LMS, fue posible monitorear el uso de los recursos LMS por parte de los estudiantes y el personal en la universidad. Además fue

Es obvio a partir del análisis de correlación que el uso de los recursos por parte de los estudiantes podría afectar el rendimiento académico. En el futuro, se sugiere que este estudio se realice con varios cursos de diferentes disciplinas para determinar si las disciplinas académicas afectan el rendimiento de los estudiantes a partir de las mismas herramientas.

## REFERENCIAS

- [1] G. Mariscal, O. Marban y C. Fernandez, "A Survey of Data Mining and Knowledge Discovery Process Models and Methods", *The Knowledge Engineering Review*, vol. 25:2, 2010, págs. 137-166.
- [2] O Marban, G. Mariscal, and J. Segovia, "A data mining and Knowledge Discovery Process Model," *Open Access Data Base www.intechweb.org*, 2009, pp. 1-16.
- [3] HI Witten y E. Frank, *Minería de datos: Tecnología y herramientas prácticas de aprendizaje automático*, 2.ª ed., Nueva York: Elsevier, 2005.
- [4] EU Okike, M. Mosanako y M. Mogorosi, "Federación de datos académicos, construcción de bases de conocimientos y análisis de datos heterogéneos para mejorar la gestión en los niveles universitarios", *Journal of Applied Information Science and Technology*, vol. 11, núm. 2, 2018, págs. 21-30.
- [5] SN Hamade, "Percepciones de los estudiantes sobre los sistemas de gestión del aprendizaje en un entorno universitario: Yahoo Groups Vs BlackBoard, Novena Conferencia Internacional sobre Tecnología de la Información, Nuevas generaciones 2012, pp. 594-599.
- [6] YB Kurata RMLP Bano, MCT Marcelo, "Eficacia de la aplicación del sistema de gestión del aprendizaje en la capacidad de aprendizaje de estudiantes terciarios en un programa de pregrado de ingeniería en Filipinas". En: Andre T. (eds) *Avances en factores humanos en capacitación, educación y ciencias del aprendizaje*. AHFE 2017. *Avances en Sistemas Inteligentes y Computación*, vol 596, 2018. Springer, Cham. [https://doi.org/10.1007/978-3-319-60018-5\\_15](https://doi.org/10.1007/978-3-319-60018-5_15).
- [7] Y. Ghlay, "Eficacia de los sistemas de gestión del aprendizaje en la educación superior: puntos de vista de profesores con diferentes niveles de actividad en LMS", *"Diario de educación superior en línea"*, vol. 3, núm. 2, 2019 págs. 29-50
- [8] C. Romero, S. Ventura y E. García, "Minería de datos en sistemas de gestión de cursos: estudio de caso y tutorial de Moodle", *"Computadoras y educación"*, vol. 51, 2008, págs. 368-384.
- [9] E. García, C. Romero, S. Ventura y C. de Castro, "Una herramienta de minería de reglas de la asociación educativa colaborativa", *Internet y la educación superior*, vol. 14, núm. 2, 2011, págs. 77-88.
- [10] AK Alhazmi y AA Rahman, "¿Por qué LMS no pudo apoyar el aprendizaje de los estudiantes en las instituciones de educación superior?" *Simposio IEEE de 2012 sobre aprendizaje electrónico, administración electrónica y servicios electrónicos*, Kuala Lumpur, 2012, pp. 1-5, doi : 10.1109/IS3e.2012.6414943.
- [11] R. Babo y A. Azevedo, *Instituciones de educación superior y sistemas de gestión del aprendizaje: adopción y estandarización*, IGI Global, 2012.
- [12] JG Boticario y OC Santos, "Problemas en el desarrollo de sistemas de gestión de aprendizaje adaptativo para instituciones de educación superior", disponible en <https://core.ac.uk/reader/55533720>. Consultado el 2 de octubre de 2020.
- [13] N. Darko-Adjei, "Percepciones de los estudiantes y uso del sistema de gestión del aprendizaje Sakai en la Universidad de Ghana. Disponible <http://ugspace.ug.edu.gh/handle/123456789/26847>.
- [14] H. Coates, R. James y GA Baldwin, "Examen crítico de los efectos de los sistemas de gestión del aprendizaje en la enseñanza y el aprendizaje universitarios", *Tert Educ Manag* 11, 19–36 (2005). <https://doi.org/10.1007/s11233-004-3567-9>
- [15] LV Ngeze, "Sistemas de gestión del aprendizaje en instituciones de educación superior en Tanzania: análisis de las actitudes y desafíos de los estudiantes hacia el uso de UDOM LMS en la enseñanza y el aprendizaje en la Universidad de Dodoma", 2016. Disponible en <https://core.ac.uk/reader/55533720>. Consultado el 2 de octubre de 2020.
- [16] MF Paulsen, "Experiencias con sistemas de gestión del aprendizaje en 113 instituciones europeas", *Journal of Educational Technology & Society*, vol. 6. No. 4, 2003, págs. 134-148
- [17] N. Fathima, DM Shannon y M Ross, "Expansión del modelo de aceptación de tecnología (TAM) para examinar el uso por parte de los profesores de los sistemas de gestión del aprendizaje (LMS) en instituciones de educación superior", *"Diario de enseñanza y aprendizaje en línea"*, vol. 11, núm. 2, junio de 2015, págs. 210.
- [18] V. Mhetre, "Algoritmos de minería de datos basados en la clasificación para predecir los estudiantes lentos promedio y rápidos en los sistemas educativos usando Weka", *ICCM*, 2017, pp. 475-479.
- [19] J. Talukdar, S. K Kalita, "Detección de cáncer de mama mediante la herramienta de minería de datos (Weka)", *"Revista internacional de investigación científica y de ingeniería"*, vol. 6, núm. 11, 2015, págs. 1124-1128.