



Técnicas de minería de datos aplicadas a la plataforma educativa Moodle

Data mining techniques applied to the Moodle educational platform

Dr. César Higinio Menacho Chiok¹

¹ Universidad Nacional Agraria La Molina, Lima, Perú. Email: cmenacho@lamoliana.edu.pe

Recepción: 29/03/2020; Aceptación: 15/06/2020

Resumen

La plataforma educativa Moodle, está siendo utilizada por muchas instituciones de educación superior para apoyar sus procesos de enseñanza-aprendizaje en forma virtual. Esta herramienta ofrece a los docentes actividades y recursos pedagógicos para diseñar ambientes educativos más interactivos y eficientes a sus estudiantes. La gran cantidad de datos generados cuando los estudiantes interactúan con Moodle, pueden ser analizados aplicando las técnicas de minería de datos (TMD); con la finalidad de proporcionar a los docentes información y conocimiento, relevante y oportuno, para apoyar y mejorar el aprendizaje de los estudiantes con Moodle. El objetivo de este estudio es presentar una metodología para aplicar las TMD a la plataforma Moodle. Los resultados indican que con notas bajas en las evaluaciones y tareas; y un tiempo de acceso bajo a Moodle, es más probable que el estudiante desaprobe el curso; además, se identificó el grupo de estudiantes con bajo rendimiento que necesitaría retroalimentación.

Palabras clave: Moodle, minería de datos educativa, árbol de clasificación, k-medias, Weka.

Abstract

The Moodle educational platform is being used by many higher education institutions to support their teaching-learning processes in a virtual way. This tool offers teachers activities and pedagogical resources to design more interactive and efficient educational environments for their students. The large amount of data generated when students interact with Moodle can be analyzed by applying data mining techniques (TMD); in order to provide teachers with, relevant and timely, information and knowledge to support and improve student learning with Moodle. The objective of this study is to present a methodology to apply TMDs to the Moodle platform. The results indicate that with low grades in the tests and assignments; and a low access time to Moodle, the student is more likely to disapprove of the course; also, the group of low-achieving students who needed feedback was identified.

Keywords. Moodle, Educational data mining, classification tree, K-Means, Weka.

Forma de citar el artículo: Menacho, C. 2020. Técnicas de minería de datos aplicadas a la plataforma educativa Moodle. Revista Tierra Nuestra 14(1):137-146(2020).

DOI: <http://dx.doi.org/10.21704/rtn.v14i1.1509>

Autor de correspondencia (*): Menacho, C. Email: cmenacho@lamoliana.edu.pe

© Los autores. Publicado por la Universidad Nacional Agraria La Molina.

El artículo es de acceso abierto y está bajo la licencia CCBY

1. Introducción

Cada vez más instituciones de educación superior, están integrando en sus sistemas educativos tradicionales los llamados Sistemas de Gestión de Aprendizaje (SGA), con la finalidad de apoyar y mejorar sus procesos de enseñanza y aprendizaje. La integración de estos SGA en los ambientes educativos, conforma nuevos paradigmas educativos que proporcionan a los estudiantes una variedad de actividades y recursos pedagógicos para permitir un aprendizaje más efectivo y eficiente; y a los profesores con una serie de herramientas que facilitan la gestión de los materiales, el seguimiento y la evaluación del aprendizaje de sus estudiantes (Flate, 2003). Sin embargo, a diferencia de la educación tradicional, los ambientes educativos apoyados con un SGA, carecen de la relación directa entre el profesor y el estudiante, por lo que es necesario que los docentes tengan información sobre el comportamiento, la interacción, el seguimiento y la evaluación del proceso de enseñanza-aprendizaje de los estudiantes; con la finalidad de tomar decisiones más efectivas de la enseñanza (Hijon, R. & Velázquez, A., 2006). Hoy en día, uno de los SGA de mayor uso es la plataforma educativa Moodle (Modular Object Oriented Developmental Learning Environment), la cual es de licencia libre y posee gran potencia y flexibilidad para crear y manejar cursos en línea basados en la web. Moodle, proporciona una serie de actividades y recursos pedagógicos que pueden ser implementados en los cursos, tales como: archivos, tareas, lecciones, cuestionarios, foros, chat, glosarios, wiki, talleres, etc. Como resultado de la interacción de los estudiantes con Moodle, se generan grandes volúmenes de datos que se encuentran en diversos tipos de formatos (texto, numéricos, imágenes, videos, voz, etc.), los cuales son almacenados en las bases de datos de la plataforma, constituyendo verdaderos repositorios de datos heterogéneos sobre el entorno educativo.

La Minería de Datos (MD) comprende un conjunto de métodos, técnicas y herramientas provenientes de los campos de la estadística y el aprendizaje automatizado, con la finalidad de extraer conocimiento de las bases de datos; descubriendo patrones relevantes y útiles. En las últimas décadas se está aplicando la MD a los datos originados en los ambientes educativos, generando una nueva comunidad emergente de investigación conocida como la Minería de Datos Educativa (MDE). La Sociedad Internacional para Minería de Datos Educativa, define la MDE de la siguiente manera: “La MDE es una disciplina emergente, concerniente con el

desarrollo de métodos para explorar los tipos de datos que provienen de entornos educativos, y el uso de esos métodos para comprender mejor a los estudiantes y los ajustes del entorno en los cuales aprenden”. Según Romero (2010), la MDE es el uso de la MD para analizar los datos educativos, con la finalidad de descubrir patrones y perfiles sobre el comportamiento del proceso cognitivo de los estudiantes (aprendizaje) y asistir a los profesores con información oportuna para mejorar los cursos impartidos (enseñanza). En (Barnes, 2009), se considera a la MDE como un área emergente de investigación interdisciplinaria que se ocupa del desarrollo de métodos y técnicas de minería de datos, con la finalidad de explorar los datos originados en ambientes educativos. La MDE, usa métodos para explorar los datos de los entornos educativos con la finalidad de comprender mejor el desempeño de los estudiantes y las condiciones en las cuales ellos aprenden (Ramaswami, 2009). Así mismo, las instituciones de educación superior tienen el compromiso de brindar educación de calidad a sus estudiantes. Es así, que deben enfrentar con una serie de problemas; tales como: el rendimiento académico, el abandono o diseción, la retroalimentación, la recomendación de materiales, la personalización, la adaptación, la falta de motivación, anomalías, etc. Una forma de lograr un mejor nivel de calidad en la educación superior es mediante la aplicación de las TMD para el descubrimiento de conocimiento en sus procesos educativos (Kumar, B. & Pal, Saurabh, 2011).

Las instituciones de educación superior, tienen el reto de explotar los grandes volúmenes de datos educativos generados en su plataforma Moodle, para proporcionar información relevante y útil a sus actores (instituciones, facultades, profesores y estudiantes), con la finalidad de apoyar su toma de decisiones más eficientes para enfrentar los problemas educativos, y así lograr, una mejor calidad en sus procesos de enseñanza y aprendizaje. Este trabajo de investigación tiene como objetivo presentar un marco metodológico para la aplicación de las técnicas de minería de datos a la plataforma educativa Moodle. La metodología se desarrolla bajo el enfoque conocido como: Descubrimiento del conocimiento en bases de datos (Knowledge Discovery in Databases: KDD). Se presenta y se describe cada una de las etapas, adaptándolas al ámbito educativo. Se presenta la aplicación de un caso de estudio, sobre un curso de Estadística que emplea diferentes actividades y recursos de la plataforma educativa Moodle. Se aplican las TMD para la clasificación (árboles de decisión C4.5), la asociación (reglas de asociación a

Priori) y el agrupamiento (k-medias) a partir de los datos recopilados de cuestionarios y tareas diseñadas para el curso en Moodle.

2. Antecedentes

La Minería de Datos Educativa (MDE), es la aplicación de la TMD en el entorno educativo. La Sociedad Internacional para Minería de Datos Educativa, define la MDE de la siguiente manera: “La MDE es una disciplina emergente, concerniente con el desarrollo de métodos para explorar los tipos de datos que provienen de entornos educativos, y el uso de esos métodos para comprender mejor a los estudiantes, y los ajustes del entorno en los cuales aprenden”. En (Barnes, 2009), se menciona que la

MDE, es un área emergente de investigación interdisciplinaria que se ocupa del desarrollo de métodos y técnicas de minería de datos, con la finalidad de explorar los datos originados en ambientes educativos. La MDE, es el uso de métodos computacionales para analizar los datos educativos, con la finalidad de descubrir patrones y perfiles sobre el comportamiento del proceso cognitivo de los estudiantes (aprendizaje) y asistir a los profesores con información oportuna para mejorar los cursos impartidos (enseñanza) (Romero, 2010). En (Baker, R. & Yacef, ., 2009), se presenta una taxonomía de los trabajos de las TMD aplicados desde el punto de vista de las áreas educativas: la mejora del modelo del estudiante, el descubrimiento del conocimiento sobre la estructura de un dominio, el estudio para un apoyo pedagógico más eficaz y factores que afectan el aprendizaje de los estudiantes.

Según Romero, C. (2006), la aplicación de las TMD a los sistemas basados en la web (e-learning), ayudan a los profesores a mejorar el proceso de enseñanza y aprendizaje; y el funcionamiento y rendimiento, usando la información generada por la interacción de los estudiantes con e-learning. En (Castro, Vellido, Nebot, & Mugica, 2007), se presenta una taxonomía de las aplicaciones de las TMD para resolver problemas educativos en los sistemas e-learning. En (Kadoic & Oreski, 2018), se analiza el impacto de las actividades implementado en Moodle para un curso de sobre la calificación final; usando los datos de los “archives log” se concluye que existe una fuerte correlación entre las actividades realizadas por los estudiantes y el rendimiento final. En (Félix, Ambrosio, Duilio, & Simoes, 2019), se analiza los datos de 1307 estudiantes correspondientes a la interacción y el uso de los contenidos y actividades de un curso en Moodle; se usa MoodlePredicta, una

herramienta para la minería de datos educativa que predice el rendimiento de los estudiantes aplicando el clasificador Naive de Bayes, en el estudio se identifica estudiantes con riesgo de abandono. En (Songsakda, C., Massudi, M., & Husniza, H., 2012), se aplican TMD para obtener un modelo que permita predecir el aprendizaje de los estudiantes que interactúan con Moodle, con la finalidad de mejorar su rendimiento e identificar aquellos con un bajo aprendizaje. En (Sael, N., Marzak, A., & Behja, H., 2013), con la finalidad de analizar el comportamiento del aprendizaje y ayudar a mejorar su evaluación en la plataforma Moodle, se aplican TM de clustering (k-medias y EM) para identificar perfiles de aprendizaje y las reglas de asociación para encontrar relaciones eventuales entre los diferentes contenidos. En (Swedan, 2012), se desarrolla la herramienta SMoodle que permite monitorear el comportamiento del aprendizaje de los estudiantes en la plataforma Moodle; aplicando las TMD k-medias, C4.5 y reglas de asociación. En (Fournier, 2011), se desarrolla una herramienta que permite a los estudiantes en tiempo real buscar sus contenidos pedagógicos y visualizar las conexiones de contenidos similares para la plataforma Moodle. Los profesores pueden retroalimentar a sus estudiantes enviando información y recursos, mejorando el rendimiento académico. En (Fahad, Khan, Shakir, & Sawar, Abu, 2012), las actividades de Moodle se organizan en nueve categorías (vistas de recursos, vistas de asignaciones, foros, etc.) y se aplican los árboles de decisión C4.5 con métricas chi-cuadrado, Gani, Relief y One-R para predecir la nota final del curso. En (Casey, K. & Gibson, P., 2010), usando los datos almacenados en los archivos “log” y las estadísticas que proporciona Moodle, se define métricas sobre las actividades y recursos (número de páginas visitadas, número de recursos visitados, número de visitas diarias, cantidad de material revisado, etc.); existe correlaciones significativas entre las métricas y la nota final de los estudiantes de los cursos dirigidos en Moodle.

3. Metodología

La metodología propuesta se basa en desarrollar el Proceso de Descubrimiento del Conocimiento en Base de Datos Educativas (PDCBE). La propuesta del PDCBE, consiste en realizar seis etapas iterativas con la finalidad de aplicar las TMD para extraer conocimiento a partir de los datos generados en la plataforma Moodle. En la Figura 1 se muestra el diagrama de las etapas propuestas.

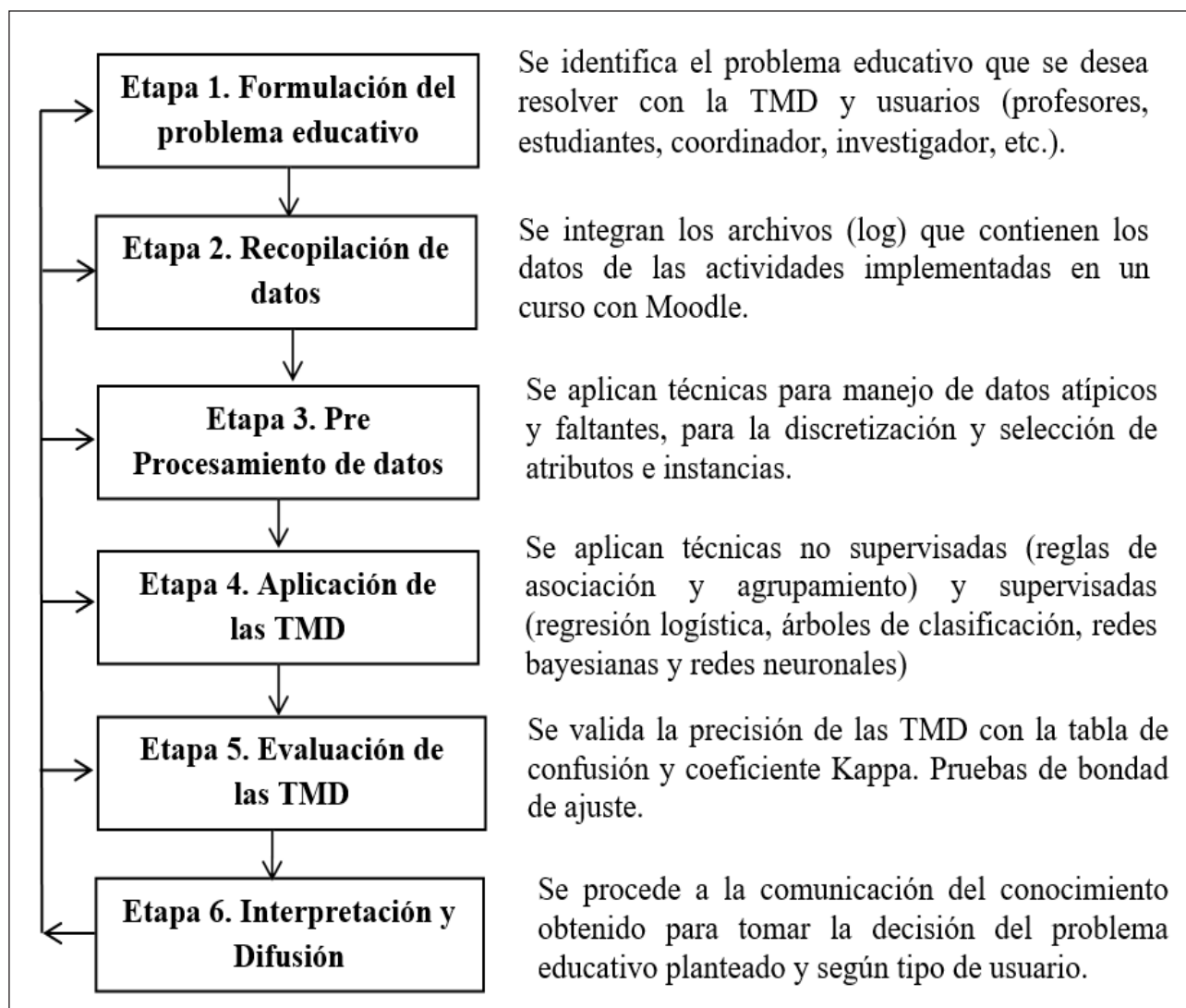


Figura 1. Proceso para el descubrimiento del conocimiento en base de datos educativas

Etapa 1. Formulación del problema educativo.

Se debe plantear con claridad el problema en el dominio educativo a resolver y el tipo de usuario (profesores, estudiantes, coordinadores académicos o investigadores) a quien estará dirigido los resultados obtenidos. Los procesos de enseñanza y aprendizaje pueden ser evaluados y analizados definiendo un conjunto de tareas educativas, tales como: el rendimiento académico, la retroalimentación, personalización del aprendizaje, recomendaciones de materiales, atraso, abandono de los estudiantes, segmentación de estudiantes, etc. (Romer, 2006).

Etapa 2. Recopilación de datos. Consiste en integrar y agregar todos los datos relevantes de

las diferentes fuentes en una única base de datos, considerando el problema educativo planteado. La plataforma Moodle, almacena todos los datos, que se generan como resultado de la interacción del estudiante con las diferentes actividades (archivos, cuestionarios, tareas, foros, encuestas, lecciones, etc.), en una base de datos relacional (My SQL) con aproximadamente 200 tablas relacionadas entre ellas. Si es necesario, se puede recopilar datos de otras bases de datos dentro de la institución de educación superior: datos académicos, datos personales, datos de admisión, datos uso bibliotecas, etc.). En la Figura 2, se muestra un modelo de datos (Data mart) que integra en un base de datos todas las tablas asociadas a las actividades de Moodle.

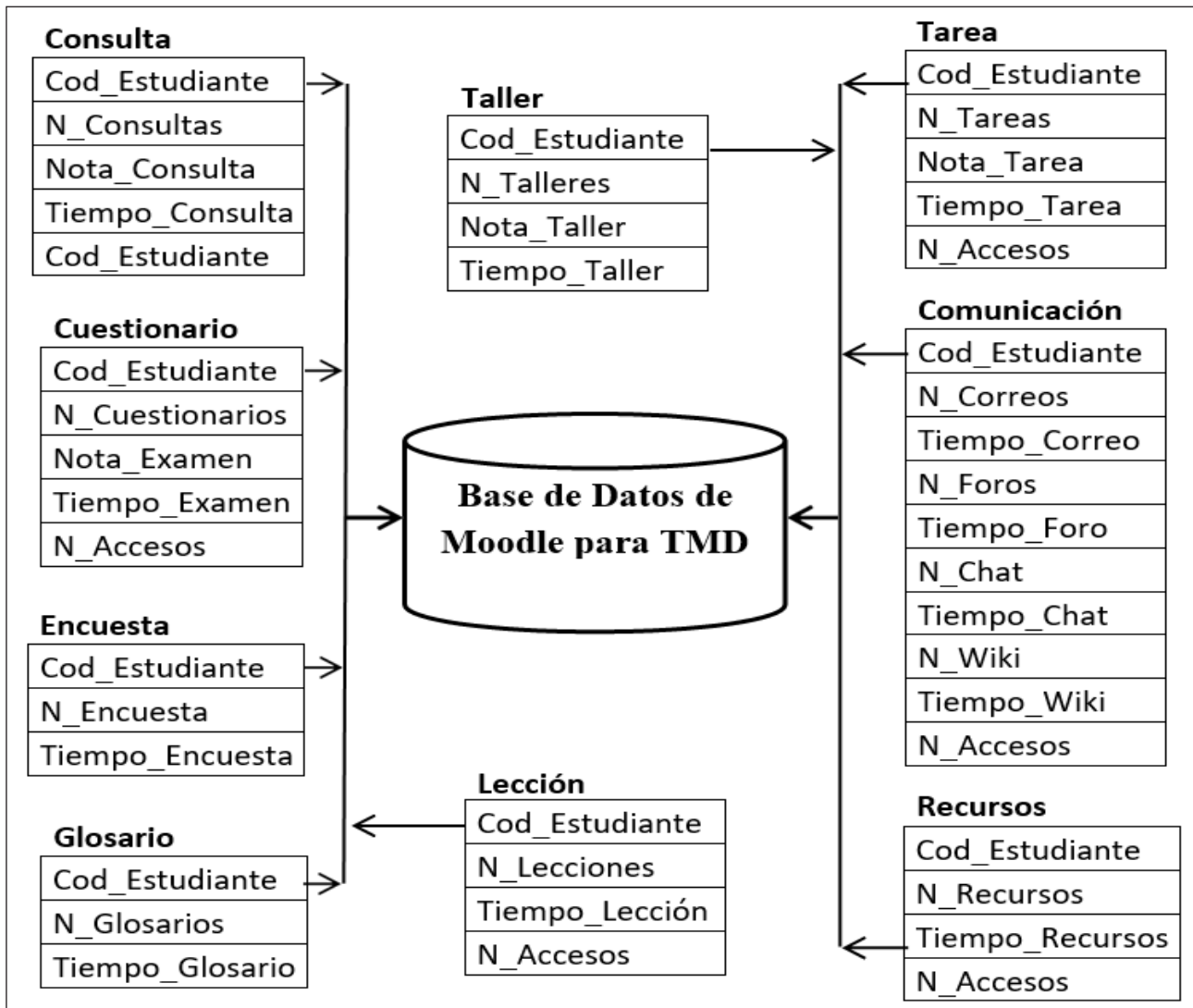


Figura 2. Modelo de datos de Moodle para aplicar las MD

Etap 3. Pre-Procesamiento. Esta etapa es muy importante, puesto que la validez del conocimiento descubierto depende en gran parte de la calidad de los datos analizados. Se aplican técnicas para la detección y el manejo de datos atípicos: estudiantes con notas altas en las evaluaciones, tiempos de acceso y foros muy altos, etc. Para datos faltantes: estudiantes que no responden todas las evaluaciones, talleres, foros, etc. Para el filtrado de datos y selección de atributos más relevantes: tipo de usuario, fechas, tipo de actividad. También se usan técnicas de discretización para convertir los datos cualitativos en cuantitativos. Las notas finales de los estudiantes de un curso pueden ser discretizadas en dos categorías: Desaprobado (si la nota es < 10), Aprobado (si la nota es ≥ 11) o aplicar una discretización con igual tamaño de intervalo: Bajo (nota es ≥ 0 y < 7), Medio (nota es ≥ 8 y < 14) o Alto (nota es ≥ 15 y ≤ 20). Se pueden crear nuevos atributos a partir de los existentes: calcular el promedio prácticas: $PR = (PC1 + PC2 + PC3 + PC4) / 4$ y el promedio

final: $PF = 0.3 * PR + 0.3 * EP + 0.4 * EF$.

Etap 4. Técnicas de minería de datos. Existen una variedad de técnicas de minería de datos que se pueden aplicar a los datos generados por la plataforma educativa Moodle. La aplicación de las TMD depende del problema educativo que se desea resolver.

- **Técnicas de asociación.** Se basa en identificar reglas de asociación (RA) usando una proposición probabilística para descubrir relaciones o dependencias más frecuentes entre grupos de variables. Su aplicación en Moodle: descubrir relaciones entre las actividades y las interacciones, patrones sobre el comportamiento y determinar el interés de ciertas actividades que realizan los estudiantes. El algoritmo que se usa se denomina APRIORI.
- **Técnicas de Agrupamiento.** El propósito es obtener grupos de individuos con características similares dentro de ellos y diferentes entre individuos de distintos grupos.

Su aplicación en Moodle es la agrupación de estudiantes por: calificación de cuestionarios, comportamiento de navegación, tiempo de acceso a las actividades, número de foros respondidos, rendimiento académico. Las técnicas de agrupamiento son: jerárquicos, k-medias, COBWEB y EM.

• **Técnicas de clasificación.** Se aplican cuando el conjunto de datos está previamente agrupado por un atributo clase. Se establece un modelo predictivo, con una variable dependiente (atributo clase) en función de un conjunto de variables predictoras, la finalidad es clasificar nuevas observaciones en alguna de las clases. En la plataforma Moodle, se aplica para predecir el rendimiento académico en función de un conjunto de variables explicativas (puntuaciones de exámenes, tiempo, número de accesos, etc.), clasificar a un estudiante de acuerdo con su interacción con las actividades, predecir el futuro abandono de estudiantes en un curso de Moodle, identificar las actividades para aprobar un curso, etc. Entre las técnicas de clasificación que se pueden aplicar se encuentran: el árbol de clasificación (ID3, C4.5, C5.0, CART y CHAID), la regresión logística, las redes bayesianas, las redes neuronales, el k vecino más cercano, las máquinas de soporte vectorial, etc.

En la Tabla 1, se presenta una propuesta de problemas educativos para la plataforma Moodle que pueden ser resueltos con la aplicación de las TMD.

Etapa 5. Evaluación de la TMD. En esta etapa se realiza la evaluación de los resultados conseguidos con la TMD. Se aplican medidas y criterios de bondad de ajuste e inferencia estadística a fin de comparar y validar los resultados de acuerdo a los parámetros del modelo o técnica propuesta. Entre las técnicas de evaluación se tiene: la matriz de confusión, la validación cruzada, curvas ROC, AIC, BIC, etc. En esta etapa, si la TMD no se ajusta a los datos o los resultados no permiten resolver el problema educativo formulado, se puede regresar a la Etapa 2 o Etapa 4 según sea el caso.

Etapa 6. Interpretación y Difusión de los resultados. En la etapa de interpretación de los resultados, se busca presentar en forma entendible y precisa los conocimientos obtenidos con la TMD aplicada. Cada TMD tiene su propia forma de presentar los resultados. Se deben identificar y analizar los resultados de acuerdo al patrón o conocimiento descubierto, enfocados a resolver el problema educativo formulado. La difusión, comprende la comunicación del conocimiento encontrado a fin de apoyar la toma de decisiones a los usuarios (profesores, estudiantes, coordinadores, etc.), con la finalidad de mejorar el proceso de enseñanza y aprendizaje de los estudiantes a través la plataforma Moodle. También, si se desea, se puede implementar los cambios basados en los resultados, a fin de que se apliquen en futuros semestres.

Tabla 1. Problemas educativos asociados a las TMD

Problema educativo a ser resuelto	TMD
1. Predecir el rendimiento académico final de los estudiantes	Clasificación
2. Predecir el abandono de estudiantes de un curso de Moodle	
3. Identificar los factores que influyen en la aprobación de un curso	
4. Predecir usos anómalos de las actividades de Moodle	
5. Agrupar a los estudiantes según su comportamiento con las actividades y recursos de Moodle	Agrupamiento
6. Agrupar a los estudiantes según sus perfiles de uso de las interacciones	
7. Agrupar a los estudiantes para realizar una retroalimentación más efectiva	
8. Identificar las relaciones entre las actividades y recurso usados en Moodle	Reglas de asociación
9. Recomendar objetos de aprendizaje según las preferencias	
10. Detectar perfiles de navegación	

4. Resultados

El caso de estudio, considera un curso de Estadística que esta implementado con una serie de actividades en la plataforma educativa Moodle. Se aplican las TMD de clasificación árboles de decisión J48 para predecir el resultado final del curso usando las reglas con la finalidad de identificar las actividades que influyen su aprobación, el algoritmo k-medias para agrupar a los estudiantes según sus notas y el uso de las actividades con Moodle con la finalidad de identificar estudiantes con bajos rendimiento y brindarles un mejor asesoramiento académico y reglas de asociación para seleccionar las mejores reglas que relacionen las actividades que realizan los estudiantes con Moodle, con la finalidad de mejorar los contenidos programáticos desarrollados del curso durante el ciclo académico.

Según la metodología propuesta en esta investigación, a continuación, se detalla cada una de las etapas:

Etap 1. Formulación del problema educativo

Los problemas educativos que se formulan para aplicar las TMD en la plataforma Moodle son:

- 1) La predicción del rendimiento académico a partir de las evaluaciones y tiempos de uso de las actividades, con la finalidad de identificar las actividades para la aprobación del curso. Se aplica árbol de decisión C4.5.
- 2) La agrupación de los estudiantes, con la finalidad de identificar estudiantes con bajo rendimiento y darles una retroalimentación adecuada. Se aplica el algoritmo el k- medias.
- 3) Las relaciones entre las diferentes actividades, con la finalidad de identificar las dependencias más relevantes. Se aplica reglas de asociación con el algoritmo A PRIORI.

Etap 2. Recopilación de datos

En la plataforma Moodle, toda la información generada por la interacción de los estudiantes con las actividades de un curso durante todo el semestre, se almacenan en archivos (log) que constituye una base de datos. Para el curso de Estadística, se ha implementado en Moodle las siguientes actividades: Cuestionarios (dos evaluaciones y cinco laboratorios) y tareas (siete tareas). Además, se recopiló los tiempos de uso de los cuestionarios, las tareas y el tiempo de uso (acceso a la plataforma). Para las notas de las evaluaciones, los laboratorios y las tareas, así como, sus respectivos tiempos de uso se calcularon los promedios respectivos. Se ha recopilado los datos de 585 alumnos matriculados en las 15 secciones que tiene el

curso de Estadística. En la Tabla 2, se presenta la descripción de las variables consideradas para esta aplicación.

Tabla 2. Descripción de las variables

Nomenclatura	Descripción
X1_Nota_Eval	Nota promedio de las evaluaciones
X2_Nota_Labo	Nota promedio de los laboratorios
X3_Nota_Tarea	Nota promedio de las tareas
X4_Tiempo_Eval	Tiempo promedio de uso de las evaluaciones
X5_Tiempo_Labo	Tiempo promedio de uso de los laboratorios
X6_Tiempo_Acceso	Tiempo de uso de Moodle
Y_Nota_Final	Nota promedio final del curso

Etap 3. Pre procesamiento

Con la finalidad de depurar los datos, se procedió a eliminar 30 observaciones con datos faltantes (estudiantes que no rindieron alguna evaluación, tarea y laboratorio). Aquellos estudiantes, a los que le faltó rendir alguna evaluación, generaron datos faltantes, los cuales se reemplazaron por el promedio dentro de cada sección. Para el algoritmo C4.5 y reglas de asociación, se aplicó la técnica de discretización. Para las notas promedio de las evaluaciones, laboratorios, tareas y final, se discretizó en dos categorías (Aprobado: $\text{Nota} \geq 13$ y Desaprobado: $\text{Nota} < 13$). Para los tiempos promedios de uso de las evaluaciones, laboratorios y Moodle, se aplicó una discretización con igual tamaño de los intervalos de clase. El número de intervalos se fijó a 3 ($k=3$), etiquetándolos como: Bajo, Medio y alto. Al finalizar el pre procesamiento resultó la base de datos de 555 estudiantes.

Etap 4. Aplicación de las TMD

Se aplican las TMD de árbol de decisión con el algoritmo C4.5, para obtener reglas de clasificación que permitan identificar las actividades que afecta en rendimiento de los estudiantes y su aprobación del curso, agrupación con el algoritmo k-medias, con la finalidad de agrupar a los estudiantes según su comportamiento de uso de la plataforma Moodle y reglas de asociación con la finalidad de identificar dependencias entre las actividades. Para obtener los resultados de las TMD se aplica el programa Weka.

a. Aplicación del árbol de decisión C4.5

Se aplicó el algoritmo J48 que el programa

Weka implementa el árbol de decisión C4.5. El árbol C4.5 se aplicó con una validación cruzada 10-folds y un factor de confianza de 0.25. En la Tabla 3, se presenta la matriz de confusión para la clasificación con el árbol C4.5. La tasa de buena clasificación tuvo un valor de una precisión del 90.1%.

Tabla 3. Matriz de confusión con C4.5

Clasificación real	Clasificación predicha		Total
	Aprobado	Desaprobado	
Aprobado	361	27	388
Desaprobado	28	139	167
Total	389	166	555

En la Figura 3, se presenta el árbol C4.5 con un tamaño de 11 nodos y 6 nodos terminales, donde se muestra las reglas que indican cuando un estudiante apruebe o desaprobe el curso de estadística en Moodle.

X1_Nota_Eval = Desaprobado
X3_Nota_Tarea = Aprobado
X2_Nota_Labo = Desaprobado: Desaprobado (17.0/3.0)
X2_Nota_Labo = Aprobado: Aprobado (73.0/25.0)
X3_Nota_Tarea = Desaprobado: Desaprobado (127.0/18.0)
X1_Nota_Eval = Aprobado
X2_Nota_Labo = Desaprobado
X3_Nota_Tarea = Aprobado: Aprobado (24.0)
X3_Nota_Tarea = Desaprobado: Desaprobado (18.0/2.0)
X2_Nota_Labo = Aprobado: Aprobado (296.0/3.0)

Figura 3. Árbol de clasificación con C4.5

b. Aplicación de k-medias

Se aplicó la técnica de agrupamiento de k-medias con el programa Weka. En la Tabla 4, se presenta

la distribución de los estudiantes del curso categorizados en tres grupos.

Tabla 4. Distribución de la agrupación de estudiantes

Grupo	Número de estudiantes	Porcentaje de estudiantes
1	213	38.0%
2	132	24.0%
3	210	38.0%
Total	555	100.0%

En la Tabla 5, se presenta los análisis de variancias para cada una de las variables y su significación. Todas las variables resultaron significativas con un nivel de 0.05, indicando que los tres grupos formados son estadísticamente diferentes, con lo cual se confirma la validación de los tres grupos de estudiantes.

En la Tabla 6, se presenta los promedios de cada una de las variables en cada grupo. En el Grupo 1, se encuentran los estudiantes con un rendimiento medio, en el Grupo 2, los de menores rendimientos y en el Grupo 3 los de mayores rendimientos; que corresponden con valores de los promedios de notas y tiempo de las evaluaciones, laboratorios y tareas intermedios, bajas y altas respectivamente.

c. Aplicación reglas de asociación

Se aplicó reglas de asociación con el algoritmo A PRIORI, con una confianza mínima del 0.9 y mínimo soporte de 0.4. En la Figura 4, se presenta las cinco primeras reglas encontradas para la aprobación y desaprobación de un estudiante en el curso.

Tabla 5. Cuadro de los ANVA para cada variable

Variables	Grupo		Error		Fc	Sig.
	C. M. Grupos	G.L.	C. M. Error	G.L.		
Auto Promedio	276,228	2	11,425	552	24,177	,000
Lab Promedio	129,898	2	6,468	552	20,083	,000
Tarea Promedio	469,005	2	4,094	552	114,555	,000
Promedio Final	272,284	2	4,106	552	66,309	,000
Tiempo Auto	27241,111	2	111,684	552	243,912	,000
Tiempo Lab	47442,930	2	70,191	552	675,912	,000
Tiempo Clase	507,330	2	22,572	552	22,476	,000

Tabla 6. Promedios para cada uno de los grupos

Grupo	Nota Promedio Evaluación	Nota Promedio Laboratorio	Nota Promedio Tarea	Tiempo Promedio Evaluación	Tiempo Promedio Laboratorio	Tiempo Acceso	Nota Promedio Final
1	13.5	15.1	13.5	53.3	44.0	46.9	14.1
2	11.8	13.6	10.9	46.6	41.9	39.2	11.0
3	15.0	16.3	15.7	62.1	65.3	48.7	15.8

1. X1_Nota_Eval=Aprobado X3_Nota_Tarea=Aprobado 122 ==> Y_Nota_Final=Aprobado
 2. X1_Nota_Eval=Aprobado X2_Nota_Labo=Aprobado X3_Nota_Tarea=Aprobado 110 ==> Y_Nota_Final=Aprobado 110
 3. X1_Nota_Eval=Aprobado X2_Nota_Labo=Aprobado X4_Tiempo_Eval=Alto 80 ==> Y_Nota_Final=Aprobado 80
 4. X1_Nota_Eval=Aprobado X3_Nota_Tarea=Aprobado X4_Tiempo_Eval=Alto 76 ==> Y_Nota_Final=Aprobado
 5. X1_Nota_Eval=Aprobado X2_Nota_Labo=Aprobado 146 ==> Y_Nota_Final=Aprobado 143
 13. X1_Nota_Eval=Desaprobado X5_Tiempo_Labo=Bajo X6_Tiempo_Acceso=Bajo 82 ==> Y_Nota_Final=Desaprobado 78
 22. X1_Nota_Eval=Desaprobado X3_Nota_Tarea=Desaprobado 117 ==> Y_Nota_Final=Desaprobado 109
 26. X1_Nota_Eval=Desaprobado X3_Nota_Tarea=Desaprobado X6_Tiempo_Acceso=Bajo 113 ==> Y_Nota_Final=Desaprobado
 31. X1_Nota_Eval=Desaprobado X4_Tiempo_Eval=Bajo X6_Tiempo_Acceso=Bajo 80 ==> Y_Nota_Final=Desaprobado 73
 40. X1_Nota_Eval=Desaprobado X5_Tiempo_Labo=Bajo 90 ==> Y_Nota_Final=Desaprobado 81

Figura 4. Reglas de asociación con APRIORI

Etapas 5. Interpretación y difusión de los resultados

- Los resultados con el árbol C4.5. Las reglas para que un estudiante del curso de Estadística desaprobe el curso son:

R1: Si desaprovea la nota de evaluación, aprueba la nota de tarea y desaprovea la nota de laboratorio.

R2: Si desaprovea la nota de evaluación y desaprovea la nota de tarea.

R3: Si aprueba la nota de evaluación, desaprovea la nota de laboratorio y desaprovea la nota de tarea.

El profesor, debe considerar que las notas de las evaluaciones y tareas son las que más afectan la desaprobación de un estudiante en el curso de Estadística.

- Los resultados con k-medias. Se obtuvo la clasificación de cada uno de los estudiantes a algún grupo (El grupo 1, están los estudiantes con un rendimiento medio, el Grupo 2 con menores y el Grupo 3 con mayores). En la Tabla 7, se presenta la asignación de los primeros 12 estudiantes a cada uno de los

grupos que el algoritmo k-medias los clasificó. El profesor, deberá dar una retroalimentación y asesoramiento a los estudiantes del Grupo 2

Tabla 7. Distribución de los estudiantes a cada grupo

Grupo	Relación de estudiantes
1	E1 E6 E8 E9 E10 E12 E13 E14 E17 E21 E22 E24 ...
2	E4 E7 E16 E18 E23 E33 E37 E38 E39 E45 E48 E49 ...
3	E0 E2 E3 E5 E11 E15 E19 E20 E25 E30 E32 E34 ...

- Reglas de asociación. Las reglas de asociación más relevantes para considerar la desaprobación de un estudiante en el curso de estadística son:

R1: Si la nota de la evaluación es desaprobatoria y el tiempo de laboratorio y acceso a Moodle son bajos.

R2: Si la nota de la evaluación y tarea es desaprobatoria.

R3: Si la nota de la evaluación y tarea es desaprobatoria y el tiempo de acceso a Moodle es bajo.

R4: Si la nota de la evaluación es desaprobatoria y el tiempo de la evaluación y acceso son bajos.

R5: Si la nota de la evaluación es desaprobatoria y el tiempo de laboratorio es bajo

El profesor, debe considerar la desaprobación de las evaluaciones y tareas y el tiempo de acceso como los factores que afectan la desaprobación de un estudiante en el curso de estadística.

5. Conclusiones

Los datos que se generan cuando los estudiantes interactúan con las actividades implementadas para un curso en la plataforma educativa Moodle, puede ser analizadas y explotadas con las TMD, con la finalidad de apoyar a los profesores y estudiantes a mejorar el proceso de enseñanza y aprendizaje. La aplicación con la metodología propuesta, con el árbol C4.5 permitió sugerir al profesor considerar los cuestionarios y

las tareas como actividades que afectan la desaprobación de un estudiante. El k-medias clasificó a los estudiantes en tres grupos, donde el Grupo 2 son los que tienen menores rendimientos y por lo tanto se sugiere al profesor brindarles una mayor retroalimentación y asesoramiento. Las reglas de asociación, identificaron dependencias entre la desaprobación de las notas de evaluaciones y tareas con el bajo tiempo de acceso a Moodle, como factores que afectan la desaprobación de los estudiantes en el curso de Estadística.

6. Referencias bibliográficas

- Baker, R., & Yacef, . (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Education Data Mining* , Volume 1, Issue 1 1: 3– 17.
- Barnes, T. D. (2009). *Data Mining 2009: 2nd International Conference on Educational Data Mining*.
- Casey, K., & Gibson, P. (2010). (m)Oodles of Data Mining Moodle to understand Student Behaviour. *International Conference on Engaging Pedagogy 2010 (ICEP10)* National University of Ireland Maynooth .
- Castro, F., Vellido, A., Nebot, & Mugica, F. (2007). Applying data mining techniques to e- learning problem. *Journal Evolution of Teaching and Learning Paradigms in Intelligent Environment.ed. Springer Berlin Heidelberg* , pp. 183-221.
- Fahad, M., Khan, Shakir, & Sawar, Abu. (2012). Using Instructive Data Mining Methods to Revise the Impact of Virtual Classroom in E-Learning - IMPRESO. *International Journal of Advanced Science and Technology*. Vol. 45 , pp. 125-134.
- Félix, I., Ambrosio, A., Duilio, J., & Simoes, E. (2019). Predicting Student Outcome in Moodle. *Conference: Academic Success in Higher Education* , pp. 1-2.
- Flate, M. P. (2003). Online education and learning management systems. *Proceedings of the 4th International Conference on e-Learning*.
- Fournier, H. K. (2011). The value of learning analytics to networked learning on a personal learning environment. *Proceedings of the First International Conference on Learning Analytics and Knowledge (LAK '11)* (pp. 104-109). New York , pp. 104-109.
- Hijon, R., & Velázquez, A. (2006). E-learning Platforms Analysis and Development of Students Tracking Functionality. *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications* , pp. 2823-2828.
- Kadoic, N., & Oreski, D. (2018). Analysis of Student Behavior and Success Based on Logs in Moodle. pp. 730-735.
- Kumar, B., & Pal, Saurabh. (2011). Mining Educational Data to Analyze Students.
- Ramaswami, M. (2009). A Study on Feature Selection Techniques in Educational Data Mining. *International working group on educational data mining* , Vol. 1, Issue 1.
- Romer, C. a. (2006). Educational data mining: A survey from 1995 to 2005. Elsevier. *Expert Systems with Applications* 33 , pp. 135-146.
- Romero, C. a. (2010). Educational Data Mining: A Review of the State-of-the-Art. *IEEE* , Vol. XX, N°. X.
- Sael, N., Marzak, A., & Behja, H. (2013). Multilevel clustering and association rule mining for learners' profiles analysis. *IJCSI International Journal of Computer Science Issues*, Vol. 10, Issue 3, No 1. , pp. 202-210.
- Songsakda, C., Massudi, M., & Husniza, H. (2012). Behavior Usage Model to Manage the Best Practice of e-Learning. *Knowledge Management International Conference* , pp. 4-6.
- Swedan, M. I. (2012). Mining YUC1 Students Learning Behavior in Moodle System Using Data Mining Techniques. *UNIASCIT*, Vol. 2(4) , pp. 318-823.