



Predicción de rendimiento del curso en línea en la educación superior con datos de LMS: Una revisión sistemática de literatura

Predicting online course performance in higher education with LMS data: A systematic literature review

 <https://doi.org/10.47230/unesum-ciencias.v9.n2.2025.202-219>

Recibido: 10-02-2025

Aceptado: 11-03-2025

Publicado: 25-05-2025

Laura Daniela Llivipuma Yáñez¹

 <https://orcid.org/0009-0001-6971-8905>

Ermenson Ricardo Ordoñez-Avila^{2*}

 <https://orcid.org/0000-0003-2583-2076>

1. Facultad de Ciencias Informáticas, Universidad Técnica de Manabí; Portoviejo, Ecuador.

2. Departamento de Sistemas Computacionales; Universidad Técnica de Manabí; Portoviejo, Ecuador.

Volumen: 9

Número: 2

Año: 2025

Paginación: 202-219

URL: <https://revistas.unesum.edu.ec/index.php/unesumciencias/article/view/950>

***Correspondencia autor:** ermenson.ordonez@utm.edu.ec

RESUMEN

Las instituciones de educación superior han incrementado el uso de sistemas de gestión de aprendizaje (LMS) para facilitar la educación en línea y optimizar los procesos formativos. Estos entornos no solo permiten el acceso remoto a contenidos y actividades, sino que también generan grandes volúmenes de datos sobre las interacciones entre estudiantes y docentes, los cuales pueden utilizarse para predecir el rendimiento en cursos en línea. Con la expansión de la educación virtual, se reconoce la necesidad de métodos precisos de predicción basados en datos registrados por los LMS. En este marco, la minería de datos educativa (EDM) se plantea como una herramienta eficaz para identificar patrones y apoyar la toma de decisiones académicas. Este artículo presenta una revisión sistemática de la literatura sobre el uso de datos de LMS para la predicción del rendimiento académico en línea, considerando técnicas y enfoques aplicados. Las técnicas más utilizadas incluyen regresión logística, árboles de decisión, redes bayesianas y redes neuronales, aplicadas al análisis de datos relacionados con accesos, calificaciones y participación. Asimismo, se identifican como métricas comunes de evaluación la exactitud, precisión y el error promedio, y como procesos clave la selección de características, normalización y validación cruzada. Este estudio proporciona una visión actualizada del uso de técnicas de aprendizaje automático en entornos virtuales, contribuyendo a futuras investigaciones sobre el análisis del rendimiento estudiantil en la educación superior en línea.

Palabras clave: Minería de datos educativos (EDM), LMS, predicción, Rendimiento, Educación superior.

ABSTRACT

Higher education institutions have increased their use of learning management systems (LMS) to facilitate online education and optimize learning processes. These environments allow remote access to content and activities and generate large volumes of data on student-teacher interactions, which can be used to predict performance in online courses. With the expansion of virtual education, the need for accurate prediction methods based on data recorded by LMSs is recognized. In this framework, educational data mining (EDM) is proposed as an effective tool to identify patterns and support academic decision-making. This article presents a systematic literature review on using LMS data for online academic performance prediction, considering applied techniques and approaches. The most commonly used techniques include logistic regression, decision trees, Bayesian networks, and neural networks, which are applied to data analysis related to access, ratings, and participation. Standard evaluation metrics are identified as accuracy, precision, and average error, and feature selection, normalization, and cross-validation are identified as key processes. This study provides an updated view of machine learning techniques in virtual environments, contributing to future research on student performance in online higher education.

Keywords: Educational data mining (EDM), LMS, Prediction, Performance, Higher education.



Creative Commons Attribution 4.0
International (CC BY 4.0)

Introducción

Los sistemas de gestión de aprendizaje en la educación superior

Las instituciones de educación superior están adoptando de manera creciente los sistemas de gestión de aprendizaje (LMS, por sus siglas del inglés Learning Management System) para ofrecer soluciones de aprendizaje en línea y optimizar las prácticas educativas mediante el aprovechamiento de sus diversas funcionalidades (Al-Hunaiyyan et al., 2020; Ashrafi et al., 2020). La incorporación de estas plataformas ha impulsado la transformación de la educación en modalidad en línea, posicionando a los LMS como una de las herramientas más útiles en entornos virtuales, al ofrecer acceso remoto a cursos y materiales educativos (Bringula et al., 2021; Zabolotniaia et al., 2020). En torno a sus beneficios, no solo ofrecen recursos educativos, sino que también generan grandes cantidades de datos que pueden ser analizados para enfoques como la predicción del rendimiento académico (Aitdaoud et al., 2023; Salihoun, 2020).

La minería de datos educativa (EDM, por sus siglas en inglés de Educational Data Mining) se ha empleado para explorar información, especialmente en problemas de predicción dentro de entornos académicos (Orozco et al., 2021). La EDM permite identificar patrones de comportamiento en los datos, lo que puede contribuir a la identificación de factores asociados con el éxito del aprendizaje en línea (Mancilla-Vela et al., 2020; Morales et al., 2019). Además, es posible explorar diferentes patrones o comportamientos de usuarios en el aula virtual, tanto de estudiantes como de docentes, con el objetivo de comprender mejor las formas de aprendizaje y así apoyar decisiones que garanticen el éxito en el proceso educativo (Yağcı, 2022). Entre los métodos aplicados en la EDM destacan la regresión logística, los árboles de decisión, las redes bayesianas y las redes neuronales, los cuales se utilizan ampliamente para predecir el rendi-

miento en un curso en línea (Job & Pandey, 2020; Ramaswami et al., 2020; Sharma & Sharma, 2020).

Este trabajo se enfoca en sintetizar y evaluar críticamente la literatura existente sobre el uso de datos de LMS para predecir el rendimiento de cursos en línea en la educación superior. El objetivo es identificar las diferentes técnicas empleadas y las variables que han sido destacadas en la literatura. Además, se busca presentar información sobre el rendimiento de los modelos predictivos, lo que permitirá valorar la efectividad de las técnicas utilizadas.

Esta revisión propone una base para futuras investigaciones que busquen aplicar métodos de minería de datos enfocados en la predicción del rendimiento en cursos en línea. No solo se consideran a los estudiantes como actores principales, sino también se exploran y analizan los datos registrados de la interacción de los profesores en los LMS.

Materiales y Métodos

Se realizó una revisión sistemática de la literatura, definida en las etapas que sugiere el modelo PRISMA: planificación; búsqueda y selección; y, extracción de información. El protocolo desarrollado en la revisión sistemática de la literatura se resume a continuación:

Planificación

En la etapa de planificación se organizaron los pasos: objetivo de búsqueda, preguntas de investigación y criterios de inclusión y exclusión. El objetivo principal de esta revisión de la literatura fue identificar estudios que integren aplicaciones de predicción enfocadas en técnicas de minería de datos en el contexto de la predicción del rendimiento del curso en línea en la educación superior. Para este propósito, se definieron cinco preguntas de investigación clave. Estas preguntas, junto con sus descripciones y motivaciones, se presentan en la Tabla 1.

Tabla 1.

Preguntas de investigación de revisión sistemática de la literatura

Preguntas de Investigación	Motivación y resultados esperados
P1: ¿Qué técnicas de aprendizaje automático se han utilizado para la predicción del rendimiento académico en entornos de educación superior en línea?	Obtener referencias de la aplicación de técnicas de minería de datos en modelado predictivo.
P2: ¿Qué datos de los sistemas de gestión de aprendizaje se han explorado para la predicción del rendimiento del curso en línea?	Identificar qué características o predictores son útiles para explicar el rendimiento del curso en línea.
P3: ¿Se ha explorado datos de estudiantes y docentes en conjunto, para la predicción del rendimiento en la educación en línea?	Examinar el uso de datos de estudiantes y docentes reportados en la literatura en relación con el análisis del rendimiento en cursos en línea.
P4: ¿Cuáles son las métricas más utilizadas para validar los modelos de predicción enfocados en el rendimiento del curso en línea?	Identificar las métricas utilizadas para evaluar el rendimiento de los modelos predictivos basados en datos de LMS.
P5: ¿Qué técnicas de aprendizaje automático se han utilizado para mejorar el ajuste de los modelos de predicción con datos del LMS?	Describir las técnicas que han permitido mejorar el ajuste de modelos predictivos basados en datos de LMS.

Se agregaron los criterios de inclusión: estudios enfocados en técnicas de minería de datos educativos en el contexto de la educación superior, publicados en el idioma inglés, artículos de revisión e investigación, enfocados principalmente en el rendimiento del curso en línea en la educación superior con datos de LMS, incluyendo técnicas de minería de datos y algoritmos de predicción. Con esto, se incluye artículos que mencionan métodos de minería de datos utilizados para la predicción, documentos relacionados con la Minería de Datos Educativos, y estudios realizados a partir del año 2019.

Por otro lado, dentro de los criterios de exclusión, se descartaron artículos que no reportan específicamente el uso de minería de datos, publicaciones anteriores al año 2019, estudios presentados en congresos, actas de conferencias, tesis de grado y posgrado, e informes. Además, se excluyeron aquellos artículos que utilizaran únicamente métodos descriptivos o estadísticos generales, así como investigaciones que no analizaran datos procedentes de sistemas de gestión de aprendizaje.

Búsqueda y selección

La fase de búsqueda y selección comenzó con la definición de términos clave enfocados en la minería de datos y sus técnicas para la predicción, además del uso de datos del LMS. Se estableció un grupo de control de artículos preliminares que sirvieron como base para normalizar e identificar términos frecuentes para la construcción de cadenas de búsqueda candidatas. Entre los términos incluidos se encuentran: Minería de Datos Educativos (EDM), LMS, Moodle, predicción, rendimiento, logs, y aprendizaje en línea. Se realizaron varias combinaciones para encontrar la cadena de búsqueda ideal con estos términos. Como resultado, la cadena seleccionada se definió como: LMS AND prediction AND performance AND logs AND "online learning"

El siguiente paso fue la depuración de los resultados de la búsqueda. La búsqueda se realizó en las bases de datos: ACM, IEEE Xplore, Web of Science (WoS), Scopus, Springer Link y Science Direct. La Tabla 3 presenta los resultados de la búsqueda, así como la sintaxis de las cadenas de búsqueda utilizadas en cada base de datos.

Tabla 2.

Aplicación de la cadena de búsqueda en las fuentes de información y resultados preliminares

No.	Base de datos	Formato de cadenas de investigación
1	ACM	[All: lms] AND [All: prediction] AND [All: performance] AND [All: logs] AND [All: "online learning"]
2	WoS	LMS AND prediction AND performance AND LOGS AND "online learning" (Topic)
3	IEEE Xplore	LMS AND prediction AND performance AND LOGS AND "online learning"
4	Science Direct	LMS AND prediction AND performance AND LOGS AND "online learning"
5	Scopus	TITLE-ABS-KEY (lms AND prediction AND performance AND logs AND "online learning")
6	Springer Link	LMS AND prediction AND performance AND LOGS AND "online learning"

La fase de identificación permitió recopilar los registros obtenidos de las bases de datos seleccionadas para la búsqueda. Durante la fase de cribado, estos registros fueron evaluados en función de los criterios de inclusión y exclusión previamente establecidos. Se determinó el número de documentos excluidos y se detallaron los motivos para cada exclusión. A los documentos restantes se les aplicó el filtro de "artículo a texto completo", y se evaluó su pertinencia. Al igual que en la fase anterior, se especificaron las razones para descartar estudios no aptos. La tercera y última fase fue la de inclusión, donde se obtuvo la muestra final de publicaciones, incluyendo registros identificados en otras fuentes relevantes para la revisión.

Los estudios primarios fueron sometidos a un proceso de evaluación de calidad, considerando tanto el enfoque general como su contribución a esta investigación. Este análisis permitió identificar elementos estructurales que cumplen con los estándares esperados para el desarrollo del estudio. La calidad de los estudios se clasificó en tres niveles: moderada, aceptable y óptima. Los estudios con una calificación del 70% o superior se consideraron de calidad óptima, aquellos con una calificación entre el 50%

y el 69% se clasificaron como aceptables, y los estudios con una calidad inferior al 50% se catalogaron como moderados.

Para evaluar la calidad, se consideraron dos grupos principales: uno que evalúa el contexto general y otro que analiza aspectos específicos del contenido. Estos grupos se desglosaron en los criterios: 1) enfoque, relacionado con el objeto de estudio, centrado en modelos predictivos aplicados a la evaluación de cursos en línea en la educación superior con datos de LMS, y 2) contenido, incluyendo la claridad de los objetivos, métodos y técnicas propuestas, datos y modelos en minería de datos educativos, así como la presentación de resultados y limitaciones.

La evaluación de la calidad de los estudios primarios fue realizada por pares, utilizando un sistema de calificación ponderado en tres niveles. El método de calificación se definió en función del cumplimiento de cada criterio mencionado, con valores asignados de la siguiente manera: 0 = no cumple, 0,5 = cumplimiento parcial, y 1 = cumplido satisfactoriamente. Con esta evaluación se reduce el sesgo en la selección de los artículos y su información.

Extracción de información

En la etapa de extracción de información se incluyeron criterios: propósito del estudio, contextos de aplicación, análisis de

resultados y conclusiones, datos de entrada y salida, métodos basados en minería de datos, técnicas utilizadas. La Tabla 3 describe detalladamente cada uno de los criterios mencionados.

Tabla 3.

Descripción de los criterios de extracción de información

Criterios	Descripción
C1. Propósito del estudio	Identificar el problema resuelto, el propósito de la investigación y su relación con el enfoque del SLR.
C2. Contextos de aplicación	Detectar casos de uso y escenarios relacionados en diversos contextos de aplicación.
C3. Técnicas de aprendizaje automático	Identificar las técnicas de aprendizaje automático empleadas en problemas de predicción en entornos de educación superior en línea.
C4. Características o predictores analizados	Identificar los datos y sus tipos, así como las muestras y los detalles del procesamiento de estos datos.
C5. Métricas de rendimiento de modelos	Reconocer las métricas que han sido utilizadas para evaluar el rendimiento de los modelos de aprendizaje automático.
C6. Técnicas de ajuste de modelos	Reconocer las técnicas que mejoran el ajuste de los modelos de predicción.

Resultados y Discusión

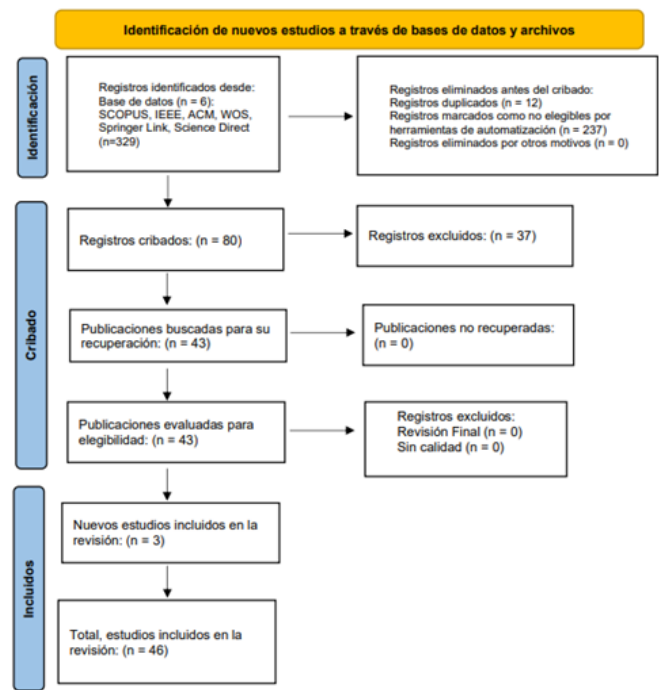
Análisis descriptivo de los resultados de la búsqueda y selección de artículos

En cuanto a los resultados de la búsqueda y selección de artículos, en la fase 1) Identificación, se procesaron inicialmente 329 estudios. Tras eliminar 12 estudios duplicados, se procedió a evaluar 317 estudios mediante la lectura de títulos, resúmenes y palabras clave. De estos, 237 estudios fueron excluidos por no cumplir con los criterios de elegibilidad descritos en la Tabla 2, y por no estar enfocados en los aspectos técnicos requeridos en esta investigación, como el campo de la computación, modelos de predicción y estudios realizados con datos de educación superior. En la fase 2) Cribado, se analizaron 80 estudios a partir de la lectura del texto completo, de los cuales 37 artículos fueron excluidos por razones motivadas por la ausencia de aplicación de técnicas de minería de datos y la ausencia de propuestas de predicción del rendimiento del curso en

línea en la educación superior. En la fase 3) Inclusión, se añadieron 3 artículos a través de métodos manuales de selección en otras bases de datos. En total, se incluyeron 46 artículos en el SLR, que permitieron responder a las preguntas de investigación. La figura 1 resume el flujo y los resultados obtenidos durante el proceso de búsqueda y selección de los artículos.

Figura 1.

Diagrama de flujo de la búsqueda y selección de artículos primarios



En la siguiente sección se describen las razones por las cuales ciertos estudios no cumplieron con los criterios de elegibilidad. La Tabla 4 detalla la justificación de estas

exclusiones, categorizadas en función de los motivos: otros niveles de educación, otros campos y objetivos de estudio, y ausencia de aplicaciones predictivas.

Tabla 4.

Justificación de la eliminación de artículos, en el cribado

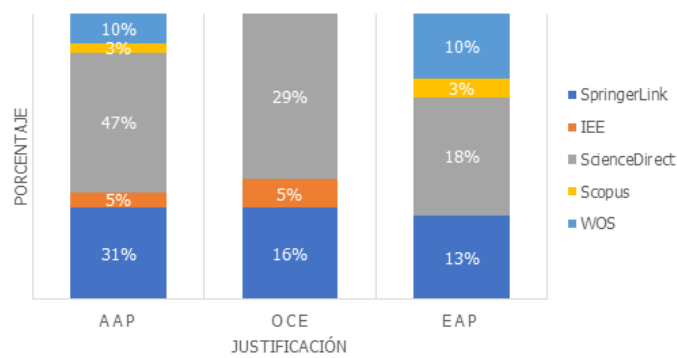
Artículos	Técnicas	Algoritmos	Continente	Países
(Contreras et al., 2020; Mancilla-Vela, et al., 2020; Morales Carrillo et al., 2019)	Clasificación	Árbol de decisión, Estadísticas descriptivas, Rede neuronal, K-Nearest Neighbors (KNN).	América del Sur	Ecuador, Chile, Colombia
(Uzir et al., 2020)	Agrupamiento, Clasificación, Regresión	K-means, Agrupamiento Jerárquico, Árboles de Decisión, Máquinas de Vectores de Soporte (SVM), K-Nearest Neighbors (KNN), Regresión Lineal	Oceanía	Australia
(Batoool et al., 2022; Cunha & Figueira, 2021; Park & Yoo, 2021; Saidani et al., 2024; Uzir et al., 2020)	Agrupamiento, Clasificación, Regresión	K-means, Agrupamiento Jerárquico, Árboles de Decisión, Máquinas de Vectores de Soporte (SVM), K-Nearest Neighbors (KNN), Regresión Lineal	Europa	Serbia, Reino Unido, Italia, Portugal
(Batoool et al., 2022; Dol & Jawandhiya, 2024; Huang & Chen,	Agrupamiento,	K-means, Agrupamiento Jerárquico, Árboles de Decisión, Random Forest, Red neuronal,		Malasia, Corea del Sur, Tailandia

2024; Huang & Zeng, 2024; Kaensar & Wongnin, 2023; Páez, 2023; Park & Yoo, 2021; Saidani et al., 2024; Uzir et al., 2020; Yoo et al., 2022)	Clasificación, Regresión	Máquinas de Vectores de Soporte (SVM), Deep learning (DNN), K-Nearest Neighbors (KNN), Regresión Lineal, Regresión logística, J48, Naïve Bays	Asia	Arabia Saudita, Pakistán, India, China
---	--------------------------	---	------	--

Algunas publicaciones no fueron finalmente seleccionadas en el grupo de estudios primarios porque analizaron datos de cursos en línea desde una perspectiva descriptiva, sin incluir un análisis predictivo, además de estar enfocadas en otras áreas y objetivos aislados de esta investigación. La Figura 2 muestra las categorías que resumen las principales razones por las que se excluyeron estos estudios en la fase de cribado.

Figura 2.

Recuento de artículos eliminados en la fase de cribado y sus justificaciones. AAP: Ausencia de aplicaciones predictivas. OCE: otros campos y objetos de estudio. EAP: otras áreas de enfoque y propósitos de investigación



Análisis de la calidad de los estudios primarios

En esta sección se presentan los resultados de la evaluación del texto completo incluido en el SLR. El número de artículos obtenidos inicialmente fueron 43, posteriormente, fueron incluidos mediante métodos manuales 3

artículos, con un total de 46 estudios primarios. La mayor proporción de estudios fue obtenida de la base de datos Scopus, tal como se presenta en la Figura 3. Mientras que la Figura 4 muestra los porcentajes de estos estudios por año de publicación, alrededor del 24% de los trabajos corresponden al año 2020.

Figura 3.

Texto completo incluido en SLR, por bases de datos

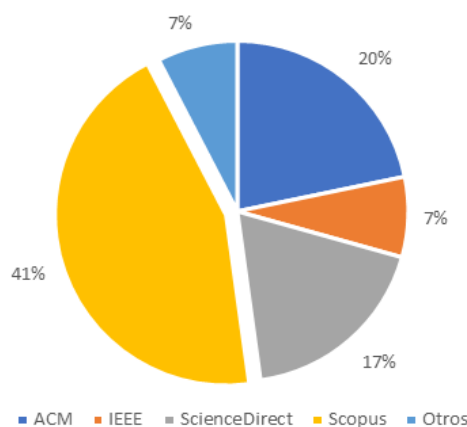
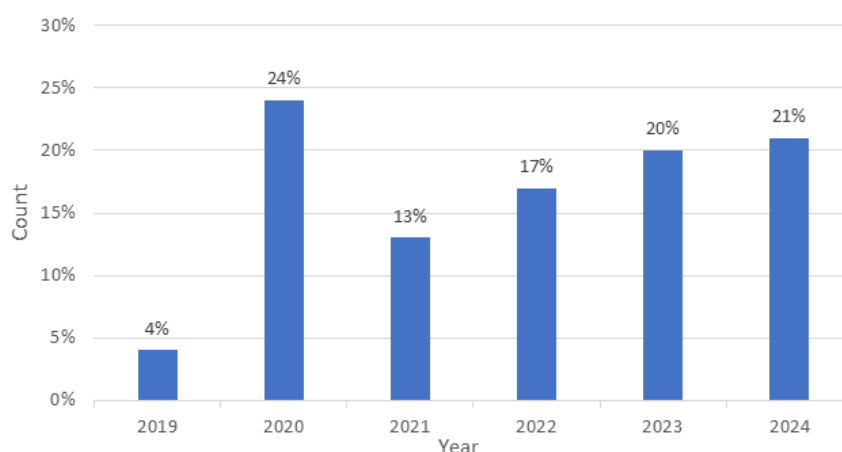


Figura 4.

Texto completo incluido en SLR, por año de publicación



La Tabla 5, presenta 23 de los estudios primarios brindan referencias a la predicción del rendimiento del curso en línea, la aplicación de la predicción en este campo y la asociación de datos LMS para la evaluación del curso en línea en la educación superior. En comparación, 15 de los 46 artículos obtuvieron una evaluación media al no brindar aportes específicos al enfoque de la predicción del curso en línea.

En cuanto a la evaluación del contenido, se destaca la ausencia de limitaciones descritas en los 29 estudios primarios; 23 de estos

trabajos incorporaron apartados o secciones que describen las limitaciones en cada investigación. Las limitaciones son aportes esenciales para este trabajo de investigación. La mayoría de las limitaciones están centradas en el acceso o disponibilidad de datos para el aprendizaje automático, así como la ingeniería de datos complejos. Finalmente, 25 de los 29 estudios primarios tienen una calificación óptima por encima del 70% en la evaluación del enfoque de investigación. Todos los estudios cumplieron satisfactoriamente los criterios de eva-

luación en la revisión de contenido con calificaciones superiores al 70%. La calidad ponderada de los estudios finales, se basa principalmente en la aplicación de varias técnicas y algoritmos que analizan un gran volumen de datos, por tratarse de los regis-

tros de Moodle y sus características complejas. Así también, destaca la utilidad de estos estudios, que en la práctica pueden ser replicados en favor de la gestión de profesores y administradores en instituciones de educación superior.

Tabla 5.
Resultados de la evaluación de la calidad de los estudios primarios

Artículos	Enfoque				Contenido				%
	E1	E2	E3		C1	C2	C3	C4	
(Contreras et al., 2020; Cheng et al., 2023; Huang & Zeng, 2024; Kaensar & Wongnin, 2023; Páez, 2023; Park & Yoo, 2021; Yoo et al., 2022)	1	1	0.5	100%	1	1	1	1	100%
(Huang & Chen, 2024; Maraza-Quispe et al., 2021; Le et al., 2020; Park & Yoo, 2021; Quispe et al., 2021; Uzir et al., 2020)	0.5	0.5	1	80%	0.5	1	1	1	100%
(Çebi & Güyer, 2020; Contreras et al., 2020; Cruz et al., 2022; Cunha & Figueira, 2021; Kew & Tasir, 2022; Luo et al., 2024; Mancilla-Vela et al., 2020)	0.5	1	1	90%	1	1	0.5	1	95%
(Bringula et al., 2021; Dias et al., 2020; Hernández-García et al., 2024; Morales Carrillo et al., 2019; Park & Yoo, 2021; Poellhuber et al., 2023; Tsiakmaki et al., 2020; Van Goidsenhoven et al., 2020; Viberg et al., 2020)	1	0.5	0.5	80%	1	1	1	1	95%
(Batool et al., 2022; Cheng et al., 2023; Dol & Jawandhiya, 2024; Huang & Zeng, 2024; Perkash et al., 2024; Rico-Juan et al., 2024; Saidani et al., 2024)	1	1	1	100%	1	1	0.5	1	95%
(Alam, 2023; Aldowah et al., 2019; Rivera Torres, 2022; Joy & Pillai, 2022; Juan, 2021; Páez, 2023; Palanci et al., 2024; Ramaswami et al., 2020; Yoo et al., 2022)	0.5	1	1	90%	0.5	1	1	1	95%

Discusión

Además de discutir y tomar como referencia la experiencia de los autores en contextos específicos de minería de datos, y la predicción del curso en línea a partir de datos LMS, se consideró mencionar algunos trabajos que fueron excluidos, por ser un aporte indirecto a esta investigación, en el sentido de información asociada en otros contextos, escenarios y enfoques que puedan sustentar el debate. A continuación, se analiza cada

contribución de los artículos para responder a las preguntas de investigación.

Sobre la pregunta de investigación P1: ¿Qué técnicas de aprendizaje automático se han utilizado para la predicción del rendimiento académico en entornos de educación superior en línea? Los hallazgos son diversos debido a los numerosos artículos que informan sobre la aplicación de técnicas de minería de datos utilizadas en la educación en línea. Sin embargo, las técnicas y algo-

ritmos identificados como resultado de las estrategias de búsqueda son mencionados en la Tabla 6, agrupados en países de forma continental.

Tabla 6.
Matriz de técnicas implementadas para la predicción del rendimiento académico en entornos de educación superior en línea

Artículos	Técnicas	Algoritmos	Continente	Países
(Contreras et al., 2020; Mancilla-Vela, et al., 2020; Morales Carrillo et al., 2019)	Clasificación	Árbol de decisión, Estadísticas descriptivas, Rede neuronal, K-Nearest Neighbors (KNN).	América del Sur	Ecuador, Chile, Colombia
(Uzir et al., 2020)	Agrupamiento, Clasificación, Regresión	K-means, Agrupamiento Jerárquico, Árboles de Decisión, Máquinas de Vectores de Soporte (SVM), K-Nearest Neighbors (KNN), Regresión Lineal	Oceanía	Australia
(Batool et al., 2022; Cunha & Figueira, 2021; Park & Yoo, 2021; Saidani et al., 2024; Uzir et al., 2020)	Agrupamiento, Clasificación, Regresión	K-means, Agrupamiento Jerárquico, Árboles de Decisión, Máquinas de Vectores de Soporte (SVM), K-Nearest Neighbors (KNN), Regresión Lineal	Europa	Serbia, Reino Unido, Italia, Portugal
(Batool et al., 2022; Dol & Jawandhiya, 2024; Huang & Chen, 2024; Huang & Zeng, 2024; Kaensar & Wongnin, 2023; Páez, 2023; Park & Yoo, 2021; Saidani et al., 2024; Uzir et al., 2020; Yoo et al., 2022)	Agrupamiento, Clasificación, Regresión	K-means, Agrupamiento Jerárquico, Árboles de Decisión, Random Forest, Red neuronal, Máquinas de Vectores de Soporte (SVM), Deep learning (DNN), K-Nearest Neighbors (KNN), Regresión Lineal, Regresión logística, J48, Naïve Bays	Asia	Malasia, Corea del Sur, Tailandia, Arabia Saudita, Pakistán, India, China

Las técnicas implementadas son algoritmos basados en Naïve Bays (Contreras et al., 2020; Páez, 2023), K-means (Huang & Chen, 2024; (Park & Yoo, 2021; Uzir et al., 2020), Agrupamiento Jerárquico (Kaensar & Wongnin, 2023), Árboles de Decisión (Morales Carrillo et al., 2019),J48 (Yoo et al., 2022), Máquinas de Vectores de Soporte (SVM) (Kaensar & Wongnin, 2023; Park & Yoo, 2021), K-Nearest Neighbors (KNN) (Saidani et al., 2024), Regresión Lineal(Cunha & Figueira, 2021), Random Forest (Batool et al., 2022; Dol & Jawandhiya, 2024), Red neuronal (Huang & Zeng, 2024; Mancilla-Vela, et al., 2020). Los estudios referidos han aprovechado la minería de datos educativos para proponer y aplicar modelos de

predicción, empleando diferentes técnicas y algoritmos.

La técnica de clasificación se utiliza ampliamente para la predicción del rendimiento del curso. Para predecir el rendimiento del curso en línea se ha utilizado Naïve Bays como una de las técnicas que dio mejores resultados (Dol & Jawandhiya, 2024; Saidani et al., 2024). En este marco, se reporta mayor efectividad con el empleo de las técnicas: K-means, Agrupamiento Jerárquico, Árboles de Decisión, Red neuronal Máquinas de Vectores de Soporte (SVM), K-Nearest Neighbors (KNN), Regresión Lineal, Regresión logística, J48, y Random Forest; siendo este último altamente valorado por su robustez

y precisión (Contreras et al., 2020; Dol & Jawandhiya, 2024; Huang & Zeng, 2024; Bringula et al., 2021).

Las redes neuronales se están volviendo cada vez más comunes debido a su capacidad para aprender de los datos (Tsiakmaki et al., 2020). Los métodos tradicionales como la regresión lineal y los árboles de decisión siguen siendo ampliamente utilizados, pero pueden ser superados en casos que requieran modelado más sofisticado (Contreras et al., 2020; Morales Carrillo et al., 2019). Estos algoritmos han demostrado su capacidad para identificar patrones en grandes conjuntos de datos, permitiendo realizar la predicción del rendimiento de un curso en línea.

Respecto a la pregunta de investigación P2: ¿Qué datos de los sistemas de gestión de aprendizaje se han explorado para la predicción del rendimiento del curso en línea? Para analizar y comparar la predicción del rendimiento académico utilizando datos de un Learning Management System (LMS) se ha reportado el uso de los siguientes indicadores: Rendimiento académico, Acceso, Tareas, Aspectos sociales y Exámenes (Maraza-Quispe et al., 2021). Por su parte, Quispe et al. (2021) emplearon en su investigación los indicadores: número de accesos, calificaciones de los cursos, tiempo empleado, número de cursos inscritos y número de actividades desarrolladas. Mientras que también se ha analizado el número de visitas, el número de publicaciones, el número de visitas al foro y el número de tareas entregadas a tiempo (Hernández-García et al., 2024; Le et al., 2020).

Datos como la cantidad de conexiones del sistema, la cantidad de clases, y los datos de calificaciones del semestre anterior también han sido utilizados (Park & Yoo, 2021). En estos trabajos, se han seleccionado distintos datos tales como de participación, calificaciones, e interacciones para poder predecir el rendimiento de un curso en línea. Además de la selección adecuada de atri-

butos, también influye que los estudiantes apliquen estrategias propias de aprendizaje, y se generen nuevos datos para el análisis.

Respecto a la pregunta P3: ¿Se ha explorado datos de estudiantes y docentes en conjunto para la predicción del rendimiento en la educación en línea? Se ha analizado datos enfocados y direccionados en la mejora del rendimiento de los cursos en línea en beneficio de los estudiantes (Le et al., 2020; Maraza-Quispe et al., 2021; Uzir et al., 2020; Yağcı, 2022). Por otro lado, se hace mención del análisis de datos de los LMS con inclinación en la facilidad del manejo de los sistemas de gestión de aprendizaje por parte de los docentes, para que los estudiantes puedan estar más interesados en el material proporcionado (Hernández-García et al., 2024; Quispe et al., 2021). Desde la perspectiva de la minería de datos, la combinación de datos de estudiantes y docentes ofrece una oportunidad para mejorar la capacidad de predicción y eficacia en entornos de aprendizaje en línea.

En respuesta a P4: ¿Cuáles son las métricas más utilizadas para validar los modelos de predicción enfocados en el rendimiento del curso en línea? En múltiples estudios se aplicaron métricas de exactitud y precisión como las técnicas que permiten determinar el mejor rendimiento de los modelos de predicción (Cunha & Figueira, 2021; Cruz et al., 2022; Mancilla-Vela, et al., 2020). También se ha evaluado modelos aplicando métricas como el promedio de error mediante el error cuadrático medio (MSE). Respecto al MSE, al principio su valor puede ser el más bajo cuando la cantidad de datos es reducida, y la precisión del modelo tiende a aumentar si se agregan más datos (Morales Carrillo et al., 2019).

Finalmente, en respuesta a P5: ¿Qué técnicas de aprendizaje automático se han utilizado para mejorar el ajuste de los modelos de predicción con datos del LMS? Se reporta el uso de la selección de características para identificar las columnas más relevan-

tes para la predicción y eliminar aquellas redundantes o irrelevantes. Así también, se emplea la normalización para escalar los datos en favor de mejorar el rendimiento de los algoritmos de aprendizaje automático; y la validación cruzada para dividir los datos en conjuntos de entrenamiento y validación para evaluar el rendimiento del modelo en datos no vistos previamente (Contreras et al., 2020; Páez, 2023; Uzir et al., 2020).

Conclusiones

Se han explorado diversas técnicas de minería de datos para predecir el rendimiento de los cursos en línea utilizando datos de sistemas de gestión de aprendizaje (LMS). En este estudio, se revisaron 46 artículos que cumplieron con los criterios de inclusión y exclusión, lo que facilitó la respuesta a las preguntas de investigación propuestas en la revisión sistemática de la literatura (SLR).

Se exploraron diferentes datos de sistemas de gestión de aprendizaje para predecir el rendimiento de los cursos en línea, incluyendo acceso, tareas, aspectos sociales y exámenes. También se han investigado datos de estudiantes y docentes juntos para entender mejor las estrategias de aprendizaje y mejorar los resultados. Para validar los modelos de predicción, se emplearon métricas como la exactitud, la precisión, el error promedio y la validación cruzada.

Las técnicas de aprendizaje automático, tales como la selección de características, la normalización de datos y la validación cruzada, fueron utilizadas para optimizar la precisión y la robustez de los modelos. Finalmente, el presente trabajo proporciona una visión a otros investigadores sobre los datos LMS que ya se han considerado y dar paso a que otros datos pueden ser posiblemente considerados para la predicción de un curso en línea también ayuda a identificar qué técnicas de minería de datos fueron consideradas en el contexto de esta investigación ofreciendo una guía útil para futuras investigaciones en este campo.

Bibliografía

- Aitdaoud, M., Namir, A., & Talbi, M. (2023). A New Pre-Processing Approach Based on Clustering Users Traces According to their Learning Styles in Moodle LMS. *International Journal of Emerging Technologies in Learning (IJET)*, 18(07), 226–242. <https://doi.org/10.3991/IJET.V18I07.37635>
- Alam, A. (2023). Improving Learning Outcomes through Predictive Analytics: Enhancing Teaching and Learning with Educational Data Mining. *Proceedings of the 7th International Conference on Intelligent Computing and Control Systems, ICICCS 2023*, 249–257. <https://doi.org/10.1109/ICICCS56967.2023.10142392>
- Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37, 13–49. <https://doi.org/10.1016/J.TELE.2019.01.007>
- Al-Hunaiyyan, A., Al-Sharhan, S., & Alhajri, R. (2020). Prospects and Challenges of Learning Management Systems in Higher Education. *Article in International Journal of Advanced Computer Science and Applications*, 11(12). <https://doi.org/10.14569/IJACSA.2020.0111209>
- Alnasyan, B., Basher, M., & Alassafi, M. (2024). The power of Deep Learning techniques for predicting student performance in Virtual Learning Environments: A systematic literature review. *Computers and Education: Artificial Intelligence*, 6, 100231. <https://doi.org/10.1016/J.CAEAI.2024.100231>
- Alshammari, A. (2024). Using analytics to predict students' interactions with learning management systems in online courses. *Education and Information Technologies*, 1–26. <https://doi.org/10.1007/S10639-024-12709-9/METRICS>
- Angskun, T., Na Sritha, K., Srithong, A., Khopolklang, N., Kamollimsakul, S., Phithak, T., & Angskun, J. (2024). Using big data to assess an affective domain for distance education. *Future Generation Computer Systems*, 160, 131–139. <https://doi.org/10.1016/J.FUTURE.2024.05.057>
- Ashrafi, A., Zareravasan, A., Rabiee Savoji, S., & Amani, M. (2020). Exploring factors influencing students' continuance intention to use the learning management system (LMS): a multi-perspective framework. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2020.1734028>

- Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review*, 37, 100489. <https://doi.org/10.1016/J.EDUREV.2022.100489>
- Batool, S., Rashid, J., Nisar, M. W., Kim, J., Kwon, H. Y., & Hussain, A. (2022). Educational data mining to predict students' academic performance: A survey study. *Education and Information Technologies* 2022, 28:1, 28(1), 905–971. <https://doi.org/10.1007/S10639-022-11152-Y>
- Bertolini, R., Finch, S. J., & Nehm, R. H. (2021). Enhancing data pipelines for forecasting student performance: integrating feature selection with cross-validation. *International Journal of Educational Technology in Higher Education*, 18(1), 1–23. <https://doi.org/10.1186/S41239-021-00279-6/FIGURES/8>
- Bognár, L., & Fauszt, T. (2022). Factors and conditions that affect the goodness of machine learning models for predicting the success of learning. *Computers and Education: Artificial Intelligence*, 3, 100100. <https://doi.org/10.1016/J.CAEAI.2022.100100>
- Bravo-Agapito, J., Romero, S. J., & Pamplona, S. (2021). Early prediction of undergraduate Student's academic performance in completely online learning: A five-year study. *Computers in Human Behavior*, 115, 106595. <https://doi.org/10.1016/J.CHB.2020.106595>
- Bringula, R., Batalla, M. Y., & Borebor, M. T. (2021). Modeling Computing Students' Perceived Academic Performance in an Online Learning Environment. *SIGITE 2021 - Proceedings of the 22nd Annual Conference on Information Technology Education*, 99–104. <https://doi.org/10.1145/3450329.3476856>
- Çebi, A., & Güyer, T. (2020). Students' interaction patterns in different online learning activities and their relationship with motivation, self-regulated learning strategy and learning performance. *Education and Information Technologies*, 25(5), 3975–3993. <https://doi.org/10.1007/S10639-020-10151-1/METRICS>
- Cheng, Z., Zhang, Z., Xu, Q., Maeda, Y., & Gu, P. (2023). A meta-analysis addressing the relationship between self-regulated learning strategies and academic performance in online higher education. *Journal of Computing in Higher Education*, 1–30. <https://doi.org/10.1007/S12528-023-09390-1/METRICS>
- Contreras, L. E., Fuentes, H. J., Rodríguez, J. I., Contreras, L. E., Fuentes, H. J., & Rodríguez, J. I. (2020). Predicción del rendimiento académico como indicador de éxito/fracaso de los estudiantes de ingeniería, mediante aprendizaje automático. *Formación Universitaria*, 13(5), 233–246. <https://doi.org/10.4067/S0718-50062020000500233>
- Cruz, E., González, M., & Rangel Ortiz, J. C. (2022). Técnicas de machine learning aplicadas a la evaluación del rendimiento y a la predicción de la deserción de estudiantes universitarios, una revisión. *Prisma Tecnológico*, ISSN-e 2076-8133, ISSN 2312-637X, Vol. 13, No. 1, 2022 (Ejemplar Dedicado a: Prisma Tecnológico), Págs. 77-87, 13(1), 77–87. <https://dialnet.unirioja.es/servlet/articulo?codigo=9064058&info=resumen&idoma=SPA>
- Cui, Y., Song, X., Hu, Q., Li, Y., Shanthini, A., & Vadi-vel, T. (2021). Big data visualization using multimodal feedback in education. *Computers & Electrical Engineering*, 96, 107544. <https://doi.org/10.1016/J.COMPELECENG.2021.107544>
- Cunha, A., & Figueira, Á. (2021a). Analysing students' interaction sequences on Moodle to predict academic performance. <http://ceur-ws.org>
- Cunha, A., & Figueira, Á. (2021b). Analysing students interaction sequences on Moodle to predict academic performance Analysing students' interaction sequences on Moodle to predict academic performance. <http://ceur-ws.org>
- Deeva, G., De Smedt, J., Saint-Pierre, C., Weber, R., & De Weerd, J. (2022). Predicting student performance using sequence classification with time-based windows. *Expert Systems with Applications*, 209, 118182. <https://doi.org/10.1016/J.ESWA.2022.118182>
- Dewi, M. A., Kurniadi, F. I., Murad, D. F., Rabiha, S. G., & Romli, A. (2023). Machine Learning Algorithms for Early Predicting Dropout Student Online Learning. 2023 IEEE 9th International Conference on Computing, Engineering and Design, ICCED 2023. <https://doi.org/10.1109/ICCED60214.2023.10425359>
- Dhanalakshmi, R., Namasivayam, B., Revathy, S., & Balasubramani, S. (2023). Analyzing Student Performance and Interaction in Online Learning Environments using LMS Log Data. 7th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2023 - Proceedings, 1036–1042. <https://doi.org/10.1109/ICECA58529.2023.10395561>

- Dirk Tempelaar, B. R. Q. N. (2021). The Contribution of Dispositional Learning Analytics to Precision Education. https://www.researchgate.net/publication/355336388_The_Contribution_of_Dispositional_Learning_Analytics_to_Precision_Education
- Dol, S. M., & Jawandhiya, P. M. (2024). Systematic Review and Analysis of EDM for Predicting the Academic Performance of Students. *Journal of The Institution of Engineers (India): Series B*, 105(4), 1021–1071. <https://doi.org/10.1007/S40031-024-00998-0/METRICS>
- Feldman-Maggor, Y., Blonder, R., & Tuvi-Arad, I. (2022). Let them choose: Optional assignments and online learning patterns as predictors of success in online general chemistry courses. *The Internet and Higher Education*, 55, 100867. <https://doi.org/10.1016/J.IHEDUC.2022.100867>
- Hamad, F., Shehata, A., & Al Hosni, N. (2024). Predictors of blended learning adoption in higher education institutions in Oman: theory of planned behavior. *International Journal of Educational Technology in Higher Education*, 21(1), 1–28. <https://doi.org/10.1186/S41239-024-00443-8/TABLES/15>
- Han, H. (2023). Fuzzy clustering algorithm for university students' psychological fitness and performance detection. *Heliyon*, 9(8), e18550. <https://doi.org/10.1016/J.HELİYON.2023.E18550>
- Hernández-García, Á., Cuenca-Enrique, C., Del-Río-Carazo, L., & Iglesias-Pradas, S. (2024). Exploring the relationship between LMS interactions and academic performance: A Learning Cycle approach. *Computers in Human Behavior*, 155, 108183. <https://doi.org/10.1016/J.CHB.2024.108183>
- Huang, Q., & Chen, J. (2024). Enhancing academic performance prediction with temporal graph networks for massive open online courses. *Journal of Big Data*, 11(1), 1–26. <https://doi.org/10.1186/S40537-024-00918-5/FIGURES/6>
- Huang, Q., & Zeng, Y. (2024). Improving academic performance predictions with dual graph neural networks. *Complex and Intelligent Systems*, 10(3), 3557–3575. <https://doi.org/10.1007/S40747-024-01344-Z/FIGURES/7>
- Imhof, C., Comsa, I. S., Hlosta, M., Parsaeifard, B., Moser, I., & Bergamin, P. (2023). Prediction of Dilatory Behavior in eLearning: A Comparison of Multiple Machine Learning Models. *IEEE Transactions on Learning Technologies*, 16(5), 648–663. <https://doi.org/10.1109/TLT.2022.3221495>
- Job, M. A., & Pandey, J. (2020). Academic Performance Analysis Framework for Higher Education by Applying Data Mining Techniques.
- Rivera Torres, K. (2022). Técnicas de clasificación para la predicción del desempeño de los estudiantes usando minería de datos. Benemérita Universidad Autónoma de Puebla. <https://hdl.handle.net/20.500.12371/18244>
- Jovanovic, J., Mirriahi, N., Gašević, D., Dawson, S., & Pardo, A. (2019). Predictive power of regularity of pre-class activities in a flipped classroom. *Computers & Education*, 134, 156–168. <https://doi.org/10.1016/J.COMPEDU.2019.02.011>
- Jovanović, J., Saqr, M., Joksimović, S., & Gašević, D. (2021). Students matter the most in learning analytics: The effects of internal and instructional conditions in predicting academic success. *Computers & Education*, 172, 104251. <https://doi.org/10.1016/J.COMPEDU.2021.104251>
- Joy, J., & Pillai, R. V. G. (2022). Review and classification of content recommenders in E-learning environment. *Journal of King Saud University - Computer and Information Sciences*, 34(9), 7670–7685. <https://doi.org/10.1016/J.JKSUCI.2021.06.009>
- Juan, S. (2021). Promoting engagement of nursing students in online learning: Use of the student-generated question in a nursing leadership course. *Nurse Education Today*, 97. <https://doi.org/10.1016/J.NEDT.2020.104710>
- Kaensar, C., & Wongnin, W. (2023). Analysis and Prediction of Student Performance Based on Moodle Log Data using Machine Learning Techniques. *International Journal of Emerging Technologies in Learning (IJET)*, 18(10), 184–203. <https://doi.org/10.3991/IJET.V18I10.35841>
- Kerruish, E. (2024). Postdigital Teaching of Critical Thinking in Higher Education: Non-Instrumentalised Sociality and Interactivity. *Postdigital Science and Education*, 1–20. <https://doi.org/10.1007/S42438-024-00456-6/METRICS>
- Kew, S. N., & Tasir, Z. (2022). Learning Analytics in Online Learning Environment: A Systematic Review on the Focuses and the Types of Student-Related Analytics Data. *Technology, Knowledge and Learning*, 27(2), 405–427. <https://doi.org/10.1007/S10758-021-09541-2/METRICS>

- Kuhlmann, S. L., Plumley, R., Evans, Z., Bernacki, M. L., Greene, J. A., Hogan, K. A., Berro, M., Gates, K., & Panter, A. (2024). Students' active cognitive engagement with instructional videos predicts STEM learning. *Computers & Education*, 216, 105050. <https://doi.org/10.1016/J.COMPEDU.2024.105050>
- Le, M. D., Nguyen, H. H., Nguyen, D. L., & Nguyen, V. A. (2020a). How to Forecast the Students' Learning Outcomes Based on Factors of Interactive Activities in a Blended Learning Course. *ACM International Conference Proceeding Series*, 11–15. <https://doi.org/10.1145/3404709.3404711>
- Le, M. D., Nguyen, H. H., Nguyen, D. L., & Nguyen, V. A. (2020b). How to Forecast the Students' Learning Outcomes Based on Factors of Interactive Activities in a Blended Learning Course. *ACM International Conference Proceeding Series*, 11–15. <https://doi.org/10.1145/3404709.3404711>
- Li, X., Zhang, Y., Cheng, H., Li, M., & Yin, B. (2022). Student achievement prediction using deep neural network from multi-source campus data. *Complex and Intelligent Systems*, 8(6), 5143–5156. <https://doi.org/10.1007/S40747-022-00731-8/FIGURES/5>
- Luo, Y., Han, X., & Zhang, C. (2024). Prediction of learning outcomes with a machine learning algorithm based on online learning behavior data in blended courses. *Asia Pacific Education Review*, 25(2), 267–285. <https://doi.org/10.1007/S12564-022-09749-6/METRICS>
- Mancilla-Vela, G., Leal-Gatica, P., Sánchez-Ortiz, A., Vidal-Silva, C., Mancilla-Vela, G., Leal-Gatica, P., Sánchez-Ortiz, A., & Vidal-Silva, C. (2020). Factores asociados al éxito de los estudiantes en modalidad de aprendizaje en línea: un análisis en minería de datos. *Formación Universitaria*, 13(6), 23–36. <https://doi.org/10.4067/S0718-50062020000600023>
- Maraza-Quispe, B., Valderrama-Chauca, E. D., Cari-Mogrovejo, L. H., & Apaza-Huanca, J. M. (2021). Predictive Model of Student Academic Performance from LMS data based on Learning Analytics. *ACM International Conference Proceeding Series*, 13–19. <https://doi.org/10.1145/3498765.3498768>
- Meletiyou-Mavrotheris, M., Katzis, K., Sofianidis, A., Stylianidou, N., Konstantinou-Katzi, P., Ting, T. T., Lin Teh, S., & Wee, M. C. (2022). Ascertaining the Online Learning Behaviors and Formative Assessments Affecting Students' Academic Performance during the COVID-19 Pandemic: A Case Study of a Computer Science Course. *Education Sciences* 2022, Vol. 12, Page 935, 12(12), 935. <https://doi.org/10.3390/EDUCSCI12120935>
- Morales Carrillo, J. J., Trujillo Utreras, V. K., Cevallos Molina, S. K., & Hiraida Monserrate, S. C. (2019). Minería de datos en educación: una revisión literaria. *Revista ESPAMCIENCIA*, ISSN-e 1390-8103, Vol. 10, No. Extra 1, 2019, Págs. 14-20, 10(1), 14–20. <https://dialnet.unirioja.es/servlet/articulo?codigo=7020087&info=resumen&idoma=ENG>
- Naseem, M., Chaudhary, K., & Sharma, B. (2022). Predicting Freshmen Attrition in Computing Science using Data Mining. *Education and Information Technologies*, 27(7), 9587–9617. <https://doi.org/10.1007/S10639-022-11018-3/METRICS>
- Orozco Iguasnia, J. B., & Villarroel Sánchez, M. de J. (2021). Aplicación de técnicas de minería de datos para predecir el desempeño académico de los estudiantes de la escuela 'Lic. Angélica Villón L.' *Revista Científica y Tecnológica UPSE (RCTU)*, 8(2), 68–75. <https://doi.org/10.26423/RCTU.V8I2.637>
- Páez, A. R. (2023). Predicción del rendimiento académico mediante selección de características de estudiantes universitarios. *Revista Electrónica Sobre Tecnología, Educación y Sociedad*, 10(19). <https://www.ctes.org.mx/index.php/ctes/article/view/801>
- Palanci, A., Yilmaz, R. M., & Turan, Z. (2024). Learning analytics in distance education: A systematic review study. *Education and Information Technologies*, 1–22. <https://doi.org/10.1007/S10639-024-12737-5/TABLES/6>
- Park, H. S., & Yoo, S. J. (2021). Early Dropout Prediction in Online Learning of University using Machine Learning. *International Journal on Informatics Visualization*, 5(4), 347–353. <https://doi.org/10.30630/JOIV.5.4.732>
- Pecuchova, J., & Drlik, M. (2023). Predicting Students at Risk of Early Dropping Out from Course Using Ensemble Classification Methods. *Procedia Computer Science*, 225, 3223–3232. <https://doi.org/10.1016/J.PROCS.2023.10.316>
- Perkash, A., Shaheen, Q., Saleem, R., Rustam, F., Villar, M. G., Alvarado, E. S., Torre Diez, I. de la, & Ashraf, I. (2024). Feature optimization and machine learning for predicting students' academic performance in higher education institutions. *Education and Information Technologies*, 1–25. <https://doi.org/10.1007/S10639-024-12698-9/METRICS>

- Poellhuber, L. V., Poellhuber, B., Desmarais, M., Leger, C., & Manh-Chien Vu, M. (2023). Cluster-Based Performance of Student Dropout Prediction as a Solution for Large Scale Models in a Moodle LMS. *ACM International Conference Proceeding Series*, 592–598. <https://doi.org/10.1145/3576050.3576146>
- Quispe, B. M., Apfata, J. E. N., Figueroa, R. C. Q., & Solis, M. A. V. (2021). Design proposal of a personalized Dashboard to optimize teaching-learning in Virtual Learning Environments. *ACM International Conference Proceeding Series*, 77–84. <https://doi.org/10.1145/3498765.3498777>
- Ramaswami, G. S., Susnjak, T., Mathrani, A., & Umer, R. (2020). Predicting Students Final Academic Performance using Feature Selection Approaches. 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering, CSDE 2020. <https://doi.org/10.1109/CSDE50874.2020.9411605>
- Rico-Juan, J. R., Cachero, C., & Macià, H. (2024). Study regarding the influence of a student's personality and an LMS usage profile on learning performance using machine learning techniques. *Applied Intelligence*, 54(8), 6175–6197. <https://doi.org/10.1007/S10489-024-05483-1/TABLES/4>
- Saidani, O., Umer, M., Alshardan, A., Alturki, N., Nappi, M., & Ashraf, I. (2024). Student academic success prediction in multimedia-supported virtual learning system using ensemble learning approach. *Multimedia Tools and Applications*, 1–26. <https://doi.org/10.1007/S11042-024-18669-Z/METRICS>
- Sáiz-Manzanares, M. C., Marticorena-Sánchez, R., Rodríguez-Díez, J. J., Rodríguez-Arribas, S., Díez-Pastor, J. F., & Ji, Y. P. (2021). Improve teaching with modalities and collaborative groups in an LMS: an analysis of monitoring using visualisation techniques. *Journal of Computing in Higher Education*, 33(3), 747–778. <https://doi.org/10.1007/S12528-021-09289-9/FIGURES/9>
- Salihoun, M. (2020). State of Art of Data Mining and Learning Analytics Tools in Higher Education. *International Journal of Emerging Technologies in Learning*, 15(21), 58–76. <https://doi.org/10.3991/IJET.V15I21.16435>
- Sharma, P., & Sharma, Dr. S. (2020). DATA MINING TECHNIQUES FOR EDUCATIONAL DATA: A REVIEW. *International Journal of Engineering Technologies and Management Research*, 5(2), 166–177. <https://doi.org/10.29121/IJETMR.V5.I2.2018.641>
- Stojanov, A., & Daniel, B. K. (2024). A decade of research into the application of big data and analytics in higher education: A systematic review of the literature. *Education and Information Technologies*, 29(5), 5807–5831. <https://doi.org/10.1007/S10639-023-12033-8/FIGURES/3>
- Toring, H., Legaspi, G., Omolon, J., Amadeo, R., Amadeo, E., Opolentissima, Q., Barina, V., Cacho, T., Illustrimo, F., & Cortes, S. (2023). Evaluation of students' satisfaction toward an adopted learning management system at Indiana Aerospace University: A structural equation modelling approach. *Asia Pacific Management Review*, 28(3), 336–346. <https://doi.org/10.1016/J.APMRV.2022.12.002>
- Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., & Ragos, O. (2020). Fuzzy-based active learning for predicting student academic performance. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/3410352.3410823>
- Uzir, N. A. A., Gašević, D., Jovanovic, J., Matcha, W., Lim, L. A., & Fudge, A. (2020). Analytics of time management and learning strategies for effective online learning in blended environments. *ACM International Conference Proceeding Series*, 392–401. <https://doi.org/10.1145/3375462.3375493>
- Van Goidsenhoven, S., Bogdanova, D., Deeva, G., Vanden Broucke, S., De Weerd, J., & Snoeck, M. (2020). Predicting student success in a blended learning environment. *ACM International Conference Proceeding Series*, 17–25. <https://doi.org/10.1145/3375462.3375494>
- Veguru, H. S. K., Naren, J., & Singam, Y. (2024). Student's Interest and Opinion Towards Online Education. *Procedia Computer Science*, 233, 590–596. <https://doi.org/10.1016/J.PROCS.2024.03.248>
- Viberg, O., Khalil, M., & Baars, M. (2020). Self-regulated learning and learning analytics in online learning environments: A review of empirical research. *ACM International Conference Proceeding Series*, 524–533. <https://doi.org/10.1145/3375462.3375483>
- Wang, F. H. (2021). Interpreting log data through the lens of learning design: Second-order predictors and their relations with learning outcomes in flipped classrooms. *Computers & Education*, 168, 104209. <https://doi.org/10.1016/J.COMPE-DU.2021.104209>
- Wang, Q., & Mousavi, A. (2023). Which log variables significantly predict academic achievement? A systematic review and meta-analysis. *British Journal of Educational Technology*, 54(1), 142–191. <https://doi.org/10.1111/BJET.13282>

- Wong, J. T., & Hughes, B. S. (2023). Leveraging learning experience design: digital media approaches to influence motivational traits that support student learning behaviors in undergraduate online courses. *Journal of Computing in Higher Education*, 35(3), 595–632. <https://doi.org/10.1007/S12528-022-09342-1/FIGURES/4>
- Xu, X., Wang, J., Peng, H., & Wu, R. (2019). Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Computers in Human Behavior*, 98, 166–173. <https://doi.org/10.1016/J.CHB.2019.04.015>
- Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1), 1–19. <https://doi.org/10.1186/S40561-022-00192-Z/TABLES/14>
- Yang, Y., Hooshyar, D., Pedaste, M., Wang, M., Huang, Y. M., & Lim, H. (2020). Predicting course achievement of university students based on their procrastination behaviour on Moodle. *Soft Computing*, 24(24), 18777–18793. <https://doi.org/10.1007/S00500-020-05110-4/METRICS>
- Yoo, J. E., Rho, M., & Lee, Y. (2022b). Online Students' Learning Behaviors and Academic Success: An Analysis of LMS Log Data From Flipped Classrooms via Regularization. *IEEE Access*, 10, 10740–10753. <https://doi.org/10.1109/ACCESS.2022.3144625>
- Yürüm, O. R., Taşkaya-Temizel, T., & Yıldırım, S. (2023). The use of video clickstream data to predict university students' test performance: A comprehensive educational data mining approach. *Education and Information Technologies*, 28(5), 5209–5240. <https://doi.org/10.1007/S10639-022-11403-Y/FIGURES/4>
- Zabolotniaia, M., Cheng, Z., Dorozhkin, E. M., & Lyzhin, A. I. (2020). Use of the LMS Moodle for an effective implementation of an innovative policy in higher educational institutions. *International Journal of Emerging Technologies in Learning*, 15(13), 172–189. <https://doi.org/10.3991/IJET.V15I13.14945>
- Zhang, L., & Yu, X. (2022). Intelligent retrieval method of mobile learning resources in the intelligent higher education system. *International Journal of System Assurance Engineering and Management*, 13(6), 3079–3091. <https://doi.org/10.1007/S13198-021-01455-7/TABLES/2>
- Zhang, S., Zhou, Z., Zheng, W. X., & Tang, X. (2024). Frequency-domain Volterra kernel-based adaptation: Formulations and algorithms. *Signal Processing*, 216, 109295. <https://doi.org/10.1016/J.SIGPRO.2023.109295>

Cómo citar: Llivipuma Yáñez, L. D. ., & Ordonez Avila, E. R. (2025). Predicción de rendimiento del curso en línea en la educación superior con datos de LMS: Una revisión sistemática de literatura. *UNESUM - Ciencias. Revista Científica Multidisciplinaria*, 9(2), 202–219. <https://doi.org/10.47230/unesum-ciencias.v9.n2.2025.202-219>