593 Digital Publisher
CEIT
ISSN 2588-0705

Predicción del rendimiento académico utilizando modelos de aprendizaje automático: Una revisión sistemática de la literatura

Predicting academic performance using machine learning models: A systematic review of the literature

Ronaldo Andre Del Carpio-Mendoza [1]
Universidad Nacional Mayor de San Marcos - Perú
ronaldo.delcarpio@unmsm.edu.pe

doi.org/10.33386/593dp.2024.6.2797

V9-N6 (nov-dic) 2024, pp 1038-1054 | Recibido: 21 de septiembre del 2024 - Aceptado: 16 de octubre del 2024 (2 ronda rev.)

1 ORCID: https://orcid.org/0009-0006-6854-038X

Cómo citar este artículo en norma APA:A

Del Carpio-Mendoza, R., (2024). Predicción del rendimiento académico utilizando modelos de aprendizaje automático: Una revisión sistemática de la literatura. 593 Digital Publisher CEIT, 9(6), 1038-1054, https://doi.org/10.33386/593dp.2024.6.2797

Descargar para Mendeley y Zotero

RESUMEN

La predicción del rendimiento académico se ha convertido en un área de creciente interés en la educación superior, debido a su potencial para identificar y apoyar a estudiantes en riesgo antes de que enfrenten dificultades académicas. Este estudio se centra en la aplicación de modelos de aprendizaje automático para predecir el rendimiento académico, explorando diferentes variables y técnicas utilizadas en investigaciones recientes. A través de una revisión sistemática de la literatura, se analizaron estudios que emplean ML para predecir el éxito académico, identificando las variables, criterios, técnicas y las metodologías más efectivas. Los resultados destacan el impacto de variables como el historial académico, factores sociodemográficos, económicos y culturales en el rendimiento estudiantil, así como la eficacia de técnicas como las redes neuronales artificiales, los árboles de decisión y las máquinas de vectores de soporte. Finalmente, se discuten las implicaciones de estos hallazgos para el desarrollo de intervenciones educativas más eficientes y personalizadas.

Palabras claves: rendimiento académico, aprendizaje automático, minería de datos educativos.

ABSTRACT

Academic performance prediction has become an area of growing interest in higher education, due to its potential to identify and support at-risk students before they face academic difficulties. This study focuses on the application of machine learning models to predict academic performance, exploring different variables and techniques used in recent research. Through a systematic review of the literature, studies that use ML to predict academic success were analyzed, identifying the most effective variables, criteria, techniques and methodologies. The results highlight the impact of variables such as academic history, sociodemographic, economic and cultural factors on student performance, as well as the effectiveness of techniques such as artificial neural networks, decision trees and support vector machines. Finally, the implications of these findings for the development of more efficient and personalized educational interventions are discussed.

Keywords: academic performance, machine learning, educational data mining.

Introducción

La enseñanza en el ámbito universitario se encuentra enfrentándose constantemente a diferentes desafíos, principalmente debido a la falta de dominio que muestran los estudiantes en diferentes asignaturas presentes a lo largo de la carrera universitaria. El problema no solo afecta el rendimiento académico, sino que también dificulta la implementación de procesos educativos eficientes que incrementen el éxito de los estudiantes. Sin las competencias necesarias, los estudiantes se encuentran en una posición vulnerable, enfrentando mayores dificultades para alcanzar sus objetivos académicos y profesionales (Ahammad et al., 2021). Lograr predecir el desempeño de los estudiantes es un objetivo crucial en la educación, ya que permite la implementación de intervenciones estratégicas antes de que los estudiantes alcancen niveles avanzados de sus estudios. La capacidad de predicción temprana es vital para mejorar los resultados académicos y reducir las tasas de deserción. En este contexto, la minería de datos educativos y el aprendizaje automático se han convertido en herramientas esenciales, proporcionando patrones útiles y datos relevantes para tomar decisiones informadas (Lebkiri et al., 2021).

El avance de la inteligencia artificial, especialmente en el campo del Machine Learning (ML), está transformando la educación, adoptando tecnologías para detectar tempranamente a los estudiantes con dificultades y permitiendo intervenciones oportunas que pueden marcar una diferencia significativa en sus resultados académicos mediante el uso de algoritmos, los cuales evidencian su efectividad para predecir el rendimiento estudiantil y extraer patrones útiles a partir de datos crudos (Doctor, 2023). En ML nos enfocamos en el desarrollo de algoritmos que permiten a las computadoras aprender y hacer predicciones basadas en datos. Aplicando modelos de clasificación para segmentar a los estudiantes, se pueden implementar estrategias de retención personalizadas, optimizando los recursos educativos y mejorando la tasa de permanencia en las instituciones de educación superior (Aráuz & Martínez 2023).

Analizar grandes volúmenes de datos educativos para extraer patrones significativos y mejorar la toma de decisiones es un proceso llamado minería de datos educacionales, el cual constituye una prometedora solución para mejorar el rendimiento académico, sin embargo, el análisis de grandes cantidades de datos y la experimentación con diferentes métodos y técnicas para encontrar un modelo fiable requiere de una inversión considerable de tiempo y esfuerzo (Cruz et al., 2022). Investigaciones anteriores lograron identificar técnicas de ML e inteligencia artificial aplicadas en contextos educativos, destacando el impacto significativo que estas tecnologías pueden tener en la educación a diferentes niveles (Quijije & Maldonado 2023).

Una de las innovaciones de mayor significancia en el campo educativo es, la implementación de sistemas de alerta temprana. Estos sistemas, basados en técnicas de Deep Learning (DL), pueden realizar predicciones precisas sobre el rendimiento estudiantil, permitiendo intervenciones tempranas y utilizando un enfoque que demuestra ser eficaz en prevenir la deserción y mejorar los resultados académicos, especialmente cuando se utilizan algoritmos avanzados como las redes neuronales artificiales (ANN), (Forero & Bennasar, 2024). Existen diferentes factores que influyen en el desempeño académico, como el sexo del estudiante, el semestre de estudios, entre otros. El problema radica en la cantidad de variables que afectan el rendimiento académico, muchas de las cuales no son fácilmente accesibles (Yadav & Deshmukh, 2023). En este sentido, se pueden lograr mejores valores de sensibilidad, especificidad y precisión balanceada utilizado algoritmos de clasificación óptimos para mejorar las predicciones (Gamboa & Salinas, 2022).

Dentro de los modelos más utilizados, se encuentran la regresión cuantílica y la integración de máquinas de vectores de soporte (SVM), los cuales son eficaces en la predicción del rendimiento académico cuando se aplican adecuadamente (de Morais et al., 2021). Utilizando datos manejables, se mostró en una de las universidades más grandes de Croacia

593 Digital Publisher
CEIT
ISSN 2588-0705

que el modelo de árbol de decisión (DT) logra ser más preciso para el análisis del rendimiento estudiantil, habiendo considerado solamente 264 datos de estudiantes (Oreški & Zamuda, 2022). Es importante generar información fácilmente gestionable para la mejora de los procesos educativos, debido a la falta de inclusión de técnicas de minería de datos en prácticas educativas comunes (Sánchez & Mateos, 2023). Existe evidencia limitada sobre la enseñanza, el aprendizaje y la utilidad del ML en entornos educativos, sin embargo, permite beneficiar de forma colateral también a educadores, como fue el caso para docentes de género masculino, con experiencia previa en IA y en educación infantil, obteniendo mejores puntuaciones y valoraciones en sus entornos laborales (Cardozo et al., 2022).

Es posible, por ejemplo, detectar porcentajes de los estudiantes que finalmente terminan rezagándose. Después de observar previamente que una proporción significativa de estudiantes se estaba quedando atrás en la escuela primaria temprana, (Alhazmi, 2022) desarrolló un modelo de clasificación que detectó al 62% y 64% de estos estudiantes.

La repetición de al menos un grado dio lugar a una búsqueda de estudiantes en situación de riesgo. También se identificó un modelo que busca comprender las dificultades que enfrentan los estudiantes en su experiencia de aprendizaje a distancia, donde se utilizaron métodos tradicionales de recolección de datos como entrevistas y cuestionarios, si bien estos métodos tienden a ser poco confiables y tardados, los mismos lograron identificar problemas como la falta de interacción, dificultades técnicas, problemas de salud mental como falta de motivación, estrés y ansiedad. En otro estudio (Morales et al., 2022) realizado en Tlaxcala, se emplearon técnicas de aprendizaje automático para predecir el logro académico en las áreas de español y matemáticas en estudiantes.

El presente estudio tiene como objetivo, investigar los modelos predictivos de ML empleados para predecir el rendimiento académico. Nuestra hipótesis plantea que los modelos predictivos de ML pueden predecir el rendimiento académico de los estudiantes, por tanto, se formularon 4 preguntas de investigación que permitieron analizar cada uno de los trabajos y discutir su relevancia con el objeto de estudio.

Este paper se encuentra organizado de la siguiente forma: sección 2 describe la metodología utilizada para la investigación, sección 3 analiza los resultados obtenidos y, por último, en la sección 4 se muestran las conclusiones.

Método

Se utilizó la metodología desarrollada por (Yadav & Deshmukh, 2023), que consiste en 4 fases:

Alcance de la revisión: Se comprende el alcance considerado un proceso de búsqueda sistemática de artículos científicos.

Búsqueda y selección de estudios: Se realiza la búsqueda y selección de estudios, utilizando los criterios de inclusión y exclusión.

Reporte de la revisión: En esta parte se generan las estadísticas del análisis.

Síntesis de la información: Se ejecuta una exploración entre los estudios elegidos, brindando respuesta a las preguntas de investigación.

Alcance de la revisión

Teniendo como finalidad determinar los modelos de ML que logren predecir el rendimiento académico en estudiantes, se plantearon las siguientes 4 preguntas de investigación.

P1: ¿Cuáles son las variables utilizadas en aprendizaje automático más efectivas para predecir el rendimiento académico de los estudiantes?

P2: ¿Qué criterios se utilizan para seleccionar estudios relevantes en la predicción del rendimiento académico mediante aprendizaje automático?

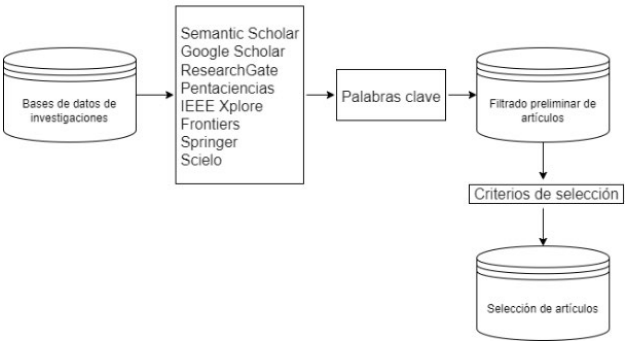
P3: ¿Qué técnicas de aprendizaje automático se han utilizado en la investigación

para predecir el rendimiento académico y cuáles son las más precisas?

P4: ¿Qué metodologías se han utilizado para desarrollar modelos de ML que lograron predecir el rendimiento académico de los estudiantes?

Para brindar respuesta a las preguntas planteadas, se realizó la búsqueda sistemática de la información en las bases de datos Google Scholar, ResearchGate, Springer, Frontiers, IEEE Xplore, Scielo, Semantic Scholar y Pentaciencias. El proceso integro es exhibido en la figura 1.

Figura 1
Proceso de revisión de artículos científicos



Se realizó la búsqueda de literatura de forma exhaustiva a través de las bases de datos de investigaciones. Continuando con el proceso, se seleccionaron trabajos a través de la implementación de los criterios de selección que son mostrados en la tabla 1.

Tabla 1
Criterios de selección

Inclusión	Exclusión
Los artículos deben contener variables, criterios, técnicas y metodologías para predecir el rendimiento académico utilizando ML	Los artículos que no guarden relación con las preguntas de investigación
Las investigaciones deben encontrarse desde el año 2020 hasta el año 2024	Las investigaciones que se encuentren fuera de las fechas indicadas
Los trabajos deben encontrarse publicados en revistas de investigación	Los trabajos que se encuentren en procesos de revisión

Búsqueda y selección de estudios

Para brindar respuesta a las preguntas planteadas, se realizó la búsqueda sistemática de la información en las bases de datos Google Scholar, ResearchGate, Springer, Frontiers, IEEE Xplore, Scielo, Semantic Scholar y Pentaciencias.

Se utilizaron los términos de búsqueda (“academic performance prediction”, “machine learning”, “educational data mining for academic performance”) desde el año 2020 hasta el año 2024.

Reporte de la revisión

Los estudios con una mayor representatividad inicialmente fueron seleccionados de las bases de datos de investigaciones ResearchGate, Google Scholar y IEEE XPLORE. Posteriormente se ejecutaron los criterios de selección, destacándose ResearchGate, Google Scholar y Springer como las bases de datos de investigaciones más representativas.

En la tabla 2, en relación al objeto de estudio, se muestran los 4230 artículos encontrados, en donde 189 de estos documentos fueron elegidos de forma preliminar y finalmente 54 artículos lograron ser seleccionados.

Tabla 2
Búsqueda de artículos

Fuente\Artículos	Encontrados	Preliminares	Seleccionados
ResearchGate	1919	80	24
Google Scholar	1746	39	11
IEEE Xplore	239	13	2
Springer	147	25	7
Semantic Scholar	143	18	2
Frontiers	18	2	1
Scielo	15	10	6
Pentaciencias	3	2	1

Síntesis de la información

Se brindará respuesta a las preguntas planteadas en esta sección, para esto analizamos los diferentes artículos de investigación.

593 Digital Publisher
CEIT
ISSN 2588-0705

Resultados

¿Cuáles son las variables utilizadas en aprendizaje automático más efectivas para predecir el rendimiento académico de los estudiantes?

Para predecir el rendimiento académico, las variables consideradas suelen derivarse de estudios sobre contextos educativos, familiares y demográficos. La investigación de Incio et al., (2023), ha revelado que factores como el nivel educativo de los padres, la etnia y variables demográficas influyen significativamente en el rendimiento estudiantil. Por ejemplo, este estudio fue realizado con estudiantes de Ingeniería Civil en el curso de Matemática II, encontrando que el nivel educativo de la madre y la etnia del estudiantado tienen un impacto notable en el rendimiento académico. Otro estudio en escuelas de Pennsylvania realizado por Chen & Ding (2023), utilizó datos educativos y demográficos, incluyendo la densidad de población y las tasas de criminalidad, para evaluar cómo estos factores afectan los resultados educativos.

También Cruz et al., (2022) y Alhazmi, (2022) presentaron una influencia significativa variables sociodemográficas en el rendimiento estudiantil. Por otra parte, analizando datos de 32,593 estudiantes en cursos universitarios en línea, el paper de Gil & Quintero (2023), destaca a las variables edad, género, región y discapacidad. Asimismo, en el Ashoka Women's Engineering College en India, el estudio de Sukanya et al., (2023) analiza variables como el área y población del condado, la densidad poblacional, y el porcentaje de estudiantes avanzados y rezagados.

El historial académico y las evaluaciones previas son robustos indicadores del rendimiento futuro y han sido un factor clave en múltiples estudios Ahammad et al., (2021), Lebkiri et al., (2021), Aráuz & Martínez, (2023), Quijije & Maldonado, (2023) y Gamboa & Salinas (2022). Un estudio de Said & Srinivasa, (2023) en Tanzania, utilizó variables como "average score" y "difference score" para predecir el rendimiento en matemáticas de estudiantes de secundaria. El análisis de Maqsood et al., (2023), basado en transcripciones de estudiantes en Ciencias de la Computación mostró variaciones en el desempeño de los cursos más desafiantes para cada grupo. Otro estudio, con 14,495 estudiantes, empleó el historial académico para predecir el abandono escolar (Gonzalez et al., 2023). Así también, las investigaciones de Morales et al., (2022) y Xie & Liu, (2023), implementaron técnicas estadísticas y de ML, en donde analizaron el rendimiento académico en cursos de ingeniería haciendo uso del historial académico.

La asistencia y la participación activa son de igual forma factores cruciales. El estudio de Zhao et al., (2023), subraya la importancia de la asistencia escolar y el tiempo libre posterior a la escuela, junto con las notas de los primeros periodos académicos. Los estudios de Selly & Anna, (2022) y Mushi & Ngondya, (2021), indican también que la ausencia escolar es un predictor clave y la participación en actividades extracurriculares también influye en el rendimiento educativo.

Variables psicológicas y emocionales, como la motivación y la actitud hacia el aprendizaje, por otra parte, son mencionadas por Lebkiri et al., (2021). La investigación de Pugosa et al., (2024), exhibe que la enseñanza heurística demuestra mejorar el desempeño y las actitudes hacia las matemáticas. Además, el estudio de Al & Ahmad, (2022), mostró una correlación significativa entre la autopercepción y el rendimiento académico.

Por otra parte, en estudios se menciona que la calidad educativa y los recursos disponibles son cruciales para el rendimiento académico, al igual que las condiciones socioeconómicas (Chen & Ding, 2023) y (Sukanya et al., 2023). Además, el comportamiento en clase, el entorno de aprendizaje y el programa de estudios se han mencionado como indicadores adicionales importantes para anticipar el éxito académico en Doctor (2023), Quijije & Maldonado, (2023), de Morais et al., (2021) y Sánchez & Mateos, (2023).

Se identifican las variables más efectivas utilizadas para predecir el rendimiento académico en la tabla 3, en donde se destacaron: el historial académico, los factores sociodemográficos, económicos y culturales (FSEC), en donde el género se encuentra considerado dentro de esta variable.

Tabla 3
Variables más efectivas para predecir el rendimiento académico

Variables de rendimiento	Referencias	Número de artículos
Historial de notas académicas	Ahammad et al., (2021), Lebkiri et al., (2021), Aráuz & Martínez (2023), Quijije & Maldonado (2023), Gamboa & Salinas (2022), Morales et al., (2022), Said & Srinivasa (2023), Maqsood et al., (2023), Gonzalez et al., (2023), Zhao et al., (2023), Selly & Anna (2022), Mushi & Ngondya (2021)	12
FSEC	Lebkiri et al., (2021), Cruz et al., (2022), Quijije & Maldonado (2023), Alhazmi (2022), Morales et al., (2022), Incio et al., (2023), Chen & Ding (2023), Sukanya et al., (2023), Zhao et al., (2023), Mushi & Ngondya (2021)	10
Comporta-miento en clase	Ahammad et al., (2021), Doctor (2023), Zhao et al., (2023), Selly & Anna (2022), Mushi & Ngondya (2021)	15
Entorno de aprendizaje	de Morais et al., (2021), Sánchez & Mateos (2023), Gil & Quintero (2023), Xie & Liu (2023)	4
Rendimiento	Ahammad et al., (2021), Aráuz & Martínez (2023), Gil & Quintero (2023), Said & Srinivasa (2023)	4
Factores psicológicos	Lebkiri et al., (2021), Pugosa et al., (2024), Al & Ahmad., (2022)	3
Edad	de Morais et al., (2021), Gil & Quintero (2023)	2
Actividades extracurricu-lares	Mushi & Ngondya (2021)	1
Discapacidad	Gil & Quintero, (2023)	1
Programa de estudios	Quijije & Maldonado, (2023)	1

¿Qué criterios se utilizan para seleccionar estudios relevantes en la predicción del rendimiento académico mediante aprendizaje automático?

Los criterios son importantes porque garantizan que los modelos de aprendizaje automático sean precisos, confiables y aplicables en contextos educativos. La privacidad y ética en el manejo de los datos junto con un análisis detallado del estado del arte son esenciales para contextualizar y validar los hallazgos de cada estudio (de Morais et al., 2021), (Sukanya et al., 2023). En la investigación realizada por Dúo et al., (2023), se evaluaron aspectos cruciales como la perspectiva y experiencia de los docentes en la implementación de proyectos basados en inteligencia artificial (IA) y ML. El trabajo de Oreški & Zamuda, (2022) se enfocó en la fiabilidad, precisión e interpretabilidad de los modelos desarrollados para la predicción del rendimiento académico durante el año 2020, 2021 y destacó la importancia de estos aspectos para asegurar que los modelos predictivos sean efectivos y comprensibles en el contexto educativo.

También son frecuentemente evaluadas características sociodemográficas como un criterio para entender su impacto en el rendimiento académico (Xie & Liu, 2023), (Zhao et al., 2023). En la investigación de Lebkiri et al., (2021), se realizó un análisis exhaustivo teniendo como criterios de selección a la fiabilidad interna de los elementos, el análisis factorial y la precisión de los modelos, reportando una precisión del 97% en la predicción de posibles fracasos académicos. El estudio de Li, (2024), utilizó hasta 30 indicadores para predecir el rendimiento académico, aplicando un modelo de regresión logística dividió el conjunto de datos en entrenamiento y prueba en una proporción de 7/3 y demostró la importancia de la amplitud y diversidad de las variables utilizadas en la predicción. En el estudio de Hussain & Jr, (2023), se evaluó el desempeño como precisión, recuperación, puntuación F1 y exactitud en la minería de datos educativos, analizando el rendimiento de varios clasificadores, incluyendo SVM, Naive Bayes (NB), DT y ANN, en la predicción del rendimiento académico, proporcionando un criterio de evaluación de los métodos de clasificación. El trabajo de Yan (2022) utilizó el criterio de selección de modelos como la regresión lineal y random forest (RF) para analizar 395 conjuntos de datos con 32 factores relacionados con el rendimiento académico, incluyendo variables como el rendimiento

593 Digital Publisher
CEIT
ISSN 2588-0705

histórico de los estudiantes, su comportamiento y la relación entre ambas.

La validez de los modelos utilizados es uno de los aspectos más importantes, asegurando la precisión y confiabilidad de las predicciones, siendo considerados en diversas investigaciones (Ahammad et al., 2021), (Aráuz & Martínez, 2023), (Quijije & Maldonado, 2023) y (Maqsood et al., 2023).

Otros estudios destacan los criterios que se enfocan en la precisión, interpretabilidad, y relevancia de las variables en los modelos predictivos. En el estudio Guanín et al., (2024), se priorizó la precisión de los algoritmos, el entendimiento de los modelos y la significancia estadística de los resultados. Este estudio destacó la importancia de identificar tempranamente a los estudiantes en riesgo y la implementación de estrategias de intervención efectivas. El estudio de Bravo et al., (2022) se realizó utilizando métodos supervisados de aprendizaje automático para predecir el rendimiento académico universitario, utilizando criterios de relevancia de las variables predictoras, la calidad de los datos, la precisión del modelo y la validación cruzada. La calidad de los datos es mencionada también como un criterio relevante en (Doctor, 2023), (Sánchez & Mateos, 2023) y (Mushi & Ngondya, 2021), mientras que la relevancia de los factores de éxito estuvo presente en la investigación de Incio et al., (2023).

Tabla 4
Criterios para seleccionar estudios relevantes

Variables de rendimiento	Referencias	Número de artículos
Validez de modelos utilizados	(Ahammad et al., 2021), (Lebkiri et al., 2021), (Aráuz & Martínez 2023), (Quijije & Maldonado 2023), (de Morais et al., 2021), (Oreški & Zamuda 2022), (Maqsood et al., 2023), (Dúo et al., 2023), (Hussain & Jr 2023), (Guanín et al., 2024)	10
Calidad de datos	(Ahammad et al., 2021), (Lebkiri et al., 2021), (Doctor 2023), (Aráuz & Martínez 2023), (Sánchez & Mateos 2023), (Mushi & Ngondya 2021), (Yan 2022), (Bravo et al., 2022)	8
Relevancia y factores de éxito	(Ahammad et al., 2021), (Lebkiri et al., 2021), (Incio et al., 2023), (Sukanya et al., 2023)	4
Metodología del studio	(Doctor 2023), (Aráuz & Martínez 2023), (Sánchez & Mateos 2023), (Bravo et al., 2022)	4
Características Sociodemográficas	(Xie & Liu 2023), (Zhao et al., 2023), (Li 2024)	3
Privacidad y ética	(de Morais et al., 2021)	1
Estado del arte	(Sukanya et al., 2023)	1

¿Qué técnicas de aprendizaje automático se han utilizado en la investigación para predecir el rendimiento académico y cuáles son las más precisas?

Existen diversas técnicas de aprendizaje automático utilizadas para predecir el rendimiento académico con alta precisión. En el estudio de Cruz et al., (2022), se han utilizado técnicas de ML y DL para predecir el rendimiento académico y la deserción estudiantil, mostrando altos niveles de precisión en la predicción. En el estudio de Ahammad et al., (2021), realizado en 2021 en Bangladesh, se aplicaron diferentes técnicas, destacan NB, K-Nearest Neighbours (K-NN), SVM, XGBoost (XGB) y Multi-layer Perceptron (MLP), siendo MLP la más precisa con un 86.25% de precisión. En el paper de Mulyana et al., (2023), se utilizó el algoritmo de RF para clasificar el rendimiento académico de estudiantes, empleando un dataset de un sistema de gestión del aprendizaje llamado Kalboard 360, demostrando una precisión del 89%, mostrando ser efectivo para manejar grandes conjuntos de datos y clasificar el rendimiento académico. En otro estudio realizado en Vietnam durante

los años académicos 2013 a 2016 por Dinh et al., (2020), se recopilaron datos de encuestas y bases académicas de estudiantes, aplicando técnicas como NB y MLP, que mostraron la mejor precisión en la predicción del rendimiento académico.

Diversos estudios utilizaron técnicas de aprendizaje automático y algoritmos avanzados para predecir el rendimiento académico, destacando la alta precisión de estos métodos. En marzo de 2023, se realizó una investigación en la que se utilizaron técnicas de aprendizaje automático como RF, Boosting y clasificadores base para predecir el rendimiento académico de estudiantes (Olukoya, 2023). La técnica más precisa fue REP Tree, con una precisión del 83.33%. Además, en 2024 el estudio de Thorat, (2024) analizó la aplicación de conceptos matemáticos como álgebra lineal y cálculo en el aprendizaje automático, destacando la precisión y relevancia de estas técnicas en la ciencia de datos moderna. En la investigación de Shi (2024), se utilizaron técnicas avanzadas como Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) y Convolutional Recurrent Neural Network (CRNN) para predecir el rendimiento académico con resultados prometedores, empleando tecnologías como Optical Character Recognition y Connectionist Text Proposal Network para la minería de caracteres en entornos educativos en 2024.

Otro estudio de Bellaj et al., (2024), aplicó técnicas como SVM, RF, LR, XGB, EVC, DT, K-NN y NB, siendo los modelos más precisos EVC y XGB. En (Abdu, 2024), se investigaron datos de la Universidad de Wollo entre 2017 y 2022, comparando diferentes algoritmos de aprendizaje automático como SVM, DT y NB, siendo SVM el que demostró la mayor precisión con un 96% de exactitud. En otro caso, Yağcı, (2022) utilizó algoritmos como RF y regresión logística para predecir calificaciones finales en un curso de Lengua Turca-I, logrando una precisión entre el 70 al 75% con RF, ANN y SVM mostrando las tasas de precisión más altas.

Otros estudios analizaron la relevancia de conceptos matemáticos y detectaron patrones de participación estudiantil utilizando algoritmos específicos. En un estudio que empleó el algoritmo PELT (Nakamura et al., 2024), se lograron altas tasas de precisión, recall y F1-score, al identificar cambios significativos en las tasas de participación estudiantil durante las clases. Este enfoque se basó en datos educativos de lectores de libros electrónicos y comportamientos de docentes. Además, en (Salles et al., 2020), se aplicaron DT, RF y algoritmos de clustering para predecir el rendimiento académico en evaluaciones matemáticas interactivas en Francia.

Las técnicas de aprendizaje automático como regresión logística, ANN, y minería de patrones suelen ser útiles para predecir la deserción y el rendimiento académico. En (Aco et al., 2023), se utilizó regresión logística, NB, Red Neuronal Perceptrón Multicapa, DT, SVM y RF, destacando la Regresión Logística como la más precisa para predecir la deserción universitaria en un dataset analizado en 2023. En otro estudio (Hoyos & Daza, 2023), realizado entre los años 2017 y 2019, se aplicaron técnicas como regresión logística para predecir la deserción estudiantil, con una sensibilidad del modelo del 61.97%, permitiendo la identificación temprana de estudiantes en riesgo. En (Contreas et al., 2023), un estudio de noviembre de 2022 utilizó aprendizaje automático supervisado para predecir el rendimiento académico a nivel universitario. Las técnicas más precisas fueron las ANN y los modelos de clasificación. En la investigación de (Czibula et al., 2022), se emplearon técnicas como regresión Tweedie, SGD y Poly en el marco IntelliDaM, logrando alta precisión en la predicción del rendimiento académico en un curso de Ciencias de la Computación en la Universidad Babeş-Bolyai en Rumania. Finalmente, (Yu & Wen, 2020) destaca el uso de Sequential Pattern Mining (SPM) para predecir acciones de estudiantes en problemas STEM, donde las predicciones de máquinas de vectores de soporte fueron más precisas que las de expertos, mostrando el potencial del aprendizaje automático en educación STEM.

593 Digital Publisher
CEIT
ISSN 2588-0705

Diversas investigaciones mencionan técnicas usualmente implementadas como DT en (Doctor, 2023), (Yadav & Deshmukh, 2023), (de Morais et al., 2021), (Oreški & Zamuda, 2022), (Cardozo et al., 2022), SVM en (Lebkiri et al., 2021), (Aráuz & Martínez, 2023), (Gamboa & Salinas 2022), (Chen & Ding, 2023), ANN en (Incio et al., 2023), (Sukanya et al., 2023). Técnicas de ensamble como RF y GB en (Gil & Quintero 2023), (Maqsood et al., 2023), (Mushi & Ngondya, 2021), técnicas de regresión como la regresión logística en (Li, 2024) y regresión lineal en (Selly & Anna, 2022). Otras técnicas como NB y K-NN en (Hussain & Jr, 2023), (Gil & Quintero, 2023), clustering en (Sánchez & Mateos, 2023), (Maqsood et al., 2023) y de aprendizaje supervisado para regresión y clasificación como es XGB en (Zhao et al., 2023).

Ver tabla 5.

¿Qué metodologías se han utilizado para desarrollar modelos de machine **learning que lograron predecir el rendimiento académico de los estudiantes?**

En varios estudios, se han empleado metodologías enfocadas en la selección de variables más relevantes, mejorar la precisión de los modelos y asegurar que los resultados sean interpretables y útiles para la toma de decisiones. En (Gamboa & Salinas, 2022), se utilizó la metodología Cross-Industry Standard Process for Data Mining (CRISP-DM) en 2022, la cual permitió desarrollar modelos predictivos utilizando algoritmos como Regresión Logística, NB y SVM con kernel lineal. Se implementó también un ensamble con un punto de corte óptimo para mejorar las predicciones. Esta metodología, que consta de seis fases: entendimiento del negocio, entendimiento de los datos, preparación de los datos, modelación, evaluación y despliegue de resultados, es mostrada en la figura 2. En (Doctor, 2023), se aplicó la metodología CRISP-DM y técnicas de minería de datos para desarrollar modelos de ML que predicen el rendimiento académico de estudiantes utilizando datos de un sistema de información académica integrado.

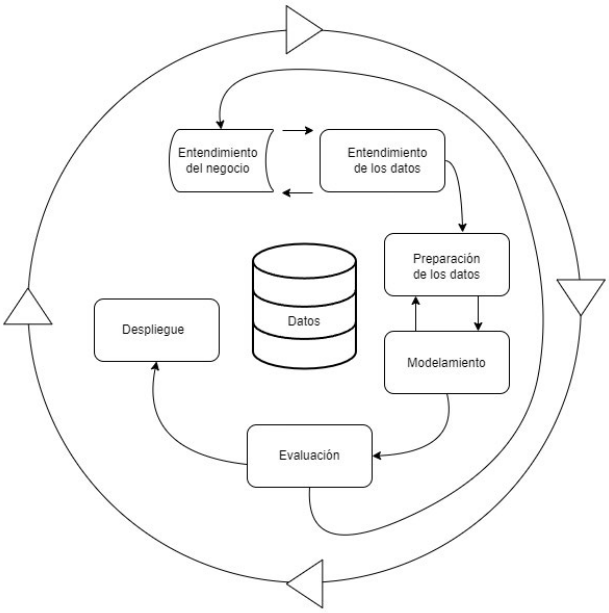
Tabla 5
Técnicas de aprendizaje automático

Variables de rendimiento	Referencias	Número de artículos
DT	Doctor (2023), Yadav & Deshmukh (2023), de Morais et al., (2021), Oreški & Zamuda (2022), Cardozo et al., (2022), Sukanya et al., (2023), Selly & Anna (2022), Mushi & Ngondya (2021), Al & Ahmad., (2022), Hussain & Jr (2023), Guanín et al., (2024), Bravo et al., (2022), Mulyana et al., (2023), Dinh et al., (2020), Shi (2024), Abdu (2024), Salles et al., (2020), Aco et al., (2023), Contreas et al., (2023)	19
SVM	Ahammad et al., (2021), Lebkiri et al., (2021), Aráuz & Martínez (2023), Cruz et al., (2022), Gamboa & Salinas (2022), Oreški & Zamuda (2022), Chen & Ding (2023), Sukanya et al., (2023), Al & Ahmad., (2022), Hussain & Jr (2023), Shi (2024), Bellaj et al., (2024), Abdu (2024), Yağcı (2022), Aco et al., (2023)	15
ANN	Doctor (2023), Aráuz & Martínez (2023), Cruz et al., (2022), Yadav & Deshmukh (2023), Oreški & Zamuda (2022), Incio et al., (2023), Chen & Ding (2023), Sukanya et al., (2023), Selly & Anna (2022), Al & Ahmad., (2022), Hussain & Jr (2023), Bravo et al., (2022), Dinh et al., (2020), Contreas et al., (2023)	14
RF, GB	Lebkiri et al., (2021), Chen & Ding (2023), Gil & Quintero (2023), Sukanya et al., (2023), Maqsood et al., (2023), Mushi & Ngondya (2021), Al & Ahmad., (2022), Yan (2022), Guanín et al., (2024), Mulyana et al., (2023), Olukoya (2023), Yağcı (2022), Salles et al., (2020), Aco et al., (2023)	14
Regresión logística	Lebkiri et al., (2021), Doctor (2023), Gamboa & Salinas (2022), de Morais et al., (2021), Chen & Ding (2023), Li (2024), Bravo et al., (2022), Olukoya (2023), Yağcı (2022), Aco et al., (2023), Contreas et al., (2023)	11
NB	Ahammad et al., (2021), Cruz et al., (2022), Yadav & Deshmukh (2023), Gamboa & Salinas (2022), Gil & Quintero (2023), Hussain & Jr (2023), Dinh et al., (2020), Abdu (2024), Aco et al., (2023)	9
K-NN	Ahammad et al., (2021), Gil & Quintero (2023), Mushi & Ngondya (2021), Bellaj et al., (2024), Abdu (2024), Yağcı (2022)	6
Clustering	Aráuz & Martínez (2023), Sánchez & Mateos (2023), Maqsood et al., (2023), Hussain & Jr (2023), Salles et al., (2020)	5
Regresión lineal	Aráuz & Martínez (2023), Selly & Anna (2022), Yan (2022), Hoyos & (Daza 2023)	4
XGB	Ahammad et al., (2021), Zhao et al., (2023), Guanín et al., (2024), Bellaj et al., (2024)	4
Variaciones de modelos	Quijije & Maldonado (2023), Said & Srinivasa (2023), Al & Ahmad., (2022), Yu & Wen (2020)	4
DL	Mulyana et al., (2023), Thorat (2024)	2
MLP	Ahammad et al., (2021)	1

En 2020, se propuso una adaptación del modelo CRISP-DM al contexto de datos educativos, denominada CRISP-EDM (Ramos et al.,2020). El proceso constituido con la metodología CRISP-DM permitió un análisis riguroso y sistemático de los datos educativos en diversas investigaciones (Lebkiri et al., 2021),

(Doctor 2023), (Quijije & Maldonado, 2023), (Gamboa & Salinas, 2022), (Guanín et al., 2024), (Bellaj et al., 2024).

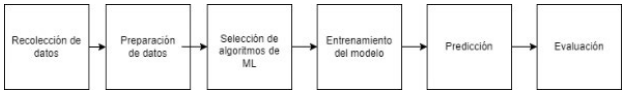
Figura 2
Metodología CRISP-DM



Además, en marzo de 2023, se realizó un análisis sobre el papel de las matemáticas en el ML (Patil et al., 2023), enfocándose en conceptos como álgebra lineal, estadísticas, cálculo y probabilidad, los cuales son fundamentales para construir modelos precisos con mínimos errores.

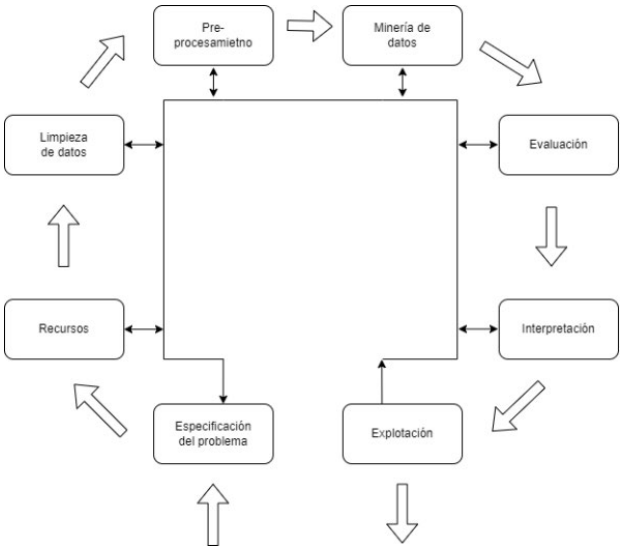
Otros métodos son basados en el análisis de datos que se enfocan en la recopilación, el análisis de datos académicos, análisis de comportamiento, fueron utilizados junto con técnicas estadísticas y algoritmos de ML en (Ahammad et al., 2021), (Aráuz & Martínez, 2023), (Maqsood et al., 2023), (Dúo et al., 2023), (Yu & Wen, 2020). En la figura 3, se muestra un patrón metodológico utilizado comúnmente por diversos estudios para procesos de ML. En el estudio (Martins et al., 2023), se han utilizado diversas metodologías como los algoritmos ANN, CNN y RNN para predecir el rendimiento académico de estudiantes, este estudio se basó en la metodología PRISMA para realizar la revisión sistemática de la literatura. La metodología fue utilizada para desarrollar modelos de ML en el ámbito educativo. Además, entre 2015 y 2021, se examinaron tendencias recientes en minería de datos educativos para predecir el rendimiento académico de estudiantes (Roslan & Chen., 2022), revisando 58 de 219 artículos de investigación de las bases de datos Lens y Scopus. Este estudio se centró en factores que influyen en el rendimiento estudiantil, como registros académicos, demografía y algoritmos de clasificación como DT, destacando la importancia de las intervenciones tempranas para mejorar el rendimiento académico.

Figura 3
Otros modelos metodológicos para procesos de ML



En (Al et al., 2023), se emplearon algoritmos de ML supervisado para analizar factores que afectan el rendimiento académico de estudiantes universitarios en situación de riesgo académico. Se utilizaron métodos de conjunto como Logit Boost, Bagging y el algoritmo J48, que fue el más eficaz con una precisión del 82.4%. Este estudio implemento la metodología Knowledge Discovery In Data Bases (KDD), cuyos subprocesos son mostrados en la figura 4, destacó también la importancia de identificar factores significativos para mejorar las políticas educativas y reducir el fracaso académico.

Figura 4
Metodología KDD



Digital Publisher
CEIT
ISSN 2588-0705

Tabla 6
Metodologías para desarrollar modelos predictivos del rendimiento académico

Variables de rendimiento	Referencias	Número de artículos
CRISP-DM	Lebkiri et al., (2021), Doctor (2023), Quijije & Maldonado (2023), Gamboa & Salinas (2022), Guanín et al., (2024), Bellaj et al., (2024), Doctor (2023), Ramos et al., (2020)	8
Otros modelos metodológicos para procesos de ML	Ahammad et al., (2021), Aráuz & Martínez (2023), Maqsood et al., (2023), Dúo et al., (2023), Yu & Wen (2020), Martins et al., (2023), Roslan & Chen., (2022)	7
Metodología estadística	Gamboa & Salinas (2022), Patil et al., (2023)	2
KKD	Al et al., (2023)	1

Discusión

En la figura 5, encontramos la distribución de Bases de datos de los 54 artículos científicos seleccionados, teniendo ResearchGate una representatividad del 44%, Google Scholar del 20%, Springer del 13%, Pentaciencias del 2%, Scielo del 11%, Frontiers del 2%, IEEE Xplore del 4% y Semantic Scholar del 4%.

Se exhibe en la figura 6, los 4230 trabajos encontrados en relación al objeto de estudio, siendo seleccionados desde las bases de datos de artículos científicos de forma preliminar, 80 de ReseachGate, 39 de Google Scholar, 13 de IEEE Xplore, 25 de Springer, 18 de Semantic Scholar, 2 de Frontiers, 10 de Scielo y 2 de Pentaciencias. Posteriormente fueron seleccionados 54 artículos, de los cuales 24 fueron de ResearchGate, 11 de Google Scholar, 2 de IEEE Xplore, 7 de Springer, 2 de Semantic Scholar, 1 de Frontiers, 6 de Scielo y 1 de Pentaciencias.

Las 10 variables consideradas en diversas investigaciones para predecir el rendimiento académico utilizando modelos de ML son mostradas en la figura 7. En donde se destacan como variables más efectivas al historial académico con una representatividad del 28% y a FSEC con una del 23%.

En la figura 8, se observan los 7 criterios más utilizados que sirven para seleccionar estudios relevantes en la predicción del rendimiento académico mediante aprendizaje automático, de los cuales la validez de modelos destaca con un 32% de representatividad y la calidad de datos con un 26%.

En la figura 9, son presentadas 13 técnicas de aprendizaje automático utilizadas para predecir el rendimiento académico. Las técnicas que tuvieron mayor representatividad fueron DT con un 17%, SVM con un 14%, ANN con un 13% y los algoritmos de ensemble RF, GB con un 13%.

Se aprecian las metodologías más utilizadas para predecir el rendimiento académico utilizando ML en la figura 10. Constituyendo un 44% de representatividad CRISP-DM, un 39% otros modelos metodológicos para procesos de ML, 11% metodologías estadísticas y 6% KDD.

Figura 5
Bases de datos de artículos científicos seleccionados

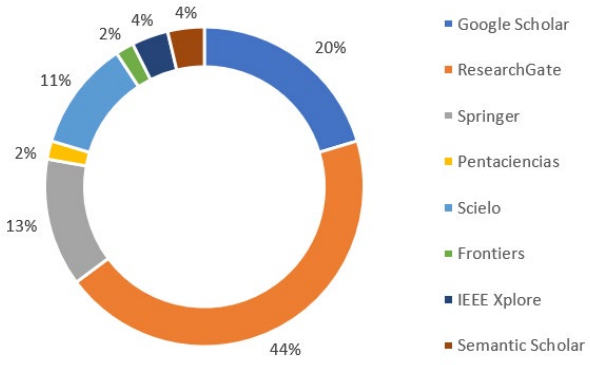


Figura 6
Artículos encontrados, preliminares y seleccionados

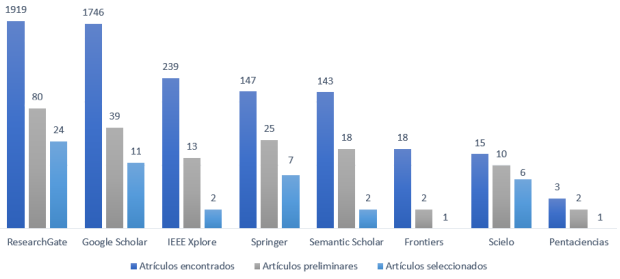


Figura 7
Variables más efectivas para predecir el rendimiento académico

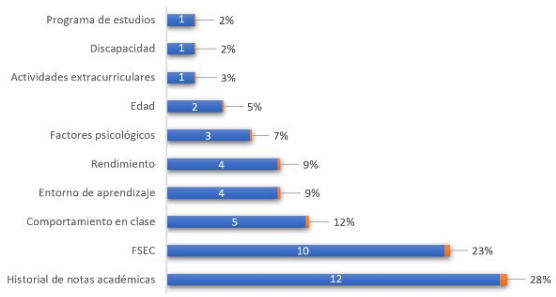


Figura 8
Criterios para seleccionar estudios relevantes

Figura 9
Técnicas de aprendizaje automático

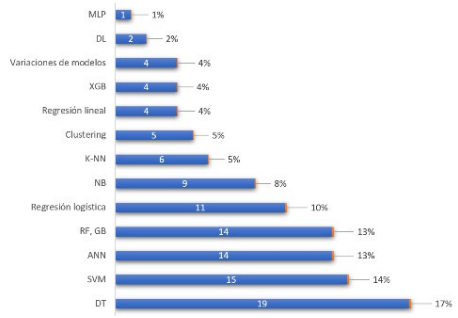
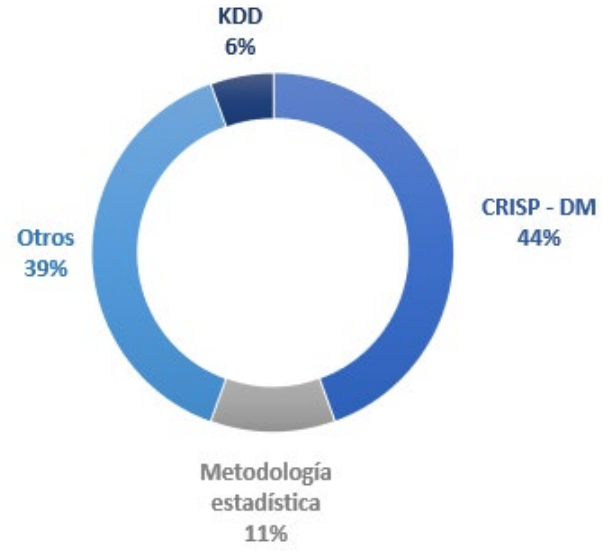


Figura 10
Criterios para seleccionar estudios relevantes



Conclusiones

El presente estudio desarrolla una revisión sistemática de la literatura en relación a la predicción del rendimiento académico en matemáticas. Fueron encontrados 4230 artículos relacionados en las diferentes bases de datos de investigaciones, siendo seleccionados 54 artículos habiendo ejecutado los criterios de selección.

Los artículos seleccionados revelan que el uso de modelos de aprendizaje automático para predecir el rendimiento académico constituye una estrategia efectiva para identificar a estudiantes en riesgo de manera temprana. Las variables más determinantes incluyen el historial académico y los factores sociodemográficos, económicos y culturales. La validez de los modelos predictivos y la calidad de los datos son cruciales para obtener predicciones precisas. Entre las técnicas de ML, DT, SVM y ANN destacan por su alta precisión en la predicción del rendimiento académico. Estos modelos no solo permiten anticipar problemas potenciales, sino también mejorar la toma de decisiones en el ámbito académico. Los hallazgos indican la importancia de personalizar las estrategias educativas en función de los perfiles individuales de los estudiantes, utilizando las predicciones de ML para diseñar planes de estudio y apoyo que

593 Digital Publisher CEIT ISSN 2588-0705

respondan a las necesidades específicas de cada alumno.

Con una predicción más precisa del rendimiento académico, las instituciones pueden dirigir sus recursos, como tutorías, programas de mentoría y asistencia financiera, a los estudiantes que más lo necesitan, maximizando el impacto de estos recursos.

Para futuras investigaciones, se recomienda realizar estudios que exploren la integración de factores como salud mental, entornos educativos y calidad docente.

Referencias bibliográficas

Abdu, E. (2024). Student Performance Prediction Using Machine Learning Algorithms. *Applied Computational Intelligence and Soft Computing*, 2024, https://doi.org/10.1155/2024/4067721

Aco, A., Hancco, B., Pérez, Yasiel. (2023). Análisis comparativo de Técnicas de Machine Learning para la predicción de casos de deserción universitaria. *RISTI-Revista Ibérica de Sistemas e Tecnologias de Informação*, 51(1), 84-98. https://doi.org/10.17013/risti.51.84-98

Ahammad, K., Chakraborty, P., Akter, E., et al. (2021). A Comparative Study of Different Machine Learning Techniques to Predict the Result of an Individual Student Using Previous Performances. *International Journal of Computer Science and Information Security (IJCSIS),* 19(1), 5-10. https://doi.org/10.5281/ZENODO.4533373

Al, L., Al, J., Tarhini, A., et al. (2023). Using machine learning to predict factors affecting academic performance: the case of college students on academic probation. *Education and Information Technologies*, 28(10), 12407-12432. https://doi.org/10.1007/s10639-023-11700-0

Al, Y., & Ahmad, N. (2022). Prediction methods on students academic performance: a review. *Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition),* 41(1), 196-217. https://doi.org/10.17605/OSF.IO/CHJF2

Alhazmi, H. (2022). Detection of students' problems in distance education using topic modeling and machine learning. *Future Internet*, 14(6), 170. https://doi.org/10.3390/fi14060170

Aráuz, D., & Martínez, J. (2023). Predicción del rendimiento académico en la UNADECA por medio de sistemas de clasificación. UNACIENCIA: *Revista de Estudios e Investigaciones*, 16(31), 17-35. https://doi.org/10.35997/unaciencia.v16i31.738

Bellaj, M., & Bendahmane, A., & Boudra, S., et al. (2024). Educational Data Mining: Employing Machine Learning Techniques and Hyperparameter Optimization to Improve Students Academic Performance. *International Journal of Online and Biomedical Engineering (iJOE)*, 20(3), 55-74. https://doi.org/10.3991/ijoe.v20i03.46287

Bravo, L., & Nieves, N., & Gonzalez, K. (2022). Prediction of University-Level Academic Performance through Machine Learning Mechanisms and Supervised Methods. *Ingeniería*, 28(1), e19514. https://doi.org/10.14483/23448393.19514

Cardozo, S., Silveira, A., & Fonseca, B. (2022). Detección temprana del riesgo escolar. Predicción de trayectorias de rezago en la educación primaria en Uruguay mediante técnicas de machine learning. *Revista latinoamericana de estudios educativos*, 52(2), 297-326. https://doi.org/10.48102/rlee.2022.52.2.391

Chen, S, & Ding, Y. (2023). A Machine Learning Approach to Predicting Academic Performance in Pennsylvania's Schools. *Social Sciences*, 12(3), 118. https://doi.org/10.3390/socsci12030118

Contreas, L., Nieves, N., & González, G., et al. (2023). Prediction of University-Level Academic Performance through Machine Learning Mechanisms and Supervised Meth-

ods. *Ingeniería*, 28(1), e19514. https://doi.org/10.14483/23448393.19514

Cruz, E., González, M., & Rangel, J. (2022). Técnicas de machine learning aplicadas a la evaluación del rendimiento ya la predicción de la deserción de estudiantes universitarios, una revisión. *Prisma Tecnológico*, 13(1), 77-87. https://doi.org/10.33412/pri.v13.1.3039

Czibula, G., Ciubotariu, G., Maier, M. I., et al. (2022). IntelliDaM: A machine learning-based framework for enhancing the performance of decision-making processes. A case study for educational data mining. *IEEE Access*, 10(1), 80651-80666. https://doi.org/10.1109/ACCESS.2022.3195531

de Morais, F, Melo, A., Moutinho, M., et al. (2021). Modelos de regressão aplicados na previsão da evasão escolar do ensino básico: uma revisão sistemática da literatura. *Anais do XXXII Simpósio Brasileiro de Informática na Educação*, 168-178. https://doi.org/10.5753/sbie.2021.218504

Dinh, H., & Cu, G., & Pham, T. (2020). An Empirical Study for Student Academic Performance Prediction Using Machine Learning Techniques. *International Journal of Computer Science and Information Security*, 20(20).

Doctor, A. (2023). A predictive model using machine learning algorithm in identifying students probability on passing semestral course. *International Journal of Computing Sciences Research*, 7(1), 1830-1856. https://doi.org/10.25147/ijcsr.2017.001.1.135

Dúo, P., Moreno, A., López, J., et al. (2023). Inteligencia Artificial y Machine Learning como recurso educativo desde la perspectiva de docentes en distintas etapas educativas no universitarias. *RiiTE Revista interuniversitaria de investigación en Tecnología Educativa*, 58-78. https://doi.org/10.6018/riite.579611

Forero, W., & Bennasar, F. (2024). Techniques and applications of Machine Learning and Artificial Intelligence in education: a systematic review. *RIED-Revista Iberoamericana de Educación a Distancia*, 27(1), 209-253. https://doi.org/10.5944/ried.27.1.37491

Gamboa, J., & Salinas, J. (2022). Predicción de la situación académica en alumnos de pregrado usando algoritmos de Machine Learning. *Perfiles*, 1(27), 4-10. https://doi.org/10.47187/perf.v1i27.142

Gil, V., & Quintero, C. (2023). Análisis de variables asociadas al rendimiento académico en cursos universitarios virtuales. *Formación universitaria*, 16(4), 33-42. https://doi.org/10.4067/s0718-50062023000400033

Gonzalez, A., Noguez, J., Neri, L., et al. (2023). Predictive analytics study to determine undergraduate students at risk of dropout. *Frontiers in Education*, 8(1). https://doi.org/10.3389/feduc.2023.1244686

Guanín, J., & Guaña, J., & Casillas, J. (2024). Predicting Academic Success of College Students Using Machine Learning Techniques. *Data*, 9(4), 60. https://doi.org/10.3390/data9040060

Hoyos, J., & Daza, G. (2023). Predictive Model to Identify College Students with High Dropout Rates. *Revista electrónica de investigación educativa*, 25(13). https://doi.org/10.24320/redie.2023.25.e13.5398

Hussain, S., & Jr, I. (2023). Significance of Education Data Mining in Student's Academic Performance Prediction and Analysis. *International Journal of Innovations in Science & Technology*, 5(1), 215-231.

Incio, F., Capuñay, D., & Estela, R. (2023). Modelo de red neuronal artificial para predecir resultados académicos en la asignatura Matemática II. *Revista Electrónica Educare*, 27(1), 338-359. https://doi.org/10.15359/ree.27-1.14516

Lebkiri, N., Daoudi, M., Abidli, Z., et al. (2021). Using machine learning for prediction students failure in Morocco: an application of the CRISP-DM methodology. *Int. J. Educ. Inf.*

593 Digital Publisher CEIT ISSN 2588-0705

Technol, 15(1), 344-352. https://doi.org/10.46300/9109.2021.15.36

Li, Y. (2024). Data Analysis of Student Academic Performance and Prediction of Student Academic Performance Based on Machine Learning Algorithms. *Communications in Humanities Research*, 32(1), 65-71. https://doi.org/10.54254/2753-7064/32/20240013

Maqsood, R., Ceravolo, P., Ahmad, M. et al. (2023). Examining students' course trajectories using data mining and visualization approaches. *Int J Educ Technol High Educ*, 20(55). https://doi.org/10.1186/s41239-023-00423-4

Martins, L, dos Santos, V., de Oliveira, A., et al. (2023). Revisão Sistemática sobre Machine Learning Aplicada a Bioacústica utilizando o Método PRISMA. *Anais da XII Escola Regional de Informática de Mato Grosso*, 251-255. https://doi.org/10.5753/eri-mt.2023.236622

Morales, M., González, J., Robles, H., et al. (2022). Algoritmos de aprendizaje automático para la predicción del logro académico. *Revista Iberoamericana para la Investigación y el Desarrollo Educativo (RIDE)*, 12(24), 35. https://doi.org/10.23913/ride.v12i24.1180

Mulyana, A., & Puspita, W., & Unjung, J. (2023). Increased accuracy in predicting student academic performance using random forest classifier. *Journal of Student Research Exploration*, 1, 94-103. https://doi.org/10.52465/josre.v1i2.169

Mushi, P., & Ngondya, D. (2021). Prediction of mathematics performance using educational data mining techniques. *International Journal of Advanced Computer Research*, 11(1), 83-102. https://doi.org/10.19101/IJACR.2021.1152024

Nakamura, K., Ishihara, M., Horikoshi, I., et al. (2024). Uncovering insights from big data: change point detection of classroom engagement. *Smart Learn. Environ*, 11(31). https://doi.org/10.1186/s40561-024-00317-6

Olukoya, B. (2023). Using ensemble random forest, boosting and base classifiers to ameliorate prediction of students academic performance. *International Journal of Advance Research, Ideas, and Innovations in Technology*. 6(1), 654.

Oreški, D., & Zamuda, D. (2022). Machine Learning Based Model for Predicting Student Outcomes. *In 12th International Conference on Industrial Engineering and Operations Management (IEOM 2022)*, 4884-4894. https://doi.org/10.46254/AN12.20220967

Patil, M., Jadhav, S., Talekar, S., et al. (2023). The role of mathematics in machine learning. *Journal of Data Acquisition and Processing*, 3(1), 1062-1073. https://doi.org/10.5281/zenodo.7702430

Pugosa, C., & Yumol, C., & Nogadas, C., et al. (2024). Effects of Heuristic Method on Students' Performance in Mathematics. *British Journal of Teacher Education and Pedagogy*, 3(1), 69-86. https://doi.org/10.32996/bjtep.2024.3.2.8

Quijije, H., & Maldonado Zuñiga, K. (2023). Técnica de minería de datos para procesos educativos en estudiantes con necesidades educativas especiales basado en un modelo predictivo. *Revista Científica Arbitrada Multidisciplinaria PENTACIENCIAS*, 5(5), 205–217. https://doi.org/10.59169/pentaciencias.v5i5.730

Ramos, J., Rodrigues, R., Silva, J.C., et al. (2020). CRISP-EDM: uma proposta de adaptação do Modelo CRISP-DM para mineração de dados educacionais. *In Anais do XXXI Simpósio Brasileiro de Informática na Educação.* https://doi.org/10.5753/cbie.sbie.2020.1092

Roslan, M., & Chen, C. (2022). Educational data mining for student performance prediction: A systematic literature review (2015-2021). *International Journal of Emerging Technologies in Learning (iJET)*, 17(5), 147-179. https://doi.org/http://dx.https://doi.org/.org/10.3991/ijet.v17i05.27685

Said, N., & Srinivasa, G. (2023). Application of Discriminant Analysis to Predict Students' Performances in Mathematics in Advanced Secondary Schools. *European Journal of Statistics*, 3(1), 1-10. https://doi.org/10.28924/ada/stat.3.8

Salles, F., Dos Santos, R. & Keskpaik, S. (2020). When didactics meet data science: process data analysis in large-scale mathematics assessment in France. *Large-scale Assess Educ*, 8(7). https://doi.org/10.1186/s40536-020-00085-y

Sánchez, R., & Mateos, J. (2023). Minería de Datos Educacionales: Descubrir tesoros ocultos durante el aprendizaje: Educational Data Mining: Discover hidden treasures during learning. *REVISTA CIENTÍFICA ECOCIENCIA*, 10(1), 18-41. https://doi.org/10.21855/ecociencia.100.830

Selly, A., & Anna, A. (2022). Learning Analytics dan Educational Data Mining pada Data Pendidikan. *JURNAL RISET PEMBELAJARAN MATEMATIKA SEKOLAH*, 6(1), 12-20. https://doi.org/10.21009/jrpms.061.02

Shi, X. (2024). Character Data Mining in Educational Scene. *Journal of Electrical Systems*, 20(1), 57-63. https://doi.org/10.52783/jes.2358

Sukanya, S., & D William, A., & Mahesh, et al. (2023). A Machine Learning Approach to Predicting Academic Performance. *International Journal of Engineering Technology and Management Sciences*, 7(1), 12-16. https://doi.org/10.46647/ijetms.2023.v07i06.003

Thorat, R. (2024). Role of Mathematics in Data Science - Machine learning. *International Journal of Scientific Research in Modern Science and Technology*, 3(3), 18-21. https://doi.org/10.59828/ijsrmst.v3i3.191

Xie, G., Liu, X. (2023). Gender in mathematics: how gender role perception influences mathematical capability in junior high school. *The Journal of Chinese Sociology*, 10(1), 10. https://doi.org/10.1186/s40711-023-00188-3

Yadav, N., & Deshmukh, S. (2023). Prediction of Student Performance Using Machine Learning Techniques: A Review. *In International Conference on Applications of Machine Intelligence and Data Analytics (ICAMIDA 2022), Atlantis Press*, 735-741. https://doi.org/10.2991/978-94-6463-136-4_63

Yağcı, M. (2022). Educational data mining: prediction of students academic performance using machine learning algorithms. *Smart Learn, Environ*, 9(11). https://doi.org/10.1186/s40561-022-00192-z

Yan, C. (2022). Research on Student Academic Performance Prediction Methods. *Highlights in Science, Engineering and Technology*, 24(1), 257-263. https://doi.org/10.54097/hset.v24i.3940

Yu, N., Wen, W. (2020). How well do teachers predict students' actions in solving an ill-defined problem in stem education: a solution using sequential pattern mining. *IEEE Access*, 8(1), 134976-134986. https://doi.org/10.1109/ACCESS.2020.3010168

Zhao, L., Ren, J., Zhang, L., et al. (2023). Quantitative analysis and prediction of academic performance of students using machine learning. *Sustainability*, 15(16), 12531. https://doi.org/10.3390/su151612531