

ALGORITMOS DE APRENDIZAJE AUTOMÁTICO EN LA PREDICCIÓN DEL RENDIMIENTO ACADÉMICO UNIVERSITARIO: UNA REVISIÓN SISTEMÁTICA

Machine Learning Algorithms in the Prediction of University Academic Performance: a Systematic Review

Fabiola Montero

Universidad de Panamá, Facultad de Informática, Electrónica y Comunicación
fabiola.monterog@up.ac.pa <https://orcid.org/0000-0002-4681-9471>

Nelson Montilla

Universidad de Panamá, Facultad de Informática, Electrónica y Comunicación
nelson.montilla@up.ac.pa <https://orcid.org/0000-0002-4220-4912>

Julio Arcia

Universidad de Panamá, Facultad de Informática, Electrónica y Comunicación
julio.arcia@up.ac.pa <https://orcid.org/0009-0006-8052-792X>

Recibido: 31-1-2024, Aceptado: 8-5-2024

DOI <https://doi.org/10.48204/3072-9696.6361>

RESUMEN

En los últimos años, el crecimiento del Aprendizaje Automático o Machine Learning ha tenido un impacto significativo en diversos campos, incluyendo la educación. En el ámbito educativo, se han aplicado técnicas o algoritmos de Aprendizaje Automático para diversos propósitos, como predecir el rendimiento académico de estudiantes de educación superior. Este artículo se enfoca en la revisión de publicaciones que muestran los resultados obtenidos al emplear algoritmos de Aprendizaje Automático para predecir el rendimiento académico de estudiantes universitarios. Se realizó una revisión sistemática en bases de datos académicas como Redalyc, Google Académico y Scielo, empleando filtros de búsqueda y palabras clave pertinentes. Posteriormente, se seleccionaron aquellos artículos más relevantes para el análisis. El propósito de este artículo de revisión es identificar los algoritmos de Aprendizaje Automático más frecuentemente utilizados y efectivos en la predicción del rendimiento académico de estudiantes universitarios, así como examinar cómo han sido implementados en distintas universidades para que sea una guía en futuras investigaciones que incluyan la aplicación de estos algoritmos con datos de estudiantes de la Universidad de Panamá. Los resultados muestran que los algoritmos más empleados para este tipo de predicción son Árboles de Decisión, K-Vecinos más Cercanos (KNN), Naive Bayes, Random Forest (Bosques Aleatorios) y Máquinas de Soporte Vectorial (SVM), así como las Redes Neuronales Artificiales, las cuales también destacan por su alta precisión. El empleo de estos algoritmos para predecir el rendimiento académico en universidades ofrece un gran potencial para reducir

la deserción, mejorar los resultados académicos, respaldar las decisiones pedagógicas y optimizar recursos educativos.

ABSTRACT

In recent years, the growth of Machine Learning has had a significant impact across various fields, including education. In the educational sphere, Machine Learning techniques or algorithms have been applied for various purposes, such as predicting the academic performance of higher education students. This article focuses on reviewing publications that present the results obtained by employing Machine Learning algorithms to predict the academic performance of university students. A systematic review was conducted using academic databases such as Redalyc, Google Scholar, and Scielo, utilizing search filters and relevant keywords. Subsequently, the most relevant articles were selected for analysis. The purpose of this review article is to identify the most frequently used and effective Machine Learning algorithms for predicting the academic performance of university students, as well as to examine how they have been implemented in different universities, serving as a guide for future research involving the application of these algorithms with student data from the University of Panama. The results show that the most commonly used algorithms for this type of prediction are Decision Trees, K-Nearest Neighbors (KNN), Naive Bayes, Random Forest, and Support Vector Machines (SVM), as well as Artificial Neural Networks, which also stand out for their high accuracy. The use of these algorithms to predict academic performance in universities offers great potential to reduce dropout rates, improve academic outcomes, support pedagogical decisions, and optimize educational resources.

PALABRAS CLAVES

Algoritmo, aprendizaje automático, aprendizaje supervisado, rendimiento académico.

KEYWORDS

Algorithm, machine learning, supervised learning, academic performance.

INTRODUCCIÓN

El Aprendizaje Automático, también conocido como Machine Learning en inglés, es un campo muy desarrollado dentro de la Inteligencia Artificial. Según Norman (2019), el Aprendizaje Automático consiste en proporcionar a una computadora un conjunto de datos para que construya un modelo capaz de predecir comportamientos futuros. Géron (2020), citado por Díaz (2021), define el Aprendizaje Automático como la “ciencia (y arte) de programar computadores para que aprendan a partir de datos” (p.182).

Para Soria et al. (2023), el Aprendizaje Automático es la disciplina que permite a las computadoras aprender sin necesidad de ser programadas con reglas específicas. Esta disciplina aplica estadística y otros métodos

matemáticos para detectar patrones en los datos y, a partir de ellos, hacer predicciones e incluso tomar decisiones, mejorando con cada intento.

Según Moreno et al. (1994) desde sus inicios, se han identificado diferentes paradigmas en el Aprendizaje Automático, como el aprendizaje deductivo, analítico, analógico, inductivo, mediante descubrimiento, algoritmos genéticos y conexionismo, entre otros. Sin embargo, de acuerdo con Díaz (2021) la disponibilidad masiva de datos y el crecimiento exponencial de Internet han impulsado el rápido desarrollo del Aprendizaje Automático, dividiéndose en categorías como aprendizaje supervisado y no supervisado, entre otras. Estos enfoques desarrollan modelos mediante algoritmos que permiten entrenarlos con una gran cantidad de datos y capacitarlo para resolver tareas específicas, como la realización de predicciones.

Aprendizaje Supervisado y No Supervisado

Los algoritmos de aprendizaje supervisado se caracterizan por el uso de un conjunto de datos etiquetados para entrenar un modelo que clasifique datos o que prevea resultados con precisión. En la medida en que se ingresan datos al modelo, este va realizando ajustes hasta que se adapte de manera correcta, permitiendo que aprenda con el tiempo (IBM, 2023).

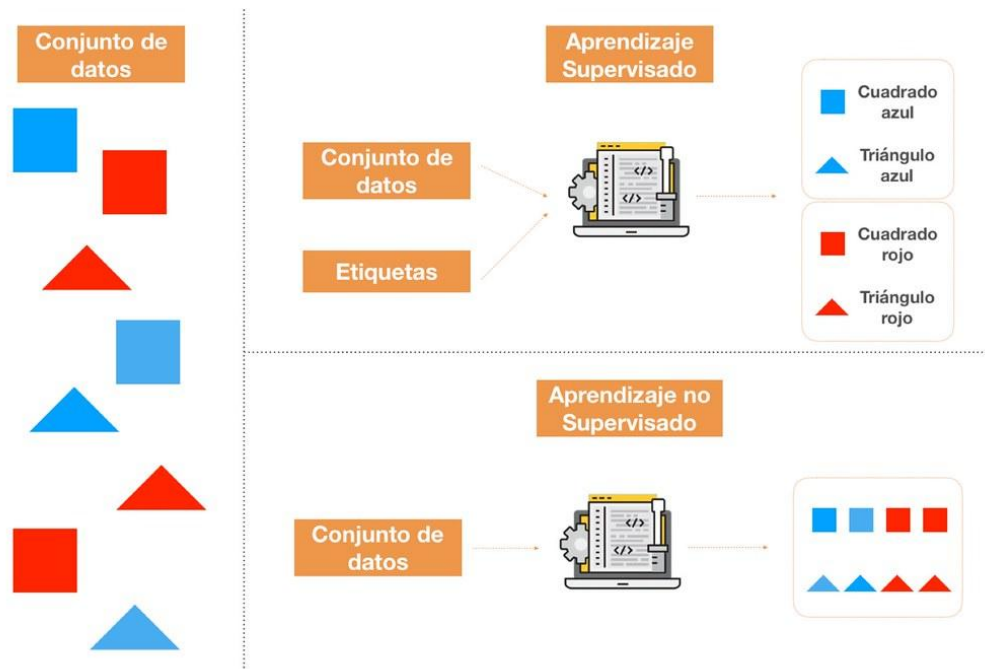
El aprendizaje supervisado utiliza datos etiquetados, donde tanto las entradas como las salidas son conocidas (González, 2018). Según Sandoval (2018), se entrena un algoritmo de aprendizaje supervisado proporcionándole características y etiquetas para que pueda realizar predicciones en el futuro. En este proceso, el modelo aprende de estos ejemplos para hacer predicciones sobre nuevos conjuntos de datos, aplicando el conocimiento adquirido a datos no vistos anteriormente.

Los algoritmos de aprendizaje no supervisado, por otro lado, analizan y agrupan conjunto de datos no etiquetados utilizando variables de entrada. Estos algoritmos descubren patrones ocultos o agrupaciones de datos sin intervención humana (IBM, 2023), identificando patrones que describan el comportamiento general de los datos. De acuerdo con Sandoval (2018), en el aprendizaje no supervisado se proporcionan solo las características, no las etiquetas, para que el algoritmo agrupe los datos según sus similitudes.

La Figura 1, ilustra la diferencia fundamental entre el aprendizaje supervisado y el no supervisado en términos de etiquetado y resultados.

Figura 1

Comparación de los tipos de aprendizaje: supervisado y no supervisado



Nota. Adaptado de Diferencia entre aprendizaje supervisado y no supervisado, por L. Gonzalez, 2018, AprendeIA (<https://aprendeia.com/diferencia-entre-aprendizaje-supervisado-y-no-supervisado/>)

En la Tabla 1, se aprecia una comparación entre los algoritmos de aprendizaje supervisado y no supervisado, destacando sus características fundamentales, de acuerdo con Serrato (2021) y González (2018, 2020). La tabla resume los aspectos clave que diferencian a estos dos tipos de aprendizaje automático, facilitando la comprensión de sus principios y su uso en diferentes escenarios.

Tabla 1

Diferencias entre aprendizaje supervisado y no supervisado

Característica	Aprendizaje Supervisado	Aprendizaje No Supervisado
Etiquetado de datos	Requiere etiquetado de datos con la salida deseada	No requiere datos etiquetados
Objetivo	Predecir resultados para nuevos datos	Descubrir patrones y estructuras en datos sin etiquetar
Datos de entrenamiento	Etiquetados con el resultado esperado	No etiquetado ni con información sobre resultados esperados

Complejidad computacional	Sencilla	Compleja
Relación entrada-salida	Aprende la relación entre la entrada y la salida	No utiliza datos de salida
Aplicaciones	Regresión, clasificación	Agrupación de datos detección de anomalías, reducción de dimensionalidad

Según, Pineda (2022), los algoritmos supervisados se pueden agrupar en las siguientes categorías:

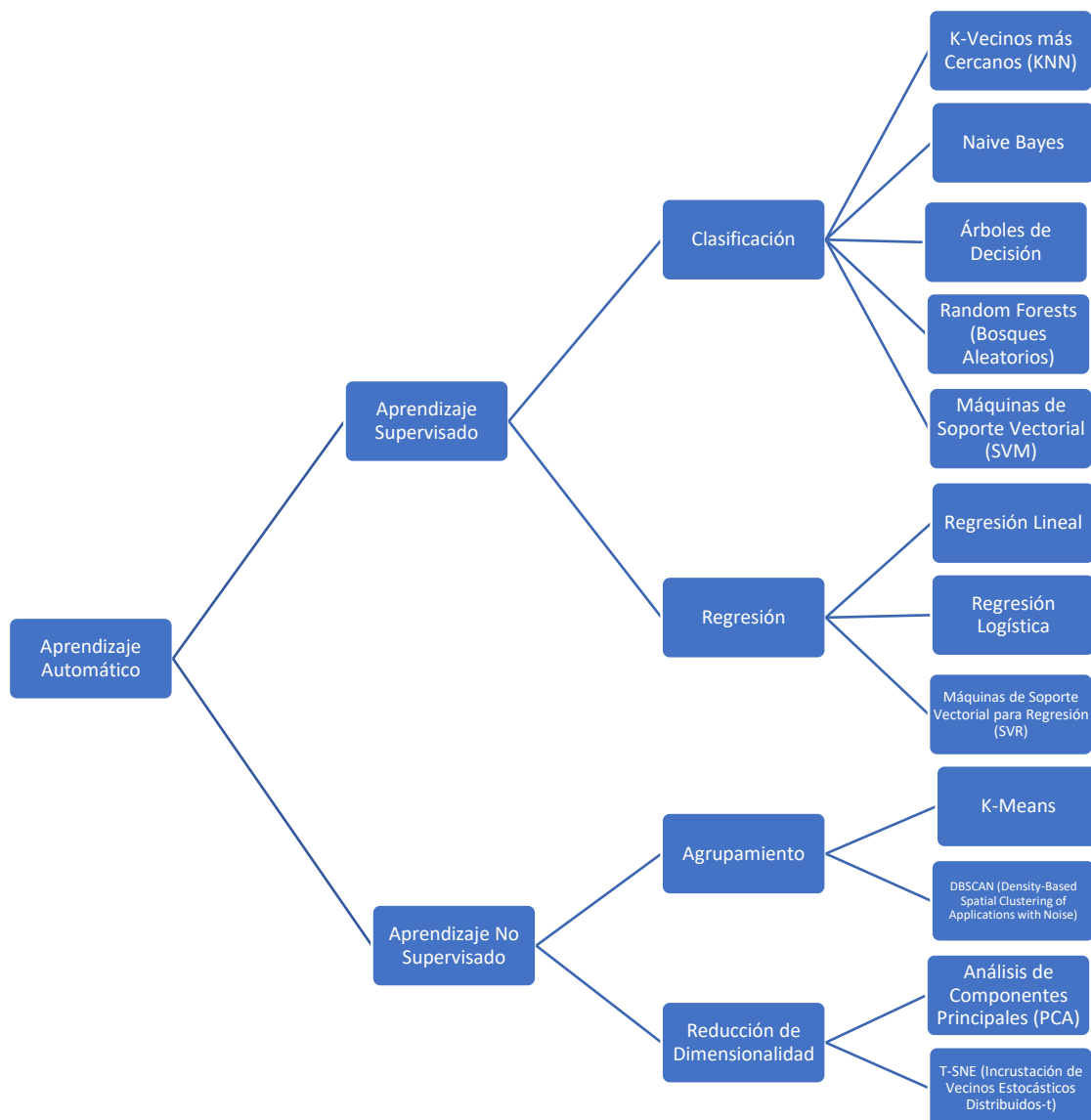
- Clasificación: se utilizan para asignar etiquetas categóricas a nuevas observaciones.
- Regresión: Se utilizan para predecir valores numéricos a partir de datos.

Por otra parte, de acuerdo con Soria et al. (2023), los algoritmos no supervisados se pueden clasificar en las siguientes categorías principales:

- Agrupamiento: se utilizan para agrupar observaciones similares en grupos.
- Reducción de dimensionalidad: se utilizan para reducir el número de variables en un conjunto de datos sin perder información importante.

En la Figura 2, se presenta una clasificación de algoritmos de aprendizaje automático, tanto supervisado como no supervisado, acompañada de ejemplos representativos de los algoritmos más comúnmente utilizados en cada categoría.

Figura 2
Clasificación de Algoritmos de Aprendizaje Automático



Los algoritmos, tanto supervisados como no supervisados, están en constante evolución. Con el tiempo, sus capacidades, aplicaciones y diversidad se han expandido, consolidándolos como una de las áreas de desarrollo más importantes en la actualidad.

Aprendizaje Profundo o Deep Learning

Además de los algoritmos de aprendizaje supervisado y no supervisado se encuentra el Aprendizaje Profundo o Deep Learning, que es un subcampo del Aprendizaje Automático que a su vez lo es de la Inteligencia Artificial y que es más poderoso que los algoritmos de aprendizaje supervisado y no supervisado, puesto que está basado en técnicas distintas que imitan el comportamiento de la

naturaleza mediante programación evolutiva. De acuerdo con, Lagares et al. (2022), el Aprendizaje Profundo es utilizado para resolver problemas que requieren modelos con mayor complejidad y que con las técnicas comunes de los aprendizajes supervisados y no supervisados no son suficientes.

Los algoritmos de Aprendizaje Profundo se distinguen por extraer conocimiento de los datos, ya que en el medio de estos algoritmos existe una estructura denominada Redes Neuronales o Redes Neuronales Artificiales que simulan la estructura biológica que encontramos en una neurona en el cerebro humano y de la misma manera como funcionan biológicamente, las neuronas del Aprendizaje Profundo procesan información y cuentan con grandes capacidades de aprendizaje. Existe una gran cantidad de tipos de algoritmos de Aprendizaje Profundo que han sido clasificados como redes con aprendizaje supervisado, no supervisado o aprendizaje por refuerzo. Para Yepes (2022), una de las Redes Neuronales más utilizadas son las Redes Neuronales Convolucionales (Convolutional Neural Network, CNN), que es una variación del Perceptrón Multicapa, pero que su aplicación se realiza en matrices bidimensionales, por lo que son muy efectivas en aplicaciones de visión artificial, clasificación y segmentación de imágenes.

Adicionalmente a las Redes Neuronales Convolucionales, están las Redes Neuronales Profundas (Deep Neuronal Network, DNN), Redes Neuronales Recurrentes (Recurrent Neuronal Network, RNN) y un subtipo de estas llamado las Redes Neuronales de gran memoria a corto plazo (Long Short Term Memory, LSTM) (Lagares et al., 2022).

De acuerdo con la tarea que deba llevarse a cabo, se puede optar por trabajar con algoritmos supervisados, no supervisados o de aprendizaje profundo e incluso combinando algoritmos de diferentes tipos si es necesario, para lograr satisfacer de manera más adecuada la necesidad y alcanzar un nivel más efectivo en los resultados.

Rendimiento Académico

En el ámbito de la educación superior, un elemento crítico es el rendimiento académico tal como resalta Enughwure y Ogbise (2020), ya que constituye el indicador primordial de éxito o fracaso del estudiante que puede ser analizado desde diferentes dimensiones de acuerdo con el contexto o la situación de interés (Zhang et al., 2021). Por lo general, el rendimiento se evalúa mediante calificaciones que se asignan en una escala que puede clasificarse como Excelente, Bueno, Regular o Deficiente, entre otras categorías posibles de acuerdo con el nivel de enseñanza o el sistema educativo. Albreiki et al. (2021) considera que la predicción del rendimiento académico de los estudiantes proporciona excelentes beneficios para aumentar las tasas de retención, una gestión eficaz de la matrícula, la gestión de exalumnos, un mercadeo dirigido mejorado y la eficacia general de la institución.

El rendimiento de los estudiantes en el proceso educativo de acuerdo con Sandra et al. (2021), se puede definir idealmente como algo que se obtiene a partir de cambios en el comportamiento de los estudiantes en función de sus experiencias, ya que los resultados de aprendizaje también son una realización

del potencial o capacidad que poseen los estudiantes. Sandra et al. (2021) sostiene que los resultados de aprendizaje de los estudiantes se pueden observar en su comportamiento y en la forma de comprensión de conocimientos, habilidades de pensamiento o habilidades motoras.

García (2020) define el rendimiento académico “como una consecuencia de la relación entre la capacidad del estudiante y su desempeño” (p. 3). Para Oliver (2007), citado por García (2020), suele existir una tendencia a asumir que el rendimiento académico de un estudiante está relacionado principalmente con sus responsabilidades y hábitos de estudio, sin profundizar en una evaluación más completa que tenga en cuenta otros factores importantes que afectan el resultado como lo indican Buenaño et al. (2019) y Ojajuni et al. (2021). Estos factores no son fáciles de identificar o analizar a simple vista y, a menudo, están ocultos en los datos de los estudiantes que las instituciones suelen recopilar y almacenar. Lamentablemente, por cuestiones como el desconocimiento, la complejidad o el volumen, estos datos no siempre se utilizan correctamente. No obstante, el uso de técnicas de Aprendizaje Automático como lo indica Issah et al. (2023) ofrece la posibilidad de descubrir información valiosa disponible en los resultados académicos obtenidos.

Para Chandana et al. (2022), al identificar diferentes factores que afectan el desempeño de aprendizaje y el desempeño de un estudiante durante su carrera académica; es posible comprender mejor las causas del bajo rendimiento académico. En los sistemas educativos es difícil rastrear el comportamiento y las características de los estudiantes ya que no existe ninguna automatización o herramienta que prediga o proyecte cómo será el desempeño académico de los estudiantes. Encontrar la correlación entre los factores que afectan el rendimiento de los estudiantes y los resultados académicos juega un papel vital en el sector educativo actual. La predicción del rendimiento para Ghazal (2020) es importante para los estudiantes, ya que ellos y sus asesores académicos pueden utilizarla para tomar decisiones informadas. Esto también puede ayudar a identificar las medidas adecuadas a tomar y crear programas de titulación personalizados que les permitan adquirir con éxito y eficacia los conocimientos necesarios para completar sus estudios de manera oportuna.

Diversos estudios sobre el rendimiento académico en estudiantes universitarios han determinado que múltiples factores pueden influir en su desempeño como lo resaltan Balaji et al. (2021) y Contreras et al. (2023). Entre estos, se destaca el realizado por Jiménez (2018), que menciona tres factores que inciden en el rendimiento académico: el sexo, el acceso a becas y el nivel de uso de las tecnologías de la información y la comunicación (TIC).

Además, según Espejel y Jiménez (2019) el rendimiento de los estudiantes universitarios también puede estar influenciado por el nivel de estudios de la madre y la ocupación del padre. La influencia del nivel de estudios de la madre se debe a que participa activamente en actividades escolares y quien coadyuva al desarrollo de habilidades cognitivas debido a la convivencia. También se pudo comprobar que a nivel universitario el rendimiento académico de los estudiantes está influenciado por la ocupación del padre, donde aquellos

con mejor ingreso pueden proporcionar fuentes y recursos físicos y electrónicos para el aprendizaje y así mejorar sus habilidades cognitivas.

En cambio, Bedregal et al. (2020) señalan que el rendimiento académico está estrechamente relacionado con la proporción entre el número de créditos aprobados y el número de créditos en los que un estudiante se matriculó. En cambio, Mora (2015) destaca que existen otros factores como lo son el entorno familiar, laboral o de salud, pero que las variables de tipo académicas son de mayor utilidad para la toma de decisiones.

Según Lunghi (2022), un fenómeno recurrente es la entrada masiva de estudiantes a programas universitarios, pero la cantidad de graduados no guarda proporción con la cantidad de estudiantes que ingresan, situación que se ha agravado en muchos países desde 2020 debido a la pandemia de COVID-19. En este sentido, Preciado et al. (2022), consideran que esta disparidad se atribuye a diversos factores que inciden en el rendimiento académico, como se mencionó previamente, lo que lleva a que los estudiantes repitan múltiples asignaturas del plan de estudio, prolongando su tiempo de estudio.

De acuerdo con Ayala (2022); Cruz et al. (2022); Valero et al. (2022) y Contreras et al. (2023), ante esta problemática, se han implementado algoritmos de Aprendizaje Automático para predecir o anticipar el rendimiento de estudiantes en distintas universidades. Los resultados han permitido la detección temprana de dificultades académicas y la toma de decisiones adecuadas, como el diseño de estrategias para asegurar el éxito académico, que es el principal objetivo de cualquier institución educativa.

Es importante resaltar que, además de la predicción del rendimiento académico, se han empleado algoritmos de Aprendizaje Automático para predecir la deserción de estudiantes universitarios. Por ejemplo, Miranda y Guzmán (2017) emplearon tres clasificadores distintos: redes bayesianas, árboles de decisión y redes neuronales, cada uno de los cuales produjo resultados determinantes para identificar la deserción estudiantil. Por otro lado, Urteaga et al. (2020) señalan que, a pesar de las ventajas de la educación en línea, tiende a experimentar altos índices de abandono. Por lo que, implementaron diversos algoritmos para desarrollar modelos predictivos de deserción en cursos virtuales ofrecidos por la Secretaría de Extensión de la Universidad Tecnológica Nacional de la República Argentina, Regional Buenos Aires, concluyendo que el algoritmo más eficaz para la creación de modelos fue una red neuronal.

Este artículo se fundamenta en la revisión de diversas publicaciones que presentan los resultados de la implementación de algoritmos de Aprendizaje Automático para predecir el rendimiento académico de estudiantes universitarios. Es esencial tener en cuenta que no existe una única solución para abordar este problema, ya que la elección del algoritmo adecuado puede variar según los datos específicos de cada institución. Por lo tanto, este artículo de revisión tiene como objetivo identificar los algoritmos de Aprendizaje Automático más frecuentemente utilizados que han demostrado una alta precisión y exactitud en la predicción del rendimiento académico de estudiantes

universitarios, además de examinar cómo se han implementado o aplicado en distintas universidades.

MATERIALES Y MÉTODOS

La metodología empleada para la elaboración del artículo es la de revisión sistemática. Fonseca y López (2014) señalan que la revisión sistemática resume de manera efectiva los resultados de investigaciones primarias, por medio de estrategias diseñadas para limitar el sesgo y el error aleatorio. Estas estrategias incluyen desde una búsqueda exhaustiva de todos los artículos relevantes hasta la identificación, selección y análisis de los estudios que se incluirán en la revisión. Posteriormente, se lleva a cabo la síntesis de la información recopilada, seguida de la interpretación de los resultados obtenidos.

Inicialmente, se definieron los objetivos de la revisión:

- Identificar los algoritmos de Aprendizaje Automático más frecuentemente utilizados y efectivos en la predicción del rendimiento académico de estudiantes universitarios.
- Examinar cómo han sido implementados o aplicados en distintas universidades para que sea una guía en futuras investigaciones que incluyan la aplicación de estos algoritmos con datos de estudiantes de la Universidad de Panamá.

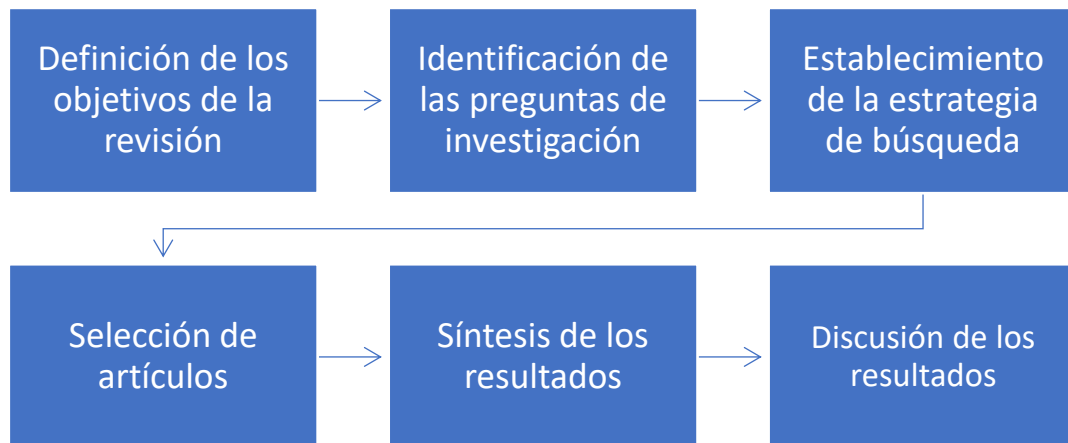
Luego, se definieron las preguntas de investigación:

- ¿Cuáles son los algoritmos de Aprendizaje Automático más utilizados y efectivos en la predicción del rendimiento académico de estudiantes universitarios?
- ¿Cómo han sido aplicados estos algoritmos en distintas universidades para predecir el rendimiento académico de estudiantes universitarios?

Posteriormente, se llevó a cabo una búsqueda exhaustiva, siguiendo criterios de selección de artículos, específicamente consultando en bases de datos académicos, incluyendo Google Académico, Redalyc, Scielo, Dialnet, así como la consulta de revistas especializadas y libros relacionados con el tema de interés. Para llevar a cabo esta búsqueda, se utilizaron palabras clave pertinentes y se aplicaron filtros de búsqueda adecuados para optimizar la selección de artículos relevantes publicados desde 2019 al 2024. En este sentido, los artículos presentados en esta revisión han seguido un proceso compuesto por seis etapas distintas, tal como se ilustra en la Figura 3.

Cada artículo encontrado fue minuciosamente revisado, resultando en la selección de 30 artículos definitivos o primarios para la presentación de los resultados y discusión. Esta selección se basó en la importancia de estos artículos en relación con las distintas aplicaciones de algoritmos de Aprendizaje Automático en la predicción del rendimiento académico de estudiantes universitarios.

Figura 3
Diagrama del proceso de revisión sistemática



RESULTADOS

A continuación, hacemos referencia a los artículos que han sido seleccionados y revisados en relación con la aplicación de algoritmos de Aprendizaje Automático en la predicción del rendimiento académico de estudiantes universitarios:

En un estudio realizado por Gutiérrez et al. (2022), se evaluaron cuatro algoritmos de Aprendizaje Automático: Árboles de Decisión, Bosques Aleatorios, Redes Neuronales y Máquinas de Soporte Vectorial (SVM), con el objetivo de predecir el rendimiento académico de estudiantes de la Facultad de Ciencias Empresariales y Económicas de la Universidad de Lima, Perú. Se utilizaron métricas como la matriz de confusión, el nivel de precisión y la curva de ROC para comparar y seleccionar el algoritmo más adecuado. Los resultados mostraron que los Bosques Aleatorios obtuvieron la mayor precisión en comparación con los otros algoritmos evaluados, con un 80%.

Por otro lado, Contreras et al. (2020) llevaron a cabo un estudio sobre la predicción del rendimiento académico en estudiantes de Ingeniería Industrial de la Universidad Distrital de Colombia. Su metodología incluyó la aplicación de algoritmos de clasificación como Árbol de Decisión, Máquinas de Soporte Vectorial (SVM), Red Neuronal (Perceptrón) y K-Vecinos más cercanos (KNN), y la evaluación de la calidad de la clasificación utilizando métricas como exactitud, precisión, recall y el puntaje F1. Se identificaron siete variables que influyen significativamente en el rendimiento académico, entre las que se incluyen la edad, género, puntajes ICFES para aptitudes matemáticas y puntaje global ICFES. Los algoritmos SVM y Red Neuronal (Perceptrón) se seleccionaron debido a su alta precisión en la predicción del rendimiento académico.

El estudio realizado por Gamboa y Salinas (2022) identifica las variables determinantes en la predicción del rendimiento académico de estudiantes universitarios después de seis semestres en la Universidad Nacional Agraria La Molina de Lima, Perú, utilizando algoritmos de Aprendizaje Automático. Se aplicó la metodología CRISP (Cross-Industry Standard Process for Data Mining) y se analizó un conjunto de datos con 622 registros. Se probaron varios algoritmos, como Regresión Logística, K-Vecinos más Cercanos (KNN), Naive Bayes, Árboles de Decisión (C5.0 y CART), Red Neuronal Perceptrón Multicapa, Máquinas de Soporte Vectorial con Kernel Lineal (SVM-L) y Kernel Radial (SVM-R), destacando Regresión Logística, Naive Bayes y SVM-L. Posteriormente, se combinaron estos tres mediante un algoritmo de ensamblaje, una técnica para fortalecer el rendimiento mediante el promedio de las probabilidades obtenidas. Los resultados resaltaron la importancia de las variables calificaciones del primer semestre, mientras que las variables sociodemográficas no fueron significativas en la predicción del rendimiento académico a largo plazo.

Rico et al. (2019) desarrollaron y aplicaron un modelo predictivo para el rendimiento académico de estudiantes de Ingeniería en el Instituto Politécnico Nacional de la Ciudad de México, utilizando el algoritmo Naive Bayes en un proceso de Descubrimiento de Conocimiento en Bases de Datos (KDD). Recopilaron datos de 122 estudiantes a través de una encuesta y utilizaron estos datos para predecir el rendimiento de 71 estudiantes adicionales. Realizaron la integración y recopilación de datos, el procesamiento de datos y la construcción del modelo predictivo, evaluando la precisión de las predicciones mediante validación cruzada. Los resultados demostraron una alta precisión en las predicciones del algoritmo Naive Bayes, incluso con un conjunto de datos limitado, lo que sugiere su utilidad en entornos universitarios con pocos datos disponibles. Además, el modelo desarrollado identifica los factores más influyentes en la aprobación y reprobación de los estudiantes desde el inicio del curso.

Gerson (2019) realizó un estudio sobre la predicción del rendimiento académico de los estudiantes de la Facultad de Ingeniería de la Universidad Nacional del Centro del Perú, utilizando modelos de Aprendizaje Automático Supervisado como la Regresión Logística y Random Forest. En esta investigación, se emplearon el lenguaje de programación Python y la librería Scikit-Learn, con el apoyo de las librerías Pandas y Numpy para el tratamiento de los datos. Los resultados mostraron que el modelo de Regresión Logística alcanzó una exactitud del 77%, mientras que el modelo de Random Forest logró un 84%.

Por su parte, Candia (2019), realizó una investigación para predecir el rendimiento académico de los estudiantes de la Universidad Nacional de San Antonio Abad del Cusco (UNSAAC). Empleando la metodología CRISP-DM, WEKA como plataforma, así como varios algoritmos supervisados como Árboles de Decisión J48, Random Forest, K-Vecinos más Cercanos (KNN), Regresión Logística y Perceptrón Multicapa. Los resultados indican que el algoritmo que proporcionó la mejor predicción fue el Random Forest con una efectividad del 69%. En segundo lugar, se ubicó la Regresión Logística con un 68%. Los factores que influyen en el rendimiento académico incluyen la nota de ingreso a

la institución, la Escuela Profesional donde estudia, el semestre, la modalidad de ingreso y la cantidad de cursos matriculados.

Castrillón et al. (2020) emplearon el algoritmo de clasificación Bayesiano J48 a través del programa WEKA para identificar los factores más influyentes en el rendimiento académico de los estudiantes. Entre estos factores se encuentran la pedagogía de los profesores, horarios de clase adecuados, una buena relación docente-estudiante, la calidad académica de los docentes y una actividad extracurricular moderada. Mediante este algoritmo, lograron generar un Árbol con los atributos más relevantes, obteniendo una efectividad del 91.67%.

En su investigación, Diaz et al. (2021) emplearon el algoritmo de Árbol de Decisión en conjunto con minería de datos, específicamente el algoritmo J48 de Árbol de Decisión del software WEKA. Utilizaron datos de 237 estudiantes de maestría de la Escuela de Postgrado de la Universidad Nacional José Faustino Sánchez Carrión en Perú, recopilados mediante cuestionarios. La precisión del modelo se evaluó con el índice Kappa de Cohen, mostrando un alto nivel de acierto en las predicciones. Los resultados obtenidos con el algoritmo de Árbol de Decisión fueron exitosos, lo que demuestra la eficacia para predecir problemas de rendimiento académico. Sin embargo, los autores sugieren mejorar la precisión del modelo priorizando ciertas variables.

Rico y Gaytán (2022) diseñaron y evaluaron modelos para predecir el rendimiento académico de un curso de matemáticas en una Universidad de México, utilizando técnicas de aprendizaje automático como Bayes Ingenuo, K-Vecinos más Cercanos (KNN) y Árbol de Decisión. La construcción y evaluación de los modelos se realizaron con el software Weka. Para el entrenamiento, se recopilaron las calificaciones de las primeras cinco actividades académicas de 260 estudiantes de cursos anteriores. Los datos de pruebas correspondieron a los registros de 112 estudiantes inscritos en el curso actual. Se compararon las predicciones con los resultados reales obtenidos por estos 112 estudiantes al finalizar el curso, para calcular la exactitud de cada técnica, terminado el curso. Bayes Ingenuo logró una exactitud del 75.9%, seguido por KNN con un 69.6%, y finalmente el Árbol de Decisión con un 60.7%.

En el estudio de Santamaría et al. (2023), se compararon varios algoritmos de Aprendizaje Automático, como Random Forest, Máquinas de Soporte Vectorial (SVM), K-Vecinos más Cercanos (KNN) y Regresión Logística, junto con un modelo de Aprendizaje Profundo basado en Redes Neuronales Artificiales (ANN). El objetivo era determinar cuál de estos modelos ofrecía la mayor precisión al predecir el rendimiento académico de estudiantes universitarios de la Universidad de Guayaquil en Ecuador. Los datos se prepararon para el aprendizaje supervisado, asignando el 80% para entrenamiento y el 20% para pruebas. Se evaluaron mediante métricas como la curva ROC, el área bajo la curva (AUC) y la métrica F1-score que indicó que todos los modelos tenían un nivel aceptable de robustez, destacando la Red Neuronal Artificial con un 89.3% de robustez en las predicciones.

Rico (2022), desarrollo y evaluó modelos predictivos para el rendimiento académico de estudiantes universitarios en México. Se incluyeron 260 estudiantes y 14 actividades académicas. Se construyó un modelo predictivo

utilizando los algoritmos Naive Bayes, K-Vecinos más Cercanos (KNN) y Árbol de Decisión, para predecir el rendimiento de 112 estudiantes en un semestre posterior. Los resultados mostraron que a medida que aumentaba el número de actividades consideradas, la exactitud de las predicciones también aumentaba. Los modelos con Naive Bayes mostraron un mejor desempeño, alcanzando una exactitud del 81.25%, en comparación con las otras dos técnicas empleadas. Esta metodología permitió identificar a los estudiantes en riesgo de reprobación del curso antes de que finalizara.

Chavez et al. (2023), propusieron un modelo de Red Neuronal Artificial para predecir el rendimiento académico, alcanzando una precisión del 93.81%. En comparación, otras técnicas como Bosques Aleatorios lograron una precisión del 92.81%, Máquina de Soporte Vectorial (SVM) un 80.03% y Naive Bayes un 92.21%. El modelo se evaluó utilizando información de al menos 32000 estudiantes de la Universidad Abierta del Reino Unido, considerando variables como la frecuencia de cursado, evaluaciones promedio, tasa de aprobación del curso, uso de materiales virtuales y clics en aulas virtuales. La arquitectura del modelo consta de tres capas, utilizando funciones de activación como la tangente hiperbólica sigmoidea para las capas de entrada y oculta, y una función lineal para la capa de salida. Este enfoque logró una efectividad de predicción del 78% y 75% para los dos primeros sujetos, respectivamente.

Carles (2023) desarrolló varios modelos de aprendizaje automático para predecir el rendimiento académico de los estudiantes universitarios de Universitat Politècnica de València en España. Con una población de 5035 alumnos, empleó algoritmos como Máquinas de Soporte Vectorial (SVM), Árboles de Decisión, Random Forest, Vecinos más Cercanos (KNN) y métodos de Regresión. Tras evaluar cada uno de estos modelos, el modelo de Árboles de Decisión fue seleccionado como el más adecuado, ya que mostró una precisión competente y una sólida capacidad de generalización en la predicción de la nota media de los estudiantes. Además, se concluyó que la nota de acceso al grado universitario es el factor más determinante del rendimiento académico global a lo largo de los años de estudio.

El estudio de El-Keiey et al. (2022) analizó el impacto de los rasgos de personalidad y la inteligencia de los estudiantes en su rendimiento académico, utilizando datos recopilados de la Facultad de Computación e Inteligencia Artificial de la Universidad de El Cairo. Se recolectaron datos de aproximadamente 300 estudiantes de diferentes niveles de pregrado. Se utilizaron algoritmos de Aprendizaje Automático, incluyendo K-Vecino más Cercano (KNN), Máquina de Soporte Vectorial (SVM), Árbol de Decisión (DT) y Naive Bayes (NB), para predecir el rendimiento. El DT mostró el mejor rendimiento, con una precisión del 90%. Se concluyó que los rasgos de personalidad son importantes para predecir el desempeño de los estudiantes universitarios.

Según Contreras et al. (2021), al emplear algoritmos supervisados como K-Vecinos más Cercanos (KNN), Máquinas de Soporte Vectorial (SVM), Naive Bayes y Árbol de Decisión para predecir el rendimiento académico, se obtiene un resultado de exactitud, del 74%, siendo el mejor resultado alcanzado

mediante el algoritmo Árbol de Decisión. Sin embargo, se observa que los métodos de ensamblaje superan a los supervisados. Se implementan algoritmos de ensamblaje que mejoran la exactitud de los clasificadores previos. Entre estos, se destacan métodos como Bagging (CART, Random Forest, ExtraTreesClassifier), métodos Boosting (AdaBoost, GBM, XGBoost, CatBoost, Light Boost) y Voting (Blending, Stacking). Los resultados indican que los algoritmos de Stacking y Blending alcanzan valores de precisión en cada semestre alrededor del 85% para entrenamiento y del 75% para prueba, lo que muestra los mejores resultados.

Chen et al. (2021), investigaron la capacidad de predecir el rendimiento académico de estudiantes universitarios basándose en sus comportamientos de lectura en un curso con libros electrónicos, utilizando un sistema desarrollado por la Universidad de Kyoto. Participaron 100 estudiantes de primer año de la Universidad de Asia en Taiwán. Se emplearon algoritmos de clasificación de Aprendizaje Automático, como Regresión Logística (LR), Bayes ingenuo Gaussiano (NB, una variante de Naive Bayes), Clasificación de Vectores de Soporte (SVC), Árbol de Decisión (DT), Bosque Aleatorio (RF) y Red Neuronal (NN), para predecir el rendimiento académico. NB destacó en la identificación de estudiantes en riesgo, mientras que LR y NN mostraron un rendimiento consistente en diferentes métricas para la predicción del rendimiento.

Bedregal et al. (2020) aplicaron Redes Neuronales y técnicas de aprendizaje supervisado para generar Árboles de Decisión, como C4.5 e ID3. De estas, el algoritmo C4.5 produjo las mejores métricas de rendimiento. Utilizaron variables como los créditos aprobados, los créditos que debería haber cursado, las calificaciones, el Promedio Ponderado Ajustado (PPA), el Índice de Rendimiento Exógeno y Endógeno (IREX, IREN), y el Índice de Rendimiento Total (IRT). Sin embargo, se encontró que la variable más significativa en la construcción del modelo es la relación entre los créditos aprobados por un estudiante y los créditos que debería haber cursado, seguida por las calificaciones.

En el caso de Ahmad et al. (2021), se propone un método para predecir el desempeño de los estudiantes de la Facultad de Ingeniería Eléctrica de la Universiti Teknologi MARA (UiTM) Terengganu, Malasia. El alcance de este artículo es la predicción de los estudiantes que podrían alcanzar la Lista del Decano (promedios de calificaciones GPA de 3.5 y superiores). Se empleó un modelo predictivo basado en una técnica de Aprendizaje Automático que utiliza Redes Neuronales Artificiales (ANN) para predecir el desempeño real de los estudiantes del primer semestre. Los hallazgos han demostrado que el modelo desarrollado podría predecir con precisión el resultado real de los estudiantes del primer semestre con éxito y con errores mínimos.

Por su parte, Ekubo y Esiefarienrhe (2022) evaluaron varias técnicas de Aprendizaje Automático para predecir el bajo rendimiento académico utilizando datos de estudiantes de la Universidad del Delta del Níger en Nigeria. El estudio aplicó la metodología CRISP-DM y la herramienta WEKA para modelado. Se probaron algoritmos clasificadores como Árbol de Decisión J48, Regresión Logística (LR), Perceptrón Multicapa (MLP) y Bayes ingenuo Gaussiano (NB).

Los resultados indicaron que MLP fue el mejor clasificador para este conjunto de datos.

Alturki y Alturki (2021) recopilaron registros de 300 estudiantes de la Facultad de Ciencias de la Información y Computación en Arabia Saudita. Se comparó el rendimiento de los siguientes algoritmos: Árboles de Decisión (C4.5, Simple CART, LADTree), Naive Bayes y Random Forest. Los resultados muestran que Naive Bayes y Random Forest logran la mayor precisión, con un 63,33% y 63%, respectivamente.

Por otro lado, Al-Alawi et al. (2023), emplearon algoritmos de Aprendizaje Automático Supervisado para investigar los factores que influyen negativamente en el rendimiento académico de estudiantes universitarios en calidad condicional. Utilizaron un conjunto de datos proporcionado por la Universidad Sultán Qaboos, que incluía 33 características y 37599 estudiantes, desde el otoño de 2009 hasta el verano de 2019. Se aplicó la metodología de Descubrimiento de Conocimiento en Bases de Datos (KDD) y después de seleccionar las características más efectivas, se construyeron modelos predictivos utilizando varios algoritmos supervisados que incluyeron Árbol de Decisión (J48 y Random Tree), Random Forest, Naive Bayes, K-Vecinos más Cercanos (KNN), Máquinas de Soporte Vectorial (SVM) y Redes Neuronales Artificiales (ANN). El algoritmo J48 de Árbol de Decisión obtuvo el mejor resultado con una precisión del 82,4%.

En cambio, Abdelmagid y Qahmash (2023), exploró la detección de patrones educativos y la predicción del rendimiento académico de estudiantes universitarios mediante la tecnología "Orange" para minería de datos. El estudio analizó los resultados de los estudiantes de la Universidad King Khalid en varios cursos para identificar los factores que influyen en su rendimiento. Se utilizaron técnicas de clasificación, agrupamiento y predicción, junto con algoritmos como Random Forest, Regresión Lineal, K-Vecinos más Cercanos (KNN) y Máquina de Soporte Vectorial (SVM). Los resultados mostraron que el algoritmo Random Forest fue el más efectivo para clasificar las puntuaciones de los estudiantes, seguido por Regresión Lineal, KNN y SVM.

A su vez, Meghji et al. (2023), proponen desarrollar un modelo de clasificación para predecir el rendimiento académico final de los estudiantes en una etapa temprana y determinar qué cursos influyen en este desempeño. Utilizan datos de estudiantes de Ingeniería de Software de la Universidad de Ingeniería y Tecnología de Mehran en Pakistán, con una muestra de 291 alumnos. Se emplearon algoritmos de clasificación como Naive Bayes (NB), Árbol de Decisión J48, Random Forest (RF) y KStar (variante del algoritmo K-Vecinos más cercanos (KNN)) para predecir el rendimiento al final de 4 años. Para cada algoritmo, se realizaron experimentos utilizando diferentes combinaciones de los datos recopilados. El clasificador de Árbol de Decisión J48 se identificó como el más efectivo para identificar cursos que podrían requerir intervención educativa.

Alturki et al. (2022), emplearon algoritmos de minería de datos para prever el rendimiento académico de estudiantes de maestría utilizando un conjunto de datos reducido y características fácilmente accesibles al inicio del programa.

Recopilaron más de 700 registros de estudiantes de la Facultad de Informática Empresarial y Matemáticas de la Universidad de Mannheim en Alemania, abarcando datos demográficos y calificaciones semestrales. Se aplicaron seis algoritmos: Regresión Logística (LR), Random Forest (RF), K-Vecinos más cercanos (KNN), Naive Bayes (NB), Máquina de Soporte Vectorial (SVM) y Redes Neuronales Artificiales (ANN). Los resultados revelaron que el modelo RF, superó a los demás algoritmos en términos de predicción del rendimiento académico de los estudiantes de maestría.

Yagci (2022) propone un nuevo modelo basado en algoritmos de Aprendizaje Automático para predecir las calificaciones de exámenes finales de estudiantes de pregrado, utilizando las calificaciones de exámenes parciales como fuente de datos. Se evaluaron y compararon varios algoritmos, utilizando datos de rendimiento académico de 1854 estudiantes en un curso de idioma turco I en una Universidad Estatal de Turquía. Los algoritmos evaluados incluyeron Random Forest (RF), Redes Neuronales (NN), Regresión Logística (LR), Máquina de Soporte Vectorial (SVM), Naive Bayes (NB) y K-Vecinos más cercanos (KNN), con el algoritmo RF logrando una precisión del 74,6% en la predicción del rendimiento académico de los estudiantes.

Ghosh et al. (2022) investigaron los factores que influyen en el rendimiento estudiantil y predicen su desempeño utilizando algoritmos de Aprendizaje Automático en la Universidad de Textiles de Bangladesh. Identificaron 11 factores significativos, incluidos aspectos psicológicos, personales y del entorno académico. Utilizaron dos algoritmos de aprendizaje supervisado, Máquina de Soporte Vectorial (SVM) y Bosques Aleatorios, para clasificar los resultados estudiantiles y predecir su desempeño. La comparación entre ambos algoritmos reveló que el modelo de Bosques Aleatorios superó al de SVM en términos de precisión, alcanzando un impresionante 96.88%, frente al 81.25% del SVM.

Alsubaie (2023), propuso un enfoque de Aprendizaje Automático para mejorar la calidad de los programas de capacitación en línea en la Universidad de Taif, utilizando la plataforma Maharat y los estándares de aprendizaje en línea del Reino de Arabia Saudita. Su objetivo fue predecir el rendimiento académico de los estudiantes basándose en su participación en el entorno de aprendizaje en línea. Después de extraer características relevantes y realizar la clasificación, se evaluaron las predicciones utilizando Máquina de Soporte Vectorial (SVM), Naive Bayes, K-Vecinos más Cercanos y Árbol de Decisión. Los resultados mostraron que SVM alcanzó una precisión del 93.2%, seguido por Naive Bayes con 92.3% en términos de predicción.

Por otro lado, Incio et al. (2023) desarrollaron y entrenaron una Red Neuronal Artificial para predecir los resultados académicos de estudiante de Ingeniería Civil en la asignatura de Matemática II en la Universidad Nacional Intercultural Fabiola Salazar Leguía de Bagua, Perú. Utilizaron la metodología CRISP-DM y el software Matlab, empleando dos algoritmos de aprendizaje supervisado: Scaled Conjugate Gradient (SCG) y Levenberg-Marquardt (LM), para comparar los resultados. La implementación de la Red Neuronal mostró mejores resultados utilizando el 85% de los datos para el entrenamiento, el 10% para la validación y el 5% para la prueba. El algoritmo ML demostró una mayor

efectividad en la predicción, alcanzando una capacidad predictiva del 86%, mientras que la red entrenada con el algoritmo SCG alcanzó una capacidad predictiva del 70%.

En su investigación sobre la predicción del rendimiento académico de los estudiantes ingresantes de la Facultad de Ingeniería en Industrias Alimentarias de la Universidad Nacional Agraria de la Selva (UNAS) de Perú, Ponce (2024) utilizó el software WEKA para analizar algoritmos como K-Vecinos más Cercanos (KNN), Random Forest (RF), Naive Bayes y Bagging. Empleando la metodología CRISP-DM y técnicas de minería de datos, logró desarrollar un modelo de aprendizaje automático que predice de manera significativa el rendimiento académico. Los indicadores considerados fueron la opción de ingreso, la nota de ingreso, el sexo del ingresante y el número de hermanos en la familia. Las pruebas de Anova de Friedman mostraron que los modelos de RF y KNN alcanzaron una exactitud del 98.4%.

Saltos y Cruz (2024) en su estudio del rendimiento académico de estudiantes de las carreras de Economía y Turismo de la Universidad Técnica de Manabí en Ecuador, emplearon la metodología CRISP-DM y ejecutaron varios algoritmos de aprendizaje automático, como Árbol de Decisión, Random Forest, Redes Neuronales y Máquina de Soporte Vectorial, para obtener el rendimiento académico. Después de evaluar los distintos algoritmos, encontraron que el más eficiente fue Random Forest, con una estimación aceptable del 86%. Este algoritmo proporciona valores precisos, lo que permitió crear un panel de control en Power BI, que muestra y analiza datos de manera clara y efectiva, incluyendo asistencia, notas, calificaciones y otros resultados académicos de los estudiantes.

En tabla 2, se aprecia la frecuencia con la que se han empleado los algoritmos de Aprendizaje Supervisado y Aprendizaje Profundo en la predicción del rendimiento académico, en esta revisión.

Tabla 2

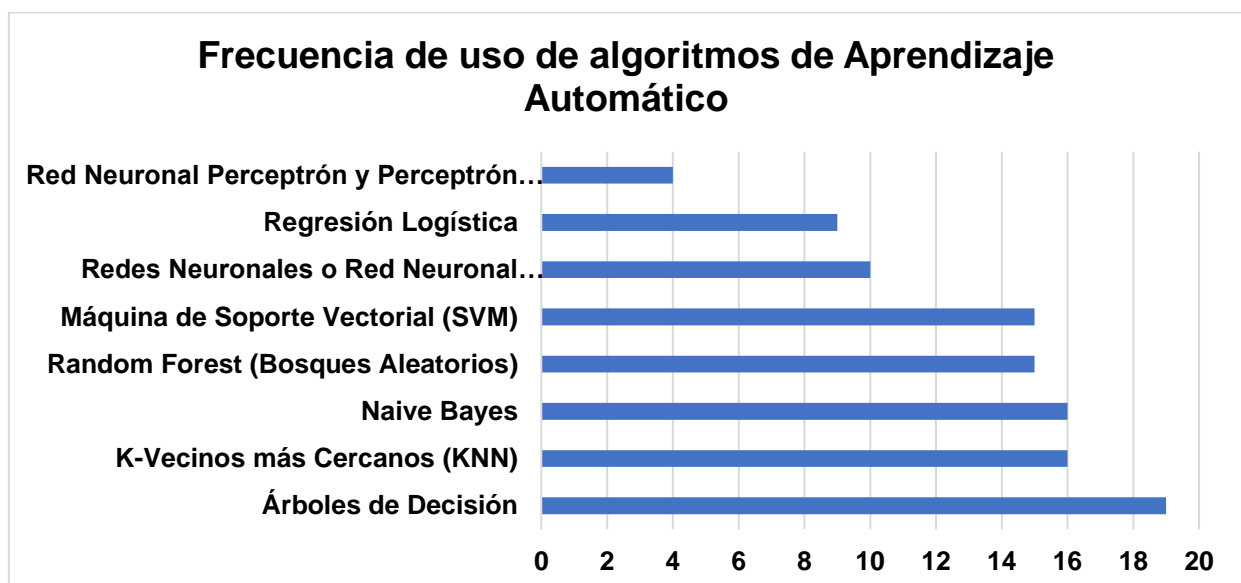
Algoritmos más utilizados para la predicción del rendimiento académico.

Algoritmo	Tipo de algoritmo	Número de veces
Árboles de Decisión	Aprendizaje Supervisado	19
K-Vecinos más Cercanos (KNN)	Aprendizaje Supervisado	16
Naive Bayes	Aprendizaje Supervisado	16
Random Forest (Bosques Aleatorios)	Aprendizaje Supervisado	15
Máquina de Soporte Vectorial (SVM)	Aprendizaje Supervisado	15
Redes Neuronales o Red Neuronal Artificial (ANN)	Aprendizaje Profundo	10
Regresión Logística	Aprendizaje Supervisado	9
Red Neuronal Perceptrón y Perceptrón Multicapa	Aprendizaje Profundo	4

Con el fin de ofrecer una representación visual más clara de esta información, se presenta un gráfico que facilita la identificación de los algoritmos más comúnmente empleados en la predicción del rendimiento académico de estudiantes universitarios. Dicho gráfico se encuentra detallado en la Figura 4.

Figura 4

Frecuencia de uso de algoritmos de Aprendizaje Automático para la Predicción del Rendimiento Académico



DISCUSIÓN

De los artículos revisados sobre la aplicación de algoritmos de Aprendizaje Automático para predecir el rendimiento académico de estudiantes universitarios, se evaluaron y compararon tanto algoritmos de Aprendizaje Supervisado como algoritmos de Aprendizaje Profundo. Entre los algoritmos de Aprendizaje Supervisado más ampliamente empleados se encuentran: Árboles de Decisión, K-Vecinos más Cercanos (KNN), Naive Bayes, Random Forest (Bosques Aleatorios) y Máquinas de Soporte Vectorial (SVM), junto con algunas de sus variantes consideradas en el estudio. Los algoritmos de Aprendizaje Profundo incluyeron los siguientes: Red Neuronal Perceptrón, Red Neuronal Perceptrón Multicapa y Redes Neuronales o Red Neuronal Artificial (ANN).

Según la revisión realizada, los algoritmos de Aprendizaje Supervisado reconocidos por su precisión en la predicción del rendimiento académico en entornos universitarios son: Random Forest (Bosques Aleatorios), Árboles de Decisión y Naives Bayes. Y de los algoritmos de Aprendizaje Profundo, tanto las Redes Neuronales Perceptrón como las Redes Neuronales Artificiales (ANN) han demostrado una notable precisión.

Lo resultados de este artículo de revisión coinciden en gran medida con los hallazgos de García (2020), quien señala que los algoritmos más utilizados para predecir el rendimiento académico son el Naive Bayes, las Redes Neuronales, los Árboles de Decisión y las Máquina de Soporte Vectorial.

Es importante destacar que durante la revisión no se encontraron artículos que presentaran resultados basados en el uso de algoritmos de Aprendizaje No Supervisado en la predicción del rendimiento académico de estudiantes universitarios. Según Serrato (2021), esto se debe a que estos algoritmos se utilizan principalmente para descubrir patrones ocultos en los datos en lugar de realizar predicciones directas. Además, los resultados de estos algoritmos suelen ser difíciles de interpretar en términos de variables específicas que influyen en el rendimiento académico.

En esta tarea, comprender las razones detrás del rendimiento de los estudiantes es fundamental para intervenir de manera efectiva. Por lo general, esto se logra de manera más eficaz mediante el uso de algoritmos de Aprendizaje Supervisado, que permiten establecer relaciones directas entre las variables y las predicciones, ofreciendo una comprensión más clara de los factores que impactan en el rendimiento académico.

CONCLUSIÓN

Los artículos revisados resaltan el predominio de algoritmos de Aprendizaje Supervisado en la tarea de predecir el rendimiento académico de estudiantes universitarios. Además, se observa la presencia de algoritmos de Aprendizaje Profundo en este ámbito.

El uso de algoritmos de Aprendizaje No Supervisado para predecir el rendimiento académico de estudiantes universitarios es menos común, en

comparación con los algoritmos de Aprendizaje Supervisado y Aprendizaje Profundo, puesto que generalmente se prefieren los algoritmos de Aprendizaje Supervisado por su orientación hacia la realización de predicciones y por la necesidad de obtener resultados precisos y de fácil interpretación.

Es evidente que los algoritmos de Aprendizaje Automático pueden considerar una amplia variedad de variables, como historiales académicos, factores socioeconómicos, características personales y datos de comportamiento, entre otros, para lograr una precisión significativamente mayor en la predicción del rendimiento de los estudiantes. A través del análisis de datos históricos de estudiantes y variables relevantes relacionadas con los factores que influyen en el rendimiento académico, estos algoritmos pueden identificar y capturar información compleja que va más allá de los métodos estadísticos tradicionales, generando así modelos predictivos altamente precisos.

Es fundamental señalar que la aplicación de algoritmos de Aprendizaje Automático en la predicción del rendimiento académico tiene un gran potencial para contribuir a la reducción de la tasa de deserción, mejorar los resultados educativos, identificar tempranamente a los estudiantes en riesgo de fracaso y respaldar la toma de decisiones en la enseñanza, así como optimizar los recursos educativos. Por lo tanto, es crucial continuar profundizando en la investigación sobre la predicción del rendimiento académico para explorar otros factores que puedan influir en el éxito de los estudiantes.

Con un enfoque equilibrado y una implementación adecuada, los algoritmos de Aprendizaje Automático tienen el potencial de generar mejoras significativas en la educación superior y en el logro académico de los estudiantes.

REFERENCIAS BIBLIOGRÁFICAS

- Abdelmagid, A. S., & Qahmash, A. I. (2023). Utilizing the Educational Data Mining Techniques "Orange Technology" for Detecting Patterns and Predicting Academic Performance of University Students. *Information Sciences Letters*, 17. doi:<http://dx.doi.org/10.18576/isl/120330>
- Ahmad, N., Hassan, N., Jaafar, H., & Enzai, N. (2021). Students' Performance Prediction using Artificial Neural Network. *IOP Conference Series: Materials Science and Engineering*, 9. doi:<https://doi.org/10.1088/1757-899X/1176/1/012020>
- Al-Alawi, L., Al Shaqsi, J., Tarhini, A., & Al-Busaidi, A. (2023). Using machine learning to predict factors affecting academic performance: the case of college students on academic probation. *Education and Information Technologies*, 26. doi:<https://doi.org/10.1007/s10639-023-11700-0>
- Albreiki, B., Zaki, N., & Alashwal, H. (2021). A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques. *Education Sciences*. doi:<https://doi.org/10.3390/educsci11090552>

- Alsubaie, M. (2023). Predicting student performance using machine learning to enhance the quality assurance of online training via Maharat platform. *Alexandria Engineering Journal*, 323-339.
- Alturki, S., & Alturki, N. (2021). Using Educational Data Mining to Predict Student's Academic Performance for Applying Early Interventions. *Journal of Information Technology Education: Innovations in Practice*, 17. doi:<https://doi.org/10.28945/4835>
- Alturki, S., Cohausz, L., & Stuckenschmidt, H. (2022). Predicting Master's students' academic performance: an empirical study in Germany. *Smart Learning Environments*, 22. doi:<https://doi.org/10.1186/s40561-022-00220-y>
- Ayala Franco, E. (2022). Sistema predictivo de rendimiento escolar de alumnos universitarios de primer ingreso. México: Dirección General de Bibliotecas y Servicios Digitales de Información. Obtenido de <https://ring.uaq.mx/handle/123456789/3866>
- Balaji, P., Alelyani, S., Qahmash, A., & Mohana, M. (2021). Contributions of Machine Learning Models towards Student Academic Performance Prediction: A Systematic Review. *Applied Sciences*. doi:<https://doi.org/10.3390/app112110007>
- Bedregal Alpaca, N., Tupacyupanqui Jaén, D., & Cornejo Aparicio, V. (2020). Análisis del rendimiento académico de los estudiantes de Ingeniería de Sistemas, posibilidades de deserción y propuestas para su retención. *Ingeniare. Revista chilena de ingeniería.*, 28(4), 668-683. Obtenido de https://www.scielo.cl/scielo.php?pid=S0718-33052020000400668&script=sci_arttext
- Buenaño-Fernández, D., Gil, D., & Luján-Mora, S. (2019). Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study. *Sustainability*. doi:<https://doi.org/10.3390/su11102833>
- Candia Oviedo, D. (2019). Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático. Cusco, Cusco, Perú. Obtenido de <https://repositorio.unsaac.edu.pe/handle/20.500.12918/4120>
- Carles Vega, J. (2023). Modelización de los factores que inciden en el rendimiento académico de los estudiantes universitarios con técnicas de estadística multivariante y de machine learning. Valencia, España: Universitat Politècnica de València. Obtenido de <http://hdl.handle.net/10251/192774>
- Castrillón, O. D., Sarache, W., & Ruiz Herrera, S. (2020). Predicción del rendimiento académico por medio de técnicas de inteligencia artificial. *Formación universitaria*, 13(1), 93-102. Obtenido de https://www.scielo.cl/scielo.php?pid=S0718-50062020000100093&script=sci_arttext

- Chandana, D. A., Sundaram, M., Bhumika, S., & Meghana, B. N. (2022). Application of Machine Learning Techniques to Evaluate the Performance of Students in an Academic Environment. *Saudi Journal of Engineering and Technology*, 327-334. doi:<https://doi.org/10.36348/sjet.2022.v07i06.008>
- Chavez, H., Chavez-Arias, B., Contreras-Rosas, S., Alvarez-Rodríguez, J., & Raymundo, C. (2023). Artificial neural network model to predict student performance using nonpersonal information. *Frontiers in Education*, 11. doi:<https://doi.org/10.3389/feduc.2023.1106679>
- Chen, C.-H., Yang, S., Weng, J.-X., Ogata, H., & Su, C.-Y. (2021). Predicting at-risk university students based on their e-book reading behaviours by using machine learning classifiers. *Australasian Journal of Educational Technology*, 15. doi:<https://doi.org/10.14742/ajet.6116>
- Contreras Bravo, L., Fuentes López, H., & Rivas Trujillo, E. (2021). Análisis del rendimiento académico mediante técnicas de aprendizaje automático con métodos de ensamble. *Revista Boletín REDIP*, 10(13), 171-190. Obtenido de <https://revista.redipe.org/index.php/1/article/view/1737/1652>
- Contreras, L., Fuentes, H., & Rodríguez, J. (2020). Predicción del rendimiento académico como indicador de éxito/fracaso de los estudiantes de ingeniería, mediante aprendizaje automático. *Formación universitaria*, 13(5), 233-246. Obtenido de https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-50062020000500233
- Contreras, L., Nieves, N., & Gonzalez, K. (2023). Predicción del rendimiento académico universitario mediante mecanismos de aprendizaje automático y métodos supervisados. *Ingeniería*, 28. doi:<https://doi.org/10.14483/23448393.19514>
- Contreras, L., Tarazona, G., & Alemán, A. (2023). *Machine Learning aplicado al rendimiento académico en educación superior: factores, variables y herramientas*. Bogotá: Editorial Universidad Distrital Francisco José de Caldas.
- Cruz, E., González, M., & Rangel, J. (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*, 13(1), 77-87. Obtenido de <https://doi.org/10.33412/pri.v13.1.3039>
- Díaz Ramírez, J. (2021). Aprendizaje Automático y Aprendizaje Profundo. *Ingeniare. Revista chilena de ingeniería*, 29(2), 180-181. Obtenido de https://www.scielo.cl/scielo.php?pid=S0718-33052021000200180&script=sci_arttext
- Díaz, B., Meleán, R., & Marín, W. (2021). Rendimiento académico de estudiantes en Educación Superior: predicciones de factores influyentes

- a partir de árboles de decisión. *Telos*, 23(3). Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=8133982>
- Ekubo, E., & Esiefarienrhe, B. (2022). Using machine learning to predict low academic performance at a Nigerian university. *The African Journal of Information and Communication*, 33.
doi:<https://doi.org/10.23962/ajic.i30.14839>
- El-Keiey, S., ElMenshawy, D., & Hassanein, E. (2022). Student's Performance Prediction based on Personality Traits and Intelligence Quotient using Machine Learning. *International Journal of Advanced Computer Science and Applications*, 8.
- Enughwure, A., & Ogbise, M. (2020). Application of Machine Learning Methods to Predict Student Performance: A Systematic Literature Review. *International Research Journal of Engineering and Technology (IRJET)*, 3405-3415.
- Espejel García, M., & Jiménez García, M. (2019). Nivel educativo y ocupación de los padres: Su influencia en el rendimiento académico de estudiantes universitarios. *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo.*, 10(19). Obtenido de https://www.scielo.org.mx/scielo.php?pid=S2007-74672019000200026&script=sci_arttext
- Fonseca, A., & López Hernández, D. (2014). Criterios para publicar artículos de revisión sistemática. *Revista de Especialidades Médico-Quirúrgicas*, 19(3), 393-399.
- Gamboa Unsihuay, J. E., & Salinas Flores, J. W. (2022). Predicción de la situación académica en estudiantes de pregrado utilizando algoritmos de aprendizaje automático. *Perfiles*, 1(27), 4-10. Obtenido de <http://ceaa.esepoch.edu.ec:8080/revista.perfiles/faces/Articulos/Perfiles27Art1.pdf>
- García Peredo, L. J. (2020). Revisión sistemática sobre la predicción del rendimiento académico en estudiantes: técnicas y algoritmos. Chiclayo, Lambayeque, Perú. Recuperado el 5 de septiembre de 2023, de https://tesis.usat.edu.pe/bitstream/20.500.12423/3635/1/TIB_GarciaPeredoLuisJesus.pdf
- Gerón, A. (2020). *Aprende Machine Learning con Scikit-Learn, Keras y TensorFlow* (segunda edición ed.). Madrid: Anaya Multimedia.
- Gerson Yovanni, O. M. (2019). Aplicación de Data Science para la Predicción del Rendimiento Académico de los Estudiantes de la Facultad de Ingeniería de Sistemas de la Universidad Nacional del Centro del Perú. Huancayo, Junín, Perú. Obtenido de <https://repositorio.uncp.edu.pe/handle/20.500.12894/5837>
- Ghazal, A. (2020). Using Machine Learning to Support Students' Academic Decisions. *Journal of Theoretical and Applied Information Technology*.

- Ghosh, S., J. F., & Ahmad, I. (2022). Application of the Classification Algorithms on the Prediction of Student's Academic Performance. *Trends in Sciences*, 16. doi:<https://doi.org/10.48048/tis.2022.5070>
- González, L. (3 de Agosto de 2018). *AprendeIA*. Recuperado el 30 de mayo de 2023, de Todo sobre aprendizaje supervisado en Machine Learning: <https://aprendeia.com/todo-sobre-aprendizaje-supervisado-en-machine-learning/>
- González, L. (15 de Junio de 2018). *AprendeIA*. Recuperado el 9 de Diciembre de 2023, de Diferencia entre aprendizaje supervisado y no supervisado: <https://aprendeia.com/diferencia-entre-aprendizaje-supervisado-y-no-supervisado/>
- González, L. (11 de Febrero de 2020). *AprendeIA*. Recuperado el 5 de Junio de 2023, de Aprendizaje no Supervisado: <https://aprendeia.com/aprendizaje-no-supervisado-machine-learning/>
- Gutiérrez Villaverde, H. E., Linares Barbero, M., Agüero Correa, Á. A., & Pérez Nuñez, J. R. (2022). Predicción de rendimiento académico de alumnos usando machine learning. Lima: Repositorio Institucional Universidad de Lima. Obtenido de <https://repositorio.ulima.edu.pe/handle/20.500.12724/16703>
- IBM. (2023). *¿Qué es el aprendizaje no supervisado?* Recuperado el 30 de mayo de 2023, de https://www.ibm.com/mx-es/topics/unsupervised-learning?mhsrc=ibmsearch_a&mhq=aprendizaje%20supervisado
- IBM. (2023). *¿Qué es el aprendizaje supervisado?* Recuperado el 30 de mayo de 2023, de https://www.ibm.com/es-es/topics/supervised-learning?mhsrc=ibmsearch_a&mhq=aprendizaje%20supervisado
- Incio Flores, F., Capuñay Sanchez, D., & Estela Urbina, R. (2023). Modelo de red neuronal artificial para predecir resultados académicos en la asignatura Matemática II. *Revista Electrónica Educare*, 27(1), 338-359. Obtenido de https://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S1409-42582023000100338&lang=es
- Issah, I., Appiah, O., Appiahene, P., & Inusah, F. (2023). A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. *Decision Analytics Journal*. doi:<https://doi.org/10.1016/j.dajour.2023.100204>
- Jiménez García, M. (2018). Análisis cuantitativo de las variables que influyen en el rendimiento universitario. *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 9(17), 623-638.
- Lagares, J. A., Díaz, N., & Barranco, C. (2022). Aprendizaje profundo: una nueva vía para convertir el dato en conocimiento. *Economía industrial*(423), 25-38.

- Lunghi, C. G. (8 de Noviembre de 2022). *Repositorio Institucional de la UNLP*. Recuperado el 15 de agosto de 2023, de <http://sedici.unlp.edu.ar/handle/10915/145726>
- Meghji, A., Mahoto, N., Asiri, Y., Alshahrani, H., Sulaiman, A., & Shaikh, A. (2023). Early detection of student degree-level academic performance using educational data mining. *PeerJ Computer Science*, 29. doi:<http://doi.org/10.7717/peerj-cs.1294>
- Miranda, M., & Guzmán, J. (2017). Análisis de la Deserción de Estudiantes Universitarios usando Técnicas de Minería de Datos. *Formación universitaria*, 10(3), 61-68. Obtenido de https://www.scielo.cl/scielo.php?pid=S0718-50062017000300007&script=sci_arttext&tlng=pt
- Mora García, R. (2015). Factores que intervienen en el rendimiento académico universitario: Un estudio de caso. *Opción*, 31(6), 1041-1063.
- Moreno, A., Armengol, E., Béjar L. Belanche, E., Cortés, U., Gavaldá, R., López, B., & M. Sánchez, M. (1994). *Aprendizaje automático*. Barcelona: Edicions UPC.
- Norman, A. (2019). *Aprendizaje Automático en acción: Un libro para el lego, guía paso a paso para los novatos*. Litres.
- Ojajuni, O., Ayeni, F., Akodu, O., Ekanoye, F., Adewole, S., Ayo, T., . . . Mbarika, V. (2021). Predicting Student Academic Performance Using Machine Learning. *Computational Science and Its Applications – ICCSA*, 481-491. doi:https://doi.org/10.1007/978-3-030-87013-3_36
- Oliver, J. (2007). Factores asociados al rendimiento académico. *Educación*, 31(1), 43-63.
- Pineda Pertuz, C. (2022). *Aprendizaje automático y profundo en Python*. Bogotá: Ra-Ma.
- Ponce Guizabalo, S. (1 de Abril de 2024). Modelo de aprendizaje automático para la predicción del rendimiento académico de los alumnos ingresantes en la facultad de Ingeniería en Industrias Alimentarias de la UNAS. Tingo María, Leoncio Prado, Perú. Obtenido de <https://repositorio.unas.edu.pe/items/d92d7ad5-89d4-4ba7-af91-c8c963bcc421>
- Preciado, G., Huerta, J., Vera, Á., & Corral, A. (2022). Causas Asociadas a La Deserción Escolar en Educación Superior. Una Revisión Sistemática Del 2010 Al 2020. *Ra Ximhai*, 18(1), 83-101. Obtenido de <https://doi.org/10.35197/rx.18.01.2022.04.jp>
- Rico Páez, A. (2022). Modelos predictivos progresivos del rendimiento académico de estudiantes universitarios. *Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 12(24).
- Rico Páez, A., & Gaytán Ramírez, N. D. (2022). Predicción del rendimiento académico utilizando las primeras actividades académicas de

estudiantes universitarios y técnicas de aprendizaje automático. *Revista Dilemas Contemporáneos: Educación, Política y Valores*(3).
doi:<https://doi.org/10.46377/dilemas.v9i3.3177>

- Rico Páez, A., Gaytán Ramírez, N. D., & Sánchez Guzmán, D. (2019). Construcción e implementación de un modelo para predecir el rendimiento académico de estudiantes. *Diálogos sobre educación. Temas actuales en investigación educativa*, 10(19).
- Saltos Mero, J., & Cruz Felipe, M. (2024). Análisis del rendimiento académico de estudiantes de las carreras Economía y Turismo con Power BI en los periodos (2021). *593 Digital Publisher CEIT*, 9(1), 762-772. Obtenido de https://www.593dp.com/index.php/593_Digital_Publisher/article/view/2162
- Sandoval, L. J. (2018). Algoritmos de Aprendizaje Automático para Análisis y Predicción de Datos. *Revista Tecnológica*(11), 36-40. Obtenido de <http://redicces.org.sv/jspui/handle/10972/3626>
- Sandra, L., Lumbangaol, F., & Matsuo, T. (2021). Machine Learning Algorithm to Predict Student's Performance: A Systematic Literature Review. *TEM Journal*, 1919-1927. doi:<https://doi.org/10.18421/TEM104-56>
- Santamaría, T., Patiño, D., González, V., & Flores, L. (2023). Implementación de técnicas de machine learning y creación de una red neuronal artificial para la predicción del rendimiento académico de estudiantes en ambientes universitarios que usan e-learning y streaming. *DYNA - Ingeniería e Industria*, 98(3), 282-287. Obtenido de <https://web.s.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=1&sid=51437559-d08c-4cff-82e6-2b1c30a9437c%40redis>
- Serrato, C. (25 de marzo de 2021). *INMEDIATUM*. Recuperado el 5 de junio de 2023, de <https://inmediatum.com/blog/estrategia/diferencia-entre-aprendizaje-supervisado-y-no-supervisado/>
- Soria Olivas, E., Sánchez, M., Gamero Cruz, R., Castillo Caballero, B., & Cano Michelena, P. (2023). *Sistemas de Aprendizaje Automático*. Madrid: RA-MA.
- Urteaga , I., Siri, L., & Garófalo, G. (2020). Predicción temprana de deserción mediante aprendizaje automático en cursos profesionales en línea. *RIED. Revista Iberoamericana de Educación a Distancia*, 23(2), 147–167. Obtenido de <https://www.redalyc.org/journal/3314/331463171008/>
- Valero, J., Navarro, Á., Larios, A., & Julca, J. (2022). Deserción universitaria evaluación de diferentes algoritmos de Machine Learning para su predicción. *Revista de ciencias sociales*, 28(3), 362-375. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=8526463>
- Yagci, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 19. doi:<https://doi.org/10.1186/s40561-022-00192-z>

Yepes Piqueras, V. (2022). El aprendizaje profundo (deep learning) en la optimización de estructuras. Valencia: Universitat Politècnica de València. Obtenido de <https://victoryepes.blogs.upv.es/2020/09/15/el-aprendizaje-profundo-deep-learning-en-la-optimizacion-de-estructuras/>

Zhang, Y., Yun, Y., An, R., Cui, J., Dai, H., & Shang, X. (2021). Educational Data Mining Techniques for Student Performance Prediction: Method Review and Comparison Analysis. *Frontiers in Psychology*.