

Análisis de algoritmos de aprendizaje automático para evaluar el rendimiento académico de estudiantes en la asignatura de matemáticas

Analysis of machine learning algorithms to evaluate the academic performance of students in the subject of mathematics

Beatriz Angelic Paredes Quispe

<https://orcid.org/0000-0002-1094-6831>

beatriz.paredes2@unmsm.edu.pe

Universidad Nacional Mayor de San Marcos,
Lima, Perú

RECIBIDO: 15/10/2024 - ACEPTADO: 05/12/2024 - PUBLICADO: 31/12/2024

RESUMEN

El rendimiento académico de los estudiantes es crucial para el desarrollo educativo y social de un país, ya que impacta directamente en las oportunidades futuras de los jóvenes y en el progreso colectivo de la sociedad. En particular, las matemáticas desempeñan un papel fundamental en la educación básica, aunque muchos estudiantes enfrentan dificultades que afectan su desempeño académico y puede limitar su desarrollo a largo plazo. Este desafío se refleja en los resultados del Programa para la Evaluación Internacional de Alumnos (PISA), que evidencian las brechas en el aprendizaje de matemáticas a nivel global. La presente investigación busca identificar el algoritmo de aprendizaje automático más preciso para predecir el rendimiento académico de los estudiantes de secundaria en matemáticas. El enfoque adoptado es cuantitativo, con un diseño no experimental y de tipo aplicado. La muestra incluyó 400 estudiantes, cuyos datos fueron obtenidos de un repositorio público de la Universidad de California de Irvine. Los datos recolectados abarcaron factores demográficos, académicos, psicológicos y sociales, entre otros, sumando un total de 30 variables clave. Estos datos fueron utilizados para entrenar y evaluar diversos algoritmos de aprendizaje automático con el fin de identificar los factores más determinantes en el proceso de aprendizaje. El algoritmo más preciso fue la Red neuronal densa, alcanzando un 74% de precisión, demostrando su efectividad para predecir el rendimiento académico. Los resultados revelaron una relación positiva y significativa entre las variables analizadas y el rendimiento académico, sugiriendo que mejoras en estas variables podrían aumentar el promedio académico de los estudiantes.

Palabras claves: Rendimiento Académico, Algoritmo de Predicción, Aprendizaje Automático, Precisión.

ABSTRACT

The academic performance of students is crucial for the educational and social development of a country, as it directly impacts the future opportunities of young people and the collective progress of society. Mathematics plays a fundamental role in basic education, yet many students face difficulties that affect their academic performance and can limit their long-term development. This challenge is reflected in the results of the Program for International Student Assessment (PISA), which highlights the gaps in mathematics learning globally. The present research seeks to identify the most accurate machine learning algorithm for predicting the academic performance of high school students in mathematics. The approach adopted is quantitative, with a non-experimental, applied-type design. The sample included 400 students, whose data were obtained from a public repository at the University of California, Irvine. The data collected covered demographic, academic, psychological and social factors, among others, totaling 30 key variables. These data were used to train and evaluate various machine learning algorithms to identify the most important determinants in the learning process. The most accurate algorithm was the Dense Neural Network, reaching 74% accuracy, demonstrating its effectiveness in predicting academic performance. The results revealed a positive and significant relationship between the variables analyzed and academic performance, suggesting that improvements in these variables could increase the students' academic average.

Keywords: Academic Performance, Prediction Algorithm, Machine Learning, Accuracy.

I. INTRODUCCIÓN

El rendimiento académico se entiende como el resultado cuantitativo de la comprensión de los contenidos de los programas de estudio, representado por medio de las calificaciones dentro de una escala convencional, que se genera del proceso de enseñanza aprendizaje (Calva et al., 2021). Según lo afirmado por Páez y Gaytán (2022), el rendimiento académico puede medirse en diferentes fases y recopilarse en informes y sistemas informáticos, permitiendo predecir resultados y tomar decisiones para mejorar el aprendizaje.

Tal como menciona Garbanzo (2013), el rendimiento académico es el resultado de la combinación de diversos factores multicausales que incluyen aspectos sociodemográficos, psicosociales, pedagógicos, institucionales y socioeconómico. El autor también señala que entre los factores que influyen en el rendimiento académico se encuentran aspectos como la motivación, la ansiedad, el entusiasmo y la autoestima de los estudiantes, así como las características del docente y la percepción del clima académico, todos los cuales están fuertemente influenciados por las decisiones tomadas en el ámbito institucional.

De acuerdo con Siquiera y Gurge (2011), el bajo rendimiento escolar se puede definir como un logro escolar inferior al esperado según la edad, las habilidades cognitivas y la enseñanza recibida por un determinado estudiante. Los autores también señalan que, más allá de las causas intrínsecas (individuales), este problema genera dificultades emocionales, como baja autoestima y desmotivación. Además, las causas extrínsecas (ambientales) incluyen factores como la preocupación familiar, que

tienen repercusiones a nivel individual, familiar, educativo y social (pp. 79-80).

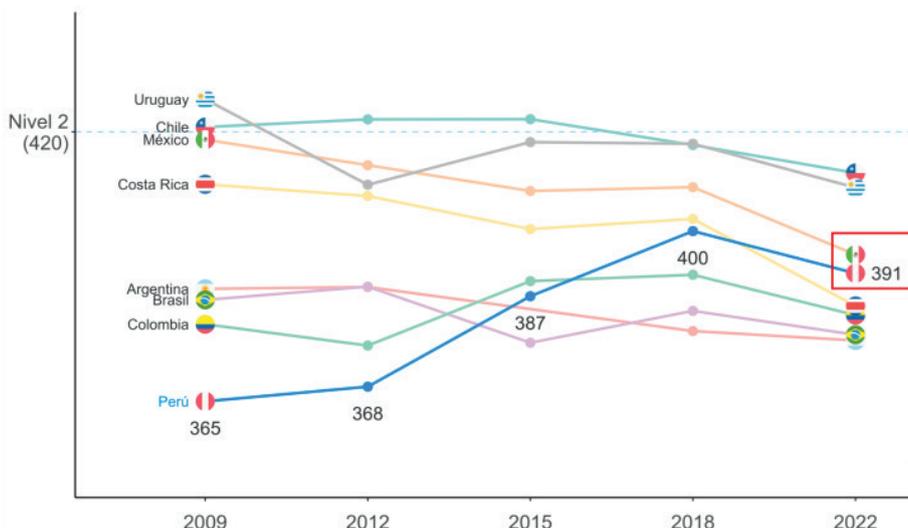
El bajo rendimiento académico en matemáticas se refleja en los resultados de la evaluación PISA, donde, entre 2018 y 2022, 44 de los 72 países participantes experimentaron una disminución en su promedio de puntos en la asignatura de matemáticas (MINEDU, 2024). Como se puede visualizar en la figura 1, en el caso de Perú, los resultados de la evaluación PISA indicaron que el país no alcanzó el Nivel 2, el cual constituye la puntuación mínima requerida para desenvolverse de manera eficaz en la sociedad contemporánea.

Según lo expuesto por Ghorbani y Ghousi (2020), diversos factores tienen un impacto en la mejora de la precisión de las predicciones, y la correcta selección de estos factores es crucial para generar un modelo eficiente. De acuerdo con Estrada et al. (2016), al implementar técnicas de aprendizaje automático para crear modelos predictivos, se pueden anticipar los rendimientos académicos e identificar los factores que influyen en el proceso de enseñanza y aprendizaje. No obstante, la aplicación de estas metodologías en este campo ha sido relativamente reciente en los países de Latinoamérica.

El aprendizaje automático proporciona diferentes técnicas o algoritmos para predecir situaciones con grandes cantidades de información que, a través de un buen procesamiento y filtrado de datos, pueden generar predicciones muy efectivas (Forero y Negre, 2024). Según lo expuesto por Romero y Ventura (2010), para realizar la predicción, se han analizado datos para desarrollar modelos predictivos mediante técnicas de aprendizaje automático, dado que las herramientas estadísticas tradicionales

Figura 1

Ministerio de Educación (2024). Presentación de resultados PISA 2022.



resultan ineficaces al abordar grandes volúmenes de datos y múltiples variables relacionadas con los estudiantes. El autor también señala que estas técnicas se enfocan en el manejo de datos para generar resultados que se presentan como decisiones, siendo especialmente útiles en el desarrollo de modelos predictivos.

De acuerdo con la investigación de Salal et al. (2019), diversos estudios han empleado técnicas de aprendizaje automático para desarrollar modelos predictivos que, utilizando algoritmos como Naïve Bayes, árboles de decisión y otros, permiten anticipar el rendimiento académico de los estudiantes. Estos avances proporcionan herramientas valiosas que facilitan la toma de decisiones en el ámbito educativo (Bravo et al., 2021).

Como se mencionó en los estudios previos, la identificación de los factores que influyen en el rendimiento académico es esencial para desarrollar modelos predictivos eficaces. En este contexto, las tecnologías de aprendizaje automático, y particularmente las redes neuronales, han demostrado ser herramientas poderosas para modelar y predecir el rendimiento académico. Las redes neuronales artificiales (RNA) tienen el potencial de facilitar este tipo de análisis debido a su destacado desempeño en tareas de predicción y clasificación. De este modo, es posible prever el rendimiento académico futuro de los estudiantes, considerando una variedad de factores socioeducativos (Longoni et al., 2010).

La mejora del rendimiento académico es una de las principales preocupaciones en el ámbito educativo, especialmente en asignaturas críticas como

matemáticas. Con el avance de las tecnologías y la disponibilidad de grandes volúmenes de datos, los algoritmos de aprendizaje automático han emergido como herramientas poderosas para analizar y predecir el rendimiento académico de los estudiantes. Sin embargo, la efectividad de estos algoritmos puede variar significativamente dependiendo de su diseño y aplicación.

Este estudio tiene como objetivo general identificar qué algoritmo de aprendizaje automático ofrece la mayor precisión en la predicción del rendimiento académico de los estudiantes de secundaria en la asignatura de matemáticas. A través de la identificación y comparación de diferentes algoritmos, se busca no solo responder esta pregunta, sino también contribuir a una comprensión más profunda de cómo los métodos de análisis predictivo pueden utilizarse para optimizar el proceso educativo y ofrecer intervenciones más efectivas.

Para abordar este objetivo, se analizarán diversos algoritmos predictivos, incluidos la Regresión Logística, Árbol de decisión y la Red Neuronal Densa. Estos modelos serán evaluados utilizando métricas clave como la precisión, exactitud, sensibilidad, F1 score y especificidad.

Se espera que los resultados contribuyan a una comprensión más profunda de las herramientas analíticas disponibles para predecir el rendimiento académico, proporcionando una base sólida para la selección de los modelos más efectivos en este contexto. Así, el estudio busca proporcionar a los docentes herramientas basadas en datos para implementar estrategias pedagógicas más efectivas

que apoyen a los estudiantes con mayores necesidades, promoviendo un aprendizaje más personalizado y mejorando los resultados académicos en matemáticas.

En la sección dos se presentan los algoritmos predictivos, y en la tercera sección se muestran los resultados de la evaluación de cada algoritmo y se muestran los resultados de las métricas clave. Finalmente, en la última sección se ofrecen las conclusiones.

1.1. Objetivos de investigación

Esta investigación tiene como objetivo general:

1. Determinar el modelo de aprendizaje automático que ofrece la mayor precisión en la predicción del rendimiento académico en matemáticas de los estudiantes de secundaria en matemáticas.

Para alcanzar este objetivo general, se han establecido los siguientes objetivos específicos:

1. Identificar los atributos predictivos más relevantes que afectan significativamente en la predicción del rendimiento académico en matemáticas para estudiantes de secundaria.
2. Determinar técnicas de preprocesamiento de datos para mejorar la predicción del rendimiento académico en matemáticas para estudiantes de secundaria.
3. Validar la precisión de los algoritmos de aprendizaje automático para determinar su efectividad en la predicción del rendimiento académico en matemáticas de los estudiantes de secundaria.

1.2. Definición de variables

1.2.1 Variable Independiente: Aprendizaje automático

El Aprendizaje automático, o Machine Learning, es una disciplina dentro de las ciencias de la computación y, específicamente, una rama de la inteligencia artificial. Este concepto fue introducido en 1959 por el informático estadounidense Arthur Samuel y se refiere a las técnicas orientadas al diseño de algoritmos que permiten a las máquinas aprender. Se considera que una máquina ha aprendido cuando, a partir de su experiencia, mejora en la realización de la tarea que se le ha asignado. Así, es posible

abordar problemas sin necesidad de programarlos de manera específica, ya que la máquina puede desarrollar una lógica propia (Arán, 2022).

1.2.2. Variable Dependiente: Rendimiento académico

Según Bricklin (2007), afirma que el rendimiento académico hace referencia a la evaluación del conocimiento adquirido en el ámbito escolar terciario o universitario. Un estudiante con buen rendimiento académico es aquel que obtiene calificaciones positivas en los exámenes que debe de rendir a lo largo de una asignatura cursada.

1.3. Trabajos relacionados

De acuerdo con Vega (2019), en su tesis de maestría “Modelo de pronóstico de rendimiento académico de alumnos en los cursos de programas de estudios básicos de la Universidad Ricardo Palma utilizando algoritmos de Machine Learning”. El acceso a la educación se considera un derecho fundamental con profundas repercusiones en las oportunidades laborales, especialmente para aquellos estudiantes con un rendimiento académico sobresaliente. Esta investigación tiene como objetivo principal utilizar algoritmos de aprendizaje automático para identificar a los estudiantes con alta probabilidad de obtener un bajo rendimiento en los 13 cursos del Programa de Estudios Básicos de la Universidad Ricardo Palma en Perú. La intención es implementar estrategias que mejoren los resultados académicos, detectando áreas de mejora y contribuyendo así al avance de la sociedad. El estudio se enfoca en la implementación, análisis y comparación de tres algoritmos: Redes Neuronales Artificiales (RNA), Gradient Boosting Machine (GBM) y XGBoosting, con el propósito de predecir el rendimiento académico mediante la anticipación del número de estudiantes aprobados y desaprobados en cada curso.

Según la investigación de Gil y Quintero (2021), en su trabajo titulado “Predicción del rendimiento académico estudiantil con redes neuronales artificiales”, se propone un enfoque que emplea una red neuronal artificial (RNA) para predecir el rendimiento académico de los estudiantes, subrayando su habilidad para procesar información y detectar relaciones no lineales entre variables. La investigación se fundamenta en datos de 395 estudiantes colombianos de educación media vocacional en la “Institución Educativa Villa del Socorro”, en Medellín. La RNA logró clasificar correctamente el 73% de las muestras, superando otras técnicas en métricas

como exactitud, recuperación, precisión y puntuación F1. La conclusión resalta que anticipar el rendimiento académico permite diseñar estrategias pedagógicas que mejoren el proceso de enseñanza y aprendizaje.

Tal como menciona Salgado et al. (2019), en su estudio titulado "Modelo para predecir el rendimiento académico basado en redes neuronales y analítica de aprendizaje", donde investigación describe un modelo predictivo de rendimiento académico usando técnicas de redes neuronales sobre un conjunto de datos reales de 300 estudiantes y las variables predictoras están basadas en un entorno virtual de aprendizaje estas variables se procesaron, utilizando el paquete "nnet" de R, que permite la creación, aplicación, visualización e implementación de redes neuronales. El diseño de la red neuronal utilizada se basó en una topología de red neuronal multicapa que estuvo determinada por una capa de entrada y una capa intermedia u oculta con funciones de activación sigmoidea tangente hiperbólica y una capa de salida con función de activación lineal, esta implementación de redes neuronales obtuvo una certeza del 75%.

De acuerdo con Lizares (2017), en su investigación titulada "Comparación de modelos de clasificación: regresión logística y árboles de clasificación para evaluar el rendimiento académico", señala que los árboles de decisión tienen una mayor efectividad al momento de clasificar y predecir. Además, concluyó lo siguiente: en primer lugar, los factores que más influyeron en el rendimiento académico fueron el tipo de colegio, las horas de estudio y el interés por la asignatura; en segundo lugar, los factores que tuvieron menos influencia fueron el sexo y la formación profesional de los padres. Finalmente, destacó que se obtuvo un 92.34% de probabilidad de reprobación en estudiantes provenientes de instituciones educativas nacionales, que no dedican más de tres horas al estudio y que muestran poco interés por el curso.

II. MATERIAL Y MÉTODO

2.1. Tipo y diseño de investigación

De acuerdo con Haro et al. (2024), el tipo de investigación de investigación aplicada es aquella "investigación dirigida a resolver problemas específicos o aplicar los conocimientos a situaciones prácticas." (p. 4). Según Hernández (2014), una investigación cuantitativa es principalmente la medición, con el objetivo de confirmar y predecir los fenómenos estudiados. Se centra en identificar

patrones y relaciones causales entre los variables, con el fin de formular y validar teorías.

2.2. Periodo y Contexto

El estudio actual se llevó a cabo en 2024, empleando datos del repositorio público de la Universidad de California de Irvine (UCI), que alberga múltiples bases de datos para aprendizaje automático. Los datos utilizados en esta investigación provienen de un conjunto específico del repositorio de aprendizaje automático de la UCI, el cual fue inicialmente recopilado en 2014.

Este estudio se desarrolla en el campo del aprendizaje automático, con un enfoque específico en la evaluación del rendimiento académico de los estudiantes en la asignatura de Matemáticas. Para llevar a cabo este análisis, se emplearon datos del repositorio público de la UCI, ampliamente reconocido en la comunidad académica como una fuente confiable de bases de datos para investigaciones en este campo. Los datos utilizados han sido aplicados en diversas investigaciones previas sobre tareas de clasificación y predicción. En este contexto, se realiza un análisis con el objetivo de predecir el rendimiento académico de los estudiantes utilizando técnicas de aprendizaje supervisado, específicamente mediante algoritmos de redes neuronales, para determinar si es posible anticipar los resultados académicos en función de diversas características.

2.3. Población

De acuerdo con Arias y Miranda (2016), la población de estudio corresponde a un conjunto de casos específicos, delimitados y accesibles, que se utilizarán como base para seleccionar la muestra, la cual debe ajustarse a una serie de criterios previamente establecidos. La población está conformada por 1000 datos provenientes de un repositorio de Repositorio de aprendizaje automático de la UCI, correspondientes a los estudiantes de educación secundaria de un colegio, quienes están siendo evaluados en la asignatura de matemáticas. Para calcular la muestra, se utiliza la siguiente ecuación (1).

$$n = \frac{Nz^2 pq}{e^2(N-1) + pqz^2} \dots (1)$$

$$n = \frac{100(2.58)^2 0.5(1-0.5)}{0.05^2 (1000-1) + 0.5(1-0.5)(2.58)^2}$$

$$n \approx 399.92$$

Donde:

n = Tamaño de muestra

z = Valor correspondiente al nivel de confianza (Para un 99% de confianza, $z = 2.58$)

p = Proporción aproximada del fenómeno en estudio en la población, se asigna una proporción del 0.5 (50%).

q = Proporción de la población de referencia que no presenta el fenómeno en estudio ($1 - p$).

e = Margen de error tolerable en este caso se asigna un margen de 0.05.

El tamaño de la muestra, calculado mediante la fórmula para una población finita, es de aproximadamente 400 registros de estudiantes de secundaria, lo que garantiza una representatividad adecuada para el análisis.

2.4. Selección de muestra

Considerando los postulados de Stratton (2021), se consideró apropiado utilizar el muestreo por conveniencia, una modalidad de muestreo no probabilístico, el cual es uno de las más comunes. En este tipo de muestreo, el investigador anuncia el estudio y los participantes se autoseleccionan si desean participar.

En este estudio, la población estuvo conformada por 400 registros de estudiantes de secundaria. Se consideró pertinente clasificarlos en subgrupos según el nivel educativo de los estudiantes, y dentro de cada grupo se eligieron participantes de manera conveniente, lo que facilitó la selección de estos. Esta estrategia de muestreo permitió asegurar una selección práctica y accesible de los estudiantes, aunque es importante señalar que la muestra no necesariamente es completamente representativa de toda la población estudiantil de secundaria.

De acuerdo con lo anteriormente mencionado, se concluye que la población está conformada por 400 registros, los cuales fueron divididos en subgrupos según el nivel educativo de los estudiantes para facilitar el análisis.

Subgrupo 1 = 80: Primero de secundaria

Subgrupo 2 = 80: Segundo de secundaria

Subgrupo 3 = 80: Tercero de secundaria

Subgrupo 4 = 80: Cuarto de secundaria

Subgrupo 5 = 80: Quinto de secundaria

Se consideró una selección proporcional de los estudiantes, dividiendo la muestra total de 400 registros de manera equitativa entre los cinco subgrupos, asignando 80 datos a cada uno, con el objetivo de asegurar una distribución balanceada para el análisis.

2.5. Selección de factores

Los datos utilizados en este estudio provienen de una fuente secundaria, específicamente del Repositorio de Aprendizaje Automático de la UCI. Este repositorio, administrado por la Facultad de Ciencias de la Computación de la UCI, contiene una base de datos públicos con archivos en formato CSV que recopilan información relevante sobre el rendimiento académico de los estudiantes.

Los datos que se incluyeron como factores relacionados con las calificaciones de los estudiantes son los siguientes:

- Grado
- Género
- Edad
- Dirección
- Tamaño de la familia
- Estado de convivencia
- Educación de la Madre
- Educación del Padre
- Trabajo de la Madre
- Trabajo del Padre
- Razón de elección del colegio
- Tutor
- Tiempo de casa al colegio
- Tiempo de estudio
- Número de repetición del curso
- Apoyo educativo adicional
- Apoyo educativo familiar
- Clases pagadas extra
- Actividades extracurriculares
- Asistencia guardería

- Desea cursar superior
- Internet en casa
- Pareja
- Calidad relación familiar
- Tiempo libre después colegio
- Salir con amigos
- Consumo alcohol en el colegio
- Consumo de alcohol fines de semana
- Estado de salud
- Número de faltas

Esta información permite analizar diversos factores que pueden influir en el rendimiento académico de los estudiantes.

2.6. Recolección de datos

El presente estudio emplea una base de datos proveniente de una fuente secundaria, específicamente del Repositorio de Aprendizaje Automático de la UCI. Administrado por la Facultad de Ciencias de la Computación de la UCI, este repositorio es una de las colecciones más amplias y accesibles de bases de datos, ampliamente utilizado en la investigación y la educación en el ámbito de la ciencia de datos y el aprendizaje automático.

La base de datos empleada en este análisis está disponible en formato CSV, lo que facilita su procesamiento y examen mediante herramientas como Microsoft Excel. Dado que Excel es una de las aplicaciones más populares para el análisis de datos, permite organizar, visualizar y realizar cálculos de manera eficaz, lo cual es crucial para explorar la información relacionada con el rendimiento académico de los estudiantes. La facilidad de uso y la versatilidad de Excel optimizan la interpretación y

el manejo de los datos, permitiendo un análisis exhaustivo y preciso.

2.7. Herramientas para el tratamiento de datos

Tras la recolección de los datos, los métodos de tratamiento se ocupan de procesar, depurar, convertir y analizar la información, preparándola para su utilización en el estudio. En la siguiente tabla se especifican los instrumentos y herramientas utilizadas en los procesos de recolección y tratamiento de los datos, destacando las principales empleadas (ver Tabla 1).

2.8. Selección de Algoritmos predictivos

La selección de los algoritmos predictivos de la presente investigación se ha basado en su amplia adopción y éxito demostrado en la resolución de problemas similares en el campo del aprendizaje automático. En particular, se eligieron tres algoritmos que se destacan tanto por su versatilidad como por su rendimiento en tareas de clasificación: Red Neuronal Densa, Árbol de Decisión y Regresión Logística.

En la investigación de Bagnato (2020), el algoritmo es un conjunto finito de instrucciones organizadas de manera lógica, diseñadas para resolver un problema específico. Consiste en seguir una secuencia de pasos o reglas previamente analizadas y definidas con el objetivo de llegar a una solución. Según Liu (2015), trata de una secuencia de pasos ordenados que ejecutan procesos específicos para responder a una pregunta planteada. En otras palabras, es un conjunto finito y estructurado de instrucciones que facilitan la resolución de problemas o la obtención de una conclusión.

Regresión Logística

De acuerdo con Domínguez (2021), afirman que la regresión logística, una de las técnicas más populares

Tabla 1

Herramientas para el tratamiento de datos

Item	Nombre	Descripción
1	Google Colabority	Es una plataforma en línea gratuita que permite combinar código ejecutable con texto, esta herramienta es compatible con múltiples bibliotecas de Python, lo que facilita el análisis y la visualización de datos de manera eficiente (Esnaola, 2022).
2	Jupyter Notebook	Es una aplicación de código abierto y se utiliza en el análisis de datos, permitiendo depuración, modelización estadística, entrenamiento de modelos de aprendizaje automático y visualización de datos (Cabrera y Díaz, 2018).
3	Python	Python es un lenguaje de programación multiparadigma que destaca por su facilidad de aprendizaje, sintaxis clara y dinámica, sus principales ventajas son su simplicidad, legibilidad y precisión en la sintaxis (Pérez, 2016).
4	NumPy Keras TensorFlow Scikit learn	Es una librería de Python que facilita la creación de matrices, lo que permite un acceso más rápido a los datos (Henriquez et al., 2021).

y versátiles del análisis de datos, es ampliamente utilizada en el aprendizaje supervisado, especialmente cuando la variable respuesta o dependiente es discreta y toma dos valores, como 0 y 1. Esta técnica se ha convertido en un estándar para modelar y analizar variables categóricas binarias.

Árbol de decisión

El árbol de decisión es un algoritmo de aprendizaje supervisado ampliamente utilizado en problemas de clasificación, conocido por su sencillez y facilidad de interpretación. Su estructura permite convertirlo directamente en un conjunto de reglas para realizar predicciones y evaluar la importancia de las variables independientes. Identifica la variable más relevante y su valor para crear grupos homogéneos, eligiendo los puntos de división que optimizan los resultados (Bernuy, 2018).

Redes Neuronales Densa

Conforme nos indica (Bueno, 2019), es un tipo de red artificial en la que cada neurona de una capa está conectada a todas las de la siguiente. Esto permite que cada neurona reciba información de todas las neuronas previas y transmita su salida a las siguientes. Estas redes se emplean en aprendizaje supervisado, destacándose en tareas como clasificación, regresión y reconocimiento de patrones, gracias a su habilidad para captar representaciones complejas de los datos.

2.9. Selección de métricas de validez de evaluación

La validez se evalúa según el nivel de confianza que el modelo tiene al cumplir con su objetivo original. El resultado de esta validez no es simplemente una "demostración de que funciona", sino que se fundamenta en los diferentes resultados obtenidos a partir de una muestra de prueba, que nos proporcionan esa "confianza". Por lo tanto, las pruebas son esenciales, ya que nos permiten identificar tanto los aspectos positivos como negativos del modelo, lo que facilita la realización de las modificaciones necesarias (García y Molina, 2012).

En otra investigación, nos indica que la validación cuantifica la incertidumbre al comparar las predicciones del modelo con datos reales para medir su precisión. Es esencial usar datos de prueba diferentes a los de entrenamiento para evitar resultados sesgados. Aunque los datos pueden contener errores, demostrar pequeñas discrepancias entre

los datos de muestra y las predicciones suele ser suficiente para validar el modelo (Kumar, 2017).

La exactitud

Se refiere a lo cerca que está el resultado de una medición del valor verdadero. En términos estadísticos, la exactitud está relacionada con el sesgo en una estimación. También se conoce como verdadero positivo (Ayala, 2021).

$$\text{Exactitud} = \frac{TP + TN}{TP + TN + FP + FN} \dots (2)$$

La precisión

Se refiere a la dispersión del conjunto de valores obtenidos a partir de mediciones repetidas de una magnitud. Cuanto menor es la dispersión mayor la precisión. Se representa por la proporción entre el número de predicciones correctas tanto positivas como negativas y el total de predicciones (Ayala, 2021).

$$\text{Precisión} = \frac{TP}{TP + FP} \dots (3)$$

Sensibilidad

Esta métrica de evaluación mide la proporción de instancias positivas correctamente identificadas. Es importante cuando se quiere identificar todos los casos positivos, incluso a costa de tener algunos falsos positivos (Ayala, 2021).

$$\text{Sensibilidad} = \frac{TP}{TP + FN} \dots (4)$$

Especificidad

El también conocida como la tasa de verdaderos negativos o TN, según la ecuación (1). Se trata de los casos negativos que el algoritmo ha clasificado correctamente, expresa cuán bien puede el modelo detectar esa clase (Ayala, 2021).

$$\text{Especificidad} = \frac{TN}{TN + FP} \dots (5)$$

F1 score

Esta métrica de evaluación es la media armónica entre la precisión y la sensibilidad. Se utiliza cuando existe un equilibrio de clases o cuando se busca un balance entre la precisión y la capacidad del modelo para identificar los casos positivos (Molina, 2024).

$$FI\ score = \frac{precisión \times sensibilidad}{precisión + sensibilidad} \dots\dots (6)$$

III. RESULTADOS

Se cumplió con el primer objetivo específico del estudio, se seleccionaron 33 factores, los cuales están fundamentados en la teoría de la sección 2.5.

Los resultados obtenidos se sustentan en la teoría de Rivero (2000), que ha evidenciado una relación directa entre el nivel socioeconómico y el rendimiento académico, de manera que los estudiantes de los sectores más desfavorecidos enfrentan mayores dificultades para desarrollar sus habilidades intelectuales, lo que se traduce en un bajo desempeño académico.

Por otro lado, el nivel académico de los padres y el apoyo familiar tienen incidencia directa en el rendimiento académico (Bandura, 1982) (González et al., 2003). De acuerdo con los estudios de Montero et al. (2007), señalan que la problemática tiene su raíz en factores individuales, sociales y culturales de los estudiantes, tales como características sociodemográficas como el género, la situación económica de la familia y el tipo de institución educativa en la que cursaron la educación secundaria.

El autor Latiesa (1992), señala la relevancia de los factores institucionales, como los horarios de clases, el número de estudiantes por grupo, la cantidad de libros en la biblioteca, el entorno institucional en cuanto a infraestructura, y los aspectos vinculados con la carrera que cursan los estudiantes. Tal como indica Cominetti y Ruiz (1997), nos indican que el rendimiento académico no debe solo tomar en cuenta como unidad de medida a través de su evaluación, sino que es fundamental considerar, además del desempeño individual del estudiante, la influencia de factores como el grupo de compañeros, el aula y el contexto educativo. Aunque las calificaciones escolares suelen considerarse un factor cercano al rendimiento académico, es necesario identificar las variables que inciden o explican cómo se distribuyen los aprendizajes entre los estudiantes.

Según la literatura, los resultados muestran una correlación significativa entre los factores identificados y el rendimiento académico de los estudiantes, cumpliendo así con el objetivo de la investigación. A continuación, se presentan los resultados de los factores a considerar en la Tabla 2.

Tabla 2

Factores clave que influyen en el rendimiento académico escolar.

Factores	Descripción	Valores
F1	Grado	Numérico
F2	Género	Binario
F3	Edad	Numérico
F4	Dirección	Numérico
F5	Tamaño de familia	Binario
F6	Convivencia	Binario
F7	Educación madre	Numérico
F8	Educación padre	Numérico
F9	Trabajo madre	Binario
F10	Trabajo padre	Binario
F11	Razón del colegio	Numérico
F12	Tutor	Binario
F13	Trayecto escolar	Numérico
F14	Tiempo de estudio	Numérico
F15	Número de repetición curso	Numérico
F16	Refuerzo educativo	Binario
F17	Refuerzo familiar	Binario
F18	Clases extra	Binario
F19	Actividades extra	Binario
F20	Asistencia infantil	Binario
F21	Desea cursar superior	Binario
F22	Internet en casa	Binario
F23	Pareja	Binario
F24	Calidad familiar	Numérico
F25	Tiempo libre	Numérico
F26	Salir con amigos	Numérico
F27	Alcoholcolegio	Numérico
F28	Alcohol postcolegio	Numérico
F29	Estado de salud	Numérico
F30	Número de faltas	Numérico

Elaboración propia.

Se cumplió con el segundo objetivo específico del estudio, se llevaron a cabo varios pasos destinados a la limpieza, depuración y transformación de los datos (ver Figura 2).

1. Se filtraron los datos para conservar únicamente los registros correspondientes a los estudiantes de secundaria en la asignatura de matemáticas.
2. Se suprimió la columna Horas_sueño_diarias debido a que contenía valores vacíos.
3. Las columnas Tareas_asignadas_semanal y Participación_clase_semanal fueron eliminadas, ya que ambas contenían un único tipo de dato. Todos los valores de esa columna son iguales o muy similares por tanto no aportan variabilidad relevante para el análisis de los datos. En el análisis de datos, la variabilidad es

importante porque permite identificar patrones, tendencias y diferencias significativas. Si todos los valores en una columna son los mismos, no existe ninguna diferencia que el modelo pueda aprender o que contribuya a la predicción, lo que la hace irrelevante para el análisis.

4. Se procedió a completar las celdas que se encontraban vacías.
5. La columna Tipo_matrícula fue eliminada por inconsistencias en sus datos.
6. Se eliminaron los registros duplicados.

Luego de realizar el preprocesamiento de los datos en el cual se hizo una limpieza normalización y depuración de la data, se procede a realizar la estandarización de los datos.

Se utilizó en Python las funciones específicas para estandarizar las variables de pandas y scikit-learn, el cual transforma las variables para que tengan

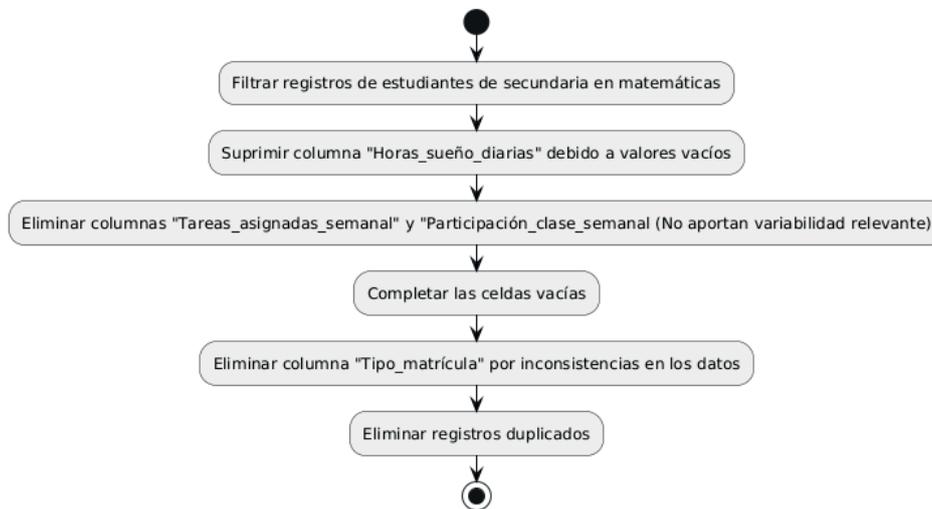
una media de 0 y una desviación estándar de 1, lo que las hace comparables entre sí y mejora la efectividad de muchos algoritmos de aprendizaje automático. En la tabla 3 se presentan cinco factores a los que se les aplicó la transformación de datos categóricos a numéricos.

Los modelos de aprendizaje automático requieren exclusivamente datos numéricos, por lo que resulta adecuado convertir todas las variables cualitativas o no numéricas. Para ello, es necesario aplicar enfoques específicos o emplear bibliotecas de Python que permitan realizar estas conversiones.

Luego se procede a evaluar la predicción en los algoritmos de aprendizaje automático y determinar cuál de ellos es el que tiene mayor porcentaje de precisión en la predicción de rendimiento académico en matemáticas de estudiantes de secundaria.

Se aplicaron y desarrollaron los algoritmos seleccionados en la sección 2.8 utilizando el lenguaje de programación Python. Se emplearon los datos

Figura 2
Pasos del preprocesamiento de datos.



Elaboración propia.

Tabla 3
Una representación parcial de la transformación de datos de los factores

Factor	Indicador	Item
Grado	Número entero entre 1 y 5	1
Género	Número binario femenino = 1 y masculino = 0	2
Edad	Número numérico entre 12 y 16	3
Dirección	Número binario 1: urbano y 0: rural	4
Tamaño de familia	Número binario 1 <= 3 y 0 > 3	5

Elaboración propia.

obtenidos del repositorio de aprendizaje automático de la UCI, los cuales fueron recopilados para conformar una base de datos pública en formato CSV que contiene información relevante sobre el rendimiento académico de los estudiantes.

Los datos corresponden a los registros académicos de estudiantes de educación secundaria en la asignatura de matemáticas, obtenidos de una muestra de 400 estudiantes seleccionados mediante muestreo por conveniencia, con el fin de asegurar la representatividad de diversos niveles académicos y contextos socioeconómicos.

El 80% de la muestra, equivalente a 320 estudiantes, se destinó al entrenamiento del modelo, mientras que el 20% restante, correspondiente a 80 estudiantes, se utilizó para la validación. Las variables consideradas para el estudio se detallan en la Tabla 4.

Tabla 4
Porcentaje de prueba y entrenamiento de la muestra.

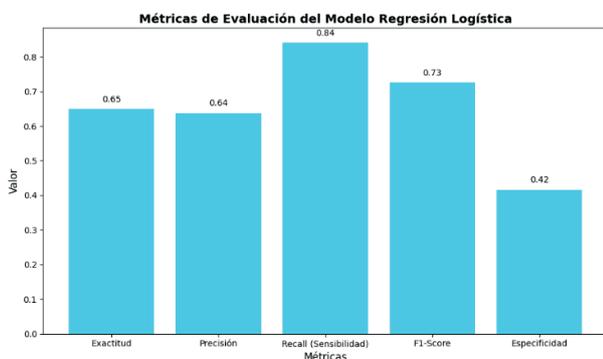
	Selección de porcentajes	Muestra total = 400
Entrenamiento	80%	320%
Prueba	20%	80%

Elaboración propia.

Se cumplió con el tercer objetivo 3, se emplearon métricas de validación para evaluar la, exactitud, precisión, sensibilidad F1 score y especificidad para determinar la efectividad de la predicción de los algoritmos de aprendizaje automático. A continuación, se presentan los resultados de los algoritmos de aprendizaje automático en relación con estas métricas de validación.

Regresión logística

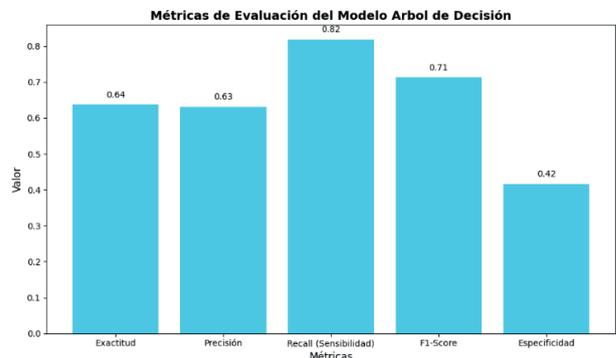
Figura 3
Tabla de evaluación de métricas del algoritmo de Regresión logística.



Elaboración propia.

Árbol de decisión

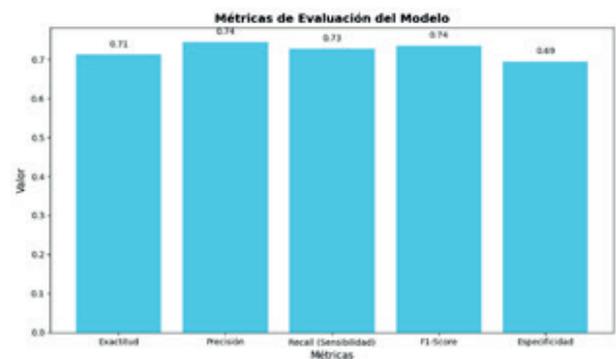
Figura 4
Tabla de evaluación de métricas del algoritmo de Árbol de decisión.



Elaboración propia.

Red neuronal densa

Figura 5
Tabla de evaluación de métricas del algoritmo de Red neuronal densa.



Elaboración propia.

En esta sección se presentan los resultados obtenidos de las métricas de evaluación al evaluar diversos modelos de aprendizaje automático para predecir el rendimiento académico de estudiantes de secundaria.

Se elaboró un cuadro comparativo Tabla 5, para mostrar los distintos factores tomados en cuenta en la evaluación de la efectividad del modelo predictivo de aprendizaje automático.

En base a las métricas evaluadas, la red neuronal densa mostró el mejor desempeño en términos de precisión, exactitud y especificidad, lo que indica que fue el modelo más confiable en general para la predicción del rendimiento académico. Sin embargo, el Árbol de Decisión y la Regresión Logística fueron más efectivos en términos de especificidad y

F1 score, siendo capaces de identificar mejor a los estudiantes con alto rendimiento.

Tabla 5

Resultados de las métricas de evaluación aplicados a los algoritmos de aprendizaje automático.

Métricas de Evaluación	Árbol de Decisión	Regresión Logística	Red Neuronal Densa
Exactitud	64%	65%	71%
Precisión	63%	64%	74%
Sensibilidad	82%	84%	73%
F1-Score	71%	73%	74%
Especificidad	42%	42%	69%

Elaboración propia.

En términos de importancia de las variables, la sensibilidad fue la variable con mayor peso en los modelos, particularmente en los árboles de decisión y la regresión logística, los cuales mostraron una relación positiva significativa con el rendimiento académico de los estudiantes. Esto sugiere que un aumento en las variables relevantes puede generar un incremento en el rendimiento de los estudiantes, especialmente en los modelos de Árbol de Decisión y Regresión Logística.

En resumen, cada modelo presentó fortalezas y debilidades según la métrica utilizada, lo que destaca la importancia de elegir el modelo adecuado en función de los objetivos específicos del análisis.

DISCUSIÓN

Los resultados obtenidos son consistentes con estudios previos que han utilizado aprendizaje automático para predecir el rendimiento académico de estudiantes de secundaria realizadas en otros países como Colombia y México. Según la investigación de Gil y Quintero (2021), el estudio se basó en datos de 395 estudiantes colombianos de educación media vocacional de la "Institución Educativa Villa del Socorro" en Medellín. En su investigación, la red neuronal artificial (RNA) logró clasificar correctamente el 73% de las muestras, superando a otras técnicas en métricas como exactitud, recuperación, precisión y puntuación F1. De manera similar, el modelo de Red neuronal de la presente investigación utilizado en el presente estudio alcanzó una precisión superior, siendo la red neuronal densa la que obtuvo el mejor rendimiento, con un 74% de precisión. Este resultado es comparable al de Gil y Quintero, quienes lograron una predicción precisa con un 73%, destacándose en las mismas métricas.

Al igual que en la investigación de Castrillón et al. (2020), los resultados del modelo de aprendizaje automático de árbol de decisión mostraron una tendencia congruente en donde el valor más alto en cuanto a la métrica de exactitud fue de 91%, emplearon 460 registros de estudiantes y el valor de la exactitud de la presente investigación se obtuvo un 64%, siendo inferior el porcentaje pero se observa que la muestra de datos utilizado por Castrillón abarca una mayor cantidad de registro de datos de alumnos.

De manera análoga a la investigación de Juárez et al. (2014), los resultados fueron igualmente consistentes, en donde participaron 104 estudiantes en su investigación y el valor obtenido en la métrica de exactitud fue de 80%, en comparación con nuestra investigación donde se obtuvo un 71% utilizando el modelo la Red neuronal.

En cuanto a las métricas de sensibilidad y especificidad, se observa que a medida que aumenta la probabilidad de aprobación de un estudiante, la sensibilidad tiende a disminuir mientras que la especificidad aumenta. Esta tendencia también fue reportada de manera similar en la investigación de Reyes et al. (2020), quienes concluyeron de manera similar respecto al comportamiento de las métricas de sensibilidad y especificidad.

CONCLUSIÓN

El rendimiento académico en matemáticas de los estudiantes de secundaria depende de una combinación de factores individuales, familiares, sociales e institucionales. Entre los más influyentes se encuentran el nivel socioeconómico, el apoyo familiar, las condiciones escolares y el contexto educativo. Los resultados destacan la importancia de abordar el rendimiento desde un enfoque integral que considere tanto las características del estudiante como su entorno.

El preprocesamiento de datos es una etapa esencial en cualquier análisis, y los seis pasos implementados han sido fundamentales para garantizar la calidad y efectividad del proceso. Cada uno de estos pasos jugó un papel crucial en la preparación de los datos, asegurando que estuvieran listos para un análisis adecuado. Su correcta ejecución permitió obtener resultados más precisos y confiables, lo que incrementó la validez y el valor del análisis realizado.

La Red Neuronal Densa mostró el mejor desempeño general, pero la Regresión Logística y el Árbol de Decisión tienen ventajas específicas en función de

las métricas como sensibilidad y exactitud, respectivamente. La elección del algoritmo más adecuado dependerá de los objetivos específicos del análisis, ya sea priorizar la identificación de estudiantes con buen rendimiento o aquellos con bajo rendimiento.

RECOMENDACIONES

Dado que la Red Neuronal Densa mostró el mejor desempeño en precisión y F1-Score, se recomienda priorizar su uso en escenarios donde se busque un balance entre la identificación de estudiantes con alto y bajo rendimiento académico. Sin embargo, es importante evaluar también la sensibilidad para asegurar que se detecten adecuadamente a los estudiantes con bajo rendimiento.

Para obtener resultados más robustos y generalizables, se recomienda realizar investigaciones futuras con muestras más grandes y diversas. Ampliar la diversidad de los estudiantes, en términos de características demográficas y de rendimiento académico, permitirá que los modelos de predicción sean más representativos y aplicables a diferentes contextos educativos.

Sería beneficioso evaluar la efectividad de los algoritmos no solo en el ámbito de matemáticas, sino también en otras asignaturas o áreas del conocimiento. Esto ayudaría a entender cómo se comportan los modelos en diferentes dominios y si pueden adaptarse a otros escenarios educativos.

En investigaciones futuras podrían centrarse en explorar y aplicar nuevas técnicas de preprocesamiento de datos para manejar datos faltantes, o la implementación de técnicas más sofisticadas para la normalización de datos. Esto podría mejorar la calidad de los datos y la precisión de los modelos.

REFERENCIAS

- [1] Arán, P. (2022). Métodos de aprendizaje automático y aplicaciones. Universidad Zaragoza, 55. Obtenido de <https://zaguan.unizar.es/record/124527/files/TAZ-TFM-2022-1444.pdf>
- [2] Arias, J., Villasís, M., y Miranda, M. (2016). El protocolo de investigación III: La población de estudio, 201-206.
- [3] Ayala, E. (2021). Predicción del rendimiento académico basado en Machine Learning, Escuela Profesional de Ingeniería de Sistemas, Ayacucho 2021. Universidad Nacional de San Cristóbal de Huamanga, 31. <http://repositorio.unsch.edu.pe/handle/UNSCH/4694>
- [4] Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2), 122-147.
- [5] Bagnato, J. (2020). Aprende Machine Learning en español. Leanpub.
- [6] Bernuy, A. (2018). Predicción del rendimiento académico mediante minería de datos en estudiantes del primer ciclo de la escuela profesional de Ingeniería de computación y sistemas, Universidad de San Martín de Porres, Lima-Perú. Universidad de San Martín de Porres.
- [7] Bravo, L., Fuentes, H., y Rivas, E. (2021). Análisis del rendimiento académico mediante técnicas de aprendizaje automático con métodos de ensamble. *Revista Boletín Redipe*, 10(13), 171-190. <https://doi.org/10.36260/rbr.v10i13.1737>
- [8] Bricklin, B. (2007). Factores que intervienen en el rendimiento académico. Psicología de bajo rendimiento. México: Pax.
- [9] Bueno, F. (2019). Redes Neuronales: Entrenamiento y Comportamiento. Universidad Complutense Madrid.
- [10] Cabrera, E., y Díaz, E. (2018). Manual de uso de Jupyter Notebook para aplicaciones docentes. Universidad Complutense de Madrid.
- [11] Calva, K., Flores, M., Porras, H., y Cabezas, A. (2021). Modelo de predicción del rendimiento académico para el curso de nivelación de la Escuela Politécnica Nacional a partir de un modelo de aprendizaje supervisado. *Latin American Journal of Computing*, 3(2). Obtenido de <https://lajc.epn.edu.ec/index.php/LAJC/article/download/264/159/>
- [12] Cominetti, R y Ruiz, G. (1997). Algunos factores del rendimiento: las expectativas y el género. Human Development Department. LCSHD Paper series, 20, The World Bank, Latin America and Caribbean Regional Office
- [13] Domínguez, M. (2021). Trabajo de fin de grado en Matemáticas. Universidad Zaragoza.
- [14] Esnaola, L. (2022). Experiencia de uso de Google Colaboratory en Sistemas Inteligentes. Recuperado de <http://repositorio.unnoba.edu.ar/xmlui/handle/23601/652>
- [15] Estrada, R. I., Zamarripa, R. A., Zúñiga, P. G., y Martínez, I. (2016). Aportaciones desde la minería de datos al proceso de captación de matrícula en instituciones de educación

- superior particulares. *Revista Electrónica Educare*, 20(3), 1-21. <http://www.redalyc.org/articulo.oa?id=194146862011>
- [16] Forero, W., y Negre, F. (2024). Técnicas y aplicaciones del machine learning e inteligencia artificial en educación: Una revisión sistemática. *Revista Iberoamericana de Educación a Distancia*, 45(27). <https://doi.org/10.5944/ried.27.1.37491>
- [17] Garbanzo, G. (2013). Factores asociados al rendimiento académico en estudiantes universitarios desde el nivel socioeconómico: Un estudio en la Universidad de Costa Rica. *Revista Electrónica Educare*, 17. Obtenido de https://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S1409-42582013000300004
- [18] García, J. y Molina, J. (2012). Técnicas de análisis de datos, 266. Recuperado de http://matema.ujaen.es/jnavas/web_recursos/archivos/weka%20master%20recursos%20naturales/apuntesAD.pdf
- [19] Gil, D., y Quintero, C. (2021). Development of an artificial neural network for predicting student academic success. *Información tecnológica*, 221-228. Obtenido de <https://dx.doi.org/10.4067/S0718-07642021000600221>
- [20] González, J., Núñez J., Álvarez L., Rocés C., González, S., González P. y Bernardo A. (2003). Adaptabilidad y cohesión familiar, implicación parental en conductas autorregulatorias, autoconcepto del estudiantes y rendimiento académico. *Psicothema*, 15 (3), 471-477. Recuperado de <http://www.psicothema.com/pdf/1090.pdf>.
- [21] Ghorbani y Ghousi, (2020). Comparación de diferentes métodos de remuestreo para predecir el rendimiento de los estudiantes mediante técnicas de aprendizaje automático, 67899-67911. *Institute of Electrical and Electronics Engineers*. DOI: 10.1109/ACCESS.2020.2986809
- [22] Haro, A. F., Chisag Pallmay, E. R., Ruiz Sarzosa, J. P., y Caicedo Pozo, E. J. (2024). Tipos y clasificación de las investigaciones. *Revista Latinoamericana de Ciencias Sociales y humanidades*, 4-11.
- [23] Hernández, R. (2014). *Metodología de la Investigación*. México: McGraw Hill Interamericana Editores S.A de C.V.
- [24] Henríquez, C., Salcedo, D., y Sánchez, G. (2021). El aprendizaje automático en entornos educativos universitarios: Caso deserción académica. *Revista Panorama*, 20(1). <https://doi.org/10.15665/rp.v20i1.2736>
- [25] Latiesa, M. (1992). La deserción universitaria, desarrollo de la escolaridad en la enseñanza superior. Éxitos y fracasos. España: Siglo XXI de España Editores S.A.
- [26] Lizares, M. (2017). Comparación de modelos de clasificación: regresión logística y árboles de clasificación para evaluar el rendimiento académico. *Universidad Mayor de San Marcos*
- [27] Liu, A. (2015). *Data Science and Data Scientist*. IBM.
- [28] Longoni, M. G., Porcel, E. A., López, M. V., y Dapozo, G. N. (2010). Red de Universidades con Carreras en Informática (RedUNCI). *Revista*, 692-701.
- [29] MINEDU. (2024). Resultados PISA 2022. Ministerio de Educación, 243. Obtenido de http://umc.minedu.gob.pe/wp-content/uploads/2024/04/Reporte_de_resultados_PISA_2022_Per%C3%BA.pdf
- [30] Molina, M. (2024). Un intruso de otro mundo: F1-score. *Revista Electrónica de AnestesiAR*.
- [31] Montero, E., Villalobos J., y Valverde B. (2007). Factores institucionales, pedagógicos, psicosociales y sociodemográficos asociados al rendimiento académico en la Universidad de Costa Rica: Un análisis multinivel. *Relieve*, 13 (2), 215-234. Recuperado de http://www.uv.es/RELIEVE /v13n2/ RELEVEv13n2_5.htm.
- [32] Páez, A., y Gaytán, N. (2022). Modelos predictivos del rendimiento académico a partir de características de estudiantes de ingeniería. *IE Revista de Investigación Educativa de la REDIECH*, 13, 1-18. <https://www.redalyc.org/journal/5216/521670731008/521670731008.pdf>
- [33] Pérez, A. (2016). *Python fácil*. Universidad Técnica del Norte.
- [34] Rivero, J. (2000). Reforma y desigualdad educativa en América Latina. *Revista Iberoamericana de educación*, 23. Recuperado de <http://rieoei.org/rie23 a03.htm>.
- [35] Romero, C., y Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (applications and reviews)*, 40(6), 601-618. <https://doi.org/10.1109/TSMCC.2010.2053532>

- [36] Salal, Y. K., Abdullaev, S. M., y Kumar, M. (2019). Educational data mining: Student performance prediction in academic. *International Journal of Engineering and Advanced Technology*, 8(4C), 54-59. Obtenido de <https://www.semanticscholar.org/paper/Educational-Data-Mining-%3A-Student-Performance-in-Salal-Abdullaev/b21fa7245581c3baad2d468cb9d706940de7e010>
- [37] Salgado, N., Beltrán, J., Escobar, C., Nicolalde, D., y Chafra, G. (2019). Modelo para predecir el rendimiento académico basado en redes neuronales y analítica de aprendizaje. Pontificia Universidad Católica del Ecuador, 258-266.
- [38] Stratton, J. (2021). Investigación de población: Estrategias de muestreo por conveniencia. Cambridge University Press, 373-374. <https://doi.org/10.1017/S1049023X21000649>
- [39] Siqueira, C., y Gurge, J. (2011). Bajo rendimiento escolar: una revisión actualizada. *Revista da Associação Médica Brasileira*, 57(1), 78-86. Obtenido de [https://doi.org/10.1016/S2255-4823\(11\)70021-2](https://doi.org/10.1016/S2255-4823(11)70021-2)
- [40] Kumar, R. (2017). *Machine Learning and Cognition in Enterprises*. India: TODD GREEN.
- [41] Vega, J. F. (2019). Modelo de pronóstico de rendimiento académico de alumnos en los cursos del programa de estudios básicos de la Universidad Ricardo Palma usando algoritmos de Machine Learning. Universidad Ricardo Palma, 358. Obtenido de <https://repositorio.urp.edu.pe/handle/20.500.14138/2914>

Financiamiento:

Propia.

Conflictos de interés:

El autor declara no tener conflictos de interés.