



*Algoritmos de Aprendizaje Automático en la Predicción del Rendimiento
Académico en la Educación Superior*

*Machine Learning Algorithms in Predicting Academic Performance in Higher
Education*

*Algoritmos de aprendizagem automática na previsão do desempenho acadêmico
no ensino superior*

María Gabriela López-Barriga ^I
gaby07lopez12@gmail.com
<https://orcid.org/0000-0003-3728-3078>

Alex Eduardo Pozo-Valdiviezo ^{II}
eduardo.pozo@esepoch.edu.ec
<https://orcid.org/0000-0003-0480-5669>

Natalia Alexandra Pérez-Londo ^{III}
nperez@esepoch.edu.ec
<https://orcid.org/0000-0001-9068-8790>

Cristina Estefanía Ramos-Araujo ^{IV}
cristina.ramos@esepoch.edu.ec
<https://orcid.org/0000-0002-8644-5814>

Correspondencia: gaby07lopez12@gmail.com

Ciencias de la Educación
Artículo de Investigación

* **Recibido:** 01 de mayo de 2024 * **Aceptado:** 07 de junio de 2024 * **Publicado:** 25 de julio de 2024

- I. Ingeniera Estadística, Investigador Externo en la Escuela Superior Politécnica de Chimborazo, Riobamba, Ecuador.
- II. Máster en Ciencias, Tecnologías, Salud Mención Matemáticas y Aplicaciones-Matemáticas para las Ciencias de la Vida, Profesor Ocasional Tiempo Completo en la Escuela Superior Politécnica de Chimborazo, Riobamba, Ecuador.
- III. Máster Universitario en Estadística Aplicada, Profesor Ocasional en la Escuela Superior Politécnica de Chimborazo, Riobamba, Ecuador.
- IV. Máster of Science in Applied Mathematics, Profesor Ocasional Tiempo Completo en la Escuela Superior Politécnica de Chimborazo, Riobamba, Ecuador.

Resumen

En la educación un problema controversial es el desempeño académico de los estudiantes lo cual está asociado con múltiples factores internos y externos que afectan a los universitarios los mismos que conllevan al éxito o fracaso del estudiante, por lo tanto, el objetivo de la presente investigación fue elaborar una revisión sistemática de los algoritmos de Aprendizaje Automático para predecir el rendimiento académico de los estudiantes universitarios. La metodología implementada tuvo un enfoque cualitativo y cuantitativo, se utilizó un diseño no experimental y la población de estudio corresponde a los artículos relacionados al rendimiento académico encontrados en las bases de datos como: Scopus, Dialnet, SciELO y ERIC. Se utilizó el diagrama de flujo PRISMA en donde se encontraron un total de 6437 publicaciones correspondiente a los últimos 10 años, posteriormente a partir de criterios de inclusión y exclusión se redujeron a 52 artículos para ser analizados. Mediante esta metodología se logró determinar que durante el 2022 hubo mayor parte de publicaciones sobre el tema las mismas que fueron desarrolladas en el continente americano. También, se comparó las métricas de evaluación como la precisión, exactitud, sensibilidad, especificidad, el puntaje F1 y la curva ROC de los algoritmos de Aprendizaje Automático obteniendo que el mejor algoritmo supervisado para predecir el rendimiento académico fue Árbol de Decisión con un valor superior al 90% en casi todos los indicadores. En este contexto se concluyó que las variables que más influyen en el rendimiento académico fueron las variables relacionadas al factor socioeconómico, familiar, demográfico, personal, institucional, académico pre-universidad y académico universitario los mismos que podrían ser recolectados para realizar un análisis del rendimiento académico en la ESPOCH.

Palabras clave: Revisión sistemática; Árbol de decisión; Aprendizaje automático; Métricas de evaluación; Rendimiento académico; Educación universitaria; Directrices prisma.

Abstract

In education, a controversial problem is the academic performance of students, which is associated with multiple internal and external factors that affect university students, which lead to the success or failure of the student, therefore, the objective of the present research was to prepare a systematic review of Machine Learning algorithms to predict the academic performance of university students. The methodology implemented had a qualitative and quantitative approach, a non-experimental

design was used and the study population corresponds to the articles related to academic performance found in databases such as: Scopus, Dialnet, SciELO and ERIC. The PRISMA flowchart was used where a total of 6,437 publications corresponding to the last 10 years were found, later based on inclusion and exclusion criteria they were reduced to 52 articles to be analyzed. Through this methodology, it was determined that during 2022 there were most publications on the subject, which were developed in the American continent. Also, the evaluation metrics such as precision, accuracy, sensitivity, specificity, F1 score and ROC curve of the Machine Learning algorithms were compared, obtaining that the best supervised algorithm to predict academic performance was Decision Tree with a value higher than 90% in almost all indicators. In this context, it was concluded that the variables that most influence academic performance were the variables related to the socioeconomic, family, demographic, personal, institutional, pre-university academic and university academic factors, which could be collected to perform an analysis of academic performance at ESPOCH.

Keywords: Systematic review; Decision tree; Machine learning; Evaluation metrics; Academic performance; University education; Prisma guidelines.

Resumo

Na educação, um problema controverso é o desempenho acadêmico dos estudantes, que está associado a múltiplos fatores internos e externos que afetam os estudantes universitários, que levam ao sucesso ou insucesso do aluno, pelo que, o objetivo desta investigação foi preparar uma revisão sistemática de algoritmos de aprendizagem automática para prever o desempenho acadêmico de estudantes universitários. A metodologia implementada teve uma abordagem qualitativa e quantitativa, foi utilizado um desenho não experimental e a população do estudo corresponde aos artigos relacionados com o desempenho acadêmico encontrados em bases de dados como: Scopus, Dialnet, SciELO e ERIC. Foi utilizado o fluxograma PRISMA onde foram encontradas um total de 6.437 publicações correspondentes aos últimos 10 anos, posteriormente, com base em critérios de inclusão e exclusão, foram reduzidas a 52 artigos a analisar; Recorrendo a esta metodologia, foi possível constatar que durante o ano de 2022 houve um maior número de publicações sobre o tema, as mesmas que foram desenvolvidas no continente americano. Além disso, foram comparadas as métricas de avaliação como a precisão, exatidão, sensibilidade, especificidade, o score F1 e a curva ROC dos algoritmos de Machine Learning, obtendo-se que o melhor algoritmo supervisionado para

prever o desempenho académico foi a Árvore de Decisão com valor superior 90% em quase todos os indicadores. Neste contexto, concluiu-se que as variáveis que mais influenciaram o desempenho académico foram as variáveis relacionadas com os fatores socioeconómicos, familiares, demográficos, pessoais, institucionais, académicos pré-universitários e académicos universitários, os mesmos que poderiam ser recolhidos para a realização de um estudo.

Palavras-chave: Revisão sistemática; Árvore de decisão; aprendizagem de máquina; Métricas de avaliação; Rendimento académico; Formação universitária; Orientações do Prisma.

Introducción

La educación universitaria ofrece la oportunidad de mejorar la calidad de vida, capacitando a los estudiantes para el campo laboral. Los docentes, capacitados en estrategias y metodologías de aprendizaje, son fundamentales para guiar a los estudiantes y extraer su potencial (Arras Vota et al., 2016; Acuña Benites et al., 2021; Chuyma Huilca et al., 2021; Moreno Durán et al., 2016; citados en Mireles y García, 2022)

El rendimiento de los estudiantes se ve afectado por factores personales, socioculturales, económicos y académicos, siendo los personales y sociales los más comunes (Torres y Rodríguez, 2006; citados en González y Guadalupe, 2017). El desempeño escolar refleja conocimientos y rasgos psicológicos adquiridos durante el proceso educativo, y se evalúa objetivamente al final del periodo académico (Bravo et al., 2021).

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 (Gutiérrez et al., 2021; citados en Páez & Gaytán, 2022). La predicción del rendimiento académico es un campo en crecimiento que emplea enfoques informáticos como análisis de aprendizaje, aprendizaje automático, minería y exploración de datos educativos (Rastrollo et al., 2020; citados en Estrada & Fuentes, 2021). Estos avances facilitan la toma de decisiones en el ámbito educativo (Murnion y Helfert, 2013; citados en Bravo et al., 2021).

Revisar la literatura es esencial para contextualizar y fundamentar la investigación, definir conceptos y teorías, desarrollar hipótesis y mejorar la metodología (Arnau & Sala, 2020). Dada la influencia de factores internos y externos en el rendimiento académico, es necesario revisar la literatura científica para identificarlos.

Aprendizaje Automático

El Aprendizaje Automático (AA) es un área de la inteligencia artificial centrada en desarrollar algoritmos que pueden aprender y mejorar a partir de los datos sin programación explícita, lo cual ahorra tiempo y esfuerzo (Sandoval, 2018). Permite a los ordenadores aprender de los datos disponibles, extrayendo patrones y generando algoritmos para categorización, predicción y explicación, en lugar de seguir instrucciones paso a paso (Bobadilla, 2020). Aunque es una técnica dentro de la inteligencia artificial, esta última incluye una gama más amplia de métodos para analizar datos, como el procesamiento del lenguaje natural. El AA se introdujo en la década de 1950 para simular aspectos de la cognición humana (Pedrero et al., 2021).

Algoritmos de Aprendizaje Automático

Los algoritmos de aprendizaje automático se dividen en tres categorías: supervisado, no supervisado y por refuerzo. El aprendizaje supervisado se denomina así porque se entrena un algoritmo con preguntas (características) y respuestas (etiquetas), permitiendo al modelo hacer predicciones futuras basadas en ese conocimiento (Sandoval, 2018). Su objetivo es relacionar variables con un resultado deseado y hacer predicciones con nuevos datos una vez terminado el entrenamiento (Pedrero et al., 2021). Los algoritmos comunes de aprendizaje supervisado son: Regresión Lineal y algoritmo de Clasificación. Dentro de los algoritmos de clasificación constan redes neuronales, máquina de Vectores de Soporte (SVM), Naïve Bayes, árbol de Decisión, K-Nearest Neighbor (KNN) y bosque Aleatorio (Random Forest) (MENASALVAS et al., 2023).

El aprendizaje no supervisado se refiere a técnicas donde se analizan datos no etiquetados para descubrir patrones ocultos, en contraste con el aprendizaje supervisado, que intenta predecir valores basados en datos etiquetados. Los dos enfoques principales en aprendizaje no supervisado son: Clustering (Agrupación y Reducción de Dimensionalidad). Ambos métodos ayudan a analizar datos complejos y a extraer información valiosa sin necesidad de etiquetas predefinidas (Romero, 2015).

El Aprendizaje por Refuerzo es una técnica de Aprendizaje Automático que permite a los agentes aprender a través de interacciones de prueba y error en un entorno dinámico. Se enfoca en optimizar el comportamiento del agente para maximizar una recompensa acumulada, en lugar de usar un conjunto de datos de entrenamiento fijo. Es especialmente útil en problemas donde crear un algoritmo de antemano es muy costoso, como en la gestión de recursos, planificación, control del tráfico, robótica, detección de intrusos, control de sistemas energéticos, y detección de fraudes.

Actualmente, se requiere un gran poder de procesamiento de datos para abordar estos problemas (Cervantes et al., 2019).

Métricas de evaluación de los algoritmos del Aprendizaje Automático

Para evaluar la calidad de un algoritmo de Aprendizaje Automático, es crucial considerar la calidad de la predicción y el grado de generalización. Estas evaluaciones se basan en métricas derivadas de la matriz de confusión. Esta matriz, utilizada en pruebas binarias, muestra los resultados de las predicciones comparadas con los valores verdaderos. Los cuatro resultados principales en la matriz de confusión son:

Falsos positivos (FP): Casos etiquetados como positivos por el algoritmo, pero en realidad son negativos.

Falsos negativos (FN): Casos etiquetados como negativos, pero en realidad son positivos.

Verdaderos negativos (TN): Casos correctamente predichos como negativos.

Verdaderos positivos (TP): Casos correctamente predichos como positivos (Pedrero et al., 2021; Arias, 2013).

Tabla 1: Matriz de confusión

		PREDICHOS	
		Positivos	Negativos
REALES	Positivos	Verdaderos Positivos (TP)	Falsos Negativos (FN) Verdaderos
	Negativos	Falsos Positivos (FP)	Negativos (TN)

Realizado por: López, María y Gualpa, Henry, 2023

Precisión: La proporción de casos positivos predichos con precisión respecto al número total de casos positivos esperados es una métrica de evaluación utilizada con frecuencia para tareas de clasificación (Contreras et al., 2020, p.239).

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

Exactitud: El número de instancias predichas correctamente dividido por el número total de registros es la forma más directa de evaluar la precisión de un clasificador (Contreras et al., 2020, p.239).

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

Sensibilidad: Esta métrica ilustra lo bien que el algoritmo predice situaciones positivas; indica la proporción de predicciones positivas que son realmente positivas en relación con el número total de predicciones reales clasificadas como positivas (Martínez, 2022, p.33).

$$\text{Sensibilidad} = \frac{TP}{TP+FN} \quad (3)$$

Puntaje F1: Es la media de la precisión y la sensibilidad, ponderadas conjuntamente. La asociación entre esta puntuación y la calidad del algoritmo se basa en la idea de que mayores porcentajes de F1 indican algoritmos más potentes. Esta puntuación tiene en cuenta tanto los falsos positivos como los falsos negativos (Silva, 2020).

$$F1 = \frac{2*Precisión*S+Sensibilidad}{Precisión+Sensibilidad} \quad (4)$$

Especificidad: Una medida de la precisión del algoritmo en la clasificación de situaciones negativas es el número de verdaderos negativos dividido por el número total de verdaderos negativos, que es lo mismo que el número de casos negativos (Camps, 2023, p.63).

$$\text{Especificidad} = \frac{TN}{TN+FN} \quad (5)$$

Curva ROC: El término hace referencia a la Característica Operativa del Receptor, que es una representación gráfica de la especificidad (eje X) frente a la sensibilidad (eje Y) de un clasificador binario. Basándose en un umbral predeterminado, se mide por el área bajo la curva (AUC), que indica lo bien que el algoritmo distingue y predice entre las dos clases. Los valores en ambos ejes van de 0 a 1, o lo que es lo mismo, de 0% a 100% en términos porcentuales (Martínez, 2022).

Metodología

Por el método de investigación el estudio fue mixto, teórico y descriptivo, debido a la combinación de resultados cuantitativos y cualitativos. Se ha realizado de manera no experimental e inductivo-deductiva, analizando el rendimiento académico en la educación superior en función de factores geográficos, económicos y sociales durante el periodo 2013-2022. La población de estudio corresponde a las publicaciones relacionados al rendimiento académico con algoritmos de Aprendizaje Automático encontrados en las diferentes bases de datos científicas durante el período 2013 – 2022.

Para el desarrollo del presente estudio se utilizó como estrategia metodológica la revisión sistemática de artículos científicos, la cual adopta técnicas claras y sistemáticas para la

identificación, cribado e inclusión de la información sobre el tema de interés garantizando que la investigación sea fiable y rigurosa. La técnica PRISMA, acrónimo de Preferred Reporting Items for Systematic Reviews and Meta-Analyses, creó una normativa internacional, se ha considerado durante la selección de los trabajos y el proceso de revisión sistemática. Los estudios fueron evaluados críticamente para comparar los algoritmos de aprendizaje automático en términos de precisión, sensibilidad y especificidad.

Resultados

Se realizó una búsqueda en diversas bases de datos usando descriptores y operadores booleanos, centrada en “predicción del rendimiento académico con algoritmos de Aprendizaje Automático”. Los resultados obtenidos se muestran en las tablas siguientes. Se aplicaron criterios adicionales de inclusión y exclusión para seleccionar los artículos finales para la investigación. Se consideraron un total de 6,437 artículos científicos sobre el rendimiento académico utilizando algoritmos de Aprendizaje Automático, publicados en los últimos 10 años. Posteriormente, se eliminaron los documentos duplicados mediante la herramienta de automatización Zotero, reduciendo la población a 6,019 artículos. A continuación, se realizó una revisión preliminar de títulos y resúmenes, eliminando 5,788 publicaciones que no estaban relacionadas con el área de educación. Asimismo, se excluyeron 88 artículos que no ofrecían acceso libre. Tras una lectura completa y de acuerdo con los criterios de exclusión establecidos, se descartaron 91 estudios adicionales. Finalmente, se incluyeron 52 artículos en la revisión sistemática.

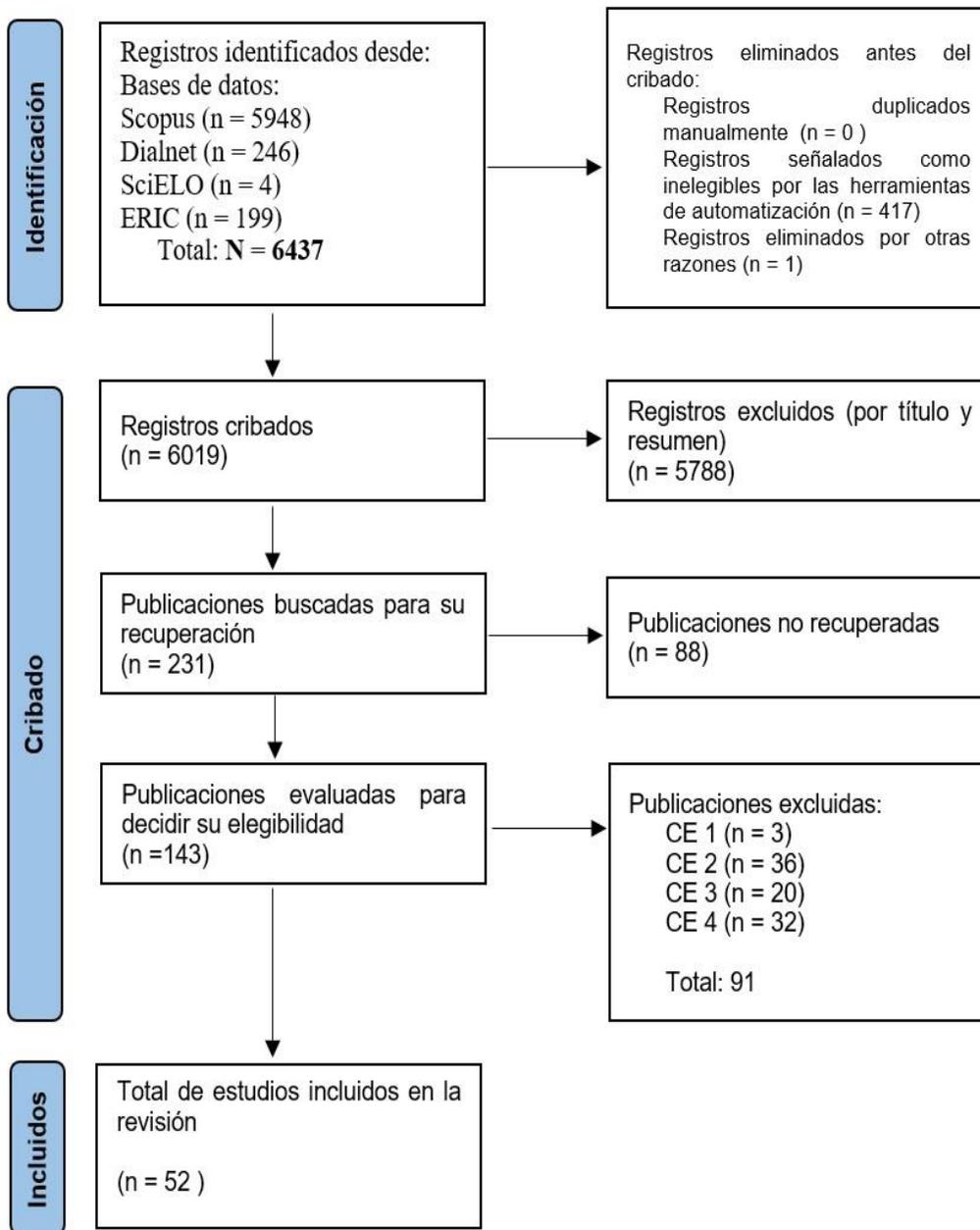
Tabla 2: Resultados de búsqueda

Bases de datos	N° de artículos
Scopus	5948
Dialnet	246
SciELO	44
ERIC	199
Total	6437

Realizado por: López, María y Gualpa, Henry, 2023

La Figura 1 muestra el diagrama de flujo que ejemplifica cómo se han encontrado y seleccionado los artículos científicos utilizando las directrices PRISMA 2020.

Figura 1: Diagrama de flujo mediante las directrices internacionales PRISMA



Realizado por: López, María y Gualpa, Henry, 2023

Se ha diseñado una serie de preguntas con el fin de evaluar la rigurosidad, la credibilidad y pertinencia de los estudios seleccionados. Para esta evaluación se ha empleado los criterios para

evaluar la calidad del estudio en referencia al autor (Wen et al., 2010) debido a que estos son aplicables a cualquier estudio.

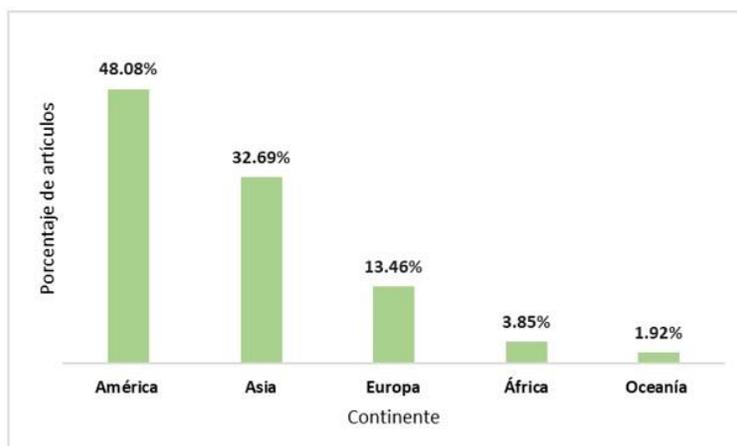
Tabla 3: Criterios para evaluar la calidad de los estudios

Pregunta	ID
¿Están bien definidos los objetivos de la investigación?	PC1
¿Existe una descripción apropiada del contexto en el que se llevó a cabo la investigación?	PC2
¿El diseño de la investigación fue apropiado y justificable?	PC3
¿Se analizan explícitamente las limitaciones del estudio?	PC4
¿Se aplica el experimento en suficientes conjuntos de datos del estudio?	PC5
¿Se compara el método de estimación propuesto con otros métodos?	PC6
¿Los hallazgos del estudio están claramente establecidos?	PC7
¿Los resultados del estudio son de gran valor para la comunidad académica?	PC8

Realizado por: López, María y Gualpa, Henry, 2023

Tras aplicar los criterios de calidad, se determinó que los 52 artículos científicos seleccionados cumplen con un nivel aceptable, puesto que el valor final de calidad es superior a 4 (50% de la puntuación correcta). Estos artículos conforman una muestra significativa que facilitará el desarrollo sólido de la investigación. Se ha identificado los artículos científicos de acuerdo al país de origen y se agrupó de acuerdo al continente al que pertenece.

Se observó que el 48% de las investigaciones sobre rendimiento académico se llevaron a cabo en el continente americano, con Colombia y México destacándose como los principales contribuyentes. En contraste, solo el 1.92% de los artículos provienen del continente oceánico. Dado que los datos no provienen de una distribución normal y la varianza entre los grupos es constante, se optó por aplicar el test no paramétrico de Kruskal-Wallis. Este test se utiliza para determinar si existen diferencias significativas en las métricas de evaluación en relación con los diferentes continentes.

Figura 2: Número de artículos clasificados por continentes

Realizado por: López, María y Gualpa, Henry, 2023

En la Tabla 4 se evidenció un valor p-valor menor que el nivel de significancia 0.05, por tanto, no existe diferencias significativas en las métricas de precisión, puntaje F1 y especificidad (p-valores de 0.1125, 0.1021 y 0.3126 respectivamente), lo que indica que la calidad predictiva de los algoritmos de aprendizaje automático clasificados por continentes es estadísticamente igual en estas métricas. Sin embargo, sí se encontraron diferencias significativas en las métricas de exactitud y sensibilidad (p-valores de 0.02878 y 0.0392 respectivamente), lo que sugiere que la calidad predictiva de al menos dos algoritmos de aprendizaje automático entre continentes es estadísticamente diferente en estas métricas.

El test de Kruscal Wallis ha mostrado la existencia de diferencia significativa entre los algoritmos, pero no se conoce en cuales difieren, por esta razón se ha aplicado el test de Nemenyi para comparar entre pares de continentes.

Precisión:

Tabla 4: Test de Kruscal Wallis de la métrica de precisión, exactitud, sensibilidad especificidad y puntaje f1

Métrica	Estadístico cuadrado	Chi-Grados de libertad	p-valor
Precisión	59.823	3	0.1125
Exactitud	90.386	3	0.02878
Sensibilidad	10.074	4	0.0392
Especificidad	35.635	3	0.3126

Puntaje 62.043 3 0.1021

Realizado por: López, María y Gualpa, Henry, 2023

Mediante el test de Nemenyi (Tabla 5 y 6), se observó que los algoritmos de aprendizaje automático presentan diferencias significativas en las métricas de exactitud y sensibilidad entre los continentes de Asia y América. Sin embargo, para el resto de los continentes, no se encontraron diferencias significativas.

Como es de interés encontrar un algoritmo predictivo del rendimiento académico en nuestro país Ecuador se ha analizado minuciosamente las métricas de evaluación de los algoritmos de Aprendizaje Automático en el continente americano.

Tabla 5: Diferencia significativa de Nemenyi de la exactitud

Continentes	p-valor
América - Africa	1.00
Asia - África	0.71
Europa - África	0.99
Asia - América	0.03
Europa – América	1.00
Europa - Asia	0.13

Realizado por: López, María y Gualpa, Henry, 2023

A continuación, se ha aplicado el test de Nemenyi para comparar entre pares de continentes.

Tabla 6: Diferencia significativa de Nemenyi de la sensibilidad

Continentes	p-valor
América - Africa	0.995
Asia - África	0.852
Europa - África	0.978

Oceanía - África	0.999
Asia - América	0.037
Europa – América	0.996
Oceanía – América	1.000
Europa - Asia	0.122
Oceanía - Asia	0.843
Oceanía - Europa	1.000

Realizado por: López, María y Gualpa, Henry, 2023

Cumpliendo con los supuestos de normalidad y homocedasticidad, aplicamos un análisis de varianza (ANOVA) para evaluar las diferencias en las métricas de evaluación de los algoritmos de aprendizaje automático en el continente americano. Los resultados indican diferencias significativas en las métricas de precisión (p-valor = 0.0206), exactitud (p-valor = 0.0144), sensibilidad (p-valor = 0.0287) y puntaje F1 (p-valor = 0.0208), todas con p-valores menores que el nivel de significancia de 0.05, lo que lleva a rechazar la hipótesis nula. Esto sugiere que, en promedio, al menos dos algoritmos difieren en estas métricas. Sin embargo, no se encontraron diferencias significativas en la métrica de especificidad (p-valor = 0.518), lo que implica que la especificidad de los algoritmos es estadísticamente igual.

Curva ROC

Al analizar las representaciones gráficas (Fig. 3) de la curva ROC y el área bajo la curva (AUC) de los algoritmos encontrados en precisión, se observó que el algoritmo Random Forest presentó un AUC de 89.60%, indicando una mayor capacidad predictiva con aproximadamente un 90% de probabilidad de realizar predicciones correctas. Dado que el valor AUC del algoritmo RF se encuentra en el intervalo [0.75, 0.9), se ha considerado un algoritmo bueno en comparación con RN, cuyo AUC se sitúa en el intervalo [0.6, 0.75), considerándose así un algoritmo regular para el cálculo de predicciones.

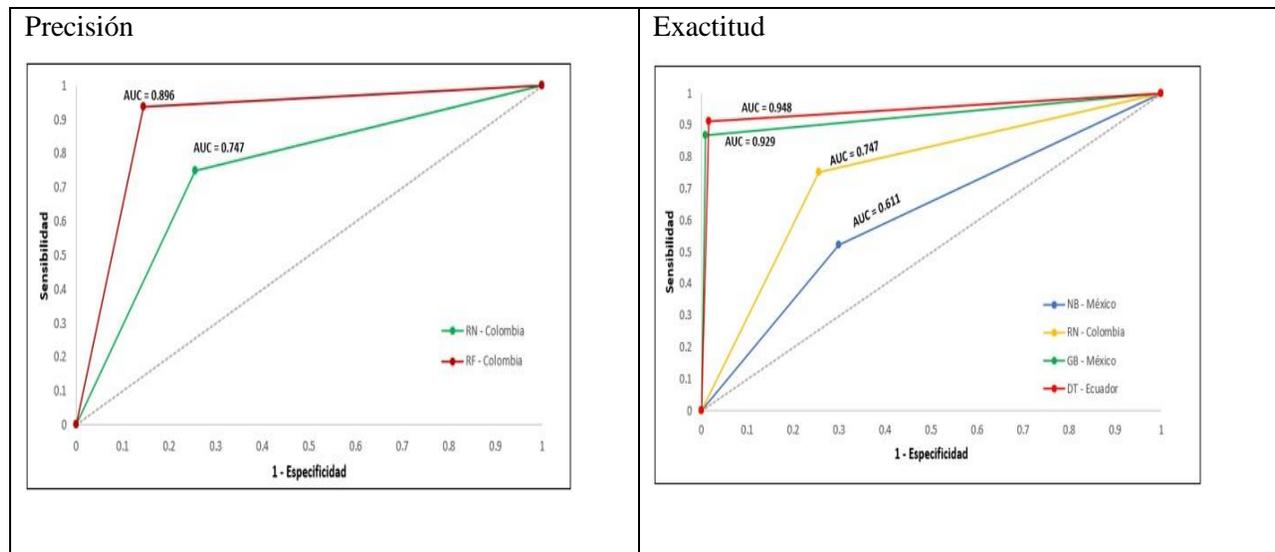
Para la métrica de exactitud, se evidenció que el algoritmo Decisión Tree presentó un AUC de 94.8%, indicando una mayor capacidad predictiva con aproximadamente un 95% de probabilidad

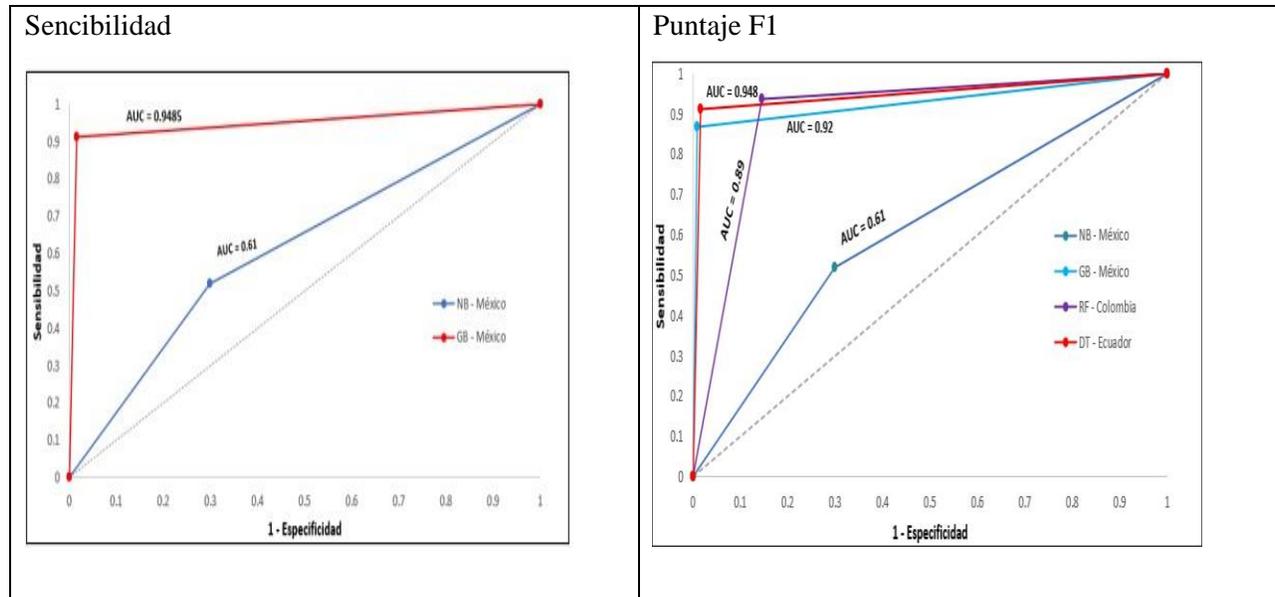
de realizar predicciones correctas. Dado que el valor AUC del algoritmo DT se encuentra en el intervalo [0.9, 0.97), se ha considerado un algoritmo muy bueno en comparación con NB, cuyo AUC se sitúa en el intervalo [0.6, 0.75), considerándose así un algoritmo regular para el cálculo de predicciones.

En cuanto a la sensibilidad, el algoritmo Gradient Boosting presentó un AUC de 94.85%, indicando una mayor capacidad predictiva con aproximadamente un 95% de probabilidad de realizar predicciones correctas. Dado que el valor AUC del algoritmo GB se encuentra en el intervalo [0.9, 0.97), se ha considerado un algoritmo muy bueno en comparación con NB, cuyo AUC se sitúa en el intervalo [0.6, 0.75), considerándose así un algoritmo regular para el cálculo de predicciones.

Para el puntaje F1, el algoritmo Decisión Tree obtuvo un AUC de 94.85%, indicando una mayor capacidad predictiva con aproximadamente un 95% de probabilidad de realizar predicciones correctas. Dado que el valor AUC del algoritmo DT se encuentra en el intervalo [0.9, 0.97), se ha considerado un algoritmo muy bueno en comparación con NB, cuyo AUC se sitúa en el intervalo [0.6, 0.75), considerándose así un algoritmo regular para el cálculo de predicciones.

Figura 3: Curvas ROC de los algoritmos Redes Neuronales y Random Forest





Realizado por: López, María y Gualpa, Henry, 2023

Discusión

La revisión sistemática revela varios estudios sobre la predicción del rendimiento académico, identificando diversos factores que afectan el desempeño de los estudiantes universitarios (García, 2020). Se analizaron diferentes algoritmos de aprendizaje automático, destacando: Random Forest, que clasificó correctamente el 97.51% de los estudiantes potenciales desertores (Páez & Ramírez, 2022) (Huguet-Torres et al., 2024).

Este resultado superó al Gradient Boosting (Chen & Yuanzhao, 2023) por una diferencia del 4.58%. Sin embargo, el Decisión Tree (Buenaño et al., 2019) mostró una exactitud del 96.55% en la clasificación de desertores y no desertores, superior al 91.23% del Random Forest.

En términos de sensibilidad, el Decisión Tree identificó el 98.37% de los estudiantes que desertaron, mientras que la especificidad del Random Forest fue del 93.51%, mayor que el 81.34% del Gradient Boosting. El puntaje F1 del Decisión Tree fue el mejor, alcanzando un 97.86%, en comparación con el 85.51% del Random Forest. El AUC de la curva ROC mostró que el Decisión Tree obtuvo un 94.85%, superior al 89.86% del Gradient Boosting, indicando un mejor umbral de conveniencia. El Decisión Tree se considera el mejor algoritmo debido a su estructura jerárquica, facilidad de interpretación, y bajo costo computacional. Es un algoritmo no paramétrico, por lo que no requiere normalización de datos y es útil tanto para clasificación como para regresión.

En cuanto a la recolección de datos, el estudio de Huguet-Torres et al. (2024) utilizó un cuestionario con 41 preguntas aplicadas a 4987 estudiantes. Buenaño et al. (2019) emplearon datos históricos de 335 estudiantes entre los semestres 2016-1 y 2018-2 en una universidad de Ecuador. Taya et al. (2022) utilizaron una muestra de 3406 estudiantes en los periodos académicos 2019-I y 2020-I, considerando variables como calificaciones finales, acceso al aula virtual, y factores institucionales. Para predecir el rendimiento académico en la ESPOCH, se consideran variables relevantes como la situación laboral del estudiante, tecnología disponible, ingreso familiar, clima familiar, nivel educativo de los padres, entre otras. En contraste, variables como el número de integrantes de la familia y la provincia tienen menor relevancia.

Conclusiones

Para esta investigación se seleccionaron cuatro bases de datos de lectura científica Scopus, SciELO, Dialnet y ERIC debido a su relevancia educativa, facilidad de navegación y capacidad para realizar búsquedas avanzadas. Estas bases de datos permitieron identificar un total de 6,437 artículos científicos publicados en los últimos 10 años. Aplicando las directrices internacionales del método PRISMA, que proporciona un enfoque estructurado para la identificación, cribado e inclusión de estudios, se logró obtener una muestra significativa de 52 artículos relevantes para esta investigación. La aplicación rigurosa del método PRISMA aseguró la calidad y pertinencia de los estudios seleccionados, proporcionando una base sólida para el análisis y las conclusiones de esta investigación.

En este estudio se compararon las métricas de evaluación de algoritmos de Aprendizaje Automático, incluyendo precisión, exactitud, sensibilidad, puntaje F1 y especificidad, utilizando pruebas de hipótesis como ANOVA, Kruskal-Wallis y Nemenyi. Los resultados revelaron que el algoritmo de Decisión Tree es el más eficaz para predecir el rendimiento académico, destacándose con valores superiores al 90% en casi todas las métricas evaluadas. Este rendimiento sobresaliente se debe a varias características del algoritmo: es relativamente fácil de entrenar, predecir e interpretar; es no paramétrico, por lo que no requiere normalización de datos; y presenta un costo computacional menor en comparación con otros algoritmos. Además, el Decisión Tree es variable para tareas de clasificación y regresión, y resulta muy útil en la exploración de datos. Su capacidad para manejar una variedad de variables cualitativas y cuantitativas, seleccionando automáticamente

las más relevantes, lo convierte en una herramienta valiosa para la predicción del rendimiento académico.

Referencias

1. Arias, R., Santa, J. & Veloza, J. (2013). Aplicación del aprendizaje automático con árboles de decisión en el diagnóstico médico. *Cultura del cuidado*, 10(1), 63-72. <https://doi.org/10.18041/1794-5232/cultrua.2013v10n1.2102>
2. Arnau Sabatés, L., & Sala Roca, J. (2013). La revisión de la literatura científica: Pautas, procedimientos y criterios de calidad. *Universitat Autònoma de Barcelona*, 1(1), 1-22. https://ddd.uab.cat/pub/recdoc/2020/222109/revliltcie_a2020.pdf
3. Bobadilla, J. (2021). *Machine learning y deep learning: usando Python, Scikit y Keras*. Ediciones de la U. <https://acortar.link/knhqbH>
4. Bravo, L., Fuentes, H., & 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>
5. Buenaño, D., Gil, D., & Luján, S. (2019). Application of machine learning in predicting performance for computer engineering students: A case study. *Sustainability*, 11(10), 2833. <https://doi.org/10.3390/su11102833>
6. Camps, N. (2023). *Mejora de modelos de predicción de sepsis en pacientes en Unidades de Cuidados Intensivos mediante técnicas de Machine Learning* (Tesis de Grado, Universidad de Alicante). <http://rua.ua.es/dspace/handle/10045/136223>
7. Cervantes, G., Vega, E., & Portilla, E. (2019). Machine Learning para Robots, del Entrenamiento Virtual a la Tarea Real. *Pädi Boletín Científico de Ciencias Básicas e Ingenierías del ICBI*, 7(Especial), 14-18. <https://doi.org/10.29057/icbi.v7iEspecial.4785>
8. Chen, S., & Yuanzhao, D. (2023). A machine learning approach to predicting academic performance in Pennsylvania's schools. *Social Sciences*, 12(3), 118. <https://doi.org/10.3390/socsci12030118>
9. 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. <http://dx.doi.org/10.4067/S0718-50062020000500233>

10. Estrada, O., & Fuentes, D. (2021). ¿ Se logra predecir el rendimiento académico? Un análisis desde la tecnología educativa. *Revista Fuentes*, 23(3), 363-375. <https://doi.org/10.12795/revistafuentes.2021.14278>
11. Garcia, L. (2020). Revisión sistemática sobre la predicción del rendimiento académico en estudiantes: técnicas y algoritmos (Tesis de Grado, Universidad Católica Santo Toribio de Mogrovejo). <http://hdl.handle.net/20.500.12423/3635>
12. González, C., & Guadalupe, E. (2017). Factores que inciden en el rendimiento académico de los estudiantes de la Universidad Politécnica del Valle de Toluca. *Revista Latinoamericana de estudios Educativos*, 47(1), 91-108. <http://ri.iberomx/handle/iberomx/4886>
13. Huguet-Torres, A., Carrero-Planells, A., Fresneda, A., Sebastián-Hernández, R., Sansó, N., Moreno-Mulet, C., & Yáñez, A. (2024). Impacto personal y académico de la pandemia por COVID-19 en los estudiantes de enfermería en España (EsE-COVID). *Medicina Balear*, 38 (5), 108-116. <https://repositori.uib.es/xmlui/handle/11201/164289>
14. Martínez, T. (2022). Comparación de modelos machine learning aplicados al riesgo de crédito. 33-34. (F. d. Universidad de Concepción, Ed.) Chile. <http://repositorio.udec.cl/xmlui/handle/11594/9846>
15. MENASALVAS, H., & et al. (2023). Newsletter trimestral de la Cátedra Idanae:Ética e Inteligencia Artificial -Management Solutions. Universidad Politécnica de Madrid. <https://blogs.upm.es/catedra-idanae/wp-content/uploads/sites/698/2020/01/Idanae-ESP-4T19-LR.pdf>
16. Mireles, M., & García, J. (2022). Satisfacción estudiantil en universitarios: una revisión sistemática de la literatura. *Revista Educación*, 46(2), 610-626. <https://revistas.ucr.ac.cr/index.php/educacion/article/view/47621>
17. Páez, A., & 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>
18. Pedrero, V., Reynaldos, K., Ureta, J., & Cortez, E. (2021). Generalidades del Machine Learning y su aplicación en la gestión sanitaria en Servicios de Urgencia. *Revista médica de Chile*, 149(2), 248-254. <http://dx.doi.org/10.4067/s0034-98872021000200248>

19. Romero, S. (2015). Uso de técnicas de machine learning para predecir el rendimiento académico de los estudiantes de la Carrera de Ingeniería Civil en Informática de la Universidad del Bío-Bío, Chillán (Tesis de Grado, Universidad del Bío Bío) <http://replib.ubiobio.cl/jspui/bitstream/123456789/2610/1/Soto%20Romero%2c%20Gaspar.pdf>
20. Sandoval, L. (2018). Algoritmos de aprendizaje automático para análisis y predicción de datos. *Revista Tecnológica*(11). http://redicces.org.sv/jspui/bitstream/10972/3626/1/Art6_RT2018.pdf
21. Silva, I. (2020). Una metodología sistemática para evaluar los modelos de predicción para la clasificación del estilo de conducción. Madrid (Tesis Doctoral, Universidad Politécnica de Madrid). <https://doi.org/10.20868/UPM.thesis.66065>
22. Taya-Acosta, E., Barraza-Vizcarra, H., Ramirez-Rejas, R., & Taya-Osorio, E. (2022). Academic Performance Evaluation Using Data Mining in Times of Pandemic: Relationship between Access to the Virtual Classroom and Grades of University Students. *TECHNO REVIEW. International Technology, Science and Society Review/Revista Internacional De Tecnología, Ciencia Y Sociedad*, 11(1), 89-106. <https://eaapublishing.org/journals/index.php/technorev/article/view/484>
23. Wen, P., Macdonald, D., Reardon, D., Cloughesy, T., Sorensen, A., Galanis, E., & Chang, S. (2010). Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *Journal of clinical oncology*, 28(11), 1963-1972. <https://ascopubs.org/doi/full/10.1200/JCO.2009.26.3541>