

# MODELADO PREDICTIVO DEL RENDIMIENTO ACADÉMICO EN ESTUDIANTES DE EDUCACIÓN BÁSICA MEDIA DE LA UNIDAD EDUCATIVA “MIGUEL DE CERVANTES”, CANTÓN PUJILÍ, PROVINCIA DE COTOPAXI, MEDIANTE ALGORITMOS DE ÁRBOLES DE DECISIÓN

## PREDICTIVE MODELING OF ACADEMIC PERFORMANCE IN SECONDARY BASIC EDUCATION STUDENTS OF THE “MIGUEL DE CERVANTES” EDUCATIONAL UNIT, PUJILÍ CANTON, COTOPAXI PROVINCE, USING DECISION TREE ALGORITHMS

Carlos Washington Segovia Segovia<sup>1</sup>, Luis Tello Oquendo<sup>2</sup>

{carlos.segovia@unach.edu.ec<sup>1</sup>, luis.tello@unach.edu.ec<sup>2</sup>}

Fecha de recepción: 20/12/2024 / Fecha de aceptación: 04/01/2025 / Fecha de publicación: 06/01/2025

**RESUMEN:** El estudio tuvo como objetivo determinar la influencia de factores personales, pedagógicos y socioeconómicos en el rendimiento académico de estudiantes de educación básica media y desarrollar un modelo predictivo basado en algoritmos de árboles de decisión. Se analizaron características individuales y del entorno social que tienen un impacto significativo en el desempeño escolar. Para verificar esta interrogante, se utilizó una metodología estadística para analizar datos recolectados de estudiantes, considerando variables como edad, género, ingresos familiares, tipo de familia, acceso a internet, horas frente a pantallas, estilo de aprendizaje, lugar de residencia, motivación y salud. Los datos se procesaron mediante técnicas de análisis multivariable, y el modelo predictivo se validó utilizando métricas de precisión y sensibilidad. Los resultados mostraron que factores como el acceso a internet, los ingresos familiares y el apoyo social influyen de manera significativa en el rendimiento académico. Por otro lado, variables como las horas frente a pantallas y los problemas de salud demostraron un impacto negativo moderado. El modelo predictivo logró clasificar con una precisión del 88%, los niveles de éxito académico, proporcionando información útil para la toma de decisiones pedagógicas. Se concluyó que las instituciones educativas deben considerar estos factores en el diseño de estrategias para mejorar el aprendizaje, reducir la deserción y fomentar la equidad educativa. Este hallazgo destaca la importancia de emplear modelos basados en datos para entender y abordar las necesidades

<sup>1</sup>Estudiante de la Escuela de Posgrado, Universidad Nacional de Chimborazo (UNACH), Riobamba 060110, Ecuador, <https://orcid.org/0009-0008-9920-3014>.

<sup>2</sup>Escuela de Posgrado, Universidad Nacional de Chimborazo (UNACH), Riobamba 060110, Ecuador, <https://orcid.org/0000-0002-5274-666X>.

educativas de los estudiantes, promoviendo un entorno de aprendizaje inclusivo, efectivo y adaptado a sus realidades.

*Palabras clave: Árboles de decisión, educación básica, factores socioeconómicos, modelado predictivo, rendimiento académico*

**ABSTRACT:** The study aimed to determine the influence of personal, pedagogical, and socioeconomic factors on the academic performance of middle school students and to develop a predictive model based on decision tree algorithms. Individual characteristics and social environment factors with a significant impact on academic performance were analyzed. To address this question, a statistical methodology was employed to analyze data collected from students, considering variables such as age, gender, family income, family type, internet access, screen time, learning style, place of residence, motivation, and health. The data were processed using multivariable analysis techniques, and the predictive model was validated using precision and sensitivity metrics. The results showed that factors such as internet access, family income, and social support significantly influence academic performance. Conversely, variables like screen use time and health problems demonstrated a moderately negative impact. The predictive model achieved good accuracy in classifying academic success levels, providing valuable insights for pedagogical decision-making. The study concluded that educational institutions should consider these factors when designing strategies to improve learning, reduce dropout rates, and promote educational equity. This finding underscores the importance of leveraging data-driven models to understand and address students' educational needs, fostering an inclusive, effective, and reality-adapted learning environment.

*Keywords: Decision trees, basic education, predictive modeling, academic performance, socioeconomic factors*

## INTRODUCCIÓN

En un esfuerzo significativo y debido a las altas tasas de deserción escolar en Ecuador, el Acuerdo Ministerial que establecía la pérdida de año lectivo para estudiantes de segundo a séptimo de básica fue derogado el 20 de mayo de 2024. Esta decisión aplica desde el año lectivo 2024-2025. La nueva normativa establece que los estudiantes de estos niveles no podrán perder el año por notas. Sin embargo, si un estudiante de quinto a séptimo de básica presenta problemas de aprendizaje, los profesores y representantes podrán decidir que repita el año. Para los estudiantes de octavo de básica hasta el bachillerato, se mantiene el examen supletorio. Aquellos estudiantes que obtuvieron una calificación final entre 4,1 y 6,99/10 puntos pueden rendir el examen para evitar la repetición. Para aprobar, deben alcanzar una calificación igual o mayor a 7/10 puntos (1).

Un sistema educativo es eficiente cuando cumple sus objetivos sin desperdiciar recursos. La repetición y la deserción escolar implican un uso ineficiente de recursos humanos y financieros. La deserción es el último eslabón del fracaso escolar, y a menudo se origina en la repetición; este fenómeno afecta principalmente a sectores pobres y rurales. La deserción tiende a ocurrir

alrededor de los 10 años, edad a la que usualmente los niños con bajos recursos comienzan a trabajar (2).

Actualmente, la minería de datos educativa ha ganado relevancia para analizar la deserción y el rendimiento académico, siguiendo el comportamiento de los estudiantes para prever posibles abandonos. Esta técnica también se emplea para mejorar el proceso educativo y gestionar actividades académicas. El estudio utiliza algoritmos como selección de atributos y árboles de decisión para identificar factores y patrones que ayuden a prevenir la deserción y mejorar el rendimiento (3).

El rendimiento académico de los estudiantes está influenciado por una combinación de factores personales, familiares, económicos y sociales. Se ha observado una conexión notable entre el bajo rendimiento académico y las situaciones de disfunción familiar. Asimismo, el apoyo familiar es fundamental para alcanzar un buen desempeño académico (4).

El rendimiento académico de los estudiantes está estrechamente relacionado con las condiciones socioeconómicas, ya que factores como los ingresos familiares, el nivel educativo de los padres y la disponibilidad de recursos educativos son determinantes clave. Los estudiantes provenientes de entornos con mayores recursos tienden a obtener mejores resultados académicos gracias a su acceso a materiales educativos, un entorno familiar más estable y mayor seguridad. Asimismo, elementos como el estrés familiar y las expectativas parentales también afectan el rendimiento académico, reflejando la influencia del contexto socioeconómico (5).

La inteligencia emocional, según (6), es una de las inteligencias múltiples que incluye habilidades como la autorregulación, la empatía y la motivación, esenciales para gestionar emociones y fortalecer las relaciones interpersonales. En el contexto educativo, la inteligencia emocional contribuye al bienestar psicológico, la regulación emocional y el pensamiento lógico, lo que influye positivamente en el rendimiento académico. Este rendimiento, afectado por factores académicos, económicos y familiares, refleja los logros estudiantiles en el proceso de enseñanza-aprendizaje (7).

Las emociones juegan un papel fundamental en el éxito educativo, tanto en entornos virtuales como presenciales. Estas emociones han redefinido el acto educativo, impulsando experiencias que integran tanto aspectos cognitivos como sensoriales. En este contexto, el aprendizaje se convierte en un proceso integral que no solo aborda el conocimiento académico, sino también el desarrollo emocional y social de los estudiantes. Las emociones permiten una mayor conexión con el contenido educativo, fomentando un ambiente en el que los estudiantes se sienten apoyados y motivados para alcanzar su máximo potencial (8).

Existe una relación positiva entre la actividad física y el rendimiento académico; la actividad física en las escuelas mejora el rendimiento de los niños, incluso superando a aquellos que no practican deporte. La actividad física regular mejora la cognición, la concentración y la autoestima, lo que favorece el aprendizaje. Estudiantes que participan en actividad física adicional muestran mejores cualidades cognitivas y comportamiento, lo que incide en su rendimiento (9).

Estudios previos han explorado la predicción del rendimiento académico utilizando diversos enfoques, entre ellos, los algoritmos de árboles de decisión. A continuación, se presentan algunos estudios destacados:

(3) investigó la predicción del rendimiento académico en estudiantes de secundaria utilizando árboles de decisión. El estudio incluyó a 219 adolescentes y analizó factores como el estatus socioeconómico, índice de masa corporal (IMC), actividad física, tiempo frente a pantallas y niveles emocionales. Los resultados identificaron seis grupos de fracaso académico y tres grupos de éxito, con precisiones del 80.11% en el entrenamiento y 81.40% en la validación. Estos hallazgos destacan la capacidad del modelo para prever el éxito o fracaso escolar, demostrando su eficacia en el entorno educativo.

(10) aplicaron modelos de inteligencia artificial en pruebas estandarizadas para mejorar el rendimiento académico en la educación superior. A través de una revisión sistemática, se analizaron 17 estudios publicados entre 2019 y 2023, donde se destacó el uso de modelos como Redes Neuronales Artificiales y Árboles de Decisión. Los resultados evidenciaron beneficios en la optimización del aprendizaje y la toma de decisiones educativas. Sin embargo, es crucial abordar las limitaciones y considerar aspectos éticos para garantizar una aplicación efectiva y responsable en el ámbito académico.

(11) examinaron el rendimiento académico de estudiantes de Economía y Turismo en la Universidad Técnica de Manabí utilizando Power BI en 2021. La obtención de datos precisos es fundamental para evaluar el desempeño académico, un proceso que puede ser complejo y arduo. La utilización de herramientas como Power BI facilita la recopilación de datos objetivos, permitiendo identificar áreas de mejora y aplicar medidas rápidas y eficaces. Mediante la metodología CRISP-DM, que consta de seis etapas —comprensión del negocio, análisis de datos, preparación, modelado, evaluación y despliegue—, se aplicaron algoritmos de aprendizaje automático, como Árboles de Decisión, Bosques Aleatorios, Redes Neuronales y Máquinas de Soporte Vectorial. El algoritmo de Random Forest resultó ser el más eficiente, generando un dashboard con estadísticas detalladas sobre los estudiantes.

(12) desarrollaron un modelo predictivo basado en árboles de decisión para identificar estudiantes con riesgo de sufrir estrés académico en la Facultad de Ciencias de la Universidad Nacional José Faustino Sánchez Carrión, durante el semestre académico 2023-II. Utilizando un diseño no experimental, con enfoque cuantitativo, nivel explicativo y aplicado, se emplearon dos instrumentos de recolección de datos: la escala Perceived Stress Scale (PSS), con 14 preguntas validadas, y un cuestionario adicional para recopilar factores diversos, compuesto por 23 preguntas. La muestra incluyó 298 estudiantes, y se ajustaron hiperparámetros como  $xval$  (5),  $minsplit$  (7),  $minbucket$  (5),  $maxdepth$  (4) y  $cp$  (0.01). El modelo alcanzó una precisión del 82,93% y un nivel de concordancia del 74,51% en la predicción del estrés académico. Este estudio demuestra la viabilidad del uso de un modelo predictivo basado en árboles de decisión para identificar estudiantes en riesgo de estrés académico.

Esta investigación contribuye de manera significativa al conocimiento existente al identificar y relacionar diversas variables que exploran aspectos abordados en estudios previos, aportando detalles específicos sobre su impacto a través de un análisis exhaustivo de correlaciones y el uso

de modelos predictivos avanzados, como el árbol de decisión. A diferencia de investigaciones previas centradas principalmente en la educación superior, este estudio considera también los niveles de educación básica, un ámbito frecuentemente ignorado. En comparación con otras investigaciones, ofrece un análisis más detallado y granular de cómo interactúan los diferentes factores, destacando su aplicabilidad práctica mediante modelos interpretables que facilitan una comprensión más profunda y accesible de los resultados. Además, permite abordar la incógnita central de la investigación: ¿Cuáles son los factores personales y académicos más significativos en la población de estudiantes de la Unidad Educativa "Miguel de Cervantes"?

El estudio proporciona valiosos hallazgos sobre el rendimiento académico utilizando técnicas avanzadas como mapas de calor y árboles de decisión. Se emplea un mapa de calor para analizar las correlaciones entre variables clave, como la relación positiva entre ingresos familiares y el rendimiento académico, y la correlación negativa entre las horas frente a pantallas y el desempeño. Además, el modelo de árbol de decisión, tras ser ajustado y validado, muestra una precisión superior al 90%. Entre los principales factores observados, la "Horas frente a Pantallas" es determinante: El 33.3% de los estudiantes que dedican más de 2 horas diarias a actividades de ocio frente a pantallas solo "Alcanza los aprendizajes requeridos", mientras que el 66.7% restante, con menos tiempo frente a pantallas, tiene mayores posibilidades de superar los aprendizajes.

Además, los ingresos familiares y la actividad extracurricular influyen positivamente en el rendimiento, destacando que los estudiantes con ingresos mayores y participación extracurricular tienen un 78.6% de probabilidades de "Superar los aprendizajes requeridos". Por otro lado, el acceso a internet y la motivación también impactan el rendimiento, siendo el primero un factor positivo y la falta de motivación un aspecto negativo.

## MATERIALES Y MÉTODOS

### Tipo de Investigación

Considerando que el enfoque cuantitativo utiliza la recolección y análisis de datos para responder preguntas, probar hipótesis y establecer patrones mediante medición, conteo y estadística (13), esta investigación aplicada con enfoque cuantitativo se centra en la construcción y validación de un modelo predictivo para estimar el rendimiento académico en estudiantes de educación básica media. Este enfoque permite analizar de manera objetiva la influencia de factores personales, pedagógicos y psicosociales en el desempeño estudiantil. Mediante un proceso riguroso de recolección y análisis de datos, se busca identificar patrones significativos y relaciones entre las variables involucradas, utilizando algoritmos de árboles de decisión como herramienta clave para el modelado predictivo. La metodología de la investigación se estructura en etapas definidas: selección de variables relevantes, preprocesamiento de datos, desarrollo del modelo y validación estadística. Este diseño garantiza no solo la precisión del modelo, sino también su aplicabilidad en contextos educativos similares, contribuyendo a la mejora de estrategias pedagógicas y a la identificación temprana de riesgos asociados a la deserción escolar.

## **Población**

La población de esta investigación está conformada por las calificaciones de los estudiantes de educación básica media de la Unidad Educativa "Miguel de Cervantes", ubicada en el Cantón Pujilí, Provincia de Cotopaxi. Este grupo incluye a todos los estudiantes matriculados durante los periodos académicos: 2022-2023 y 2023-2024, quienes presentan una variedad de características personales, pedagógicas y psicosociales relevantes para el análisis del rendimiento académico, y cuyo número alcanza los 30 alumnos en cada periodo académico.

En este estudio, se trabaja con el total de estudiantes de educación básica media legalmente matriculados es de 30. Al trabajar con la totalidad de los estudiantes, se garantiza una cobertura completa y detallada del fenómeno de estudio, eliminando la necesidad de un muestreo adicional y asegurando la precisión y representatividad de los resultados. Este enfoque permite un análisis directo del impacto de los diversos factores sobre el rendimiento académico, minimizando sesgos derivados de la selección de una submuestra.

## **Recopilación de datos**

En esta investigación para el modelado predictivo del rendimiento académico en estudiantes de educación básica media, la recopilación de datos se realizó mediante encuestas dirigidas a los estudiantes. La encuesta es una técnica que emplea procedimientos estandarizados de investigación, para recoger y analizar de manera sistemática diversos datos (14).

Las encuestas abordaron múltiples variables relevantes para el análisis del rendimiento académico, como: Edad del estudiante, Género, Ingresos Familiares Mensuales, Tipo de Familia, Personas en el Hogar, Actividad Extracurricular, Acceso a Internet, Horas Frente a Pantallas, Estilo de Aprendizaje, Zona, Problemas de Salud, Motivación, Deseo de Continuar Estudios y Nivel Académico (usadas como base para evaluar el modelo predictivo).

Esta recopilación de datos ofrece un panorama integral de las variables personales, pedagógicas y socioeconómicas que pueden influir en el rendimiento académico de los estudiantes, permitiendo desarrollar un modelo predictivo robusto basado en estos factores.

## **Encuesta**

La investigación implementó una encuesta como instrumento de recolección de datos diseñado para analizar comprensivamente los factores que influyen en el rendimiento académico de los estudiantes. La encuesta, estructurada metodológicamente en ocho secciones, aborda múltiples dimensiones que potencialmente impactan el desempeño escolar, permitiendo una evaluación sistemática y multifactorial de las condiciones estudiantiles.

El diseño metodológico contempló variables demográficas, socioeconómicas y contextuales, estratificadas en categorías específicas que permiten un análisis detallado. La sección de información general incluye variables como edad, género y nivel socioeconómico, clasificado en tres rangos: bajo (ingresos <\$250 mensuales), medio (\$250-\$500 mensuales) y alto (>\$500 mensuales). Esta estratificación permite comprender la relación entre condiciones económicas y rendimiento académico.

La estructura del instrumento abarca dimensiones críticas como composición familiar, participación en actividades extracurriculares, acceso tecnológico y estilos de aprendizaje. Se consideraron variables como tipos de familia (nuclear, monoparental, extendida), número de habitantes en el hogar, acceso a internet y tiempo de exposición a pantallas, categorizado en intervalos que van desde menos de una hora hasta más de cuatro horas diarias. El diseño metodológico priorizó la obtención de información mediante preguntas cerradas y algunas abiertas, equilibrando la recopilación de datos cuantitativos y cualitativos.

Entre las consideraciones metodológicas se contemplaron limitaciones inherentes, como el potencial sesgo de deseabilidad social y la dependencia de información autorreportada. No obstante, el instrumento fue concebido para proporcionar una perspectiva integral, permitiendo analizar la compleja interrelación de factores que configuran el rendimiento académico.

La participación fue establecida sobre bases voluntarias, con un compromiso explícito de utilizar la información únicamente con fines académicos y de investigación, asegurando la transparencia y rigurosidad del proceso de recolección de datos.

### **Análisis estadístico**

El análisis estadístico realizado en este estudio empleó el algoritmo de árboles de decisión CART. Este algoritmo fue seleccionado debido a su capacidad para manejar datos complejos y proporcionar modelos interpretables, lo que es esencial para comprender las relaciones entre las variables involucradas en el rendimiento académico (15).

### **Proceso de análisis:**

*El lenguaje de programación utilizado en esta investigación fue Python, con el apoyo de librerías como Pandas, NumPy, Seaborn, Matplotlib y Scikit-learn para el procesamiento y análisis de datos. El script realiza transformaciones categóricas y de rangos, y elimina columnas irrelevantes para optimizar el análisis. Se genera una matriz de correlación entre variables, visualizada mediante un mapa de calor con Seaborn. Los datos se dividen en conjuntos de entrenamiento y prueba utilizando Scikit-learn. La configuración de Pandas permite mostrar todos los datos, facilitando un análisis detallado. El modelo principal es un Árbol de Decisión ajustado con el criterio de Gini y una profundidad máxima de 4. Se lleva a cabo una validación cruzada estratificada con StratifiedKFold, y el modelo se evalúa mediante precisión, reporte de clasificación y matriz de confusión. A continuación, se detalla cada aspecto del análisis:*

1. **Preprocesamiento de datos:** Los datos recolectados fueron sometidos a un riguroso proceso de preprocesamiento, incluyendo limpieza, normalización y manejo de valores atípicos. Este paso es crucial para asegurar la calidad y precisión de los datos analizados. Los datos fueron tratados utilizando un mapeo estructurado para convertir variables categóricas en valores numéricos, lo que facilita el análisis estadístico y la modelización predictiva. Cada variable fue asignada a un valor según su correspondiente correspondencia: el género se codificó como 0 para 'Femenino' y 1 para 'Masculino'; los ingresos familiares mensuales se categorizan como 0 para 'Menos de 250 USD', 1 para 'Entre 250 a 500 USD' y 2 para 'Más de 500 USD'. El tipo de familia se codificó como 0 para 'Monoparental' y 1 para 'Nuclear',

mientras que las personas en el hogar se dividieron en 0, 1 y 2 según el número de integrantes. La actividad extracurricular fue categorizada como 0 para 'No' y 1 para 'Sí', y el acceso a internet se asignó como 0 para 'No' y 1 para 'Sí'. Asimismo, las horas frente a pantallas se clasificaron en 0 para 'Hasta 1 hora', 1 para 'Entre 1 y 2 horas' y 2 para 'Hasta 4 horas'. El estilo de aprendizaje fue mapeado como 1 para 'Auditivo' y 2 para 'Visual', y la zona como 1 para 'Urbana' y 2 para 'Rural'. Los problemas de salud se codificaron como 0 para 'No' y 1 para 'Sí', mientras que la motivación fue categorizada en 0 para 'Graduarme con honores', 1 para 'Aprender' y 2 para 'Conseguir buen empleo'. Finalmente, el deseo de continuar estudios se asignó como 1 para 'Sí' y 0 para 'No'. Además, la variable edad fue segmentada en rangos y codificada en valores específicos según corresponda. Este enfoque permite un análisis comprensivo y un modelado efectivo basado en datos estructurados.

2. Entrenamiento del modelo: Una vez preprocesados, los datos fueron utilizados para entrenar el modelo CART. Este algoritmo segmenta los datos en grupos homogéneos mediante divisiones sucesivas, optimizando la relación entre las variables independientes y el rendimiento académico. La fórmula general del árbol es:

$$Rendimientoacadémico = f(Variable_1, Variable_2, \dots, Variable_n) \quad (1)$$

donde  $f$  representa la función de predicción basada en las variables seleccionadas.

3. Validación cruzada: Para evaluar la precisión del modelo y prevenir el sobreajuste, se aplicó una validación cruzada. En este proceso, se divide el conjunto de datos en subconjuntos, donde algunos se utilizan para el entrenamiento y otros para la validación. La fórmula para la validación cruzada es (Scikit-learn developers, 2018):

$$Precisión = \frac{1}{k} \sum_{i=1}^k Error_i \quad (2)$$

donde  $k$  representa el número de pliegues o subconjuntos.

Resultados y evaluación: Los resultados del análisis mostraron que el modelo CART fue altamente eficaz para identificar los principales factores que influyen en el rendimiento académico. Además, la precisión del modelo permitió predecir el desempeño estudiantil con alta exactitud, facilitando la intervención temprana en casos de riesgo. Este enfoque no solo optimiza la toma de decisiones pedagógicas, sino que también contribuye a la reducción de la deserción escolar al proporcionar información detallada sobre áreas específicas a mejorar.

## RESULTADOS

### Mapa de calor

El mapa de calor, ilustrado en la figura 1, se basa en una matriz de correlación, que permite analizar la relación entre diversas variables relacionadas con aspectos demográficos, familiares y

académicos de los estudiantes. Cada celda de la matriz muestra el coeficiente de correlación entre pares de variables, proporcionando una visión cuantitativa de la asociación entre ellas. Los valores en la matriz van desde -1 hasta 1, donde valores cercanos a 1 indican una fuerte relación positiva, valores cercanos a -1 reflejan una relación negativa significativa, y valores cercanos a 0 sugieren una relación débil o nula (16). Esta representación visual ayuda a identificar cómo factores como ingresos familiares, estilo de aprendizaje y motivación influyen mutuamente en el rendimiento académico y las experiencias educativas.

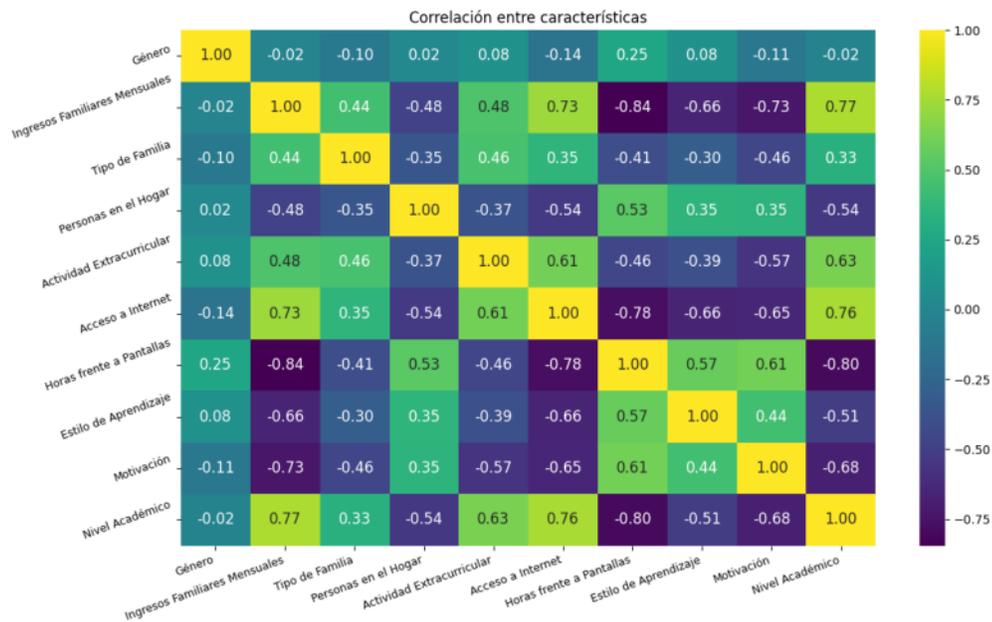


Figura 1. Mapa de calor que representa la correlación entre variables influyentes en el rendimiento académico.

La matriz de correlación presentada en la Figura 1, ofrece un análisis detallado de cómo diferentes factores están relacionados entre sí en el contexto educativo y familiar de los estudiantes. Es importante considerar que factores como la "zona", "problemas de salud", "edad" y "deseos de continuar estudios" fueron eliminados de la matriz debido a la homogeneidad en estos aspectos dentro de la población estudiada. A continuación, se destacan algunas interpretaciones relevantes:

#### Eliminación de factores:

1. Zona: Al vivir todos los niños en una zona rural cercana a la unidad educativa, no se presentan variaciones significativas en este aspecto.
2. Problemas de Salud: No se identificaron problemas de salud graves entre los estudiantes, lo que elimina la necesidad de incluir esta variable en el análisis.
3. Deseos de Continuar Estudios: Todos los estudiantes manifestaron deseos uniformes de continuar sus estudios, lo que suprime variaciones en este factor.
4. Edad: La falta de diversidad en esta variable reduce su relevancia en la identificación de patrones o diferencias significativas entre los estudiantes, ya que la mayoría se encuentra

en un rango de edad muy cercano. Por lo tanto, su inclusión no aporta información diferenciadora para el modelo.

### Árbol de decisión

En este estudio se empleó un árbol de decisión para modelar la relación entre las características de los datos y la variable objetivo. Los árboles de decisión son una herramienta popular en el ámbito de la minería de datos y la inteligencia artificial debido a su capacidad para modelar relaciones no lineales de manera interpretable. Estos modelos dividen el espacio de características en regiones homogéneas basadas en reglas simples, lo que permite entender fácilmente el proceso de toma de decisiones (17).

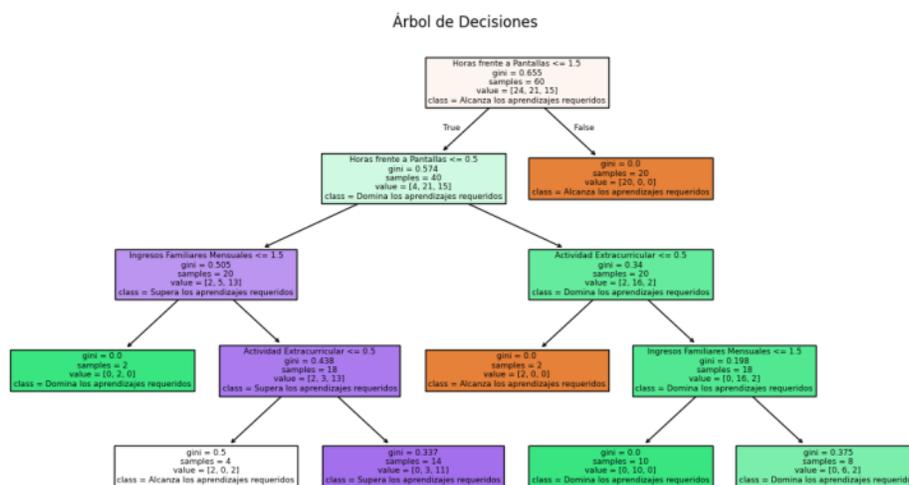


Figura 2. Representación gráfica del árbol de decisión utilizado en el estudio.

En la Figura 2 se puede observar el árbol de decisión resultante, el modelo evaluado muestra un alto nivel de precisión, con una precisión de validación cruzada promedio de 0.83 y una precisión final del modelo de 0.88. Además, al analizar los resultados en cada uno de los pliegues, se observa una variabilidad entre 0.75 y 0.92, indicando que el modelo mantiene un desempeño consistente en diferentes subdivisiones de los datos. El reporte de clasificación muestra cómo se distribuye el rendimiento según las categorías de desempeño académico: Alcanza los aprendizajes requeridos, Domina los aprendizajes requeridos y Supera los aprendizajes requeridos.

Los valores de precisión, recall y f1-score para cada categoría reflejan un buen equilibrio entre los distintos indicadores de desempeño. Por ejemplo, la categoría "Alcanza los aprendizajes requeridos" muestra una precisión del 0.92 y una sensibilidad del 1.00, lo que sugiere una alta capacidad para identificar correctamente los casos positivos. Asimismo, la categoría "Domina los aprendizajes requeridos" muestra una precisión del 0.90 y una sensibilidad de 0.86, destacando una buena identificación de este grupo. Finalmente, la categoría "Supera los aprendizajes requeridos" presenta una precisión del 0.79 con una sensibilidad de 0.73. Esto sugiere que el

modelo es efectivo en la clasificación de los aprendizajes requeridos y puede ser un recurso valioso para la evaluación del rendimiento académico en contextos similares.

### Matriz de confusión

La matriz de confusión ilustrada en la figura 3 muestra cómo el modelo clasifica los resultados en tres categorías distintas: "Alcanza los aprendizajes requeridos", "Domina los aprendizajes requeridos" y "Supera los aprendizajes requeridos". De acuerdo con los valores presentados, el modelo clasifica correctamente la mayoría de los casos en cada categoría. Por ejemplo, logró identificar correctamente 24 casos en la categoría "Alcanza los aprendizajes requeridos". Sin embargo, también se observa cierta confusión en las clasificaciones, como los 3 casos incorrectamente clasificados como "Supera los aprendizajes requeridos" dentro de la categoría "Domina los aprendizajes requeridos". Por lo que la matriz de confusión refleja un desempeño satisfactorio del modelo.

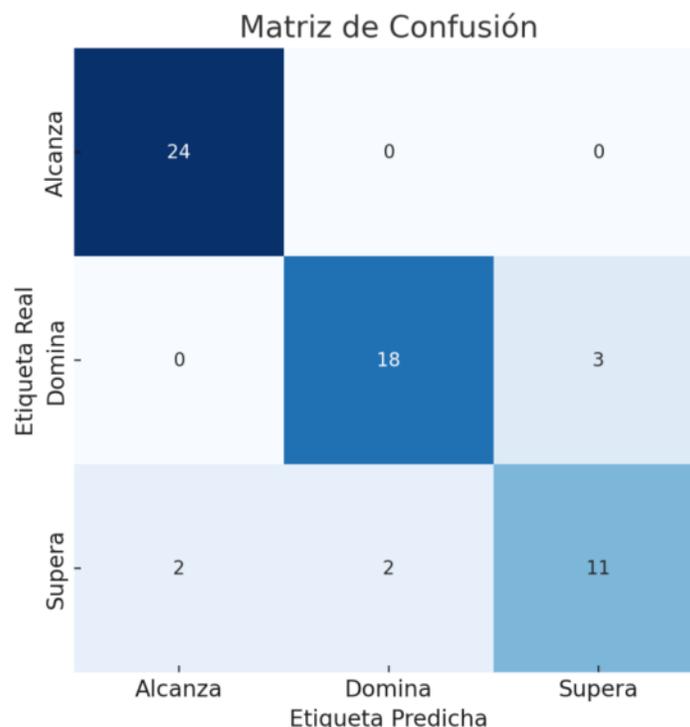


Figura 3. Matriz de confusión

## DISCUSIÓN

### Interpretación de la matriz de correlación

La matriz de correlación muestra diversas relaciones entre las variables analizadas y cómo estas influyen en el nivel académico de los estudiantes. En particular, se destaca la correlación positiva entre los ingresos familiares mensuales y el nivel académico (0.77), indicando que a medida que aumentan los ingresos, también lo hace el desempeño académico. Esta relación sugiere que los

recursos económicos pueden proporcionar acceso a mejores oportunidades educativas, tecnologías y apoyo adicional, lo que contribuye a un mejor rendimiento académico. Además, la correlación positiva con acceso a internet (0.76) subraya la relevancia de las tecnologías digitales como un factor determinante para el nivel académico, ya que un acceso eficiente a estas herramientas facilita la realización de tareas académicas y la interacción educativa (18).

Por otro lado, la correlación negativa con horas frente a pantallas (-0.80) indica que un uso excesivo de dispositivos electrónicos está asociado a un menor nivel académico. Esto podría estar relacionado con la distracción o el tiempo reducido dedicado a actividades académicas más productivas. Asimismo, se observa una relación moderada entre la actividad extracurricular y el nivel académico (0.63), lo que sugiere que la participación en estas actividades complementarias contribuye positivamente al desarrollo de habilidades y al rendimiento académico (19).

Finalmente, la correlación con la motivación académica también es significativa (-0.68), mostrando que estudiantes más motivados suelen alcanzar niveles superiores en su rendimiento. En conjunto, la matriz resalta cómo factores como ingresos, acceso a tecnología, participación extracurricular y motivación impactan directamente en el desempeño académico, mostrando la complejidad de las interacciones que afectan el éxito educativo.

### **Interpretación del árbol de decisión**

El árbol de decisiones clasifica a los estudiantes basándose en su tiempo frente a pantallas, ingresos familiares mensuales y participación en actividades extracurriculares. La clasificación comienza evaluando las "Horas frente a Pantallas" como criterio principal, dividiéndose en dos ramas principales: una para valores menores o iguales a 1.50 (tiempo menor o igual a dos horas diarias frente a pantallas) y otra para valores mayores a 1.50 (tiempo mayor a más de 2 horas diarias frente a pantallas).

En la primera rama, donde las "Horas frente a Pantallas menor o igual que 1.50", se profundiza en factores adicionales como ingresos familiares mensuales y actividad extracurricular. Si los ingresos son bajos o iguales a 1.50 (ingresos familiares menores a \$500), los estudiantes se clasifican como "Domina los aprendizajes requeridos", que es el rango más bajo de calificaciones. En contraste, si los ingresos son mayores, la actividad extracurricular se convierte en el siguiente criterio, resultando en clasificaciones como "Alcanza" o "Supera los aprendizajes requeridos". Esto sugiere que las actividades extracurriculares potencian el rendimiento académico, especialmente cuando los ingresos son medios o altos (20).

En la segunda rama, donde las "Horas frente a Pantallas mayor que 1.50" (tiempo mayor a más de 2 horas diarias frente a pantallas), todos los casos se agrupan en la categoría "Alcanza los aprendizajes requeridos". Esto indica que, a partir de un umbral alto de horas frente a pantallas, los estudiantes tienen una clasificación uniforme en cuanto a rendimiento académico, sin distinciones adicionales basadas en otros factores.

## **CONCLUSIONES**

La variable “horas frente a pantallas” es el principal factor determinante en los niveles de aprendizaje de los estudiantes. Aquellos que dedican más de dos horas diarias a actividades de ocio frente a pantallas caen exclusivamente en la categoría “Alcanza los aprendizajes requeridos”, representando el 33.3% de los estudiantes (10 de 30 muestras), sin lograr superarla. En contraste el 66.7% restante que limita su tiempo frente a pantallas a menos de 2 horas diarias muestra un desempeño más diverso, con una mayor posibilidad de alcanzar o superar los aprendizajes requeridos, destacando el impacto positivo de un uso moderado de dispositivos electrónicos.

Los factores "Ingresos Familiares Mensuales" y "Actividad Extracurricular" tienen una influencia significativa en el rendimiento académico. En el grupo que dedica menos de 1 hora diaria a actividades de ocio frente a pantallas y proviene de familias con ingresos menores a 500 USD, el 100% de los estudiantes "Dominan los aprendizajes requeridos", aunque no logran superarlos. En cambio, cuando los ingresos son mayores y los estudiantes participan en actividades extracurriculares, su rendimiento mejora considerablemente, ya que el 78.6% de ellos "Superan los aprendizajes requeridos". Esto resalta el impacto positivo de estas actividades en el desarrollo académico de los estudiantes.

El 65% de los estudiantes con mayores recursos económicos logran "Superar los aprendizajes requeridos", lo que resalta cómo los factores económicos influyen en el rendimiento académico. Además, el acceso a internet, con una correlación de 0.76, impulsa el rendimiento de 70% de los estudiantes, facilitando el acceso a recursos educativos y mejorando su desempeño. En contraste, la falta de motivación académica, reflejada en una correlación de -0.68, tiene un impacto negativo significativo en el rendimiento, demostrando cómo la motivación insuficiente puede limitar el potencial académico de los estudiantes

## DECLARACIÓN DE INTERÉS (OPCIONAL)

Se declara que no existen conflictos de interés de naturaleza alguna con la presente investigación. Durante la realización de este estudio, se han seguido principios éticos y metodológicos estrictos para asegurar la objetividad y la integridad de los resultados, evitando cualquier sesgo o influencia externa que pueda afectar la validez y la fiabilidad de los hallazgos. Además, se han tomado todas las medidas necesarias para mantener la independencia en la recolección, análisis y presentación de datos, garantizando que la investigación se desarrolle de manera imparcial y transparente.

## REFERENCIAS BIBLIOGRÁFICAS

1. Ministerio de Educación del Ecuador. (2024). Acuerdo Ministerial Nro. MINEDUC-MINEDUC-2024-00031-A. "Ministerio de Educación. Quito, Ecuador."
2. Moreno, D. M., & González, A. M. (2005). Deserción escolar. *Revista Internacional de Psicología*, 6(1), 1–3.
3. Villarrasa-Sapiña, I. (2024). Predicción del rendimiento académico en educación secundaria mediante el análisis de árboles de decisión. *Revista de Psicología Educativa*, 22(3), 25–40.

5. Real-Delor, R. E., Tirado, A. G., Ojeda, I. A. M., Muñóz, E. E. C., Cáceres, E. D. C., Almeida, M. J. C., & Vera, G. R. G. (2024). Factores asociados al rendimiento académico en estudiantes universitarios de Latinoamérica en 2023. *Investigación en Educación Médica*, 13(51), 42–52.
6. Montesdeoca, M. C. (2024). La influencia del entorno socioeconómico en el rendimiento académico. *Dominio de las Ciencias*, 10(2), 1488–1498.
7. Rodríguez-Barboza, D. J. R. (2024). Inteligencia Emocional como Factor Determinante en el Rendimiento Académico en Estudiantes. *Revista Tecnológica-Educativa Docentes 2.0*, 17(1), 400-411.
8. Cañas, F. A. C., Serna, M. D. S. G., & Pabón, D. C. A. (2024). Incidencia de las emociones en el rendimiento académico y el desarrollo de la autonomía en estudiantes de educación superior virtual: estudio en tres universidades colombianas. *Panorama*, 18(34), 103-118.
9. Ramírez, W., Vinaccia, S., & Ramón Suárez, G. (2004). El impacto de la actividad física y el deporte sobre la salud, la cognición, la socialización y el rendimiento académico: una revisión teórica. *Revista de estudios sociales*, (18), 67-75.
10. Vega-Malagón, G., Ávila-Morales, J., Vega-Malagón, A. J., Camacho-Calderón, N., Becerril-Santos, A., & Leo-Amador, G. E. (2014). Paradigmas en la investigación. Enfoque cuantitativo y cualitativo. *European Scientific Journal*, 10(15).
11. López Tejada, V., & Pérez Guarachi, J. F. (2011). Técnicas de recopilación de datos en la investigación científica. *Revista de Actualización Clínica Investiga*, 10, 485.
12. Sangüesa, R. (2005). Métodos de construcción de árboles de decisión para regresión y clasificación (CART). In *Data mining* (p. 12). Editorial UOC.
13. 1García, E. M., López, C. C., Rivas, J. A. M., & Capistran, D. L. A. (2024). Evaluación de Algoritmos de Aprendizaje Supervisado usando Modelos Binarios para Clasificación de Análisis de Sentimiento: Evaluation of Supervised Learning Algorithms Using Binary Models for Sentiment Analysis Classification. *Tecnología Educativa Revista CONAIC*, 11(1), 92-97.
14. Pertuz, C. M. P. (2022). *Aprendizaje automático y profundo en python*. Ra-Ma Editorial.
15. Origel-Rivas, C. G., Lara, E. R., Barrera, I. A., & Eleuterio, R. A. (2020). Redes neuronales artificiales y árboles de decisión para la clasificación con datos categóricos. *Res. Comput. Sci.*, 149(8), 541-554.
16. Gardner, H. (1987). La teoría de las inteligencias múltiples. *Santiago de Chile: Instituto Construir*. Recuperado de <http://www.institutoconstruir.org/centro superacion/La%20Teor%EDa%20de,20,287-305>.
17. Villarrasa-Sapiña, I. (2024). Predicción del rendimiento académico en educación secundaria mediante el análisis de árboles de decisión.
18. Morales, N. O., & García, P. A. O. (2024). Aplicación de modelos de inteligencia artificial en pruebas estandarizadas para la optimización del rendimiento académico en educación superior. *European Public & Social Innovation Review*, 9, 1-21.
19. Mero, J. S., & Felipe, M. C. (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.

20. Lopez Quiroz, L. A., & Soto Salazar, J. G. (2024). Árboles de decisión para la predicción temprana de estrés académico en estudiantes de la Facultad de Ciencias, 2023.
21. Scikit-learn developers. (2018, July 27). *scikit-learn user guide*. Release 0.19.2. [https://scikit-learn.org/0.19/\\_downloads/scikit-learn-docs.pdf](https://scikit-learn.org/0.19/_downloads/scikit-learn-docs.pdf)