

Optimization of Inclusive Educational Programs through Statistical Modeling and Data Mining

Optimización de Programas Educativos Inclusivos mediante Modelación Estadística y Minería de Datos

Mayuri Viviana Pico Gordón
<https://orcid.org/0009-0006-7838-6232>
mayuri_pico@hotmail.com
Escuela superior Politecnica de Chimborazo
Ecuador

Abstract.- This article presents an analysis of the optimization of inclusive education programs using advanced statistical modeling and data mining techniques. The objective is to improve educational equity by identifying factors that influence academic performance and evaluating the effectiveness of educational interventions. Models such as Random Forest and Gradient Boosting are used to predict educational outcomes, showing moderate performance, with Gradient Boosting slightly superior. Key factors identified include prior academic performance, available resources, and the absence of special educational needs (SEN). The importance of integrating advanced statistical and analytical methods with ethical and contextual considerations to ensure inclusive and sustainable education policies is highlighted. The study concludes that these approaches allow for a better understanding of the impact of educational variables and support informed decision-making.

Keywords: *education, data, mining, models, optimization.*

Resumen.- El artículo presenta un análisis sobre la optimización de programas educativos inclusivos mediante el uso de modelado estadístico avanzado y técnicas de minería de datos. El objetivo es mejorar la equidad educativa al identificar factores que influyen en el rendimiento académico y evaluar la efectividad de las intervenciones educativas. Se emplean modelos como Random Forest y Gradient Boosting para predecir resultados educativos, mostrando un desempeño moderado, con una ligera superioridad del Gradient Boosting. Los factores clave identificados incluyen el rendimiento académico previo, los recursos disponibles y la ausencia de necesidades educativas especiales (NEE). Se destaca la importancia de integrar métodos estadísticos y analíticos avanzados con consideraciones éticas y contextuales para garantizar políticas educativas inclusivas y sostenibles. El estudio concluye que estos enfoques permiten una mejor comprensión del impacto de las variables educativas y apoyan la toma de decisiones informada.

Palabras clave: *Educación, Datos, Minería, Modelos, Optimización.*

1. Introducción

La optimización de la evaluación de programas educativos inclusivos mediante técnicas avanzadas de modelado estadístico y minería de

datos es un campo de estudio crítico que tiene como objetivo mejorar la equidad educativa para estudiantes de diversas habilidades y antecedentes [1]. La educación inclusiva se esfuerza por garantizar que todos los estudiantes reciban oportunidades de aprendizaje

equitativas, pero a menudo enfrenta desafíos como la falta de recursos, la escasez de formación especializada para los educadores y las diversas necesidades de los estudiantes.

Estos desafíos requieren métodos de evaluación robustos para evaluar la efectividad de las prácticas inclusivas y para informar la política y la práctica educativa. Los avances recientes en modelización estadística y minería de datos proporcionan herramientas poderosas para analizar conjuntos de datos educativos grandes y complejos [2]. Al emplear metodologías como el modelado multinivel, los enfoques bayesianos y las técnicas de aprendizaje automático, los investigadores y responsables de políticas pueden obtener valiosos conocimientos sobre el rendimiento estudiantil, identificar estrategias de enseñanza efectivas y predecir resultados para diversas poblaciones estudiantiles.

Estos métodos permiten una comprensión matizada de los factores que influyen en el éxito educativo, lo que en última instancia conduce a una toma de decisiones informada que promueve la inclusión y la equidad en la educación. Sin embargo, la integración de estas técnicas avanzadas también plantea consideraciones éticas significativas, incluyendo la privacidad de los datos y los posibles sesgos en la toma de decisiones algorítmica [3].

La dependencia de conjuntos de datos propietarios y la brecha digital pueden exacerbar las desigualdades existentes, destacando la importancia de las prácticas éticas y el acceso equitativo en la evaluación educativa [4]. Además, la efectividad de estos modelos puede variar en diferentes contextos educativos, lo que requiere un refinamiento y validación continuos para asegurar su aplicabilidad en diversos entornos.

Por todo ello, se puede decir que optimizar la evaluación de programas educativos inclusivos mediante técnicas avanzadas de modelado

estadístico y minería de datos es crucial para mejorar los resultados educativos [5]. A medida que el campo evoluciona, es esencial equilibrar el uso de herramientas analíticas sofisticadas con consideraciones éticas y un compromiso con la inclusividad, asegurando que todos los estudiantes se beneficien de los avances en la evaluación y la política educativa.

La educación inclusiva tiene como objetivo proporcionar oportunidades de aprendizaje equitativas para todos los estudiantes, independientemente de sus habilidades o discapacidades. Sin embargo, este enfoque educativo enfrenta varios desafíos que pueden obstaculizar su efectividad. Los problemas clave incluyen la falta de formación especializada para los maestros, recursos insuficientes y la necesidad de sistemas de apoyo adecuados en las aulas para acomodar diversas necesidades de aprendizaje [6].

Comprender estos desafíos es esencial para desarrollar soluciones efectivas y mejorar el entorno general de la educación inclusiva. Los factores que influyen en la educación inclusiva van más allá de la dinámica del aula; también abarcan la demografía estudiantil, los planes de estudio y la calidad de la enseñanza. La investigación indica que datos de antecedentes como la educación de los padres, los ingresos familiares y el registro del hogar pueden mejorar significativamente los modelos predictivos del rendimiento educativo [7].

Estos hallazgos sugieren que una comprensión integral del panorama educativo es crucial para abordar las diversas dimensiones de la educación inclusiva. Estos hallazgos sugieren que una comprensión integral del panorama educativo es fundamental para abordar las diversas dimensiones de la educación inclusiva. En estudios recientes, se han empleado diversas metodologías para analizar los desafíos y las eficiencias de los sistemas educativos. Técnicas como el modelado estadístico y la minería de

datos han surgido como herramientas valiosas para extraer información de grandes conjuntos de datos. Estas metodologías no solo ayudan a identificar los factores que contribuyen a una educación inclusiva exitosa, sino que también asisten en la evaluación de la eficiencia instruccional y los resultados educativos [8].

Al aprovechar técnicas avanzadas de análisis de datos, los educadores y los responsables de políticas pueden obtener una comprensión más clara de la efectividad de los programas educativos inclusivos y tomar decisiones informadas que promuevan la equidad en la educación.

Por lo tanto, abordar los problemas multifacéticos asociados con la educación inclusiva es crucial para el desarrollo continuo y la optimización de las prácticas y sistemas educativos. Por lo tanto, abordar los problemas multifacéticos asociados con la educación inclusiva es crucial para el desarrollo continuo y la optimización de las prácticas y sistemas educativos [9].

Conceptos Clave

Modelado Estadístico en la Educación

El modelado estadístico desempeña un papel fundamental en la comprensión y evaluación de los programas educativos, particularmente en el contexto de la educación inclusiva. Se refiere al proceso de utilizar marcos matemáticos para identificar relaciones dentro de conjuntos de datos relevantes para entornos educativos, como métricas de rendimiento estudiantil e información demográfica [10].

Los modelos estadísticos se construyen para proporcionar información sobre los patrones de comportamiento entre los estudiantes, lo que permite a los educadores y administradores tomar decisiones basadas en datos [11].

Estos modelos a menudo incorporan tanto variables aleatorias como no aleatorias, lo que permite una interpretación matizada de los resultados educativos. También pueden resaltar variaciones en el rendimiento de los estudiantes, lo cual es esencial para adaptar las estrategias educativas a las necesidades de los diversos aprendices [12].

Técnicas de Minería de Datos en la Evaluación Educativa

La Minería de Datos Educativos (EDM) es un campo en auge que se centra en desarrollar métodos para extraer información significativa de grandes conjuntos de datos educativos [13], para lo cual se emplean diversas técnicas como la clasificación, el agrupamiento y la regresión para identificar relaciones y tendencias ocultas entre los estudiantes y los procesos educativos.

Al aprovechar estas técnicas de minería de datos, los educadores pueden mejorar su comprensión de las necesidades de los estudiantes y aumentar la efectividad de los programas. Al aprovechar estas técnicas de minería de datos, los educadores pueden mejorar su comprensión de las necesidades de los estudiantes y aumentar la efectividad de los programas. Además, el EDM facilita el análisis predictivo, que puede prever los resultados de los estudiantes e identificar a aquellos en riesgo, permitiendo intervenciones proactivas.

Integración de Métodos Estadísticos y Minería de Datos

La integración de modelos estadísticos avanzados y técnicas de minería de datos ofrece herramientas poderosas para evaluar programas educativos inclusivos. Por ejemplo, el uso de modelos multinivel puede ayudar a evaluar la efectividad de las intervenciones en diferentes poblaciones estudiantiles, teniendo en cuenta las variaciones en las experiencias de aprendizaje individuales [14].

Este enfoque integral no solo ayuda a identificar prácticas efectivas, sino que también apoya el refinamiento continuo de las estrategias educativas en respuesta a los nuevos conocimientos derivados de los datos [15].

Marcos de Evaluación

Modelos psicométricos en la evaluación

En el contexto de la evaluación de programas educativos inclusivos, los modelos psicométricos desempeñan un papel crucial al permitir la medición del aprendizaje de los estudiantes y la efectividad del programa. Los modelos psicométricos estándar, como el Modelado de Respuesta al Ítem (MRI), proporcionan un marco para comparar a los examinados en diferentes pruebas y estimar los parámetros de los ítems basándose en grupos variados de estudiantes [16].

Estos modelos facilitan el análisis de qué tan bien las evaluaciones educativas reflejan los conocimientos y habilidades de los estudiantes, mejorando así la evaluación de la eficiencia instruccional [17].

Triángulo de Evaluación

El proceso de evaluación está guiado por el "triángulo de evaluación," que consiste en tres elementos interconectados: cognición, observación e interpretación. La cognición se refiere a los constructos o objetivos de aprendizaje que las evaluaciones buscan medir. La observación implica recopilar datos a través de tareas de evaluación, como las calificaciones asignadas por los maestros en diversas tareas [18].

Finalmente, la interpretación se refiere a los métodos estadísticos utilizados para analizar los datos recopilados, a menudo empleando modelos de medición para extraer inferencias significativas sobre el aprendizaje de los

estudiantes y la efectividad del programa. Esta relación triádica asegura que las evaluaciones estén bien coordinadas y produzcan conclusiones válidas sobre el rendimiento estudiantil [19].

Enfoques Bayesianos

Los avances recientes en la metodología estadística, particularmente el modelado bayesiano, han comenzado a influir en las prácticas de evaluación educativa. Estos enfoques permiten interpretaciones más matizadas de los datos de los estudiantes al incorporar índices diagnósticos en los modelos de medición. Por ejemplo, el uso de Modelos de Respuesta al Ítem multidimensionales permite a los educadores ir más allá de las estadísticas resumidas individuales, proporcionando así una visión más rica del rendimiento estudiantil [20].

Los métodos bayesianos también facilitan la integración de diversas fuentes de evidencia, mejorando la robustez general de las evaluaciones en contextos educativos inclusivos [21].

Integración de la Minería de Datos Educativos

Las técnicas de Minería de Datos Educativos (EDM) se utilizan cada vez más para evaluar de manera efectiva los programas de instrucción. EDM proporciona a las partes interesadas—educadores, estudiantes, organizaciones e investigadores—herramientas para analizar grandes conjuntos de datos, lo que conduce a métodos de enseñanza mejorados, experiencias de aprendizaje individualizadas y una asignación optimizada de recursos dentro de las instituciones educativas [22].

Al utilizar técnicas avanzadas de modelado estadístico y minería de datos, los evaluadores pueden obtener información útil a partir de datos educativos complejos, mejorando así la

efectividad de los programas educativos inclusivos [23].

Técnicas Avanzadas de Modelado Estadístico

Las técnicas avanzadas de modelado estadístico desempeñan un papel crucial en la optimización de la evaluación de programas educativos inclusivos. Estos métodos permiten a los investigadores analizar estructuras de datos complejas y relaciones, proporcionando una comprensión más profunda de los resultados educativos [24].

Modelado Multinivel

El modelado multinivel (MLM) es una técnica estadística avanzada que permite a los investigadores particionar la varianza a través de diferentes niveles de análisis [25], como los estudiantes anidados dentro de las escuelas. Esta flexibilidad ayuda a acomodar la naturaleza jerárquica de los datos educativos, mejorando la precisión de los hallazgos. El MLM es particularmente útil para evaluar el impacto de diversos factores en el rendimiento estudiantil mientras se controlan las influencias contextuales, como el estatus socioeconómico (SES).

Métodos Bayesianos

Los métodos bayesianos han surgido como un enfoque poderoso en el modelado estadístico, particularmente para las evaluaciones educativas. Estos métodos permiten la incorporación de conocimientos previos y la actualización continua de las estimaciones a medida que se dispone de nuevos datos. Esta adaptabilidad es beneficiosa para manejar las incertidumbres inherentes a los datos educativos, como las variaciones en el rendimiento de los estudiantes en diferentes contextos [26].

Las redes de inferencia bayesiana, por ejemplo, permiten a los investigadores modelar interrelaciones complejas entre variables, mejorando la interpretabilidad de los datos relacionados con los resultados educativos [27].

Teoría de Respuesta al Ítem

La Teoría de Respuesta al Ítem (IRT) ofrece un marco robusto para comprender cómo funcionan los diferentes ítems de prueba en diversas poblaciones. Permite la comparación del rendimiento de los examinados en diferentes evaluaciones al predecir las propiedades de los ítems basándose en las características de los examinados [28].

Este método es particularmente valioso en entornos educativos inclusivos, ya que puede revelar cómo los estudiantes diversos interactúan con las herramientas de evaluación, facilitando prácticas de evaluación más equitativas [29].

Técnicas de Aprendizaje Automático

Las técnicas de aprendizaje automático, incluyendo el aprendizaje supervisado y no supervisado, proporcionan herramientas innovadoras para analizar datos educativos. Los algoritmos de aprendizaje supervisado, como la regresión y los árboles de decisión, predicen resultados basados en datos históricos, mientras que los métodos no supervisados, como el agrupamiento, identifican patrones dentro de los conjuntos de datos sin resultados preetiquetados [12].

El uso de estos algoritmos en el análisis de políticas permite evaluaciones más matizadas de los programas educativos, lo que permite a los responsables de políticas adaptar las intervenciones basándose en las tendencias y relaciones identificadas [30].

Regresión Cuantil

La regresión cuantílica es otra técnica avanzada que evalúa cómo la relación entre las variables difiere en varios puntos de la distribución del resultado, en lugar de centrarse únicamente en la media [31].

Este enfoque es particularmente relevante en la investigación educativa, ya que ayuda a identificar cómo los diferentes niveles de financiamiento escolar se relacionan con el rendimiento entre diversos grupos demográficos de estudiantes, incluidos aquellos de entornos desfavorecidos. Este enfoque es particularmente relevante en la investigación educativa, ya que ayuda a identificar cómo los diferentes niveles de financiamiento escolar se relacionan con el rendimiento entre diversos grupos demográficos de estudiantes, incluidos aquellos de entornos desfavorecidos [32].

Técnicas de Minería de Datos en Educación

La minería de datos educativos (EDM) es un campo emergente que se centra en el desarrollo de métodos para analizar grandes volúmenes de datos de entornos educativos con el fin de extraer información significativa y mejorar los procesos de toma de decisiones [33].

Se emplean varias técnicas avanzadas de minería de datos en el ámbito de la educación, incluyendo la clasificación, el agrupamiento, la regresión y la minería de reglas de asociación, cada una con propósitos distintos para comprender el comportamiento y el rendimiento académico de los estudiantes [34].

Técnicas de Clasificación

La clasificación es un método de aprendizaje supervisado utilizado para categorizar datos en clases predefinidas. En contextos educativos, los algoritmos de clasificación ayudan a predecir los resultados de los estudiantes basándose en datos

históricos. Los algoritmos comunes incluyen árboles de decisión, máquinas de soporte vectorial (SVM) y redes neuronales. Los árboles de decisión, en particular, son populares debido a su interpretabilidad y capacidad para proporcionar reglas de "si-entonces" que pueden guiar a los educadores en la toma de decisiones informadas sobre las intervenciones para los estudiantes y la asignación de recursos [7].

Por ejemplo, los investigadores han utilizado árboles de decisión para modelar el rendimiento estudiantil, prediciendo eficazmente el éxito académico basándose en varios predictores, como la asistencia y las calificaciones previas. Por ejemplo, los investigadores han utilizado árboles de decisión para modelar el rendimiento estudiantil, prediciendo eficazmente el éxito académico basándose en varios predictores, como la asistencia y las calificaciones previas [35].

Modelado Predictivo

El modelado predictivo integra diversas técnicas de minería de datos para prever el rendimiento futuro de los estudiantes. Al utilizar algoritmos como los k-vecinos más cercanos (kNN) y el análisis de regresión, los educadores pueden construir modelos que anticipen los desafíos académicos y las oportunidades de intervención. Estos modelos predictivos no solo ayudan a identificar a los estudiantes en riesgo, sino que también permiten a las instituciones educativas asignar recursos de manera más efectiva y mejorar el rendimiento académico general [36].

Métodos de Agrupamiento

El agrupamiento es otra técnica fundamental empleada en EDM que agrupa a los estudiantes según características compartidas. Al identificar agrupaciones, los educadores pueden discernir diferentes estilos de aprendizaje y patrones de comportamiento entre los estudiantes, lo que permite el desarrollo de marcos educativos

personalizados. Este método puede ser fundamental para mejorar la educación comunitaria y fomentar entornos de aprendizaje adaptativos [37].

Las técnicas de agrupamiento permiten a los instructores categorizar a los estudiantes, facilitando intervenciones específicas y experiencias de aprendizaje personalizadas [38].

Análisis de Regresión

El análisis de regresión se utiliza a menudo para modelar las relaciones entre diversas variables educativas y el rendimiento estudiantil [39]. Permite a los educadores predecir resultados continuos, como las calificaciones de los exámenes o las tasas de graduación, basándose en una combinación de variables independientes. Esta técnica es beneficiosa para comprender el impacto de diferentes factores en el éxito estudiantil y para tomar decisiones basadas en datos respecto a los ajustes curriculares y la distribución de recursos.

Minería de Reglas de Asociación

La minería de reglas de asociación es un componente crítico del EDM, centrado en descubrir relaciones interesantes entre variables dentro de grandes conjuntos de datos [40].

Esta técnica ayuda a identificar correlaciones entre los comportamientos de los estudiantes, como la actividad en línea, y los resultados académicos, permitiendo a los educadores obtener información sobre estrategias de enseñanza efectivas y el compromiso estudiantil. Por ejemplo, las reglas de asociación pueden revelar patrones en las interacciones de los estudiantes con los recursos en línea y cómo estos comportamientos se correlacionan con sus calificaciones finales, proporcionando información útil para mejorar los resultados educativos [41].

2. Materiales y Métodos

2.1 Modelos Estadísticos

Análisis Exploratorio

El análisis exploratorio se llevó a cabo mediante técnicas avanzadas de minería de datos para identificar patrones y relaciones en los datos. Se aplicaron algoritmos de clustering, como K-means y DBSCAN, para segmentar a los estudiantes en grupos según su rendimiento académico y bienestar emocional.

Además, se utilizó el algoritmo Apriori para el análisis de reglas de asociación, lo que permitió detectar relaciones significativas entre variables como el rendimiento académico, los recursos disponibles y la formación docente. Para simplificar la estructura de los datos y preservar la máxima variabilidad, se empleó el análisis de componentes principales (PCA).

Modelos Jerárquicos Multinivel

Se implementaron modelos lineales mixtos (HLM) para analizar la variabilidad entre estudiantes y escuelas, considerando la estructura anidada de los datos (estudiantes dentro de escuelas). Estos modelos incorporaron efectos aleatorios para capturar las diferencias entre contextos escolares y evaluaron el impacto de factores escolares y de los programas educativos en los resultados académicos y socioemocionales de los estudiantes.

Aprendizaje Automático

Se emplearon dos enfoques de aprendizaje automático para modelar y predecir los resultados educativos:

Random Forest: Este modelo ensamblado se utilizó para predecir el rendimiento académico y el bienestar emocional, identificando las

características más relevantes mediante la evaluación de la importancia de las variables.

Gradient Boosting Machine: Este modelo de boosting optimizó las predicciones mediante un proceso iterativo que minimizó el error de predicción, destacando los factores críticos a través del análisis de importancia de características.

Ambos modelos se evaluaron utilizando métricas de rendimiento como el coeficiente de determinación ((R^2)), el error absoluto medio (MAE) y la raíz del error cuadrático medio (RMSE).

Validación y Métricas

Para garantizar la robustez de los modelos, los datos se dividieron en conjuntos de entrenamiento (80%) y prueba (20%). Además, se aplicó validación cruzada para evaluar la generalización de los modelos. Las métricas de evaluación incluyeron (R^2), MAE y RMSE, lo que permitió medir con precisión el desempeño predictivo y la capacidad explicativa de los modelos.

Integración de Métodos

Este enfoque metodológico combina técnicas estadísticas tradicionales con herramientas modernas de aprendizaje automático, ofreciendo una perspectiva integral sobre los factores que influyen en el éxito educativo y el desarrollo socioemocional de los estudiantes. Los resultados obtenidos proporcionan una base sólida para la toma de decisiones basada en datos, orientada a optimizar los programas educativos inclusivos.

2.2 Datos utilizados

Descripción de la Base de Datos Simulada

La base de datos educativa simulada integra múltiples dimensiones del contexto educativo, estructurada en cuatro componentes principales:

Características del Estudiante: Identificador único del estudiante, Edad (6-18 años), género (M/F), nivel socioeconómico (Bajo/Medio/Alto), tipo de NEE (Ninguna/Física/Intelectual/TEA/Sensorial), rendimiento académico previo, nivel de apoyo familiar y acceso a servicios de salud.

Características del Contexto Escolar: Identificador único de la escuela, tamaño de la institución (Pequeña/Mediana/Grande), ubicación (Urbana/Rural), nivel de formación docente y Cultura inclusiva de la institución.

Características del Programa Educativo: tipo de intervención (Adaptaciones curriculares/Apoyos personalizados/Formación docente), duración del programa (3-24 meses), intensidad del programa (1-20 horas semanales) y recursos disponibles.

Resultados Educativos: rendimiento en matemáticas (1-10), rendimiento en lenguaje (1-10), bienestar emocional (1-5) y participación en clase (1-5).

La base de datos simula estuvo conformado por conjunto de 10,000 registros, incorporando distribuciones probabilísticas realistas basadas en literatura educativa actual. Las variables categóricas se codifican mediante one-hot encoding para su uso en los modelos estadísticos y de aprendizaje automático. La estructura jerárquica de los datos refleja la realidad educativa, donde los estudiantes están anidados dentro de instituciones específicas.

Esta base de datos simulada proporciona una representación detallada y estructurada del contexto educativo, permitiendo el análisis de las interacciones entre factores individuales,

contextuales e institucionales en el éxito educativo y el desarrollo socioemocional.

3. Resultados

Rendimiento de los Modelos Predictivos

Los modelos de aprendizaje automático empleados en este estudio muestran un nivel moderado de precisión en la predicción de los resultados académicos. El modelo Random Forest alcanzó un coeficiente de determinación (R^2) de 0.13, con un error absoluto medio (MAE) de 0.94 y una raíz del error cuadrático medio (RMSE) de 1.16. Por su parte, el modelo Gradient Boosting mostró un desempeño ligeramente superior, con un R^2 de 0.16, un MAE de 0.92 y un RMSE de 1.14. Este mejoramiento sugiere que Gradient Boosting tiene una mayor capacidad para identificar patrones complejos dentro del conjunto de datos (Figura 1).

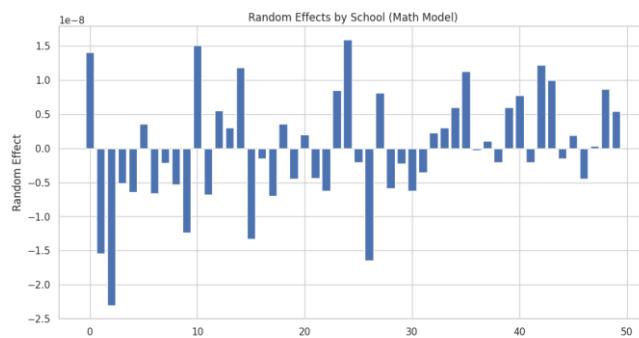


Fig 1. Efectos aleatorizados de modelos.

Factores Críticos Identificados por Random Forest

Mediante el análisis de importancia de variables realizado con el modelo Random Forest (Figura 2), se identificaron tres niveles de influencia sobre los resultados académicos.

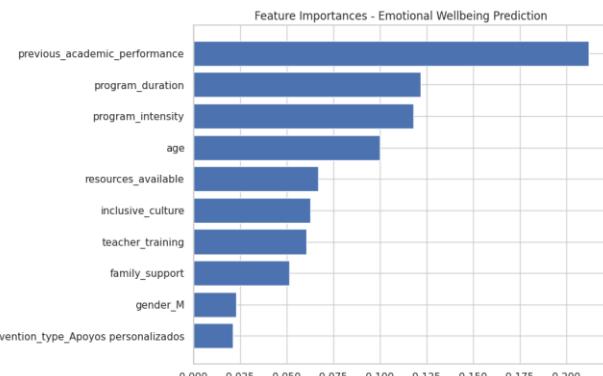


Fig 2. análisis de importancia de variables realizado con el modelo Random Forest.

Primer Nivel de Importancia: Rendimiento académico previo (19.3%), duración del programa (11.4%), intensidad del programa (10.9%) y recursos disponibles (10.9%). Segundo Nivel de Importancia: edad del estudiante (9.0%), cultura inclusiva (5.7%), apoyo familiar (5.7%) y formación docente (5.5%) y factores adicionales: ausencia de Necesidades Educativas Especiales (NEE) (5.2%), género masculino (2.0%), y ubicación urbana (2.0%).

Factores Críticos Identificados por Gradient Boosting

El modelo Gradient Boosting arrojó un perfil distinto en cuanto a la relevancia relativa de los factores, destacando claramente algunos elementos clave (Figura 3).

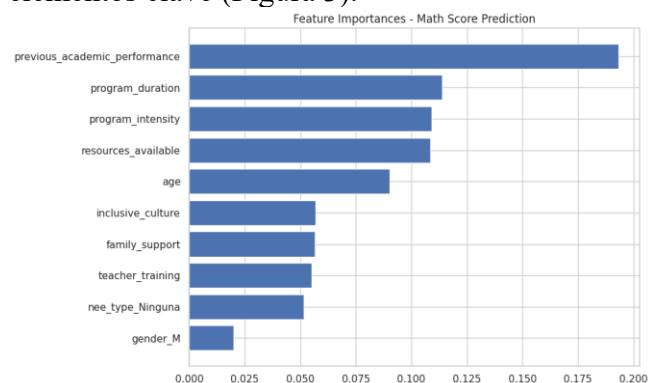


Fig 3. análisis de importancia de variables realizado con el modelo gradient boosting.

Primer Nivel de Importancia: recursos disponibles (52.5%), ausencia de NEE (24.8%) y rendimiento académico previo (11.7%). segundo nivel de importancia: intensidad del programa (2.1%), duración del programa (1.6%), edad del estudiante (1.3%) y Apoyo familiar (1.3%), así como factores Contextuales Menos Relevantes: cultura inclusiva (0.9%), formación docente (0.8%) y tamaño de la escuela (0.3–0.4%).

Los resultados reflejan diferencias significativas en la forma en que ambos modelos ponderan los distintos factores. Gradient Boosting resalta especialmente la disponibilidad de recursos y la ausencia de necesidades educativas especiales como los aspectos más influyentes, lo cual subraya la importancia de contar con infraestructura adecuada y apoyo especializado para maximizar el éxito académico.

Por otro lado, Random Forest presenta una distribución más equilibrada (Figura 4) de las variables relevantes, dando particular énfasis al rendimiento previo del estudiante y a las características estructurales del programa educativo, tales como su duración e intensidad.

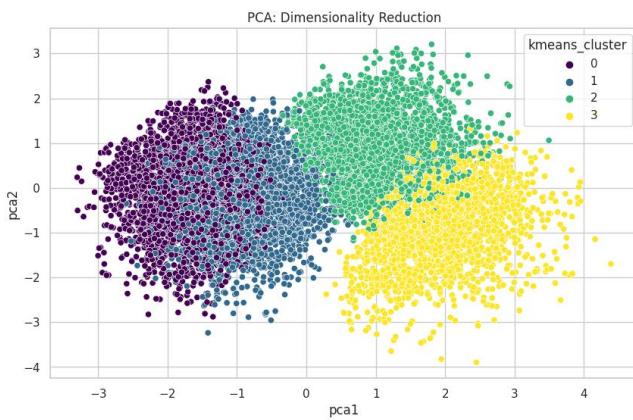


Fig 4. Divergencia entre modelos analizada mediante PCA.

La divergencia observada entre los dos modelos (Figura 5) sugiere que existen diferentes dinámicas interactivas entre las variables, lo que amerita un análisis más profundo para

comprender mejor estas relaciones y su impacto real en el desempeño escolar.

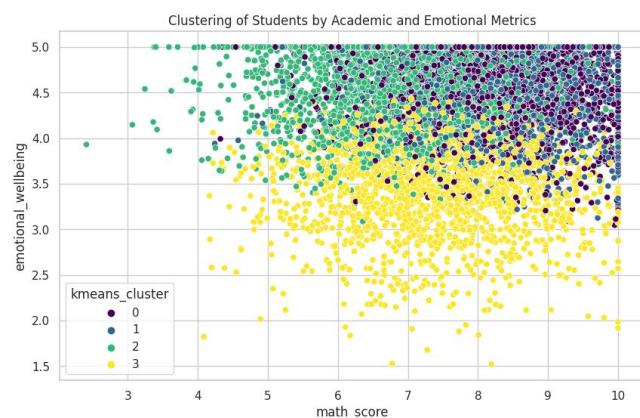


Fig 5. Divergencia entre modelos analizada mediante análisis de clúster.

4. Discusión

Los modelos predictivos empleados en este estudio muestran un nivel moderado de precisión al predecir resultados académicos, lo cual es coherente con hallazgos previos en el ámbito del educational data mining (EDM), donde la complejidad y variabilidad inherente a los datos educativos limita la capacidad predictiva absoluta [1][26]. El modelo Random Forest alcanzó un coeficiente de determinación (R^2) de 0.13, con un error absoluto medio (MAE) de 0.94 y una raíz del error cuadrático medio (RMSE) de 1.16. Por su parte, Gradient Boosting mostró un desempeño ligeramente superior, con un R^2 de 0.16, MAE de 0.92 y RMSE de 1.14. Esta mejora sugiere que Gradient Boosting tiene mayor capacidad para identificar patrones complejos dentro del conjunto de datos, posiblemente debido a su enfoque iterativo y ponderado en la corrección de errores [42].

El análisis de importancia de variables realizado mediante Random Forest reveló tres niveles de influencia sobre los resultados académicos. En primer lugar, destacaron el rendimiento académico previo (19.3%), la duración del

programa (11.4%), la intensidad del programa (10.9%) y los recursos disponibles (10.9%). Estos hallazgos coinciden con investigaciones previas que han señalado la relevancia del historial académico como predictor robusto del éxito escolar [43], así como la importancia de factores estructurales del programa educativo, tales como su diseño curricular y carga horaria [44].

En segundo nivel de importancia se encontraron la edad del estudiante (9.0%), seguida por la cultura inclusiva (5.7%), el apoyo familiar (5.7%) y la formación docente (5.5%). Estos factores reflejan aspectos contextuales que influyen indirectamente en el desempeño académico, y son consistentes con estudios que resaltan el impacto positivo de ambientes escolares inclusivos y del acompañamiento familiar en el desarrollo educativo [45]. Finalmente, entre los factores adicionales se identificaron la ausencia de necesidades educativas especiales (NEE) (5.2%), seguido por el género masculino (2.0%) y la ubicación urbana (2.0%).

Por otro lado, Gradient Boosting arrojó un perfil distinto en cuanto a la relevancia relativa de los factores. En primer lugar, destacó claramente la disponibilidad de recursos (52.5%), seguida por la ausencia de NEE (24.8%) y el rendimiento académico previo (11.7%). Este resultado subraya la importancia crítica de contar con infraestructura adecuada y apoyo especializado para maximizar el éxito académico, especialmente en contextos inclusivos donde las barreras para el aprendizaje pueden ser múltiples y heterogéneas [3][15][16]. Los factores en segundo nivel de importancia incluyeron la intensidad del programa (2.1%), la duración del programa (1.6%), la edad del estudiante (1.3%) y el apoyo familiar (1.3%). Por último, entre los factores contextuales menos relevantes se encontraron la cultura inclusiva (0.9%), la formación docente (0.8%) y el tamaño de la escuela (0.3–0.4%). La divergencia observada

entre ambos modelos sugiere que existen diferentes dinámicas interactivas entre las variables, lo cual merece un análisis más profundo para comprender mejor estas relaciones y su impacto real en el desempeño escolar [46].

Por su parte Random Forest ofrece una distribución más equilibrada de las variables relevantes, Gradient Boosting parece enfocarse más en aquellos factores que tienen una relación directa y fuerte con los resultados académicos, particularmente los relacionados con los recursos institucionales y la presencia o ausencia de necesidades educativas especiales.

Estos hallazgos también reflejan una tensión recurrente en el campo de la minería de datos educativos: cómo balancear la precisión predictiva con la interpretabilidad contextual, "la toma de decisiones basada en datos no solo requiere modelos estadísticamente sólidos, sino también una comprensión profunda de los contextos educativos en los que operan [47]. Esto implica que, aunque Gradient Boosting pueda ofrecer mejores métricas técnicas, Random Forest proporciona una visión más amplia de los factores que podrían ser útiles para diseñar intervenciones educativas multicausales.

5. Conclusiones

Los resultados del presente análisis reflejan la complejidad de los factores que inciden en el rendimiento académico y destacan la utilidad de los modelos de aprendizaje automático —en particular Random Forest y Gradient Boosting— como herramientas para identificar patrones predictivos y priorizar variables clave en entornos educativos. Ambos modelos mostraron un desempeño moderado en términos predictivos, con Gradient Boosting presentando una leve superioridad, lo cual sugiere su mayor capacidad para capturar relaciones no lineales y complejas entre las variables.

Desde la perspectiva de la importancia relativa de los factores, se observó una divergencia clara entre ambos modelos. Mientras Random Forest ofreció una distribución más equilibrada que incluye aspectos cognitivos, contextuales y estructurales, Gradient Boosting resaltó predominantemente la disponibilidad de recursos institucionales y la ausencia de necesidades educativas especiales como los determinantes más influyentes. Este hallazgo pone de relieve la relevancia de garantizar condiciones materiales adecuadas y apoyo especializado en el marco de una educación inclusiva efectiva.

La diferencia metodológica entre ambos enfoques subraya la importancia de considerar múltiples perspectivas analíticas al momento de diseñar políticas educativas o intervenciones pedagógicas. Si bien Gradient Boosting puede ser preferible desde un punto de vista técnico por su mayor ajuste estadístico, Random Forest proporciona una visión más holística que permite comprender mejor los distintos niveles de influencia sobre el éxito académico.

Referencias:

- [1] Rao, Y. S. N., & Chen, C. J. (2024). Bibliometric insights into data mining in education research: A decade in review. *Contemporary Educational Technology*, 16(2), ep502. <https://doi.org/10.30935/cedtech/14333>
- [2] Cerezo, R., Lara, J. A., Azevedo, R., & Romero, C. (2024). Reviewing the differences between learning analytics and educational data mining: Towards educational data science. *Computers in Human Behavior*, 154, 108155. <https://doi.org/10.1016/j.chb.2024.108155>
- [3] Yin, S., Li, H., Teng, L., Laghari, A. A., & Estrela, V. V. (2024). Attribute-based multiparty searchable encryption model for privacy protection of text data. *Multimedia Tools and Applications*, 83(15), 45881-45902. <http://dx.doi.org/10.1007/s11042-023-16818-4>
- [4] Arriola Navarrete, Ó. (2023). La Brecha Digital En La Revolución Industrial 4.0 Oportunidad Y Reto Para Las Bibliotecas. *Revista Interamericana de Bibliotecología*, 46(3). <http://dx.doi.org/10.17533/udea.rib.v46n3e345719>
- [5] Batool, S., Rashid, J., Nisar, M. W., Kim, J., Kwon, H. Y., & Hussain, A. (2023). Educational data mining to predict students' academic performance: A survey study. *Education and Information Technologies*, 28(1), 905-971. <http://dx.doi.org/10.1007/s10639-022-11152-y>
- [6] Mirakhur, Z., Chen, C., & Schwarz, S. (2022). A case study in teaching inclusive teaching. To Improve the Academy: A Journal of Educational Development, 41(2). <https://doi.org/10.3998/tia.386>
- [7] Yuang, Y., Yun, Y., An, R., Cui, J., Dai, H., & Shang, X. (2021). Educational data mining techniques for student performance prediction: Method review and comparison analysis. *Frontiers in Psychology*, 12, 698490. <https://doi.org/10.3389/fpsyg.2021.698490>
- [8] Ampadu, Y. B. (2023). Handling big data in education: A review of educational data mining techniques for specific educational problems. *AI, Computer Science and Robotics Technology*, 2. <https://doi.org/10.5772/acrt.17>
- [9] Wray, E., Sharma, U., & Subban, P. (2022). Factors influencing teacher self-efficacy for inclusive education: A systematic literature review. *Teaching and Teacher Education*, 117, 103800. <https://doi.org/10.1016/j.tate.2022.103800>
- [10] Gao, P., Li, J., & Liu, S. (2021). An introduction to key technology in artificial intelligence and big data driven e-learning and e-

- education. *Mobile Networks and Applications*, 26(5), 2123-2126. <http://dx.doi.org/10.1007/s11036-021-01777-7>
- [11] Zapata-Cardona, L., & Martínez-Castro, C. A. (2023). Statistical modeling in teacher education. *Mathematical Thinking and Learning*, 25(1), 64-78. <http://dx.doi.org/10.1080/10986065.2021.1922859>
- [12] Sher, J. A., & Hart, S. A. (2022). Methodological decisions and their impacts on the perceived relations between school funding and educational achievement. *Frontiers in Education*, 7, 1043471. <https://doi.org/10.3389/feduc.2022.1043471>
- [13] Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1), 11. <https://doi.org/10.1186/s40561-022-00192-z>
- [14] Baek, C., & Doleck, T. (2023). Educational data mining versus learning analytics: A review of publications from 2015 to 2019. *Interactive Learning Environments*, 31(6), 3828-3850. <http://dx.doi.org/10.1080/10494820.2021.1943689>
- [15] Alwarthan, S. A., Aslam, N., & Khan, I. U. (2022). Predicting student academic performance at higher education using data mining: a systematic review. *Applied Computational Intelligence and Soft Computing*, 2022(1), 8924028. <http://dx.doi.org/10.1155/2022/8924028>
- [16] Chen, Y., Schönlieb, C. B., Liò, P., Leiner, T., Dragotti, P. L., Wang, G., ... & Yang, G. (2022). AI-based reconstruction for fast MRI—A systematic review and meta-analysis. *Proceedings of the IEEE*, 110(2), 224-245. <http://dx.doi.org/10.1109/JPROC.2022.3141367>
- [17] Cao, C., Cui, Z. X., Wang, Y., Liu, S., Chen, T., Zheng, H., ... & Zhu, Y. (2024). High-frequency space diffusion model for accelerated MRI. *IEEE Transactions on Medical Imaging*, 43(5), 1853-1865. <http://dx.doi.org/10.48550/arXiv.2208.05481>
- [18] Yu, H. (2024). Research on the urban green transportation development level evaluation based on the triangular pythagorean fuzzy multiple attribute decision making. *Journal of Intelligent & Fuzzy Systems*, 46(3), 6279-6297. <http://dx.doi.org/10.3233/JIFS-232579>
- [19] Zhan, J., Deng, J., Xu, Z., & Martínez, L. (2023). A three-way decision methodology with regret theory via triangular fuzzy numbers in incomplete multiscale decision information systems. *IEEE Transactions on Fuzzy Systems*, 31(8), 2773-2787. <https://ieeexplore.ieee.org/document/10018542>
- [20] Jyothi, P., & Singh, A. R. (2023). Deep learning models and traditional automated techniques for brain tumor segmentation in MRI: a review. *Artificial intelligence review*, 56(4), 2923-2969. <http://dx.doi.org/10.1007/s10462-022-10245-x>
- [21] Lye, A., Cicirello, A., & Patelli, E. (2021). Sampling methods for solving Bayesian model updating problems: A tutorial. *Mechanical Systems and Signal Processing*, 159, 107760. <http://dx.doi.org/10.1016/j.ymssp.2021.107760>
- [22] Okewu, E., Adewole, P., Misra, S., Maskeliunas, R., & Damasevicius, R. (2021). Artificial neural networks for educational data mining in higher education: A systematic literature review. *Applied Artificial Intelligence*, 35(13), 983-1021. <http://dx.doi.org/10.1080/08839514.2021.1922847>

- [23] Chango, W., Lara, J. A., Cerezo, R., & Romero, C. (2022). A review on data fusion in multimodal learning analytics and educational data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(4), e1458. <http://dx.doi.org/10.1002/widm.1458>
- [24] Lemay, D. J., Baek, C., & Doleck, T. (2021). Comparison of learning analytics and educational data mining: A topic modeling approach. *Computers and Education: Artificial Intelligence*, 2, 100016. <https://doi.org/10.1016/j.caai.2021.100016>
- [25] Lin, H. M., Wu, J. Y., Liang, J. C., Lee, Y. H., Huang, P. C., Kwok, O. M., & Tsai, C. C. (2023). A review of using multilevel modeling in e-learning research. *Computers & Education*, 198, 104762. <http://dx.doi.org/10.1016/j.compedu.2023.104762>
- [26] Gallardo-Gomez, D., del Pozo-Cruz, J., Noetel, M., Alvarez-Barbosa, F., Alfonso-Rosa, R. M., & del Pozo Cruz, B. (2022). Optimal dose and type of exercise to improve cognitive function in older adults: A systematic review and bayesian model-based network meta-analysis of RCTs. *Ageing research reviews*, 76, 101591. <https://doi.org/10.1016/j.arr.2022.101591>
- [27] Moradi, M. (2023). A Bayesian Model and Bayesian Classification on the Data Obtained from Children's Educational Activity in the IoT Environment. *Transactions on Machine Intelligence*, 6(3), 126-136. <https://doi.org/10.47176/TMI.2023.126>
- [28] Hermita, N., Putra, Z. H., Alim, J. A., Wijaya, T. T., Anggoro, S., & Diniya, D. (2022). Elementary Teachers' Perceptions on Genially Learning Media Using Item Response Theory (IRT). *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 1-20. <http://dx.doi.org/10.23917/ijolae.v4i1.14757>
- [29] Spann, D. J., Cicero, D. C., Straub, K. T., Pellegrini, A. M., & Kerns, J. G. (2024). Examining measures of schizotypy for gender and racial bias using item response theory and differential item functioning. *Schizophrenia Research*, 272, 120-127. <http://dx.doi.org/10.1016/j.schres.2024.08.015>
- [30] Chen, W., Shen, Z., Pan, Y., Tan, K., & Wang, C. (2024). Applying machine learning algorithm to optimize personalized education recommendation system. *Journal of Theory and Practice of Engineering Science*, 4(01), 101-108. [https://doi.org/10.53469/jtpes.2024.04\(01\).14](https://doi.org/10.53469/jtpes.2024.04(01).14)
- [31] Peng, L. (2021). Quantile regression for survival data. *Annual review of statistics and its application*, 8(1), 413-437. <https://doi.org/10.1146/annurev-statistics-042720-020233>
- [32] Liu, F., Umair, M., & Gao, J. (2023). Assessing oil price volatility co-movement with stock market volatility through quantile regression approach. *Resources Policy*, 81, 103375. <http://dx.doi.org/10.1016/j.resourpol.2023.103375>
- [33] Roslan, M. B., & Chen, C. (2022). Educational data mining for student performance prediction: A systematic literature review (2015-2021). *International Journal of Emerging Technologies in Learning (iJET)*, 17(5), 147-179. <http://dx.doi.org/10.3991/ijet.v17i05.27685>
- [34] Ijaz, S., Safdar, T., & Sanaullah, M. (2020). Educational data mining: A review and analysis of student's academic performance. In I. S. Bajwa, T. Sibalija, & D. N. A. Jawawi (Eds.), *Intelligent Technologies and Applications* (pp. 510–523). Springer. https://doi.org/10.1007/978-981-15-5232-8_44

- [35] Yu, J. (2021). Academic Performance Prediction Method of Online Education using Random Forest Algorithm and Artificial Intelligence Methods. *International Journal of Emerging Technologies in Learning*, 15(5). <http://dx.doi.org/10.3991/ijet.v16i05.20297>
- [36] Kang, S. (2021). K-nearest neighbor learning with graph neural networks. *Mathematics*, 9(8), 830. <http://dx.doi.org/10.3390/math9080830>
- [37] Palacios, C. A., Reyes-Suárez, J. A., Bearzotti, L. A., Leiva, V., & Marchant, C. (2021). Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in Chile. *Entropy*, 23(4), 485. <http://dx.doi.org/10.3390/e23040485>
- [38] Shu, X., & Ye, Y. (2023). Knowledge Discovery: Methods from data mining and machine learning. *Social Science Research*, 110, 102817. <https://doi.org/10.1016/j.ssresearch.2022.102817>
- [39] Gupta, M., & Kaul, S. (2024). Ai in inclusive education: A systematic review of opportunities and challenges in the indian context. *MIER Journal of Educational Studies Trends and Practices*, 429–461. <https://doi.org/10.52634/mier/2024/v14/i2/2702>
- [40] Papadogiannis, I., Wallace, M., & Karountzou, G. (2024). Educational data mining: A foundational overview. *Encyclopedia*, 4(4), 1644–1664. <https://doi.org/10.3390/encyclopedia4040108>
- [41] Hikmawati, E., Maulidevi, N. U., & Surendro, K. (2021). Minimum threshold determination method based on dataset characteristics in association rule mining. *Journal of Big Data*, 8, 1-17. <http://dx.doi.org/10.21203/rs.3.rs-728509/v1>
- [42] Ali, Z. A., Abduljabbar, Z. H., Tahir, H. A., Sallow, A. B., & Almufti, S. M. (2023). eXtreme gradient boosting algorithm with machine learning: A review. *Academic Journal of Nawroz University*, 12(2), 320-334. <http://dx.doi.org/10.25007/ajnu.v12n2a1612>
- [43] Staneviciene, E., Gudoniene, D., Punys, V., & Kukstys, A. (2024). A case study on the data mining-based prediction of students' performance for effective and sustainable e-learning. *Sustainability*, 16(23), 10442. <https://doi.org/10.3390/su162310442>
- [44] Manjarres, A. V., Sandoval, L. G. M., & Suárez, M. J. S. (2018). Data mining techniques applied in educational environments: Literature Review. *Digital Education Review*, 235–266. <https://doi.org/10.1344/der.2018.33.235-266>
- [45] Joaquina De Borba, N., Adriana Santin Ferreira, V., Palhares Dos Santos, T., & Carvalho, S. (2024). Inclusive education: Strategies and impact on contemporary society. *Revista Gênero e Interdisciplinaridade*, 5(03), 182–191. <https://doi.org/10.51249/gei.v5i03.2073>
- [46] Jiao, P., Ouyang, F., Zhang, Q., & Alavi, A. H. (2022). Artificial intelligence-enabled prediction model of student academic performance in online engineering education. *Artificial Intelligence Review*, 55(8), 6321-6344. <http://dx.doi.org/10.1007/s10462-022-10155-y>
- [47] Jiang, Y., Huang, Y., & Li, Q. (2024). Data-driven instructional decision-making: Framework, instrument, and process. *World Journal of Educational Research*, 11(3), p72. <https://doi.org/10.22158/wjer.v11n3p72>

Contribución de los Autores Individuales en la Elaboración de un Artículo Científico (Política de Ghostwriting)

Todos los autores participaron equitativamente del desarrollo del artículo.

Fuentes de Financiamiento para la Investigación Presentada en el Artículo Científico o para el Artículo Científico en sí

No se recibió financiación para la realización de este estudio.

Conflicto de Intereses

Los autores declaran no tener ningún conflicto de interés relevante con el contenido de este artículo.

Licencia de Atribución de Creative Commons 4.0 (Atribución 4.0 Internacional, CC BY 4.0)

Este artículo se publica bajo los términos de la Licencia de Atribución de Creative Commons 4.0

<https://creativecommons.org/licenses/by/4.0/deed.es>