

Evaluación predictiva de las habilidades en razonamiento cuantitativo en ingeniería

Predictive Evaluation of Quantitative Reasoning Skills in Engineering

Artículo de investigación | Research Article | Artigo de investigação

Fecha de recepción: 4 de octubre de 2024

Fecha de aceptación: 23 de mayo de 2025

Fecha de disponibilidad en línea: septiembre de 2025

doi: 10.11144/Javeriana.m18.ehrc

ENRIQUE DE LA HOZ-DOMINGUEZ ✉

enriquedelahoz@unimagdalena.edu.co

UNIVERSIDAD DEL MAGDALENA, COLOMBIA

 ORCID: <https://orcid.org/0000-0003-2531-6389>

ROHEMI ZULUAGA-ORTIZ

rohemi.zuluaga@unisinu.edu.co

UNIVERSIDAD DEL SINÚ, COLOMBIA

 ORCID: <https://orcid.org/0000-0002-7245-8761>

CARLOS GARCÍA-YERENA

cgarciaey@unimagdalena.edu.co

UNIVERSIDAD DEL MAGDALENA, COLOMBIA

 ORCID: <https://orcid.org/0000-0002-9973-552X>

Para citar este artículo | To cite this article | Para citar este artigo

Delahoz-Dominguez, E., Zuluaga-Ortiz, R. & García-Yerena, C. (2025). Evaluación predictiva de las habilidades en razonamiento cuantitativo en ingeniería. *magis, Revista Internacional de Investigación en Educación*, 18, 1-16. <https://doi.org/10.11144/Javeriana.m18.ehrc>



Resumen

Esta investigación presenta un modelo para analizar y predecir el rendimiento en habilidades de razonamiento cuantitativo en estudiantes de ingeniería en Colombia. La población estudiada incluyó a 12 411 estudiantes de ingeniería para el año 2020. Se utilizaron como variables de entrada las competencias en matemáticas, ciencias, inglés, lectura y sociales obtenidas en la prueba estandarizada SABER 11, mientras que la variable de respuesta fue el desempeño en razonamiento cuantitativo de la prueba SABER PRO. Se realizó un análisis descriptivo considerando las variables de género, régimen del colegio y situación laboral de los estudiantes. Posteriormente, se implementó un modelo de *random forest*, identificando que las competencias en matemáticas y biología son las de mayor impacto parcial en la predicción del desempeño en razonamiento cuantitativo. El modelo predictivo alcanzó un RMSE de 10,95 y un R^2 de 69 %, demostrando su capacidad para pronosticar de manera efectiva el rendimiento en esta competencia clave.

Palabras clave

Análisis de datos, rendimiento escolar, evaluación del estudiante, valor agregado

Abstract

This research presents a model to analyze and predict the performance in quantitative reasoning skills among engineering students in Colombia. The study population included 12 411 engineering students for the year 2020. The input variables used were the competencies in mathematics, science, English, reading, and social studies obtained in the standardized test SABER 11, while the response variable was the performance in quantitative reasoning from the SABER PRO. A descriptive analysis was conducted, considering the variables of gender, school system, and student employment status. Subsequently, a random forest model was implemented, identifying that competencies in mathematics and biology have the greatest partial impact on predicting performance in quantitative reasoning. The predictive model achieved an RMSE of 10.95 and an R^2 of 69 %, demonstrating its ability to forecast performance in this key competency effectively.

Keywords

Data analysis, academic performance, student evaluation, added value

Descripción del artículo | Article description | Descrição do artigo

Este artículo de investigación se enmarca dentro del área de las analíticas del aprendizaje y se centra en el desarrollo y validación de un modelo predictivo para estimar el rendimiento en la competencia de razonamiento cuantitativo en estudiantes de ingeniería en Colombia, utilizando datos de las pruebas estandarizadas SABER 11 y SABER PRO. La investigación emplea técnicas de aprendizaje automático, específicamente un modelo *random forest*, para analizar la relación entre diversas competencias académicas y el desempeño en razonamiento cuantitativo. El trabajo no reporta información específica sobre financiación, pero involucra la colaboración entre la Universidad del Magdalena y la Universidad del Sinú.

Introducción

El presente artículo está enmarcado dentro del área de investigación alrededor de las analíticas del aprendizaje, los cuales promueven el uso de los datos generados en los diferentes niveles del proceso educativo con el fin de implementar herramientas que soporten la toma de decisiones objetivas para los diferentes grupos de interés del proceso (Gasevic *et al.*, 2014): estudiantes, gestores educativos, profesores, entidades gubernamentales, agencias de acreditación, entre otras. Consecuentemente, el objetivo del trabajo es analizar y predecir los resultados en la competencia de razonamiento cuantitativo del examen de Estado para la formación en ingeniería en el año 2020 en Colombia, sobre la base de los correspondientes resultados en el examen de Estado para la educación secundaria en el año 2016 (ver tabla 1).

El ejercicio de alinear el resultado del examen nacional universitario (SABER PRO) con el examen de la escuela secundaria hace posible establecer relaciones compatibles de evolución y la determinación del valor agregado que una universidad genera en la formación de los profesionales. En ese sentido, contar con información objetiva sobre las condiciones que podrían impactar positiva o negativamente en la formación académica de un estudiante constituye una previsión estratégica para determinar la planeación y asignación de los recursos educativos en función de maximizar la evolución académica de los estudiantes y consecuentemente mejorar la percepción del servicio educativo ofrecido por parte de la universidad (Delahoz-Dominguez *et al.*, 2020a).

La necesidad del estudio tiene dos fundamentos. En primer lugar, está la perspectiva del estudiante, el cual al terminar la escuela secundaria se ve abocado a una de las decisiones fundamentales para su vida: ¿qué estudiar en la universidad? Diferentes investigaciones han evidenciado la ausencia de herramientas objetivas para la toma de esta importante decisión. Por ejemplo, Kamal *et al.* (2024) muestran que los alumnos escogen la carrera a estudiar

basados en factores subjetivos como la cercanía, el costo o la presión de su grupo familiar o de su círculo social. Además, otros autores (Columbu *et al.*, 2021; Gwelo, 2019) han evidenciado que una discrepancia entre las expectativas de los estudiantes y su percepción del servicio educativo son factores críticos para determinar el abandono estudiantil universitario.

En segundo lugar, está la perspectiva de los gestores educativos a nivel universitario y gubernamental. Para los primeros, la creación de nuevas universidades y la educación en línea han generado un escenario altamente competitivo donde la innovación y la oferta de servicios diferenciadores resultan fundamentales para garantizar la estabilidad y el futuro de las instituciones (Hart & Rodgers, 2024). Para el sector gubernamental, es primordial estimar la evolución de los estudiantes para poder identificar instituciones de excelencia sobre las cuales se puedan replicar buenas prácticas y amplificar el impacto de los recursos invertidos en educación (Visbal-Cadavid *et al.*, 2017).

Paralelamente, diferentes autores han evidenciado cómo en los estudios universitarios de ingeniería las asignaturas con un alto componente cuantitativo se convierten en causantes de deserción estudiantil (Berens *et al.*, 2019; Eichler & Gradwohl, 2021; Heidrich *et al.*, 2018). Por lo tanto, es pertinente proyectar el desempeño futuro en esta área para generar planes de acción que permitan recomendar, intervenir y mejorar el proceso de relacionamiento de los estudiantes con la universidad.

Tradicionalmente, los resultados de las pruebas estatales de evaluación de la calidad de la educación han sido utilizados como herramienta para caracterizar los colegios, universidades y programas académicos, además de servir como un estimador de los conocimientos adquiridos por los estudiantes durante su paso por una institución (Aparicio *et al.*, 2019). Otro aspecto de utilidad de las pruebas estandarizadas nacionales es que sirven como criterio para el otorgamiento de becas a los estudiantes (McLaughlin *et al.*, 2023). Específicamente en el caso colombiano, en el periodo de 2015-2020 se han desarrollado dos grandes planes de becas para acceder a la educación universitaria, el programa Ser Pilo Paga y Generación E, los cuales han beneficiado a más de cuarenta mil estudiantes con una inversión por parte del Estado de dos billones de pesos colombianos por año (Londoño-Vélez *et al.*, 2020). Sin embargo, los beneficios económicos otorgados por las becas mencionadas no son libres de contraprestación, de modo que el estudiante becado está obligado a culminar los estudios profesionales que ha escogido en una universidad pública o privada en un periodo de tiempo establecido, y en el caso de que la persona abandone los estudios, tendrá que afrontar una deuda proporcional a los desembolsos que se le han dado como parte del crédito-beca. En ese sentido, esta investigación intenta aprovechar las estructuras actuales de

evaluación con las que cuenta el Estado colombiano para generar información objetiva que pueda servir como orientación y guía a los gestores educativos y a los estudiantes al momento de tomar decisiones basadas en datos.

Así, la premisa fundamental del estudio está soportada por las preguntas de investigación: ¿Es posible predecir el resultado de la competencia de razonamiento cuantitativo de los exámenes universitarios basándose en los resultados de los exámenes nacionales estandarizados de la educación secundaria para estudiantes de ingeniería? ¿Cuáles de las competencias académicas evaluadas en el examen estatal de la educación secundaria tienen mayor impacto en el desempeño de dicha competencia del examen nacional universitario para estudiantes de ingeniería?

La hipótesis global del estudio es que a través de métodos de aprendizaje automático es posible determinar las relaciones de asociación entre las dimensiones académicas de un examen estandarizado con el desempeño en la competencia de razonamiento cuantitativo. Las hipótesis específicas son: 1) la competencia de matemáticas evaluada en el examen de la educación secundaria tiene un impacto positivo en el resultado de la competencia de razonamiento cuantitativo del examen universitario; 2) la competencia de biología tiene un impacto positivo en el desempeño de la competencia de razonamiento cuantitativo; y 3) el desempeño de la métrica R^2 del modelo predictivo es superior al 60 %.

Tabla 1
Estructura de los módulos del examen SABER 11

Modulo	Competencias	Estándar de evaluación
Matemáticas	Interpretar y representar	Utiliza una variedad de recursos para expresar datos cuantitativos o matemáticos, que van desde descripciones textuales y tablas hasta gráficos, diagramas y esquemas.
	Formular y ejecutar	Analiza la precisión y la eficiencia de los enfoques y métodos matemáticos utilizados para resolver problemas.
	Argumentar	Examina críticamente los métodos y enfoques matemáticos utilizados en la resolución de problemas, confirmando su validez.
Sociales y ciudadanas	Desarrollar pensamiento social	Interpreta las estructuras conceptuales analizando sus características y ámbitos de aplicación.
		Identifica los componentes de ubicación y cronología de los fenómenos sociales, los conflictos y las normas colectivas.
	Interpretar y analizar perspectivas	Se sitúa en un marco histórico y evalúa críticamente el uso de referencias y discursos.
		Interpreta las posturas de diversos sectores y comunidades dentro de un contexto social.
Aplicar pensamiento reflexivo y sistémico	Examina el papel y el impacto de las ciencias sociales en la dinámica comunitaria.	
	Reconoce la naturaleza compleja e interconectada de los conflictos y las posibles soluciones.	

Continúa

Modulo	Competencias	Estándar de evaluación	
Lectura crítica	Identificar y entender los contenidos locales de un texto	Analiza componentes discursivos específicos.	
		Encuentra eventos y figuras que se explican en diversos formatos narrativos.	
	Comprender la articulación textual	Analiza la arquitectura textual y las funciones seccionales.	
		Caracteriza los escenarios propuestos.	
		Deconstruye conexiones entre segmentos.	
		Identifica las ideas o afirmaciones presentes en un texto informativo.	
	Reflexionar a partir de un texto y evaluar su contenido	Calcula vínculos en elementos no secuenciales.	
		Analiza argumentos persuasivos para propuestas.	
		Crea diálogos intertextuales.	
		Determina las cargas axiomáticas.	
	Biología	Explicar fenómenos	Analiza los recursos retóricos.
			Sitúa el contenido en su marco de referencia.
Evalúa el uso de elementos naturales y tecnológicos en las dimensiones ecológica, médica y socioeconómica.			
Utilizar comprensivamente el conocimiento científico		Interpreta los mecanismos naturales utilizando principios y pruebas científicas.	
		Crea representaciones de procesos incorporando variables, conexiones teóricas y resultados.	
Indagar		Establece conexiones entre los fenómenos medioambientales y los marcos científicos.	
		Determina las características de los fenómenos naturales mediante el análisis crítico.	
		Reconoce la importancia de la investigación en la creación de conocimiento.	
		Extrae conclusiones combinando conocimientos y pruebas empíricas.	
Inglés		Demostrar habilidades comunicativas	Conecta patrones de datos para confirmar predicciones.
	Utiliza habilidades cognitivas y procedimentales al evaluar predicciones.		
	Marco Común Europeo de Referencia para las lenguas (MCER)		

Fuente: elaboración propia.

Metodología

La presente investigación es de tipo explicativa, al buscar las relaciones intrínsecas que permiten estimar cómo las variables de entrada influyen en el desempeño de la competencia de razonamiento cuantitativo (Oates, 2005). Puesto que utiliza datos históricos, se considera retrospectiva y no experimental, dado que no existe alteración de las variables de entrada al modelo.

Los datos utilizados corresponden a recopilación de las bases de datos de las pruebas SABER PRO y SABER 11 realizada por Delahoz-Dominguez *et al.* (2020b). La población de estudio corresponde a 12 411 estudiantes de ingeniería que presentaron el examen SABER PRO para el año 2020.

Para los propósitos del estudio, se utilizó el modelo de aprendizaje automático *random forest* (Breiman, 2001), el cual representa un sistema de ensamblaje enfocado en la recurrencia y la construcción de múltiples árboles de decisión mediante el mecanismo de agregación (Wright & Ziegler, 2017). Este método crea un conjunto de árboles predictivos, con un subconjunto diferente de predictores utilizados por cada instancia. Cada árbol genera una clasificación autónoma, que luego se combina con interpolación numérica para formar la predicción final. Esta arquitectura de predicción individual, que posteriormente se combina para crear modelos robustos, optimiza la precisión predictiva y de clasificación al ser menos sensible a los valores atípicos que los árboles por sí solos. La implementación de *random forest* incluye la selección automática de variables, lo que permite procesar conjuntos de datos de alta dimensión sin necesidad de realizar previamente pasos de reducción dimensional. Además, el algoritmo cuantifica la relevancia específica de cada atributo mediante permutaciones controladas, posibilitando evaluar variables individuales sin relación con sus interacciones con otros predictores del modelo.

Medidas de desempeño

El coeficiente de determinación R^2 y el error cuadrático medio (RMSE) fueron seleccionados como métricas para evaluar el desempeño del modelo *random forest*.

El R^2 es una métrica que indica qué proporción de la variabilidad en la variable dependiente (real) puede ser explicada por el modelo de regresión. El R^2 puede tomar valores entre 0 y 1, siendo 1 la capacidad del modelo para el 100 % de la variabilidad de los datos, y 0 un indicador de que el modelo es tan bueno como el resultado que puede entregar el promedio de la variable de respuesta. Puede calcularse mediante la siguiente ecuación:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\rho_n(i) - \hat{\rho}_n(i))^2}{\sum_{i=1}^n (\rho_n(i) - \bar{\rho}_n)^2} \quad (1)$$

El RMSE indica qué tan lejos, en promedio, están las predicciones del modelo respecto a los valores reales. Al estar en las mismas unidades de la variable de interés, permite una interpretación directa: un RMSE de 5 significa que las predicciones fallan en promedio por 5 unidades.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{\rho}_n(i) - \rho_n(i))^2} \quad (2)$$

Resultados

Este apartado se divide en dos secciones: la primera muestra un análisis exploratorio de los datos, donde a través de estadísticas descriptivas se representan detalladamente las características de los estudiantes, universidades y programas académicos que componen la base de datos utilizada. En segundo lugar, se presenta el resultado de los modelos de aprendizaje automático.

Análisis exploratorio de los datos

Como se evidencia en la tabla 2, en las carreras de ingeniería estudiadas el porcentaje de hombres es superior al de mujeres, excepto para las carreras de ingeniería industrial e ingeniería química, donde las mujeres representan mayoría. Es de resaltar el caso de ingeniería mecatrónica, donde el 91 % de los estudiantes son hombres.

Tabla 2
Género por programa de ingeniería

Carrera	Género		
	Femenino	Masculino	Total
ING_CIVIL	1273 (37 %)	2207 (63 %)	3480
ING_AERONÁUTICA	12 (27 %)	32 (73 %)	44
ING_AUTOMATIZACIÓN	16 (30 %)	37 (70 %)	53
ING_INDUSTRIAL	2810 (52 %)	2596 (48 %)	5406
ING_ELÉCTRICA	61 (22 %)	217 (78 %)	278
ING_MECATRÓNICA	11 (9 %)	105 (91 %)	116
ING_ELECTRÓNICA	175 (20 %)	721 (80 %)	896
ING_MECÁNICA	137 (12 %)	999 (88 %)	1136
ING_QUÍMICA	543 (54 %)	456 (46 %)	999

Fuente: elaboración propia.

En cuanto al régimen del colegio de donde provienen los alumnos, en la tabla 3 se encuentra una distribución heterogénea entre las carreras: se pasa de una proporción balanceada en ingeniería civil, a carreras donde la proporción de estudiantes procedentes de colegios privados es mayoritaria, como ingeniería química, siendo el caso de mayor proporción de estudiantes de colegios públicos el de ingeniería mecatrónica.

Tabla 3

Régimen del bachillerato por programa de ingeniería

Régimen del colegio		
Carrera	Colegios privados	Colegios públicos
ING_CIVIL	1725 (49,6 %)	1755 (50,4 %)
ING_AERONÁUTICA	25 (56,8 %)	19 (43,2 %)
ING_AUTOMATIZACIÓN	24 (45,3 %)	29 (54,7 %)
ING_INDUSTRIAL	2869 (53,1 %)	2537 (46,9 %)
ING_ELÉCTRICA	135 (48,6 %)	143 (51,4 %)
ING_MECATRÓNICA	45 (38,8 %)	71 (61,2 %)
ING_ELECTRÓNICA	401 (44,8 %)	495 (55,2 %)
ING_MECÁNICA	652 (57,4 %)	484 (42,6 %)
ING_QUÍMICA	686 (68,7 %)	313 (31,3 %)

Fuente: elaboración propia.

En la tabla 4 se aprecia cómo la mayoría de los estudiantes no trabaja al momento del tomar el examen; se identifica a ingeniería aeronáutica como la carrera con mayor proporción de estudiantes trabajando.

Tabla 4

Relación de estudiantes trabajando por programas de ingeniería

Estudiantes que trabajan		
Carrera	Trabajan	No trabajan
ING_CIVIL	143 (4,1 %)	3337 (95,9 %)
ING_AERONÁUTICA	4 (9,1 %)	40 (90,9 %)
ING_AUTOMATIZACIÓN	3 (5,7 %)	50 (94,3 %)
ING_INDUSTRIAL	220 (4,1 %)	5186 (95,9 %)
ING_ELÉCTRICA	11 (4 %)	267 (96 %)
ING_MECATRÓNICA	10 (8,6 %)	106 (91,4 %)
ING_ELECTRÓNICA	36 (4 %)	860 (96 %)
ING_MECÁNICA	46 (4 %)	1090 (96 %)
ING_QUÍMICA	27 (2,7 %)	972 (97,3 %)

Fuente: elaboración propia.

En la tabla 5, se muestra el resumen de promedios y desviaciones para cada una de las competencias evaluadas en el examen SABER 11 que sirven como variables de entrada a los modelos predictivos. Se resalta la carrera de ingeniería eléctrica, la cual presenta resultados promedio superiores al resto de las carreras estudiadas.

Tabla 5

Resumen estadístico de las variables de entrada

Carrera	MAT	DesvMAT	LC	DesvLC	BIO	DesvBIO	CS	DesvCS	ING	DesvING
ING_CIVIL	64,2	11,7	60,1	9,9	63,4	10,8	60,3	10,1	60,1	13,7
ING_AERONÁUTICA	60,0	9,1	57,8	6,4	60,2	8,6	58,5	8,3	57,9	13,3
ING_AUTOMATIZACIÓN	68,2	10,6	63,5	9,5	66,5	9,3	62,1	9,0	63,8	12,4
ING_INDUSTRIAL	61,9	11,4	59,6	9,8	61,7	10,7	59,2	9,8	60,5	14,1
ING_ELÉCTRICA	71,8	11,5	66,2	9,9	70,3	10,8	65,9	9,7	66,0	14,0
ING_MECATRÓNICA	60,7	9,6	55,8	9,7	60,4	10,3	57,0	10,1	55,3	10,8
ING_ELECTRÓNICA	67,6	12,0	62,7	10,1	67,4	11,1	63,1	10,0	63,8	14,2
ING_MECÁNICA	67,7	12,4	62,1	10,4	67,1	11,9	62,2	10,2	64,9	15,3
ING_QUÍMICA	69,4	10,4	65,5	8,9	70,2	9,9	65,6	9,2	69,2	13,1

Fuente: elaboración propia.

Modelo predictivo para la competencia de razonamiento cuantitativo

En la tabla 6 se presenta el resultado del modelo *random forest* y la clasificación de las variables según su importancia en la predicción de la competencia de razonamiento cuantitativo, de mayor a menor. Consecuentemente, las variables con la mayor importancia para predecir el resultado de la variable RC son matemáticas (MAT_S11) y biología (BIO_S11).

Tabla 6

Relación de aportes parciales por variables del modelo *random forest*

VARIABLES	MAT_S11	BIO_S11	LC_S11	CS_S11	ING_S11
Aporte	147,14	93,39	42,74	27,79	9,03
Orden de importancia	1	2	3	4	5

Fuente: elaboración propia.

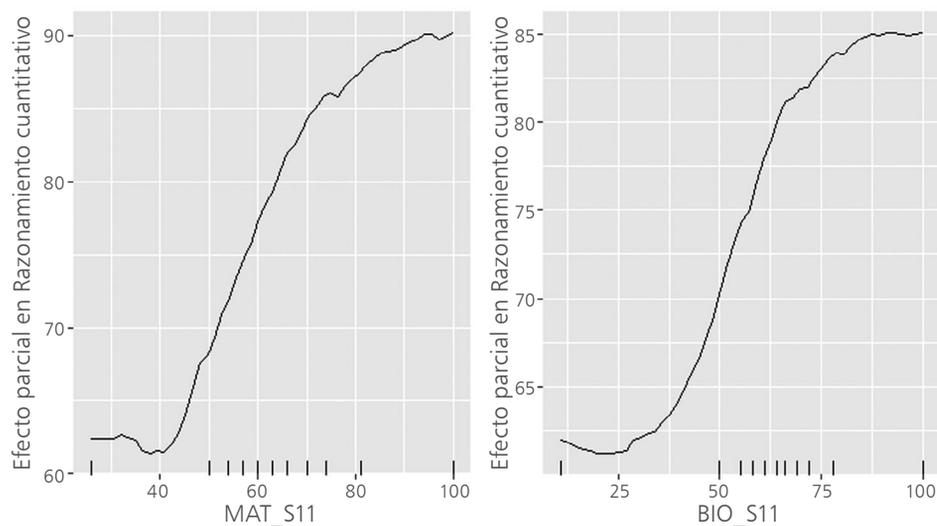
A continuación, se realiza un análisis particular del efecto que tienen estas variables, quitando el efecto conjunto de las otras variables predictoras sobre la predicción del desempeño en la competencia de razonamiento cuantitativo del examen SABER PRO. Para este propósito se presenta la figura 1, donde el eje X representa el puntaje en matemáticas en la prueba SABER 11 y el eje Y muestra el efecto parcial en el rendimiento de la competencia de razonamiento cuantitativo. Se observa que a medida que aumenta el puntaje en matemáticas (MAT_S11), el efecto sobre el rendimiento en razonamiento cuantitativo también aumenta de manera significativa, especialmente a partir de un puntaje de 50 en adelante, mostrando un crecimiento

rápido hasta cerca de 90 puntos en el eje Y. Esto indica que las competencias en matemáticas tienen un impacto positivo y considerable en razonamiento cuantitativo.

Seguidamente, para la competencia de biología se observa un comportamiento similar al de matemáticas, aunque con un crecimiento algo más moderado. A partir de un resultado global de 40 puntos en biología, el efecto sobre razonamiento cuantitativo empieza a aumentar significativamente, alcanzando su valor más alto en 85 puntos en el eje Y. Esto muestra que las competencias en biología también contribuyen positivamente al rendimiento en razonamiento cuantitativo, aunque con un impacto ligeramente menor en comparación con matemáticas.

Figura 1

Aporte parcial de las competencias en matemáticas y biología



Fuente: elaboración propia.

Validación del modelo

Para validar el enfoque de predicción del presente trabajo se presentan los resultados de desempeño del modelo de *random forest* para predecir efectivamente el resultado de la competencia de razonamiento cuantitativo a nivel universitario para estudiantes de bachillerato. En primer lugar, la tabla 7 evidencia que la variabilidad en los valores de RMSE y R^2 entre los subconjuntos de datos es relativamente baja, sugiriendo que el modelo *random forest* es robusto y consistente en la mayoría de las particiones del conjunto de datos. Además, la consistencia en los valores de R^2 cercanos a 0,70 y los RMSE alrededor de 10,5 indican que el modelo mantiene un equilibrio entre precisión predictiva y capacidad explicativa.

Tabla 7
Resultados de la fase de validación cruzada

RMSE	Rsquared	Resample
10,25	0,71	Fold_1
10,51	0,68	Fold_2
10,82	0,70	Fold_3
10,74	0,70	Fold_4
10,69	0,68	Fold_5
11,00	0,68	Fold_6
9,85	0,76	Fold_7
10,60	0,70	Fold_8
10,61	0,67	Fold_9
11,32	0,63	Fold_10

Fuente: elaboración propia.

Finalmente, se utilizaron 400 árboles, con una profundidad máxima de 5 niveles cada uno, debido a que esta combinación arrojó los mejores resultados: un error (RMSE) de 10,95 y un poder explicativo (R^2) de 0,69, lo cual sugiere que el modelo tiene una capacidad predictiva aceptable.

Discusión

Los hallazgos de la presente investigación se centran en la importancia del razonamiento cuantitativo como habilidad pilar para la formación en ingeniería. Como señalan Salazar-Fernández *et al.* (2019) en los resultados de un trabajo aplicado en la Universidad Austral de Chile, un bajo rendimiento matemático inicial es un predictor crítico de la deserción estudiantil, lo que respalda nuestra conclusión de que evaluar el rendimiento mediante esta competencia permite identificar a los estudiantes que podrían necesitar apoyo académico.

Lo anterior se refuerza al considerar que, como señalan Eichler y Gradwohl (2021), el desempeño preuniversitario influye significativamente en el éxito profesional posterior. Los resultados de la presente investigación concuerdan con esta posición, destacando que las proyecciones académicas no solo deben vincularse al rendimiento escolar, sino también emplearse como herramienta para mejorar trayectorias laborales.

Un contrapunto valioso surge de Bosch *et al.* (2021), cuyo trabajo revela que las actividades educativas aplicadas tardíamente en el ciclo formativo pueden impactar más el rendimiento final que los antecedentes académicos previos. Esto sugiere que las intervenciones oportunas mantienen su relevancia incluso con bases formativas sólidas.

Por otro lado, en el ámbito metodológico es importante resaltar la investigación realizada por Kongara y Sathyanarayana (2019), quienes lograron

una precisión de predicción del 88,6 % en matemáticas utilizando la técnica de *random forest*. Si bien su enfoque clasifica el rendimiento en categorías discretas (a diferencia de nuestro modelo continuo), su éxito valida la utilidad de esta técnica. De igual manera, Soto-Acevedo *et al.* (2023) demostraron la viabilidad de integrar evaluaciones estandarizadas mediante la combinación de las pruebas SABER 11 y SABER PRO para predecir el rendimiento en ingeniería ($AUC = 0,82$).

Paralelamente, la transversalidad de esta competencia fue demostrada por Mayes *et al.* (2020), quienes confirmaron que el pensamiento cuantitativo mejora la formación universitaria en el área de biología (ciencias naturales), lo cual fortalece la idea de una conexión interdisciplinaria al encontrar correlaciones significativas entre el razonamiento cuantitativo (SABER PRO) y las habilidades en el área de la biología (SABER 11). Esto respalda los hallazgos de Tossavainen *et al.* (2021) sobre la manera en que las deficiencias matemáticas afectan el desempeño en ingeniería.

Sin embargo, en la investigación de Zakariya *et al.* (2023) no se evidencian correlaciones significativas entre la formación matemática previa, las estrategias de aprendizaje y el rendimiento en ingeniería, una discrepancia que representa la complejidad del fenómeno estudiado.

En este sentido, el modelo de la presente investigación de bosque aleatorio ($R^2 = 0,69$, $RMSE = 10,95$) demostró una alta tasa predictiva, coincidiendo con Nachouki *et al.* (2023) en cuanto a la utilidad de este algoritmo para discriminar variables que contribuyen al rendimiento. La fuerte correlación entre las habilidades matemáticas en la educación media (SABER 11) y el rendimiento posterior es pertinente, lo que resalta la necesidad de fortalecer estas habilidades en la secundaria. Ahora bien, aunque la influencia de las habilidades en biología fue menos significativa, sus refuerzos siguen siendo valiosos en las etapas iniciales.

Finalmente, estos resultados proporcionan bases para el diseño de intervenciones en la educación media. Asimismo, como estrategia de retención y mejora del rendimiento académico en la educación superior es necesario contemplar un plan para el fortalecimiento de estas competencias básicas, focalizando a los estudiantes que obtuvieron un rendimiento inferior en SABER 11, especialmente en matemáticas.

Por su parte, se reconoce como una limitación la necesidad de contemplar el análisis de las variables asociadas a los factores socioeconómicos para lograr una comprensión más completa del rendimiento universitario.

Conclusiones

La presente investigación tuvo como objetivo la evaluación de un modelo de predicción que permitiera definir los elementos claves que influyen

en el rendimiento académico. Teniendo en cuenta lo anterior, se construyó una base de información para el análisis de este problema identificado.

En este orden de ideas, el modelo se construyó a partir del algoritmo *random forest*. Los resultados indican que las habilidades en matemáticas y biología son los predictores más significativos del éxito en el razonamiento cuantitativo. Esta correlación sugiere que el fortalecimiento de estas habilidades durante la educación media tendrá un impacto directo en las habilidades cuantitativas universitarias.

Además, el modelo de la presente investigación demostró una alta eficiencia predictiva ($R^2 = 0,69$), lo que confirma su utilidad fiable para proyectos académicos. Esta capacidad es especialmente útil en entornos educativos complejos, donde el algoritmo *random forest* destaca por procesar grandes cantidades de datos y capturar interacciones no lineales entre variables.

Por último, la investigación genera su aporte en la frontera del conocimiento de modelos para el análisis de la calidad en la educación, en la medida en que el actual modelo diagnostica y logra predecir escenarios para la mejora continua.

Sobre los autores

Enrique Delahoz-Dominguez es docente de planta de la Universidad del Magdalena, Colombia, e investigador en temas de inteligencia artificial aplicada temas de educación.

Rohemi Zuluaga-Ortiz es ingeniero industrial, magister en ingeniería y doctorando en estadística, optimización y matemática aplicada. Docente en la Universidad del Sinú, Colombia. Investigador en áreas de analítica de datos, eficiencia y productividad.

Carlos García-Yerena es docente de educación física de la Universidad del Magdalena, Colombia, e investigador en el área de la educación, salud, entrenamiento y del deporte.

Referencias

- Aparicio, J., Cordero, J. M., & Ortiz, L. (2019). Measuring efficiency in education: The influence of imprecision and variability in data on DEA estimates. *Socio-Economic Planning Sciences*, 68, 100698. <https://doi.org/10.1016/j.seps.2019.03.004>
- Berens, J., Schneider, K., Gortz, S., Oster, S., & Burghoff, J. (2019). Early Detection of Students at Risk—Predicting Student Dropouts Using Administrative Student Data from German Universities and Machine Learning Methods. *Journal of Educational Data Mining*, 11(3), 1-41. <https://eric.ed.gov/?id=EJ1241620>

- Bosch, E., Seifried, E., & Spinath, B. (2021). What successful students do: Evidence-based learning activities matter for students' performance in higher education beyond prior knowledge, motivation, and prior achievement. *Learning and Individual Differences, 91*, 102056. <https://doi.org/10.1016/j.lindif.2021.102056>
- Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32.
- Columbu, S., Porcu, M., & Sulis, I. (2021). University choice and the attractiveness of the study area: Insights on the differences amongst degree programmes in Italy based on generalised mixed-effect models. *Socio-Economic Planning Sciences, 74*, 100926. <https://doi.org/10.1016/j.seps.2020.100926>
- Delahoz-Dominguez, E. J., Guillen-Ibarra, & Fontalvo-Herrera, T. (2020a). Análisis de la acreditación de calidad en programas de ingeniería industrial y los resultados en las pruebas nacionales estandarizadas, en Colombia. *Formación Universitaria, 13*(1), 127-134. <https://doi.org/10.4067/S0718-50062020000100127>
- Delahoz-Dominguez, E., Zuluaga, R., & Fontalvo-Herrera, T. (2020b). Dataset of academic performance evolution for engineering students. *Data in Brief, 30*, 105537. <https://doi.org/10.1016/j.dib.2020.105537>
- Eichler, A., & Gradwohl, J. (2021). Investigating Motivational and Cognitive Factors which Impact the Success of Engineering Students. *International Journal of Research in Undergraduate Mathematics Education, 7*(3), 417-437. <https://doi.org/10.1007/s40753-020-00127-4>
- Gasevic, D., Rose, C., Siemens, G., Wolff, A., & Zdrahal, Z. (2014). *Learning analytics and machine learning*. En S. Teasley, & A. Pardo (Eds.), *LAK 2014: Fourth International Conference on Learning Analytics and Knowledge* (pp. 287-288). Association for Computing Machinery (ACM). <https://doi.org/10.1145/2567574.2567633>
- Gwelo, A. S. (2019). Determinants of career choice among university students. *MOJEM: Malaysian Online Journal of Educational Management, 7*(1), Article 1.
- Hart, P. F., & Rodgers, W. (2024). Competition, competitiveness, and competitive advantage in higher education institutions: A systematic literature review. *Studies in Higher Education, 49*(11), 2153-2177. <https://doi.org/10.1080/03075079.2023.2293926>
- Heidrich, L., Victória Barbosa, J. L., Cambruzzi, W., Rigo, S. J., Martins, M. G., & dos Santos, R. B. S. (2018). Diagnosis of learner dropout based on learning styles for online distance learning. *Telematics and Informatics, 35*(6), 1593-1606. <https://doi.org/10.1016/j.tele.2018.04.007>
- Kamal, N., Sarker, F., Rahman, A., Hossain, S., & Mamun, K. A. (2024). Recommender System in Academic Choices of Higher Education: A Systematic Review. *IEEE Access, 12*, 35475-35501. <https://doi.org/10.1109/ACCESS.2024.3368058>
- Kongara, D., & Sathyanarayana, N. (2019). Relief-F and Budget Tree Random Forest Based Feature Selection for Student Academic Performance Prediction. *International Journal of Intelligent Engineering and Systems, 12*, 30-39. <https://doi.org/10.22266/ijies2019.0228.04>

- Londoño-Vélez, J., Rodríguez, C., & Sánchez, F. (2020). Upstream and Downstream Impacts of College Merit-Based Financial Aid for Low-Income Students: Ser Pilo Paga in Colombia. *American Economic Journal: Economic Policy*, 12(2), 193-227. <https://doi.org/10.1257/pol.20180131>
- Mayes, R., Long, T., Huffling, L., Reedy, A., & Williamson, B. (2020). Undergraduate Quantitative Biology Impact on Biology Preservice Teachers. *Bulletin of Mathematical Biology*, 82(6), 63. <https://doi.org/10.1007/s11538-020-00740-z>
- McLaughlin, J. E., Morbitzer, K., Meilhac, M., Poupart, N., Layton, R. L., & Jarstfer, M. B. (2023). Standards needed? An exploration of qualifying exams from a literature review and website analysis of university-wide policies. *Studies in Graduate and Postdoctoral Education*, 15(1), 19-33. <https://doi.org/10.1108/SGPE-11-2022-0073>
- Nachouki, M., Mohamed, E. A., Mehdi, R., & Abou Naaj, M. (2023). Student course grade prediction using the random forest algorithm: Analysis of predictors' importance. *Trends in Neuroscience and Education*, 33, 100214. <https://doi.org/10.1016/j.tine.2023.100214>
- Oates, B. J. (2005). *Researching information systems and computing*. Sage.
- Salazar-Fernandez, J. P., Sepúlveda, M., & Muñoz-Gama, J. (2019). Influence of Student Diversity on Educational Trajectories in Engineering High-Failure Rate Courses that Lead to Late Dropout. *2019 IEEE Global Engineering Education Conference (EDUCON)*, Dubai, United Arab Emirates, pp. 607-616. <https://doi.org/10.1109/EDUCON.2019.8725143>
- Soto-Acevedo, M., Abuchar-Curi, A. M., Zuluaga-Ortiz, R. A., & Delahoz-Dominiguez, E. J. (2023). A Machine Learning Model to Predict Standardized Tests in Engineering Programs in Colombia. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, 18(3), 211-218. <https://doi.org/10.1109/rita.2023.3301396>
- Tossavainen, T., Rensaa, R. J., Haukkanen, P., Mattila, M., & Johansson, M. (2021). First-year engineering students' mathematics task performance and its relation to their motivational values and views about mathematics. *European Journal of Engineering Education*, 46(4), 604-617. <https://doi.org/10.1080/03043797.2020.1849032>
- Visbal-Cadavid, D., Martínez-Gómez, M., & Guijarro, F. (2017). Assessing the efficiency of public universities through DEA. A case study. *Sustainability*, 9(8), 1416.
- Wright, M. N., & Ziegler, A. (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, 077(i01). https://econpapers.repec.org/article/jssjstsof/v_3a077_3ai01.htm
- Zakariya, Y. F., Nilsen, H. K., Bjørkestøl, K., & Goodchild, S. (2023). Analysis of relationships between prior knowledge, approaches to learning, and mathematics performance among engineering students. *International Journal of Mathematical Education in Science and Technology*, 54(6), 1015-1033. <https://doi.org/10.1080/0020739X.2021.1984596>