

Predicción del éxito en el examen de habilitación profesional: un modelo de regresión logística basado en variables multifactoriales

Predicting success in the professional qualification exam: a logistic regression model based on multifactorial variables

Saul Yasaca Pucuna*

Escuela Superior Politécnica de Chimborazo
Riobamba-Ecuador.
syasaca@epoch.edu.ec
<https://orcid.org/0000-0001-8851-8258>

Juan Diego Erazo Rodríguez

Escuela Superior Politécnica de Chimborazo
Riobamba-Ecuador.
juan.erazo@epoch.edu.ec
<https://orcid.org/0000-0003-0152-5645>

*Correspondencia:

syasaca@epoch.edu.ec

Cómo citar este artículo:

Yasaca, S., & Erazo, J. (2025). Predicción del éxito en el examen de habilitación profesional: un modelo de regresión logística basado en variables multifactoriales. *Revista de Investigación Educativa Niveles*, 2(2), 36-47.
<https://doi.org/10.61347/rien.v2i2.77>

Recibido: 31 de julio de 2025

Proceso de evaluación:

31 de julio al 1 de septiembre de 2025

Aceptado: 2 de septiembre de 2025

Publicado: 5 de septiembre de 2025

Resumen: La tasa irregular de aprobación en el Examen de Habilitación Profesional (EHEP) afecta a estudiantes de la carrera de Medicina de la Escuela Superior Politécnica de Chimborazo, examen obligatorio para ingresar al año de prácticas y ejercer la profesión en Ecuador. El objetivo de esta investigación fue desarrollar un modelo de regresión logística que predijera el éxito en el EHEP con base en el análisis de múltiples variables. Se aplicó la metodología de Ciencia del Diseño que consta de tres ciclos: relevancia (revisión sistemática con PRISMA), diseño (desarrollo del modelo en Python utilizando la muestra de 117 estudiantes y el estándar CRISP-DM) y rigor (validación del modelo mediante métricas de rendimiento). La variable predictora más influyente fue las horas de estudio. El modelo de regresión logística obtuvo una exactitud de 80.36 %, un F1-score de 0,643 y AUC de 0.710 en la curva de Precisión-Sensibilidad, indicativos de su capacidad para distinguir las clases especialmente desbalanceadas. Los hallazgos no solo confirman la complejidad del rendimiento académico, sino que abren una base para trabajos futuros. Además, el modelo desarrollado ofrece una herramienta útil para que las instituciones educativas implementen estrategias de prevención y apoyo personalizado a estudiantes en riesgo de no aprobar el EHEP, contribuyendo así a mejorar las tasas de éxito.

Palabras clave: Ciencia del diseño, CRISP-DM, examen, habilitación profesional, PRISMA, regresión logística.

Abstract: The irregular pass rate on the Professional Qualification Exam (EHEP) affects medical students at the Escuela Superior Politécnica de Chimborazo, a mandatory exam for entering the internship year and practicing medicine in Ecuador. The objective of this research was to develop a logistic regression model that would predict success on the EHEP based on the analysis of multiple variables. The Design Science methodology was applied, consisting of three cycles: relevance (systematic review with PRISMA), design (model development in Python using a sample of 117 students and the CRISP-DM standard), and rigor (model validation using performance metrics). The most influential predictor variable was hours of study. The logistic regression model achieved an accuracy of 80.36%, an F1-score of 0.643, and an AUC of 0.710 on the Precision-Sensitivity curve, indicative of its ability to distinguish particularly unbalanced classes. The findings not only confirm the complexity of academic performance but also lay the groundwork for future work. In addition, the model developed offers a useful tool for educational institutions to implement prevention strategies and personalized support for students at risk of failing the EHEP, thus contributing to improved success rates.

Keywords: CRISP-DM, design science, exam, logistic regression, PRISMA, professional qualification.

Copyright: Derechos de autor 2025 Saul Yasaca Pucuna, Juan Diego Erazo Rodríguez.



Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial 4.0.

1. Introducción

El Examen de Habilitación para el Ejercicio Profesional (EHEP), requisito obligatorio en Ecuador para estudiantes de las carreras de medicina, odontología y enfermería que deben cumplir un año de prácticas en hospitales, previo a ejercer profesionalmente, fue implementado y aplicado por el Consejo de Aseguramiento de la Calidad de la Educación Superior desde 2020 (CACES, 2024).

Según la Ley Orgánica de Educación Superior, Artículo 104, el CACES tiene la responsabilidad de desarrollar el EHEP en carreras que comprometan el interés público, especialmente en áreas relacionadas con la vida, salud y seguridad de la ciudadanía. Este examen, de gran impacto y alcance nacional, evalúa objetivamente las competencias profesionales necesarias para ejercer legalmente en el país (CACES, 2024).

Pese a su importancia, el EHEP presenta una irregular tasa de aprobación, un desafío tanto para los egresados como para las instituciones responsables de su formación. Esta situación subraya la necesidad de desarrollar modelos predictivos que identifiquen y mitiguen los factores que influyen en el éxito académico en este tipo de pruebas.

La literatura señala una amplia variedad de factores que influyen en el rendimiento académico universitario (Alyahyan & Düştegör, 2020), que incluyen variables sociodemográficas, psicosociales, emocionales, nivel de satisfacción, consumo de sustancias, ausentismo, aspectos académicos, ambiente universitario, calidad de relación estudiante-docente y horas de estudio (Merchán-Clavellino et al., 2019).

Estudios a nivel internacional como el de Carhuaricra (2023) identifican factores asociados con el desempeño en exámenes de habilitación para el ejercicio médico, y señalan variables relevantes como edad, estado civil, número de hijos, capacitación continua, orden de mérito y sede hospitalaria. No obstante, su enfoque se limita a un número reducido de variables y a un contexto específico, lo que evidencia la necesidad de un análisis más exhaustivo e integral. Además, otros estudios exploran los factores que influyen en el rendimiento académico en estudiantes universitarios (Gutiérrez-Monsalve et al., 2021; Pastor, 2024), destacando la relevancia del estrés, la calidad del sueño y las estrategias de aprendizaje (Garbanzo, 2007).

En Ecuador la literatura sobre los factores asociados al rendimiento asociado al EHEP es escasa. Escobar-Jiménez & Torres-Rentería (2021) analizan la relación entre el éxito educativo y las condiciones socioeconómicas en los exámenes de habilitación médica, pero estas investigaciones aún no permiten diseñar o implementar estrategias de intervención amplias y efectivas. Por tanto, resulta vital desarrollar modelos predictivos que identifiquen tempranamente a los estudiantes en riesgo y faciliten la implementación de programas de apoyo personalizados. Aunque se han aplicado modelos de regresión logística para la predicción del rendimiento académico en distintos contextos universitarios (Pérez et al., 2023), pocos estudios se han enfocado específicamente en el EHEP a nivel nacional.

Es así como existe la necesidad de comprender y mitigar los factores que inciden en la baja tasa de aprobación del examen EHEP. A través del análisis de variables académicas, emocionales y de estilo de vida, se busca no solo predecir el éxito en el examen, sino proporcionar herramientas para mejorar la preparación estudiantil, ofrecer acompañamiento personalizado y optimizar los programas académicos, contribuyendo además al conocimiento científico sobre el rendimiento en exámenes profesionales.

En este contexto se plantea la siguiente pregunta de investigación: ¿Cuáles son las variables multifactoriales más significativas que permiten predecir el éxito en el EHEP y cómo puede un modelo

de regresión logística estimar la probabilidad de aprobación? Por lo tanto, el objetivo de este estudio fue desarrollar un modelo de regresión logística que predijera el éxito en el EHEP con base en el análisis de múltiples variables.

Para ello se plantearon los objetivos específicos: Identificar las variables más influyentes en la aprobación del EHEP mediante un análisis exploratorio y estadístico para comprender su impacto en el desempeño estudiantil; Desarrollar un modelo de regresión logística que utilice las variables significativas para predecir la probabilidad de éxito en el EHEP; y Validar del modelo mediante métricas de rendimiento como son exactitud, precisión, sensibilidad, F1-score y curvas de Precisión-Sensibilidad, asegurando su confiabilidad.

Se fundamenta en teorías del aprendizaje autorregulado, que resaltan la motivación y la autoeficacia como elementos fundamentales para el logro académico (Gálvez-Gamboa et al., 2024). Asimismo, se considera el síndrome de Burnout como un factor que influye negativamente en el rendimiento académico (Álvarez-Raigoza et al., 2024). Estas teorías proporcionan un marco conceptual para analizar cómo diversas variables psicométricas, junto con factores académicos y de estilo de vida, afectan el desempeño en el EHEP. Desde una perspectiva metodología cuantitativa, la investigación busca analizar patrones, frecuencias y correlaciones que permitan desarrollar un modelo predictivo generalizable (Aguilar-Reyes et al., 2025).

2. Metodología

La presente investigación adoptó un enfoque cuantitativo con un diseño preexperimental predictivo orientado a anticipar un resultado binario: la aprobación o no del EHEP. Para ello, se empleó la regresión logística, técnica estadística ampliamente utilizada para modelar la probabilidad de ocurrencia de eventos dicotómicos en función de variables multifactoriales. Esta metodología es especialmente adecuada para analizar la influencia combinada de múltiples factores en la probabilidad de éxito académico.

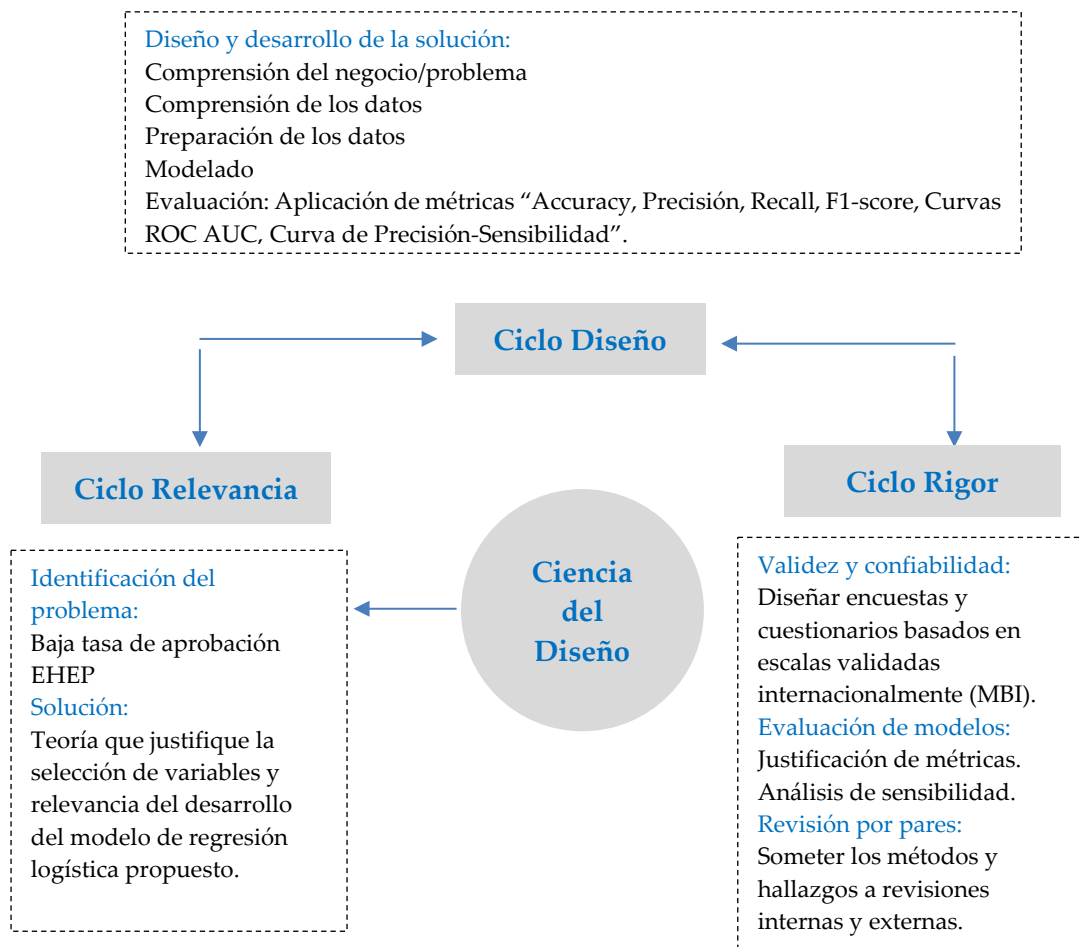
El modelo incorpora factores académicos, psicométricos y de estilo de vida, tales como calificaciones, técnicas y horas de estudio, antecedentes de depresión, niveles de ansiedad, motivación, autoeficacia, síndrome de Burnout, calidad y cantidad de sueño (Suardiaz-Muro et al., 2020). Este enfoque identifica los factores más influyentes para construir un modelo predictivo que facilite la detección temprana de riesgos y oriente sobre estrategias de intervención efectivas (Sánchez-Cubo et al., 2023). Esta selección se alinea con investigaciones que demuestran la importancia del promedio de calificaciones (GPA) y factores psicosociales como predictores comunes del desempeño académico, así como su relevancia para la elegibilidad de graduación, becas y acceso a posgrados.

El estudio se desarrolló en la Escuela Superior Politécnica de Chimborazo (ESPOCH), específicamente en la carrera de Medicina. La población objetivo estuvo constituida por estudiantes que se preparan para rendir el EHEP. La muestra se seleccionó mediante muestreo aleatorio estratificado, asegurando la representatividad de los diferentes.

Para la construcción y validación del modelo predictivo, se utilizó la metodología de Ciencia del Diseño (Design Science Research, DSR), enfocada en la creación y evaluación rigurosa de artefactos tecnológicos o conceptuales que resuelvan problemas reales mediante soluciones prácticas y sistemáticas. Este marco se estructura en tres ciclos principales: relevancia, diseño y rigor, los cuales garantizan un proceso ordenado y exhaustivo a lo largo de la investigación (figura 1).

Figura 1

Metodología Ciencia de Diseño, DSR



Ciclo de relevancia (Relevance cycle)

En el ciclo de relevancia se define el problema central: las bajas tasas de aprobación del EHEP. Como parte de este ciclo, se llevó a cabo una revisión sistemática de la literatura especializada, siguiendo el protocolo PRISMA (Page et al., 2021). Esta revisión identificó factores asociados al rendimiento académico (Flores-Cohaila, 2022) y la predicción del éxito en exámenes profesionales, focalizándose en variables reconocidas como calificaciones, técnicas y horas de estudio, variables psicométricas (Agha et al., 2023) y aspectos de estilo de vida (Rodríguez et al., 2021). La variable dependiente o *outcome* se definió como el éxito en el EHEP, categorizada dicotómicamente en aprobado (1) o no aprobado (0).

Durante este ciclo, se destacó el valor de integrar diferentes tipos de variables como características o *features* para el entrenamiento del modelo, permitiendo captar la complejidad multifactorial del fenómeno. La regresión logística, por su capacidad para manejar variables continuas y categóricas, resulta especialmente útil para determinar el efecto individual de cada predictor sobre la probabilidad de aprobar, facilitando además una interpretación clara y práctica para la toma de decisiones educativas.

Ciclo del diseño (Design cycle)

El ciclo de diseño se desarrolló siguiendo el estándar CRISP-DM (Cross Industry Standard Process for Data Mining), que contempla seis fases interrelacionadas. Inicialmente, en la fase de comprensión del negocio, se planteó el objetivo integral de desarrollar un modelo capaz de predecir el éxito en el EHEP y

responder a preguntas clave sobre los factores que influyen en la aprobación, la identificación temprana de estudiantes en riesgo y la aplicación de los resultados para mejorar el desempeño académico.

Posteriormente, en la fase de comprensión de los datos, se recopilaron y exploraron diversas fuentes, incluyendo encuestas para variables psicométricas y de estilo de vida, cuestionarios piloto estructurados y registros académicos universitarios. Se realizó un análisis preliminar para identificar y corregir valores atípicos, evaluar la representatividad de los datos y explorar relaciones iniciales entre variables.

La preparación de los datos consistió en integrar, limpiar y transformar los datos para su análisis posterior. Se abordó el manejo de valores faltantes mediante imputación, se codificaron variables categóricas o se podría utilizar técnicas como One-Hot Encoding para dichas variables y se estandarizaron variables numéricas relevantes con métodos de escalamiento (StandardScale y MinMaxScaler). Además, se implementaron técnicas de sobremuestreo y submuestreo para equilibrar las clases en caso de desbalance significativo, asegurando un entrenamiento más robusto del modelo. Finalmente, se seleccionaron variables relevantes mediante análisis de correlación visualizado con mapas de calor.

En la fase de modelado, el enfoque principal desarrolló el modelo de regresión logística binaria mediante la biblioteca Scikit-learn de Python, entrenado con el conjunto de datos preparado.

La etapa de evaluación implicó validar el rendimiento del modelo a través de métricas específicas para clasificación binaria, tales como la precisión (accuracy), sensibilidad (recall), F1-score, matriz de confusión, curva ROC y su respectiva área bajo la curva (AUC). Asimismo, se consideró la curva de Precisión-Sensibilidad, especialmente relevante para problemas con clases desbalanceadas, lo que permitió una valoración más precisa del desempeño predictivo en la detección de casos positivos.

Finalmente, en la fase de implementación, el modelo entrenado en python permite introducir datos de nuevos estudiantes y obtener predicciones en tiempo real. Además, se planteó un monitoreo continuo del rendimiento del modelo para asegurar su estabilidad y precisión en diferentes escenarios y a lo largo del tiempo.

Ciclo de Rigor (Rigor cycle)

El ciclo de rigor aseguró la solidez técnica y científica del artefacto construido. Para garantizar validez y confiabilidad, los instrumentos de recolección de datos, como encuestas y cuestionarios, se diseñaron con base en escalas validadas internacionalmente, tales como el Maslach Burnout Inventory (MBI). La evaluación del modelo se sustentó en la selección justificada de métricas, priorizando el *recall* para identificar con precisión estudiantes en riesgo, aun considerando cierto nivel de falsos positivos.

Asimismo, se realizó un análisis de sensibilidad para evaluar el impacto de variaciones en las variables predictoras sobre las predicciones del modelo. Finalmente, se sometió todo el proceso metodológico, el modelo desarrollado y sus resultados a revisión por pares, involucrando expertos y especialistas para recibir retroalimentación crítica que fortaleciera la calidad y solidez del estudio.

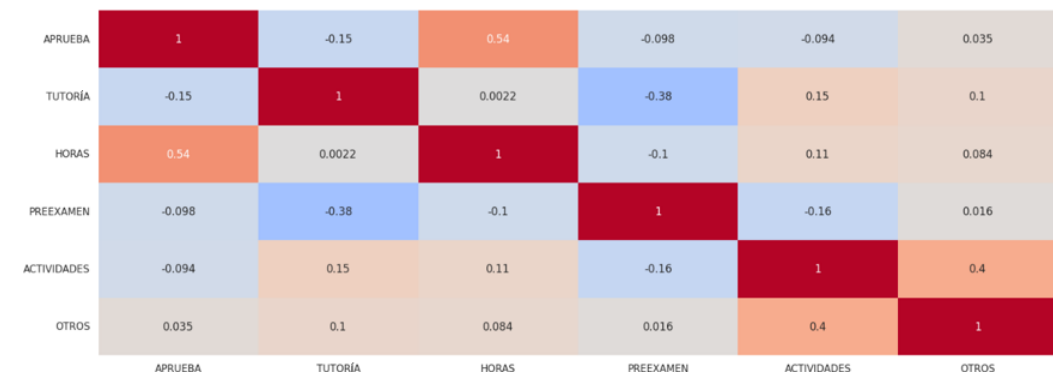
3. Resultados

Análisis de la correlación de variables

Se utilizó un mapa de calor (*heatmap*) para visualizar la matriz de correlación entre las variables seleccionadas y la variable dependiente "APRUEBA". Las variables demográficas, psicométricas y de estilo de vida no fueron consideradas en este análisis debido a su baja incidencia y problemas de convergencia en el modelo (figura 2).

Figura 2

Mapa de calor de correlación



La variable “HORAS” de estudio mostró una correlación positiva moderada (0.54) con la probabilidad de aprobar el examen, lo que indica que dedicar más tiempo al estudio incrementa significativamente las posibilidades de éxito. En contraste, las variables “TUTORÍA” (-0.15), “PREEXAMEN” (-0.098) y “ACTIVIDADES” (-0.094) registraron correlaciones negativas leves, sugiriendo que un mayor número de tutorías, mejores calificaciones en preexámenes y mayor participación en actividades curriculares se asocian ligeramente con una menor probabilidad de aprobar.

Además, la variable “OTROS” (0.035) tuvo una correlación muy baja, prácticamente sin relación lineal con la aprobación. Se observó también una correlación negativa moderada entre “TUTORÍA” y “PREEXAMEN” (-0.38), lo que posiblemente indica que estudiantes con más tutorías obtienen calificaciones menores en los preexámenes, y una correlación positiva moderada entre “ACTIVIDADES” y “OTROS” (0.4).

Por lo tanto, para la construcción y desarrollo del modelo de aprendizaje para predecir la variable “APRUEBA”, la variable “HORAS” resulta probablemente la característica más influyente para predecir. Mientras que las variables “TUTORÍA”, “PREEXAMEN” y “ACTIVIDADES” podrían aportar información adicional, aunque con relaciones más complejas que requieren análisis profundos. La variable “OTROS” parece tener bajo impacto.

Análisis del modelo regresión logística

Se desarrolló un modelo de regresión logística para predecir la variable binaria “APRUEBA”, enfocándose en estimar la probabilidad de aprobar el EHEP (figura 3).

Las variables predictoras incluidas fueron “TUTORÍA”, “HORAS”, “PREEXAMEN”, “ACTIVIDADES” y “OTROS”, con una muestra de 117 estudiantes. La optimización mediante máxima verosimilitud (MLE) fue exitosa, convergiendo en seis iteraciones, lo que refleja un proceso computacional eficiente.

El ajuste del modelo, evaluado mediante el pseudo R-cuadrado, fue moderado (0.3160), indicando que el modelo explica aproximadamente un 31.6 % de la variabilidad en la probabilidad de aprobar. La log-verosimilitud del modelo (-50.453) mejoró considerablemente frente al modelo nulo (-73.760), y la prueba de razón de verosimilitud (LLR) arrojó un valor extremadamente bajo (6.81e-09), confirmando la significancia estadística del modelo.

En cuanto a los coeficientes específicos, “HORAS” mostró un efecto positivo fuerte y altamente significativo (coeficiente 2.835, $p < 0.001$), señalando que, a mayor tiempo de estudio, mayor es la probabilidad de aprobar. Las variables “PREEXAMEN” (-0.061, $p = 0.020$) y “ACTIVIDADES” (-0.639,

$p = 0.027$) presentaron efectos negativos significativos, mientras que "TUTORÍA" tuvo un coeficiente negativo marginalmente significativo (-0.616 , $p = 0.076$). Por último, la variable "OTROS" no resultó significativa (coeficiente 0.304 , $p = 0.240$), lo que sugiere que su aporte al modelo es limitado.

Figura 3

Resultado Regresión Logística

```

Optimization terminated successfully.
Current function value: 0.431220
Iterations 6
Logit Regression Results
Dep. Variable: APRUEBA      No. Observations: 117
Model: Logit              Df Residuals: 111
Method: MLE               Df Model: 5
Date: Mon, 24 Mar 2025    Pseudo R-squ.: 0.3160
Time: 03:27:07           Log-Likelihood: -50.453
converged: True          LL-Null: -73.760
Covariance Type: nonrobust LLR p-value: 6.810e-09

      coef  std err   z   P>|z| [0.025 0.975]
-----+-----
const  1.6054   1.340   1.198  0.231 -1.021  4.232
TUTORIA -0.6161   0.347  -1.776  0.076 -1.296  0.064
HORAS   2.8356   0.537   5.283  0.000  1.784  3.887
PREEXAMEN -0.0610   0.026  -2.325  0.020 -0.112 -0.010
ACTIVIDADES -0.6396   0.289  -2.213  0.027 -1.206 -0.073
OTROS   0.3040   0.259   1.175  0.240 -0.203  0.811
    
```

Análisis de la matriz de confusión y métricas en regresión logística

La matriz de confusión mostró que el modelo clasificó correctamente 97 observaciones como aprobadas (verdaderos positivos) y 38 como no aprobadas (verdaderos negativos). Se identificaron 14 falsos positivos, es decir, casos predichos como aprobados que en realidad no lo fueron, y 19 falsos negativos, en los que el modelo no logró detectar algunos estudiantes que sí aprobaron.

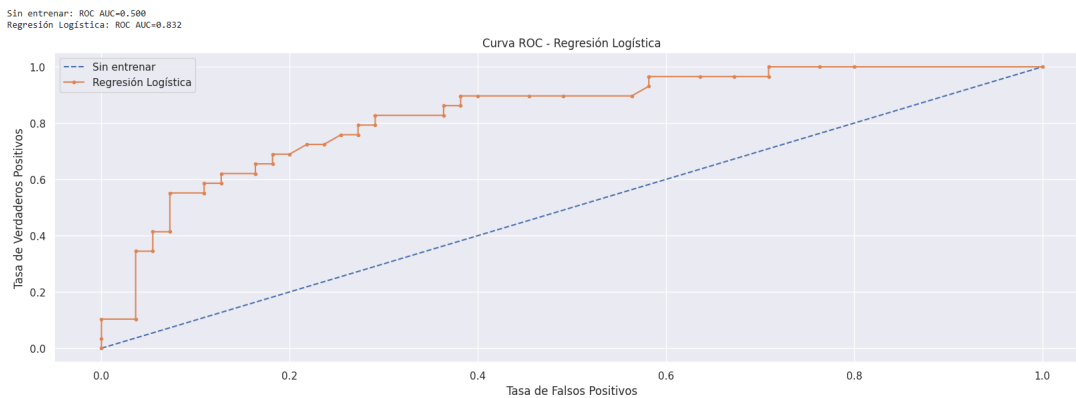
El desempeño general del modelo fue favorable, con una exactitud del 80.36 %, lo que refleja la proporción total de predicciones correctas. La precisión fue del 87.38 %, indicando que la mayoría de las predicciones positivas fueron acertadas, y la sensibilidad alcanzó 83.62 %, lo que representa el porcentaje de casos positivos correctamente identificados. El F1-score, que combina precisión y sensibilidad, fue de 85.38 %, mostrando un balance adecuado entre ambas métricas. Sin embargo, la presencia de 19 falsos negativos evidencia que el modelo puede tener limitaciones para identificar todos los casos positivos, lo cual es importante considerar en aplicaciones prácticas.

Análisis de curvas ROC AUC en regresión logística

Para evaluar la capacidad discriminativa del modelo se analizó la curva ROC (figura 4) y se obtuvo un área bajo la curva (AUC) de 0.832, lo que denota un rendimiento notablemente superior al azar y confirma la habilidad del modelo para clasificar correctamente casos positivos y negativos. La optimización del modelo utilizó el algoritmo L-BFGS para estimar los coeficientes óptimos, contribuyendo a su capacidad predictiva en datos no observados previamente.

Figura 4

Curvas ROC-AUC



Dado el desbalance en las clases de la muestra, se complementó la evaluación con la curva de Precisión-Sensibilidad (figura 5), más adecuada para estos escenarios. En esta curva se obtuvo un F1-score de 0.643 y un AUC de 0.710, valores que indican un rendimiento aceptable del modelo para diferenciar entre estudiantes que aprueban y no aprueban el EHEP. Estos resultados en conjunto validan que el modelo de regresión logística es efectivo para la predicción del éxito en el examen, con un buen balance entre precisión y capacidad de detección.

Figura 5

Curva de Precisión-Sensibilidad



4. Discusión

El modelo de regresión logística multifactorial desarrollado en este estudio demostró ser una herramienta eficaz para predecir el éxito de los estudiantes de medicina de la ESPOCH en el EHEP. A diferencia de investigaciones que abordan dicho examen desde una perspectiva unidimensional, este enfoque integró variables académicas, psicométricas y de estilo de vida, ofreciendo una visión holística de los factores que realmente inciden en el desempeño.

El número de horas del estudio mostró un fuerte efecto positivo en la probabilidad de aprobar el examen, coincidiendo con los hallazgos de Expósito et al. (2025). Esta concordancia reafirma que la dedicación al estudio es un factor clave para el rendimiento en exámenes profesionales. Asimismo, la naturaleza negativa de las variables "TUTORÍAS" y "PREEXÁMENES" refleja una dinámica compleja, donde los estudiantes que buscan más apoyos tienden a presentar dificultades académicas previas, lo que sugiere que estas intervenciones actúan más como indicadores de vulnerabilidad que de protección, una situación también descrita por Rojas et al. (2025).

En el contexto regional, los resultados encuentran respaldo en la revisión sistemática ejecutada por Flores-Cohaila (2022) sobre factores asociados con el desempeño en el Examen Nacional de Medicina (ENAM) de Perú, destacando la multifactorialidad del rendimiento académico, incluyendo variables como el promedio de calificaciones, el entorno de prácticas profesionales, el estatus académico regular, y estrategias de aprendizaje autorregulado. La implicación práctica es que estudiantes y docentes inciden en los resultados mediante la mejora de hábitos de estudio y manejo de la ansiedad, lo que fortalece el valor de incluir variables psicométricas y conductuales en nuestro modelo.

De manera complementaria, el estudio de Agha et al. (2023) aportó evidencia sobre la importancia del bienestar subjetivo y habilidades académicas asociadas con el éxito en estudiantes de medicina. Aspectos como la confianza, la percepción de eficacia del instructor y la adaptación personal se vincularon significativamente con un alto promedio de calificaciones, respaldando así la inclusión de variables psicométricas como la autoeficacia y la motivación en nuestro análisis.

En cuanto a variables relacionadas con el estilo de vida, el estudio de Rodriguez et al. (2021) evidencia la alta prevalencia de somnolencia diurna excesiva y mala calidad del sueño entre estudiantes de medicina en Ecuador, aunque sin una asociación estadísticamente significativa con el rendimiento académico. Este hallazgo es consistente con nuestro resultado, donde la calidad del sueño no emergió como un predictor relevante en el modelo. Sin embargo, dada la frecuencia de estos trastornos del sueño en la población estudiada, sigue siendo importante promover intervenciones que mejoren la salud integral del estudiante, más allá del impacto inmediato en su desempeño académico.

Comparando con investigaciones recientes, como la de Barahona-Anguisaca et al. (2024) y Del Carpio-Mendoza (2024), que combinaron modelos estadísticos y aprendizaje automático, nuestro modelo de regresión logística presenta ventajas en la interpretabilidad y uso práctico para identificar variables significativas en la predicción del éxito en el EHEP. La integración de variables multifactoriales, que van más allá de las calificaciones tradicionales, representa un aporte clave para diseñar estrategias personalizadas de acompañamiento, intervenciones académicas y programas de apoyo que potencien las tasas de aprobación.

Por lo tanto, esta investigación robustece el conocimiento sobre los factores que inciden en el rendimiento en exámenes profesionales de alta relevancia social, valida la utilidad de la regresión logística como método predictivo y consolida la importancia de considerar un enfoque multifactorial que abarque ámbitos académicos, psicométricos y de estilo de vida para optimizar la preparación estudiantil.

5. Conclusiones

El estudio desarrolló un modelo de regresión logística capaz de predecir el éxito (aprobación o no) en el EHEP aplicado a estudiantes de medicina de la ESPOCH, integrando características académicas, psicométricas y de estilo de vida. Para ello, se utilizó una metodología basada en la DSR y el estándar CRISP-DM para la construcción de un artefacto validado y confiable.

Los resultados evidencian que la variable "HORAS" de estudio es el predictor más determinante para la aprobación, destacando la importancia del esfuerzo y la disciplina del estudiante en su desempeño académico. Por otro lado, se identificaron relaciones negativas significativas con las variables "TUTORÍA" y "PREEXÁMENES", lo que sugiere la necesidad de investigaciones futuras que profundicen en el entendimiento de estos fenómenos y en cómo influyen en el éxito en el EHEP.

El modelo final de regresión logística mostró un desempeño robusto, al alcanzar una exactitud del 80.36 %, un F1-score de 0.643 y un área bajo la curva (AUC) de 0.710 en la curva de Precisión-

Sensibilidad. Estos resultados confirman su capacidad para discriminar eficazmente entre clases desbalanceadas, convirtiéndose en una herramienta útil para la predicción y detección temprana de estudiantes con mayor riesgo de no aprobar.

Referencias

- Agha, S., Alzayed, A. A., Alfuraih, T. A., Alenazi, F. T., Alomair, M. I., & Masuadi, E. (2023). Association of the Academic Performance of Undergraduate Medical Students With Positive Well-Being, Intelligence, and Factors of Academic Success. *Cureus*, 15(12), e50077. <https://doi.org/10.7759/CUREUS.50077>
- Aguilar-Reyes, J. E., Mejía-Peñañiel, E. F., Morocho-Barrionuevo, T. P., & Velasco, G. M. (2025). Estudio del rendimiento académico mediante la comparación de modelos de regresión y árboles de clasificación. *Telos*, 27(1), 94-115. <https://n9.cl/ldnu0r>
- Álvarez-Raigoza, K. L., García-Peña, Á. A., Rodríguez, J. A., Gómez-Restrepo, C., & Cita, J. E. (2024). Síndrome de burnout en estudiantes de cardiología y cardiólogos en Colombia. *Revista Colombiana de Cardiología*, 31(4), 218-229. <https://doi.org/10.24875/RCCAR.23000113>
- Alyahyan, E., & Düşteğör, D. (2020). Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17(1), 1-21. <https://n9.cl/j4uic>
- Barahona-Anguisaca, D. M., Vega-Calvas, P. A., Moyota-Paguay, A. R., & Porrás-Ramírez, L. I. (2024). Análisis de modelos estadísticos para predecir el éxito académico en estudiantes universitarios. *MQRInvestigar*, 8(2), 2951-2969. <https://doi.org/10.56048/MQR20225.8.2.2024.2951-2969>
- CACES. (2024). *Examen de habilitación profesional*. <https://www.caces.gob.ec/examen-de-habilitacion-profesional/>
- Carhuaricra, I. N. (2023). *Factores relacionados al rendimiento académico del examen nacional de medicina en internos de una universidad privada de Lima, 2022* [Tesis de grado, Universidad Privada San Juan Bautista]. Repositorio institucional. <https://n9.cl/po4p2>
- Del Carpio-Mendoza, R. A. (2024). Predicción del rendimiento académico utilizando modelos de aprendizaje automático: una revisión sistemática de la literatura. *593 Digital Publisher CEIT*, 9(6), 1038-1054. <https://doi.org/10.33386/593dp.2024.6.2797>
- Escobar-Jiménez, C., & Torres-Rentería, S. (2021). Éxito educativo y condiciones socioeconómicas: los exámenes de habilitación para el ejercicio profesional de la medicina en Ecuador. *Revista Iberoamericana de Educación Superior*, 12(35), 132-149. <https://doi.org/10.22201/IISUE.20072872E.2021.35.1086>
- Expósito, A., Díaz, M. T., & Montero, I. L. (2025). Modelo de regresión logístico en el proceso de habilitación profesional. *La Ciencia al Servicio de la Salud y Nutrición*, 16(1), 74-80. <https://cssn.esPOCH.edu.ec/index.php/v3/article/view/419>
- Flores-Cohaila, J. A. (2022). Factors associated with medical students' scores on the National Licensing Exam in Peru: A systematic review. *Journal of Educational Evaluation for Health Professions*, 19. <https://doi.org/10.3352/JEEHP.2022.19.38>
- Gálvez-Gamboa, F., Pinochet-Quiroz, P., Lepe-Martínez, N., & Cabrera, H. (2024). Experiencia de aprendizaje autorregulado y su efecto en el rendimiento académico de universitarios de primer año. *Revista Ecuatoriana de Neurología*, 33(1), 47-54. <https://doi.org/10.46997/REVECUATNEUROL33100047>

- Garbanzo, G. M. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública. *Revista Educación*, 31(1), 43. <https://doi.org/10.15517/REVEDU.V31I1.1252>
- Gutiérrez-Monsalve, J. A., Garzón, J., & Segura-Cardona, A. M. (2021). Factores asociados al rendimiento académico en estudiantes universitarios. *Formación universitaria*, 14(1), 13-24. <https://doi.org/10.4067/S0718-50062021000100013>
- Merchán-Clavellino, A., Martínez-García, C., Salguero-Alcañiz, M. P., Paíno-Quesada, S., & Alameda-Bailén, J. R. (2019). Indicadores de calidad en la Educación Superior: análisis de los factores psicosociales de los estudiantes. *Revista de Psicología y Educación*, 14(1), 27. <https://doi.org/10.23923/RPYE2019.01.169>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas. *Revista Española de Cardiología*, 74(9), 790-799. <https://doi.org/10.1016/J.RECESP.2021.06.016>
- Pastor, J. C., Cachay, L. F., & Moreno, S. E. (2024). Factors influencing the academic performance of peruvian university students. *EduSol*, 24(88), 96-105. <https://n9.cl/vtqhy>
- Pérez, M., Mejía, O., Serrano, C., Suescún, S., Mogollón-Alaguna, O., & León, F. (2023). Analysis of academic performance from a binary logistic regression model. *Innovaciencia*, 11(1), 1-14. <https://doi.org/10.15649/2346075X.3423>
- Rodríguez, M., Ascuntar, N., González, P., & Fors, M. (2021). Excessive daytime somnolence in a sample of Ecuadorian undergraduate medical students and its relationship with academic performance. *Cogent Education*, 8(1). <https://doi.org/10.1080/2331186X.2020.1870800>
- Rojas, F., Figueroa, G. S., & Gallegos, N. (2025). Identificación temprana de estudiantes en riesgo de reprobación mediante la tutoría y su impacto académico. *Espacio I+D, Innovación más Desarrollo*, 14(40), 30-43. <https://n9.cl/1th5w>
- Sánchez-Cubo, F., Mondéjar-Jiménez, J., & García-Pozo, A. (2023). Evaluación del desajuste educativo en la industria hostelera española. *Investigaciones Turísticas*, (26), 235-250. <https://doi.org/10.14198/INTURI.23653>
- Suardiaz-Muro, M., Morante-Ruiz, M., Ortega-Moreno, M., Ruiz, M. A., Martín-Plasencia, P., & Vela-Bueno, A. (2020). Sueño y rendimiento académico en estudiantes universitarios: revisión sistemática. *Revista de Neurología*, 71(2), 43-53. <https://doi.org/10.33588/RN.7102.2020015>

Transparencia

Conflicto de interés

Los autores no reportamos conflicto de intereses.

Fuente de financiamiento

Los autores declaran que no hubo financiamiento para la realización de este estudio.

Contribución de autoría

Saul Yasaca Pucuna: Conceptualización, software, validación, análisis formal, investigación, gestión de datos, visualización, redacción - preparación del borrador original, redacción - revisión y edición, financiamiento, administración del proyecto, recursos, supervisión.

Juan Diego Erazo Rodríguez: Conceptualización, metodología, validación, análisis formal, investigación, visualización, redacción - preparación del borrador original, redacción - revisión y edición, financiamiento, recursos.

Los autores contribuyeron activamente en el análisis de los resultados, revisión y aprobación del manuscrito final.