

Modelo predictivo para identificar el síndrome de burnout en estudiantes de medicina

Predictive Model to Identify Burnout Syndrome in Medical Students

Sandra Catalina Correa Herrera¹

Especialista en evaluación clínica y tratamiento de trastornos emocionales y afectivos, miembro del grupo Armonía, Centro de Investigación y Atención Psicosocial HANAMI, Colombia
Correo electrónico: scatalinacorrea@gmail.com
ORCID: <https://orcid.org/0000-0003-4998-1228>

Luisa Fernanda Benítez²

Doctora en Salud Pública
Universidad Militar Nueva Granada (UMNG), Colombia
Correo electrónico: luisa.benitez@unimilitar.edu.co
ORCID: <https://orcid.org/0000-0002-4706-8526>

Recibido: 30/07/2025

Evaluado: 18/03/2026

Aprobado: 7/04/2026

- 1 CRediT: conceptualización, curación de datos, análisis formal, metodología, software y redacción borrador original.
- 2 CRediT: conceptualización, obtención de financiamiento, metodología, recursos, redacción borrador original, redacción, revisión y edición.
- * Para citar este artículo: Correa-Herrera, S. C. y Benítez, L. F. (2026). Modelo predictivo para identificar el síndrome de burnout en estudiantes de medicina. *Informes Psicológicos*, 26(1), 71-86. <https://doi.org/10.18566/infpsic.v26n1a05>

Resumen

El síndrome de *burnout* es un problema en estudiantes de medicina debido a la alta carga académica y emocional. El uso de técnicas de aprendizaje automático puede facilitar la identificación temprana y apoyar procesos de intervención. El objetivo fue desarrollar un modelo para predecir el síndrome de *burnout* en estudiantes, por medio de técnicas de *machine learning*. Se incluyó información de 81 estudiantes de medicina de Bogotá (Colombia) y se generaron 140 valores simulados, que formaron parte del conjunto total analizado. Se realizó una preselección de características numéricas relevantes, que fueron procesadas con el método SelectKBest. Se utilizó Random Forest optimizado mediante búsqueda de hiperparámetros. La evaluación empleó métricas de precisión, recall y F1-score. El modelo desarrollado alcanzó una precisión del 77 % en la clasificación de los niveles del síndrome de *burnout* en los participantes. En conclusión, el modelo desarrollado demostró ser efectivo para predecir el síndrome de *burnout*.

Palabras clave:
Aprendizaje automático, Estudiantes, Modelos predictivos, Salud mental,
Síndrome de burnout.

Abstract

Burnout syndrome represents a significant concern among medical students due to the high academic and emotional demands of their training. The use of machine learning techniques can facilitate early identification and support intervention processes. The objective of this study was to develop a model to predict burnout syndrome in students using machine learning techniques. Data from 81 medical students in Bogotá, Colombia were included, along with 140 simulated values incorporated into the total dataset analyzed. A preselection of relevant numerical features was conducted using the SelectKBest method. An optimized Random Forest model was implemented through hyperparameter tuning. Model performance was evaluated using precision, recall, and F1-score metrics. The developed model achieved 77% accuracy in classifying burnout syndrome levels among participants. In conclusion, the model demonstrated effectiveness in predicting burnout syndrome.

Keywords:
machine learning; students; predictive models; mental health; burnout syndrome.

Introducción

El síndrome de *burnout* fue inicialmente conceptualizado en contextos laborales y definido como una respuesta psicológica generada por sobrecarga laboral y elevado agotamiento (Gil-Monte, 2007), caracterizada por agotamiento emocional, despersonalización y baja realización personal (Maslach y Jackson, 1981). Aunque esta definición se refiere específicamente al ámbito laboral, investigaciones posteriores han extendido el estudio del *burnout* a contextos académicos y universitarios. A diferencia del *burnout* laboral, que en la CIE-11 se define como un fenómeno ocupacional asociado al estrés crónico en el lugar de trabajo, el *burnout* académico hace referencia al desgaste que experimentan los estudiantes frente a las demandas de estudio. El *burnout* académico mantiene la estructura tridimensional propuesta por el modelo de Maslach: (a) agotamiento emocional ante la carga académica; (b) cinismo o desafección hacia las tareas y responsabilidades académicas, y (c) baja eficacia académica, entendida como la percepción de ser poco competente o ineficaz como estudiante.

Se ha encontrado que este síndrome es particularmente frecuente, siendo los estudiantes de la carrera de medicina unos de los más afectados (Almutairi et al., 2022) debido a las exigencias en su contexto académico, pues, además de la presión por el rendimiento, también está la exposición a contextos clínicos sumamente demandantes y con alta carga de responsabilidad (Dyrbye et al., 2006). Se ha encontrado que entre el 45 % y el 60 % de los estudiantes de medicina presentan algún grado de síndrome de *burnout*, afectando la salud mental y física, el desempeño académico y las relaciones interpersonales del individuo (IsHak et al., 2013).

A pesar de la evidencia acumulada sobre la alta prevalencia del síndrome de *burnout* en estudiantes de medicina (Almutairi et al., 2022; IsHak et al., 2013), los estudios se han enfocado en describir su incidencia y en analizar factores asociados de tipo académico, emocional o institucional (Caballero et al., 2007; Rosales-Ricardo et al., 2021). Sin embargo, existe un vacío en la aplicación de herramientas tecnológicas, como los modelos predictivos basados en *machine learning*, para la detección temprana de este síndrome en contextos universitarios latinoamericanos. Esta falta de investigaciones limita la posibilidad de contar con metodologías que permitan identificar perfiles de riesgo de manera automatizada y precisa.

En los últimos años, los modelos predictivos que utilizan lenguajes de programación y de *machine learning* se han convertido en herramientas indispensables en los ámbitos de la psicología y la salud mental, dado que ofrecen medidas que dirigen a los profesionales de la salud a diagnósticos más eficientes y rápidos, que proporcionan una ayuda en el análisis y la detección temprana de diversos síndromes y trastornos psicológicos (Papini et al., 2023; Saito et al., 2022). Uno de los lenguajes de programación más utilizado es Python, debido a su versatilidad y a la amplia cantidad de librerías que facilitan el tratamiento de datos, incluso cuando estos son extensos (Müller y Guido, 2016; Pedregosa et al., 2011).

Esta investigación tuvo como objetivo crear un modelo predictivo en Python junto con técnicas del *machine learning*, para identificar los niveles del síndrome de *burnout* en estudiantes universitarios de medicina. Con esto se busca contribuir al desarrollo de métodos

de detección automática y facilitar la evaluación del nivel de riesgo de las personas, con el fin de optimizar el tiempo para alcanzar el diagnóstico y agilizar intervenciones psicológicas adecuadas por parte del profesional de salud.

Método

Diseño

Este estudio es observacional analítico de corte transversal y está orientado al desarrollo y evaluación de un modelo predictivo para clasificar niveles del síndrome de *burnout* en estudiantes de medicina mediante técnicas de *machine learning*. Este tipo de estudio se escogió debido a que permite analizar la relación entre variables en un momento específico del tiempo (Hernández-Sampieri et al., 2014). Nos permitió posteriormente crear en Python un modelo basado en datos. En ese sentido, el diseño se centró en la construcción del modelo y su validación interna con los datos disponibles.

Participantes

La muestra en esta investigación fue de 81 estudiantes de medicina de la Universidad Militar Nueva Granada (UMNG), en la ciudad de Bogotá (Colombia), quienes participaron voluntariamente. Los estudiantes tenían edades comprendidas entre los 18 y 23 años, con un promedio de 19,4 años. De ellos, 60 eran mujeres, 20 hombres y uno se identificó como persona de género fluido. De manera complementaria, el 98 % era estudiante y el 2 % restante estudiante y trabajador.

El tipo de muestreo empleado fue no probabilístico por conveniencia, dado que los participantes fueron seleccionados a partir de su accesibilidad y disposición para participar en el estudio, de acuerdo con los criterios de inclusión establecidos. Los datos de los participantes fueron recolectados por medio de formularios *online*, los cuales tenían preguntas acerca de sus hábitos de vida, carga académica, percepción del estrés y datos socioeconómicos, así como preguntas relacionadas con el síndrome de *burnout*.

Criterios de inclusión: se incluyeron los estudiantes que diligenciaron completamente el formulario, tras otorgar su consentimiento informado y aceptar participar.

Criterios de exclusión: se excluyeron los individuos cuyos registros contenían valores nulos o inconsistentes durante la depuración previa al análisis, así como aquellos que reportaron diagnósticos psicológicos o médicos graves que impidieran su participación segura en el estudio.

Consideraciones éticas

El presente estudio se clasifica como una investigación sin riesgo, de acuerdo con lo establecido en el artículo 11, literal a) de la Resolución 8430 de 1993 del Ministerio de Salud de Colombia, ya que empleó cuestionarios y análisis de datos sin intervención o modificación intencionada de variables biológicas, fisiológicas, psicológicas o sociales de los participantes. Asimismo, la participación fue voluntaria y todos los estudiantes fueron informados sobre los objetivos de la investigación, garantizándose el consentimiento informado previo a la diligencia del cuestionario. Se respetó en todo momento el principio de confidencialidad, asegurando que la información recopilada se tratara de forma agregada y sin posibilidad de identificar individualmente a los participantes.

Instrumentos

Maslach Burnout Inventory- Human Services Survey (MBI-HSS): validado en el contexto académico colombiano por Hederich-Martínez y Caballero-Domínguez (2016), se clasifican los niveles de cada dimensión del síndrome de *burnout*: agotamiento, cinismo y autoeficacia, con adecuados índices de consistencia interna para cada subescala, reportando un Alfa de Cronbach de .77 para agotamiento, .72 para cinismo y .82 para eficacia. Los niveles para cada dimensión se clasifican tomando como referencia los percentiles definidos en dicha validación, donde los valores obtenidos por cada estudiante se agrupan en los siguientes rangos: muy bajo (por debajo del percentil 5), bajo (entre el percentil 5 y 25), medio-bajo (percentil 25 a 50), medio-alto (percentil 50 a 75), alto (percentil 75 a 95), muy alto (por encima del percentil 95) (véanse los valores específicos en la Tabla 1). A partir de esto, se establece la presencia de síndrome de *burnout* cuando las dimensiones de agotamiento y cinismo se encuentran en un estado alto o muy alto, y la dimensión de autoeficacia se encuentra baja o muy baja (Hederich-Martínez y Caballero-Domínguez, 2016).

Tabla 1

Clasificación de niveles por dimensión del burnout según percentiles

Nivel	Agotamiento	Cinismo	Autoeficacia
Muy bajo	< 0.4	< 0.2	< 2.83
Bajo	0.5 – 1.2	0.3 – 0.5	2.83 – 3.83
Medio-bajo	1.3 – 2.0	0.6 – 1.24	3.84 – 4.5
Medio-alto	2.1 – 2.8	1.25 – 2.25	4.51 – 5.16
Alto	2.9 – 4.5	2.26 – 4.0	5.17 – 5.83
Muy alto	> 4.5	> 4.0	> 5.83

Nota. Tomado de validación del cuestionario Maslach Burnout Inventory-Student Survey (MBI-SS) en contexto académico colombiano. Fuente: Hederich-Martínez y Caballero-Domínguez (2016).

Procedimiento

Con el fin de facilitar la comprensión del proceso de clasificación del síndrome de *burnout*, se presenta un diagrama de decisión ilustrativo, basado en la combinación de las dimensiones de agotamiento, cinismo y autoeficacia (ver Figura 1). Este esquema permite visualizar, de manera simplificada, cómo la interacción entre estos factores conduce a la asignación de una categoría específica.

Figura 1
Diagrama de decisión ilustrativo para la clasificación del síndrome de burnout según criterios de agotamiento, cinismo y autoeficacia



Nota. El diagrama presenta ejemplos de rutas de decisión y no incluye todas las posibles combinaciones entre las variables evaluadas. Su propósito es ilustrar la lógica general de clasificación. Cada color en el diagrama representa una ruta específica dentro del proceso de decisión. Fuente: Elaboración propia.

Luego de esto, siguiendo criterios de la literatura, se realizó una clasificación propia en Excel a partir de los resultados obtenidos en los niveles de agotamiento, cinismo y autoeficacia de cada participante, con el fin de definir niveles de tendencia hacia la presencia del síndrome de *burnout* en la muestra, para poder caracterizar y tener una visión más específica de la información, tal como se describe en la Tabla 2.

Tabla 2

Criterios de clasificación de la tendencia hacia síndrome de burnout en cuatro niveles según sus síntomas

Clasificación	Criterio
Burnout	Tres síntomas presentes
Tendencia a burnout nivel 2	Dos síntomas críticos presentes
Tendencia a burnout nivel 1	Un síntoma crítico presente
No burnout	Ningún síntoma crítico presente

Nota. Se consideró síntoma crítico el agotamiento alto o muy alto, el cinismo alto o muy alto, y la autoeficacia baja o muy baja. Fuente: Elaboración propia.

Posteriormente, con el objetivo de incrementar la cantidad de información disponible para alimentar el modelo en desarrollo, se generó un proceso de aumento de datos mediante duplicación con modificaciones, creando una muestra de 140 datos adicionales mediante simulaciones realizadas usando la herramienta de Excel, basadas en los patrones identificados en la muestra de los participantes reales. Con lo anterior se buscó complementar la información disponible, generando casos adicionales similares a los observados en la muestra real. A continuación, se realizó un proceso de depuración de datos, eliminando valores nulos o vacíos, seguido de su normalización para garantizar la calidad del análisis.

Dentro del modelo también se utilizó una técnica de selección de características, con el fin de disminuir la dimensionalidad y mejorar su rendimiento, lo que permitió identificar las variables más relevantes para predecir el síndrome de *burnout*. En esta investigación se empleó el método de SelectKBest, con el que se pueden filtrar variables relevantes en función de su relación con la variable objetivo. Para esto, se utilizó como función de puntuación el estadístico F de ANOVA (f_{classif}), que encuentra y evalúa la dependencia y relación entre cada variable, y el objeto que se tiene (en este caso es el síndrome de *burnout*). Lo anterior permitió a su vez reducir el riesgo de sobreajuste (*overfitting*), dado que permite limitar el modelo a las variables con mayor poder discriminativo (Guyon y Elisseeff, 2003).

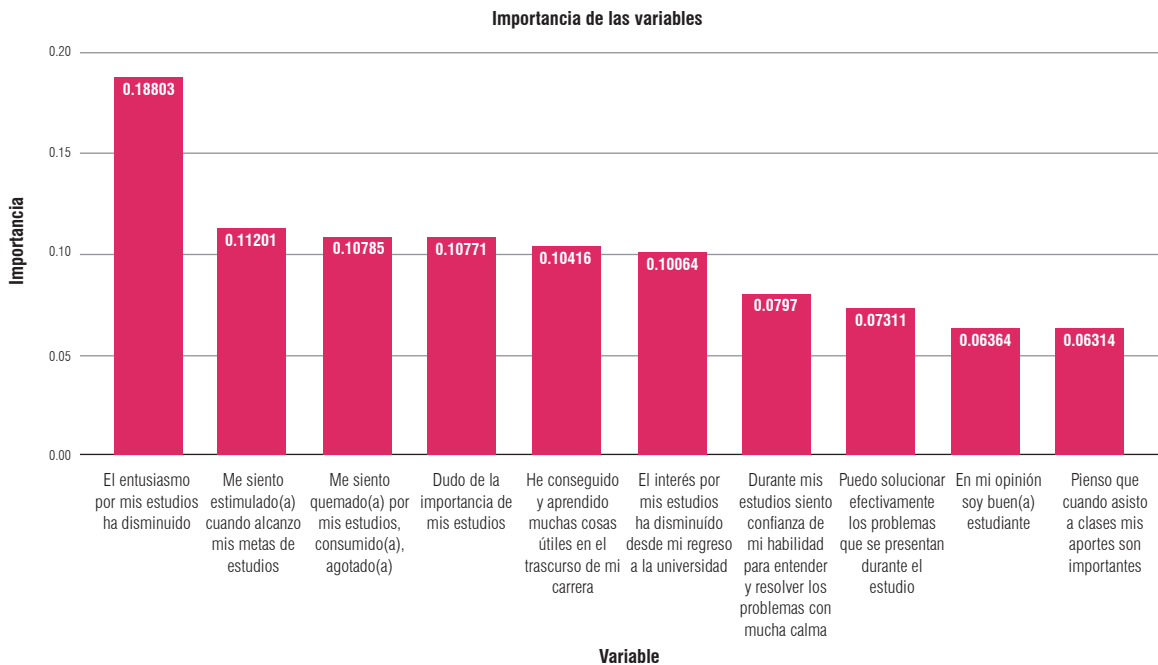
Luego, se dividió en dos grupos el conjunto de datos procesado: el primero, compuesto por el 80 % de los datos, se utilizó para entrenar el modelo en Python; mientras que el segundo grupo, con el 20 % restante, se empleó como conjunto de prueba. Uno de los algoritmos utilizado para la creación del modelo fue Random Forest, dado que tiene la capacidad de manejar bases de datos con múltiples variables, además de que puede realizar excelentes generalizaciones con los datos disponibles (Breiman, 2001).

Se utilizó GridSearchCV, una herramienta de validación cruzada, diseñada para identificar las combinaciones óptimas de parámetros ajustables. Estos parámetros, también conocidos como hiperparámetros, no se derivan directamente de los datos proporcionados al modelo, sino que deben definirse antes del proceso de entrenamiento, influyendo significativamente en su capacidad de generalización y precisión (Bergstra y Bengio, 2012).

En el caso del modelo Random Forest, los hiperparámetros configurados para regular su tamaño, complejidad y profundidad fueron `n_estimators`, `max_depth`, `min_samples_split` y `min_samples_leaf`. Lo anterior permite encontrar una configuración del modelo que equilibre entre precisión y la capacidad de generalización, tratando de evitar que el modelo se ajuste en exceso a los datos de entrenamiento.

Además del entrenamiento del modelo, se realizó un análisis de la importancia de las características, a partir del clasificador creado y optimizado por Random Forest. Esto permite identificar y visualizar las variables que contribuyen de manera más significativa en la predicción del síndrome de *burnout* en el modelo. Para ello, se empleó la métrica conocida como disminución media de la impureza (*Mean Decrease in Impurity [MDI]*), que evalúa la mejora en la pureza de los nodos al dividir los datos mediante cada variable seleccionada durante la construcción de los árboles (Breiman, 2001). Las variables fueron previamente seleccionadas mediante el método `SelectKBest` y se graficaron en orden descendente, según su relevancia, para facilitar la comprensión de los factores más determinantes en la clasificación (ver Figura 2).

Figura 2
Importancia de variables seleccionadas por orden para la predicción del síndrome de burnout



Fuente: Elaboración propia.

Resultados

Los resultados obtenidos muestran que el modelo tiene una precisión general del 77 % al clasificar las personas según su nivel de síndrome de *burnout*, es decir, clasifica correctamente aproximadamente 77 de cada 100 casos evaluados. El modelo tuvo un mejor desempeño en la detección de casos en “Tendencia a *burnout* nivel 2”, con un recall del 94 % y una precisión de 83 %, lo cual es de gran importancia a la hora de tomar decisiones clínicas (el recall hace referencia a la capacidad del modelo para identificar correctamente a los estudiantes que realmente pertenecen a esta categoría, mientras que la precisión indica qué proporción de las predicciones para esta categoría son correctas). Este desempeño podría estar influido por la mayor cantidad de datos disponibles en esta categoría en comparación con las demás, lo que facilita un aprendizaje más efectivo del modelo en esta área específica.

De igual forma, se identifican limitaciones en las categorías con menor representación en la base de datos actual, como ocurre con la clasificación de “No *burnout*” y “*Burnout*”. En estas categorías se reporta una precisión del 100 %, lo que indica que, cuando el modelo clasifica a una persona como “No *burnout*”, es altamente probable que sea correcto. Sin embargo, el recall es del 33 %, lo que significa que el modelo no logra identificar la mayoría de los casos de “No *burnout*” y de “*Burnout*”, cometiendo errores al excluir algunos que deberían pertenecer a esta clasificación. El F1-score para estas clasificaciones fue de 0.50, reflejando un balance moderado entre precisión y sensibilidad, aunque limitado por el bajo recall.

El macro promedio (*macro avg*), que calcula el promedio simple de precisión, recall y F1-score para todas las clases sin ponderar el tamaño de cada clase, reportó valores de 0.79 para precisión, 0.65 para recall y 0.59 para F1-score. Estos resultados reflejan que, si bien la precisión es alta, la recuperación (recall) y el equilibrio entre precisión y recuperación (F1-score) aún presentan áreas de mejora (ver Tabla 3).

Tabla 3
Reporte de clasificación del modelo

Clase	Precisión	Recall	F1-score	Soporte (n)
Burnout	1.00	0.33	0.50	9
No <i>burnout</i>	1.00	0.33	0.50	3
Tendencia a <i>burnout</i> nivel 1	0.33	1.00	0.50	2
Tendencia a <i>burnout</i> nivel 2	0.83	0.94	0.88	31

Fuente: Elaboración propia.

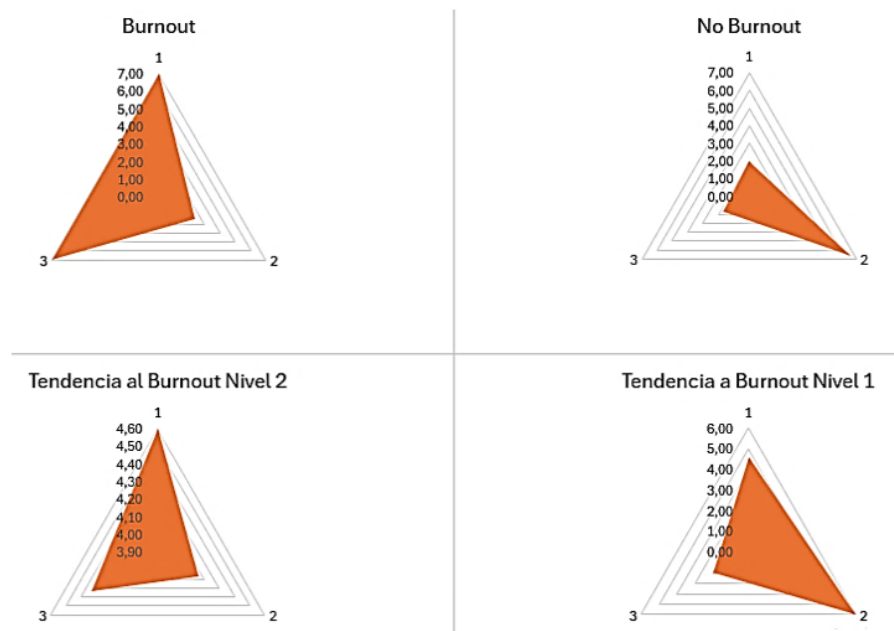
El promedio ponderado (*weighted average*) considera tanto el desempeño individual en cada clase como la cantidad de ejemplos en el conjunto de datos analizado. Esto implica que las clases con mayor número de casos (mayor soporte) tienen un impacto más significativo en el cálculo final que aquellas con pocos ejemplos. Bajo esta premisa, el modelo alcanzó una precisión ponderada de 0.85, lo que significa que, en promedio y considerando el peso de cada clase, el 85 % de las predicciones fueron correctas. Por otro lado, el recall ponderado fue de 0.78, indicando que, en general, el modelo identificó correctamente el 78 % de los casos reales. Finalmente, el F1-score ponderado fue de 0.76, evidenciando un balance razonable entre precisión y sensibilidad al tener en cuenta la distribución de ejemplos en cada clase.

Por otro lado, el análisis de importancia de las características empleado luego del Random Forest fue una herramienta que permitió observar cuáles variables tuvieron un mayor peso en el modelo al predecir los niveles del síndrome de *burnout* (Figura 2). De las variables seleccionadas en el método SelectkBest, se observó que las relacionadas con el entusiasmo por los estudios, la confianza al resolver problemas académicos y el sentimiento de logro en las metas académicas fueron las que más influyeron a lo largo del entrenamiento del modelo.

Para facilitar la interpretación de los resultados del modelo, la Figura 3 presenta perfiles de estudiantes pertenecientes a cada categoría de clasificación. Estos perfiles permiten visualizar la configuración diferencial de las dimensiones de agotamiento, cinismo y autoeficacia en cada nivel de riesgo de *burnout*.

Figura 3

Gráficos de radar de perfiles representativos de riesgo de burnout en estudiantes según clasificación del modelo



Nota. Cada gráfico corresponde a un participante aleatorio de cada categoría de clasificación. En los diagramas de radar, el eje 1 representa el agotamiento, el eje 2 la autoeficacia y el eje 3 el cinismo. Fuente: Elaboración propia.

Discusión

La presente investigación ofrece una contribución respecto a la literatura sobre el uso de algoritmos de *machine learning* para la detección del síndrome de *burnout* en poblaciones universitarias, especialmente en carreras que tienen altos porcentajes de estrés y demanda en los estudiantes (Grządzielewska, 2021; Kuadey et al., 2024; Pereira et al., 2025; Tapio, 2025). Debido a su precisión del 77 %, este modelo predictivo demostró potencial para ser una herramienta para asistir a los trabajadores de la salud en evaluaciones psicológicas. El desempeño en la clasificación de “Tendencia a *burnout* nivel 2” destaca su potencial utilidad para identificar casos que aún no cumplen los criterios diagnósticos de *burnout*, pero que presentan un alto riesgo en contextos clínicos y reales, lo que representa una valiosa herramienta para los profesionales de la salud al facilitar intervenciones oportunas y eficaces (Yi et al., 2024).

El modelo Random Forest empleado en esta investigación demuestra consistencia con estudios previos que resaltan su eficacia en tareas de clasificación en contextos clínicos y educativos (Baniadamdizaj y Baniadamdizaj, 2023; Fernández-Delgado et al., 2014). Además, su capacidad para manejar bases de datos extensas y con múltiples características lo posiciona como una opción robusta frente a otros algoritmos, como las máquinas de vectores de soporte (SVM, por sus siglas en inglés) (Gibert et al., 2020).

Asimismo, el uso de datos simulados fue importante en el desarrollo de este estudio, pues permitió reducir el desbalance en las clasificaciones iniciales del conjunto de datos. Este procedimiento ha sido validado en investigaciones que informan de la importancia del *data augmentation* en problemas de clasificación de clases minoritarias, especialmente en campos de la salud donde la obtención de datos puede ser en ocasiones limitada por razones éticas o logísticas (Chawla et al., 2002; Sufi, 2024). Pese a los conocimientos previos, esta investigación presenta una limitación que persistió pese a la presencia de datos simulados, observable en la categoría “No *burnout*”, que alcanzó un recall de 33 %. Esta situación es consistente con estudios que indican que los modelos tienden a tener sesgos hacia clases mayoritarias (Feher et al., 2024; Haixiang et al., 2017).

Los resultados actuales muestran importantes limitaciones en la clasificación de categorías con baja representación, tanto en los datos simulados como en la muestra obtenida mediante formularios en línea, como es el caso de las clasificaciones “No *burnout*” y “Tendencia a *burnout* nivel 1”. Este desequilibrio en las clasificaciones genera un desbalance en los datos ingresados al modelo de aprendizaje. Para abordar esta limitación, se plantea la recolección de datos reales provenientes de un mayor número de estudiantes de medicina. Paralelamente, se considera aumentar la generación de datos simulados, con el objetivo de ampliar la diversidad y el volumen del conjunto de datos disponible. Adicionalmente, se plantea aplicar técnicas de sobremuestreo, como SMOTE, y explorar otros modelos, como XGBoost o redes neuronales, que podrían capturar patrones más complejos y proporcionar información adicional. Además, es importante considerar como limitación el tamaño de la muestra, compuesta por 81 estudiantes provenientes de una única institución educativa. Esto puede restringir la capacidad de generalización de los resultados a

otras poblaciones universitarias con contextos académicos y sociodemográficos distintos. Asimismo, el modelo fue evaluado únicamente mediante validación interna, por lo que no se cuenta con evidencia sobre su desempeño en cohortes externas independientes.

El análisis de la importancia de las variables realizado en esta investigación respalda hallazgos previos que identifican los principales factores determinantes del síndrome de *burnout* en poblaciones universitarias (Caballero et al., 2007). Entre estos, destacan la disminución del entusiasmo por los estudios, una baja percepción de autoeficacia y el agotamiento emocional, lo que subraya el papel crucial de las variables emocionales y motivacionales en el desarrollo del *burnout* académico (Rosales-Ricardo et al., 2021). De acuerdo con este análisis, las variables “El entusiasmo por mis estudios ha disminuido”, “Me siento estimulado(a) cuando alcanzo mis metas de estudios” y “Me siento quemado(a) por mis estudios” fueron las que más contribuyeron a la predicción del *burnout*. Esto confirma la relevancia de estos ítems, lo que a su vez ofrece una oportunidad para mejorar este instrumento de evaluación en versiones posteriores. La selección de las preguntas más relevantes permitiría construir versiones más breves y eficientes del instrumento. Asimismo, facilitaría su uso en contextos educativos y clínicos, sin llegar a afectar su validez diagnóstica.

En esta misma línea, Hederich-Martínez y Caballero-Domínguez (2016), en la validación colombiana del MBI-SS, eliminaron el ítem “He aprendido muchas cosas interesantes durante mi carrera”, debido a su baja comunalidad y escasa carga factorial. Esta enmienda mejoró la consistencia interna de la escala de autoeficacia, evidenciando cómo el análisis de los ítems puede contribuir a la mejora de los cuestionarios. En ese sentido, la detección de las variables más importantes mediante técnicas de *machine learning* también puede apoyar futuras revisiones psicométricas del MBI-SS u otros instrumentos, facilitando el desarrollo de versiones optimizadas.

Los resultados de este trabajo son consistentes con estudios previos, aunque más modestos en términos de exactitud global. Por ejemplo, Baniadamdizaj y Baniadamdizaj (2023) emplearon nueve algoritmos de clasificación para predecir el nivel de *burnout* en docentes a partir del Maslach Burnout Inventory (MBI), encontrando que modelos como el análisis discriminante lineal y cuadrático alcanzaban precisiones superiores al 90 % en algunas dimensiones del síndrome. En comparación, el modelo Random Forest de este estudio obtuvo una precisión global del 77 % en cuatro categorías de *burnout* académico, con un mejor desempeño en la clasificación de casos en “Tendencia a *burnout* nivel 2”. Esta diferencia puede explicarse, en parte, por el menor tamaño muestral, la mayor complejidad del problema de clasificación (cuatro niveles en lugar de tres o un enfoque binario) y el uso de datos simulados para mitigar el desbalance de clases. De manera similar, Tapio (2025) modeló el *burnout* académico en estudiantes universitarios mediante árboles de decisión, identificando variables como el estrés percibido, las estrategias de afrontamiento y los patrones de sueño como predictores relevantes, y destacando la necesidad de explorar modelos más avanzados como Random Forest para mejorar la precisión.

De acuerdo con lo anterior, el modelo predictivo desarrollado puede ser una herramienta valiosa para la identificación del síndrome de *burnout* en estudiantes de medicina, particularmente en niveles severos. No obstante, su efectividad en clases menos representadas y

su capacidad de generalización a otras poblaciones aún representan desafíos importantes. En este sentido, se hace necesario incorporar procesos de validación externa que permitan evaluar la robustez y estabilidad del modelo en diversos contextos educativos. Se recomienda aumentar la diversidad y el tamaño de las muestras en futuras investigaciones, así como considerar el desarrollo de sistemas híbridos que integren datos cuantitativos y cualitativos, para mejorar la sensibilidad diagnóstica y la aplicabilidad clínica del modelo.

Conclusiones

En la presente investigación se desarrolló un modelo predictivo basado en el método Random Forest para clasificar la presencia o ausencia del síndrome de *burnout* en los estudiantes de medicina, así como dos niveles intermedios previos a su aparición, que pueden servir como indicadores de riesgo. El modelo presentó una precisión global del 77 %, con un mejor rendimiento en la clasificación de estudiantes en “Tendencia a *burnout* nivel 2”, en donde se obtuvo un recall del 94 % y F1-score de 88 %.

Adicionalmente, se mejoró la representación de las clases en el modelo gracias a la incorporación de datos simulados, lo cual fortaleció el entrenamiento y compensó el desbalance inicial de la base de datos real. La estrategia de selección de características mediante el método SelectKBest permitió identificar las variables más relevantes para el diagnóstico, mejorando el desempeño del modelo y facilitando su interpretación clínica.

Asimismo, el análisis de la importancia de las variables también permitió identificar factores clave relacionados con el entusiasmo por los estudios, la percepción de autoeficacia y el agotamiento emocional, reforzando los hallazgos previos sobre los principales determinantes del síndrome de *burnout* en contextos universitarios (Caballero et al., 2007).

Financiación: producto derivado del proyecto INV-MED-3951, financiado por la Vicerrectoría de Investigaciones de la UMNG.

Declaración de conflicto de intereses: los autores declaran que no existe ningún conflicto de intereses en relación con este estudio.

Agradecimientos: agradecemos a la UMNG por apoyar nuestras investigaciones, especialmente al Fondo de Investigaciones de la universidad, a la Vicerrectoría de Investigaciones y la Facultad de Medicina y Ciencias de la Salud. Agradecemos también al Centro de Investigación y Atención Psicosocial HANAMI por su apoyo en la ejecución de este proyecto, en especial a Laura Daniela Ardila, por sus ideas y colaboración para construir este artículo.

Referencias

- Almutairi, H., Alsubaie, A., Abduljawad, S., Alshatti, A., Fekih-Romdhane, F., Husni, M. y Jahrami, H. (2022). Prevalence of burnout in medical students: A systematic review and meta-analysis. *International Journal of Social Psychiatry*, 68(6), 1157-1170. <https://doi.org/10.1177/00207640221106691>
- Baniadamdzaj, S. y Baniadamdzaj, S. (2023). Prediction of Iranian EFL teachers' burnout level using machine learning algorithms and Maslach Burnout Inventory. *Iran Journal of Computer Science*, 6(1), 1-12. <https://doi.org/10.1007/s42044-022-00112-x>
- Bergstra, J. y Bengio, Y. (2012). Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, 13(1), 281-305. <https://dl.acm.org/doi/abs/10.5555/2188385.2188395>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32. <https://link.springer.com/article/10.1023/a:1010933404324>
- Caballero, C. C., Abello, R. y Palacio, J. (2007). Relación del burnout y el rendimiento académico con la satisfacción frente a los estudios en estudiantes universitarios. *Avances en Psicología Latinoamericana*, 25(2), 98-111. <https://dialnet.unirioja.es/servlet/articulo?codigo=2542727>
- Chawla, N. V., Bowyer, K. W., Hall, L. O. y Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321-357. <https://doi.org/10.1613/jair.953>
- Dyrbye, L. N., Thomas, M. R. y Shanafelt, T. D. (2006). Systematic review of depression, anxiety, and other indicators of psychological distress among us and Canadian medical students. *Academic Medicine*, 81(4), 354-373. <https://academic.oup.com/academicmedicine/article-abstract/81/4/354/8354952>
- Feher, G., Kapus, K., Tibold, A., Banko, Z., Berke, G., Gacs, B., Varadi, I., Nyulas, R. y Matuz, A. (2024). Mental Issues, internet addiction and quality of life predict burnout among hungarian teachers: A machine learning analysis. *BMC Public Health*, 24(1), 2322. <https://doi.org/10.1186/s12889-024-19797-9>
- Fernández-Delgado, M., Cernadas, E., Barro, S. y Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, 15(1), 3133-3181. https://www.jmlr.org/papers/volume15/delgado14a/delgado14a.pdf?source=post_page-----
- Gibert, D., Mateu, C. y Planes, J. (2020). The rise of machine learning for detection and classification of malware: Research developments, trends and challenges. *Journal of Network and Computer Applications*, 153, 102526. <https://www.sciencedirect.com/science/article/pii/S1084804519303868>
- Gil-Monte, P. R. (2007). El síndrome de quemarse por el trabajo (burnout) como fenómeno transcultural. *Informació Psicològica*, (91-92), 4-11. <https://www.informaciopsicologica.info/revista/article/view/259>
- Grządzielewska, M. (2021). Using machine learning in burnout prediction: A survey. *Child and Adolescent Social Work Journal*, 38(2), 175-180. <https://link.springer.com/article/10.1007/s10560-020-00733-w>

- Guyon, I. y Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157-1182. <https://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf>
- Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H. y Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, 73, 220-239. <https://doi.org/10.1016/j.eswa.2016.12.035>
- Hederich-Martínez, C. y Caballero-Domínguez, C. C. (2016). Validación del cuestionario Maslach Burnout Inventory-Student Survey (MBI-SS) en contexto académico colombiano. *Revista CES Psicología*, 9(1), 1-15. <https://revistas.ces.edu.co/index.php/psicologia/article/view/3511>
- Hernández-Sampieri, R., Fernández, C. y Baptista, P. (2014). *Metodología de la investigación* (6ª ed.). McGraw-Hill.
- IsHak, W., Nikraves, R., Lederer, S., Perry, R., Ogunyemi, D. y Bernstein, C. (2013). Burnout in Medical Students: A Systematic Review. *The Clinical Teacher*, 10(4), 242-245. <https://doi.org/10.1111/tct.12014>
- Kuadey, N. A., Ankora, C., Tahiru, F., Bensah, L., Agbesi, C. C. M. y Bolatimi, S. O. (2024). Using Machine Learning Algorithms to Examine the Impact of Technostress Creators on Student Learning Burnout and Perceived Academic Performance. *International Journal of Information Technology*, 16(4), 2467-2482. <https://link.springer.com/article/10.1007/s41870-023-01655-3>
- Maslach, C. y Jackson, S. E. (1981). The Measurement of Experienced Burnout. *Journal of Organizational Behavior*, 2(2), 99-113. <https://doi.org/10.1002/job.4030020205>
- Müller, A. C. y Guido, S. (2016). *Introduction to machine learning with Python: A guide for data scientists*. O'Reilly Media, Inc. [https://books.google.com.co/books?hl=es&lr=&id=1-4IDQAAQBAJ&oi=fnd&pg=PP1&dq=M%C3%BCller,+A.+C.,+y+Guido,+S.++\(2016\).+Introduction+to+machine+learning+with+Python:+a+guide+for+data+scientists.+O%27Reilly+Media,+Inc.+&ots=29ITILNM0X&sig=Ka0KsbSjSKRp6O11eNWqGqh_q-k&redir_esc=y#v=onepage&q&f=false](https://books.google.com.co/books?hl=es&lr=&id=1-4IDQAAQBAJ&oi=fnd&pg=PP1&dq=M%C3%BCller,+A.+C.,+y+Guido,+S.++(2016).+Introduction+to+machine+learning+with+Python:+a+guide+for+data+scientists.+O%27Reilly+Media,+Inc.+&ots=29ITILNM0X&sig=Ka0KsbSjSKRp6O11eNWqGqh_q-k&redir_esc=y#v=onepage&q&f=false)
- Papini, S., Iturralde, E., Lu, Y., Greene, J. D., Barreda, F., Sterling, S. A. y Liu, V. X. (2023). Development and Validation of a machine learning model using electronic health records to predict trauma-and stressor-related psychiatric disorders after hospitalization with sepsis. *Translational Psychiatry*, 13(1), 400. <https://www.nature.com/articles/s41398-023-02699-6>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. y Duchesnay, É. (2011). Scikit-Learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?source=post_page
- Pereira, M. G., Santos, M., Magalhães, R., Rodrigues, C., Araújo, O. y Durães, D. (2025). Burnout risk profiles in psychology students: An exploratory study with machine learning. *Behavioral Sciences*, 15(4), 505. <https://doi.org/10.3390/bs15040505>
- Rosales-Ricardo, Y., Rizzo-Chunga, F., Mocha-Bonilla, J. y Ferreira, J. P. (2021). Prevalence of burnout syndrome in university students: A systematic review. *Salud Mental*, 44(2), 91-102. <https://dialnet.unirioja.es/servlet/articulo?codigo=7903549>
- Saito, T., Suzuki, H. y Kishi, A. (2022). Predictive modeling of mental illness onset using wearable devices and medical examination data: Machine learning approach. *Frontiers in Digital Health*, 4, 861808. <https://doi.org/10.3389/fdgth.2022.861808>

- Sufi, F. (2024). Generative Pre-Trained Transformer (GPT) in research: A systematic review on data augmentation. *Information*, 15(2), 99. <https://doi.org/10.3390/info15020099>
- Tapio, R. P. (2025). Predicting burnout in college students: A machine learning approach using decision tree and psychometric data. *Asian Journal of Probability and Statistics*, 27(5), 50-60. <https://doi.org/10.9734/ajpas/2025/v27i5754>
- Yi, L., Shuai, T., Zhou, J., Cheng, L., Jiménez-Herrera, M. F. y Tian, X. (2024). Development and validation of a machine learning-based predictive model for compassion fatigue in Chinese nursing interns: A cross-sectional study utilizing latent profile analysis. *BMC Medical Education*, 24(1), 1495. <https://doi.org/10.1186/s12909-024-06505-9>