

Modelo predictivo multivariable de afectación psicosocial en estudiantes universitarios: identificación temprana mediante factores académicos, familiares y sociales

Benjamín Roldan Polo-Escobar
<https://orcid.org/0000-0001-5056-9957>
benjamin.polo@untrm.edu.pe
Universidad Nacional Toribio Rodríguez de
Mendoza
Chachapoyas, Perú

Renzo Enrique Polo-Moreano
<https://orcid.org/0009-0009-8747-8974>
20193257@aloe.ulima.edu.pe
Universidad de Lima
Facultad de Ciencias Empresariales y Económicas
Ciudad de Lima, Perú

Rodolfo Cornejo
<https://orcid.org/0000-0001-9325-6512>
recpesq@gmail.com
Universidad Nacional del Callao
Facultad de Ingeniería Pesquera y Alimentos
Instituto del Mar del Perú (IMARPE)
Callao, Perú

*Autor de correspondencia: benjamin.polo@untrm.edu.pe

Recibido: (02/05/2026), Aceptado: (01/09/2026)

Resumen. Este estudio desarrolló un modelo predictivo multivariable para identificar tempranamente afectación psicosocial en estudiantes universitarios. Se analizaron datos de 440 estudiantes mediante regresión logística, Random Forest y Gradient Boosting, integrando factores académicos, familiares y sociales. Los modelos alcanzaron valores de AUC cercanos a 0,80, con mejor discriminación para Random Forest (AUC = 0,805). Los factores sociales mostraron la mayor capacidad predictiva individual, aunque la integración de los tres dominios mejoró la representación del riesgo. Los resultados respaldan el uso de estos modelos como herramientas de apoyo para detección y evaluación temprana.

Palabras clave: afectación psicosocial, estudiantes universitarios, aprendizaje automático, predicción multivariable, factores de riesgo.

Multivariable Predictive Model of Psychosocial Impairment in University Students: Early Identification Using Academic, Family, and Social Factors

Abstract. This study developed a multivariable predictive model for the early identification of psychosocial impairment among university students. Data from 440 students were analyzed using logistic regression, Random Forest, and Gradient Boosting, integrating academic, family, and social factors. The models achieved AUC values close to 0.80, with Random Forest showing the highest discriminative performance (AUC = 0.805). Social factors demonstrated the strongest individual predictive capacity, although the integration of all three domains improved the representation of risk. The findings support the use of these models as decision-support tools for early detection and subsequent assessment.

Keywords: psychosocial impairment, university students, machine learning, multivariable prediction, risk factors.



I. INTRODUCCIÓN

La experiencia universitaria combina exigencias académicas, adaptación social, expectativas familiares y condiciones económicas que pueden converger en formas de malestar psicológico con repercusiones sobre el rendimiento, la permanencia y el bienestar estudiantil. La disponibilidad de datos abiertos sobre factores de riesgo psicosocial y resultados derivados de la Depression Anxiety Stress Scales-21 (DASS-21) en estudiantes de educación superior de Bangladesh permite estudiar estas interacciones mediante análisis secundarios reproducibles [1]. En este contexto, investigaciones recientes han documentado una presencia relevante de depresión, ansiedad y estrés y su asociación con características académicas, personales y sociales [2], [3], [4]. La afectación psicosocial se entiende aquí como un estado de malestar emocional expresado mediante síntomas de estrés, ansiedad y depresión, sin asumir un diagnóstico clínico. Esta delimitación es consistente con el uso de la DASS-21 en población no clínica y con sus propiedades psicométricas en estudiantes universitarios [5], [6], [7], [8], [9].

La evidencia indica que la afectación psicosocial responde a la combinación de demandas y recursos distribuidos en distintos ámbitos. Una menor funcionalidad familiar y un apoyo social insuficiente se han asociado con mayores niveles de ansiedad y depresión [10], mientras que perfiles con mayor estrés académico y síntomas depresivos tienden a coexistir con menor apoyo percibido [11]. En Bangladesh también se han observado relaciones entre apoyo social, estrés, ansiedad y depresión [12]. A ello se suman factores como la mala calidad del sueño y las experiencias de ciberacoso, relacionados con resultados adversos de salud mental en universitarios [13], [14]. En conjunto, estos hallazgos justifican analizar de manera integrada factores académicos, familiares y sociales, evitando enfoques fragmentados que pueden subestimar sus interacciones.

Aunque buena parte de la literatura se ha centrado en estimaciones de prevalencia y análisis asociativos, los enfoques predictivos permiten identificar combinaciones de variables capaces de discriminar tempranamente perfiles de mayor afectación. Estudios recientes han aplicado regresión y algoritmos de aprendizaje automático para predecir depresión, ansiedad y estrés en universitarios, incluidos análisis realizados en Bangladesh [3], [15]. Las revisiones disponibles confirman el crecimiento de estas aproximaciones, pero también advierten sobre problemas de sobreajuste, selección inadecuada de variables e interpretabilidad limitada [16], [17]. En consecuencia, esta investigación desarrolla y evalúa un modelo predictivo multivariable para la identificación temprana de afectación psicosocial en estudiantes universitarios, integrando factores académicos, familiares y sociales y priorizando tanto el desempeño predictivo como la interpretación de los factores con mayor contribución al riesgo estimado.

II. MARCO TEÓRICO

A. *Afectación psicosocial y malestar emocional en estudiantes universitarios*

La afectación psicosocial en el ámbito universitario puede conceptualizarse como la expresión de un desequilibrio entre las demandas experimentadas y los recursos personales, familiares y sociales disponibles para afrontarlas. Su manifestación no es necesariamente homogénea: algunos estudiantes pueden presentar principalmente tensión y sobrecarga, mientras que otros muestran síntomas ansiosos o depresivos. Esta heterogeneidad hace necesario trabajar con indicadores capaces de representar diferentes dimensiones del malestar sin reducir el fenómeno a una sola categoría diagnóstica. La DASS-21 ofrece una estructura breve para evaluar depresión, ansiedad y estrés y ha mostrado niveles adecuados de consistencia y validez en muestras universitarias y no clínicas [5], [6], [7], [8], [9]. En este marco, sus dimensiones constituyen indicadores de afectación emocional útiles para construir una variable de resultado orientada a detección temprana, siempre que los puntos de corte se empleen con finalidad de tamizaje y no como sustitutos de evaluación clínica.

El carácter multidimensional del fenómeno implica además que síntomas semejantes pueden surgir bajo configuraciones de riesgo diferentes. Dos estudiantes con niveles equivalentes de estrés pueden diferir sustancialmente en carga académica, apoyo familiar, dificultades económicas o experiencias de conflicto interpersonal. Por ello, un enfoque centrado únicamente en la intensidad del síntoma aporta información limitada para la prevención. La identificación temprana requiere reconocer qué combinación de exposiciones antecede o acompaña con mayor frecuencia los niveles

relevantes de afectación y distinguir, dentro de esa combinación, factores potencialmente modificables que puedan orientar intervenciones universitarias.

B. Factores académicos y condiciones asociadas

Las demandas académicas constituyen una fuente central de presión durante la formación superior. La carga de trabajo, las evaluaciones, el temor al fracaso, la competencia entre pares y la incertidumbre respecto del desempeño pueden incrementar la percepción de exigencia y reducir la capacidad de recuperación cuando se acumulan con otras tensiones. Estudios realizados en Bangladesh y otros contextos universitarios han identificado asociaciones entre variables académicas y síntomas de depresión, ansiedad o estrés [2], [3], [4], [18],[19], mientras que revisiones sobre estrés académico respaldan su naturaleza multifactorial [20]. Estas relaciones no deben interpretarse como deterministas: una demanda académica elevada puede ser tolerable cuando existen recursos suficientes, pero adquirir mayor peso cuando coincide con dificultades familiares, económicas o sociales.

Dentro de las condiciones vinculadas con la vida académica, el sueño ocupa una posición particularmente relevante porque refleja tanto hábitos individuales como presiones del entorno. La evidencia sintetizada en universitarios muestra una alta frecuencia de mala calidad del sueño y asociaciones entre estrés y deterioro del descanso [13]. En términos predictivos, variables relacionadas con sueño, presión por evaluaciones, desempeño y experiencias de fracaso pueden actuar como señales de vulnerabilidad, especialmente cuando su información se integra con indicadores de apoyo, convivencia y exposición a conflictos.

Factores familiares y sociales

El entorno familiar puede funcionar como recurso protector o como fuente adicional de presión. La funcionalidad familiar, la percepción de apoyo y la calidad de la comunicación se han relacionado con resultados emocionales en estudiantes universitarios. En estudiantes de primer año, la disfunción familiar y un menor apoyo social se asociaron con mayor presencia de ansiedad y depresión aun considerando la situación económica [10]. En Bangladesh, el apoyo social también ha mostrado asociaciones con estrés, ansiedad y depresión [12]. Estas evidencias justifican incluir en un modelo multivariable variables relativas a expectativas familiares, conflictos, apoyo percibido y dificultades económicas, debido a que su contribución puede variar cuando se analizan junto con las demandas académicas.

Las relaciones sociales externas a la familia constituyen otro componente esencial. El apoyo de amistades y personas significativas puede amortiguar el impacto de las demandas, mientras que el aislamiento, el rechazo, la violencia o el acoso pueden intensificar el malestar. Los perfiles de estudiantes con mayor estrés académico y sintomatología depresiva presentan con mayor frecuencia niveles reducidos de apoyo social percibido [11]. Asimismo, la revisión de Arif et al. [14] identificó resultados adversos de salud mental entre universitarios expuestos a ciberacoso, incluidos depresión, ansiedad y estrés. En consecuencia, el riesgo psicosocial debe entenderse como una configuración relacional en la que la presencia de recursos protectores puede modificar el peso de las exposiciones negativas.

Modelos predictivos multivariantes para identificación temprana

Los modelos predictivos multivariantes permiten estimar la probabilidad de un resultado a partir de la contribución simultánea de múltiples características. En salud mental universitaria, la regresión logística conserva especial utilidad por su interpretabilidad y por la posibilidad de expresar asociaciones mediante probabilidades u *odds ratios*, mientras que algoritmos como Random Forest o métodos de *boosting* pueden capturar relaciones no lineales e interacciones complejas. Estudios recientes han mostrado que estas técnicas pueden discriminar niveles de depresión, ansiedad y estrés en población universitaria [3], [15]. Sin embargo, el rendimiento aparente de un algoritmo no es suficiente para justificar su utilidad: la literatura de revisión destaca la necesidad de validación adecuada, control del sobreajuste y evaluación mediante métricas que vayan más allá de la exactitud global [16], [17].

Para una estrategia de identificación temprana, la utilidad del modelo depende tanto de su capacidad discriminativa como de su sensibilidad para detectar estudiantes con afectación relevante y de su calibración al estimar probabilidades. Además, el modelo debe conservar una relación transparente con los factores que pueden ser reconocidos y atendidos por las instituciones. Por

esta razón, resulta metodológicamente pertinente comparar un modelo estadístico interpretable con algoritmos de mayor flexibilidad, evaluar su desempeño mediante validación interna y examinar la importancia relativa de los predictores. Esta combinación permite evitar una oposición artificial entre explicación y predicción y orientar el análisis hacia una herramienta de tamizaje sustentada en factores académicos, familiares y sociales observables.

III. METODOLOGÍA

Se desarrolló un estudio cuantitativo, observacional, transversal y de alcance predictivo, orientado a construir y validar internamente un modelo multivariable para la identificación temprana de afectación psicosocial en estudiantes universitarios. Se utilizaron datos secundarios anonimizados procedentes del conjunto *Dataset on Psychosocial Risk Factors and DASS-21-Based Depression, Anxiety, and Stress Levels Among University Students in Bangladesh*, disponible públicamente en Mendeley Data bajo licencia CC BY 4.0 [1]. La base contiene información sociodemográfica, académica, familiar y social de estudiantes de instituciones públicas y privadas de educación superior.

El conjunto original estuvo integrado por 440 registros y 82 variables. Antes del análisis se realizó una auditoría de estructura, consistencia y completitud. Se identificaron únicamente 14 valores ausentes en toda la matriz de datos, concentrados en las variables prácticas religiosas ($n = 13$) y recepción de asesoramiento psicológico ($n = 1$). Debido a que estas variables no formaron parte del conjunto principal de predictores, no fue necesario imputar valores para el modelo definitivo. Se conservaron los 440 participantes disponibles.

La afectación psicosocial se operacionalizó utilizando las categorías de estrés, ansiedad y depresión derivadas de la escala *Depression Anxiety Stress Scales-21* (DASS-21), instrumento ampliamente empleado para evaluar síntomas emocionales en población universitaria y con evidencia favorable de confiabilidad y validez [5], [6], [7]. La base proporcionó por separado las categorías DASS_Stress, DASS-Anxiety y DASS-Depression, clasificadas como normal, leve, moderada, severa y extremadamente severa.

Para el análisis predictivo se construyó una variable binaria de resultado denominada afectación psicosocial relevante. Se asignó el valor 1 cuando el estudiante presentó una clasificación moderada, severa o extremadamente severa en al menos una de las tres dimensiones del DASS-21, y el valor 0 cuando las tres dimensiones permanecieron en nivel normal o leve. Este criterio permitió identificar manifestaciones de intensidad suficiente para justificar una alerta temprana sin equiparar cualquier sintomatología leve con afectación psicosocial relevante.

Mediante esta operacionalización, 260 estudiantes (59,1%) fueron clasificados con afectación psicosocial relevante y 180 (40,9%) sin afectación relevante. La distribución permitió desarrollar el proceso de clasificación sin necesidad de aplicar técnicas artificiales de sobremuestreo en el análisis principal.

Por otra parte, la selección de predictores se realizó *a priori* de acuerdo con el objetivo del estudio y se organizó en tres dominios principales: académico, familiar y social. Se incorporaron adicionalmente variables sociodemográficas básicas como covariables de ajuste. La selección conceptual previa evitó utilizar procedimientos exclusivamente automáticos para decidir qué variables debían ingresar inicialmente al modelo.

El dominio académico incluyó año académico, tipo de institución, fracaso frecuente en exámenes, preocupación por el fracaso académico, presión familiar relacionada con la selección académica y el rendimiento, satisfacción personal con el desempeño académico, presión por competencia académica, dificultades para continuar los estudios y participación en actividades extracurriculares. También se consideraron las expectativas de logro y la inseguridad respecto al futuro profesional cuando su naturaleza correspondía directamente al contexto educativo.

El dominio familiar incorporó tipo de familia, nivel socioeconómico, antecedentes familiares de trastornos mentales, fallecimiento de un miembro de la familia, discusiones frecuentes entre los padres, divorcio familiar, tutores excesivamente estrictos, conflictos familiares, presencia de adicciones en miembros de la familia, experiencias traumáticas durante la infancia, calidad de las

relaciones familiares, apoyo emocional, satisfacción con las relaciones familiares, percepción de trato parental desigual y presión para satisfacer expectativas familiares.

El dominio social comprendió apoyo de amigos, conflictos interpersonales, experiencias de intimidación o abuso verbal por pares, ruptura de relaciones afectivas, engaño o abuso por personas cercanas, trato desigual por docentes, insulto o acoso docente, percepción de inferioridad en las relaciones interpersonales, temor a comunicarse con personas nuevas, restricciones sociales, inseguridad residencial, insatisfacción con el entorno de residencia y diferentes formas de violencia interpersonal.

Se excluyeron deliberadamente como predictores principales las variables que representaban manifestaciones psicológicas demasiado próximas al desenlace —por ejemplo, “sentirse estresado con frecuencia”— debido al riesgo de fuga de información y de sobreestimar artificialmente la capacidad predictiva. Asimismo, las tres variables DASS utilizadas para construir el desenlace fueron excluidas completamente de la matriz de predictores.

A. Preparación y codificación de los datos

Las variables fueron sometidas a revisión semántica y estandarización antes del modelado. Las respuestas binarias se codificaron como 0 y 1; las variables ordinales basadas en frecuencia se transformaron respetando su dirección creciente, mientras que las variables categóricas nominales fueron representadas mediante codificación *one-hot*. Las categorías sin orden natural no recibieron puntuaciones numéricas arbitrarias. La edad fue depurada mediante la homogeneización de respuestas textuales equivalentes —por ejemplo, “20+”, “20 years” o expresiones similares— y su transformación a valores numéricos. Las variables continuas utilizadas en modelos sensibles a escala fueron estandarizadas exclusivamente dentro de cada partición de entrenamiento para impedir transferencia de información desde los datos de validación. Todo el preprocesamiento se integró dentro de una misma tubería analítica (*pipeline*), de manera que las transformaciones fueran reproducibles y se aplicaran independientemente durante cada ciclo de validación.

B. Desarrollo de los modelos predictivos

Se compararon tres estrategias complementarias. En primer lugar, se estimó una regresión logística penalizada, utilizada como modelo interpretable de referencia y para cuantificar la contribución independiente de los predictores. En segundo lugar, se implementó Random Forest debido a su capacidad para representar relaciones no lineales e interacciones sin requerir especificarlas previamente. Finalmente, se utilizó un modelo de *gradient boosting* para evaluar si una combinación secuencial de árboles permitía mejorar la discriminación respecto de los modelos anteriores. La comparación entre métodos estadísticos y algoritmos de aprendizaje automático se fundamentó en su utilización creciente para detectar riesgo psicológico en población universitaria [3], [16], [17].

La complejidad de los modelos fue controlada para evitar sobreajuste, particularmente relevante debido a la relación entre el tamaño muestral y el número potencial de predictores. En la regresión logística se utilizó regularización para reducir la influencia de coeficientes inestables. En los modelos basados en árboles, la profundidad, número de árboles, tasa de aprendizaje y parámetros de regularización fueron optimizados únicamente a partir de los datos de entrenamiento.

C. Validación interna y evaluación del rendimiento

La validación se realizó mediante validación cruzada estratificada de 10 pliegues, conservando en cada partición una proporción semejante de estudiantes con y sin afectación psicosocial relevante. La selección de hiperparámetros y las transformaciones de preprocesamiento se ejecutaron dentro del proceso de validación para reducir el optimismo derivado de utilizar información del conjunto de evaluación durante el entrenamiento. El desempeño de cada modelo se evaluó mediante el área bajo la curva ROC (AUC-ROC), sensibilidad, especificidad, exactitud, precisión, valor predictivo negativo y puntuación F1. Debido a que el objetivo principal fue la identificación temprana de estudiantes potencialmente afectados, la interpretación no se limitó a la exactitud global. Se otorgó especial atención a la sensibilidad, con el propósito de minimizar falsos negativos, sin asumir que una mayor sensibilidad justificara una reducción desproporcionada de la especificidad.

La calibración se evaluó mediante la concordancia entre probabilidades predichas y proporciones observadas, complementada con el Brier score. Para las métricas principales se estimaron intervalos de confianza del 95 % mediante remuestreo *bootstrap*. El umbral de clasificación no se fijó automáticamente en 0,50; se examinó el equilibrio entre sensibilidad y especificidad y, adicionalmente, un escenario orientado a detección temprana en el que se priorizó la sensibilidad.

D. Interpretabilidad y relevancia de los predictores

La interpretación del modelo estadístico se realizó mediante razones de momios (*odds ratios*, OR) e intervalos de confianza del 95%. En los modelos de aprendizaje automático se examinó la importancia de las variables utilizando métodos independientes del algoritmo, evitando interpretar únicamente las medidas internas de importancia generadas por los árboles. El objetivo fue identificar los factores que aportaban mayor información predictiva y determinar si pertenecían predominantemente a los dominios académico, familiar o social. La interpretación de la importancia predictiva no se consideró equivalente a causalidad. Debido al carácter transversal de los datos, las asociaciones identificadas indican capacidad de clasificación o predicción dentro de la muestra estudiada, pero no permiten establecer que los factores detectados sean causas de estrés, ansiedad o depresión.

E. Análisis complementarios y sensibilidad

Como análisis complementario se examinó individualmente la distribución de estrés, ansiedad y depresión según sus cinco categorías originales. Asimismo, se efectuaron análisis de sensibilidad modificando el criterio de afectación para verificar si la jerarquía de los predictores y el rendimiento general de los modelos se mantenían ante definiciones alternativas del desenlace. Se comparó además el rendimiento del modelo completo con modelos restringidos a cada dominio, académico, familiar y social. Esta estrategia permitió determinar si la integración multivariable aportaba capacidad discriminativa adicional respecto del empleo aislado de un solo grupo de factores y evaluar la contribución relativa de cada dimensión al proceso de identificación temprana.

F. Consideraciones éticas y reproducibilidad

La investigación utilizó exclusivamente datos secundarios anonimizados y disponibles públicamente, sin contacto con los participantes ni acceso a información que permitiera su identificación. El conjunto de datos fue utilizado de acuerdo con las condiciones establecidas por sus autores y la licencia de distribución correspondiente [1]. Los análisis se desarrollaron mediante procedimientos reproducibles de procesamiento y modelado computacional, conservando la separación entre datos de entrenamiento y validación durante todas las etapas del flujo analítico. Los resultados del modelo se interpretaron como herramienta de identificación y apoyo a la toma de decisiones y no como procedimiento diagnóstico. Una probabilidad elevada de afectación psicosocial representa una señal de alerta susceptible de orientar procesos posteriores de evaluación profesional, pero no constituye por sí misma evidencia clínica individual.

IV. RESULTADOS

A. Caracterización de la muestra y definición del desenlace

El análisis incluyó 440 estudiantes universitarios, cuyas características generales se resumen en la Tabla 1. La afectación psicosocial relevante estuvo presente en 260 participantes (59,1%), mientras que 180 (40,9%) permanecieron en niveles normales o leves en las tres dimensiones DASS-21. Entre los componentes evaluados, la ansiedad presentó la mayor frecuencia de afectación moderada o superior, seguida de depresión y estrés, lo que evidencia que el desenlace compuesto integró perfiles emocionales heterogéneos y no estuvo determinado por una única dimensión.

Tabla 1. Características generales de la muestra y distribución de la afectación psicosocial

Variable	Categoría / estadístico	n	%	Observación
Edad	Media $\pm$ DE	440	–	21,80 $\pm$ 2,40; rango 16–40 años
Sexo	Masculino	305	69,3	–
Sexo	Femenino	134	30,5	–
Sexo	Otro	1	0,2	–
Área de residencia	Pueblo	121	27,5	–
Área de residencia	Aldea	151	34,3	–
Área de residencia	Ciudad	168	38,2	–
Nivel socioeconómico	Medio	248	56,4	–
Nivel socioeconómico	Medio bajo	122	27,7	–
Nivel socioeconómico	Medio alto	41	9,3	–
Nivel socioeconómico	Bajo	25	5,7	–
Nivel socioeconómico	Alto	4	0,9	–
Tipo de institución	Pública	248	56,4	–
Tipo de institución	Privada	192	43,6	–
Año académico	1. ^{er} año	147	33,4	–
Año académico	2. ^{do} año	116	26,4	–
Año académico	3. ^{er} año	99	22,5	–
Año académico	4. ^{to} año	78	17,7	–
Afectación psicosocial relevante	Sí	260	59,1	Moderada/severa/extrema en $\geq$ 1 dimensión DASS-21
Afectación psicosocial relevante	No	180	40,9	Normal/leve en las tres dimensiones
Estrés DASS-21	Moderado o superior	110	25,0	–
Ansiedad DASS-21	Moderado o superior	214	48,6	–
Depresión DASS-21	Moderado o superior	182	41,4	–

B. Rendimiento comparativo de los modelos predictivos

Los tres modelos mostraron una capacidad discriminativa consistente, con valores de AUC próximos a 0,80 y sin diferencias de magnitud suficiente para indicar una superioridad marcada de una única estrategia. La regresión logística alcanzó un AUC de 0,795 (IC 95 %: 0,751–0,838), Random Forest obtuvo el valor más alto, con 0,805 (IC 95%: 0,762–0,844), y Gradient Boosting registró 0,799 (IC 95%: 0,752–0,839). La proximidad de los intervalos de confianza y de las curvas ROC sugiere que una parte sustancial de la señal predictiva pudo ser capturada tanto por un modelo lineal regularizado como por algoritmos capaces de representar relaciones no lineales. La Tabla 2 y la Figura 1 resumen esta comparación.

Tabla 2. Rendimiento predictivo de los modelos en validación cruzada estratificada

Modelo	AUC (IC95%)	Sensibilidad (IC95%)	Especificidad (IC95%)	Exactitud	Precisión	VPN	F1	Brier	Umbral medio
Regresión logística	0,795 (0,751–0,838)	0,765 (0,713–0,814)	0,667 (0,603–0,735)	0,725	0,768	0,663	0,767	0,178	0,545
Random Forest	0,805 (0,762–0,844)	0,735 (0,683–0,786)	0,739 (0,674–0,798)	0,736	0,803	0,658	0,767	0,176	0,589
Gradient Boosting	0,799 (0,752–0,839)	0,754 (0,700–0,804)	0,722 (0,655–0,788)	0,741	0,797	0,670	0,775	0,176	0,595

La regresión logística presentó la sensibilidad más alta del análisis principal (0,765), aunque con una especificidad menor (0,667). Random Forest mostró el balance más simétrico entre sensibilidad (0,735) y especificidad (0,739), además de la mayor precisión positiva (0,803). Gradient Boosting alcanzó la mayor exactitud global (0,741) y el mayor F1 (0,775), con sensibilidad de 0,754 y especificidad de 0,722. Estos resultados indican que las diferencias entre algoritmos dependieron más del equilibrio entre falsos positivos y falsos negativos que de cambios importantes en la discriminación global.

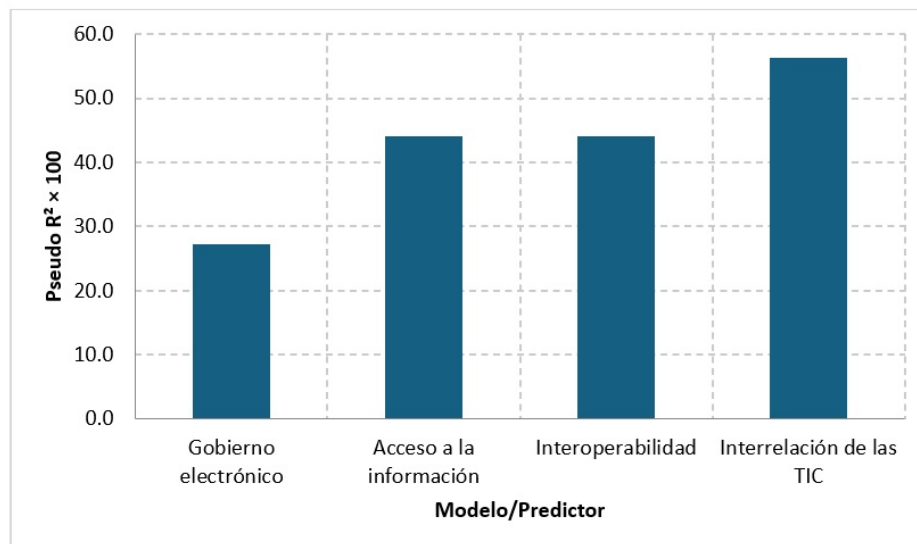


Fig. 1. Curvas ROC de los modelos predictivos

La calibración fue igualmente semejante entre los modelos. Los Brier scores se situaron entre 0,176 y 0,178, con el valor más bajo para Gradient Boosting (0,176), seguido muy de cerca por Random Forest (0,176). Las curvas de calibración mantuvieron una correspondencia razonable entre las probabilidades estimadas y las proporciones observadas, aunque se identificaron desviaciones locales en los intervalos de probabilidad intermedia y alta. La Figura 2 muestra que estas desviaciones no siguieron un patrón uniforme de sobreestimación o subestimación en los tres algoritmos.

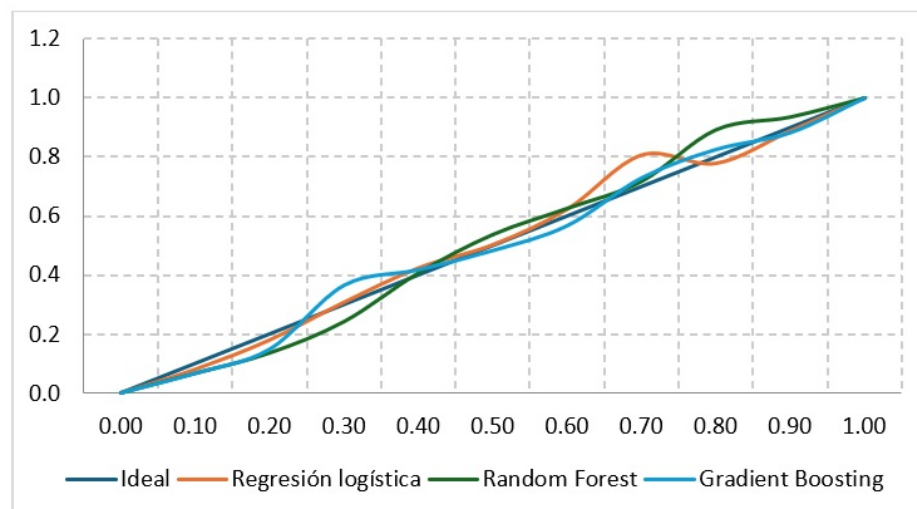


Fig. 2. Calibración de las probabilidades predichas

C. Contribución de los dominios académico, familiar y social

El análisis restringido por dominios mostró que la información social tuvo la mayor capacidad discriminativa individual, con un AUC de 0,788 (IC 95%: 0,745–0,829). El dominio académico alcanzó un AUC de 0,769 (IC 95%: 0,723–0,813), mientras que el dominio familiar presentó un rendimiento menor, aunque todavía superior al azar, con 0,703 (IC 95%: 0,655–0,751). La integración simultánea de los tres dominios y las covariables produjo un AUC de 0,795 (IC 95%: 0,754–0,836) en la regresión logística, como se observa en la Tabla 3.

Tabla 3. Capacidad discriminativa de los dominios de predictores (regresión logística)

Dominio	AUC	IC 95 % inferior	IC 95 % superior
Académico	0.769	0.723	0.813
Familiar	0.703	0.655	0.751
Social	0.788	0.745	0.829
Integrado	0.795	0.754	0.836

El incremento del modelo integrado respecto del dominio social aislado fue moderado, lo que indica que una proporción importante de la señal predictiva estaba contenida en las experiencias interpersonales y sociales. Sin embargo, la incorporación de factores académicos y familiares añadió información complementaria y permitió representar perfiles de riesgo más amplios. Este comportamiento respalda una interpretación multivariable: la afectación psicosocial no se concentró en un único ámbito, aunque los factores sociales mostraron una contribución particularmente relevante dentro de la muestra.

D. Predictores con mayor contribución al riesgo estimado

La importancia por permutación del modelo Random Forest identificó como predictor dominante la percepción de inferioridad en las relaciones de amistad, cuya permutación produjo una disminución media del AUC de 0,0286. A continuación se ubicaron el temor de no alcanzar las metas vitales (0,0128), haber sido engañado por amistades (0,0108), el complejo de inferioridad (0,0086), la percepción de sesgo docente (0,0066) y haber sufrido intimidación o abuso verbal por pares u otras personas (0,0061). También aparecieron entre los predictores principales la edad, el nivel socioeconómico, la percepción de no alcanzar los objetivos propuestos, las dificultades para continuar los estudios, el apoyo de amistades y la participación en actividades extracurriculares. La Figura 3 muestra la jerarquía de las doce variables con mayor contribución.

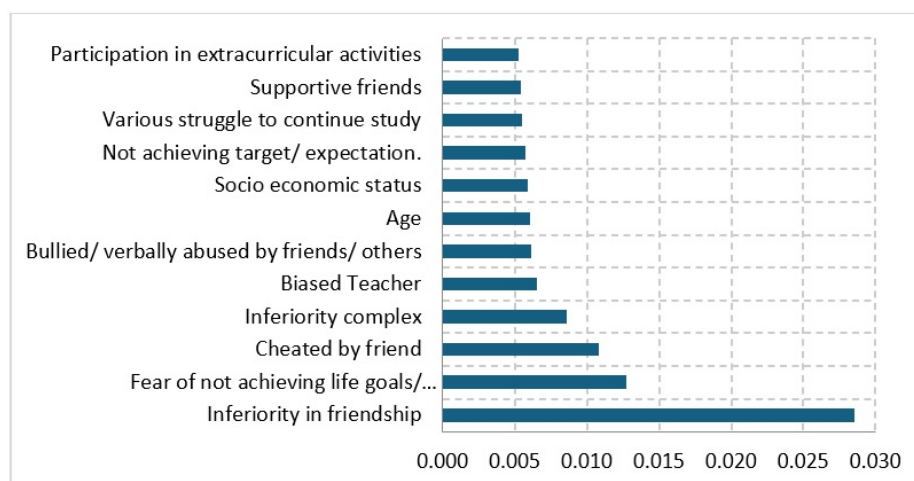


Fig. 3. Importancia de los principales predictores

La regresión logística regularizada identificó mayores probabilidades de afectación asociadas con el temor de no alcanzar las metas vitales (OR = 1,442; IC 95%: 1,208–1,701), la percepción de no cumplir los objetivos propuestos (OR = 1,324; IC 95%: 1,081–1,574), experiencias traumáticas en la infancia (OR = 1,301; IC 95%: 1,094–1,536), preocupación por el fracaso académico (OR = 1,243; IC 95%: 1,010–1,508) y fracaso frecuente en exámenes (OR = 1,178; IC 95%: 1,030–1,337), mientras que no experimentar inferioridad en las amistades se asoció con menor probabilidad estimada de afectación (OR = 0,654; IC 95%: 0,564–0,765). La Tabla 4 resume las asociaciones principales. La concordancia parcial con la importancia por permutación respalda la estabilidad de factores vinculados con autovaloración interpersonal, expectativas de logro, experiencias adversas y presión académica, interpretados como predictores y no como relaciones causales.

Tabla 4. Principales asociaciones del modelo de regresión logística regularizada

Variable	Categoría	Referencia	OR	IC95% inferior	IC95% superior
Sentimiento de inferioridad en la amistad	Nunca	Siempre	0,654	0,564	0,765
Temor a no alcanzar las metas/expectativas de vida	Sí	No	1,442	1,208	1,701
No alcanzar la meta/expectativa	Sí	No	1,324	1,081	1,574
Complejo de inferioridad	Nunca	Siempre	0,759	0,649	0,873
Experiencia traumática en la infancia	Sí	No	1,301	1,094	1,536
Sexo	Masculino	Femenino	0,800	0,670	0,946
Preocupación por el fracaso académico	Sí	No	1,243	1,010	1,508
Engañado/a por un amigo	Nunca	Siempre	0,814	0,686	0,960
Nivel socioeconómico	Medio	Bajo	0,817	0,691	0,976
Prefiere estar solo/a la mayor parte del tiempo	Sí	No	1,191	0,989	1,426

E. Escenario orientado a identificación temprana

Como análisis complementario se evaluó un escenario en el que el umbral de clasificación se desplazó para priorizar la detección de estudiantes afectados. La regresión logística incrementó su sensibilidad hasta 0,838, con una especificidad de 0,600; Gradient Boosting alcanzó una sensibilidad de 0,823 y especificidad de 0,633; y Random Forest mantuvo una sensibilidad de 0,773 con especificidad de 0,711. La Tabla 5 resume este escenario. El ajuste confirma el compromiso esperado entre sensibilidad y especificidad: una estrategia de alerta temprana puede reducir falsos negativos, pero a costa de incrementar el número de estudiantes señalados para evaluación posterior.

Los resultados muestran que el problema puede abordarse con un rendimiento predictivo moderado y estable, sin depender de un único algoritmo. Random Forest ofreció la mayor discriminación global, mientras que la regresión logística conservó ventajas de interpretabilidad y sensibilidad. La convergencia entre modelos y la relevancia recurrente de factores sociales y académicos apoyan el uso del sistema como herramienta de priorización para evaluación posterior, no como procedimiento diagnóstico autónomo.

Tabla 5. Escenario secundario orientado a identificación temprana con prioridad de sensibilidad

Modelo	Sensibilidad	Especificidad	Exactitud	Precisión	VPN	F1	Umbral medio
Regresión logística	0,838	0,600	0,741	0,752	0,720	0,793	0,460
Random Forest	0,773	0,711	0,748	0,794	0,684	0,784	0,562
Gradient Boosting	0,823	0,633	0,745	0,764	0,713	0,793	0,501

CONCLUSIONES

El modelo predictivo multivariable permitió identificar afectación psicosocial relevante en estudiantes universitarios con un rendimiento discriminativo moderado y estable. Los tres enfoques evaluados alcanzaron valores de AUC próximos a 0,80, lo que demuestra que la combinación de factores académicos, familiares y sociales contiene información útil para la identificación temprana de estudiantes potencialmente vulnerables.

Los factores sociales concentraron la mayor capacidad predictiva individual, especialmente aquellos relacionados con autovaloración interpersonal, temor al fracaso, experiencias adversas y

relaciones con pares. No obstante, la incorporación conjunta de variables académicas y familiares mejoró la representación del riesgo, confirmando que la afectación psicosocial responde a una configuración multivariable y no a un único ámbito de exposición.

Los resultados respaldan el uso de modelos predictivos como herramientas de apoyo para priorizar procesos de evaluación y acompañamiento universitario, pero no como procedimientos diagnósticos.

Debido al carácter transversal de los datos y a la procedencia específica de la muestra, se requiere validación externa en otras poblaciones universitarias antes de considerar su aplicación institucional generalizada.

REFERENCIAS

- [1] T. R. Shaha, M. Begum, T. Podder, A. Siddika, T. Naznin, and M. T. Ahmed, "Dataset on psychosocial risk factors and dass-21-based depression, anxiety, and stress levels among university students in bangladesh," 2026.
- [2] M. Kamruzzaman, A. Hossain, M. A. Islam, M. S. Ahmed, E. Kabir, and M. N. Khan, "Exploring the prevalence of depression, anxiety, and stress among university students in bangladesh and their determinants," *Clinical Epidemiology and Global Health*, vol. 28, p. 101677, 2024.
- [3] M. E. Hasan *et al.*, "Prevalence, associated factors, and machine learning-based prediction of depression, anxiety, and stress among university students: a cross-sectional study from bangladesh," *Journal of Health, Population and Nutrition*, vol. 44, no. 1, p. 361, 2025.
- [4] M. M. Hossain, M. A. Alam, and M. H. Masum, "Prevalence of anxiety, depression, and stress among students of jahangirnagar university in bangladesh," *Health Science Reports*, vol. 5, p. e559, 2022.
- [5] M. D. Manzar, M. Salahuddin, D. Nureye, F. Z. Kashoo, M. M. Noohu, J. S. Alotaibi, M. S. Alamri, and M. D. Griffiths, "Depression, anxiety, and stress scale-21 (dass-21): Further psychometric exploration using robust item response theory and classical theory measures among university students," *PLoS ONE*, vol. 20, no. 7, p. e0325238, 2025.
- [6] A. Osman, J. L. Wong, C. L. Bagge, S. Freedenthal, P. M. Gutierrez, and G. Lozano, "The depression anxiety stress scales-21 (dass-21): Further examination of dimensions, scale reliability, and correlates," *Journal of Clinical Psychology*, vol. 68, no. 12, pp. 1322-1338, 2012.
- [7] J. D. Henry and J. R. Crawford, "The short-form version of the depression anxiety stress scales (dass-21): Construct validity and normative data in a large non-clinical sample," *British Journal of Clinical Psychology*, vol. 44, no. 2, pp. 227-239, 2005.
- [8] Z. Antúnez and E. V. Vinet, "Escala de depresión, ansiedad y estrés (dass-21): Validación de la versión abreviada en estudiantes universitarios chilenos," *Terapia Psicológica*, vol. 30, no. 3, pp. 49-55, 2012.
- [9] J. J. B. R. Aruta, "Screening psychological symptoms in filipino university students during the covid-19 pandemic: Translation and structural validation of the filipino version of the dass-21," *Psychology in the Schools*, vol. 61, no. 8, pp. 3243-3262, 2024.
- [10] A. E.-P. no, M. F. Rivadeneira, J. Alvarado, M. Murtagh, S. Guijarro, L. Alomoto, and G. C. narejo, "Anxiety and depression in first-year university students: the role of family and social support," *Frontiers in Psychology*, vol. 15, p. 1462948, 2024.
- [11] M. Kotyśko and J. Frankowiak, "The importance of perceived social support for symptoms of depression and academic stress among university students—a latent profile analysis," *PLoS ONE*, vol. 20, no. 5, p. e0324785, 2025.
- [12] S. Howlader, S. Abedin, and M. M. Rahman, "Social support, distress, stress, anxiety, and depression as predictors of suicidal thoughts among selected university students in bangladesh," *PLOS Global Public Health*, vol. 4, no. 4, p. e0002924, 2024.

- [13] G. Nakie *et al.*, “Sleep quality and associated factors among university students in africa: a systematic review and meta-analysis study,” *Frontiers in Psychiatry*, vol. 15, p. 1370757, 2024.
- [14] A. Arif, M. A. Qadir, R. S. Martins, and H. M. A. Khuwaja, “The impact of cyberbullying on mental health outcomes amongst university students: A systematic review,” *PLOS Mental Health*, vol. 1, no. 6, p. e0000166, 2024.
- [15] C. E. Morr, M. Jammal, I. Bou-Hamad, S. Hijazi, D. Ayna, M. Romani, and R. Hoteit, “Predictive machine learning models for assessing lebanese university students’ depression, anxiety, and stress during covid-19,” *Journal of Primary Care & Community Health*, vol. 15, p. 21501319241235588, 2024.
- [16] B. L. Schaab *et al.*, “How do machine learning models perform in the detection of depression, anxiety, and stress among undergraduate students? a systematic review,” *Cadernos de Saúde Pública*, vol. 40, no. 11, p. e00029323, 2024.
- [17] A. Daza, N. Saboya, J. I. Necochea-Chamorro, K. Z. Ramos, and Y. del R. Vásquez Valencia, “Systematic review of machine learning techniques to predict anxiety and stress in college students,” *Informatics in Medicine Unlocked*, vol. 43, p. 101391, 2023.
- [18] L. Carmona, C. Costa, S. Gascón, G. Ribeiro, and M. J. Chambel, “Prevalence and risk factors for anxiety, stress and depression among higher education students in portugal and brazil,” *Journal of Affective Disorders Reports*, vol. 17, p. 100825, 2024.
- [19] G. X. D. Tan, X. C. Soh, A. Hartanto, A. Y. H. Goh, and N. M. Majeed, “Prevalence of anxiety in college and university students: An umbrella review,” *Journal of Affective Disorders Reports*, vol. 14, p. 100658, 2023.
- [20] Iqra, “A systematic review of academic stress intended to improve the educational journey of learners,” *Methods in Psychology*, vol. 11, p. 100163, 2024.