Predictive model to anticipate academic failure in students at the Technological University of Panama starting from the eighth week of the semester

Sergio Cervantes[1], Jose Villarreal[1], América Weng-Kong[1], Gustavo Quintero-Barreto[1], Aranzazu Berbey-Alvarez[1]

[1]Universidad Tecnológica de Panamá, Republica de Panamá, sergio.cervantes@utp.ac.pa, jose.villarreal@utp.ac.pa, America.weng@utp.ac.pa, Gustavo.quintero@utp.ac.pa; Aranzazu.berbey@utp.ac.pa

Abstract– ***Academic failure in higher education remains a persistent challenge in Panama, particularly at the Universidad Tecnológica de Panamá (UTP), where high rates of course failure and dropout negatively affect both institutional efficiency and student trajectories. This study proposes the development of a predictive model using machine learning techniques to identify at-risk students by the eighth week of the semester. A structured database of 115,081 academic records was analyzed, resulting in 58,491 valid entries for model training and testing. Logistic regression, decision tree, and random forest algorithms were evaluated and compared using metrics such as AUC, accuracy, recall, and F1-score. The results indicate that the random forest model achieved the best performance, outperforming the other algorithms in sensitivity and balance between true positives and false positives. Consequently, its implementation is recommended as the basis of an institutional early warning system, complemented by logistic regression for interpretability purposes. The main contribution of this research lies in providing an automated and reliable tool that supports academic decision-making, optimizes institutional resources, and promotes a more equitable and efficient higher education system.***

Keywords-- ***Retention, academic failure, data processing, educational technology.***

Modelo predictivo para anticipar el fracaso académico en estudiantes de la Universidad Tecnológica de Panamá a partir de la octava semana del semestre

Sergio Cervantes[1], Jose Villarreal[1], América Weng-Kong[1], Gustavo Quintero-Barreto[1], Aranzazu Berbey-Alvarez[1]
[1]Universidad Tecnológica de Panamá, Republica de Panamá, sergio.cervantes@utp.ac.pa, jose.villarreal@utp.ac.pa, America.weng@utp.ac.pa, Gustavo.quintero@utp.ac.pa; Aranzazu.berbey@utp.ac.pa

Resumen– ***El fracaso académico en la educación superior constituye un desafío persistente en Panamá y particularmente en la Universidad Tecnológica de Panamá (UTP), donde las altas tasas de reprobación y deserción afectan la eficiencia institucional y la trayectoria estudiantil. Este estudio propone el desarrollo de un modelo predictivo basado en técnicas de aprendizaje automático, capaz de identificar estudiantes en riesgo a partir de la octava semana del semestre. Se analizó una base de datos estructurada con 115,081 registros académicos, de los cuales se seleccionaron 58,491 válidos para el entrenamiento y prueba de los modelos. Se evaluaron algoritmos de regresión logística, árbol de decisión y Random Forest, comparando su desempeño mediante métricas como AUC, precisión, recall y F1-score. Los resultados muestran que el Random Forest alcanzó el mejor rendimiento, destacándose por su capacidad para identificar correctamente a estudiantes en riesgo, superando a los demás modelos en sensibilidad y equilibrio entre aciertos y falsos positivos. Estos hallazgos permiten recomendar su implementación como base para un sistema institucional de alerta temprana, complementado con la regresión logística para fines de interpretabilidad. El aporte principal de la investigación radica en proveer una herramienta automatizada y confiable que fortalezca la toma de decisiones académicas, optimice los recursos institucionales y promueva una educación superior más equitativa y eficiente.***

Palabras clave-- ***Retención estudiantil, fracaso académico, procesamiento de datos, tecnología educativa***.

I. Introducción

El fracaso académico en la educación superior representa un desafío persistente para las universidades, especialmente en contextos donde la equidad y la eficiencia institucional enfrentan limitaciones estructurales. En Panamá, y particularmente en la Universidad Tecnológica de Panamá (UTP), esta problemática se refleja en altas tasas de reprobación y deserción, afectando tanto la trayectoria estudiantil como la gestión de los recursos institucionales [1].

El rendimiento académico está determinado por múltiples factores —académicos, socioeconómicos, psicológicos, demográficos, pedagógicos e institucionales—, lo que exige enfoques integrales para su análisis [2]. En este sentido, hay experiencias previas institucionales sobre las causas de la deserción universitaria [3], el seguimiento estadístico de una asignatura de ingeniería [4], la situación de la educación superior postpandemia [5], modelo predictivo basado en IA y el contexto para identificar estudiantes en riesgo de deserción universitaria en Panamá [6], autorregulación del aprendizaje [2] y estudio sobre los factores psicosociales que intervienen en la permanencia de los estudiantes universitarios [7].

En este contexto, el uso de las herramientas propias de la ciencia de datos y el aprendizaje automático ofrecen nuevas oportunidades para anticipar los riesgos y así poder diseñar intervenciones tempranas. En este sentido, de acuerdo con Luzuriaga-Pavón y Castillo-Páez [8] concluyó que el uso de modelos predictivos no solo ayuda a comprender los determinantes de la deserción, sino que también permite asignar de manera más eficiente los recursos institucionales y diseñar estrategias de apoyo, tales como programas de tutorías, asesorías personalizadas o intervenciones focalizadas sobre todo en aquella asignaturas críticas [9] con alta carga cognitiva y apoyos económicos y psicológicos para los estudiantes de carreras de ingenierías. En este hilo de ideas, Guzman *et al.,* [10] indicó que es necesario fortalecer los programas de apoyo financiero, tutorías académicas y orientación psicológica para mejorar la retención estudiantil. Otras estratégicas corresponde a términos de transporte y servicios de apoyo a los estudiantes [11]. Las instituciones deben adoptar estrategias basadas en datos que permitan la identificación temprana de estudiantes en riesgo y la implementación de intervenciones efectivas para la implementación de mecanismos focalizado de apoyo temprano al estudiante.

Esta investigación propone el desarrollo de un modelo predictivo basado en datos institucionales históricos, capaz de identificar, desde las etapas tempranas del semestre, a los estudiantes con alta probabilidad de fracaso académico. El objetivo es fortalecer la toma de decisiones académicas, mejorar la permanencia estudiantil y contribuir a una educación superior más equitativa y eficiente, en línea con los objetivos estratégicos de la UTP, las tendencias internacionales en analítica educativa [12] y los objetivos de desarrollo sostenible [13].

En América Latina, la permanencia estudiantil en la educación superior ha estado condicionada por factores estructurales como: la planificación a largo plazo, la

desigualdad social, la gestión de los recursos y la estabilidad institucional. Aunque Panamá ha mostrado avances en desarrollo y gobernabilidad, persisten brechas significativas en el acceso equitativo a la educación [14].

La pandemia de COVID-19 profundizó estas desigualdades, afectando especialmente a los estudiantes de los sectores más vulnerables. Investigaciones recientes destacan cómo las condiciones socioeconómicas y la falta de políticas adaptadas aumentan el riesgo de abandono educativo [15]. En Panamá, aún no existe una estrategia nacional integral para prevenir la deserción universitaria, y factores como el acceso desigual a la tecnología y la conectividad siguen limitando la continuidad académica [16]. En Barsallo-Alvarado [17] se indicaron una serie de factores como causantes del abandono universitario aspectos como situaciones individuales o de vida, la estructura del sistema educativo en oposición con la orientación institucional, problemas económicos, falta de orientación previa para la admisión universitaria, situaciones familiares, problemas de transporte, distancias, la seguridad personal, el crimen, las drogas, el aislamiento, la falta de valores, y la cultura de conformismo. En Vejerano y Trujillo- González [11] se indican aspectos como la influencia de la accesibilidad y la logística para acceder al centro universitario.

Ante este panorama, los modelos predictivos basados en datos emergen como herramientas clave para anticipar el fracaso académico y facilitar las intervenciones oportunas, promoviendo una educación superior más inclusiva y eficiente.

El fracaso académico en la UTP no solo representa un reto institucional, sino también una oportunidad para implementar soluciones innovadoras. Este estudio propone un modelo predictivo que, a partir de la octava semana del semestre, identifique a estudiantes en riesgo, permitiendo intervenciones como tutorías, orientación académica o ajustes curriculares.

La propuesta se alinea con los objetivos estratégicos de la UTP en cuanto a innovación, calidad y toma de decisiones basadas en evidencia [12]. La integración de técnicas de *machine learning (ML)* permitirá automatizar procesos de monitoreo y anticipar problemáticas complejas con mayor precisión.

Desde una perspectiva social, el proyecto busca reducir desigualdades educativas, considerando que el fracaso académico en la educación superior suele estar vinculado a factores más allá del rendimiento cognitivo, como condiciones socioeconómicas, de salud física y mental, tecnológicas o motivacionales. Su implementación práctica puede generar un impacto positivo tanto en la comunidad estudiantil como en la gestión institucional, además de aportar conocimiento relevante al campo de la analítica educativa en Panamá.

II. MARCO TEÓRICO

A. El fenómeno del fracaso académico y la intervención temprana

El fracaso académico o deserción estudiantil representa un desafío significativo para las instituciones de educación superior, afectando la eficiencia institucional y el desarrollo profesional de los estudiantes. Este fenómeno es reconocido como multifactorial, resultado de la compleja interacción entre elementos individuales, académicos, socioeconómicos y contextuales [1].

La literatura especializada, incluyendo estudios realizados en la Universidad Tecnológica de Panamá (UTP) [1], señala que la deserción universitaria se origina por una combinación de problemáticas sociales (pobreza, sobreempleo, falta de expectativas), familiares (influencia del ambiente familiar) y personales/psicológicas (estado emocional, actitud hacia el abandono), acceso a un transporte eficiente en áreas como Arraiján [11], distancias y seguridad [17] y en algunos casos problemas de salud física y mental [7].

Para contrarrestar este problema, el enfoque se ha desplazado hacia la detección proactiva mediante los Sistemas de Alerta Temprana (SAT) [18]. Estos sistemas buscan identificar a estudiantes en riesgo en las primeras etapas del semestre, haciendo que la predicción temprana, como la propuesta a partir de la octava semana, sea una herramienta crucial para la implementación de estrategias de intervención oportunas [8].

B. Factores clave que influyen en el fracaso académico

El desarrollo de un modelo predictivo eficaz requiere la identificación de variables predictoras relevantes[19]. Estas se agrupan en tres categorías principales que deben nutrir la base de datos del modelo:

--Factores académicos: Las variables históricas y actuales como las calificaciones, índice académico, el desempeño en asignaturas clave, y la adaptación al ritmo universitario son fundamentales para el modelado [8], [11], [20]

--Factores socioeconómicos: Factores como el nivel de ingresos, el origen geográfico [8], la red de apoyo familiar, las condiciones de estudio y el impacto del estado emocional son determinantes en el rendimiento [21]. Estudios en la UTP confirman que el factor socioeconómico está estrechamente ligado a la deserción [1], [3].

C. Diseño de la base de datos y tecnología analítica

La implementación del modelo exige una base de datos robusta que facilite el análisis predictivo [22].

--Integración de datos: La base de datos debe ser diseñada para integrar la información proveniente de múltiples fuentes: académica (notas, créditos), sociodemográfica, psicológica y, de manera crucial, la actividad digital/conductual del estudiante [18], [20].

--*Machine learning* y analítica: El uso de *machine learning* permite, a través de algoritmos, identificar patrones ocultos en el comportamiento de los datos masivos y elaborar las predicciones necesarias para automatizar las decisiones de alerta [22]. El preprocesamiento de estos datos es clave para garantizar el rendimiento de los modelos.

D. Evaluación y aplicación de modelos predictivos

El objetivo central del proyecto es evaluar la precisión y aplicabilidad de distintos algoritmos de aprendizaje automático para la clasificación binaria (éxito/fracaso académico). En la primera tabla se repasa estos modelos y se determina la utilidad del modelo revisado para el proyecto.

E. Métricas de evaluación clave

En el contexto de un SAT, no basta con la precisión global del modelo. Es fundamental centrarse en dos aspectos clave:

--Sensibilidad: El porcentaje de estudiantes en riesgo (clase positiva) que son correctamente identificados por el modelo. Una alta sensibilidad es crucial para que las intervenciones lleguen a quienes realmente las necesitan [23].

--Especificidad: El porcentaje de estudiantes sin riesgo (clase negativa) que son clasificados correctamente. Mantiene la confianza en el sistema al evitar falsas alarmas [23].

F. Aspectos éticos

La investigación con datos estudiantiles demanda adherencia a los principios del CIBio-UTP [24]. La confidencialidad se garantiza mediante la seudonimización/anonimización de los registros, desvinculando la identidad de los atributos de riesgo utilizados por el modelo, es decir, interesan las características del estudiante no quien es él o la estudiante.

El uso de algoritmos de *Machine Learning* impone la obligación ética de la objetividad, minimizando el riesgo de sesgo algorítmico en la selección de datos que pueda inducir a la discriminación. Además, el principio de transparencia requiere que las predicciones sean explicables y auditables. Esto permite justificar las alertas de riesgo (*Random Forest*) para intervenciones focalizadas y justas.

Finalmente, el consentimiento informado se aborda mediante la formalización de un acuerdo institucional que legitima el uso de los datos administrativos e históricos de la UTP para los fines explícitos de la investigación y mejora de la retención de los estudiantes universitarios [24].

G. Aspectos legales

La investigación se enmarca en la legislación panameña, con estricto cumplimiento de la Ley 81 de 2019 sobre Protección de Datos Personales [25], vigilada por la ANTAI [26].Se observan rigurosamente los siguientes principios:

--Licitud y consentimiento: El tratamiento de los datos se ampara en un Convenio institucional formal con la UTP que autoriza el uso de la información exclusivamente para la investigación y el desarrollo del modelo predictivo.

--Finalidad y proporcionalidad: Los datos se recolectan únicamente para el fin específico de predecir el fracaso académico. Se garantiza la proporcionalidad al incluir solo aquellas variables que demostraron ser predictivas según el análisis exploratorio y la importancia de variables del modelo.

--Confidencialidad y seguridad: Se aplican medidas técnicas y organizativas (anonimización y acceso restringido) para proteger la integridad y confidencialidad de los registros académicos.

III. METODOLOGÍA

El presente estudio adopta un enfoque cuantitativo, no experimental, de carácter longitudinal y retrospectivo, basado en el análisis de registros institucionales de la Universidad Tecnológica de Panamá (UTP).

A. Diseño de variables y muestra

La variable dependiente es el fracaso académico, definida como la reprobación de materias o del semestre completo. La reprobación de las materias corresponde a lo definido en el Estatuto Universitario de la UTP como fracaso [27]. El sistema de calificaciones de la Universidad Tecnológica de Panamá se expresa por letras con el siguiente significado para fracaso, letra F en el rango numérico de menos de 61/100 [27]. Las variables independientes incluyen factores demográficos, académicos, socioeconómicos e institucionales.

Se utilizaron registros académicos correspondientes al año 2024, así como al año 2025 hasta el primer semestre, con un total de 115,081 perfiles. Para evaluar la utilidad de estos datos en el entrenamiento del modelo predictivo, se aplicaron diversas técnicas descritas en el apartado metodológico. Posteriormente, se empleó una proporción estándar de 80 % para entrenamiento y 20 % para pruebas, con el fin de garantizar una adecuada validación del desempeño de los modelos.

B. Proceso metodológico

El proceso de construcción y selección del modelo predictivo comprende cuatro etapas secuenciales:

--Recolección y la limpieza de los datos: Esta etapa incluye la depuración de registros incompletos, la estandarización de formatos, el tratamiento de valores faltantes si es necesario y la codificación de variables categóricas.

--Análisis exploratorio de datos (EDA): Se realiza un análisis mediante estadística descriptiva, cálculo de correlaciones y visualizaciones. Esto permite identificar patrones, caracterizar a la población en riesgo y establecer las relaciones entre las variables.

--Modelado y entrenamiento de modelos candidatos: Se entrenan y evalúan diversos algoritmos de clasificación binaria—regresión logística, árboles de decisión y *Random Forest*—utilizando validación cruzada estratificada (ver tabla I).

TABLA I DEFINICIONES DE ALGORITMOS CANDIDATOS

Algoritmo	Definición y mecanismo	Utilidad en el proyecto
Regresión Logística	Modelo estadístico que predice la probabilidad de un evento binario (éxito/fracaso) aplicando la función sigmoide a una combinación lineal de variables, transformando valores continuos en probabilidades entre 0 y 1 [8]	Ofrece interpretabilidad clara, mostrando cómo cada factor influye en la probabilidad de fracaso.

Árboles de Decisión	Modelo jerárquico que divide el conjunto de datos mediante reglas de decisión basadas en atributos (usando métricas como la ganancia de información o el índice de *Gini*) hasta llegar a una predicción de clase [20].	Proporciona reglas de decisión fácilmente comprensibles, útiles para la intervención institucional.
Random Forest	Técnica de ensamblado que combina múltiples árboles de decisión entrenados con subconjuntos aleatorios de datos. Selecciona aleatoriamente subconjuntos de características en cada nodo para mejorar la precisión y evitar el sobreajuste [2 1] .	Reconocido por su alta precisión y robustez [28], y ofrece medidas de importancia de variables [29], ayudando a identificar los factores más influyentes.

--Evaluación de desempeño y selección dinámica: El rendimiento de cada modelo se evalúa con las métricas de clasificación estándar: precisión, *recall* (sensibilidad*), F1-score, ROC-AUC (Receiver Operating Characteristic - Area Under the Curve)* y el análisis de la matriz de confusión.

C. Estrategia de selección y aplicación del modelo

La estrategia de aplicación se basa en la evaluación continua del desempeño de los modelos sobre los datos históricos más recientes. En lugar de seleccionar un único modelo fijo, se mantienen los modelos entrenados como candidatos. La información

Para cada nuevo período de predicción (la octava semana del semestre), se selecciona el modelo que demuestre el mejor equilibrio entre sensibilidad (*recall* de la clase en riesgo, priorizando la detección de falsos negativos) y precisión/*ROC-AUC* sobre el conjunto de datos de validación actual. De este modo, se garantiza que la UTP aplique el modelo más adecuado para ese semestre en particular, asegurando la máxima eficacia predictiva para la intervención temprana.

D. Descripción de la muestra final

Como parte del desarrollo del modelo predictivo, se realizó un riguroso proceso de recolección, limpieza y análisis exploratorio de datos (EDA), utilizando registros institucionales proporcionados por la Dirección de Planificación Universitaria (DIPLAN) de la UTP en conjunto con la Dirección General de Tecnología de la Información y Comunicaciones (DITIC) que es la entidad técnica que se encarga del soporte informático, infraestructura digital y el mantenimiento del Sistema de Matrícula UTP donde se ingresan dichos datos [30][31] y la Secretaría General de la Universidad Tecnológica de Panamá (UTP) es el órgano oficial encargado de administrar, organizar, custodiar y registrar los datos de matrícula e historial académico de los estudiantes [32]. La base original contenía 78 variables y un total de 115,081 registros estudiantiles, correspondientes a los años 2024 y 2025. Tras aplicar criterios de depuración y estandarización, se obtuvo una muestra final compuesta por 58,491 registros válidos. Los criterios de depuración y estandarización de la matrícula en la Universidad Tecnológica de Panamá (UTP) forman parte del conjunto de reglas lógicas, académicas y administrativas integradas en su Sistema de Matrícula Web. Las reglas lógicas corresponden a la validación de identidad unificada, control de concurrencia por cita de matrícula, asignación automática de deuda y la integración del correo institucional. Las reglas académicas corresponden al bloqueo por requisitos previos, límite de créditos permitidos, regla de asignaturas fundamentals y la restricción de puntaje mínimo de ingreso. Las reglas administrativas corresponden a control de solvencia financiera, regla de retiro total excluyente de deuda, estatus de condicionalidad y la validación de admisión. Esta muestra fue dividida en dos subconjuntos: 46,792 registros para entrenamiento del modelo y 11,699 para pruebas, es decir, una relación de 79.45% de entrenamiento del modelo y 20% para pruebas. A partir del análisis exploratorio, se identificaron las variables más relevantes para la predicción del fracaso académico, las cuales se pueden observar en la figura 1.

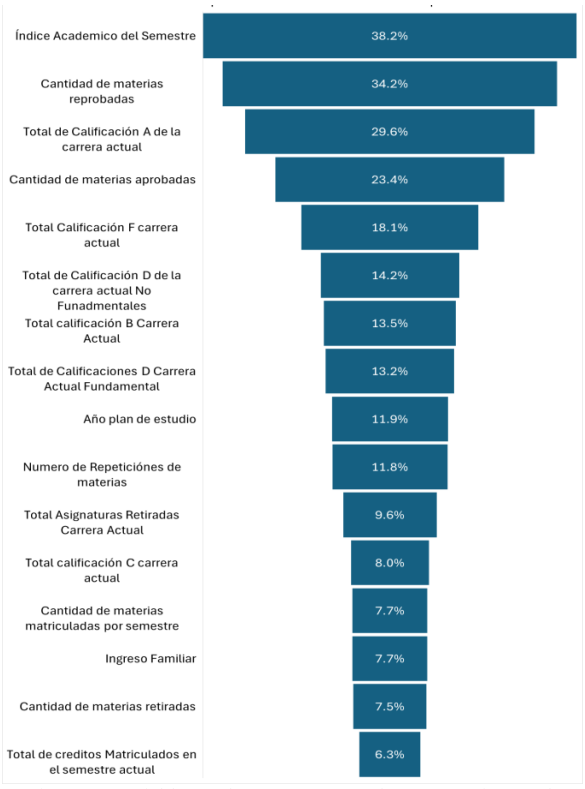


Fig. 1. Variables relevantes para el entrenamiento de modelos predictivos.

A partir del análisis EDA, se identificaron las variables con mayor poder predictivo para anticipar el riesgo de fracaso académico. Entre ellas destacan el Índice Académico del Semestre (38.2%) y la cantidad de materias reprobadas (34.2%), seguidas por el Total de calificaciones A en la carrera actual (29.6%) y la cantidad de materias aprobadas (23.4%). También se consideran variables como el total de calificaciones F y D, el número de repeticiones de materias, el año del plan de estudio, el ingreso familiar, y el Total de

créditos matriculados en el semestre actual, entre otras. Estas variables fueron seleccionadas por su alta correlación con el desempeño académico y su capacidad para alimentar los algoritmos de clasificación binaria utilizados en el modelo.

Se garantizó que las variables utilizadas para realizar la predicción correspondieran únicamente a información disponible hasta la octava semana del semestre mediante la habilitación de una ventana temporal al docente en el Sistema web de matrícula de la Universidad Tecnológica de Panamá durante la novena semana de clase del semestre [33].

E. Entrenamiento

Para el entrenamiento de los modelos predictivos se utilizó la librería *GridSearchCV* de Python, con el objetivo de optimizar los hiperparámetros de cada algoritmo mediante validación cruzada estratificada. En todos los casos se optó por mantener un *class_weight* balanceado, lo que permitió compensar la desproporción entre clases y mejorar la sensibilidad del modelo ante casos de fracaso académico. Las configuraciones óptimas obtenidas se muestran a continuación en la tabla II.

TABLA II HIPERPARÁMETROS ÓPTIMOS PARA CADA MODELO

Algoritmo	Hiperparámetros
Regresión logística	***C=0.1**:* Controla la regularización del modelo. ***max_iter=1000**:* Establece el número máximo de iteraciones para que el algoritmo converja, especialmente importante cuando se usa regularización. ***penalty='l1'**:* Aplica regularización L1 *(Lasso),* que favorece la selección de variables al generar coeficientes cero para atributos menos relevantes. ***solver='liblinear'**:* Algoritmo de optimización adecuado para problemas pequeños y con penalización *L1.*
Árbol de decisión	***criterion='entropy'**:* Utiliza la ganancia de información como criterio de división, lo que permite una segmentación más informativa. ***max_depth=5:*** Limita la profundidad del árbol para evitar el sobreajuste y mejorar la interpretabilidad. ***min_samples_leaf=1**:* Establece el número mínimo de muestras requeridas en una hoja, permitiendo mayor granularidad. ***min_samples_split=2:*** Número mínimo de muestras necesarias para dividir un nodo, configurado en su valor más bajo para máxima flexibilidad.
Random Forest	***bootstrap=True**:* Habilita el muestreo con reemplazo, esencial para la construcción de múltiples árboles independientes. ***max_depth=10:*** Controla la profundidad máxima de cada árbol, reduciendo el riesgo de sobreajuste. ***max_features='sqrt'**:* Selecciona aleatoriamente la raíz cuadrada del número total de características en cada división, promoviendo diversidad entre árboles. ***min_samples_leaf=1**:* Permite hojas con una sola muestra, útil para capturar patrones específicos. ***min_samples_split=2**:* Permite dividir nodos con al menos dos muestras, maximizando la capacidad de aprendizaje. ***n_estimators=300:*** Número de árboles en el bosque, aumentando la estabilidad y precisión del modelo.

IV RESULTADOS

Una de las métricas clave para evaluar el desempeño de los modelos predictivos fue la curva *ROC*, la cual permite visualizar la capacidad de los algoritmos para distinguir entre estudiantes en riesgo de fracaso académico y aquellos que no lo están. La **curva ROC** -*Receiver Operating Characteristic*- es un gráfico estadístico que evalúa el rendimiento de un modelo de clasificación binaria o prueba diagnóstica. Representa la sensibilidad (Verdaderos Positivos) frente a la tasa de falsos positivos (1- Especificidad) en todos los umbrales de corte posibles [34].

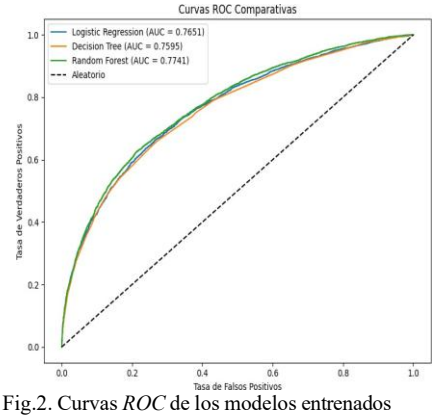


Fig.2. Curvas *ROC* de los modelos entrenados

En la figura 2 se presentan las curvas *ROC* comparativas para los tres modelos entrenados: Regresión Logística, Árbol de Decisión y *Random Forest*.

La curva *ROC* grafica la tasa de verdaderos positivos frente a la tasa de falsos positivos, y su desempeño se resume mediante el área bajo la curva (*AUC*). Cuanto más se acerque la curva al vértice superior izquierdo, mejor será la capacidad del modelo para realizar predicciones correctas. Los resultados obtenidos fueron los siguientes:

--*Random Forest*: *AUC* = 0.7741, el mejor desempeño entre los modelos evaluados.

--Regresión logística: *AUC* = 0.7651, con una capacidad predictiva sólida y buena interpretabilidad.

--Árbol de decisión: *AUC* = 0.7595, ligeramente inferior, pero con reglas de decisión fácilmente comprensibles.

La línea diagonal discontinua sirve como punto de comparación para evaluar si el modelo tiene capacidad real de discriminación. Cualquier *AUC* superior a 0.5 indica que el modelo predice mejor que el azar. El hecho de que todos los modelos superen significativamente esta línea confirma su utilidad para la detección temprana del fracaso académico. Cuanto más se acerque a 1, más discriminativo será, y cuanto más se aproxime a 0,5, menos discriminativo será. De acuerdo con Martin-Pérez y Pérez Martin [34] una curva con un valor de 0,75 se considera que tiene una capacidad discriminativa media. Además de la curva *ROC*, se evaluó el desempeño de los modelos mediante sus respectivas matrices de confusión, las cuales permiten analizar la capacidad de clasificación binaria en términos de verdaderos positivos (VP), falsos positivos (FP), verdaderos negativos (VN) y falsos negativos (FN). (Ver tabla III).

TABLA III. MATRIZ DE CONFUSIÓN

Real	VP (cuadrante I)	FP (cuadrante II)
	VN (cuadrante III)	FN (cuadrante IV)
	Predicción	

De acuerdo Murel y Kavlakoglu [35], una plantilla de matriz de confusión estándar para un clasificador binario está conformada de la siguiente manera: «*el cuadrante I proporciona el número de verdaderos positivos (VP), que es el número de predicciones correctas para la clase positiva. El cuadrante II representa el número de falsos negativos (FN), los casos positivos reales que se predijeron erróneamente como negativos. El cuadrante III corresponde a los falsos positivos (FP), aquellos casos de clase negativa identificados incorrectamente como casos positivos (errores estadísticos tipo I. El cuarto cuadrante muestra el número de verdaderos negativos (VN), que son las instancias de clase negativa reales que se predijeron con precisión como negativas. La suma de cada uno de estos valores proporcionaría el número total de predicciones del modelo*». En la figura 3 se pueden observar estos resultados de los 3 modelos*: Logistic regression, decisión tree* y *random forest.RE*

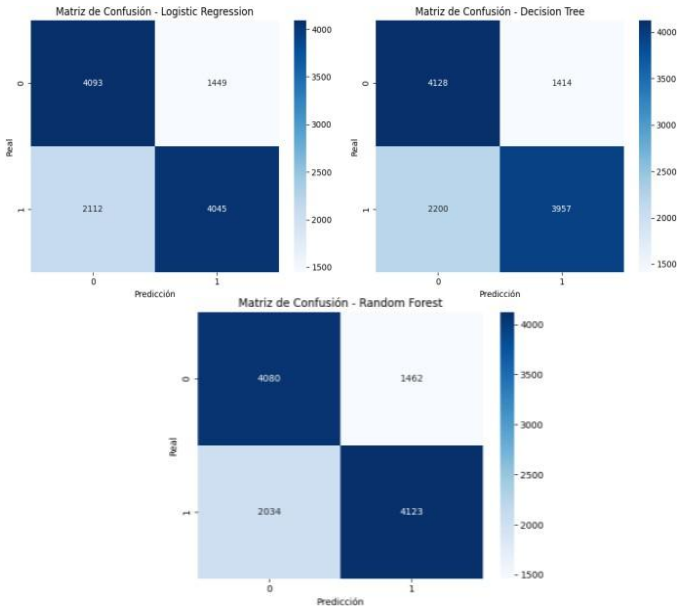


Fig. 3. Resultados de los modelos *Logistic regression, decisión tree* y *random forest.RE*

Estos resultados muestran que el modelo de *Random Forest* logra el mayor número de verdaderos positivos (4,123), lo que indica una mejor capacidad para identificar correctamente a los estudiantes en riesgo. Aunque presenta un número ligeramente mayor de falsos positivos (1,462), este modelo también reduce los falsos negativos (2,034), en comparación con los falsos negativos de los modelos *logistic regression* (2112) y los falsos negativos del modelo decisión tree (2200), lo cual es crucial en el contexto de intervención temprana, ya que permite actuar sobre más casos reales de riesgo académico y así tomar acciones correctivas a tiempo (ver figura 3). En conjunto, los resultados confirman que *Random Forest* ofrece el mejor equilibrio entre sensibilidad y precisión, siendo el modelo más adecuado para la implementación de un sistema de alerta temprana en la UTP.

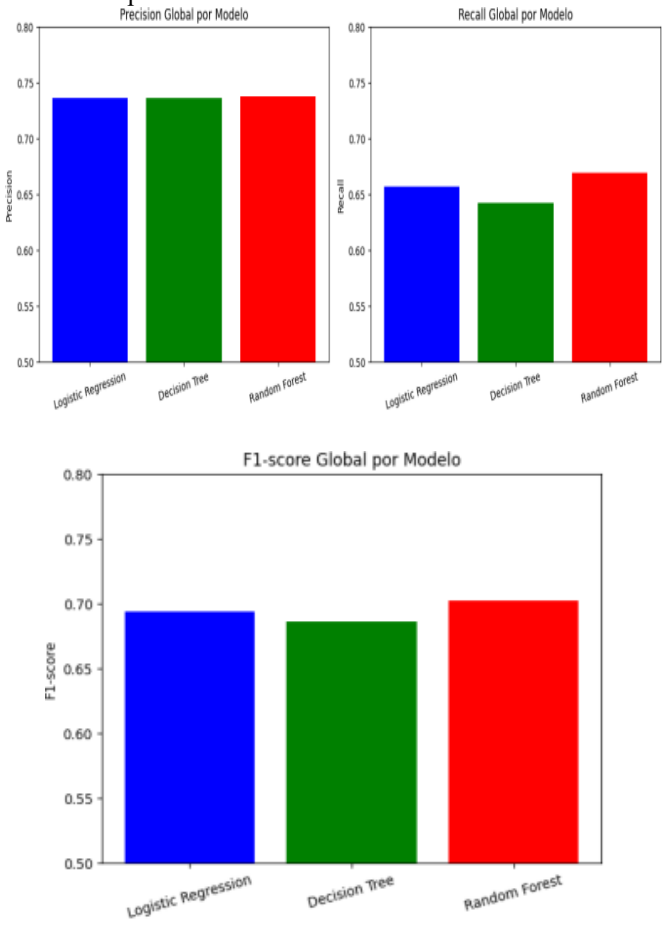


Fig.4. Barras de *Precision, Recall & F1-score.*

--Precisión global por modelo: El primer gráfico compara la precisión de los tres modelos evaluados. La precisión mide la proporción de predicciones positivas que fueron realmente correctas, es decir, cuántos de los estudiantes identificados como en riesgo efectivamente lo estaban. En este caso, los tres modelos presentan valores similares, cercanos al 75 %, lo que indica que, en general, cuando los modelos predicen riesgo, suelen acertar. Sin embargo, se observa que la regresión logística tiene una ligera desventaja respecto a los otros dos, lo que podría deberse a su naturaleza lineal y menor capacidad para capturar relaciones complejas entre variables (Ver figura 4).

--Sensibilidad global por modelo (*Recall*): El segundo gráfico muestra el desempeño de los modelos en términos de sensibilidad, también conocida como *recall*. Esta métrica es especialmente relevante en el contexto de alerta temprana, ya que mide la capacidad del modelo para identificar correctamente a los estudiantes que realmente están en riesgo. Aquí, el *Random Forest* se destaca con un valor cercano al 70 %, lo que indica que logra detectar

una mayor proporción de casos positivos. La regresión logística se ubica en segundo lugar, mientras que el árbol de decisión presenta el valor más bajo. Esto sugiere que, aunque los tres modelos tienen precisión similar, el *Random Forest* es más eficaz para evitar falsos negativos, lo cual es crucial para la intervención educativa oportuna (Ver figura 4).

--*F1-score* global por modelo: El tercer gráfico presenta el *F1-score*, una métrica que combina precisión y sensibilidad en una sola medida armónica. Esta métrica es útil cuando se busca un equilibrio entre identificar correctamente los casos positivos y evitar falsas alarmas. Nuevamente, el *Random Forest* lidera con un *F1-score* cercano al 70 %, lo que confirma su superioridad en términos de rendimiento global. La regresión logística le sigue de cerca, mientras que el árbol de decisión queda ligeramente rezagado. Estos resultados refuerzan la idea de que, aunque todos los modelos son funcionales, el *Random Forest* ofrece el mejor balance entre aciertos y cobertura, siendo el más adecuado para la implementación de un sistema de alerta temprana en la semana ocho del semestre (Ver figura 4).

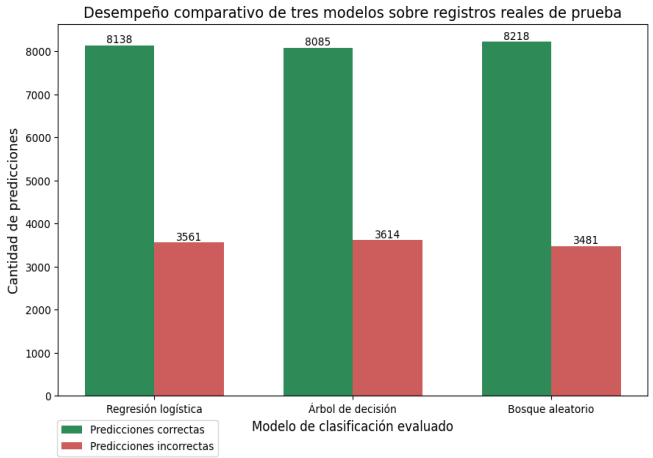


Fig.5. Desempeño comparativo de los tres modelos sobre 90 registros reales de prueba

La figura 5 presenta un resumen visual del desempeño de los tres modelos de clasificación aplicados sobre una muestra de 90 registros reales, siendo estos los modelos de regresión logística, árbol de decisión y bosque aleatorio. Se comparan las cantidades de predicciones correctas e incorrectas, lo que permite observar con claridad las diferencias de rendimiento entre ellos. El modelo de *Random Forest (bosque aleatorio)* destaca por tener la mayor cantidad de aciertos (66) y la menor de errores (24), seguido por el árbol de decisión con 56 aciertos y 34 errores y, finalmente, la regresión logística con 66 aciertos y 24 errores. Esta representación complementa los análisis anteriores al ofrecer una perspectiva directa sobre la eficacia práctica de cada algoritmo en condiciones reales. Para el modelo *Random Forest* la relación porcentual corresponde a 73.33 % (aciertos) frente a 26.66 % errores. El segundo lugar lo ocupa el modelo de árbol de decisión con una relación porcentual de 62.22 % (aciertos) frente a 37.77% (errores) y el tercer lugar lo ocupa el modelo de relación logística con 52.22% (aciertos) frente a 47.78% (errores).

En la figura 6 se probaron los modelos con todos los datos de prueba para comprobar el comportamiento con grandes cantidades de datos.

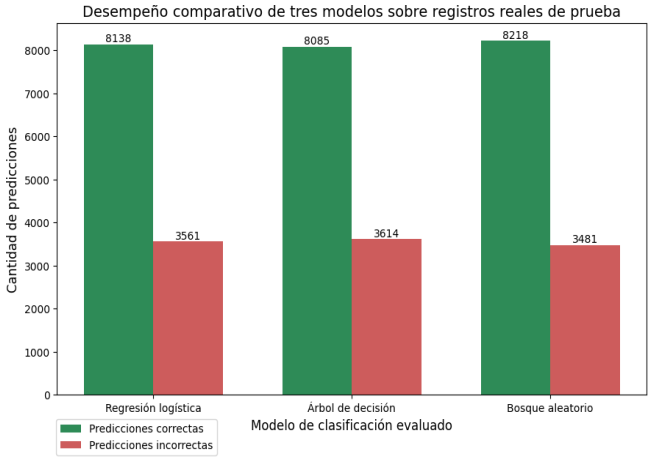


Fig.6. Desempeño comparativo de los tres modelos sobre todos los registros reales de prueba.

A diferencia de la figura 5, en la figura 6 se puede observar claramente que los tres modelos logran predecir correctamente alrededor de 8,000 registros, lo que representa una mejora significativa en la estabilidad de los resultados al aumentar el tamaño de la muestra.

El modelo de *Random Forest* se mantiene como el más preciso, superando ligeramente a los demás con 8,218 aciertos frente a 8,138 de la regresión logística y 8,055 del árbol de decisión. Esta consistencia en el rendimiento sugiere que, incluso con un conjunto de datos más amplio, el *Random Forest* conserva su ventaja predictiva, aunque el margen de diferencia entre los modelos se reduce. Además, se evidencia que todos los modelos mantienen una proporción relativamente estable entre aciertos y errores, lo que refuerza la validez del enfoque metodológico utilizado. De forma complementaria se puede apreciar que el modelo *Random Forest* es el que tiene el valor más bajo de errores con 3481, seguido del modelo de regresión logística con 3561 errores y el tercer lugar es el modelo de árbol de decisión con 3614 errores.

Con base a las variables identificadas como las más relevantes para la predicción del modelo y las necesidades institucionales en la figura 7, se ha diseñó una base de datos relacional para optimizar el proceso de predicción de fracaso y de esta manera facilitar todo el proceso de entrenamiento semestralmente.

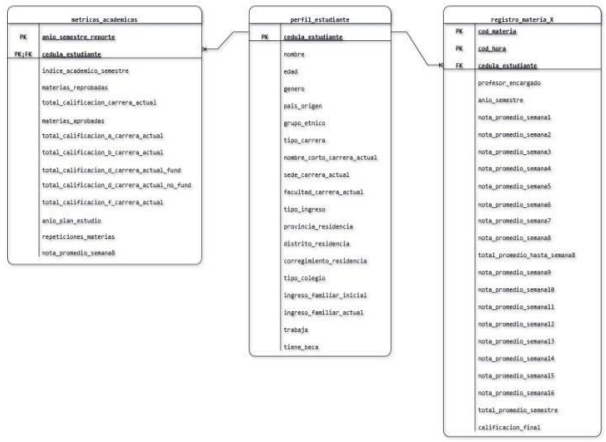


Fig.7. Diseño de base de datos relacional para el control de notas, estudiantes y métricas.

El diseño presentado en la figura 7, contempla tres tablas principales: perfil_estudiante, registro_materia_X y metricas_academicas, las cuales se integran mediante claves únicas para cada estudiante. En cumplimiento con los lineamientos internos de los sistemas universitarios [31], se ha adoptado la cédula del estudiante como llave primaria, dado que este identificador es utilizado oficialmente como referencia principal en los sistemas administrativos de la UTP. La cédula fue utilizada sin vinculación con el nombre del estudiante para garantizar el proceso de anonimización de los datos de acuerdo a las Políticas de privacidad establecidas y Uso del App de la UTP [36] en conjunto con la Ley 81 del 2019 para la protección de datos [37] y la normative de la Oficina de Protección de Datos de la Procuraduría de la Administración [38]. La cédula no se difundió, ni se publicó en los reportes emitidos. La tabla perfil_estudiante almacena información sociodemográfica relevante para el análisis de riesgo, mientras que registro_materia_X permite el seguimiento semanal del rendimiento académico por asignatura. Finalmente, metricas_academicas consolida indicadores clave como el índice académico general, número de materias reprobadas y promedio ponderado. Esta estructura facilita la implementación de modelos predictivos que, a partir de la octava semana del semestre, identifican a estudiantes con alta probabilidad de fracaso académico, habilitando intervenciones tempranas por parte de las autoridades educativas.

V Discusión

El modelo de regresión logística se caracteriza por su simplicidad y capacidad interpretativa. En este estudio, obtuvo un *AUC* de 0.7651, lo que indica un desempeño aceptable en la discriminación entre clases. Según la matriz de confusión, logró 4,045 verdaderos positivos y 4,093 verdaderos negativos, aunque también presentó 1,449 falsos positivos y 2,112 falsos negativos, lo que refleja una precisión moderada. En el conjunto reducido de prueba, identificó correctamente a 47 de los 90 estudiantes, mientras que en el conjunto grande logró 8,138 aciertos de 11,699 registros (69.56%). Aunque es sensible a relaciones no lineales y a la calidad de los datos, su transparencia lo convierte en una herramienta útil para explicar decisiones institucionales.

El árbol de decisión mostró un desempeño ligeramente inferior en términos de *AUC* obtuvo 0.7595, menor que el modelo de regresión logística, pero logró una estructura jerárquica comprensible para usuarios no técnicos. En la matriz de confusión, obtuvo 3,957 verdaderos positivos y 4,128 verdaderos negativos, con 1,414 falsos positivos y 2,200 falsos negativos. En el conjunto reducido tuvo 56 aciertos de 90 registros, y en el conjunto grande 8,085 aciertos de 11,699 registros de prueba (69.11%), superado levemente por la regresión logística. Aunque puede sobreajustarse si no se controla su profundidad, en este caso se logró un equilibrio adecuado entre interpretabilidad y rendimiento.

El modelo de *Random Forest* fue el que presentó el mejor desempeño general. Su *AUC* fue de 0.7741, el más alto entre los tres modelos. En la matriz de confusión, alcanzó 4,123 verdaderos positivos y 4,080 verdaderos negativos, con 1,462 falsos positivos y 2,034 falsos negativos. En el conjunto reducido logró 66 aciertos de 90 registros, y en el conjunto grande 8,218 aciertos (70.25%), siendo el más preciso. Además, destacó en las métricas de *recall* y *F1-score*, lo que evidencia su capacidad para capturar patrones complejos. Aunque su interpretabilidad es menor, su robustez lo posiciona como el candidato más adecuado para implementar un sistema de alerta temprana en la Universidad Tecnológica de Panamá.

Como consecuencia de los resultados obtenidos, se considera oportuno presentar la siguiente línea de acciones:

--Modelo predictivo principal y complementarios: Se recomienda implementar el modelo de *Random Forest* como base para el sistema de alerta temprana en la Universidad Tecnológica de Panamá, dada su superioridad en precisión y capacidad de generalización. Como herramienta auxiliar, se sugiere mantener el modelo de regresión logística, el cual puede servir de apoyo en la justificación de decisiones ante instancias institucionales. Para mejorar la interpretación de los resultados por parte de usuarios no técnicos, estos modelos pueden complementarse con visualizaciones de importancia de variables y explicaciones simplificadas.

---Repositorio institucional único: Crear un repositorio central donde se almacenen datos académicos validados y depurados, con reglas claras de control de calidad (validación de rangos, formatos y campos obligatorios), acceso seguro segmentado por perfiles y protocolos de respaldo y trazabilidad. En este sentido, la gestión y calidad de los datos debe revisarse regularmente para asegurar que los datos primarios para el análisis se encuentren de forma completa y sean coherentes.

---Diccionario de datos institucional: Estandarizar nombres de variables, formatos de fecha, escalas de notas e identificadores únicos, acompañado de un manual de referencia para garantizar uniformidad entre facultades y carreras. Ofrecer capacitaciones regulares al personal de las secretarias académicas de las facultades, coordinadores de carreras, personal de los vicedecanatos académicos y demás

gestores sobre la plataforma institucional para el manejo de los datos.

---Interoperabilidad de sistemas: Integrar los sistemas de notas, matrícula, becas y asistencia, evitando duplicidad de registros y así mejorar la eficiencia administrativa mediante los procesos de recolección y actualización de información.

---Notas y seguimiento: Incentivar a los docentes a reportar las calificaciones promedio por materia en la semana 8, a fin de identificar tempranamente a los estudiantes en riesgo de fracaso académico. Además, establecer como política académica institucional la actualización semanal o quincenal de calificaciones en el sistema, con alertas automáticas en caso de retrasos.

---**Automatización de reportes:** Diseñar *dashboards* institucionales (ej. Power BI o Looker Studio) que permitan visualizar en tiempo real el avance de notas, asistencia y desempeño por curso, evitando la dependencia de reportes manuales en Excel dispersos.

---Monitoreo de indicadores clave: Definir y dar seguimiento a métricas como el porcentaje de docentes con notas actualizadas, porcentaje de materias con vacíos de datos y tiempo promedio de carga de calificaciones, usándolos para auditoría interna y mejora continua. Potenciar el uso de pruebas o exámenes con calificaciones automáticas mediante las herramientas como cuestionarios y otras en las plataformas Moodle institucionales, sobre todo en aquellas asignaturas donde sean conveniente.

--Recursos humanos y capacitación: Como parte de las recomendaciones institucionales para la implementación del sistema de alerta temprana, se sugiere asignar al menos dos encargados de seguimiento del sistema: un estadístico o un analista de datos y un desarrollador con experiencia en Python y técnicas de *machine learning*. Asimismo, es fundamental ofrecer capacitaciones a docentes y administrativos sobre la importancia de la calidad de datos y su impacto en los procesos académicos, así como entrenamiento en el uso de las herramientas institucionales para la carga y gestión de información.

--Mantenimiento y actualización del modelo: Finalmente, se considera fundamental continuar evaluando y ajustando los modelos predictivos conforme se disponga de nuevos datos, asegurando así su vigencia, confiabilidad y efectividad en el tiempo. Para facilitar este proceso de reentrenamiento, es importante contar con los equipos adecuados que permitan gestionar eficientemente los datos, ajustar los algoritmos y validar los resultados obtenidos.

VI. CONCLUSIONES

El análisis comparativo entre los modelos de regresión logística, árbol de decisión y *Random Forest* revela diferencias significativas en cuanto a precisión, capacidad de interpretación y robustez. La regresión logística, aunque moderadamente precisa, destaca por su transparencia y facilidad de interpretación, lo que la hace útil para explicar decisiones institucionales. El árbol de decisión ofrece una estructura comprensible y un rendimiento ligeramente superior en conjuntos pequeños, pero es susceptible al sobreajuste si no se regula adecuadamente. Por su parte, el *Random Forest* demostró ser el modelo más eficaz, con el mayor número de aciertos tanto en conjuntos reducidos como amplios, además de sobresalir en métricas como el *AUC*, el *recall* y el *F1-score*. Su capacidad para capturar relaciones complejas entre variables lo convierte en la opción más robusta para tareas de predicción del riesgo de fracaso académico. Como trabajos futuros se recomienda validar el modelo con datos de otro semestre y realizar una prueba piloto en la universidad. Esto permitiría comprobar si las alertas generadas ayudan realmente a reducir la reprobación y mejorar el apoyo a los estudiantes.

AGRADECIMIENTOS

Los autores expresan su agradecimiento a la Universidad Tecnológica de Panamá (UTP) por el respaldo institucional brindado durante el desarrollo de este trabajo. Asimismo, se reconoce el valioso apoyo y orientación del Ing. Juan J. Morán S., director de Planificación Universitaria de la UTP.

REFERENCIA

[1] R. Quintero-Cortez, “Estudio de los factores de deserción a nivel superior de los estudiantes de primer ingreso de turno matutino de la carrera de Ingeniería Electromecánica de la Universidad Tecnológica de Panamá,” Universdiad de Panamá, 2024. [Online]. Available: https://up-rid.up.ac.pa/8233/

[2] D. Vega-Vega and R. Castro, “Relación de la autorregulación del aprendizaje en las dificultades académicas a nivel universitario The Relationship between Self-Regulated Learning and Academic Difficulties at the University Level,” *Rev. Latinoam. Calid. Educ.*, vol. 3, no. 2, pp. 154–165, 2026, doi: https://doi.org/10.70625/rlce/561.

[3] A. Pérez Aguirre, M. De Gracia, M. Aguirre, I. Ordas, and A. Berbey-Álvarez, “La Deserción En La Educación Superior En Panamá Y Sus Causas,” *Rev. Académica Creat. e Innovación en Educ.*, vol. 2, no. 1, pp. 41–55, 2023, doi: 10.47300/2953-3015-v2i1-04.

[4] A. Berbey-Alvarez, “Statistical analysis of the behavior of a subject in an engineering career,” *Proc. LACCEI Int. Multi-conference Eng. Educ. Technol.*, vol. 2019-July, no. September, 2019, doi: 10.18687/LACCEI2019.1.1.16.

[5] A. Labrador, M. Sánchez, L. Henríquez, and A. Berbey-Alvarez, “Percecepción de la educación superior en Panamá,” in *2022: Actas del VII Congreso de Investigación, Desarrollo e Innovación de la Universidad Internacional de Ciencia y Tecnología*, UNICYT, Ed., Panamá, República de Panamá: UniCyT, 2022, pp. 286–293. [Online]. Available: https://revistas.unicyt.org/index.php/actasidi-unicyt/article/view/64

[6] L. Arenales, V. Villarreal, and J. J. Saldana-barrios, “Predictive model based on artificial intelligence and contextual awareness to identify students at risk of dropping out of university in Panama,” *Ing. Solidar.*, vol. 2, no. 1, 2025, doi: 10.16925/2357-6014.2025.01.08.

[7] G. Spence-Diaz, “Factores psicosociales que intervienen en la permanencia de los estudiantes de la Licenciatura en Psicología,” *Horiz. académico*, vol. 5, no. 2, pp. 85–99, 2025, [Online]. Available: https://horizonteacademico.org/index.php/horizonte/article/view/122

[8] M. Luzuriaga-Pavón and S. Castillo-Paez, “Modelado del índice de deserción estudiantil en Carreras de Ingeniería en una Institución de Educación Superior,” *Rev. Cient. Multidiscip.*, vol. 6, no. 2, pp. 228–242, 2026, doi: 10.60100/rcmg.v6i2.733.

[9] J. Trujillo-gonzález and N. Cosme-Toribio, “Riesgo estructural y fracaso académico en asignaturas críticas de un programa

universitario de matemática," *Rev. oradores*, vol. 23, no. mayo, pp. 14–24, 2025, doi: 10.37594/oratores.n23.1987.

[10] Y. B. Guzman, U. Andina, U. Andina, L. V. Vera, and U. Andina, "Factores de riesgo para el abandono universitario: una revisión de la literatura," *Tribunal*, vol. 6, no. 14, pp. 847–864, 2026.

[11] C. Vejerano and J. Trujillo-González, "Veinte Años de Historia : Análisis de Graduados y Desertores en la Licenciatura en Matemática en Panamá," *Invest. Pens. Crit*, vol. 13, no. 2, pp. 52–60, 2025.

[12] Universidad Tecnológica de Panamá, "Objetivos estratégicos." Universidad Tecnológica de Panamá, Panamá, República de Panamá, p. 3, 2020. [Online]. Available: https://utp.ac.pa/documentos/2020/pdf/objetivos-estrategicos-pdi-2018-2030_.pdf

[13] Programa de las Naciones Unidas para el Desarrollo, "¿Qué son los Objetivos de Desarrollo Sostenible?," Los ODS en acción. Accessed: Aug. 18, 2023. [Online]. Available: https://www.undp.org/es/sustainable-development-goals#:~:text=de Desarrollo Sostenible%3F-,Los Objetivos de Desarrollo Sostenible (ODS)%2C también conocidos como,disfruten de paz y prosperidad.

[14] J. Loxton, "The Puzzle of panamanian exceptionalism," *J. Democr.*, vol. 33, no. 1, pp. 85–99, 2022, doi: 10.1353/jod.2022.0006.

[15] M. B. Herrero and A. Greco, "Investigación soberana para el buen gobierno en salud. Desafíos y oportunidades," *Debate Público*, vol. 14, no. 28, pp. 147–159, 2024, [Online]. Available: https://ri.conicet.gov.ar/handle/11336/252891

[16] J. B. Sokol, "Revisión de los Shocks en la Economía Review of Shocks in the Economy," *Invest. Pens. Crit*, vol. 11, no. 1, pp. 46–61, 2023, [Online]. Available: https://revistas.usma.ac.pa/ojs/index.php/ipc/article/view/345/532

[17] G. Barsallo-Alvarado, "Deserción Universitaria en Panamá en el Siglo XXI University Dropout in Panama in the 21st Century Abandon universitaire au Panama au XXIe siècle Deserção Universitária no Panamá no século XXI," *Contact*, vol. 5, no. 1, pp. 70–90, 2025, doi: 10.48204/contacto.v5n1.9001.

[18] P. Macías-Galeas, "Más allá de la tecnología: Gestión académica y sisitemas de alerta temprana en la retención estudiantil en la educación superior virtual," *Rev. cientiífica mutlidisciplinaria*, vol. 6, no. 11, pp. 127–137, 2025, doi: 10.56124/ubm.v6i11.014 jul.

[19] P. J. (2026). Euan-Catzin, A. J., & Canto-Herrera, "Análisis psicométrico de una escala para medir la intención de deserción universitaria : un estudio piloto Psychometric analysis of a scale to measure university dropout intention : A pilot study," *Rev. Panam. Pedagog.*, vol. 41, no. 2026, pp. 1–17, 2025, [Online]. Available: https://revistas.up.edu.mx/RPP/article/download/3588/2971

[20] P. Parrilla-Cañadas, "Análisis del abandono universitario a partir de datos académicos y de actividad," Universitat Politécnica de Valencia, 2025. [Online]. Available: https://riunet.upv.es/entities/publication/735867aa-7733-4396-9ebc-f69f56f66845

[21] F. Palma Bolaños, "Desarrollo de un modelo predictivo mediante el uso de algoritmos de aprendizaje supervisado que permita predecir el rendimiento académico de estudiantes de pregrado en base a factores socioeconómicos y psicológicos," Escuela Politécnica Nacional, 2025. [Online]. Available: https://bibdigital.epn.edu.ec/handle/15000/26287

[22] L. Polo-Ahumada and D. Medina-Reyes, "Modelo Analítico Para La Predicción De La Deserción Estudiantil A Nivel De Pregrado En La Universidad Autónoma Del Caribe," Universidad Autónoma del Caribe, 2022. [Online]. Available: https://repositorio.uac.edu.co/handle/11619/4137

[23] E. Mejía-Zamora, "Modelo matemático para predecir el grado de deserción de los estudiantes en el Instituto Superior Tecnológico Bolívar." Universidad Técnica de Ambato, Ecuador, Ambato, Ecuador, 2023. [Online]. Available: https://repositorio.uta.edu.ec/items/50b6a991-c7b4-44ed-b23c-40191ddad534

[24] Universidad Tecnológica de Panamá, "Lineamiento de ética para la investigación académica," Consejo de investigación, postgrado y extensión. Accessed: Sep. 27, 2025. [Online]. Available: https://utp.ac.pa/sites/default/files/documentos/2025/pdf/utp lineamiento-etica-investigacion-academica.pdf

[25] República de Panamá, "Ley 81 de 26 de marzo de 2019 sobre Protección de Datos Personales 2019." Gobierno de Panamá, Panamá, República de Panamá, 2019. [Online]. Available: http://www.gacetaoficial.gob.pa/pdfTemp/28743_A/GacetaNo_28743a%0A_20190329.pdf

[26] ANTAI, "Reglamentan Ley 81 de Protección de Datos Personales 2021." Gobierno de Panamá, Panamá, República de Panamá, 2021. [Online]. Available: https://www.antai.gob.pa/reglamentan%0Aley-81-de-proteccion-de-datos%0Apersonales/

[27] Universidad Tecnológica de Panamá, "Estatuto de la Universidad Tecnológica de Panamá," no. 1. Universidad Tecnológica de Panamá, Panamá, República de Panamá, pp. 1–74, 2025. [Online]. Available: https://utp.ac.pa/sites/default/files/documentos/2026/pdf/utp-estatuto-universitario-modificado-cgu-8-5-2025.pdf

[28] y M. P. P. Kravets, V. Pasichnyk, "Mathematical Model of Logistic Regression for Binary Classification. Part 1. Regression Models of Data Generalization," *J. Lviv Polytech. Natl. Univ. Inf. Syst. Netw.*, vol. 15, p. 290, 2024, [Online]. Available: https://www.researchgate.net/publication/384477565_Mathematical_Model_of_Logistic_Regression_for_Binary_Classification_Part_1_Regression_Models_of_Data_Generalization

[29] Y. Lu, T. Ye, and J. Zheng, "Decision Tree Algorithm in Machine Learning," in *2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, IEEE, Ed., Dalian, China: IEEE, 2022, pp. 1014–1017. doi: 10.1109/AEECA55500.2022.9918857.

[30] Universidad Tecnológica de Panamá, "Dirección General de Tecnología de la Información y Comunicaciones," Conoce la UTP. Accessed: Sep. 13, 2026. [Online]. Available: https://utp.ac.pa/direccion-general-de-tecnologia-de-la-informacion-y-comunicaciones

[31] Universidad Tecnológica de Panamá, "Políticas y Lineamientos para el Uso Aceptable de los Recursos de Tecnología de Información y Comunicaciones de la Universidad Tecnológica de Panamá." Universidad Tecnológica de Panamá, Panamá, República de Panamá, p. 9, 2009.

[32] Universidad Tecnológica de Panamá, "Secretaria General," Conoce la UTP. Accessed: Sep. 13, 2026. [Online]. Available: https://utp.ac.pa/secretaria-general

[33] Universidad Tecnológica de Panamá, "Matrícula UTP." Universidad Tecnológica de Panamá, Panamá, República de Panamá, 2026. [Online]. Available: https://matricula.utp.ac.pa/session/cuenta/validar/jigZlfXUf$XZTdndTa0fQnjbTQfm

[34] J. Marínez-Pérez and P. Pérez- Martin, "La curva ROC," *Form. Contin. Metodol. y técnicas*, vol. 49, no. 1, pp. 1–3, 2023, doi: 10.1016/j.semerg.2022.101821.

[35] J. Murel and E. Kavlakoglu, "¿Qué es una matriz de confusión?," Artificial intelligence. Accessed: Jun. 29, 2026. [Online]. Available: https://www.ibm.com/es-es/think/topics/confusion-matrix

[36] U. T. de Panamá, "Políticas de Privacidad y Uso del App UTP," App UTP. Accessed: Sep. 13, 2026. [Online]. Available: https://utp.ac.pa/politicas-de-privacidad-y-uso-del-app-utp

[37] Asamblea Nacional de Panamá, "Ley 81 de 26 de marzo de 2019. Sobre protección de datos personales." Gaceta Oficial Digital, Panamá, República de Panamá, p. 35, 2019. [Online]. Available: https://s3-legispan.asamblea.gob.pa/legispan/NORMAS/2010/2019/LEY/Administrador Legispan_28743-A_2019_3_29_ASAMBLEANACIONAL_81.pdf

[38] Procuraduría de la Administración. Ministerio Público., "Resolución N° PA/DS 020-2024 de 20 de febrero de 2024," no. 29991. Gaceta Oficial Digital, Panamá, República de Panamá, pp. 1–3, 2024. [Online]. Available: https://www.gacetaoficial.gob.pa/storage/gacetas/2024/03/29991/103391.pdf