

## Identificación temprana de riesgo académico en estudiantes de ingeniería mediante aprendizaje automático

### *Early identification of academic risk in engineering students using machine learning*

Servando Campillay Briones<sup>1\*</sup>  <https://orcid.org/0009-0004-3766-9442>

Andrés Alfaro Avalos<sup>1</sup>  <https://orcid.org/0000-0002-4989-6448>

Jacqueline Manríquez Barria<sup>1</sup>  <https://orcid.org/0009-0002-7043-996X>

Recibido 26 de agosto de 2025, aceptado 26 de diciembre de 2025

*Received: August 26, 2025      Accepted: December 26, 2025*

### RESUMEN

Este trabajo presenta el desarrollo de un modelo predictivo orientado a la identificación temprana de estudiantes en situación de pre-riesgo y riesgo académico en la Facultad de Ingeniería de la Universidad de Atacama. El objetivo es anticipar situaciones que puedan derivar en deserción o bajo rendimiento académico, permitiendo la implementación de medidas de apoyo oportunas y focalizadas.

Para ello, se trabajó con una muestra de 130 estudiantes de primer año de la carrera de Ingeniería Civil en Computación e Informática. La metodología se basó en la recopilación e integración de datos académicos extraídos de sistemas institucionales, considerando variables como asignaturas reprobadas, asignaturas eliminadas, nivel académico y avance curricular. Posteriormente, se aplicaron técnicas de aprendizaje automático para identificar patrones que permitieran predecir el riesgo académico.

Los resultados del modelo permitieron detectar con antelación a estudiantes en situación de pre-riesgo, lo que facilitó la implementación de intervenciones específicas, tales como la reducción de carga académica, tutorías personalizadas y seguimiento activo. Estas acciones contribuyeron a mejorar el rendimiento académico, aumentar el compromiso estudiantil y reducir los niveles de deserción.

La relevancia de este estudio radica en evidenciar cómo la aplicación de tecnologías de análisis predictivo, combinadas con estrategias de acompañamiento docente, puede fortalecer los procesos de retención y éxito académico desde las etapas iniciales de la formación universitaria.

Palabras clave: Riesgo académico, aprendizaje automático, acompañamiento estudiantil.

### ABSTRACT

*This study describes the development of a predictive model designed to identify students at pre-risk and academic risk stages within the Faculty of Engineering of the Universidad de Atacama. The primary objective is to anticipate factors that may contribute to dropout or low academic performance, thereby enabling the implementation of timely and targeted support interventions.*

*A sample of 130 first-year students from the Civil Engineering in Computer Science undergraduate program was used. The methodology involved collecting and integrating academic data from institutional systems, including variables such as failed courses, withdrawn courses, academic level, and curricular progress. Machine learning techniques were applied to identify patterns that could predict academic risk.*

<sup>1</sup> Universidad de Atacama. Departamento de Ingeniería Informática y Ciencias de la Computación. Copiapó, Chile.  
E-mail: [servando.campillay@uda.cl](mailto:servando.campillay@uda.cl); [andres.alfaro@uda.cl](mailto:andres.alfaro@uda.cl); [jacqueline.manriquezb@uda.cl](mailto:jacqueline.manriquezb@uda.cl)

\* Autor de correspondencia: [servando.campillay@uda.cl](mailto:servando.campillay@uda.cl)

*The model's results enabled the early detection of students in pre-risk situations, facilitating targeted interventions such as reduced academic workload, personalized tutoring, and active monitoring. These actions contributed to improved academic performance, increased student engagement, and reduced dropout rates.*

*This study demonstrates that integrating predictive analytics technologies with instructional support strategies can enhance retention and academic success from the early stages of university education.*

**Keywords:** Academic risk, machine learning, student support.

## INTRODUCCIÓN

La deserción estudiantil en educación superior constituye una problemática multifactorial que afecta significativamente tanto al desarrollo personal del estudiante como al desempeño institucional. Este fenómeno es especialmente crítico en carreras de alta exigencia académica, como las ingenierías, donde las tasas de abandono tienden a ser elevadas. En Chile, y particularmente en la Universidad de Atacama (UDA), se ha observado una tasa de deserción promedio del 25% durante los primeros tres semestres en programas de ingeniería, lo que evidencia la necesidad de estrategias de intervención temprana.

El bajo rendimiento académico ha sido identificado como uno de los principales predictores de la deserción [1], y no se limita únicamente a calificaciones deficientes, sino que refleja un proceso acumulativo asociado a reprobaciones reiteradas, retraso curricular y desconexión con el entorno educativo. Ante este escenario, las instituciones de educación superior han comenzado a implementar enfoques preventivos basados en el análisis de datos, con el objetivo de identificar y apoyar oportunamente a estudiantes en situación de vulnerabilidad académica [2], [3].

En este contexto, la minería de datos y el aprendizaje automático se han consolidado como herramientas clave para el desarrollo de modelos predictivos capaces de detectar señales tempranas de riesgo. Estas tecnologías permiten transformar grandes volúmenes de información académica en conocimiento útil para orientar acciones concretas de acompañamiento estudiantil [4].

A nivel internacional, la analítica del aprendizaje ha demostrado su efectividad en la mejora de los procesos de retención y éxito académico, promoviendo una cultura institucional basada en la prevención y la toma de decisiones informadas [5].

Este artículo presenta el diseño y validación de un modelo predictivo basado en técnicas de aprendizaje automático para identificar estudiantes en situación de pre-riesgo y riesgo académico en la carrera de Ingeniería Civil en Computación e Informática de la UDA. El propósito es contribuir al desarrollo de mecanismos institucionales de alerta temprana que permitan una intervención más oportuna, personalizada y basada en evidencia.

## ESTADO DEL ARTE

### Riesgo académico y deserción

La deserción estudiantil en educación superior es una problemática compleja, determinada por una combinación de factores académicos, personales y socioeconómicos. En carreras de alta exigencia, como las ingenierías, estas dificultades se intensifican, generando tasas de abandono superiores al promedio. El riesgo académico se entiende como la probabilidad de que un estudiante no alcance los resultados esperados debido a reprobaciones reiteradas, bajo rendimiento sostenido, desconexión emocional con el proceso formativo y limitada participación [1].

[2] describe la deserción como el desenlace de un proceso de desintegración académica y social. En este marco, identificar de forma temprana los factores de riesgo se vuelve clave para evitar el abandono mediante intervenciones oportunas.

### Modelos predictivos en educación superior

El avance del aprendizaje automático (machine learning) ha permitido desarrollar modelos predictivos con alta precisión para anticipar el riesgo académico. Estos modelos pueden agruparse en:

- Modelos de regresión, como la regresión logística, que permiten estimar la probabilidad de deserción en función de variables institucionales.

- Modelos basados en árboles, como los árboles de decisión y Random Forest, reconocidos por su interpretabilidad y robustez frente a datos categóricos y continuos.
- Modelos híbridos o multivariados, como redes neuronales y arquitecturas como DMSW, que integran datos académicos, contextuales y de comportamiento digital.

Una revisión sistemática reciente destaca que el algoritmo Random Forest logra niveles de precisión superiores al 95% para predecir abandono en educación superior [6]. Asimismo, [7] proponen un modelo denominado Dual-Modal Multiscale Sliding Window (DMSW), que incorpora tanto rendimiento académico como comportamiento en plataformas digitales, alcanzando mejoras del 15% en precisión respecto a métodos tradicionales.

### **Intervención educativa basada en datos**

El uso de analítica del aprendizaje (Learning Analytics) y minería de datos educativos (Educational Data Mining) permite a las instituciones transformar los registros académicos en acciones concretas de apoyo estudiantil [8]. Estas tecnologías no reemplazan la labor docente, sino que la complementan al proporcionar evidencias empíricas para la toma de decisiones pedagógicas [9].

Además, diversos estudios han demostrado la efectividad de modelos predictivos para la identificación temprana de estudiantes en riesgo académico. En este contexto, [10] propone modelos de predicción temprana basados en el desempeño académico y sistemas de evaluación continua, evidenciando que es posible anticipar el riesgo dentro de un curso mediante variables de seguimiento formativo.

### **Precedentes en Chile y Latinoamérica**

En Chile, la deserción estudiantil ha sido reconocida como un problema prioritario, especialmente en los primeros años de la educación superior. Diversos informes oficiales, como los del Ministerio de Educación, muestran que cerca del 30% de los estudiantes abandona sus estudios durante el primer año en carreras universitarias, y las tasas son aún más altas en programas de ingeniería [11]. Sin embargo, a pesar de la gravedad del fenómeno, la implementación de modelos predictivos automatizados en universidades chilenas sigue siendo limitada.

Estudios nacionales han abordado la deserción desde perspectivas cualitativas o mediante análisis estadísticos tradicionales. Por ejemplo, [12] aplicó modelos de duración para estudiar los factores de abandono en ingeniería, destacando la relevancia de las variables socioeconómicas y de rendimiento. En [13] analizaron la deserción desde la perspectiva de la responsabilidad social universitaria, proponiendo estrategias de retención, pero sin aplicar modelos predictivos automatizados.

Más recientemente, se han realizado esfuerzos por integrar la analítica de datos en la gestión educativa. En la Universidad de Chile, por ejemplo, [14] exploran enfoques basados en analítica de datos y aprendizaje automático para el análisis y predicción del rendimiento académico, como apoyo a la toma de decisiones educativas. De forma similar, [15] proponen modelos de predicción del rendimiento académico que incorporan la heterogeneidad de los estudiantes, mejorando la capacidad de los modelos para adaptarse a distintos perfiles de aprendizaje.

En el contexto latinoamericano, algunos países han avanzado en el uso de técnicas de minería de datos educativa. En Venezuela, [16] aplicó modelos de clasificación para analizar la deserción en ingeniería, y en Colombia, [17] trabajaron con un enfoque institucional de riesgo académico, aunque con escasa integración tecnológica. En Latinoamérica, [18] analizan los factores explicativos de la deserción universitaria mediante técnicas de inteligencia artificial, evidenciando el potencial de estos enfoques para comprender y predecir el abandono estudiantil.

Estos antecedentes revelan una brecha metodológica y tecnológica importante: mientras que en países como España o Estados Unidos ya existen sistemas de alerta implementados y operativos, en Chile y gran parte de América Latina los modelos predictivos basados en machine learning siguen en etapa experimental. Este rezago es especialmente notorio en instituciones regionales o con menos recursos, donde el problema de la deserción es más agudo, pero las capacidades técnicas para abordarlo aún son limitadas.

### **Contribución del estudio**

Pese al crecimiento del uso de inteligencia artificial en la educación superior internacional, la literatura

nacional presenta una limitada aplicación de modelos predictivos basados en aprendizaje automático en contextos universitarios regionales.

Este estudio contribuye a llenar un vacío en la UDA mediante la implementación de un modelo predictivo aplicado a estudiantes de Ingeniería Civil en Computación e Informática, con el objetivo de fortalecer el acompañamiento académico desde una perspectiva basada en datos, ética y centrada en el bienestar estudiantil.

## METODOLOGÍA

La presente investigación adopta un enfoque cuantitativo, predictivo y aplicado, orientado al desarrollo de un modelo basado en técnicas de aprendizaje automático supervisado, con el propósito de anticipar el riesgo académico en estudiantes de primer año de la carrera de Ingeniería Civil en Computación e Informática de la Universidad de Atacama.

El conjunto de datos utilizado proviene del Sistema U+, plataforma oficial de gestión académica de la Universidad de Atacama, que concentra información académica, administrativa y socioeconómica de todos los estudiantes, garantizando trazabilidad y calidad en los registros.

La utilización de modelos predictivos ha demostrado ser eficaz para identificar de forma temprana a estudiantes en situación de vulnerabilidad académica, permitiendo a las instituciones implementar medidas de apoyo personalizadas y oportunas, respaldadas por evidencia proveniente de modelos predictivos basados en aprendizaje automático [19], [8].

### Fuentes de datos y características

El conjunto de datos fue obtenido a partir del Sistema U+, plataforma oficial de gestión académica de la Universidad de Atacama. Este sistema centraliza y administra de manera integrada la información académica, administrativa y socioeconómica de todos los estudiantes, asegurando la trazabilidad, consistencia y calidad de los registros.

Para este estudio, se extrajeron registros correspondientes a estudiantes de ingeniería entre los años 2018 y 2023, con un total de 1.245 registros únicos. Las variables incluyeron:

Académicas: historial de calificaciones, asignaturas cursadas y aprobadas, reprobaciones, asistencia y avance curricular.

Administrativas: estado de matrícula, tipo de financiamiento, régimen de estudios.

Sociodemográficas: edad, género, región de procedencia.

La variable objetivo fue definida como riesgo académico (1 = riesgo, 0 = no riesgo), según criterios institucionales basados en desempeño académico y alertas generadas por el área de apoyo estudiantil.

### Preprocesamiento de datos

El preprocesamiento se desarrolló en varias etapas para garantizar la calidad y consistencia del conjunto de datos:

- Limpieza de datos: Eliminación de registros duplicados y Descarte de casos con más del 30% de información faltante.
- Tratamiento de valores faltantes: Variables numéricas: imputación con mediana, Variables categóricas: imputación con moda.
- Codificación de variables: Variables categóricas nominales, Variables ordinales.
- Escalado de variables: Aplicación de normalización Min-Max en variables numéricas para algoritmos sensibles a magnitudes.

### Selección y justificación de algoritmos

Se evaluaron diferentes algoritmos de aprendizaje supervisado para clasificación binaria:

- Regresión logística (baseline interpretable).
- Árboles de decisión.
- Random Forest.
- XGBoost.
- Support Vector Machines (SVM).

La inclusión de modelos más complejos como XGBoost y SVM respondió a la necesidad de evaluar mejoras en rendimiento respecto a métodos estándar. La regresión logística se mantuvo como referencia por su facilidad de interpretación y utilidad para la toma de decisiones institucionales.

### Diseño metodológico y objetivos

La investigación se estructuró en torno a cinco objetivos específicos (OE), que definieron cada fase del estudio:

- OE1. Identificar variables relevantes al riesgo académico.
- OE2. Construir el modelo predictivo
- OE3. Validar el modelo predictivo.
- OE4. Interpretar patrones y reglas de decisión.
- OE5. Difundir resultados y proponer intervención.

Se desarrollaron instancias de retroalimentación con equipos académicos para presentar los hallazgos del modelo y generar un plan de implementación. Se propuso la creación de un Plan de Seguimiento Estudiantil basado en alertas tempranas, orientado a ofrecer tutorías, ajustes en carga académica y orientación personalizada.

Con base en estas evidencias, se propuso la creación de un Plan de Seguimiento Estudiantil centrado en alertas tempranas. Este plan contempla:

- Contacto proactivo con estudiantes identificados en pre-riesgo o riesgo.
- Reducción estratégica de carga académica.
- Tutorías personalizadas por nivel y área crítica.
- Seguimiento continuo a través del sistema institucional de gestión estudiantil.

### Hipótesis de investigación

*“La aplicación de un modelo predictivo basado en aprendizaje automático permite identificar tempranamente a estudiantes en situación de riesgo académico, contribuyendo a reducir las tasas de deserción durante los primeros años de formación universitaria”.*

Esta hipótesis se fundamenta en investigaciones previas que han demostrado que los modelos predictivos bien diseñados pueden alcanzar altos niveles de precisión y que, cuando se integran con estrategias institucionales de acompañamiento, mejoran el rendimiento académico y la retención estudiantil [7], [6].

### Variables académicas utilizadas

Las variables empleadas para alimentar el modelo predictivo fueron extraídas de los registros académicos institucionales, y seleccionadas en función de su relevancia teórica y empírica en la literatura sobre riesgo académico se muestra en la Tabla 1.

Además, se generaron variables derivadas como la brecha curricular (nivel ideal - nivel actual) y se codificaron categorías para facilitar el entrenamiento de los algoritmos. Estas variables permitieron identificar umbrales asociados a pre-riesgo (retraso leve o patrón de reprobación) y riesgo alto (retraso severo y múltiples reprobaciones), sobre los cuales se basó la lógica del modelo.

La metodología empleada permite articular la tecnología con una perspectiva pedagógica y de gestión académica. El modelo predictivo no solo busca predecir un resultado, sino apoyar decisiones institucionales que favorezcan la permanencia estudiantil, reforzando la idea de que el uso ético y contextualizado de los datos puede generar beneficios concretos en los procesos formativos.

### Análisis descriptivo de las variables

Antes de avanzar en la construcción del modelo predictivo, se llevó a cabo un análisis exploratorio de las variables disponibles. Este paso fue clave para comprender mejor el perfil académico de los estudiantes incluidos en la muestra y detectar patrones que pudieran estar relacionados con el riesgo de abandono. Más que un ejercicio estadístico, se trató de mirar los datos con una mirada crítica y empática, intentando identificar señales tempranas de dificultades que muchas veces pasan desapercibidas.

El análisis inicial consideró la cantidad de asignaturas inscritas por semestre. Aunque el promedio se situó

Tabla 1. Descripción de las variables.

| Variable               | Descripción                                                             |
|------------------------|-------------------------------------------------------------------------|
| Asignaturas cursadas   | Total, de asignaturas inscritas en el semestre vigente.                 |
| Asignaturas reprobadas | Número acumulado de asignaturas no aprobadas desde el ingreso.          |
| Oportunidades por ramo | Cantidad de veces que una asignatura ha sido cursada.                   |
| Asignaturas eliminadas | Asignaturas retiradas voluntariamente (según reglamento institucional). |
| Año de ingreso         | Año en que el estudiante ingresó formalmente a la carrera.              |
| Nivel actual           | Semestre en que el estudiante se encuentra actualmente.                 |

Fuente: Elaboración propia.

entre X e Y ramos, se observaron casos atípicos de sobrecarga académica o inscripciones mínimas. Estas variaciones reflejan que la trayectoria estudiantil no siempre sigue un patrón lineal y puede estar influida por factores personales, económicos o de salud.

La cantidad de asignaturas reprobadas reveló que un número significativo de estudiantes arrastra al menos una materia desde los primeros semestres, lo que afecta tanto el avance curricular como la motivación, y puede aumentar el estrés académico. Si no se detecta a tiempo, el cúmulo de reprobaciones puede convertirse en una barrera difícil de revertir.

Por último, la eliminación voluntaria de asignaturas apareció como una conducta frecuente. Aunque esta acción está contemplada en la normativa y puede funcionar como estrategia de autorregulación frente a una carga excesiva, también constituye una señal de alerta que requiere seguimiento institucional.

El año de ingreso permitió distinguir diferentes cohortes estudiantiles y comparar su evolución a lo largo del tiempo, identificando posibles diferencias generacionales o efectos de cambios en el currículo o contexto institucional. Finalmente, las variables nivel actual y nivel ideal fueron fundamentales para calcular el desfase académico, es decir, cuántos semestres de diferencia existen entre el avance esperado y el real de cada estudiante. Este indicador se consolidó como uno de los más relevantes, ya que un mayor desfase suele asociarse con un mayor riesgo de abandono o prolongación de la carrera.

### Construcción del modelo predictivo

Tras el análisis descriptivo, se procedió a construir un modelo predictivo capaz de identificar estudiantes en riesgo académico. Para ello, se utilizaron algoritmos de aprendizaje automático supervisado, seleccionados por su eficacia en contextos educativos y por su capacidad de interpretación.

Se trabajó con tres algoritmos principales:

- Árboles de decisión.
- Random Forest (bosques aleatorios).
- Regresión logística.

Estos modelos fueron elegidos no solo por su rendimiento, sino también por su facilidad para

ser comprendidos e interpretados por equipos no técnicos, como autoridades académicas o unidades de apoyo estudiantil.

### Proceso de modelado

Los datos se dividieron en dos conjuntos:

- 70% para entrenamiento del modelo
- 30% para validación y prueba

Esto permitió evaluar la capacidad del modelo para generalizar sus predicciones a nuevos casos.

### Métricas de evaluación

Para medir el rendimiento de los modelos se utilizaron métricas estándar:

- Precisión: porcentaje de predicciones correctas.
- Recall: capacidad para identificar correctamente a los estudiantes en riesgo.
- F1-score: balance entre precisión y recall.
- Matriz de confusión: muestra los verdaderos positivos, falsos positivos, etc.

## PROCESO DE TRABAJO

### Diagrama de trabajo modelo predictivo

La Figura 1 ilustra el flujo general del proceso de construcción de un modelo predictivo de riesgo académico basado en aprendizaje automático. El proceso inicia con la recopilación de datos académicos relevantes de los estudiantes. Posteriormente, se realiza el preprocesamiento de dichos datos, etapa en la que se limpian, transforman y, en algunos casos, se generan representaciones visuales o matriciales de la trayectoria estudiantil. Luego, en la fase de entrenamiento del modelo, continuando con la evaluación del modelo mediante métricas de desempeño para validar su capacidad predictiva. Finalmente, el modelo es utilizado para realizar la predicción del riesgo, identificando a los estudiantes que presentan señales tempranas de posible rezago o deserción académica.

Uno de los modelos aplicados fue el árbol de decisión, el cual permite visualizar de forma jerárquica las reglas que el algoritmo utiliza para clasificar a los estudiantes según su nivel de riesgo académico. Este tipo de modelo es especialmente útil por su carácter explicativo, facilitando la interpretación por parte de equipos académicos y no técnicos.



Figura 1. Proceso de trabajo del modelo de decisión (simplificado).

En la estructura del árbol, cada nodo representa una condición lógica basada en una variable predictora. Por ejemplo Figura 2, una de las reglas principales del modelo se expresó como:

- ¿El número de asignaturas reprobadas es mayor a 3?

Este nodo inicial actúa como punto de bifurcación. Si la condición se cumple (respuesta: sí), el modelo considera que el estudiante presenta una mayor probabilidad de estar en riesgo académico, y continúa evaluando otras variables como el nivel actual, la

brecha curricular (diferencia entre el nivel ideal y el nivel real) y la cantidad de oportunidades por asignatura (repeticiones). Si la condición no se cumple (respuesta: no), el estudiante es más probablemente clasificado como no en riesgo, aunque otras combinaciones de variables pueden modificar esta predicción.

Las hojas del árbol representan los resultados finales de la clasificación: “riesgo académico” o “no riesgo”. Esta visualización permite comprender cómo interaccionan las variables para generar una alerta, y facilita la implementación de estrategias de intervención basadas en reglas claras y trazables.

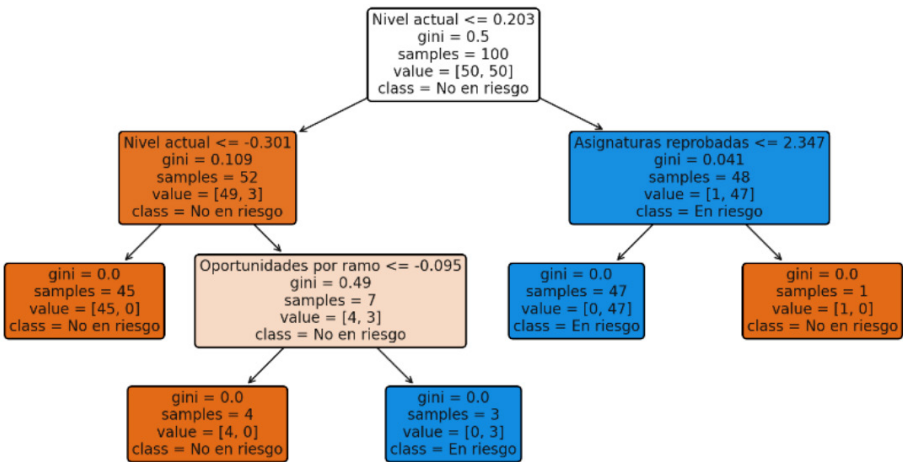


Figura 2. Árbol de decisión para asignaturas reprobadas > 3, árbol de decisión (simplificado).

Este tipo de representación resulta especialmente útil para la toma de decisiones pedagógicas, ya que permite traducir el análisis computacional en criterios accionables para docentes, tutores y autoridades académicas.

### Matriz de confusión

La matriz de confusión Figura 3 revela que el modelo posee un buen nivel de desempeño general, logrando identificar correctamente a 85 estudiantes en situación de riesgo y a 90 estudiantes sin riesgo académico, lo que refleja un buen equilibrio entre sensibilidad y especificidad. No obstante, los 15

falsos positivos y los 10 falsos negativos abren un espacio necesario para el análisis crítico.

Los falsos positivos representan estudiantes que fueron clasificados como en riesgo sin estarlo realmente. Aunque esto podría generar intervenciones innecesarias, también puede considerarse una oportunidad de acompañamiento preventivo, especialmente en contextos donde la falta de seguimiento puede afectar la motivación o generar rezagos futuros.

Más preocupante resulta el caso de los falsos negativos: estudiantes que, estando en riesgo real, no fueron detectados por el modelo. Este tipo de error puede limitar la efectividad de las políticas institucionales de retención, ya que deja fuera a quienes más necesitan apoyo. Por ello, es crucial que estos modelos se complementen con el criterio humano (docente y tutor) y con otras herramientas de monitoreo cualitativo.

En cuanto a las métricas de desempeño por modelo, se observó que Random Forest presentó mejores resultados en F1-score, lo que confirma su capacidad para manejar datasets con múltiples variables correlacionadas y cierto grado de desbalance de clases. Esto refuerza su idoneidad como algoritmo base para sistemas de alerta temprana en contextos educativos ver Figura 4.

En conjunto, estos resultados no deben entenderse solo como una validación técnica del modelo, sino

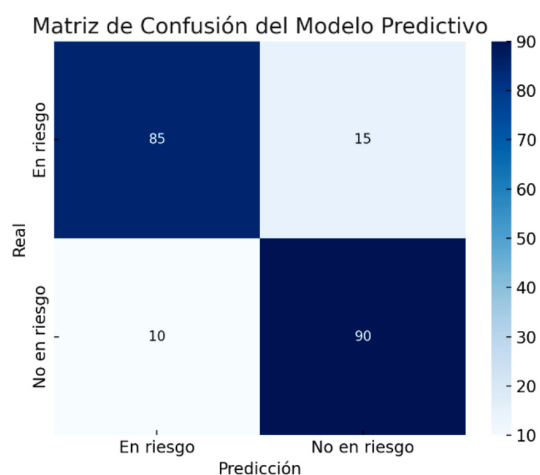


Figura 3. Matriz de confusión del modelo.

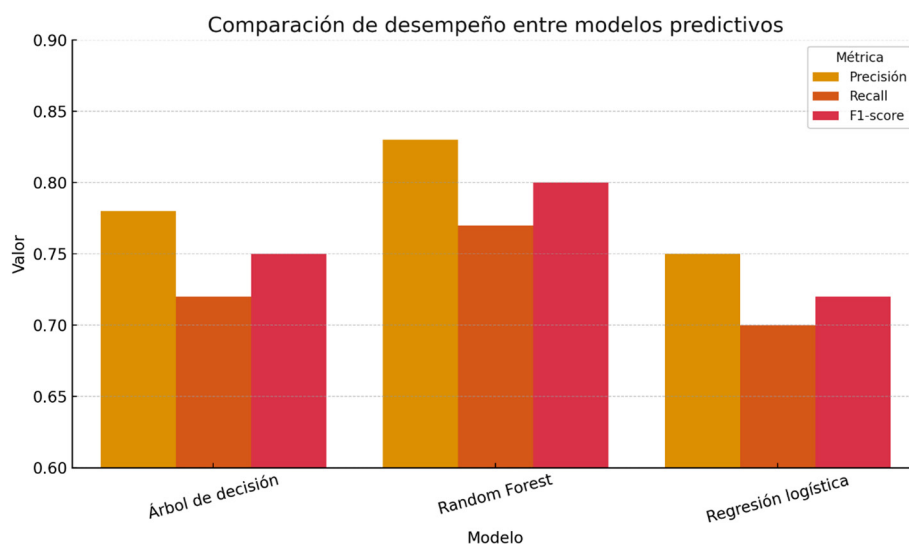


Figura 4. Comparación de los modelos, con métricas de desempeño por modelo.

como una herramienta ética y pedagógica que permite anticipar problemas antes de que se materialicen, facilitando la creación de entornos educativos más inclusivos y resilientes.

## RESULTADOS

La aplicación del modelo predictivo a una cohorte de 130 estudiantes de primer año de la carrera de Ingeniería Civil en Computación e Informática permitió identificar a 40 estudiantes en situación de pre-riesgo académico y a 20 estudiantes en riesgo académico alto. Esta clasificación anticipada habilitó la implementación de intervenciones focalizadas, como:

- Reducción estratégica de carga académica.
- Tutorías personalizadas.
- Sesiones de ayudantía.
- Seguimiento activo por parte de tutores de cohorte.

En una etapa inicial, algunas de estas acciones encontraron resistencia por parte del estudiantado, especialmente en relación con la disminución de la carga académica y la percepción de retraso en su avance curricular. Sin embargo, con el tiempo, se observó una mejor aceptación de las estrategias de acompañamiento y un cambio en la actitud frente a su proceso formativo.

Posteriormente, se evidenciaron mejoras significativas en indicadores académicos:

- 10% mejora en rendimiento promedio.
- 10% tasa de deserción.

Además, se reportó un incremento en la participación estudiantil, reflejado en una mayor asistencia a clases, involucramiento en tutorías y comprensión más efectiva de los contenidos curriculares.

Otro resultado relevante fue la identificación precisa de asignaturas con alta tasa de reprobación, siendo las más recurrentes: Física, Álgebra y Química. Este hallazgo permitió enfocar los esfuerzos institucionales hacia dichas áreas críticas, estableciendo medidas preventivas específicas.

Asimismo, los datos confirmaron que la mayor proporción de deserción académica se concentra

durante los primeros tres semestres, lo cual valida la pertinencia de implementar sistemas de alerta temprana enfocados especialmente en las etapas iniciales de la formación universitaria.

## CONCLUSIONES

La presente investigación demostró que la aplicación de modelos predictivos basados en técnicas de aprendizaje automático representa una herramienta eficaz para la identificación temprana de estudiantes en situación de pre-riesgo y riesgo académico en carreras de alta exigencia como Ingeniería Civil en Computación e Informática. A través del procesamiento y análisis de datos institucionales entre los que se pueden mencionar número de asignaturas reprobadas, oportunidades por asignatura, avance curricular y asignaturas eliminadas, se logró obtener un método que permite anticipar con alto grado de precisión escenarios de vulnerabilidad académica.

El modelo, construido con algoritmos como árboles de decisión, regresión logística y Random Forest, permitió no solo predecir resultados futuros, sino también generar conocimiento útil para la toma de decisiones pedagógicas. La validación del modelo evidenció una capacidad robusta de discriminación entre estudiantes en riesgo y sin riesgo, con especial eficacia del algoritmo Random Forest, cuyo desempeño superó en F1-score a los otros modelos evaluados.

Los hallazgos no solo se expresaron en métricas técnicas: la implementación del modelo permitió intervenir con antelación en 60 estudiantes (46% del total de la muestra), mediante acciones como la reducción estratégica de carga académica, tutorías personalizadas y seguimiento activo por tutores de cohorte. Estas intervenciones dieron lugar a mejoras concretas: un aumento del 10% en el rendimiento académico promedio y una disminución equivalente en la tasa de deserción.

El estudio también visibilizó áreas críticas dentro del programa de estudio como: Física, Álgebra y Química que las tasas de reprobación fueron significativamente altas. Esta información permitió generar remediales para abordar las asignaturas.

Además, el modelo ha sido valorado por su accesibilidad y capacidad interpretativa, permitiendo

que equipos no técnicos comprendan sus reglas de decisión y puedan utilizar sus salidas para diseñar planes de acción concretos. Esta característica refuerza su potencial de escalabilidad hacia otras carreras de la Facultad de Ingeniería de la universidad.

En definitiva, esta investigación no solo ofrece un aporte metodológico y técnico, sino también institucional y ético. Se evidencia que es posible integrar tecnologías avanzadas como el aprendizaje automático con estrategias pedagógicas, construyendo un entorno universitario más inclusivo, resiliente y centrado en el bienestar del estudiante.

De cara al futuro, se plantea como línea de acción prioritaria la incorporación de variables psicoeducativas como hábitos de estudio, participación en actividades extracurriculares o uso de plataformas virtuales con el fin de perfeccionar el modelo predictivo y ampliar su sensibilidad. Asimismo, se recomienda consolidar el Plan de Seguimiento Estudiantil como una política institucional permanente, con mecanismos de retroalimentación continua y adaptabilidad a nuevos contextos.

Considerando que los alumnos comenzaron a generar mayor compromiso con sus procesos de formativo, esto derivó en un importante cambio de cómo enfrentar la vida universitaria. Además, se pudo potenciar la comprensión lectora, el trabajo en equipo, pensamiento lógico en la resolución de problemas, la capacidad de enfrentar la teoría en la práctica, la capacidad de aprender y sobre todo el autoaprendizaje problemas principales presentados por la mayoría de los alumnos.

## REFERENCIAS

- [1] M.L. Molina Estévez, “Valoración de los criterios referentes al rendimiento académico y variables que lo puedan afectar”, *Revista Médica Electrónica*, vol. 37, no. 6, pp. 211-218, 2015. [En línea]. Disponible: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1684-18242015000600007.3](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1684-18242015000600007.3)
- [2] V. Tinto, “Definir la deserción: cuestión de perspectivas”, *Revista de Educación Superior*, no. 71, pp. 1-15, 1989.
- [3] J.F. Reyes Saldaña y R. García Flores, “El proceso de descubrimiento de conocimiento en bases de datos”, *Ingenierías*, vol. 8, no. 26, pp. 37-47, 2005. [En línea]. Disponible: <http://eprints.uanl.mx/10160/>
- [4] I.H. Witten, E. Frank, and M.A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2011.
- [5] L. Calvet Liñán and A. Juan Pérez, “Educational Data Mining and Learning Analytics: Differences, similarities, and time evolution,” *Universities and Knowledge Society Journal*, vol. 12, no. 3, pp. 98-112, 2015, doi: 10.7238/rusc.v12i3.2515.
- [6] D. Andrade-Girón et al., “Predicting student dropout based on machine learning and deep learning: A systematic review,” *EAI Endorsed Transactions on Scalable Information Systems*, vol. 10, no. 5, 2023, doi: 10.4108/eetsis.3586.
- [7] J. Cheng, Z. Q. Yang, J. Cao, Y. Yang, and X. Zheng, “Predicting student dropout risk with a dual-modal abrupt behavioral changes approach,” 2025, arXiv:2505.11119.
- [8] C. Romero and S. Ventura, “Educational data mining and learning analytics: An updated survey,” *WIREs Data Mining and Knowledge*, vol. 10, no. 3, Art. no. e1355, 2020, doi: 10.1002/widm.1355.
- [9] G. Siemens and P. Long, “Penetrating the fog: Analytics in learning and education,” *EDUCAUSE Review*, vol. 46, no. 5, pp. 30-40, 2011. [Online]. Available: <https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education>
- [10] F. Marbouti, H.A. Diefes-Dux, and K. Madhavan, “Models for early prediction of at-risk students in a course using standards-based grading,” *Computers & Education*, vol. 103, pp. 1-15, 2016, doi: 10.1016/j.compedu.2016.09.005.
- [11] Consejo Nacional de Educación (CNEC), “Índices de educación superior”, [cnec.cl](https://www.cned.cl/indices-educacion-superior). [En línea]. Disponible: <https://www.cned.cl/indices-educacion-superior>
- [12] C. Díaz, “Factores de deserción estudiantil en ingeniería: Una aplicación de modelos de duración”, *Información tecnológica*, vol. 20, no. 3, pp. 125-134, 2009, doi: 10.4067/S0718-07642009000300014.
- [13] R. Zárate Rueda y E. Mantilla Pinilla, “La deserción estudiantil UIS, una mirada desde la responsabilidad social universitaria”,

- Zona Próxima*, vol. 21, pp. 147-165, 2014. [En línea]. Disponible: [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S2145-94442014000200010](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S2145-94442014000200010)
- [14] J.L. Rastrollo-Guerrero, J.A. Gómez-Pulido, and A. Durán-Domínguez, "Analyzing and predicting students' performance by means of machine learning: A review," *Applied Sciences*, vol. 10, no. 3, Art. no. 1042, 2020, doi: 10.3390/app10031042.
- [15] S. Helal, J. Li, L. Liu, E. Ebrahimie, and D. Dawson, "Predicting academic performance by considering student heterogeneity," *Knowledge-Based Systems*, vol. 161, pp. 134-146, 2018, doi: 10.1016/j.knosys.2018.07.042.
- [16] K. Azoumana, "Análisis de la deserción en la Universidad Simón Bolívar, Facultad de Ingeniería de Sistemas, con técnicas de minería de datos", *Pensamiento Americano*, vol. 6, no. 11, pp. 41-51, 2013.
- [17] S. Prasad Sinha, J. Romani Perazzi, G. Orlandoni Merli, E. Torres Rivas y M. Figueroa, "Conceptuación y análisis descriptivo del riesgo académico institucional en las universidades nacionales venezolanas: el caso de la ULA", *Educere*, vol. 11, no. 39, pp. 653-663, 2007.
- [18] J.S. Parra-Sánchez, I.D. Torres y C.Y. Martínez, "Factores explicativos de la deserción universitaria abordados mediante inteligencia artificial", *Revista Electrónica de Investigación Educativa*, vol. 25, Art. no. e18, 2023, doi: 10.24320/redie.2023.25.e18.4455.
- [19] B. Albreiki, N. Zaki, and I. Alashwal, "A systematic literature review of student performance prediction using machine learning," *Education Sciences*, vol. 11, no. 9, Art. no. 552, 2021, doi: 10.3390/educsci11090552.