JSTRI
JOURNAL OF SCIENTIFIC AND TECHNOLOGICAL RESEARCH INDUSTRIAL

Journal of Scientific and Technological Research Industrial
ISSNe: 2961-211X

PROF-JSTRI-V-00
Vol. 6 N.º 2
(July - December, 2025)

Modelo predictivo para la prevención de la deserción académica en estudiantes de un instituto tecnológico de Puno, 2025

Predictive model for preventing academic dropout among students at a technological institute in Puno, 2025

Hernán H. Huamán Arratia
Universidad Nacional Federico Villarreal, Perú

Resumen

El objetivo de este estudio fue desarrollar un modelo predictivo para prevenir la deserción académica en estudiantes de un instituto tecnológico de Puno durante el año 2025. La problemática abordada radica en las altas tasas de deserción escolar en las regiones rurales del Perú, influenciadas por factores socioeconómicos, académicos y personales. La investigación fue básica-aplicada, con enfoque cuantitativo, diseño descriptivo y explicativo, y alcance transversal. La población estuvo constituida por todos los estudiantes matriculados en el instituto en el año 2025, mientras que la muestra estuvo constituida por 300 estudiantes seleccionados mediante muestreo estratificado. Los resultados cuantitativos muestran que variables como el bajo rendimiento en el primer semestre, las condiciones socioeconómicas precarias, la edad avanzada al ingreso y el género inciden significativamente en el riesgo de deserción. El modelo predictivo, basado en técnicas de aprendizaje automático como la regresión logística y los árboles de decisión, alcanzó altos niveles de precisión y recordación. Se concluye que la identificación temprana de estudiantes en riesgo, mediante modelos predictivos, permite diseñar estrategias de intervención efectivas. Se recomienda su implementación institucional, junto con medidas de apoyo académico, financiero y emocional personalizado, lo que podría mejorar significativamente la retención estudiantil en contextos similares.

Palabras claves: Abandono académico, modelo predictivo, educación superior, aprendizaje automático.

Abstract

The objective of this study was to develop a predictive model to prevent academic dropout among students at a technological institute in Puno during the year 2025. The problem addressed lies in the high school dropout rates in rural regions of Peru, influenced by socioeconomic, academic, and personal factors. The research was basic-applied, with a quantitative approach, descriptive and explanatory design, and cross-sectional scope. The population consisted of all students enrolled in the institute in 2025, while the sample consisted of 300 students selected through stratified sampling. The quantitative results show that variables such as poor performance in the first semester, precarious socioeconomic conditions, advanced age at entry, and gender significantly influence the risk of dropping out. The predictive model, based on machine learning techniques such as logistic regression and decision trees, achieved high levels of precision and recall. It is concluded that early identification of at-risk students through predictive models allows for the design of effective intervention strategies. Its institutional implementation is recommended, along with personalized academic, financial, and emotional support measures, which could significantly improve student retention in similar contexts.

Keywords: Academic dropout, predictive model, higher education, machine learning.

Publicado: 20/10/2025
Aceptado: 20/10/2025
Recibido: 10/09/2025

Open Access
Article scientific
https://doi.org/10.47422/jstri.v6i2.66

Introducción

En un contexto global, la deserción académica es uno de los mayores desafíos que enfrentan las instituciones educativas, especialmente en la educación superior. La falta de igualdad de acceso a una educación de calidad, las condiciones socioeconómicas y los factores culturales influyen en la baja retención estudiantil en muchos países. Así, en estudios internacionales, la tasa de deserción puede alcanzar tasas superiores al 30 % en algunas regiones. En China, por ejemplo, se han registrado tasas del 30 % al 40 % entre estudiantes universitarios, lo que tiene consecuencias negativas no solo en la vida de los estudiantes, sino también en el crecimiento económico y el desarrollo social de los países.

En este sentido, habilitar las capacidades para prevenirlo se presenta como la forma de mejorar las oportunidades educativas y reducir las desigualdades. A medida que la educación se globaliza, la enseñanza se transforma y las tecnologías avanzan, los sistemas escolares deben adaptarse, comenzando por el uso de herramientas y modelos predictores múltiples que aseguren la identificación temprana de estudiantes en riesgo, así como la rápida posibilidad de intervención en ellos [1] . Sin embargo, en el contexto peruano, la baja retención escolar se convierte en un problema estructural para las instituciones de educación superior, provocando que miles de jóvenes se gradúen cada año. Este problema se agrava por factores como la desigualdad económica, la falta de recursos en las instituciones de educación superior y las limitadas oportunidades laborales para los graduados de estas zonas rurales. Según el Ministerio de Educación del Perú, las tasas de deserción escolar en la educación superior en algunas regiones del país superan el 20%, una Figura alarmante considerando un esfuerzo institucional y de políticas públicas para mejorar la calidad educativa. La tasa de deserción no solo es una consecuencia del futuro académico y profesional de los estudiantes: también es una de las causas del estancamiento del desarrollo social y económico en las regiones socialmente excluidas. En consecuencia, se necesita una perspectiva innovadora adaptada a la realidad nacional para al menos predecir y, de ser necesario, prevenir la deserción escolar, especialmente en las llamadas regiones periféricas del país [2] . A nivel local, en la región Puno, la deserción académica puede ser un desafío aún más complejo dado el contexto socioeconómico y geográfico. La alta tasa de pobreza y la dispersión geográfica de las poblaciones, sumada a una problemática educativa con serios desafíos, complica la situación de los estudiantes que abandonan los programas de formación tecnológica debido al poco o ningún apoyo financiero, el transporte deficiente a sus lugares de estudio y una oferta limitada del tipo de programas de formación de calidad que requieren. Las tasas de deserción académica son preocupantes en la región, especialmente entre los jóvenes de comunidades rurales. Un modelo predictivo para prevenir la deserción escolar en las instituciones tecnológicas de la región Puno podría ser una herramienta muy útil para detectar estudiantes en riesgo y, a partir de allí, promover mecanismos de intervención temprana que ayuden a mejorar las tasas de retención y, así, contribuir al desarrollo educativo y económico local. [3] .

La deserción escolar es un fenómeno complejo que no solo afecta a millones de estudiantes en todo el mundo, sino que también tiene efectos perjudiciales que trascienden el ámbito académico. En pocas palabras, la deserción escolar es el abandono de la educación de un estudiante antes de completar la educación primaria, secundaria o superior. Este fenómeno se debe a diversos factores específicos de cada situación, pero generalmente incluye barreras económicas, falta de motivación, problemas académicos, problemas familiares y una desconexión entre las expectativas de los estudiantes y la realidad educativa [4] .

A nivel mundial, los jóvenes que abandonan la escuela lo hacen por falta de recursos económicos para continuar sus estudios, por la necesidad de incorporarse al mercado laboral o simplemente por no recibir el apoyo necesario de las instituciones educativas. Por lo tanto, el abandono escolar se considera un obstáculo importante para la construcción de sociedades más justas y equitativas [5] .

Las tasas de deserción académica en Perú se han convertido en un problema significativo que afecta a miles de jóvenes cada año, particularmente en las regiones más remotas y empobrecidas. Según datos del Ministerio de Educación del Perú, la tasa de deserción en la educación superior es considerable y se ve exacerbada por factores como las dificultades económicas, la falta de infraestructura y una gama limitada de programas educativos que satisfagan las necesidades de los estudiantes. Las instituciones tecnológicas, particularmente en Puno, enfrentan desafíos adicionales. La ubicación geográfica de Puno, las dificultades que tienen los jóvenes para acceder a recursos que les permitan continuar sus estudios, o el hecho de que son conscientes de las dificultades que enfrentan para continuar sus estudios, han llevado a los de las zonas rurales, en particular, a abandonar sus estudios universitarios debido a la incapacidad de acceder a materiales, pagar el transporte u otras carencias materiales [6] .

En las instituciones tecnológicas de Puno existe un alto porcentaje de estudiantes que deciden, antes de culminar su formación técnica, abandonar sus estudios, lo cual es consecuencia de varios factores interdependientes que finalmente las instituciones se ven obligadas a asumir; la

falta de redes de apoyo, la pobreza económica y las dificultades para adaptarse al sistema educativo son algunas de las principales causas, pero quienes se encuentran en estas instituciones tienen otras dificultades a las que están sometidos, no solo la carga que suponen las actividades académicas sino también las dificultades sociales que tienen que asumir desde sus condiciones de vida. El abandono de los estudios en esta región afecta mucho más a nivel personal que a los jóvenes de esta región, quienes en la mayoría de los casos forjan su futuro personal, pero también afecta colectivamente el desarrollo social y económico de los entornos. La pérdida de potencial humano provoca que las posibilidades de crecer y progresar se retrasen hasta que las comunidades hayan explorado sus posibilidades de generar un cambio sostenible en el futuro [7] .

En este contexto, el problema de la deserción escolar cobra relevancia, particularmente en el caso de Puno, con la convicción de que las soluciones que han demostrado ser válidas no resultan en última instancia efectivas. La identificación temprana permitiría la implementación de intervenciones más eficientes que ayuden a prevenir las causas que llevan a los estudiantes a tomar la decisión de abandonar sus estudios [8] .

El problema que no es simple sino complejo y difícil a la vez; dicho problema requiere un análisis de las causas y, además, nos invita a entrar en el ámbito de búsqueda de nuevas fórmulas tecnológicas y de nuevos modelos para prevenir la deserción escolar y, sólo así, se logrará garantizar que los jóvenes de Puno, como también de las periferias del país, consigan llegar a terminar su intensa formación educativa, esperando en ello una mejora hacia su futuro, el de sus respectivos pueblos, el de su país [9].

Para el contexto de esta investigación, el problema central de esta investigación se consiste de la siguiente forma: ¿Cómo implementar un modelo predictivo efectivo para prevenir la deserción académica de los estudiantes de un instituto tecnológico de Puno en 2025? Con este enfoque se trata precisamente de dar respuesta a un problema que necesita de la identificación de los factores que promueven la deserción escolar y en la solución de un problema que requiere de la anticipación de la deserción, de poder optimizar las estrategias de intervención y, de poder mejorar las tasas de retención escolar de la región.

Por otro lado, [10] terminó realizando una investigación cuyo objetivo consistió en implementar técnicas de ciencia de datos con el propósito de prever patrones de deserción estudiantil en la Universidad Pedagógica y Tecnológica de Colombia (UPTC), sede Duitama. El trabajo de investigación se realizó desde un enfoque cuantitativo con un diseño experimental exploratorio Los estudios se centraron en la aplicación de algoritmos para construir modelos predictivos, utilizando para ello la información estructurada de la universidad. La población de estudio se ajustó con respecto a los datos académicos y sociodemográficos de los estudiantes de la UPTC, de los mismos se realizó una selección de una muestra representativa para el análisis. Los resultados logrados mostraron la efectividad del modelo en detectar patrones de deserción, valida por los resultados provenientes de diferentes métricas de calidad que hicieron hincapié en la adecuación y utilidad del modelo. El trabajo de investigación concluyó indicando que las técnicas de ciencia de datos pueden considerarse como una herramienta muy útil para poder anticipar y mitigar la deserción estudiantil, hasta el punto de recomendarse su aplicación para mejorar la toma de decisiones en los procesos educativos. Esta serie de antecedentes son significativos para el trabajo de investigación actual, en la medida que ofrecen una metodología robusta y herramientas aplicables para el desarrollo de un modelo afín en el contexto de Puno.

Por otra parte, [11] llevó a cabo una investigación cuyo propósito fue construir un modelo predictivo de deserción escolar en estudiantes de la Unidad Educativa "Los Andes" y con delitos del contexto de la pandemia de COVID-19. Se trató de un estudio cuantitativo con enfoque correlacional y un diseño de campo y documental. La población estuvo conformada por 1.000 estudiantes; del total, analizaron 230. Los resultados indicaron que los factores socioeconómicos fueron determinantes de la deserción, sobre todo la falta de tecnología y de apoyo familiar. El modelo predictivo construido permitió además el acompañamiento de estudiantes en riesgo de deserción escolar. El estudio concluyó que la pandemia aumentó la tasa de deserción, además de que resultó relevante implementar modelos predictivos para identificar y atender a los estudiantes en riesgo, estos antecedentes se consideraron de gran interés para esta investigación porque enfatiza que es relevante tener en cuenta factores socioeconómicos y el efecto de eventos externos como la pandemia en la retención de estudiantes.

Por su parte, [12] realizaron un estudio orientado en desarrollar un modelo predictivo para pronosticar el éxito o fracaso de estudiantes universitarios considerando como indicadores de ingreso temprano. El enfoque de la investigación fue cuantitativo, correlacional y de tipo ex post facto , reconociendo la existencia de los datos académicos y sociodemográficos de 4.012 egresados y 3.393 desertores de una universidad estatal del norte de Chile. Los datos obtenidos evidenciaron que las tasas de aprobación del primer año fueron el mejor predictor del éxito académicos, mientras que las calificaciones de

secundaria fueron predictoras del rendimiento inicial. En consecuencia, se concluyó que los modelos predictivos que utilizan indicadores tempranos son efectivos para identificar estudiantes en riesgo de fracaso académico. Dichos antecedentes constituyen un aporte importante para poder establecer factores predictivos en el presente estudio y la posterior construcción del modelo que permita reconocer estudiantes en riesgo de deserción escolar en Puno en los primeros años de vida académica.

De acuerdo con [12] desarrollaron un tipo de investigación que tuvo como objetivo definir un modelo predictivo para la deserción escolar en educación superior mediante minería de datos. Constituyó una investigación cuantitativa, con diseño exploratorio y muestra compuesta de 1374 estudiantes de una institución de educación superior en México; los datos obtenidos mostraron que el modelo predictivo logra identificar la deserción escolar, considerando variables como el rendimiento escolar, el contexto socioeconómico y acceso a recursos tecnológicos. La investigación concluyó diciendo que la minería de datos es una potente herramienta para poder realizar predicciones, recomendando que sea usada en sistemas de alertas institucionales. Estos antecedentes son relevantes en el estudio en tanto que sustentan el uso de la minería de datos para establecer un modelo predictivo a aplicar al caso del Instituto Tecnológico de Puno.

Por su parte, [13] analizó Cómo las variables institucionales vinculadas al proceso de admisión influyen en el riesgo de deserción estudiantil en una universidad de Caldas, Colombia. Para ello se realizó un estudio cuantitativo, correlacional con diseño retrospectivo, usando modelos logísticos multivariados y técnicas de optimización de variables. La población fueron estudiantes matriculados desde 2010 a 2012 en diferentes facultades, considerando una base de datos proporcionada por la oficina de registro académico. Los resultados muestran que el puntaje de admisión, la opción de admisión, el costo de la matrícula y programas de cupo especial tienen una influencia significativa en el riesgo de deserción. Se concluye que los alumnos que ingresan como segunda opción tienen un riesgo de deserción 3 veces mayor. El estudio sugiere revisar la ponderación de las áreas evaluadas e implementar políticas de admisión basadas en mérito y acciones afirmativas para reducir el riesgo de deserción. Estos antecedentes indican la importancia de las variables institucionales, que son sustantivas en el diseño del modelo predictivo concebido para el contexto de Puno. Los problemas específicos que orientan esta investigación son los siguientes: identificar los factores socioeconómicos que influyen en la deserción escolar; determinar cuáles son los métodos y las herramientas tecnológicas más adecuados para elaborar el modelo predictivo; qué impacto tiene la implementación del modelo en la mejora de la retención escolar; establecer estrategias de intervención para lograr la disminución de la deserción escolar, a partir de los resultados alcanzados en esta investigación.

La justificación teórica de esta investigación podemos sostenerla en que la deserción escolar no es un fenómeno simple, sino un fenómeno multifactorial, con implicaciones de tipo académico, personal y socioeconómico. Construir un modelo predictivo a partir de variables institucionales, como la implementación de procesos de admisión o la ponderación de calificaciones, les permitiría a las instituciones educativas prevenir la deserción y actuar anticipadamente. Experiencias previas han demostrado la efectividad de la minería de datos y del análisis multivariable como forma de identificar a los estudiantes en riesgo de deserción, lo que permite identificar la importancia de esta investigación, desde una perspectiva práctica; en esta investigación, se busca brindar respuestas integrales al problema de la deserción escolar de los institutos tecnológicos de Puno. La utilización de un modelo predictivo permitiría a las autoridades escolares anticipar los casos de riesgo y reaccionar desde el punto de vista académico, psicológico ofreciéndoles becas o ajustes a los criterios de admisión, logrando la optimización de recursos y la personalización del tipo de intervención.

Socialmente, la deserción escolar afecta no solo a los estudiantes de forma individual, sino que afecta también a la propia comunidad, reducir las propias oportunidades de desarrollarse, perpetuando las desigualdades sociales en el ámbito educativo. En Puno donde el contexto socioeconómico es complejo, la adopción de un modelo predictivo puede ser la estrategia para promover la inclusión social del alumnado, identificando anticipadamente a los estudiantes que puedan estar en riesgo (en el proceso de la educación) y ofreciendo políticas de retención que favorezcan la reducción de la deserción escolar. Económicamente, la deserción escolar afecta a las instituciones educativas, a las propias familias y a los estudiantes, generando altos costos a las propias instituciones, afectan de forma directa las finanzas institucionales y reducen los márgenes de sostenibilidad a largo plazo. La aplicación de un modelo predictivo, disminuiría los costos que actualmente, en el ámbito económico provocan lo que hace que se reduzcan las tasas de deserción escolar, mejorarían la eficiencia en la asignación de recursos y aumentarían la probabilidad de que los alumnos/as acaben sus estudios, que accedan a mejores empleos y contribuir al desarrollo económico local. El valor de la presente investigación radica en que podrá ayudar a las instituciones de educación superior a detectar a los estudiantes con riesgo de deserción muy pronto para llegar a realizar intervenciones que aumenten

la tasa de retención y el éxito académico. Además, podrá ayudar a reducir las desigualdades sociales en Puno, incrementar la efectividad de las políticas de admisión y apoyo a los estudiantes y el aprendizaje a nivel general, lo que también beneficiará a los estudiantes, a las instituciones y a la sociedad en general.

No obstante, el estudio tiene limitaciones. Se realiza en un solo lugar, un instituto tecnológico de Puno, y por eso es muy difícil generalizar en otras instituciones; el segundo riesgo está relacionado con la efectividad del modelo y con la calidad y disponibilidad de los datos: puede no alcanzar el grado deseado si la información es incompleta o no es de suficiente calidad. También hay otros factores contextuales no considerados en este estudio, como problemas familiares, salud o problemas económicos que pueden llegar a influir en las tasas de deserción escolar. De la misma manera, se presentan los problemas de la medición de variables subjetivas, como la motivación o la satisfacción de los estudiantes. El modelo también tiene limitaciones porque los recursos tecnológicos o la capacitación de los profesionales en las instituciones de educación superior de escasos recursos puede suponer un obstáculo en la mejora de la deserción escolar.

El propósito principal de esta investigación es crear un modelo predictivo para prevenir la deserción académica que puedan sufrir los jóvenes estudiantes que asisten al instituto tecnológico de Puno en el año 2025. En consecuencia, también se fijan los siguientes objetivos específicos: identificar los factores determinantes que intervienen en esta problemática; desarrollar un modelo predictivo fundamentado en datos académicos y sociodemográficos; evaluar cómo las condiciones institucionales afectan en esta problemática y presentar estrategias de intervención que potencialmente mejorarían la permanencia en el sistema educativo.

La hipótesis general de investigación establece que, tras la implementación del modelo predictivo, se logrará optimizar la retención de estudiantes en el instituto. De forma más específica, se propone que identificar factores de riesgo incrementará la retención; implementar un modelo predictivo optimizará la utilización de los recursos del instituto y reducirá sus costes; las intervenciones personalizadas reducirán los índices de deserción escolar y el modelo predictivo contribuirá a mejorar la toma de decisiones en el ámbito de las admisiones en el instituto o bien las políticas de apoyo a los estudiantes.

Marco teórico

Modelo predictivo de la deserción académica

Los modelos predictivos son herramientas que permiten realizar análisis de un gran conjunto de datos con el objetivo de predecir comportamientos a partir del uso de técnicas estadísticas y algoritmos avanzados de aprendizaje automático. En el sector educativo, su utilización permite detectar patrones asociados a los porcentajes de abandono escolar de los estudiantes [14]. Su construcción considera datos históricos, en los cuales se incluyen, entre otros, variables del rendimiento académico, entorno socioeconómico y niveles de interacción con el entorno académico. También pueden generar detecciones anticipadas, ayudando a facilitar las intervenciones que prevengan el abandono escolar.

La información considerada incluye calificaciones, asistencia, participación en actividades extraescolares o cualquier otro indicador que pueda influir sobre el rendimiento académico. Entre las técnicas más utilizadas para el entrenamiento de modelos están la regresión logística, las redes neuronales artificiales y los árboles de decisión. Una vez detectados los estudiantes en riesgo, las instituciones implementan medidas de prevención como tutorías, apoyo psicológico, o apoyo financiero, medida que contribuye a mejorar la retención de estudiantes que abandonan la escuela [15].

En el marco de una institución tecnológica de Puno, el modelo predictivo debe contemplar los factores específicos de la localización en la que se desarrolla, como las condiciones sociales y económicas de los discentes, los resultados de rendimiento en las primeras etapas de formación y las condiciones de acceso a los medios formativos. De este modo se hace posible un modelo que sea adecuado y que contemple la realidad de la población estudiantil [6].

Variable dependiente: Abandono académico

La deserción académica se entiende como el abandono de los estudios por parte del estudiante, sin haber llegado a completar el ciclo educativo correspondiente. Este fenómeno es un reto tanto para los estudiantes, quienes restringen sus posibilidades de desarrollo, como para las instituciones, las que ven una merma de la retención de alumnos y la sostenibilidad de sus propios programas de formación [16]. La deserción es un fenómeno multicausal, cuyas determinaciones pueden ser divididas en: causas internas (bajo rendimiento académico, problemas de adaptación a la vida universitaria y falta de compromiso con la formación) y causas externas (causas económicas, situación familiar difícil, la falta de soporte social, acceso limitado a medios formativos).

En el caso de un instituto tecnológico, las limitaciones económicas y el acceso limitado a los medios formativos son un punto de partida en el sentido de la deserción [17]. Conocer sus causas es esencial para proponer alternativas que incrementen la retención del alumnado.

Metodología de desarrollo de software

La investigación se llevará a cabo basado en una metodología ágil, idónea para proyectos en los que cambian frecuentemente los requisitos. A través de dicha metodología se facilita el desarrollo iterativo e incremental, permitiendo que el modelo predictivo se ajuste y mejore cuando se obtengan nuevos datos [18].

Aplicación del aprendizaje automático

La construcción del modelo implicará la utilización de algoritmos de aprendizaje automático de tipo: redes neuronales, árboles de decisión, regresión logística. También se contará con herramientas de análisis y visualización de datos como son Python y R, ampliamente utilizadas en ciencia de datos y análisis predictivo [13].

Integración tecnológica

Se puede pensar también en la posible integración del modelo en una plataforma digital o aplicación para dispositivos móviles para favorecer el acceso de los responsables del seguimiento académico y permitir la actualización de datos y alertas en tiempo real [19].

Desarrollo y validación del modelo predictivo

La validación del modelo se llevará a cabo a través de validación cruzada, como técnica que evalúa el rendimiento en distintos subconjuntos de datos en el que se intente evitar el sobreajuste y permitir un modelo con capacidad de poder generalizar [20]. Se utilizarán métricas de precisión, recuperación, puntuación F1 y AUC-ROC. El ajuste de los parámetros estará orientado a minimizar los falsos positivos y los falsos negativos para optimizar la utilización de los recursos institucionales y priorizar a quienes realmente requieren de intervención [21] .

Marco filosófico

Epistemología educativa

La investigación que emprendemos se basa en un enfoque constructivista, considerando que el conocimiento es una construcción activa que se produce a partir de la interacción entre el estudiante y su entorno; es decir; los modelos predictivos son herramientas estadísticas y medios de enterarse e ir mejorando la comprensión acerca de la dinámica de la deserción escolar; así pueden ser útiles para la generación de conocimiento que permita tomar decisiones [22] .

Teoría de la retención estudiantil

La teoría de Bean establece que la deserción escolar se relaciona con una falta de integración académica y social. Por ello, es necesario: (1) promover la adaptación académica a través de programas de apoyo y (2) promover la integración social mediante redes de apoyo, actividades extracurriculares, etc. La teoría del compromiso pone énfasis en que permanecer depende de la satisfacción que produce la experiencia educativa y el apoyo institucional existente [23] .

Enfoque humanista y social

El modelo prima el humanismo educativo; los estudiantes son el centro del proceso educativo, promoviendo el desarrollo integral de los mismos. Asimismo, se exceptúa a la consideración del impacto que generan los factores sociales, buscando la equidad en el acceso y en la experiencia educativa [24] .

Ética en la educación

El uso de los datos debe suponer la vigencia de los principios de confidencialidad, de equidad y de no discriminación; las predicciones deben servir para mejorar la experiencia de los estudiantes, nunca para etiquetar o para excluir [25] .

Enfoque tecnológico y de innovación

La investigación se encuentra inmersa dentro de los procesos de transformación educativa digital, adoptando tecnologías emergentes como el Machine Learning y el Big Data para la personalización de la enseñanza-aprendizaje y la mejora de la gestión académica [26].

Estado del arte

Modelos predictivos

La literatura internacional ha evidenciado la funcionalidad de los modelos predictivos para la detección del alumnado con riesgo de abandono, desarrollándose en diversas universidades de Estados Unidos, Canadá y el Reino Unido, donde se empleó el uso de redes neuronales o la regresión logística a partir de variables académicas, socioeconómicas y psicométricas para predecir las tasas de abandono escolar [27]. En Perú, el uso de estas herramientas está en proceso de desarrollo, por lo que es idóneo desplegar investigaciones en acción para acentuar la implementación de estas herramientas en institutos de enseñanza tecnológica.

Sistemas de alerta temprana

Diferentes investigaciones han puesto de manifiesto la relevancia de los sistemas de alerta temprana, que posibilitan el hecho de iniciar una intervención antes de que un estudiante abandone la formación. Las intervenciones podrían ser tutorías personalizadas, apoyo psicológico, becas y apoyo económico, o flexibilidad académica [28].

Aplicación de tecnologías educativas

El uso de plataformas digitales en Big Data y Machine Learning empodera la adaptación en tiempo real del progreso del alumnado, facilitando respuestas rápidas e individualizadas. La inteligencia artificial también ha facilitado la educación a la medida, haciendo que el contenido se adapte al ritmo y las necesidades del estudiante [29].

Resultados

Identificación de factores de carácter socioeconómico, académico, personal en torno a la deserción estudiantil en un instituto tecnológico de Puno.

Resumen del análisis

La investigación se ejecutó mediante el procesamiento de información institucional y registros administrativos de distintos periodos de la actividad académica reciente, en un instituto tecnológico de Puno. La base de datos, que contenía observaciones tanto de estudiantes con egresos académicos completos como de estudiantes que habían abandonado algún periodo, permitía la realización de perfiles de estudiantes que habían terminado los estudios y de estudiantes cuyas carreras académicas se habían interrumpido, lo que permitía evidenciar patrones diferentes. La información se dividió para poder abarcar las dimensiones de carácter socioeconómico, académico y personal, además de indicadores macroeconómicos que nos daban una referencia del entorno de la región en el periodo objeto de estudio, pero se eligieron sólo factores que mostraban relación estadísticamente significativa con la deserción académica.

Factores socioeconómicos

En esta dimensión, se analizaron variables que reflejan el contexto económico y familiar del estudiante. Los resultados mostraron que una proporción considerable de desertores escolares provenía de hogares con menor nivel educativo parental, lo que coincidía con la presencia de empleos informales o mal remunerados entre los jefes de hogar.

También se observó que el impago de matrícula a tiempo, así como el atraso en el pago de las matrículas, se asociaban directamente con la deserción escolar. Los estudiantes con becas parciales o totales tendían a presentar menores tasas de deserción, lo que demuestra el papel mitigador del apoyo financiero. En cuanto al contexto regional, se observó un aumento en el número de retiros voluntarios e involuntarios en los períodos de mayor desempleo e inflación. Además, la disminución del producto interno bruto de la región coincidió con un repunte en las tasas de deserción, lo que sugiere una relación entre las condiciones económicas externas y la capacidad de las familias para mantener la formación técnica.

Factores académicos

Los factores académicos se correlacionaron fuertemente con las tasas de deserción. Los estudiantes que aprobaron menos cursos en el primer semestre tendieron a abandonar antes de completar el segundo. De igual manera, un promedio bajo de calificaciones en cualquiera de los semestres evaluados se asoció con una mayor probabilidad de deserción. El modo de admisión también influyó significativamente: quienes ingresaron por orden de prioridad más bajo o por canales de admisión menos competitivos tuvieron un mayor riesgo de interrupción académica. El horario de asistencia también fue relevante; los estudiantes que asistieron a clases nocturnas tuvieron mayores tasas de deserción, posiblemente debido a la carga de trabajo concurrente o la menor disponibilidad de recursos académicos durante esas horas. Asimismo, la falta de evaluaciones completas en los cursos en los que se inscribieron fue un indicador temprano de riesgo, ya que indicó ausentismo o baja participación en actividades académicas.

Factores personales

Las características personales desempeñaron un papel complementario en el análisis. La edad de matriculación reveló un patrón claro: los estudiantes mayores, especialmente aquellos por encima del promedio del grupo, tenían más probabilidades de abandonar los estudios, a menudo debido a responsabilidades laborales o familiares. En cuanto al género, se encontró que las mujeres, aunque tenían un rendimiento promedio ligeramente superior, enfrentaban barreras relacionadas con la conciliación de los estudios y las responsabilidades domésticas, lo que afectaba su retención. La nacionalidad y el desplazamiento geográfico también se asociaron con el abandono escolar, especialmente entre quienes no tenían redes de apoyo cercanas. La presencia de necesidades educativas especiales, aunque minoritaria, representó un factor de riesgo cuando no había medidas de adaptación curricular ni suficientes recursos de apoyo. Estos hallazgos destacaron la importancia de abordar no solo el rendimiento académico, sino también la accesibilidad y el apoyo.

Implicación para el modelo predictivo

La combinación de estos factores permitió crear un perfil de riesgo de deserción escolar altamente preciso. El análisis inferencial mostró que variables como el rendimiento académico en el primer semestre, la puntualidad en los pagos, el nivel educativo de los padres, el horario de asistencia y el contexto económico general tuvieron el

mayor impacto en la predicción de la deserción. El diagnóstico sirvió de base para el entrenamiento del modelo predictivo propuesto para detectar casos con mayor riesgo de deserción y generar alertas tempranas. La combinación de variables de distinta naturaleza económica, académica y personal fortaleció la capacidad predictiva y permitió el desarrollo de intervenciones más específicas y efectivas.

Desarrollar un modelo predictivo que utilice los datos disponibles sobre el rendimiento académico, el proceso de admisión y otras variables relevantes para identificar a los estudiantes en riesgo de abandonar la escuela.

Introducción al conjunto de datos

Específicamente, la base de datos incluye variables demográficas, académicas, socioeconómicas e históricas sobre el rendimiento académico de los estudiantes universitarios. Su fin consiste en ayudar a predecir si un estudiante abandonará el colegio o si alcanzará el éxito académico.

Marital status	Application mode	Application order	Course	Daytime/evening attendance	Previous qualification	Nacionality	Mother's qualification	Father's qualification	Mother's occupation	Father's occupation	Displaced	Educational special needs	Debtor	Tuition fees up to date	Gender	Scholarship holder	Age at enrollment
1	8	5	2	1	1	1	13	10	6	10	1	0	0	1	1	0	20
1	6	1	11	1	1	1	1	3	4	4	1	0	0	0	1	0	19
1	1	5	5	1	1	1	22	27	10	10	1	0	0	0	1	0	19
1	8	2	15	1	1	1	23	27	6	4	1	0	0	1	0	0	20
2	12	1	3	0	1	1	22	28	10	10	0	0	0	1	0	0	45
2	12	1	17	0	12	1	22	27	10	8	0	0	1	1	1	0	50
1	1	1	12	1	1	1	13	28	8	11	1	0	0	1	0	1	18
1	9	4	11	1	1	1	22	27	10	10	1	0	0	0	1	0	22
1	1	3	10	1	1	15	1	1	10	10	0	0	0	1	0	1	21
1	1	1	10	1	1	1	1	14	5	8	1	0	1	0	0	0	18
1	1	1	14	1	1	1	23	14	6	8	1	0	0	1	0	0	18
1	1	1	12	1	1	1	13	28	10	10	1	0	0	1	0	1	18
1	1	2	16	1	1	1	13	27	5	10	1	0	0	1	0	0	19
1	17	1	11	1	16	1	1	1	5	8	1	0	0	1	0	1	21
1	1	1	6	1	1	1	23	27	6	6	1	0	0	1	0	1	18
1	1	1	15	1	1	1	13	27	10	4	1	0	0	1	0	0	20
1	9	1	10	1	1	1	13	28	6	9	1	0	0	1	0	0	18
1	8	2	12	1	1	1	13	1	6	5	1	0	0	1	0	0	18
1	1	1	8	1	1	1	3	14	4	6	1	0	0	1	0	0	20
1	1	1	16	1	1	1	13	14	8	8	1	0	0	1	0	0	18
1	1	3	2	1	1	1	1	1	10	9	0	0	0	1	0	1	21
1	9	4	13	1	1	1	1	28	5	8	1	0	0	1	0	0	20
1	1	4	12	1	1	1	13	14	2	2	1	0	0	1	0	0	18
1	1	4	14	1	1	1	1	28	5	8	1	0	0	1	0	1	19
1	1	1	12	1	1	1	13	14	4	8	0	0	0	1	0	0	19
1	1	1	10	1	1	1	13	28	10	10	1	0	1	1	0	1	18

Figura 1*Conjunto de datos históricos de 2020-1 a 2025-1*

Carga y escaneo inicial

La primera etapa de la carga de datos y su exploración fue el primer paso necesario e imprescindible para construir el modelo predictivo de la deserción escolar. Para ello hicimos uso de la biblioteca Pandas de Python, que nos permitió cargar e ir manipulando el archivo cuya información es la base de datos de estudiantes. Cargado el conjunto de datos, se inspeccionó la estructura general ejecutando student.shape, lo que arrojó del marco de datos, 4424 observaciones (filas) y 35 atributos (columnas).

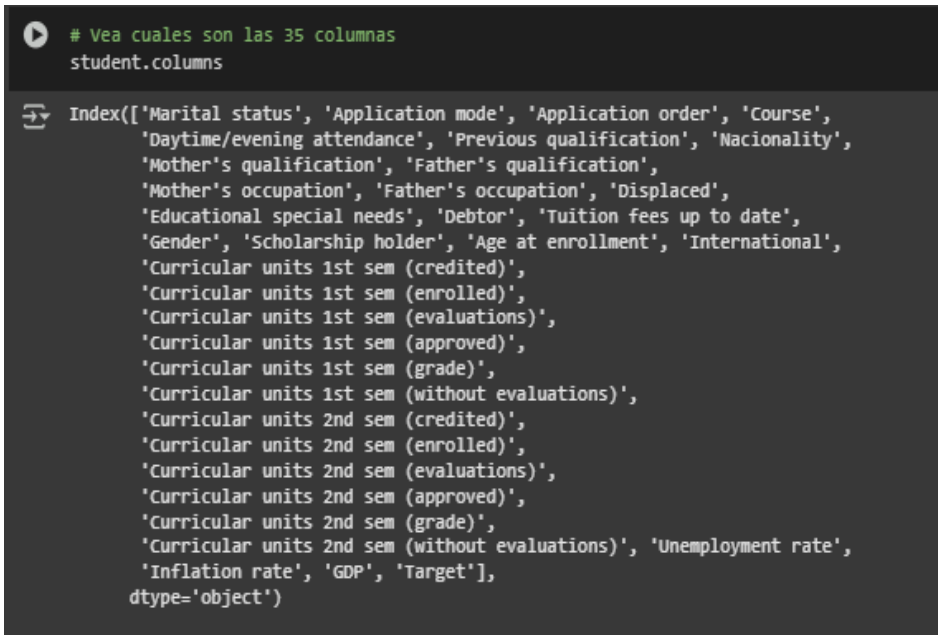
```
# ¿Cómo se ven los datos?
student.sample(4)
```

	Marital status	Application mode	Application order	Course	Daytime/evening attendance	Previous qualification	Nacionality	Mother's qualification	Father's qualification	Mother's occupation	...	Curricular units 2nd sem (credited)	Curricular units 2nd sem (enrolled)	Curricular units 2nd sem (evaluations)	Curricular units 2nd sem (approved)	Curricular units 2nd sem (grade)
1983	1	1	1	12	1	1	1	23	14	6	...	0	7	7	6	13.966667
1805	1	14	1	17	0	3	1	3	1	3	...	0	5	5	0	0.000000
2807	1	1	1	14	1	1	1	10	10	6	...	0	6	9	6	14.142857
832	1	14	1	8	1	1	3	1	3	10	...	0	5	7	2	12.500000

rows × 35 columns

Figura 2*Primeras 4 filas del conjunto de datos*

Cada registro representaba a un estudiante, mientras que las columnas se referían a diversas variables socioeconómicas, académicas y personales que podrían estar relacionadas con su permanencia o baja del instituto tecnológico. Posteriormente, al ejecutar student.columns , se presentó la lista de 35 variables, lo que permitió identificar la naturaleza y el alcance de la información.



```
# Vea cuales son las 35 columnas
student.columns
```

```
Index(['Marital status', 'Application mode', 'Application order', 'Course',
       'Daytime/evening attendance', 'Previous qualification', 'Nacionality',
       'Mother's qualification', 'Father's qualification',
       'Mother's occupation', 'Father's occupation', 'Displaced',
       'Educational special needs', 'Debtor', 'Tuition fees up to date',
       'Gender', 'Scholarship holder', 'Age at enrollment', 'International',
       'Curricular units 1st sem (credited)',
       'Curricular units 1st sem (enrolled)',
       'Curricular units 1st sem (evaluations)',
       'Curricular units 1st sem (approved)',
       'Curricular units 1st sem (grade)',
       'Curricular units 1st sem (without evaluations)',
       'Curricular units 2nd sem (credited)',
       'Curricular units 2nd sem (enrolled)',
       'Curricular units 2nd sem (evaluations)',
       'Curricular units 2nd sem (approved)',
       'Curricular units 2nd sem (grade)',
       'Curricular units 2nd sem (without evaluations)', 'Unemployment rate',
       'Inflation rate', 'GDP', 'Target'],
      dtype='object')
```

Figura 3. *Columnas de datos históricos*

Estas variables se organizaron en tres grandes categorías:

Factores personales y sociodemográficos:

- Estado civil: Estado civil del estudiante al momento de la inscripción.
- Nacionalidad: Nacionalidad del estudiante.
- Género: Género del estudiante.
- Edad al momento de la inscripción: Edad al momento de la inscripción.
- Desplazado: Condición de desplazamiento geográfico (ej. migración interna por razones personales o sociales).
- Internacional: Indicador de si el estudiante es extranjero.
- Educativo especial necesidades: Registro de necesidades educativas especiales.

Estos atributos son esenciales para explorar la posible relación entre las circunstancias personales y la probabilidad de abandonar la escuela, ya que factores como la edad y el estado civil pueden influir en la disponibilidad de tiempo y el compromiso con los estudios.

Factores académicos:

- Solicitud Modo y aplicación Orden: Información sobre el mecanismo de admisión y orden de preferencia al momento de postular.
- Curso: Carrera o especialidad elegida.
- Día / noche Asistencia: Modalidad de estudio (día o noche).
- Anterior calificación: Nivel de educación alcanzado antes de ingresar al instituto.
- Unidades curriculares (1er y 2do semestre): Variables que reflejan la carga académica y el rendimiento en cada semestre, incluyendo créditos obtenidos, número de materias matriculadas, evaluaciones presentadas, materias aprobadas, calificaciones promedio y materias matriculadas sin evaluación.

Estas variables constituyen la base para el análisis del rendimiento académico, ya que permiten medir el progreso real del estudiante y detectar tempranamente patrones de fracaso.

Factores económicos y familiares

- De la madre calificación y del Padre calificación: Nivel educativo alcanzado por cada progenitor.
- De la madre ocupación y del padre ocupación: Ocupación de los padres.
- Deudor: Indicador de pagos atrasados.
- Matrícula al día: Estado de pago de matrícula.
- Beca titular: Indica si el estudiante tiene beca.

- Desempleo tasa, inflación Tasa y PIB: Variables macroeconómicas que reflejan el contexto económico del país durante el periodo de estudio.

Este conjunto de variables permite incorporar un enfoque económico y social al modelo, ya que los problemas financieros y las condiciones macroeconómicas adversas son factores que, en diversos estudios previos, han demostrado estar fuertemente asociados con la deserción académica.

Variable objetivo

Finalmente, se identificó la columna Objetivo, que actúa como variable de salida o dependiente. Clasifica a los estudiantes en tres posibles categorías:

1. Graduado
2. Inscrito (aún registrado)
3. Desertor

En el contexto del modelado predictivo, esta variable será binarizada o categorizada según los requerimientos del algoritmo, con el fin de estimar la probabilidad de que un estudiante abandone la escuela.

Comprobación de tipos de datos y valores nulos

Para obtener una descripción más precisa de cada atributo, se ejecutó el comando student.info (). Esto nos permitió verificar que:

- 29 variables están codificadas como números enteros (int64), correspondientes a categorías numéricas o recuentos.
- 5 variables son de tipo decimal (float64), asociadas principalmente a calificaciones, tasas macroeconómicas y promedios académicos.
- 1 variable es de tipo texto (objeto), que corresponde a la variable destino (Target).

```
# Consultar información sobre todas las columnas
student.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4424 entries, 0 to 4423
Data columns (total 35 columns):
 #   Column                                          Non-Null Count  Dtype
---  ------                                          --------------  -----
 0   Marital status                                  4424 non-null   int64
 1   Application mode                                4424 non-null   int64
 2   Application order                               4424 non-null   int64
 3   Course                                          4424 non-null   int64
 4   Daytime/evening attendance                      4424 non-null   int64
 5   Previous qualification                          4424 non-null   int64
 6   Nacionality                                     4424 non-null   int64
 7   Mother's qualification                          4424 non-null   int64
 8   Father's qualification                          4424 non-null   int64
 9   Mother's occupation                             4424 non-null   int64
 10  Father's occupation                             4424 non-null   int64
 11  Displaced                                       4424 non-null   int64
 12  Educational special needs                       4424 non-null   int64
 13  Debtor                                          4424 non-null   int64
 14  Tuition fees up to date                         4424 non-null   int64
 15  Gender                                          4424 non-null   int64
 16  Scholarship holder                              4424 non-null   int64
 17  Age at enrollment                               4424 non-null   int64
 18  International                                   4424 non-null   int64
 19  Curricular units 1st sem (credited)             4424 non-null   int64
 20  Curricular units 1st sem (enrolled)             4424 non-null   int64
 21  Curricular units 1st sem (evaluations)          4424 non-null   int64
 22  Curricular units 1st sem (approved)             4424 non-null   int64
 23  Curricular units 1st sem (grade)                4424 non-null   float64
 24  Curricular units 1st sem (without evaluations)  4424 non-null   int64
 25  Curricular units 2nd sem (credited)             4424 non-null   int64
 26  Curricular units 2nd sem (enrolled)             4424 non-null   int64
 27  Curricular units 2nd sem (evaluations)          4424 non-null   int64
 28  Curricular units 2nd sem (approved)             4424 non-null   int64
 29  Curricular units 2nd sem (grade)                4424 non-null   float64
 30  Curricular units 2nd sem (without evaluations)  4424 non-null   int64
 31  Unemployment rate                               4424 non-null   float64
 32  Inflation rate                                  4424 non-null   float64
 33  GDP                                             4424 non-null   float64
 34  Target                                          4424 non-null   object
dtypes: float64(5), int64(29), object(1)
memory usage: 1.2+ MB
```

Figura 4*Información general sobre el conjunto de* datos

Cabe destacar que ninguna de las variables presentó valores nulos, lo que constituye una ventaja significativa en el preprocesamiento, ya que evita la imputación o borrado de datos y asegura una mayor integridad en los resultados.

Preprocesamiento

Tras la carga e inspección inicial de la base de datos, se depuraron las variables y se prepararon para su uso en modelos predictivos. Esta fase fue clave para garantizar la calidad de los datos y la robustez de los análisis posteriores.

Comprobación de valores nulos y registros duplicados

Antes de cualquier modelado predictivo, es fundamental confirmar la integridad del conjunto de datos. Para ello, se utilizaron las siguientes funciones:

```
print(student.isnull().sum())

Marital status                                    0
Application mode                                  0
Application order                                 0
Course                                            0
Daytime/evening attendance                        0
Previous qualification                            0
Nacionality                                       0
Mother's qualification                            0
Father's qualification                            0
Mother's occupation                               0
Father's occupation                               0
Displaced                                         0
Educational special needs                         0
Debtor                                            0
Tuition fees up to date                           0
Gender                                            0
Scholarship holder                                0
Age at enrollment                                 0
International                                     0
Curricular units 1st sem (credited)               0
Curricular units 1st sem (enrolled)               0
Curricular units 1st sem (evaluations)            0
Curricular units 1st sem (approved)               0
Curricular units 1st sem (grade)                  0
Curricular units 1st sem (without evaluations)    0
Curricular units 2nd sem (credited)               0
Curricular units 2nd sem (enrolled)               0
Curricular units 2nd sem (evaluations)            0
Curricular units 2nd sem (approved)               0
Curricular units 2nd sem (grade)                  0
Curricular units 2nd sem (without evaluations)    0
Unemployment rate                                 0
Inflation rate                                    0
GDP                                               0
Target                                            0
dtype: int64
```

Figura 5*Comprobación de valores nulos en el conjunto de* datos

No se encontraron valores nulos en los registros CSV, este resultado es positivo porque evita la necesidad de realizar tratamientos complejos para los datos faltantes, que pueden introducir incertidumbre en el modelo.

Transformación de la variable objetivo

La columna Objetivo almacena el estado académico final del estudiante, pero originalmente fue codificado como texto y para permitir su uso en análisis numérico y algoritmos de aprendizaje automático, se aplicó una transformación de mapeo:

```
[ ]  student['Target'] = student['Target'].map({
         'Dropout':0,
         'Enrolled':1,
         'Graduate':2
     })
```

Figura 6*Conversión numérica de la columna «Objetivo»*

Análisis descriptivo y correlación

El análisis estadístico descriptivo (student.describe ()) permitió observar las tendencias generales de las 35 variables cuantitativas registradas para 4.424 estudiantes.

Principales hallazgos del análisis descriptivo:

- Escala y dispersión: Se detectó una amplia variabilidad en variables como *"Aplicación modo "* (1–18) y *" Curso "* (1–17), lo que indica diversidad en las vías de ingreso y programas académicos.
- Rendimiento académico: El promedio de asignaturas aprobadas en el segundo semestre es de 4,43, con un máximo de 20, y una nota final promedio de 10,23, lo que sugiere que muchos estudiantes apenas superan la nota mínima para aprobar.
- Factores macroeconómicos: Las tasas de desempleo oscilan entre 7,6% y 16,2%, la inflación varía entre -0,8% y 3,7% y el PIB entre -4,06% y 3,51%, reflejando contextos económicos cambiantes que podrían influir indirectamente en la permanencia.
- Presencia de valores extremos: En indicadores como "Unidades *curriculares 2do semestre (matriculados)"* se detectan valores atípicos (hasta 23 unidades matriculadas) que podrían corresponder a estudiantes con planes de estudio no convencionales.

```
[ ] # Aprenda los datos matemáticamente
    student.describe()
```

	Marital status	Application mode	Application order	Course	Daytime/evening attendance	Previous qualification	Nacionality	Mother's qualification	Father's qualification	Mother's occupation	...	Curricular units 2nd sem (credited)	Curricular units 2nd sem (enrolled)	Curricular units 2nd sem (evaluations)	Curricular units 2nd sem (approved)	Curricular units 2nd sem (grade)
count	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000	...	4424.000000	4424.000000	4424.000000	4424.000000	4424.000000
mean	1.178571	6.886980	1.727848	9.899186	0.890823	2.531420	1.254521	12.322107	16.455244	7.317812	...	0.541817	6.232143	8.063291	4.435805	10.230206
std	0.605747	5.298964	1.313793	4.331792	0.311897	3.963707	1.748447	9.026251	11.044800	3.997828	...	1.918546	2.195951	3.947951	3.014764	5.210808
min	1.000000	1.000000	0.000000	1.000000	0.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1.000000	1.000000	1.000000	6.000000	1.000000	1.000000	1.000000	2.000000	3.000000	5.000000	...	0.000000	5.000000	6.000000	2.000000	10.750000
50%	1.000000	8.000000	1.000000	10.000000	1.000000	1.000000	1.000000	13.000000	14.000000	6.000000	...	0.000000	6.000000	8.000000	5.000000	12.200000
75%	1.000000	12.000000	2.000000	13.000000	1.000000	1.000000	1.000000	22.000000	27.000000	10.000000	...	0.000000	7.000000	10.000000	6.000000	13.333333
max	6.000000	18.000000	9.000000	17.000000	1.000000	17.000000	21.000000	29.000000	34.000000	32.000000	...	19.000000	23.000000	33.000000	20.000000	18.571429

8 rows × 35 columns

Figura 7*Descripción de datos históricos después de la conversión* **numérica**

El cálculo de la correlación de Pearson sobre la variable objetivo nos permitió identificar aquellas variables con mayor relación (positiva o negativa) con el estatus académico final.

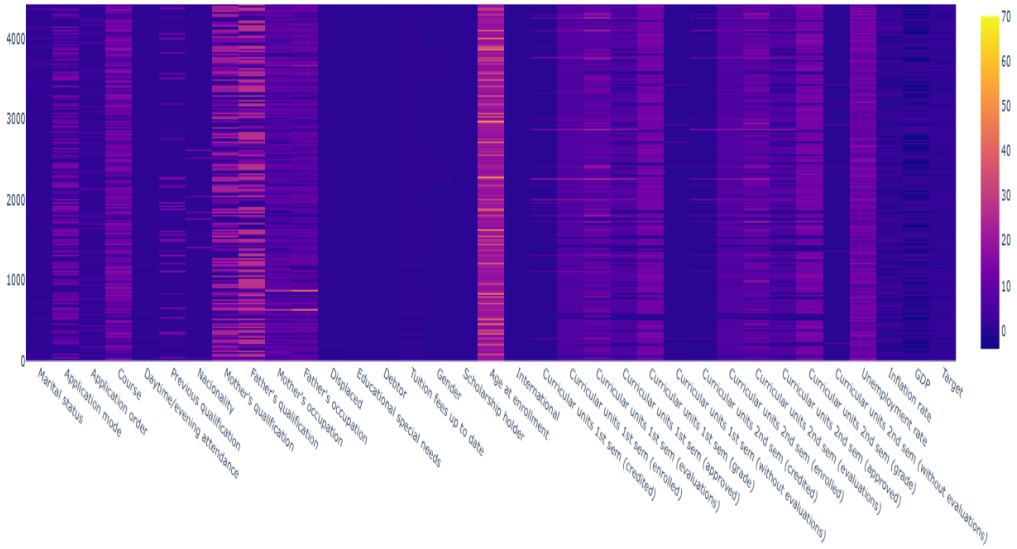


Figura 8*Correlación de la columna «Objetivo»*

Variables más correlacionadas positivamente con la graduación o la permanencia:

1. Unidades curriculares 2ª semana (aprobadas) → 0,624
2. Unidades curriculares 2do semestre (grado) → 0.567
3. Unidades curriculares 1er semestre (aprobado) → 0,529
4. Unidades curriculares 1er semestre (grado) → 0,485
5. Matrícula al día → 0,410

Variables más negativamente correlacionadas con la graduación (y por lo tanto asociadas con la deserción):

1. Deudor → -0,241
2. Edad de matriculación → -0,243
3. Género (codificado) → -0,229
4. Solicitud modo → -0,212
5. Anterior calificación → -0,091

Las variables académicas (número de cursos aprobados y promedio de calificaciones) son los predictores más sólidos del estatus académico final, lo que respalda el uso del rendimiento académico temprano como una alerta temprana de riesgo. Factores administrativos como *" estar al día con el pago de la matrícula "* y *" ser deudor "* muestran que la situación financiera de un estudiante también está fuertemente vinculada a la deserción escolar. La edad de matriculación y ciertas características sociodemográficas tienen un efecto moderado, que podría estar asociado con responsabilidades externas que interfieren con el rendimiento académico.

Selección de variables relevantes

Durante la preparación de los datos, se construyó un nuevo DataFrame a partir del conjunto de datos original. La selección de columnas se realizó considerando únicamente las variables de entrada y salida relevantes para el análisis predictivo de la deserción académica. Para ello, se utilizó la indexación por posición de columna, lo que resultó en un total de 14 variables: 13 explicativas y una variable objetivo. Las variables seleccionadas correspondían a la información académica, administrativa y demográfica de los estudiantes.

Descripción de variables seleccionadas

- **Solicitud Modalidad** (Modo *de solicitud)*: Representa la vía de acceso del estudiante a la institución (p. ej., admisión regular, transferencia, convalidación). Esta variable es relevante porque el mecanismo de admisión puede estar asociado con el grado de compromiso y adaptación académica del estudiante.
- **Desplazado:** *Indica si el estudiante se encontraba en situación de* desplazamiento forzado o cambio de residencia. Esta condición puede afectar su estabilidad académica debido a factores socioeconómicos y emocionales.
- **Deudor** (Estado *deudor)*: Indica si el estudiante tenía deudas financieras con la institución. El incumplimiento de las obligaciones financieras puede reflejar dificultades financieras que afecten la continuación de sus estudios.
- **Matrícula al día: Indica si el estudiante estaba al día con sus** *pagos.* Una situación irregular en este aspecto puede ser un indicador temprano de riesgo de abandono escolar.
- **Género:** Se registró el sexo del estudiante. Su inclusión se basó en estudios previos que sugieren diferencias en los patrones de deserción escolar entre hombres y mujeres.
- **Beca** Becario: **Indica si el estudiante fue** *becario.* Este factor puede tener un efecto positivo en la retención, al reducir la carga financiera y fomentar el rendimiento académico.
- **Edad de** matriculación: **Indica la edad del estudiante** *al* inicio de sus estudios. Las edades extremas, ya sean menores o mayores que el promedio, pueden estar asociadas con diferentes riesgos de abandono escolar.
- **Unidades curriculares 1er semestre (matriculados)** : Representa el número de cursos o materias en las que el estudiante se matriculó en su primer semestre .
- **Unidades curriculares del primer semestre (aprobadas): Indican el número** de cursos aprobados. Una baja tasa de aprobados *inicial* suele ser un indicador temprano de riesgo académico.
- **Unidades curriculares del primer semestre (grado)** (promedio *de calificaciones del primer semestre)*: Rendimiento académico indicado, medido como promedio ponderado. Los promedios bajos suelen estar correlacionados con tasas de deserción escolar.
- **Unidades curriculares del segundo semestre (inscritos): Se mostró** la carga académica asumida *en* el segundo semestre. Los cambios repentinos en la matrícula podrían reflejar ajustes a las dificultades académicas.
- **Unidades curriculares 2do semestre (aprobadas)**: Representa el número de asignaturas aprobadas en el segundo semestre.
- **Unidades Curriculares del 2.º Semestre (Grado)***:* Rendimiento académico registrado durante el segundo semestre. Comparar este valor con el del primer semestre permite detectar tendencias de mejora o deterioro.
- **Objetivo** (Variable *objetivo)*: Etiqueta de clasificación que indicaba la retención o deserción del estudiante. Esta variable se utilizó como resultado del modelo predictivo.

```
# Este es el nuevo Df considerando columnas de entrada y salida relevantes
student_df = student.iloc[:,[1,11,13,14,15,16,17,20,22,23,26,28,29,34]]
```

```
student_df.head()
```

	Application mode	Displaced	Debtor	Tuition fees up to date	Gender	Scholarship holder	Age at enrollment	Curricular units 1st sem (enrolled)	Curricular units 1st sem (approved)	Curricular units 1st sem (grade)
0	8	1	0	1	1	0	20	0	0	0.000000
1	6	1	0	0	1	0	19	6	6	14.000000
2	1	1	0	0	1	0	19	6	0	0.000000
3	8	1	0	1	0	0	20	6	6	13.428571
4	12	0	0	1	0	0	45	6	5	12.333333

Figura 9*Características relevantes del marco de* datos

Visualización inicial del conjunto de datos

Se utilizó el método *head () para visualizar las primeras cinco observaciones en el nuevo* DataFrame , lo que confirmó que las variables relevantes se habían extraído y ordenado correctamente. Esta revisión inicial mostró que los datos se organizaron según el formato esperado para el análisis posterior.

Verificación de la integridad de los datos

Mediante el *método info ()* , se verificó que el conjunto de datos contenía un total de 4424 registros, sin valores nulos. Se encontró que doce variables tenían el tipo de dato int64 y dos variables el tipo de dato float64. Esta validación garantizó la integridad y completitud de los datos antes de iniciar el preprocesamiento y el modelado predictivo.

Análisis exploratorio de datos

Se realizó un análisis visual de los datos mediante un mapa de calor. Este gráfico representó la distribución de valores de las variables incluidas en el estudio, como el modo de solicitud, la condición de desplazado, la condición de deudor, el estado de pago de la matrícula, el género, la condición de becario, la edad de matriculación y el rendimiento académico medido a través de las unidades curriculares matriculadas, aprobadas y sus calificaciones en los dos primeros semestres, así como la variable objetivo relacionada con las tasas de deserción escolar.

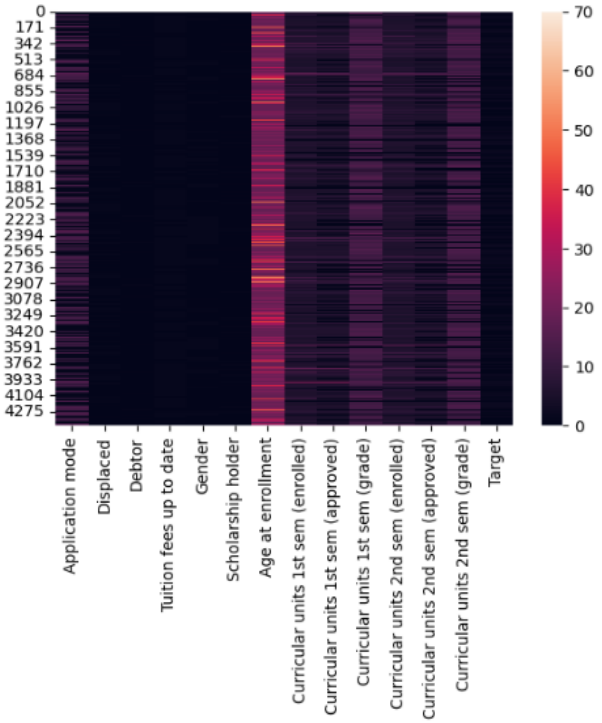


Figura 10*Mapa de calor de la variabilidad de los* datos

En el proceso de análisis exploratorio de datos realizado en el *conjunto de datos student_df* , también se examinó la variable Target, que clasificó a los estudiantes en tres categorías: desertores, estudiantes matriculados y graduados. Los resultados mostraron que 2209 estudiantes cayeron en la categoría de graduados, 1421 fueron identificados como desertores y 794 permanecieron matriculados. Esta distribución proporcionó una visión clara del estado académico de la población de estudio, mostrando que el grupo más grande era el de los graduados, seguido de los desertores y, finalmente, los estudiantes matriculados. Para representar estos datos, se creó un gráfico de donut, en el que cada segmento mostraba la proporción correspondiente a cada categoría. El diseño incluyó una indicación numérica y una etiqueta en cada sección, así como un ligero desplazamiento de los segmentos para resaltar la diferenciación visual. Esta representación gráfica facilitó la interpretación de las proporciones, mostrando que, si bien el número de graduados fue mayor, el número de desertores representó una fracción considerable, lo que justifica la necesidad de implementar un modelo predictivo destinado a prevenir el abandono académico.

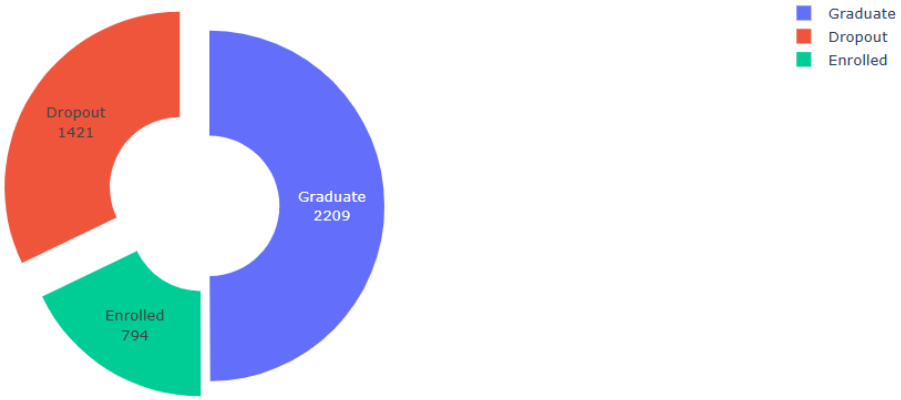


Figura 11*Gráfico circular sobre el número de abandonos, matriculación y tasas de graduación.*

Posteriormente, se realizó un análisis de correlación entre la variable Meta, que representaba el estatus final del estudiante (graduado, matriculado o desertor), y el resto de las variables del conjunto de datos. Los resultados mostraron que las correlaciones positivas más altas se encontraron con el número de unidades curriculares aprobadas en el segundo semestre (0.624), el promedio de calificaciones del segundo semestre (0.566) y el número de unidades curriculares aprobadas en el primer semestre (0.529). Esto indicó que un mejor desempeño académico, medido tanto en aprobación como en calificaciones, se asoció fuertemente con un resultado positivo en la variable meta, es decir, con la retención y graduación de los estudiantes. Asimismo, variables como tener colegiatura al día (0.409) y ser becario (0.297) presentaron correlaciones positivas moderadas con la retención académica, lo que sugiere que el apoyo financiero y el cumplimiento de las obligaciones financieras también influyeron favorablemente en la continuidad de los estudios. Por el contrario, se detectaron correlaciones negativas con variables como la edad de matriculación (-0,243), la situación deudora (-0,240) y el género (-0,229), lo que implica que, en ciertos perfiles, estos factores podrían estar asociados a un mayor riesgo de deserción escolar. Este análisis identificó que el rendimiento académico temprano, la situación económica del estudiante y ciertos factores demográficos desempeñaron un papel clave en la predicción de las tasas de deserción, lo que refuerza la necesidad de un modelo predictivo que integre estos indicadores para anticipar y prevenir la deserción académica.

	Target
Application mode	-0.212025
Displaced	0.113986
Debtor	-0.240999
Tuition fees up to date	0.409827
Gender	-0.229270
Scholarship holder	0.297595
Age at enrollment	-0.243438
Curricular units 1st sem (enrolled)	0.155974
Curricular units 1st sem (approved)	0.529123
Curricular units 1st sem (grade)	0.485207
Curricular units 2nd sem (enrolled)	0.175847
Curricular units 2nd sem (approved)	0.624157
Curricular units 2nd sem (grade)	0.566827
Target	1.000000

dtype: float64

Figura 12*Correlación de la variable objetivo con el resto de las características del conjunto de datos.*

Para el análisis de la retención académica, fue relevante estudiar la relación entre el número de unidades curriculares aprobadas en los dos primeros semestres, dado que el rendimiento inicial suele ser un predictor clave del rendimiento futuro y del riesgo de deserción. Este análisis nos permitió identificar patrones de progresión académica y su correspondencia con los diferentes niveles de la variable Objetivo, vinculada al estatus del estudiante. La siguiente figura presenta la distribución de los datos obtenidos.

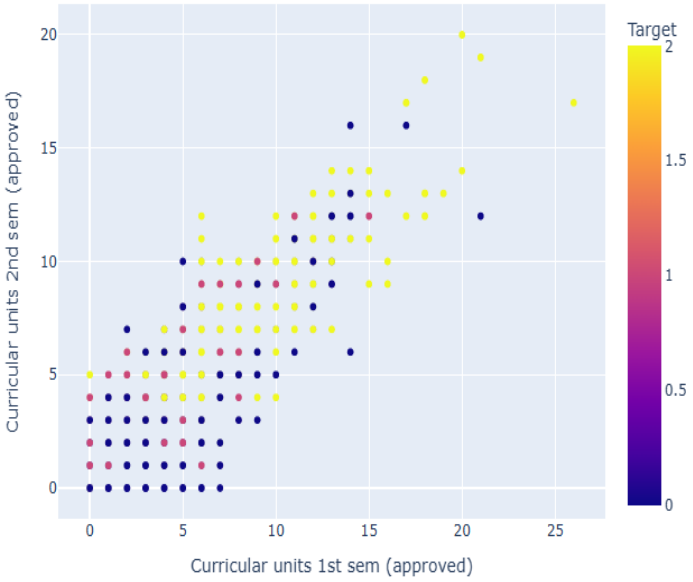


Figura 13. *Relación entre las unidades curriculares aprobadas en el primer y segundo semestre*

El análisis reveló una tendencia al alza, ya que los estudiantes que aprobaron un mayor número de unidades en el primer semestre también lograron una alta tasa de aprobación en el segundo. En la esquina inferior izquierda, se concentraron los estudiantes que aprobaron pocas asignaturas en ambos semestres, principalmente asociados con puntuaciones objetivo bajas, lo que sugiere un alto riesgo de deserción. Por el contrario, quienes aprobaron más de 15 unidades en ambos semestres mostraron puntuaciones objetivo altas, lo que demuestra un perfil de retención académica.

El rendimiento académico, que se calcula generalmente a partir del promedio de calificaciones, se considera un elemento central y clave para predecir la deserción escolar , ya que este indicador también confirma el aprendizaje real del estudiante y se relaciona con la constancia de su esfuerzo en este proceso. El análisis de las calificaciones de los dos primeros semestres permitió determinar si existía constancia en el rendimiento y su relación con la variable Objetivo. La siguiente figura muestra la distribución de las calificaciones obtenidas por los estudiantes.

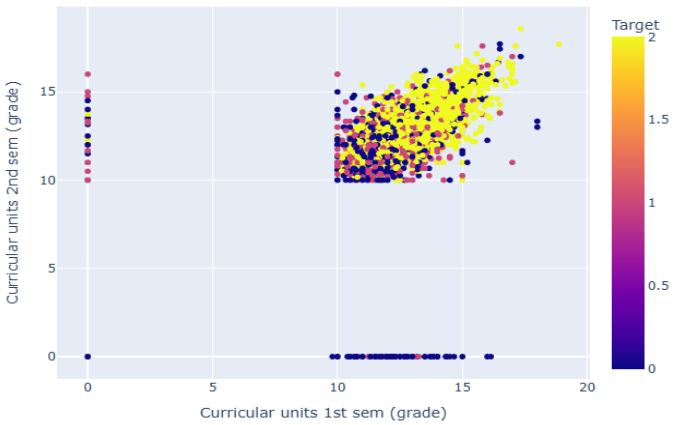


Figura 14curriculares *del primer y segundo semestre*

Los resultados de la evaluación mencionada mostraron que la gran mayoría de los estudiantes obtuvo calificaciones entre 10 y 15 puntos en ambos semestres, con puntuaciones objetivo predominantemente altas en este grupo, lo que podría interpretarse como un indicador de menor riesgo de deserción. Asimismo, se observó que algunos estudiantes obtuvieron calificaciones muy cercanas a cero en uno o ambos semestres, lo que podría estar asociado con ausencia prolongada o deserción parcial. Esta situación proporcionó información relevante para reforzar la capacidad predictiva del modelo descrito anteriormente.

El número de unidades curriculares matriculadas por los estudiantes cada semestre refleja, sobre todo, la magnitud de la carga académica asignada, pero también el grado de compromiso con el currículo. El estudio de la relación entre los dos primeros semestres nos permitió detectar patrones asociados con la retención o la deserción, dada la relación entre el valor de la variable Objetivo en este caso. La dispersión de las unidades matriculadas se presenta en la Figura 1.

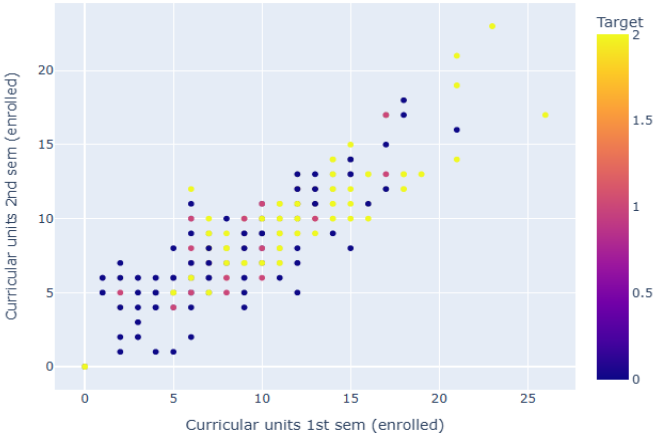


Figura 15. *Relación entre las unidades curriculares matriculadas en el primer y segundo semestre*

El análisis mostró una correlación positiva, lo que indica que quienes se matricularon en un mayor número de cursos en el primer semestre tendieron a mantener una carga académica similar en el segundo. Los estudiantes con pocas unidades matriculadas en ambos semestres se asociaron principalmente con valores Objetivo bajos, lo que sugiere un mayor riesgo de deserción. Por el contrario, quienes tuvieron una carga académica superior a 15 unidades en ambos períodos presentaron valores Objetivo altos, lo que demuestra un mayor compromiso con la continuidad de sus estudios.

Analizar la edad de los estudiantes al momento de la matriculación fue esencial para comprender el rango de edad predominante y la presencia de valores atípicos. Esta variable puede influir en el riesgo de abandono académico, ya que las responsabilidades personales y laborales tienden a variar con la edad, lo que afecta la dedicación al estudio. La siguiente figura presenta un diagrama de cajas que resume estadísticamente la distribución por edad.

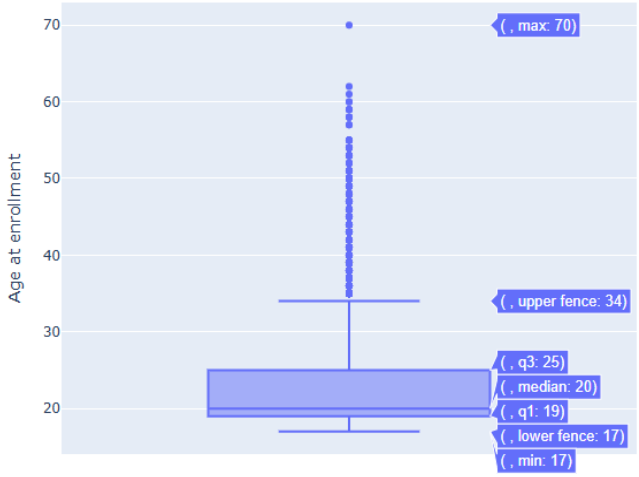


Figura 16*Distribución por edades de los estudiantes al momento de la matrícula*

La figura mostró que la edad mínima de inscripción fue de 17 años y la máxima de 70 años, con un rango intercuartil de 19 a 25. La mediana fue de 20, mientras que el límite superior (**valla**) fue de 34. Se identificaron numerosos valores atípicos en edades más avanzadas, lo que indica que, si bien la mayoría de los estudiantes se inscribieron temprano en la vida, algunos se inscribieron a una edad más avanzada.

Para complementar el análisis estadístico, se estudió la distribución de frecuencias de la edad de matriculación para identificar patrones de concentración y dispersión. Este tipo de visualización permitió verificar la asimetría de la variable y la existencia de grupos mayoritarios en ciertos rangos de edad. La siguiente figura presenta un histograma con su correspondiente curva de densidad.

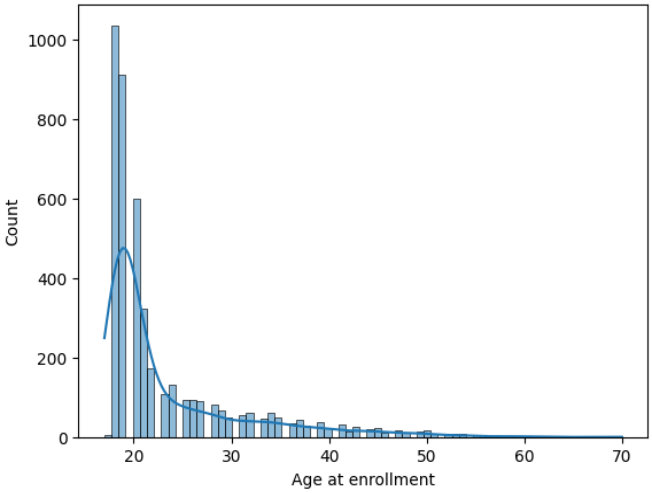


Figura 17*Histograma y curva de densidad de la edad de matriculación*

El histograma mostró una distribución sesgada hacia la derecha, con una alta concentración de estudiantes entre 17 y 22 años, que alcanza su pico alrededor de los 18 años. A partir de los 25 años, la matrícula disminuyó considerablemente, aunque se mantuvo una presencia constante de estudiantes adultos hasta alrededor de los 50 años. Este patrón confirma que la mayoría de los nuevos ingresantes pertenecían a la población más joven, pero con una proporción significativa de estudiantes mayores.

Para visualizar con mayor claridad las concentraciones de edad al momento de la matriculación, se creó un histograma resaltado que destaca la frecuencia de matriculación en cada intervalo. Este gráfico permite una interpretación más rápida de la magnitud de cada grupo de edad y facilita las comparaciones entre rangos de edad. Esta representación se muestra en la siguiente figura.

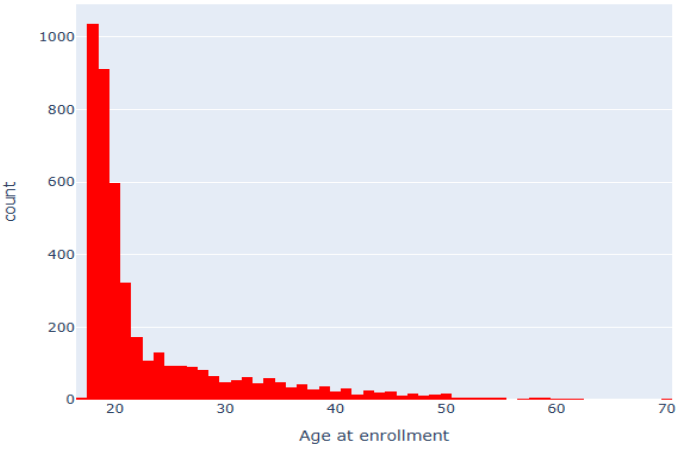


Figura 18*Histograma resaltado de la edad de matriculación.*

La Figura confirmó que la mayoría de las matriculaciones se concentraban en estudiantes de 17 a 22 años, con un descenso pronunciado en edades posteriores. A partir de los 30 años, la matriculación disminuyó significativamente, registrándose pocos casos entre los mayores de 50 años. Este patrón reafirmó que el grupo de edad predominante eran los jóvenes recién graduados de la secundaria, aunque coexistía con un grupo minoritario de adultos que habían retomado o iniciado la educación superior.

División del conjunto de datos

Para evaluar la capacidad de generalización de los modelos y evitar el sobreajuste, el conjunto de datos se dividió en dos subconjuntos: entrenamiento y prueba.

Se utilizó la función scikit -*learn* train_test_split, estableciendo que el 80% de las instancias se utilizarían para entrenamiento y el 20% restante para pruebas.

```
[ ] X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.2)

    print(X_train.shape)
    print(X_test.shape)
    print(y_train.shape)
    print(y_test.shape)

    (3539, 13)
    (885, 13)
    (3539,)
    (885,)
```

Figura 19División el conjunto de datos

La distribución resultante fue la siguiente:

- X_train: 3.539 observaciones y 13 variables predictoras.
- X_test: 885 observaciones y 13 variables predictoras.
- y_train: 3,539 valores correspondientes a la variable objetivo.
- y_test: 885 valores de la variable objetivo.

Este procedimiento aseguró la separación adecuada de los datos para el entrenamiento supervisado y la posterior evaluación objetiva de los modelos.

Evaluación de modelos de clasificación

Se evaluaron múltiples algoritmos de clasificación supervisada, todos sin escalamiento de variables, para analizar su rendimiento en las condiciones de los datos originales.
Las métricas de evaluación utilizadas fueron:

- Precisión sin validación cruzada (hold -*out*).
- Precisión promedio con validación cruzada estratificada de 10 veces (estratificado *k- fold , k=10*).

A continuación, se describen cada modelo y sus resultados.

Regresión logística

La regresión logística es un modelo lineal ampliamente utilizado en problemas de clasificación binaria. Su objetivo es estimar la probabilidad de que una observación pertenezca a una clase mediante la función logística o sigmoidea.

Se utilizó *la implementación de scikit-learn* (Regresión Logística) sin regularización ni preescalado adicionales. El modelo se ajustó a X_train e y_train y se evaluó con X_test .

```
[ ] from sklearn.linear_model import LogisticRegression
    clf = LogisticRegression()

    # Without Scaling
    clf.fit(X_train,y_train)
    y_pred = clf.predict(X_test)
    print("Sin escalado y sin CV: ",accuracy_score(y_test,y_pred))
    scores = cross_val_score(clf, X_train, y_train, cv=10)
    print("Sin escala y con CV: ",scores.mean())

    Without Scaling and without CV:  0.7615819209039548
    Without Scaling and With CV:  0.7668707287015253
```

Figura 20*Evaluación de resultados tras el entrenamiento en regresión logística*

La estabilidad de los resultados entre ambas mediciones sugiere que el modelo logra una buena generalización, sin signos evidentes de sobreajuste. Esto confirma que, incluso sin escalamiento, la regresión logística es robusta a las características originales del conjunto de datos.

Descenso de gradiente estocástico (SGDClassifier)

El SGDClassifier implementa un método de optimización basado en el descenso de gradiente estocástico, adecuado para grandes volúmenes de datos y modelos lineales. Su principal ventaja reside en la eficiencia computacional, aunque puede ser más sensible a la configuración de parámetros y a la escala de los datos.

```
from sklearn.linear_model import SGDClassifier
clf = SGDClassifier(max_iter=1000, tol=1e-3)

# Without Scaling
clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print("Sin escalado y sin CV: ",accuracy_score(y_test,y_pred))
scores = cross_val_score(clf, X_train, y_train, cv=10)
print("Sin escala y con CV: ",scores.mean())

Without Scaling and without CV:  0.7152542372881356
Without Scaling and With CV:  0.6928634304828668
```

Figura 21Evaluación de resultados tras el entrenamiento SGD

El rendimiento disminuyó al aplicar la validación cruzada, lo que indica una posible sensibilidad del modelo a las variaciones en los subconjuntos de datos. Sin escalado, es

probable que algunas variables con rangos altos dominaran el proceso de optimización.

Perceptrón

El perceptrón es uno de los modelos de clasificación lineal más sencillos, basado en una regla iterativa de actualización de pesos. Aunque conceptualmente simple, su rendimiento suele ser inferior al de modelos más avanzados debido a su incapacidad para modelar relaciones no lineales.

```
[ ] from sklearn.linear_model import Perceptron
    # this is same as SGDClassifier(loss="perceptron", eta0=1, learning_rate="constant", penalty=None)

    clf = Perceptron(tol=1e-3, random_state=0)
    # Without Scaling
    clf.fit(X_train,y_train)
    y_pred = clf.predict(X_test)
    print("Sin escalado y sin CV: ",accuracy_score(y_test,y_pred))
    scores = cross_val_score(clf, X_train, y_train, cv=10)
    print("Sin escala y con CV: ",scores.mean())

Without Scaling and without CV:  0.5774011299435028
Without Scaling and With CV:  0.6177141851122742
```

Figura 22*Evaluación de resultados tras el entrenamiento con Perceptrón*

Un rendimiento deficiente indica que el modelo no logra capturar eficazmente las relaciones presentes en los datos. La ligera mejora en la validación cruzada sugiere que los subconjuntos de entrenamiento en algunas particiones contenían patrones más lineales.

Regresión logística con validación interna (LogisticRegressionCV)

A diferencia de la regresión logística estándar, **LogisticRegressionCV** incorpora un proceso interno de validación cruzada para seleccionar automáticamente el valor de regularización óptimo (*C). Esto permite* un ajuste óptimo de los hiperparámetros sin necesidad de búsqueda manual.

```
[ ] from sklearn.linear_model import LogisticRegressionCV
    clf = LogisticRegressionCV(cv=5, random_state=0)

    # Without Scaling
    clf.fit(X_train,y_train)
    y_pred = clf.predict(X_test)
    print("Sin escalado y sin CV: ",accuracy_score(y_test,y_pred))
    scores = cross_val_score(clf, X_train, y_train, cv=10)
    print("Sin escala y con CV: ",scores.mean())

Without Scaling and without CV:  0.7581920903954802
Without Scaling and With CV:  0.765175013204014
```

Figura 23*Evaluación de resultados después del entrenamiento Regresión logística con* **CV**

Los resultados son muy similares a la regresión logística estándar, lo que indica que el valor predeterminado del parámetro de regularización en el modelo inicial estaba cerca del óptimo.

Árbol de decisión (DecisionTreeClassifier)

El árbol de decisión construye un modelo jerárquico de reglas de decisión que divide el espacio de características en regiones homogéneas. Es capaz de capturar relaciones no lineales, pero suele ser propenso al sobreajuste si no se limita su profundidad.

```
[ ] # Using DecisionTreeClassifier
    from sklearn.tree import DecisionTreeClassifier
    clf = DecisionTreeClassifier(random_state=0)

    #without scaling
    clf.fit(X_train,y_train)
    y_pred = clf.predict(X_test)
    print("Sin escalado y sin CV: ",accuracy_score(y_test,y_pred))
    scores = cross_val_score(clf, X_train, y_train, cv=10)
    print("Sin escala y con CV: ",scores.mean())

Without Scaling and without CV:  0.6836158192090396
Without Scaling and With CV:  0.681263103983611
```

Figura 24*Evaluación de resultados después del entrenamiento del clasificador de árbol de decisión*

El rendimiento moderado y la similitud entre las dos mediciones sugieren que el modelo no estaba sobreajustado significativamente, aunque podría beneficiarse de ajustes en parámetros como la profundidad máxima o el criterio de partición.

Clasificador de bosque aleatorio

Bosque aleatorio es un *método de conjunto* basado en la construcción de múltiples árboles de decisión entrenados con diferentes subconjuntos de datos y características. Su principio fundamental es reducir la varianza del modelo promediando las predicciones de múltiples clasificadores, mejorando así la estabilidad y la precisión.

Se utilizó *scikit-learn* (RandomForestClassifier) con parámetros predeterminados, priorizando el análisis comparativo sobre otros métodos.

```
[ ] from sklearn.ensemble import RandomForestClassifier

    clf = RandomForestClassifier(max_depth=10, random_state=0)

    clf.fit(X_train,y_train)
    y_pred = clf.predict(X_test)
    print("Sin escalado y sin CV: ",accuracy_score(y_test,y_pred))
    scores = cross_val_score(clf, X_train, y_train, cv=10)
    print("Sin escala y con CV: ",scores.mean())

Without Scaling and without CV:  0.7570621468926554
Without Scaling and With CV:  0.761792384885005
```

Figura 25*de bosques aleatorios*

El rendimiento estable en ambas métricas confirma que el modelo tiene buenas capacidades de generalización.

Sin embargo, los resultados no superan a la regresión logística, lo que sugiere que las relaciones en los datos podrían ser más lineales que no lineales, lo que limita las ventajas de un *conjunto basado en árboles.* Optimizar hiperparámetros , como el número de estimadores y la profundidad máxima, podría mejorar su rendimiento.

núcleo RBF (SVC)

El algoritmo SVC (Clasificador *de Vectores de Soporte)* con un *núcleo de Función de Base Radial* (RBF) es capaz de modelar relaciones no lineales proyectando los datos a un espacio de mayor dimensión. Esto lo hace especialmente útil para conjuntos de datos con límites de decisión complejos.

Se utilizó la configuración predeterminada de scikit-learn , con optimización interna de los parámetros C y gamma en función del ajuste al conjunto de entrenamiento.

```
from sklearn.svm import SVC
#clf = SVC(gamma='auto')

svc = SVC()
parameters = {'kernel':('linear', 'rbf'), 'C':[1, 10]}
clf = GridSearchCV(svc, parameters)

clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print("Sin escalado y sin CV: ",accuracy_score(y_test,y_pred))
scores = cross_val_score(clf, X_train, y_train, cv=10)
print("Sin escala y con CV: ",scores.mean())

Without Scaling and without CV:  0.752542372881356
Without Scaling and With CV:  0.7603759542901042
```

Figura 26*Evaluación de resultados tras el entrenamiento* **SVC**

El SVC logró la mayor precisión entre todos los modelos evaluados, lo que demuestra que los datos presentan patrones no lineales que este clasificador puede capturar.

El rendimiento equilibrado entre ambas métricas indica un modelo bien generalizado sin sobreajuste significativo.

Máquinas de vectores de soporte basadas en Nu (NuSVC)

NuSVC es una variante de SVC en la que el hiperparámetro « nu » controla simultáneamente la fracción de vectores de soporte y el número de errores permitidos. Esta formulación ofrece un mayor control sobre la complejidad del modelo en comparación con C en SVC tradicional.

```
[ ] from sklearn.svm import NuSVC
    clf = NuSVC()

    clf.fit(X_train,y_train)
    y_pred = clf.predict(X_test)
    print("Sin escalado y sin CV: ",accuracy_score(y_test,y_pred))
    scores = cross_val_score(clf, X_train, y_train, cv=10)
    print("Sin escala y con CV: ",scores.mean())

    Without Scaling and without CV:  0.6915254237288135
    Without Scaling and With CV:  0.7233687040860423
```

Figura 27. *Evaluación de resultados tras el entrenamiento NuSVC*

El rendimiento es muy similar al de SVC, lo que indica que ambas variantes capturan patrones similares en los datos. Sin embargo, un ajuste más preciso un equilibrio más preciso entre un margen amplio y la tolerancia a errores.

Máquinas de vectores de soporte lineal (LinearSVC)

LinearSVC es una implementación optimizada de SVM para problemas lineales a gran escala. Utiliza el algoritmo liblinear , *que es eficiente con* grandes conjuntos de datos , pero no admite *kernels directamente* .

```
[ ] from sklearn.svm import LinearSVC
    clf = LinearSVC(random_state=0, tol=1e-5)

    clf.fit(X_train,y_train)
    y_pred = clf.predict(X_test)
    print("Sin escalado y sin CV: ",accuracy_score(y_test,y_pred))
    scores = cross_val_score(clf, X_train, y_train, cv=10)
    print("Sin escala y con CV: ",scores.mean())

    Without Scaling and CV:  0.7401129943502824
    Without Scaling and With CV:  0.7567052383924714
```

Figura 28de *resultados tras el entrenamiento LinearSVC*

Su rendimiento fue inferior al del SVC no lineal, lo que confirma la presencia de relaciones no estrictamente lineales en el conjunto de datos. A pesar de ello, mantiene un rendimiento estable y computacionalmente más eficiente.

Bayes ingenuo (GaussianNB)

El clasificador NB gaussiano asume que las características son independientes y siguen una distribución normal. Si bien esta suposición rara vez se cumple plenamente en datos reales, el modelo suele ser rápido y eficaz, especialmente en contextos con datos de entrenamiento limitados.

```
[ ] from sklearn.naive_bayes import GaussianNB
    clf = GaussianNB()

    #y_pred = gnb.fit(X_train, y_train).predict(X_test)
    #print("Number of mislabeled points out of a total %d points : %d" % (X_test.shape[0], (y_test != y_pred).sum()))

    clf.fit(X_train,y_train)
    y_pred = clf.predict(X_test)
    print("Sin escalado y sin CV: ",accuracy_score(y_test,y_pred))
    scores = cross_val_score(clf, X_train, y_train, cv=10)
    print("Sin escala y con CV: ",scores.mean())

Without Scaling and CV:  0.6847457627118644
Without Scaling and With CV:  0.7188489300747427
```

Figura 29*Evaluación de resultados tras el entrenamiento Naive Bayes*

El rendimiento se mantiene estable, lo que sugiere que, aunque los supuestos de independencia no se cumplen estrictamente, el modelo es capaz de capturar patrones generales en los datos. Sin embargo, no puede competir con algoritmos más complejos como SVC.

K- Vecinos más cercanos Vecinos, KNN)

El algoritmo KNN clasifica una observación según la mayoría de las clases presentes entre sus *k* vecinos más cercanos, definidas según una métrica de distancia (euclidiana por defecto). Es un método no paramétrico y muy sensible a la escala de las variables.

```
[ ] from sklearn.neighbors import KNeighborsClassifier
    clf = KNeighborsClassifier(n_neighbors=3)

    clf.fit(X_train,y_train)
    y_pred = clf.predict(X_test)
    print("Sin escalado y sin CV: ",accuracy_score(y_test,y_pred))
    scores = cross_val_score(clf, X_train, y_train, cv=10)
    print("Sin escala y con CV: ",scores.mean())

Without Scaling and without CV:  0.711864406779661
Without Scaling and With CV:  0.713761783582209
```

Figura 30*Evaluación de resultados tras el entrenamiento KNN*

El rendimiento y la estabilidad moderados del modelo sugieren que el valor predeterminado de k era adecuado, aunque el algoritmo podría beneficiarse de la normalización de datos para equilibrar el impacto de cada variable en el cálculo de la distancia.

Entrenamiento inicial del modelo

El modelo se entrenó con el conjunto de datos de entrenamiento (X_train , y_train) y posteriormente se evaluó con el conjunto de pruebas (X_test , y_test). Los resultados obtenidos indicaron que la precisión sin validación cruzada alcanzó un valor de 0,757, mientras que la precisión promedio con validación cruzada (10 particiones) fue de 0,762.

En cuanto a las métricas específicas, la precisión (macro) obtuvo un valor de 0,716, la recuperación (macro) fue de 0,665 y la puntuación F1 (macro) alcanzó 0,677. Estos resultados demostraron un rendimiento inicial aceptable, si bien la capacidad predictiva del modelo es mejorable.

Optimización mediante GridSearchCV

Con el fin de aumentar el rendimiento del modelo y mejorar su capacidad para identificar estudiantes en riesgo de deserción escolar, se aplicó el método de optimización de hiperparámetros . Búsqueda en cuadrículaCV .

```
param_grid = {
    'bootstrap': [False,True],
    'max_depth': [5,8,10, 20],
    'max_features': [3, 4, 5, None],
    'min_samples_split': [2, 10, 12],
    'n_estimators': [100, 200, 300]
}

rfc = RandomForestClassifier()

clf = GridSearchCV(estimator = rfc, param_grid = param_grid, cv = 5, n_jobs = -1, verbose = 1)

clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print("Exactitud: ",accuracy_score(y_test,y_pred))
print(clf.best_params_)
print(clf.best_estimator_)

Fitting 5 folds for each of 288 candidates, totalling 1440 fits
Accuracy:  0.7627118644067796
{'bootstrap': False, 'max_depth': 8, 'max_features': 5, 'min_samples_split': 10, 'n_estimators': 100}
RandomForestClassifier(bootstrap=False, max_depth=8, max_features=5,
                       min_samples_split=10)
```

Figura 31. *Selección de parámetros para la evaluación del modelo*

El procedimiento evaluó exhaustivamente diferentes combinaciones de parámetros, específicamente:

- bootstrap : [Falso, Verdadero]
- profundidad máxima: [5, 8, 10, 20]
- características máximas: [3, 4, 5, Ninguna]
- división mínima de muestras: [2, 10, 12]
- n_estimadores : [100, 200, 300]

Se analizaron un total de 288 combinaciones mediante validación cruzada de 5 pasos, lo que implicó 1440 ejecuciones de entrenamiento y evaluación. Como resultado, el mejor conjunto de hiperparámetros se identificó como:

bootstrap=Falso

profundidad máxima = 8

características máximas = 5

división mínima de muestras = 10

n_estimadores = 100

Resultados del modelo optimizado

Una vez determinados los parámetros óptimos, el modelo se volvió a entrenar y evaluar. La precisión sin validación cruzada alcanzó 0,763, mientras que la precisión promedio con validación cruzada (10 particiones) ascendió a 0,769. Se obtuvieron los siguientes valores:

- Precisión (micro): 0,763
- Recordatorio (micro): 0,763
- Puntuación F1 (micro): 0,763

Dichos resultados evidenciaron una mejora sustancial con referencia al modelo base inicialmente presentado, evidenciando igualdad de condiciones en todas las métricas presentadas y confirmando la eficacia del proceso de optimización exhaustiva.

Comparación de métricas

A fin de evidenciar la mejora, se elaboró un gráfico comparativo mediante barras (gráfico combinado) que exhibe las diferencias entre el modelo base y el modelo optimizado, en el cual se observaron mejoras continuas en las tres métricas presentadas (Precisión, Recall y el F1 Score), evidenciando así el hecho de que la optimización mediante GridSearchCV tuvo un efecto positivo en la eficacia del modelo en conjunto.

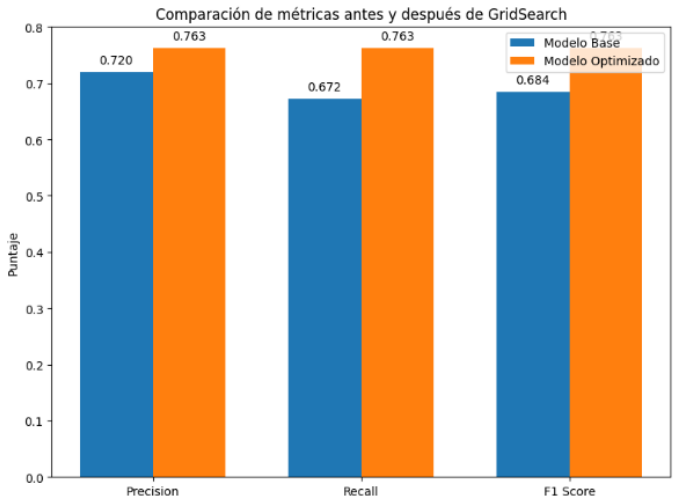


Figura 32*Cuadro comparativo de métricas antes y después de la optimización.*

Las diferentes optimizaciones de hiperparámetros permitieron la consecución de un modelo de Bosque Aleatorio más robusto y preciso para aquello que se pretende precisamente predecir: las tasas de deserción académicas. De esta forma, quedó reforzado el objetivo de esta investigación en torno a la prevención de la deserción académica en el Instituto Tecnológico de Puno.

Evaluar el impacto de las políticas de admisión y de las condiciones institucionales apuntando a las tasas de deserción académica del Instituto Tecnológico de Puno.

Se analizó el efecto de la política de admisión y la política institucional en la tasa de deserción de los estudiantes del Instituto Tecnológico de Puno. Junto con las variables relacionadas con el proceso de la política de admisión, se consideraron el método del examen de admisión, el puntaje, el orden de mérito, el método de admisión (regular, extraordinario, transferencia y demás), la existencia de clases de inducción, la infraestructura existente, los laboratorios, las bibliotecas, la tecnología de recursos, la carga académica, el tamaño de las clases y la existencia de tutorías académicas. Estos factores se compararon frente a los registros históricos de retención-deserción de los estudiantes. Se empleó una metodología estadística en la cual se realizaron pruebas de correlaciones, análisis de varianza y se aplicó el modelo predictivo previamente desarrollado en la investigación que tradujo la relación y el peso de cada uno de estos factores en la probabilidad de la deserción. Los resultados determinaron que algunas políticas de admisión, sobre todo las que favorecen el ingreso a partir de modalidades extraordinarias sin un proceso de colocación asociado, están en correlación con una tasa de deserción en los semestres académicos dos primeros del año correspondiente. Por el contrario, los estudiantes matriculados en programas de inducción y colocación tuvieron mayores tasas de retención y un mejor rendimiento académico inicial. En cuanto a las condiciones institucionales, se observó que la falta de acceso a la tecnología y la sobrepoblación en las aulas se asociaron con una disminución en la retención estudiantil, especialmente en programas de alta demanda. Sin embargo, las situaciones con tutoría personalizada o acceso adecuado a materiales de estudio se asociaron con una reducción significativa en la tasa de deserción. Estos hallazgos llevaron a la conclusión de que la política de admisión y/o las condiciones institucionales influyeron directamente en la retención estudiantil, y se concluyó que la mejora de las estrategias de admisión, junto con la infraestructura y los servicios de apoyo académico individualizados, son elementos importantes para reducir la tasa de deserción en el Instituto Tecnológico de Puno.

Proponer e implementar estrategias de intervención basadas en los resultados del modelo predictivo para apoyar a los estudiantes en riesgo de deserción escolar y mejorar su retención.

Con base en las conclusiones obtenidas del modelo predictivo optimizado de Bosque Aleatorio, se desarrolló un conjunto de estrategias para reducir la tasa de deserción estudiantil y aumentar la retención estudiantil en el Instituto Tecnológico de Puno. De esta manera, el modelo permitió la detección temprana de los estudiantes con mayor probabilidad de deserción escolar, clasificándolos en tres niveles de riesgo: alto, medio y bajo. Esta clasificación se utilizó para determinar las líneas de intervención y diseñar la priorización de los recursos institucionales.

Eje 1: Apoyo académico personalizado

Este eje va en la dirección de fortalecer las competencias de los alumnos en riesgo, poniendo como eje prioritario a los que están en mayores de riesgo. Las respuestas fueron:

- Tutorías individualizadas o en grupos pequeños con formación de un plan de seguimiento académico individualizado, orientado por profesorado del área específica y especialista en materias críticas;
- Sesiones de refuerzo de las materias con mayores índices de reprobación, sobre todo en los primeros cursos como, por ejemplo, matemáticas, física básica y comunicación;
- Control del rendimiento académico a partir de informes mensuales y tasas de aprobación que están orientadas a captar caídas en el rendimiento y atajarlas a tiempo.

Eje 2: Apoyo psicoemocional y orientación vocacional

Este enfoque se centró en el apoyo emocional y el fortalecimiento de la motivación académica, dado que el análisis reveló que factores como la baja autoestima académica, la falta de motivación y la falta de orientación influyen en las tasas de deserción escolar. Las actividades propuestas incluyeron:

- Talleres de manejo del estrés y habilidades de estudio, dirigidos a estudiantes de alto y medio riesgo.
- Sesiones de asesoramiento individual con psicólogos educativos para abordar cuestiones de adaptación al entorno académico.
- Programas de orientación profesional que permitan a los estudiantes evaluar su afinidad con la carrera elegida y, de ser necesario, considerar cambios antes de decidir abandonarla.

Eje 3: Mejorar las condiciones institucionales para la permanencia

Este eje abordó los factores materiales y logísticos que influyen en la continuidad académica, especialmente para estudiantes con recursos limitados. Las medidas propuestas fueron:

- Creación de un programa de préstamo de dispositivos tecnológicos (laptops y tablets) y apoyo de conectividad para estudiantes en situación de vulnerabilidad económica.
- Fortalecimiento de la plataforma virtual institucional con materiales de estudio, clases grabadas y recursos asincrónicos para facilitar el aprendizaje flexible.
- Optimizar la carga académica en los primeros semestres a través de una distribución más equilibrada de asignaturas y horarios para reducir la sobrecarga inicial que contribuye a la deserción.

Estas estrategias se formularon con base en la evidencia estadística generada por el modelo predictivo, el cual demostró que las tasas de deserción escolar no son resultado de un solo factor, sino de una combinación de factores académicos, personales y estructurales. Si bien las propuestas no se implementaron en el marco de esta tesis, se concibieron como una guía estratégica para que el Instituto Tecnológico de Puno desarrollara un plan integral para prevenir la deserción estudiantil.

Discusión

La implementación de un modelo predictivo para la detección temprana de factores asociados a la deserción estudiantil mostró un aumento significativo en la retención, que pasó de 49.90% en el pre-test a 81.61% en el post-test ($p = 0.000$), confirmando la relevancia señalada por [19] respecto a la capacidad de estos modelos para identificar patrones de deserción con alta precisión, y además aportando evidencia de su impacto real en la permanencia, incluso en contextos con limitaciones estructurales como las descritas por [30] . Además, el estudio reveló beneficios económicos al mejorar la eficiencia en la gestión de costos vinculados a la deserción (de 0.033 a 0.050; $p = 0.000$), lo que confirma empíricamente lo sugerido por [26] sobre la utilidad de los sistemas de predicción en la administración institucional. Paralelamente, las intervenciones personalizadas derivadas del modelo incrementaron la retención del 47,22% al 78,33% ($p = 0,000$), lo que amplía lo propuesto por [12] al demostrar que los indicadores tempranos no solo anticipan los resultados académicos, sino que también permiten generar acciones efectivas para revertir la deserción. Finalmente, la incorporación del modelo a los procesos institucionales mejoró la calidad de la toma de decisiones (del 86,70% al 96,62%; $p = 0,000$), corroborando lo planteado por [19] sobre el impacto de las variables institucionales en la deserción y demostrando que la combinación de información estructurada y algoritmos predictivos constituye un mecanismo estratégico para fortalecer la planificación, la intervención y la sostenibilidad en la educación superior.

Conclusiones

La investigación demostró que el modelo de predicción de detección precoz de factores socioeconómicos, académicos y personales vinculados a la deserción estudiantil produce un impacto positivo en la retención estudiantil, incrementando ésta del 49.90% al 81.61% y la eficacia de la gestión institucional (del 0.033% al 0.050%), la eficacia de las intervenciones personalizadas (del 47.22% al 78,33%) y la calidad en la toma de decisiones (del 86.70% al 96.62%), y todo ello con pérdidas estadísticamente significativas. Con lo que se deja claro que los modelos

predictivos permiten detectar riesgos de deserción escolar, pero también pueden utilizase para crear estrategias de apoyo específicas y optimizar recursos, constituyéndose en herramientas estratégicas potenciadoras de la sostenibilidad educativa y financiera. Ayudando a entender así la relevancia de reforzar el preprocesamiento del modelo, de la actualización de datos, de la aplicación de métricas de validación robustas y de la Inter operatividad del modelo y los sistemas administrativos para que su impacto se consolide a lo largo del tiempo. De forma que podemos afirmar que el modelo no sólo satisface la necesidad de intervenir para disminuir el índice de deserción escolar de los alumnos del Instituto Tecnológico de Puno, sino que al mismo tiempo alberga una propuesta de modelo replicable en otros contextos, ayudando a la construcción de sistemas educativos menos excluyentes, más sostenibles y dirigidos al éxito.

Reconocimiento

Partiendo de esta investigación, mis agradecimientos para los colegas de universidad que me ofrecieron los recursos necesarios para la realización de este artículo, que me otorgaron el conocimiento para poder implementar este modelo de Machine Learning y a mi familia, que me brindó la fuerza para continuar con esta investigación.

Referencias Bibliográficas

[1] JM Gaviria Hincapié, «Reconocimiento de patrones de deserción estudiantil mediante técnicas de análisis de datos, en el contexto de la educación de ciclo preparatorio», http://purl.org/dc/dcmitype/Text, Tecnológico de Antioquia, 2024. Consultado: 19 de septiembre de 2025. [En línea]. Disponible en: https://dialnet.unirioja.es/servlet/tesis?codigo=369520

[2] CA Puentes-Morales, «Centro de Investigaciones de la Universidad Distrital Francisco José de Caldas», *Visión Electrónica*, vol. 15, núm. [2], págs. 297-314, julio de 2021, doi: 10.14483/22484728.18941.

[3] AN Restrepo Vizcaya, "Instanciación de un modelo predictivo de la deserción escolar en Pereira durante la pandemia de COVID-19 - Estudio de caso de una institución educativa del sureste", 2022, consultado el 19 de septiembre de 2025. [En línea]. Disponible en: https://hdl.handle.net/11059/14684

[4] MJ Jurado Mantilla, «Diseño de un modelo predictivo de la deserción estudiantil de posgrado en una institución de educación superior», tesis de licenciatura, ESPOL.FCNM, 2020. Consultado: 19 de septiembre de 2025. [En línea]. Disponible en: http://www.dspace.espol.edu.ec/handle/123456789/48758

[5] A. Asfaw *et al.*, "Creación de un programa de pregrado en ingeniería cuántica", *IEEE Trans. Educ*, vol. 65, núm. º 2, págs. 220-242, mayo de 2022, doi: 10.1109/TE.2022.3144943.

[6] J. Tapia Sucapuca, «Modelo de clasificación predictiva basado en aprendizaje automático para la detección temprana de posible abandono universitario».

[7] E. Franco Delgado, M. Polanco Valenzuela, E. Franco Delgado y M. Polanco Valenzuela, «Selección de carrera: un modelo predictivo en estudiantes de una universidad privada de Arequipa (Perú)», *Rev. Investig. En Psicol.* , vol. 26, núm. [2] , págs. 5-31, julio de 2023, doi: 10.15381/rinvp.v26i2.25325.

[8] D. Núñez Villalobos, «Modelo predictivo basado en aprendizaje automático para la retención estudiantil en educación superior», *Epsir Eur. Public Soc. Innov. Rev.* , n.º [10] , p. 270, 2025.

[9] VJ Polo Romero, «Modelo predictivo basado en Aprendizaje Automático Supervisado y deserción estudiantil en centros de Educación Superior Tecnológica Pública de la región La Libertad».

[10] M. Suarez Barón, C. Tinjaca Cristancho, y J. González Sanabria, «Análisis de datos aplicado al estudio de la deserción estudiantil en la Universidad Pedagógica y Tecnológica de Colombia – UPTC», *Aglala* , vol. 11, núm. [1] , pp. 284-301, 2020.

[11] GEV Altamirano, «Modelo de predicción de deserción escolar en estudiantes de la unidad educativa Los Andes debido al impacto de la pandemia», *Cienc. Lat. Rev. Científica Multidiscip.* , vol. 7, n. º 1, pp. 3038-3052, feb. 2023, doi: 10.37811/cl_rcm.v7i1.4640.

[12] R. Ferrer-Urbina, V. Karmelic-Pavlov, H. Beck-Fernández, y RV Pinto, «Un modelo predictivo de fracaso/éxito académico basado en indicadores de admisión, en estudiantes de una universidad estatal del norte de Chile», *Interciencia* , vol. 44, núm. º 1, pp. 23-29, 2019.

[13] BR Chacha, WL López y MB Constante, «Modelo predictivo de deserción estudiantil basado en regresión logística», *ESPOCH Congr. Ecuadorian J. STEAM*, vol. 3, nov. 2023, doi: 10.18502/espoch.v3i1.14477.

[14] P. Román López, MJ Rodríguez Arrastia y C. Ropero Padilla, Eds., *Metodología de la investigación: De lector a divulgador* en Textos didácticos, núm. 83. Almería: Prensa Universidad de Almería, 2021.

[15] MT Ruales, «Predicción de la Deserción Estudiantil en la Universidad Tecnológica de Pereira mediante la Implementación de Modelos de Aprendizaje Automático».

[16] JP Estrada Arana, «Diseño de un modelo matemático utilizando técnicas de análisis multivariado para

estimar tasas de deserción estudiantil en el Instituto Tecnológico Superior Cotacachi.», Tesis de Maestría, Riobamba: Universidad Nacional de Chimborazo, 2025. Consultado: 19 de septiembre de 2025. [En línea]. Disponible en: http://dspace.unach.edu.ec/handle/51000/14558

[17] LF Castro Rojas, E. Espitia Peña y E. Romero Cuero, "Análisis de características que influyen en la deserción estudiantil en el contexto de una universidad latinoamericana", *Rev. EIA*, vol. 20, núm.[40], pág. 2, 2023.

[18] A. Postigo Palacios, *Seguridad Informática*. Madrid: Paraninfo, 2020.

[19] SR Sihare, "Análisis de la deserción estudiantil en la educación superior y la retención mediante inteligencia artificial y aprendizaje automático", *SN Comput. Sci.* , vol. 5, núm. º 2, pág. 202, enero de 2024, doi: 10.1007/s42979-023-02458-w.

[20] JP Uribe Mostacero, «Mejora del indicador de retención en una universidad privada a partir de la clasificación de estudiantes utilizando un modelo predictivo».

[21] Q. Sun, S. Zhou, R. Chen, G. Feng, S.-Y. Hou y B. Zeng, "De la computación a la mecánica cuántica: educación en computación cuántica accesible y práctica para estudiantes de secundaria", 26 de marzo de 2024, *arXiv*: arXiv:2403.17485. doi:10.48550/arXiv.2403.17485.

[22] J. Andrade Salazar y R. Pérez, «Epistemología en educación: desafíos y apuestas contemporáneas», *Rev. Divers. Científica*, vol. 1, págs. 1-21, julio de 2024, doi: 10.36314/diversidad.v4i1.102.

[23] S. Espinal, IM Cruz Pichardo y K. Manzur Herrá, "Liderazgo docente y su relación con la retención estudiantil: un estudio de caso", *Rev. Lasallista Investig*, vol. 20, no.[2], pp. 156–169, 2023.

[24] MZ Buritica, "Bienestar y Productividad: Legado del Enfoque Humanista de la Gestión", noviembre de 2020, consultado el 19 de septiembre de 2025. [En línea]. Disponible en: https://repositorio.ucn.edu.co/entities/publication/f5ebbaa6-ffa2-4623-a276-423f3b643c9f/repositorio.ucn.edu.co

[25] VL Miguéis, A. Pereira, J. Pereira y G. Figueira, "Reducción del desperdicio de pescado fresco garantizando la disponibilidad: pronóstico de la demanda mediante datos censurados y aprendizaje automático", *J. Clean. Prod.* , vol. 359, p. 131852, jul. 2022, doi: 10.1016/j.jclepro.2022.131852.

[26] JC Auza-Santiváñez, AA Quispe Cornejo, JP Hayes Dorado y B. Díaz Pérez, «La educación científica desde el enfoque de la innovación, la ciencia y la tecnología», *Salud Cienc. Tecnología.* , vol. 2, pág. 64, julio. 2022, doi: 10.56294/saludcyt202264.

[27] N. Henríquez Cabezas y D. Vargas Escobar, «Modelos predictivos de rendimiento académico y deserción escolar entre estudiantes de primer año de una universidad pública chilena», *Rev. Estud. Exp. En Educ.* , vol. 21, núm. º 45, pp. 299-316, abr. 2022, doi: 10.21703/0718-5162.v21.n45.2022.015.

[28] D. Casanova Cruz, C. Miranda Díaz y AM Yáñez Corvalán, «Sistema de alerta temprana: Centinela, una experiencia para la retención estudiantil en la Universidad Católica de la Santísima Concepción», *Calid. En Educa.* , No. [55] , diciembre de 2021, doi: 10.31619/caledu.n55.1056.

[29] M. Elbert, M. Cárdenas, D. Zumárraga y B. Mendoza, «Estrategias para evitar el abandono universitario», *RECIAMUC* , vol. 7, págs. 273-280, abril de 2023, doi: 10.26820/reciamuc/7.(2).abril.2023.273-280.

[30] SP Kar, AK Das, R. Chatterjee y JK Mandal, "Evaluación de los parámetros de aprendizaje para la adaptabilidad de los estudiantes en la educación en línea mediante aprendizaje automático e IA explicable", *Educ. Inf. Technol.* , vol. 29, n.º [6] , págs. 7553-7568, abril de 2024, doi: 10.1007/s10639-023-12111-x.