



UNIVERSIDAD TÉCNICA DE COTOPAXI

DIRECCIÓN DE POSGRADO

MAESTRÍA EN CIENCIA DE DATOS

MODALIDAD: PROYECTO DE INVESTIGACIÓN

Título:

ANÁLISIS PREDICTIVO DE LA DESERCIÓN ESTUDIANTIL DE LA UNIVERSIDAD TÉCNICA DE COTOPAXI, UTILIZANDO TÉCNICAS DE MACHINE LEARNING MEDIANTE APRENDIZAJE SUPERVISADO, PARA LA TOMA DE DECISIONES OPORTUNAS

Trabajo de titulación previo a la obtención del título de Magister en Ciencia de Datos

Autor:

Ing. Vera Machuca Carlos Alberto

Tutor:

MSc. Verónica del Consuelo Tapia Cerda

LATACUNGA – ECUADOR

2025

APROBACIÓN DEL TUTOR

En mi calidad de Tutora del Trabajo de Titulación “Análisis predictivo de la deserción estudiantil de la Universidad Técnica de Cotopaxi, utilizando técnicas de Machine Learning mediante aprendizaje supervisado, para la toma de decisiones oportunas “, presentado por Vera Machuca Carlos Alberto, para optar por el título Magíster en Ciencias de Datos.

CERTIFICO

Que dicho Trabajo de Titulación ha sido revisado en todas sus partes y se considera que reúne los requisitos y méritos suficientes para ser sometido a la presentación para la valoración por parte del Tribunal de Lectores que se designe y su exposición y defensa pública.

Latacunga, abril del 2025



MS.c. Verónica del Consuelo Tapia Cerda

CC: 0502053697

APROBACIÓN TRIBUNAL

El Trabajo de Titulación: Análisis predictivo de la deserción estudiantil de la Universidad Técnica de Cotopaxi, utilizando técnicas de Machine Learning mediante aprendizaje supervisado, para la toma de decisiones oportunas, ha sido revisado, aprobado y autorizada su impresión y empastado, previo a la obtención del título de Magíster en Ciencias de Datos. El trabajo reúne los requisitos de fondo y forma para que el estudiante pueda presentarse a la exposición y defensa.

Latacunga, abril del 2025



PHD. José Augusto Cadena Moreano
C.C.: 0501552798
Presidente del tribunal



Mg. Edison Patricio Bedón Salazar
C.C.: 0502253271
Lector 2



Mg. Luis René Quisaguano Collaguazo
C.C.: 1721895181
Lector 3

DEDICATORIA

A Dios, por darme la fortaleza, la perseverancia y la sabiduría para recorrer este camino, a mi esposa, mi mayor inspiración y mi más firme apoyo, gracias por estar siempre a mi lado, por creer en mis sueños como si fueran tuyos y por brindarme tu luz en este camino, esta meta alcanzada también es tuya. A mis padres, por inculcarme los valores del esfuerzo, la perseverancia y el compromiso gracias a cada uno de ustedes, por ser mi fortaleza y mi mayor inspiración.

Vera Machuca Carlos Alberto

AGRADECIMIENTO

Quiero expresar mi más sincero agradecimiento a la Técnica de Cotopaxi por brindarme la oportunidad de realizar mis estudios de posgrado y por proporcionar un ambiente académico de excelencia que ha sido fundamental en mi desarrollo profesional. A la Dirección de Posgrados, mi gratitud por su constante apoyo y guía durante todo el proceso académico. A mi tutora, MSc. Verónica del Consuelo Tapia Cerda, mi más profundo agradecimiento por su dedicación, paciencia y sabiduría. Sus consejos, críticas constructivas y acompañamiento han sido esenciales para la culminación de este trabajo.

Vera Machuca Carlos Alberto

RESPONSABILIDAD DE AUTORÍA

Quien suscribe, declara que asume la autoría de los contenidos y los resultados
Obtenidos en el presente Trabajo de Titulación.

Latacunga, abril del 2025



Ing. Vera Machuca Carlos Alberto
120548683-8

RENUNCIA DE DERECHOS

Quien suscribe, cede los derechos de autoría intelectual total y/o parcial del presente Trabajo de Titulación a la Universidad Técnica de Cotopaxi.

Latacunga, abril del 2025



Ing. Vera Machuca Carlos Alberto
120548683-8

AVAL DEL PRESIDENTE

Quien suscribe, declara que el presente Trabajo de Titulación: Análisis predictivo de la deserción estudiantil de la Universidad Técnica de Cotopaxi, utilizando técnicas de Machine Learning mediante aprendizaje supervisado, para la toma de decisiones oportunas, contiene las correcciones a las observaciones realizadas por los miembros del Tribunal en la predefensa.

Latacunga, abril del 2025



.....

PhD. José Augusto Cadena Moreano
C.C.: 0501552798

UNIVERSIDAD TÉCNICA DE COTOPAXI

DIRECCIÓN DE POSGRADO

MAESTRÍA EN CIENCIAS DE DATOS

Título: Análisis predictivo de la deserción estudiantil de la Universidad Técnica de Cotopaxi, utilizando técnicas de Machine Learning mediante aprendizaje supervisado, para la toma de decisiones oportunas.

Autor: Ing. Vera Machuca Carlos Alberto

Tutor: MS.c. Verónica del Consuelo Tapia Cerda

RESUMEN

La presente investigación tiene como objetivo desarrollar un modelo predictivo utilizando técnicas de Machine Learning mediante aprendizaje supervisado para identificar las variables que tienen mayor influencia en la deserción estudiantil en la Universidad Técnica de Cotopaxi (UTC) durante el periodo académico octubre 2024 – febrero 2025. La hipótesis propuesta busca alcanzar una confiabilidad superior al 90% en la predicción, lo que se logra con un 93.06% de precisión mediante el uso de algoritmos como Random Forest y Máquinas de Soporte Vectorial (SVM). Estos resultados proponen estrategias de intervención temprana basadas en datos para mejorar las tasas de retención estudiantil. El estudio se basa en datos académicos, socioeconómicos y demográficos recolectados a través de una encuesta web implementada con Django y SQLite, dirigida a estudiantes de los primeros ciclos de la UTC. Los modelos predictivos fueron entrenados y evaluados utilizando herramientas como Google Colab, Python y bibliotecas como Scikit-Learn, Pandas y Matplotlib, validando su desempeño con métricas como la matriz de confusión y la división de datos en conjuntos de entrenamiento y prueba. Los hallazgos obtenidos destacan que los factores socioeconómicos son los que más influyen en la deserción estudiantil. Además, el enfoque metodológico ágil adoptado en el desarrollo del modelo predictivo convierten la presente propuesta en una referencia significativa para futuras investigaciones relacionadas con el uso de inteligencia artificial (IA) en el ámbito educativo. Finalmente, este trabajo contribuye en la gestión universitaria de la UTC a través de la reducción de la deserción a más de fortalecer el reconocimiento institucional al ofrecer soluciones basadas en información.

PALABRAS CLAVE: Machine Learning, Random Forest, máquinas de soporte vectorial (SVM), análisis de datos, Scikit-Learn.

UNIVERSIDAD TÉCNICA DE COTOPAXI

DIRECCIÓN DE POSGRADO

MAESTRÍA EN CIENCIA DE DATOS

Title: Predictive analysis of student desertion at the Technical University of Cotopaxi by using Machine Learning techniques through supervised learning for timely decision-making.

Author: Ing. Vera Machuca Carlos Alberto

Tutor: MS.c. Verónica Consuelo Tapia Cerda


ABSTRACT

This research aims to develop a predictive model using supervised machine learning techniques to identify the variables that have the greatest influence on student dropout rates at the Technical University of Cotopaxi (UTC) during the academic period from October 2024 to February 2025. The proposed hypothesis seeks to achieve a reliability greater than 90% in prediction, which is achieved with 93.06% accuracy through the use of algorithms such as Random Forest and Support Vector Machines (SVM). These results propose data-driven early intervention strategies to improve student retention rates. The study is based on academic, socioeconomic, and demographic data collected through a web survey implemented with Django and SQLite, aimed at students in the first cycles of the UTC. The predictive models were trained and evaluated using tools such as Google Colab, Python, and libraries such as Scikit-Learn, Pandas, and Matplotlib, validating their performance with metrics such as the confusion matrix and data splitting into training and test sets. The findings highlight that socioeconomic factors are the most influential factors in student dropout rates. Furthermore, the agile methodological approach adopted in developing the predictive model makes this proposal a significant reference for future research related to the use of artificial intelligence (AI) in education. Finally, this work contributes to the university management of the UTC by reducing dropout rates and strengthening institutional recognition by offering data-based solutions.

KEYWORDS: Machine Learning, Random Forest, Support Vector Machines (SVM), data analysis, Scikit-Learn.

AVAL DE TRADUCCIÓN

José Fernando Toaquiza Chancusig con cédula de identidad número: 0502229677
Licenciado en: Ciencias de la Educación especialización Inglés con número de registro de la SENESCYT:1020 – 07 - 747854; **CERTIFICO** haber revisado y aprobado la traducción al idioma Inglés del resumen del trabajo de investigación con el título: **“ANÁLISIS PREDICTIVO DE LA DESERCIÓN ESTUDIANTIL DE LA UNIVERSIDAD TÉCNICA DE COTOPAXI, UTILIZANDO TÉCNICAS DE MACHINE LEARNING MEDIANTE APRENDIZAJE SUPERVISADO, PARA LA TOMA DE DECISIONES OPORTUNAS.”** de: Vera Machuca Carlos Alberto aspirante a Magíster en **Ciencia de Datos.**



Mg. José Fernando Toaquiza Chancusig
CI: 0502229677

La Maná, abril del 2025

ÍNDICE DE CONTENIDOS

PORTADA.....	i
APROBACIÓN DEL TUTOR.....	ii
APROBACIÓN TRIBUNAL	iii
DEDICATORIA	iv
AGRADECIMIENTO	v
RESPONSABILIDAD DE AUTORÍA	vi
RENUNCIA DE DERECHOS.....	vii
AVAL DEL PRESIDENTE.....	viii
RESUMEN.....	ix
ABSTRACT.....	x
AVAL DE TRADUCCIÓN	xi
ÍNDICE DE CONTENIDOS	xii
ÍNDICE DE TABLAS	xvi
ÍNDICE DE GRÁFICOS	xvii
INTRODUCCIÓN	2
Justificación.....	3
Planteamiento del Problema.....	4
Objetivos	5
Objetivo General	5
Objetivos Específicos.....	5
CAPITULO I. FUNDAMENTACIÓN TEÓRICA.....	6
1.1. Análisis Predictivo	6
1.1.1. Definición y Propósito.....	6
1.1.2. Aplicaciones en Diferentes Áreas	6
1.1.3. Importancia y Beneficios del Análisis Predictivo en la Educación.....	6
1.2. Deserción Estudiantil	7
1.2.1. Definición.....	7
1.2.2. Causas y Factores	8
1.2.2.2. Factores socioeconómicos.....	9
1.2.2.3. Factores psicológicos	9
1.2.2.4. Factores institucionales	10
1.2.3. Impacto de la deserción estudiantil en las universidades.....	11

1.3.	Modelos teóricos sobre la deserción estudiantil.....	11
1.3.1.	Modelos sociopsicológicos	11
1.3.2.	Modelos integrales de deserción	11
1.4.	Técnicas de Machine Learning.....	12
1.4.1.	Introducción	12
1.4.2.	Tipos de Aprendizaje en Machine Learning	12
1.5.	Algoritmos de aprendizaje supervisado más utilizados.....	14
1.5.1.	Regresión Logística.....	14
1.5.2.	Máquinas de Soporte Vectorial (SVM)	14
1.5.3.	Árboles de Decisión y Bosques Aleatorios	15
1.5.4.	Redes Neuronales.....	15
1.6.	Evaluación de modelos predictivos	16
1.6.1.	Métodos de validación cruzada.....	16
1.6.2.	Matriz de Confusión.....	17
1.7.	Aplicación de Machine Learning en el Ámbito Educativo	17
1.7.1.	Revisión de Estudios Previos	17
1.7.2.	Modelos Predictivos Basados en Datos Académicos.....	18
1.7.3.	Modelos Predictivos Considerando Factores Socioeconómicos y Demográficos	19
1.7.4.	Desafíos y Limitaciones del Análisis Predictivo en el Contexto Educativo	19
1.8.	Análisis Predictivo para la Toma De Decisiones	19
1.8.1.	Definición de Toma de Decisiones Basada en Datos.....	19
1.8.2.	Análisis Predictivo en la Toma de Decisiones Educativas	20
1.9.	Desarrollo de Sistemas de Análisis Predictivo.....	21
1.9.1.	Metodologías Ágiles	21
1.9.2.	Metodología Kanban.....	21
1.9.3.	Metodología XP	21
1.9.4.	Metodología Scrum.....	22
1.9.5.	Herramientas de desarrollo	22
1.10.	Hipótesis	25
CAPITULO II. MATERIALES Y MÉTODOS.....		26
1.11.	Caracterización de la Institución	26
1.12.	Diseño de la Investigación	28

1.12.1.	Tipo de Investigación	28
1.13.	Enfoque de Investigación	28
1.14.	Método de Investigación	29
1.15.	Población y Muestra	29
1.16.	Técnicas de Investigación	30
1.16.1.	Encuesta	30
1.17.	Instrumento de Investigación	31
1.17.1.	Cuestionario	31
1.17.2.	Diseño del Instrumento	31
1.18.	Metodología de Desarrollo Basada en Prácticas Ágiles.....	32
1.19.	Roles del Equipo	33
1.20.	Etapas del Proyecto	33
1.21.	Herramientas de apoyo al desarrollo	34
1.22.	Proceso de Negocio	36
1.23.	Equipos Utilizados	37
1.24.	Arquitectura del Sistema	38
1.25.	Formato de Historias de Usuario.....	39
1.26.	Formato para la Pila del Producto	40
1.27.	Formato para ejecutar pruebas del sistema.....	41
CAPITULO III. PROPUESTA		42
3.1.	Etapa de Planificación	42
3.1.1.	Historias de usuario.....	42
3.1.2.	Pila del Producto	44
3.1.3.	Pila del Sprint.....	46
3.2.	Etapa de Ejecución	47
3.2.1.	Diseños de Interfaz Gráfica de Usuario	47
3.2.2.	Modelo de Base de Datos.....	51
3.2.3.	Análisis de los Datos.....	53
3.2.4.	Modelado Predictivo	66
3.2.5.	Análisis Exploratorios Previos.....	66
3.2.6.	Procesamiento del DataSet.....	69
3.2.7.	Definición de la Variable Objetivo	69
3.2.8.	Conversión de Variables Categóricas en Numéricas	70
3.2.9.	Selección de Técnicas de Machine Learning Supervisado	73

3.2.10.	Entrenamiento del Modelo con RandomForest.....	74
3.2.11.	Predicción y Evaluación del Modelo <i>RandomForest</i>	75
3.2.12.	Matriz de Confusión del Modelo <i>RandomForest</i>	76
3.2.13.	Eficiencia del Modelo <i>RandomForest</i>	77
3.2.14.	Entrenamiento del Modelo con Máquina de Soporte de Vectores..	77
3.2.15.	Influencia de las Variables para el Modelo Predictivo	80
3.2.16.	Comprobación de la Hipótesis	81
3.2.17.	Importancia de Dividir los Datos para Entrenamiento y Pruebas ...	81
3.2.18.	Estrategias de Mejora con Base a los Hallazgos Obtenidos	82
	CONCLUSIONES GENERALES	84
	RECOMENDACIONES	85
	BIBLIOGRAFÍA	86
	ANEXOS	91

ÍNDICE DE TABLAS

Tabla 1: Importancia del análisis predictivo en Educación	7
Tabla 2: Beneficios del análisis predictivo	7
Tabla 3: Oferta Académica de la UTC por campus y extensiones	26
Tabla 4: Razones para adaptar las prácticas ágiles de Scrum en el proyecto	32
Tabla 5: Herramientas de desarrollo para el análisis predictivo de deserción	35
Tabla 6: Características del equipo utilizado para el desarrollo	38
Tabla 7: Formato para definir las historias de usuario	40
Tabla 8: Formato para definir la pila del producto	40
Tabla 9: Formato para ejecutar las pruebas del sistema.....	41
Tabla 10: Historia de Usuario No. HU001.....	43
Tabla 11: Historia de Usuario No. HU002.....	43
Tabla 12: Historia de Usuario No. HU003.....	44
Tabla 13: Pila de producto del sistema de análisis predictivo	45
Tabla 14: Planificación de la Pila del Sprint.....	46
Tabla 15: Análisis e interpretación de los datos obtenidos	54
Tabla 16: Clasificación binaria para la variable objetivo.....	70
Tabla 17: Nomenclatura usada en la matriz de correlación de variables.....	72
Tabla 18: Comparativa de Técnicas de Machine Learning.....	73
Tabla 19: Configuración del modelo de análisis predictivo.....	74
Tabla 20: Análisis del modelo predictivo SVM.....	78
Tabla 21: Estrategias de mejora para evitar la deserción estudiantil.	82

ÍNDICE DE GRÁFICOS

Gráfico 1: Matriz de confusión	17
Gráfico 2: Total de carreras por campus o extensión de la UTC	27
Gráfico 3: Estructura del conjunto de datos considerados como población	29
Gráfico 4: Etapas y artefactos de software.....	34
Gráfico 5: Modelo de negocio del sistema.....	37
Gráfico 6: Arquitectura del sistema de análisis predictivo	39
Gráfico 7: Interfaz de Inicio del Sistema	47
Gráfico 8: Interfaz de validación de estudiantes	48
Gráfico 9: Interfaz del formulario de deserción estudiantil	49
Gráfico 10: Interfaz de autenticación.....	50
Gráfico 11: Interfaz de gestión de estudiantes	50
Gráfico 12: Menú de opciones del sistema predictivo.....	51
Gráfico 13: Modelo de datos del sistema de análisis predictivo.....	52
Gráfico 14: Modelos diseñados en Django para el formulario de deserción estudiantil.....	53
Gráfico 15: Resumen de datos recopilados mediante el cuestionario web	54
Gráfico 16: Código de análisis para determinar sesgo en la deserción estudiantil	67
Gráfico 17: Distribución de respuestas sobre la probabilidad de deserción	68
Gráfico 18: Manejo de valores nulos y/o atípicos.....	69
Gráfico 19: Estandarización de datos.....	69
Gráfico 20: Cabecera del dataframe expandido	69
Gráfico 21: Conversión de variables categóricas a numéricas.....	70
Gráfico 22: Mapeo Manual de categorías a valores binarios.....	70
Gráfico 23: Matriz de correlación de variables.....	71
Gráfico 24: Entrenamiento del modelo de análisis de la deserción estudiantil	75
Gráfico 25: Predicción y evaluación del modelo Random Forest.....	76
Gráfico 26: Matriz de confusión del modelo predictivo RandomForest	76
Gráfico 27: Codificación del modelado predictivo con SVM	77
Gráfico 28: Métricas de evaluación del modelo SVM.....	78
Gráfico 29: Matriz de confusión del modelo predictivo SVM	79

ÍNDICE DE ANEXOS

Anexo A: Cuestionario sobre la Deserción Estudiantil	92
Anexo B: Codificación del Análisis Predictivo	97
Anexo C: Código de la Matriz de Confusión.....	100
Anexo D: Hoja de Vida del Investigador.....	102

INFORMACIÓN GENERAL

Título del Trabajo de Titulación:

Análisis predictivo de la deserción estudiantil de la Universidad Técnica de Cotopaxi, utilizando técnicas de Machine Learning mediante aprendizaje supervisado, para la toma de decisiones oportunas.

Línea de Investigación:

Tecnología de la información y las comunicaciones, robótica, automatización y optimización de sistemas.

- Los proyectos de investigación que se enmarquen en esta línea tendrán como objetivos desarrollar tecnologías y herramientas informáticas de apoyo a la incorporación de planes y programas de desarrollo, utilizando las TICs en la optimización y organización de procesos para la creación de sistemas informáticos y de control, así como la introducción de la inteligencia de negocios, la inteligencia artificial y la física - matemática aplicada a la modelación y simulación de sistemas.

Sub-Línea:

Inteligencia artificial e inteligencia de negocios

Grupo de Investigación:

Desarrollo Tecnológico para Sistemas de Información Automatizados

Introducción

La presente investigación tiene por objetivo desarrollar un modelo predictivo utilizando técnicas de Machine Learning mediante aprendizaje supervisado para determinar las variables que influyen en la deserción estudiantil en la Universidad Técnica de Cotopaxi (UTC), siendo un aporte innovador en el ámbito académico que busca mejorar las tasas de retención. Surge la necesidad de obtener datos reales del origen de la deserción, para comprender y abordar los procesos académicos de los estudiantes y el cumplimiento de los objetivos institucionales.

Se utiliza modelos de Machine Learning como Random Forest y Máquinas de Soporte Vectorial (SVM) para predecir la deserción estudiantil con un 93.06% de confiabilidad, permitiendo intervenciones tempranas basadas en datos, la combinación de datos académicos, demográficos y socioeconómicos permitieron desarrollar modelos que identifican patrones de deserción y brindan información valiosa sobre los factores subyacentes que contribuyen a este fenómeno. En el Capítulo I se revisa literatura científica sobre la deserción estudiantil y técnicas de Machine Learning, en el Capítulo II se detalla la metodología de recolección y análisis de datos académicos, socioeconómicos y demográficos, mientras que en el Capítulo III se presenta la implementación del modelo predictivo, aplicando técnicas supervisadas de Machine Learning sobre un conjunto de datos facilitados por estudiantes de los primeros ciclos de las carreras de la UTC a través de una encuesta web sistematizada con Django y SQLite.

Se entrenaron y evaluaron modelos con diferentes algoritmos, validando su desempeño mediante métricas como la matriz de confusión y la separación en datos de entrenamiento y pruebas. El desarrollo se realizó con el entorno Google Colab, el lenguaje de programación Python, las bibliotecas Scikit-Learn, Pandas, Matplotlib y NumPy. Este estudio resulta clave para aportar en la gestión universitaria, reducir la deserción y mejorar el reconocimiento institucional, se concluye que los factores socioeconómicos son los que más influyen en la deserción estudiantil y que el enfoque metodológico ágil adoptado convierten la presente propuesta en una referencia para futuras investigaciones en educación superior.

Justificación

La deserción estudiantil es un fenómeno ocasionado por diferentes factores académicos, socioeconómicos y/o demográficos que afecta a los estudiantes universitarios, pero influye también en las instituciones de educación superior y la sociedad en general. En el caso de la Universidad Técnica de Cotopaxi, la visión profunda y el enfoque en este desafío son cruciales para mantener sus objetivos escolares y ayudar en el crecimiento a largo plazo de la región. Así mismo, este fenómeno limita el acceso de las personas a oportunidades de aprendizaje y superación académica que a futuro minimizará sus ofertas de trabajo y desarrollo personal.

Por el contrario, el empleo de tácticas derivadas del análisis predictivo contribuirá en mejorar la retención de los estudiantes al brindar información clave para que la universidad pueda tomar decisiones eficaces de forma anticipada y preventiva contra el abandono escolar. Es importante señalar que mantener a los estudiantes dentro del sistema educativo superior es fundamental para su éxito individual lo que también repercute en la calidad y reputación de la institución. La anticipación de la deserción favorece la asignación eficiente de los recursos institucionales, al evidenciar a los alumnos en peligro de deserción.

De esta manera, la UTC puede canalizar una ayuda particularizada e individualizada, evitando la falta de recurso en alumnos de bajos niveles económicos de tal modo que concentren su atención en tener un rendimiento académico adecuado sin preocupaciones ajenas a la actividad universitaria. Es decir, el análisis predictivo facilita la identificación de factores específicos que contribuyen a la deserción de cada estudiante, esto posibilita el diseño de estrategias personalizadas que aborden las necesidades individuales, creando un entorno académico más favorable y de apoyo.

La implementación de técnicas avanzadas como Machine Learning muestra el compromiso de la Universidad Técnica de Cotopaxi con la innovación y fortalece su reputación como institución proactiva y orientada a la mejora continua considerando que este estudio aumentará el conocimiento en el campo de la

educación superior. Los resultados y los métodos creados pueden servir como guías para otras instituciones que enfrentan dificultades similares.

Al abordar la deserción estudiantil, la institución afirma su dedicación a la igualdad en la educación, se esfuerza por disminuir las discrepancias en la admisión y la permanencia de los estudiantes de diversos grupos demográficos a más de fomentar el bienestar, el desempeño institucional y el progreso socioeconómico regional de los estudiantes.

Planteamiento del Problema

La deserción estudiantil es un problema significativo al que deben enfrentarse todas las instituciones educativas a nivel global, de acuerdo con diferentes estudios, la tasa de deserción afecta a la calidad educativa, al desarrollo social y económico de las regiones y países, tal es el caso de Estados Unidos que registra una tasa de deserción del 52%, esto hace que la deserción estudiantil se considere un inconveniente estructural del sistema educativo en general considerando que las estadísticas muestran cómo este fenómeno afecta la permanencia y la graduación de los estudiantes [1].

Las universidades del Ecuador enfrentan una problemática persistente de abandono estudiantil en la educación superior, al respecto [2] propone que la deserción puede ser de manera voluntaria o no voluntaria, siendo la falta de orientación vocacional uno de los factores más influyentes. A pesar de los esfuerzos por mejorar el acceso a la educación, las tasas de deserción en las universidades siguen siendo altas, especialmente en instituciones públicas localizadas cerca a zonas rurales. Otros factores que afectan la continuidad académica son la falta de apoyo financiero, la insuficiencia de programas de retención estudiantil, la desigualdad en la calidad educativa y las dificultades socioeconómicas [3].

La Universidad Técnica de Cotopaxi no se aleja de esta realidad, dado a que el abandono escolar es una problemática significativa que ocasiona la disminución del acceso a la educación superior y mantiene las desigualdades socioeconómicas.

Existen factores influyentes en la deserción estudiantil que pueden agruparse en tres amplias categorías: personales, escolares y socio familiares, así mismo la falta de desarrollo en capacidades como atención, confianza, seguridad y autoestima afectan las condiciones de desempeño académico en la UTC [4]. Dado a que en esta universidad se tiene un número considerable de estudiantes provenientes de zonas rurales que a su vez enfrentan diversas barreras socioeconómicas, se requieren estudios que permitan proponer políticas universitarias dinámicas que se adapten a las realidades cambiantes de los alumnos para contribuir en su permanencia académica.

Objetivos

Objetivo General

Desarrollar un modelo predictivo utilizando técnicas de Machine Learning mediante aprendizaje supervisado para determinar las variables que influyen en la deserción estudiantil en la Universidad Técnica de Cotopaxi durante el periodo académico octubre 2024 – marzo 2025.

Objetivos Específicos

- Consultar literatura científica mediante fuentes bibliográficas primarias para recopilar información relacionada con la aplicación de Machine Learning en ámbitos universitarios.
- Realizar un análisis preliminar de los datos académicos, demográficos y socioeconómicos de los estudiantes de la UTC a través de un sistema en ambiente web para identificar las variables más relevantes que influyen en la deserción estudiantil.
- Implementar modelos de Machine Learning supervisado utilizando los datos recolectados y analizar estadísticamente las variables recopiladas para identificar patrones y relaciones significativas con la deserción estudiantil.
- Evaluar el desempeño de los modelos predictivos desarrollados mediante un conjunto de datos de entrenamiento y prueba para determinar el algoritmo con mayor precisión.

CAPITULO I. FUNDAMENTACIÓN TEÓRICA

1.1. Análisis Predictivo

1.1.1. Definición y Propósito

El análisis predictivo comprende un conjunto de técnicas que utilizan datos históricos y contemporáneos para predecir comportamientos futuros. Se basa en métodos matemáticos, estadísticos y de aprendizaje automático (Machine Learning) para modelar patrones y tendencias, generando así predicciones cuantitativas o cualitativas.

En el ámbito educativo, el análisis predictivo se aplica con el propósito principal de revelar el rendimiento académico de los estudiantes, permitiendo anticipar posibles dificultades o éxitos. Este propósito general se traduce en la capacidad de predecir resultados futuros, como el rendimiento en exámenes, la probabilidad de abandono escolar o la necesidad de apoyo personalizado [5].

1.1.2. Aplicaciones en Diferentes Áreas

En [6] detalla la utilización el uso extensivo de análisis predictivo dentro del contexto más amplio de (Big Data Analytics), en ciudades inteligentes (por ejemplo, gestión del tráfico, predicción de delitos, optimización de recursos).

En [5] se explica la aplicación y viabilidad de realizar este tipo de análisis dentro de la educación. Específicamente, para predecir el rendimiento de los estudiantes en instituciones de educación superior utilizando técnicas de minería de datos educativos y análisis de aprendizaje por vídeo.

1.1.3. Importancia y Beneficios del Análisis Predictivo en la Educación

La importancia y beneficios del análisis predictivo en la educación a continuación se exponen en las Tablas 1 y 2 respectivamente:

Tabla 1: Importancia del análisis predictivo en Educación [5]

Importancia	Descripción
Identificar a los estudiantes en riesgo	Predecir qué estudiantes son propensos a fallar permite una intervención temprana y apoyo, mejorando las tasas de éxito de los estudiantes.
Optimizar la asignación de recursos	Los datos permiten a las IES asignar recursos de manera más eficaz, orientando el apoyo hacia los estudiantes que más lo necesitan.
Mejorar la enseñanza y el aprendizaje	Analizar las interacciones de los estudiantes con el contenido de vídeo y otras plataformas de aprendizaje proporciona información valiosa para que los instructores perfeccionen sus estrategias de enseñanza.
Mejorar la toma de decisiones	Proporciona información basada en datos para que los profesores, estudiantes y administradores tomen decisiones más informadas sobre el diseño del currículo, la asignación de recursos y el apoyo a los estudiantes.

Tabla 2: Beneficios del análisis predictivo [5]

Beneficios	Descripción
Apoyo específico	Los recursos pueden centrarse en los usuarios más necesitados, maximizando el impacto de los servicios de apoyo o además de impulsar donde se necesita realizar cambios específicos.
Mejor gestión de los recursos	Una asignación más eficiente de los recursos permite ahorrar costos y mejorar la eficiencia institucional en general.
Experiencia mejorada	Las experiencias de aprendizaje personalizadas, adaptadas a las necesidades individuales del estudiante, resultan en un entorno de aprendizaje más eficaz y de apoyo.

Aunque no se centra directamente en la educación, [6] ilustra ampliamente los beneficios similares en el contexto de la planificación y gestión urbanas, haciendo hincapié en la toma de decisiones proactiva basada en datos sobre enfoques reactivos.

1.2. Deserción Estudiantil

1.2.1. Definición

La deserción estudiantil o escolar es un hecho social y educativo que se observa en varias sociedades, a pesar de que es crucial señalar que la mayoría de los estudios y análisis se han enfocado en alumnos de nivel primario y secundario, existiendo escasa información sobre el abandono escolar en nivel superior, existe una diferencia entre los estudios de abandono escolar en los niveles primario y secundario y la educación superior [7].

Es importante tener en cuenta que es alarmante este hecho social, dado que las situaciones y retos a lo que se enfrentan los estudiantes universitarios, afectados por los elementos de la vida adulta, las obligaciones laborales, económicas y la independencia académica. Es importante cubrir este vacío de investigación para tener maneras más eficaces de las demandas particulares de alumnos de nivel superior y ayudar a incrementar los índices de retención en este nivel educativo que es vital.

1.2.2. Causas y Factores

La deserción estudiantil es un fenómeno influenciado por múltiples factores que afectan la continuidad y éxito académico de los estudiantes. Entre las principales causas se encuentra la poca capacitación en el manejo de herramientas tecnológicas, asimismo, la falta de accesibilidad y las dificultades para desplazarse, especialmente en zonas sin conexión a internet, representan barreras significativas, otros aspectos relevantes incluyen los problemas laborales que dificultan la dedicación al estudio y la falta de orientación pedagógica [8].

La diversidad de factores que inciden en la deserción estudiantil evidencia la complejidad del problema y la importancia de enfrentarlo desde distintas perspectivas. Tomando en cuenta por la falta de capacitación tecnológica, por los obstáculos financieros y las restricciones de acceso que impactan en el desempeño escolar, siendo también en la motivación y en el sentimiento de pertenencia de los alumnos, enfrentar estos problemas de forma proactiva podría generar un cambio notable en las tasas de retención y triunfo académico.

1.2.2.1. Factores académicos

La educación juega un papel fundamental en el desarrollo y el progreso de cualquier sociedad, esto último les permite promover valores culturales, crear conocimiento, que es la base del progreso y la convivencia entre los ciudadanos; también cooperan en el avance de la ciencia y la tecnología que permite desarrollar habilidades para interactuar de manera apropiada en la sociedad, la educación es la fuente del conocimiento cultural, espiritual, intelectual, moral y de todos aquellos componentes que particularizan a los seres pensantes y transformadores de los contextos [9].

Indiscutiblemente, la educación es un elemento crucial para el crecimiento individual y grupal, dado que va más allá de la mera impartición de saberes, al impulsar valores, promover la coexistencia ciudadana y estimular el avance científico y tecnológico. Su influencia en el fomento de habilidades sociales y capacidades individuales la transforma en un impulsor esencial para el bienestar y el progreso constante de las comunidades.

1.2.2.2. Factores socioeconómicos

La deserción escolar cuenta con distintos factores socioeconómicos los cuales crean elevados costos sociales y privados, algunos factores no son sencillos de estimar, pero entre ellos se mencionan los que proceden de una fuerza de trabajo menos competente y más difícil de calificar, por ejemplo, cuando las personas no tienen base en ciertos niveles de educación no logran aprovechar los beneficios de programas de entrenamiento ofrecidos por el estado o por las empresas, lo que termina generando un problema de especie de analfabetismo [10].

El abandono escolar tiene altos problemas sociales y económicos, como una fuerza laboral menos capacitada, baja productividad y menor crecimiento económico, también aumenta la dependencia de programas sociales y las desigualdades económicas, afectando la movilidad social y la cohesión comunitaria. Además, limita la participación ciudadana debilitando la democracia, esto da a conocer la necesidad de implementar estrategias que promuevan una educación inclusiva y accesible.

1.2.2.3. Factores psicológicos

Dentro de estos factores se presentan las dificultades emocionales y el comportamiento que provocan en cada alumno, estos problemas no solo afectan el aprendizaje sino también la salud mental de los estudiantes, la mayor parte de los alumnos que manifiestan dificultades emocionales y conductuales tienen leves alteraciones en su desarrollo cognitivo, psicomotor o emocional, lo cual les puede llevar ser asignados a condiciones diagnósticas específicas como: retardo mental, síndrome de déficit atencional o trastornos [11].

Los problemas emocionales y de comportamiento en los alumnos demuestran la relevancia de enfocar la educación de forma integral, teniendo en cuenta la dimensión académica, emocional y social. Es alarmante cómo elementos internos y externos, vinculados con el ambiente educativo, pueden restringir su desarrollo integral. Esto resalta la importancia de una capacitación de los profesores más empática y adaptable, con tácticas pedagógicas inclusivas y recursos adecuados para satisfacer la variedad de necesidades. La cooperación entre familias, instituciones educativas y expertos es fundamental para generar un entorno que promueva el bienestar emocional y el triunfo escolar de los alumnos.

1.2.2.4. Factores institucionales

Se denomina un factor institucional a la organización y funcionamiento de la institución educativa, esto en base a módulos de gestión, módulos pedagógicos, la oferta de programas, entre otros. Cuando la institución posee deficiencias al proveer al alumno de elementos básicos para su desarrollo, el mismo puede generar rechazo a la institución, favoreciendo a la posibilidad de desertar. Acorde a la variable institucional se puede diferenciar 5 factores básicos:

- Normativa académica.
- Financiación estudiantil.
- Recursos universitarios.
- Calidad de programa.
- Relaciones interpersonales.

El componente institucional es esencial en la experiencia educativa, dado que fallos en la administración, recursos o programas pueden impactar de manera adversa en la dedicación del alumno, incrementando la probabilidad de abandono escolar. Elementos como las regulaciones académicas, el financiamiento estudiantil, la excelencia de los programas y las relaciones interpersonales deben ser considerados con prioridad para asegurar un ambiente que impulse y respalde el crecimiento integral de los estudiantes.

1.2.3. Impacto de la deserción estudiantil en las universidades

La deserción estudiantil es un tema que aborda a la mayor parte de las instituciones de educación superior de toda Latinoamérica [12]. Este fenómeno puede ser el resultado de varios factores, voluntarios o involuntarios, que tienen consecuencias no sólo para los estudiantes que abandonan, sino también para sus familias, las instituciones y para la sociedad.

El abandono estudiantil en la educación superior es un problema que afecta a los estudiantes como instituciones y a la sociedad en general, su impacto recae a los estudiantes y a sus familias, también afecta a los ingresos, eficiencia y calidad de las universidades. Este proceso está determinado por la interacción de factores personales juega un papel crucial en la decisión de permanecer o abandonar sus estudios, lo que refuerza la importancia de generar entornos educativos que apoyen y motiven al estudiante.

1.3. Modelos teóricos sobre la deserción estudiantil

Existen posturas que plantean diversos modelos: algunos tienden a dar prioridad a un agente u otro (institución/alumno) y otros proponen modelos de interrelación dinámica entre estos [13].

1.3.1. Modelos sociopsicológicos

El abandono de los estudios del adolescente se puede deber a una o diversas causas, siendo las principales las económicas, personales, sociales, familiares, educativas y motivacionales [14].

Este modelo socio psicológico se basa específicamente en las personas que rodean a la persona afectada buscando una solución con ayuda de la familia, es aquel que va mediante el factor psicológico que este puede ir causando a lo largo del tiempo a causa de la atención.

1.3.2. Modelos integrales de deserción

El modelo integral de deserción es aquel que puede llevarse a cabo como el entrenamiento es decir como una manera para encontrar la habilidad social

mediante actividades que dan valor a los estudiantes y a la vez encuentran una fortaleza en los videojuegos utilizándose como una alternativa hacia el estrés.

Las estrategias incluyeron la comprensión de la carrera, las condiciones de la misma, motivación a estudiantes, y la mejora de las condiciones ambientales, en la etapa dos de “Entrenamiento” se encontraban las habilidades sociales, desarrollo de las actividades, y así mismo, buscar espacios de estudio, finalmente, la etapa tres, llamada: “Valoración”, donde se deberían trabajar espacios de encuentro, seguimiento académico de los estudiantes, y fortalecimiento entre relación docente-estudiante, la creación del modelo fue una oportunidad para minimizar los niveles de deserción y monitorear los estudiantes en su proceso académico [15].

1.4. Técnicas de Machine Learning

1.4.1. Introducción

El machine learning o aprendizaje automático es la ciencia que permite que las máquinas o sistemas “aprendan” a partir de los datos. Es decir, en lugar de programar, paso a paso, al igual que ocurre en el enfoque de programación convencional, el área de machine learning está diseñado para desarrollar algoritmos genéricos que pueden extraer patrones de varios tipos de datos [16].

Por consiguiente, el Machine Learning hace referencia a la ciencia que permite a los computadores aprender y actuar como seres humanos, mejorando su capacidad de aprendizaje de forma autónoma. Una aplicación de la tecnología de IA en este sentido es el comercio minorista, donde se está aplicando como innovación en todo el ciclo de productos y servicios.

1.4.2. Tipos de Aprendizaje en Machine Learning

1.4.2.1. Aprendizaje Supervisado

El tipo más simple de Machine Learning es el modelo de aprendizaje supervisado, el cual consiste en aprender un modelo utilizando datos que están primeramente etiquetados, para luego poder asignar una etiqueta a datos no etiquetados. De este modo se puede decir que es supervisado en vista de que el programador establece etiquetas las cuales son las salidas esperadas en relación de cada dato que fue

ingresado, es así que el programador es quien muestra las clases o etiquetas en las que clasifica algo [17].

El aprendizaje supervisado es aquel que aprende y relaciona patrones entre la entrada y la salida, esto se da cuando los datos obtenidos tienen una etiqueta. Los datos etiquetados son un conjunto de datos que contienen características propias las cuales se relacionan mediante la identificación de patrones.

1.4.2.2. Aprendizaje No Supervisado

Es un modelo predictivo entrenado el cual comparte características similares al aprendizaje supervisado, pero la diferencia está en la comprensión que se da en datos no clasificados o etiquetados puesto a que descubre patrones de ejemplos similares entre grupos de datos [18].

En síntesis, el aprendizaje no supervisado es aquel que utiliza algoritmos para agrupar conjuntos de datos o encontrar patrones que se encuentran ocultos sin la necesidad del factor humano.

1.4.2.3. Aprendizaje por refuerzo

El aprendizaje por refuerzo (AR), en inglés “*reinforcement learning*” (RL), es una técnica del campo de la inteligencia artificial que imita la forma en la que los humanos aprenden a realizar tareas mediante recompensas que refuerzan la conducta [19]. Normalmente se identifican tres elementos: un entorno, un agente y un intérprete; el agente, teniendo en consideración el estado del entorno y las recompensas obtenidas previamente, selecciona la acción con la cual estima que va a obtener una mayor recompensa, esta acción produce un cambio en el entorno; el intérprete observa el efecto de la acción en el entorno y proporciona información al agente sobre el nuevo estado y la recompensa a la última acción ejecutada, cerrando así el lazo de control [19].

El aprendizaje por refuerzo, en el campo de la inteligencia artificial es muy ambiguo en vista de que se puede considerar que humaniza bastante el concepto de que debe haber una recompensa para que aprenda, lo mismo que pasa con los bebés al sentir la necesidad de alcanzar algo aprenden a moverse que con el tiempo es gatear y a su vez caminar todo con el fin de conseguir eso que quiere. De igual modo, el

concepto de “premio” no es muy claro pero dependería bastante del contexto que se vaya a utilizar.

1.5. Algoritmos de aprendizaje supervisado más utilizados

1.5.1. Regresión Logística

El análisis de regresión logística tradicional se utiliza ampliamente en problemas de clasificación binaria, pero tiene muchas iteraciones y requiere mucho tiempo para entrenar grandes volúmenes de datos, lo que lo hace ineficiente en esos casos. Un estudio realizado por [20] analiza el modelo matemático de la regresión logística, define la función de error, utiliza el método de descenso de gradiente para calcular los coeficientes de regresión y mejora la función sigmoidea. Así, se reduce el número de iteraciones, manteniendo la precisión y mejorando el efecto de clasificación, además, se aplica a un modelo de evaluación de vehículos para predecir la aceptación de un automóvil por parte de los consumidores [20].

Aquí expande una solución para mejorar la regresión logística tradicional, esto con el fin de solucionar problemas de clasificaciones binarias, se enfoca en reducir el número de iteraciones a solo las necesarias, esto aceleraría el proceso sin reducir la precisión. Mediante la optimización de la función “sigmoidea” y se usan métodos como descenso de gradientes.

1.5.2. Máquinas de Soporte Vectorial (SVM)

Los modelos de soporte vectorial (SVM) son sistemas de aprendizaje automatizado cuyos procesos de entrenamiento están controlados por un agente externo y están propiamente relacionados con problemas de clasificación binaria o multiclase mediante la inducción de un separador lineal ya sea en el espacio original de los ejemplos de entrada, si éstos son separables o cuasi-separables o en un espacio transformado (denominado espacio de características) si los ejemplos no son separables linealmente en el espacio original [20].

Se destaca que las máquinas de soporte vectorial son un método de algoritmo muy poderoso al poder clasificar y aplicar regresión al encontrar la frontera o línea (a lo

que se le puede decir hiperplano) óptima para separar las clases, incluso en casos donde los datos son no lineales esto lo hace mediante el uso de kernels.

1.5.3. Árboles de Decisión y Bosques Aleatorios

Los árboles de decisión proporcionan una gran explicabilidad dado que pueden ser comprendidos por expertos en aplicaciones prácticas al brindar un modelo más cercano al lenguaje humano, a diferencia de los métodos de caja negra como lo son las redes neuronales, las máquinas de soporte vectorial y los bosques aleatorios [21].

El modelo bosque aleatorio (*Random Forest*) se forma de un conjunto de árboles de decisión individuales, utilizando distintas muestras de datos de entrenamiento. Es utilizado para entrenar datos en base a información histórica para generar tendencias y clasificaciones posibles. El modelo se forma a través del conjunto de predicciones creadas de cada árbol individual.

Una de las principales ventajas de bosque aleatorio es la precisión de sus resultados, pues con una base de datos presenta resultados bastante exactos, es un algoritmo que funciona muy bien para analizar big data generando tendencias y clasificaciones de los datos analizados, además de responder bien ante casos de datos perdidos sin alterar gravemente el resultado [22]. Para utilizar el modelo del bosque aleatorio es necesario una base que son los árboles de decisión que proporcionan una gran explicabilidad, lo que da una oportunidad de no utilizar las opciones de caja cerrada lo que vienen siendo las redes neuronales o las máquinas de soporte vectorial.

1.5.4. Redes Neuronales

Son modelos matemáticos que simulan el funcionamiento del sistema nervioso, los cuales están compuestos por un conjunto de unidades las cuales son denominadas neuronas o nodos estos enlazados unos con otros, el primer modelo de red neuronal fue presentado por McCulloch y Pitts basado en un modelo computacional de actividad nerviosa; estos ayudan a obtener un modelo no explícito que relaciona un

conjunto de variables de salida con un conjunto de variables de entrada, de esta manera, permiten predecir el valor de salida a partir de valores de entrada del modelo [23].

Las redes neuronales artificiales son una solución matemática al intentar reproducir el sistema nervioso, tiene neuronas que representan un escalón o umbral prefijado; permiten obtener un modelo no explícito que relaciona un conjunto de variables de salida con un conjunto de variables de entrada, lo que mediante los valores de entrada permite predecir cuales son los valores de salida, lo que por medio de la fase de entrenamiento podemos adaptarla al nivel de datos que va a procesar.

1.6. Evaluación de modelos predictivos

1.6.1. Métodos de validación cruzada

En [24] se plantea que la validación cruzada divide el conjunto de datos tanto para entrenamiento, como para pruebas, el entrenamiento de modelos requiere de un gran conjunto de datos para mejorar el desempeño de este, es por eso que resulta, en cierta medida, un desafío conseguir los datos para el entrenamiento, ya sea por su cantidad o calidad. Es allí en donde la validación cruzada es útil, dado a que del conjunto de datos que se tiene puede dividirse según se requiera para tener un conjunto de entrenamiento y otro para las pruebas, así se logra maximizar el uso de los datos disponibles y se garantiza que el modelo sea más robusto.

El proceso de validación cruzada empieza ajustando los valores de los hiperparámetros que se quieren evaluar, luego dividiendo el conjunto de entrenamiento en subconjuntos del mismo tamaño llamados folds en kfolds; por lo general son cinco folds que tienen el 20% de los datos; se entrenan 5 modelos tomando un fold como conjunto de validación de forma iterativa desde el primer fold al último fold, después se obtiene el promedio de los 5 valores de la métrica de interés [24].

La validación cruzada solventa el problema de los datos al entrenar un modelo, gracias a que con un solo conjunto es posible entrenar 5 modelos y obtener la métrica que sea de interés, a su vez se puede validar los modelos con el mismo

conjunto, por lo que los datos se aprovechan de forma máxima para construir un modelo robusto.

1.6.2. Matriz de Confusión

Al evaluar un modelo se debe tomar en cuenta la matriz de confusión, según [25] la idea general es contar el número de veces que las instancias de la clase A son clasificadas como clase B; al entrenar un modelo no se tiene éxito en el 100% de los casos, por lo que es importante evaluar el desempeño del mismo para poder saber con exactitud el porcentaje de casos de éxito y fallos que pueda tener el modelo analizando los: TN (True Negatives), FP (False Positives), FN (False Negatives) y TP (True Positives) como se presenta en el Gráfico 1.

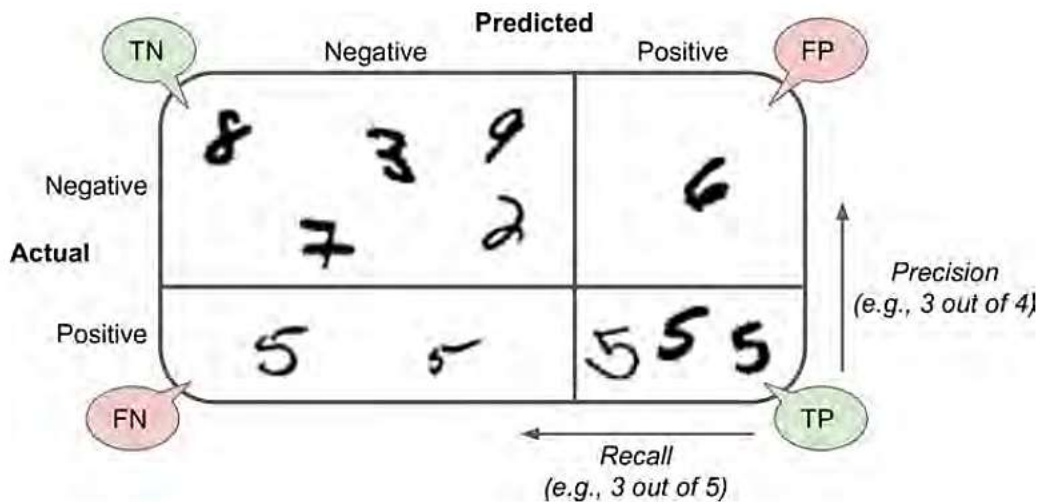


Gráfico 1: Matriz de confusión [25]

1.7. Aplicación de Machine Learning en el Ámbito Educativo

1.7.1. Revisión de Estudios Previos

En [26] se menciona que factores sociodemográficos, programa de estudio, beneficios obtenidos al ingresar, historial académico y rendimiento en el primer semestre de estudio son claves a la hora de modelar predicciones de deserción estudiantil. Además, el mejor algoritmo predictor según [26] fue el “random forest”, ya que tuvo una probabilidad de 0.83 y por capturaba el 34% de deserción real.

El modelo predictivo usando random forest se convierte en una herramienta útil para detectar de forma temprana una posible deserción estudiantil y evitarla,

evidenciando así que se pueden usar los datos para muchas aplicaciones útiles que brinden soluciones a la sociedad.

1.7.2. Modelos Predictivos Basados en Datos Académicos

La predicción del rendimiento de los estudiantes, se puede decir que es uno de los temas más importantes para los contextos de aprendizaje como las escuelas y las universidades, ya que ayuda a diseñar mecanismos eficaces que mejoran los resultados académicos y evitan el abandono escolar [27]. Varios estudios explican cómo distintos factores afectan la capacidad de un estudiante para tener éxito como son: graduarse, la preparación preuniversitaria, etc. Lograron analizar el impacto de dichos factores en tasas de graduaciones de una universidad usando tres modelos predictivos los cuales son: Máquinas de Vector de Soporte (SVMs), Procesos Gaussianos (GPs) y Máquinas de Boltzmann Profundo (DBMs) [28].

- Máquinas de Vector de Soporte: Es una técnica de clasificación de aprendizaje supervisado que tiene su origen en los años 90 con Vladimir Vapnik, y co-autores, y ha ido ganando popularidad desde entonces en muchos campos como la Estadística y la Minería de Datos [29]. El SVMs es una técnica poderosa y efectiva ya que es debido a su sólida base en matemáticas, lo cual permite poder generar modelos de clasificación precisos y generalizables.
- Procesos Gaussianos: Es aquella que ofrece un enfoque probabilístico en el aprendizaje supervisado y pueden ser usados en problemas de clasificación y regresión [30]. Esto quiere decir que son útiles para los problemas de clasificación y regresión ya que permite manejar la duda que dejan las predicciones ya que no son solo valores si no los intervalos de confianza.
- Máquinas de Boltzmann Profundo: Es una red neuronal estocástica que consiste en dos capas: una capa visible que contiene la entrada conocida y una capa oculta que contiene las variables latentes [31]. Esta estructura logra modelar relaciones difíciles en los datos con el fin de poder obtener características profundas ya que son útiles en el aprendizaje no supervisado.

Con estos 3 modelos en su estudio ellos alcanzaron un 86% de precisión para decir que hay cero demoras para la graduación la cual se demora de un año o demora de 2 o más años.

1.7.3. Modelos Predictivos Considerando Factores Socioeconómicos y Demográficos

La implementación de modelos predictivos, han demostrado su capacidad para enriquecer estrategias de prevención y mejorar las proyecciones de intervenciones. Si bien los estudios de predicción del crimen son limitados en América Latina, existen precedentes que muestran cómo estos enfoques pueden ser aplicados con éxito [32]. En ese artículo los autores proponen el uso de modelos de Machine Learning avanzados y el método SHAP para mejorar la interpretabilidad de los modelos. Su enfoque se basa en 7 teorías criminológicas, y seleccionan cuidadosamente las variables para la predicción de crímenes, esto permite que puedan tener una comprensión más profunda de los factores críticos para la predicción del crimen. Cabe recalcar que el método SHAP se base en un juego de ganancias en la cual todos los jugadores cooperan entre todos la cual van a tener una ganancia en común.

1.7.4. Desafíos y Limitaciones del Análisis Predictivo en el Contexto Educativo

En un estudio de [33] se basaron en un modelo predictivo basados en ML la cual se utilizan para prever el rendimiento académico de los estudiantes, basándose en datos históricos y en tiempo real, como calificaciones anteriores, participación en clase y datos sociodemográficos. El modelo les permitió obtener resultados con la finalidad de poder identificar a los alumnos en riesgo de bajo desempeño o abandono, esto permitió que los profesores puedan intervenir oportunamente con varias estrategias de apoyo.

1.8. Análisis Predictivo para la Toma De Decisiones

1.8.1. Definición de Toma de Decisiones Basada en Datos

La toma de decisiones basada en datos (Data-Driven Decision Making, DDDM) es un enfoque que utiliza datos objetivos y herramientas analíticas para guiar las decisiones en las organizaciones. En lugar de depender de intuiciones o experiencias pasadas, se enfoca en hechos cuantificables que proporcionan una visión clara sobre los procesos y resultados. Esto no solo aumenta la precisión en la

toma de decisiones, sino que también permite prever posibles resultados a partir del análisis de datos históricos y tendencias actuales [34].

La toma de decisiones basada en datos ayuda a gestionar sus procesos y estrategias al centrarse en datos concretos y análisis detallados, ya que mediante los datos podemos analizar y reemplazar la dependencia de suposiciones o intuiciones, este enfoque permite decisiones más informadas, alineadas con los hechos y las tendencias. Esta técnica o enfoque también fomenta una cultura organizacional más transparente, ya que las decisiones se fundamentan en información clara y accesible para todos los niveles de la organización, ya que el mundo cada vez más impulsado por los datos ya que no solo optimiza los procesos internos, sino que también mejora la capacidad de respuesta ante una acción.

1.8.2. Análisis Predictivo en la Toma de Decisiones Educativas

El análisis predictivo es una herramienta poderosa en la toma de decisiones educativas, permite identificar patrones en los datos y predecir resultados futuros, facilitando así intervenciones oportunas y basadas en evidencia, un ejemplo es la mejora del rendimiento y retención de estudiantes: Según investigaciones recientes, el análisis predictivo permite a las instituciones educativas identificar a los estudiantes en riesgo de bajo rendimiento o deserción, estos modelos utilizan datos históricos, como calificaciones y participación en actividades, para predecir el éxito futuro y brindar apoyo adicional a los estudiantes que más lo necesitan [35]. En particular, los modelos de aprendizaje automático, como los árboles de decisión y la regresión logística, son eficaces para predecir el abandono estudiantil y facilitar la implementación de estrategias preventivas [36].

La relevancia del análisis predictivo está en que facilita intervenciones oportunas y fundamentadas en evidencia, mejorando la gestión del rendimiento y la retención estudiantil. Por ejemplo, mediante el análisis de datos históricos, como calificaciones y participación en actividades, las instituciones educativas pueden identificar a los estudiantes en riesgo de bajo rendimiento o deserción, permitiendo que reciban el apoyo necesario a tiempo.

1.9. Desarrollo de Sistemas de Análisis Predictivo

1.9.1. Metodologías Ágiles

Una metodología de desarrollo de forma general consiste en un conjunto de principios, prácticas y procedimientos que guían el proceso de implementación de sistemas informáticos. Particularmente las metodologías ágiles se caracterizan por su enfoque flexible, iterativo e incremental que prioriza la entrega rápida de valor al cliente, la colaboración continua con los usuarios y sobre todo la capacidad de adaptarse a los cambios. Esto se consigue al dividir el proyecto de software en ciclos cortos denominados iteraciones, donde se efectúan entregas parciales pero funcionales con la finalidad de tener una retroalimentación constante por parte de los usuarios para promover la mejora continua.

En el desarrollo de sistemas de análisis predictivo, dichas metodologías ágiles son pertinentes dada la complejidad y naturaleza cambiante de los datos, además este tipo de sistemas requieren constantes ajustes en la selección de variables, el entrenamiento de modelos y validación de resultados. A continuación, se revisan metodologías ágiles como Kanban, XP y Scrum.

1.9.2. Metodología Kanban

La metodología Kanban se refiere a técnicas de representación visual de información para mejorar la eficiencia en la ejecución de las tareas de un proyecto [37], es por ello por lo que Kanban cumple el rol tanto de enfoque ágil como de herramienta y su objetivo principal es cumplir un conjunto de reglas definidas.

Esta metodología permite observar el flujo de trabajo dentro de un proyecto mediante un conjunto de normas que indican como organizar y mostrar las tareas por medio de un tablero Kanban. Una ventaja para destacar de Kanban es la facilidad de integración con otros métodos de enfoque de gestión de proyectos, lo cual hace que esta sea una metodología flexible.

1.9.3. Metodología XP

La metodología Extreme Programming (XP) fue desarrollada por Kent Beck, obtiene su nombre del proceso de tomar la mejor práctica y luego llevarla a su

máximo nivel [38]. XP se enfoca en mejorar la comunicación entre el equipo y el cliente, asegurando la mejora continua. Es así que [39] explica los principios de esta metodología, las cuales incluyen: las pruebas constantes, programación en parejas, la retroalimentación continua y el proceso de trabajo en XP.

Esta metodología busca la adaptabilidad del desarrollo del software con el objetivo de gestionar la complejidad de los requisitos del software a medida que estos van cambiando. Por lo tanto, XP cubre estos aspectos para que no afecten el avance del proyecto.

1.9.4. Metodología Scrum

La metodología Scrum se basa en un enfoque ágil que trabaja con iteraciones cortas llamados sprints, los cuales cumplen un periodo de tiempo de máximo un mes, dentro de los sprints incluyen varios elementos como lo son: Daily Scrum, el trabajo de desarrollo, revisión y retrospectiva del sprint [39]. Scrum es perfecto para proyectos en entornos complejos donde se requiere una entrega rápida. Por ello cuenta con 3 roles específicos como lo son el dueño o también denominado Product Owner quien representa a los interesados del proyecto, que es el responsable de mejorar el valor del producto y el equipo de desarrollo.

El Scrum Master quien facilita el proceso y asegura que el equipo cumpla con las normas de Scrum, pero no gestiona directamente el desarrollo del producto. Y por último el Equipo de Desarrollo o Team que cumplen con la responsabilidad de convertir los requisitos del producto backlog en trabajo funcional. Es evidente de la metodología Scrum se enfoca en entregar el mejor producto, pero en ciclos de tiempo corto. Promoviendo flexibilidad y mejora continua con la finalidad de adaptarse rápidamente a cambios. Un punto por destacar de esta metodología es que facilita la colaboración entre el equipo de trabajo con los interesados.

1.9.5. Herramientas de desarrollo

El desarrollo de inteligencia artificial, de forma específica, un sistema de análisis predictivo involucra la utilización de distintas herramientas tecnológicas avanzadas que permitan manipular grandes volúmenes de datos, la aplicación de algoritmos

de Machine Learning y la visualización de resultados. Dichas herramientas deben ser eficientes sobre todo en el procesamiento de información para entrenar modelos de predicción y presentar los resultados de manera accesible e interactiva para la toma de decisiones informadas. En los apartados siguientes se detallan alternativas sólidas y vanguardistas para la creación de este tipo de modelos predictivos basados en aprendizaje automático supervisado.

1.9.5.1. Lenguaje de Programación Python

Es un lenguaje de programación que se destaca por su sintaxis sencilla además de ser un lenguaje de alto nivel, interpretado, multiparadigma y orientado a objetos, es utilizado para realizar cualquier tipo de programa desde aplicaciones de escritorio hasta aplicaciones web; es un lenguaje interpretado, lo que ofrece ventajas como la velocidad de desarrollo e inconvenientes como una velocidad más baja al ser ejecutado [40].

Dentro de la programación, actualmente es uno de los lenguajes más usados por muchos desarrolladores gracias a su versatilidad ha logrado que sea ideal al momento de desarrollar aplicaciones, se puedan usar diferentes bibliotecas para una mayor legibilidad en los proyectos inclusive con inteligencia artificial.

1.9.5.2. Google Colab

Es una herramienta que proporciona Google para escribir y ejecutar código Python desde el navegador, siendo de gran ayuda y utilidad para trabajos de aprendizaje automático, los recursos de este son limitados lo cual es necesario para que Colab, pueda dar dichos recursos gratuitamente, siendo prohibidas las acciones asociadas a operaciones informáticas en bloques [41]. Es importante recalcar que Google Colab es ideal para proyectos de aprendizaje automático, y para personas que no cuentan con un hardware potente, logrando ser una herramienta efectiva en el ámbito de desarrollo de proyectos de análisis predictivo.

1.9.5.3. Django

Al implementar un modelo predictivo, Django se puede utilizar para construir una interfaz web que permita a los usuarios interactuar con el modelo y obtener predicciones en tiempo real. Django proporciona una estructura para definir

modelos de datos, vistas y plantillas, lo que simplifica el desarrollo de la interfaz de usuario y la comunicación con el modelo predictivo [42].

Django se constituye en una herramienta ideal para el análisis predictivo al contar con una colección de módulos que son a su vez reutilizables y pueden ser compartidos por los diferentes desarrolladores de manera segura mejorando la curva de aprendizaje.

1.9.5.4. Scikit-Learn

Es una biblioteca de código abierto para el análisis de datos escrito en Python, se basa en otras bibliotecas de Python como: NumPy, SciPy y matplotlib. Expone una amplia variedad de algoritmos de machine learning, tanto para aprendizaje supervisado como no supervisado, usando una interfaz consistente [43].

En la actualidad el uso de este tipo de bibliotecas es esencial para mejorar los conocimientos en machine learning, al ser eficiente y con una gran familiaridad con otras herramientas, mejorando el flujo de trabajo y con un enfoque profundo en el ámbito educativo.

1.9.5.5. SQLite

SQLite, un sistema de gestión de bases de datos relacionales (DBMS) de enorme éxito, sirve como base de datos preferida para sistemas integrados, el alto rendimiento de SQLite en estos sistemas está impulsado por una serie de decisiones de diseño deliberadas, que en conjunto enfatizan compatibilidad de hardware y baja sobrecarga de software [43]. Además de ser una herramienta muy conocida, cabe recalcar que es una opción viable para proyectos de análisis predictivo de escala mediana dado a que a medida que va creciendo la base de datos, el requisito de memoria también aumenta.

1.9.5.6. DBeaver

Es una herramienta de gestión de bases de datos, desarrollada sobre JAVA, potente que ofrece una interfaz gráfica de usuario (GUI) para interactuar con diversas bases de datos, es de código abierto y es compatible con una amplia variedad de sistemas de gestión de bases de datos, incluyendo MySQL, PostgreSQL, SQLite, Oracle, SQL Server y muchos más [44]. DBeaver es una herramienta muy poderosa siempre

y cuando se la complemente adecuadamente, cumpliendo con una solución más directa para explorar bases de datos, de manera que se pueda realizar un análisis predictivo adecuado.

1.10. Hipótesis

Si se aplican técnicas de Machine Learning sobre datos académicos, socioeconómicos y demográficos, entonces se podrá predecir la deserción estudiantil con una precisión superior al 90%.

CAPITULO II. MATERIALES Y MÉTODOS

1.11. Caracterización de la Institución

La Universidad Técnica de Cotopaxi nace el 24 de enero de 1995, fecha en la que inicia de manera autónoma con la formación de profesionales con sentido humanista. En la actualidad cuenta con su campus matriz ubicado en el cantón Latacunga, sector San Felipe, campus Salache y dos extensiones localizadas en los cantones de Pujilí y La Maná, cada una de estas cuenta con diferentes carreras de acuerdo con las facultades de Ciencias Agropecuarias y Recursos Naturales (CAREN), Facultad de Ciencias de la Ingeniería y Aplicadas (CIYA), Facultad de Ciencias Administrativas y Económicas (CAYE), Facultad de Ciencias Sociales Artes y Educación (CSAYE) que se detallan en la Tabla 3.

Tabla 3: Oferta Académica de la UTC por campus y extensiones [45]

CANTÓN	CAMPUS/TIPO	CARRERAS
Latacunga	Matriz	<ul style="list-style-type: none"> • Contabilidad y Auditoría • Administración de Empresas • Gestión de la Información Gerencial • Mercadotecnia • Gestión del Talento Humano • Economía • Finanzas • Diseño Gráfico Interactivo • Animación Digital • Diseño Gráfico • Psicología Social • Comunicación • Trabajo Social • Electricidad • Electromecánica • Industrial • Sistemas de Información • Hidráulica • Software

CANTÓN	CAMPUS/TIPO	CARRERAS
	Salache	<ul style="list-style-type: none"> • Agroindustrial • Agronomía • Turismo • Ambiente • Medicina Veterinaria • Agropecuarias • Biotecnología
Pujilí	Extensión	<ul style="list-style-type: none"> • Pedagogía de la Lengua y Literatura • Pedagogía de las Ciencias Experimentales (Matemática y Física) • Educación Básica • Educación Inicial • Pedagogía del Idioma Inglés
La Maná	Extensión	<ul style="list-style-type: none"> • Administración de Empresas • Contabilidad y Auditoría (Híbrida) • Electromecánica • Sistemas de la Información • Turismo (Híbrida) • Agroindustrial • Agronomía

Con base a los datos de la Tabla 3, se puede notar que la mayor parte de carreras se encuentran en el Campus Matriz seguido por la extensión de La Maná, por su parte el Campus Salache y extensión Pujilí tienen el mismo número de carreras. En el Gráfico 2 se detalla lo indicado de forma visual.

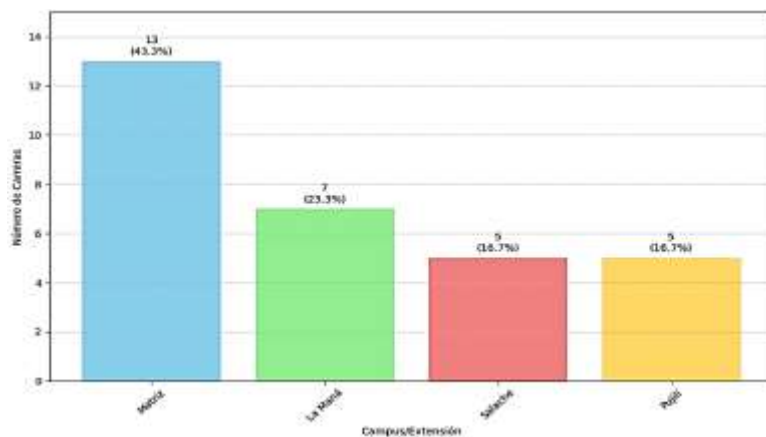


Gráfico 2: Total de carreras por campus o extensión de la UTC [45]

La Dirección de Tecnologías de la Información y Comunicaciones de la UTC es la responsable de gestionar la información de los estudiantes, para lo cual en su base de datos los agrupa en tres sedes. Los campus San Felipe y Salache en conjunto representan la sede Matriz (MUTC), mientras que para cada extensión se tiene una sede independiente, es decir se cuenta con la sede Pujilí (EPUJI) y también la sede La Maná (EMANA). Cada periodo académico los alumnos se matriculan en sus respectivas carreras para cursar las asignaturas correspondientes y de acuerdo con el Sistema de Gestión Académica de la universidad.

1.12. Diseño de la Investigación

1.12.1. Tipo de Investigación

La investigación es de carácter descriptivo dado a que este tipo de estudios buscan referir características de fenómenos o poblaciones, estableciendo asociaciones entre variables [46]. Es decir, esta tipología de investigación permite estudiar de forma detallada la problemática de la deserción estudiantil en la Universidad Técnica de Cotopaxi. En ese sentido, se busca evidenciar las variables que influyen en dicha deserción estudiantil mediante la aplicación de técnicas de aprendizaje automático que permitan predecir patrones negativos en los alumnos para una oportuna toma de decisiones por parte de los directivos de la universidad.

1.13. Enfoque de Investigación

El enfoque cuantitativo utiliza la recolección y el análisis de datos numéricos para identificar patrones empleando métodos estadísticos para el análisis de los datos [46]. Teniendo en cuenta que el trabajo propuesto requiere la recopilación y análisis de datos y variables numéricas destinadas a construir y evaluar un modelo predictivo, se optó por considerar el enfoque cuantitativo. Además, se puede señalar que el análisis predictivo y la aplicación de técnicas de aprendizaje automático están basadas en el procesamiento de datos cuantitativos para hacer inferencias y predicciones centradas en la realidad.

Por otro lado, dicho enfoque cuantitativo permite medir y evaluar la deserción estudiantil con precisión a través de datos numéricos. Con esto se puede desarrollar

un modelo que proporcione resultados objetivos y cuantificables, de igual modo se debe considerar que los algoritmos de aprendizaje automático requieren datos estructurados y numéricos capaces de ser analizados matemáticamente o mediante procesos estadísticos avanzados para que las predicciones sean más precisas.

1.14. Método de Investigación

El método inductivo es un método de investigación con énfasis en la detección de fenómenos, que comienza con la observación de un fenómeno y procede a derivar teorías o patrones generalizables a partir de él; por otro lado, la investigación deductiva abarca teorías y luego las pone a prueba utilizando datos empíricos [47].

Para el análisis predictivo de la deserción estudiantil en la Universidad Técnica de Cotopaxi, se usa una combinación de los métodos de investigación deductivo e inductivo. El deductivo permite considerar teorías existentes sobre la deserción estudiantil y aplicar modelos predictivos para validarlas, esto permite desarrollar un marco teórico sólido y probarlo mediante datos específicos.

El método inductivo, por su parte, facilita la recopilación y análisis de datos específicos para identificar patrones y tendencias, lo que ayuda a desarrollar nuevas teorías y modelos predictivos adaptados a la realidad de la universidad.

1.15. Población y Muestra

La población son los estudiantes de todos los campus y extensiones de la UTC matriculados en nivelación y primer ciclo durante el periodo académico octubre 2024 – marzo 2025, siendo un total de 1.727 (mil setecientos veinte y siete) alumnos. En el Gráfico 3 se presenta la estructura propuesta para determinar la población mediante una visualización tabular.



UNIVERSIDAD TÉCNICA DE COTOPAXI							
NIVELACIÓN Y PRIMER NIVEL							
2024-2025							
Fecha de reporte: 13/11/2024 09:00 Permanencia: [001]							
CEDULA	APELLIDOS	NOMBRES	PER	SEDE	FACU	CARRERA	NIVEL

Gráfico 3: Estructura del conjunto de datos considerados como población

En vista de que la población es difícil de manejar en su totalidad, se procede a calcular una muestra con el 95% de confianza y un margen de error del 5% conforme la formula y los cálculos presentados a continuación:

$$n = \frac{N\sigma^2 Z^2}{(N - 1)e^2 + \sigma^2 Z^2} \quad (1)$$

$$n = \frac{1.727(0,5)^2(1,96)^2}{(1.727 - 1)(0,05)^2 + (0,5)^2(1,96)^2} \quad (2)$$

$$n = \frac{1.727(0,9604)}{4,315 + 0,9604} \quad (3)$$

$$n = \frac{1.658,6108}{5,2754} \quad (4)$$

$$n = 314,40 \equiv \mathbf{315} \quad (5)$$

Por lo tanto, con los cálculos efectuados anteriormente mediante la fórmula de muestreo aleatorio simple basado en la desviación estándar de la población se obtiene que es viable trabajar con 315 estudiantes. Cabe señalar que en la formula se considera: la variable n como el tamaño de la muestra, el parámetro N como el tamaño de la población, el símbolo σ es la desviación estándar, Z es el valor del nivel de la confianza y la letra e representa el error muestral.

1.16. Técnicas de Investigación

1.16.1. Encuesta

La encuesta es una técnica de recolección de datos que implica la formulación y administración de un conjunto de preguntas a una muestra representativa de individuos con el fin de recopilar información sobre una problemática en particular, suele ser ampliamente utilizada debido a que permite obtener datos de modo rápido y eficaz [48]. Se optó por la técnica de la encuesta considerando que permite recolectar datos directos de los estudiantes de la Universidad Técnica de Cotopaxi, dichos datos son procesados para proporcionar información importante relacionada con las experiencias, percepciones y factores que pueden determinar la decisión de

continuar o abandonar los estudios superiores. La encuesta también permite identificar variables adicionales a las disponibles en el sistema de gestión académico de la UTC como la carga de trabajo, el apoyo social y circunstancias personales que podrían influir en la deserción estudiantil.

Otro factor relevante para optar por la encuesta es la posibilidad de tener una mayor participación y retroalimentación de los estudiantes generando un sentido de inclusión durante el proceso investigativo con el objetivo de obtener respuestas más sinceras y reflexivas, lo que incrementa la validez de los datos recopilados. Por otro lado, el diseño de una encuesta permite analizar numéricamente las respuestas y correlacionarlas para establecer posibles razones que desencadenan la deserción estudiantil y de este modo contribuir en la definición de estrategias preventivas y facilitar la toma de decisiones informadas.

1.17. Instrumento de Investigación

1.17.1. Cuestionario

De acuerdo con [49] el cuestionario es un instrumento para la recogida de información, diseñado para cuantificarla y universalizarla. Para el presente proyecto se considera dicho instrumento de investigación dado a que permite recolectar datos de forma sistémica y estandarizada contribuyendo en la comparación de respuestas de varios estudiantes. Con este instrumento se consigue un formato uniforme con preguntas útiles que hacen factible un análisis cuantitativo y/o estadístico requerido por las técnicas de machine learning.

Se usa preguntas cerradas y escalas de Likert para facilitar la cuantificación de respuestas relacionadas con los factores que pueden influir en la deserción estudiantil. Entre estos factores se consideran aspectos académicos, sociales y personales obteniendo así una visión integral de la problemática, a más que se puede recolectar datos a gran escala de manera rápida y eficiente.

1.17.2. Diseño del Instrumento

El instrumento (Anexo A) considera estudios previos realizados por [50] que proponen la extracción de perfiles de deserción estudiantil, a partir de datos

socioeconómicos en la Universidad Técnica de Ambato y [51] que establecen que algunos factores relevantes que influyen en la disertación estudiantil en la Universidad Politécnica Estatal del Carchi son: la edad, número de miembros en el hogar y nivel de estudio.

1.18. Metodología de Desarrollo Basada en Prácticas Ágiles

El presente proyecto cuenta con un equipo de trabajo y un tiempo de desarrollo limitado lo que hace que la alternativa más pertinente para llevar un proceso de ingeniería de software ordenado sea la aplicación de prácticas ágiles basadas en el marco de trabajo Scrum de acuerdo con los lineamientos presentados en la Tabla 4.

Tabla 4: Razones para adaptar las prácticas ágiles de Scrum en el proyecto

No	Detalle	Descripción
1	Iteraciones y Flexibilidad	Scrum propone la división del trabajo en Sprint, lo que permite ajustar el sistema a los objetivos a medida que se avanza en el desarrollo. Esto resulta útil para el presente proyecto dado a que los hallazgos iniciales pueden resultar cambiantes e influir entapas posteriores.
2	Enfoque en el Valor	El desarrollo ágil da prioridad a la entrega de avances funcionales en cada sprint por lo que se propone centrarse en los aspectos críticos del análisis predictivo como la recolección de un conjunto de datos para el entrenamiento y la selección de las técnicas de machine learning que ofrezcan resultados más relevantes.
3	Colaboración y Comunicación	Es importante fomentar la comunicación entre los involucrados con la problemática de la deserción estudiantes para evidenciar aspectos relevantes y obtener retroalimentación oportuna mediante revisiones periódicas.
4	Documentación Adaptativa	En lugar de generar una extensa documentación, con las prácticas ágiles es posible elaborar los artefactos de software más pertinentes, dándole valor al software funcional sobre una documentación exhaustiva.
5	Enfoque en Resultados	Las prácticas ágiles fomentan el mantener el foco en los objetivos del proyecto con el fin de avanzar adecuadamente en las actividades requeridas para desarrollar un análisis predictivo que contribuya de forma positiva en contrarrestar la deserción estudiantil.

No	Detalle	Descripción
6	Integración de Técnicas de IA	Se propone aplicar diferentes técnicas de machine learning en sprints cortos para evaluar los resultados obtenidos y ajustar el sistema a las necesidades de la Universidad Técnica de Cotopaxi.
7	Enfoque al Usuario	La finalidad del presente estudio es desarrollar un análisis útil para la toma de decisiones en la UTC por tanto es necesario mantener a los involucrados como parte del proceso asegurando que el resultado sea relevante y aplicable a lo largo del tiempo.

1.19. Roles del Equipo

Es crucial definir las responsabilidades del equipo para alcanzar el éxito en el desarrollo del sistema de análisis predictivo, para ello como se ha establecido se siguen prácticas ágiles que de acuerdo con Scrum se debe establecer dos roles: Scrum Máster y Development Team. Es decir que, el investigador asume el rol de Development Team y la tutora el rol de Scrum Máster

Con la organización del equipo se asegura un funcionamiento organizado del proyecto dado a que cada miembro tiene responsabilidades claras que contribuyen en la colaboración y la comunicación de todos los involucrados. Esto permite evitar confusiones sobre quién debe realizar una tarea específica y también permite que el equipo se enfoque en sus fortalezas individuales mejorando de este modo la eficiencia y calidad del trabajo. Así mismo, se tiene un progreso de trabajo ágil hacia los objetivos del proyecto de análisis predictivo de la deserción estudiantil.

1.20. Etapas del Proyecto

Se realiza un desarrollo flexible y adaptativo que permite al equipo ajustar el desarrollo según los resultados y la retroalimentación recibida por parte del Scrum Master, se inicia con la definición del product backlog que se constituye en una base sólida que guía el trabajo del equipo con base a los objetivos planteados. Se planifica cada uno de los sprints para promover que el equipo de desarrollo se enfoque en entregar funcionalidades relevantes en cada iteración, por otro lado, se debe realizar reuniones de trabajo para fomentar la comunicación constante y la resolución rápida de dudas o inconvenientes que surgen a lo largo del proyecto con

la idea de resolver conflictos de manera temprana. La revisión y retrospectiva es otro aspecto que debe llevarse a cabo al terminar cada sprint, por un lado, para evaluar el progreso, pero también es importante recibir retroalimentación que brinde oportunidades de mejora continua. En este proyecto de titulación seguir un proceso de desarrollo ágil iterativo y flexible es adecuado por los cambios significativos que pueden aparecer conforme se vaya indagando información relacionada con la deserción estudiantil en la UTC se propone el esquema de la Gráfico 4.

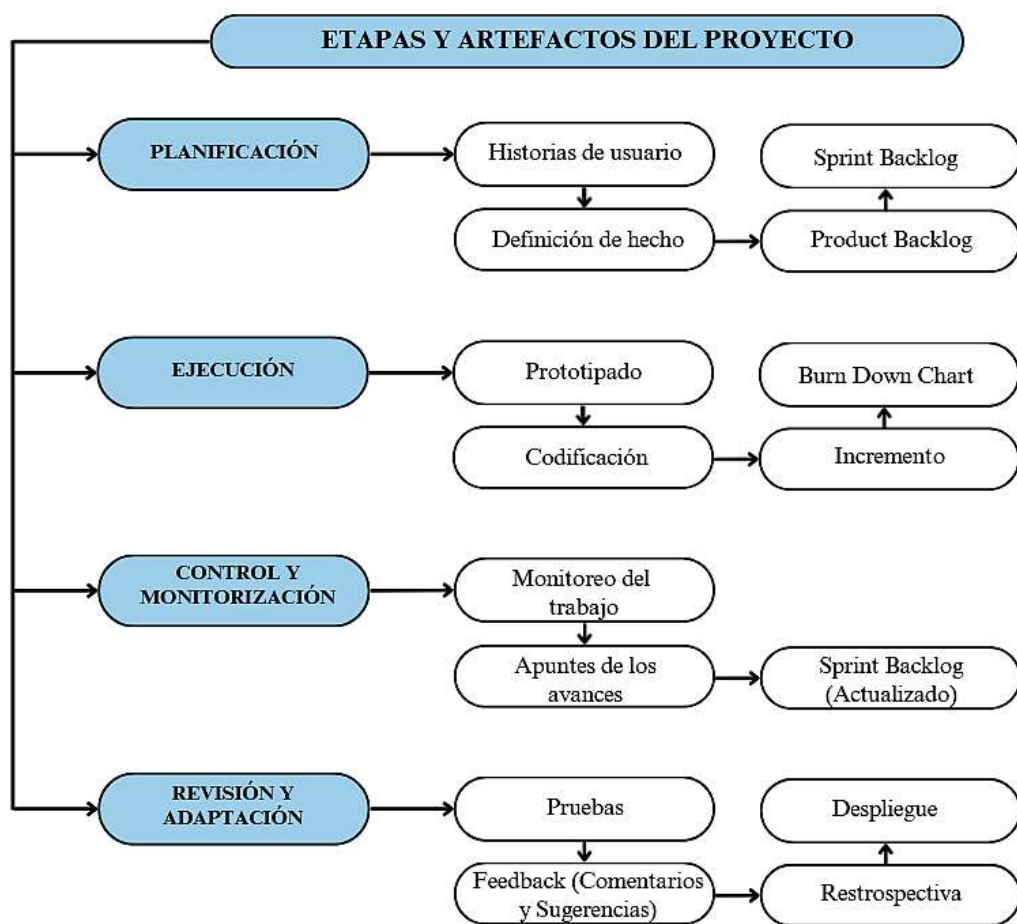


Gráfico 4. Etapas y artefactos de software considerados para el proyecto bajo un enfoque de desarrollo ágil

1.21. Herramientas de apoyo al desarrollo

La utilización de herramientas de desarrollo web y análisis de datos es fundamental para el proceso de análisis predictivo propuesto dado a que existen tecnologías que permiten la creación ágil de aplicaciones eficientes y escalables para el manejo de

grandes volúmenes de datos, así como también la implementación de modelos de aprendizaje automático para la predicción de la deserción estudiantil. Django es el framework Python más destacable dada su estructura robusta para el desarrollo de sistemas con interfaces intuitivas que faciliten la recolección, análisis y visualización de datos. Otra ventaja relevante de Python es que dispone de bibliotecas como Scikit-learn y Pandas que proporcionan potentes capacidades para el análisis y modelado de datos. En la Tabla 5 se especifican las características del software requerido.

Tabla 5: Herramientas de desarrollo para el análisis predictivo de deserción estudiantil

No	Nombre	Versión	Descripción
1	Python	3.10.12	Es el lenguaje de programación más versátil para la implementación de algoritmos de aprendizaje automático y análisis de datos.
2	Django	5.0.6	Framework de desarrollo de sistemas web usado para generar el formulario de recolección de datos y las interfaces de visualización de resultados
3	Scikit-learn	1.5.2	Ofrece diferentes algoritmos de aprendizaje automático para construir modelos eficientes para clasificación y predicción. Además, ofrece funcionalidades para la validación y evaluación de modelos de inteligencia artificial.
4	Pandas	2.1.4	Biblioteca para la manipulación y análisis de datos mediante estructuras (DataFrames), permite extraer, limpiar y transformar datos provenientes de diferentes fuentes como archivos de excel, CSV o bases de datos.
5	Matplotlib	3.7.1	Permite la visualización de datos mediante gráficas como histogramas, pasteles, barras, etc. Con esto se comunica de manera entendible los hallazgos relacionados con patrones vinculados a la deserción estudiantil.
6	Google Colab	0.0.1a2	Plataforma en línea para ejecutar análisis de datos, aprendizaje automático y desarrollo de proyectos de inteligencia artificial.
7	Visual Studio Code	1.92.2	Editor de código ligero y personalizable para proyectos con Python y Django, cuenta con varias extensiones que agilitan la escritura de código, la depuración y la gestión de entornos virtuales.

No	Nombre	Versión	Descripción
8	GIT	2.32.0	Sistema de control de versiones que permite llevar un seguimiento detallado de todas las versiones del código para gestionar los cambios y revertir de ser necesario versiones en el sistema predictivo.
9	GitLab	17.0	Proporciona un espacio de almacenamiento de código seguro proporcionando una copia de seguridad en la nube para evitar pérdida de datos y acceso al repositorio desde cualquier lugar.
10	Trello	Free	Plataforma en línea para gestionar las actividades en forma visual mediante tableros, listas y tarjetas que facilitan la organización de tareas en el proyecto.
11	Bootstrap	5.3	Es un framework front-end que proporciona un conjunto de componentes y estilos visuales predefinidos lo que permite diseñar interfaces gráficas atractivas de forma rápida.
12	SQLite	3	Sistema de gestión de bases de datos relacionales donde se registra las respuestas de los estudiantes para un análisis posterior mediante Python.

1.22. Proceso de Negocio

El proceso inicia activando el flujo del estudiante quien es el primero en interactuar con el sistema para responder el formulario diseñado para recabar sus criterios personales, esta etapa es crucial debido a que la calidad de las respuestas influye en la validez de los análisis posteriores; la recolección de datos provenientes de los estudiantes permite identificar factores que influyen en la deserción. El siguiente flujo es donde el sistema toma el control almacenando las respuestas en una base de datos para disponer de un repositorio organizado y accesible para futuros análisis, de igual forma el sistema emite un mensaje de confirmación al estudiante con el objetivo de mejorar la experiencia de usuario informándole que su accionar ha sido exitoso.

Se continúa con la ejecución del análisis predictivo que es realizado de forma automática por el sistema con base a los datos recopilados, esto permite identificar patrones y tendencias que pueden indicar el riesgo de deserción de un estudiante. Dichos resultados son cruciales para la toma de decisiones informada por parte de las autoridades con la finalidad de que sea posible establecer estrategias que

contrarresten el abandono académico en la universidad. Luego el sistema genera gráficos estadísticos con los resultados predictivos obtenidos, estas visualizaciones son vitales para comunicar los hallazgos de manera entendible a los administradores y demás interesados en la gestión de la deserción estudiantil.

Finalmente, los usuarios pueden descargar informes en diferentes formatos para un análisis más profundo de los patrones encontrados que están influyendo en la decisión de continuar o abandonar la universidad por parte de los estudiantes. En el Gráfico 5 se presenta el modelado de forma esquematizada.

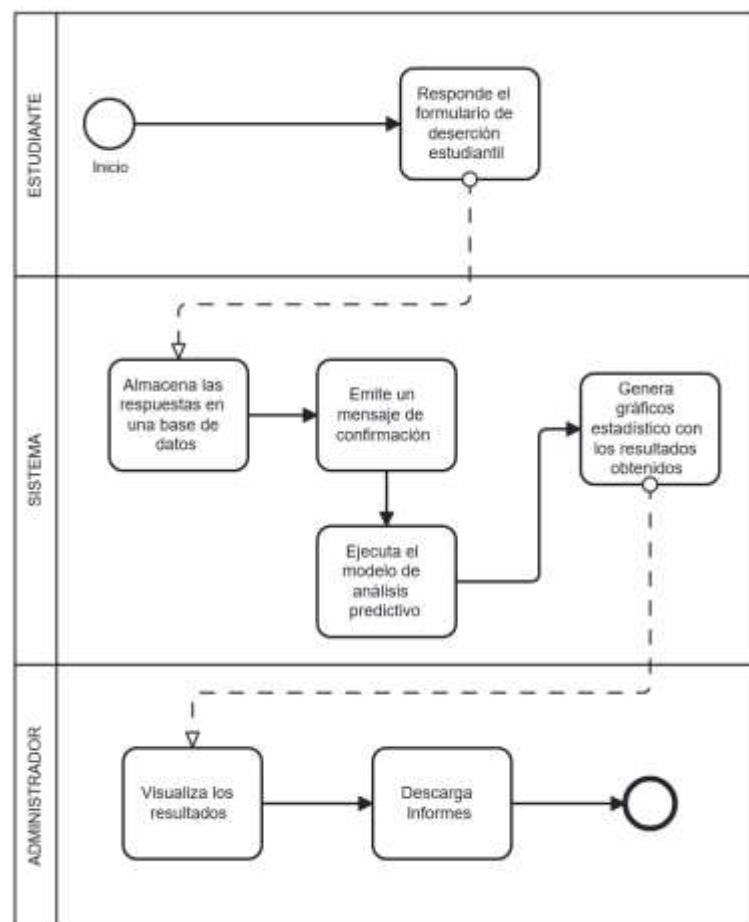


Gráfico 5. Modelo de negocio del sistema de análisis predictivo de la deserción estudiantil

1.23. Equipos Utilizados

El desarrollo fluido del sistema depende también de los componentes de hardware con los que se implementa cada una de las funcionalidades del mismo, es por ello

que se emplean un computador con características robustas (Tabla 6) para efectuar análisis de aprendizaje automático.

Tabla 6: Características del equipo utilizado para el desarrollo

No	Característica del Computador	Detalle
1	Marca	Asus
2	Sistema Operativo	Windows 11 / 64 Bits
3	Procesador	Core i7 / 10ma Generación
4	Memoria RAM	16 GB
5	Almacenamiento	Disco Solido de 512GB

1.24. Arquitectura del Sistema

La aplicación se desarrolla mediante las capas modelo, template y vista de acuerdo con lo establecido por el framework Django, en la arquitectura descrita en el Gráfico 6 se muestra la interacción entre los usuarios y el sistema para efectuar el análisis y presentación de resultados. Esta propuesta tiene un enfoque modular que facilita el mantenimiento y la escalabilidad, a más de ello, se asegura que cada componente pueda implementarse de manera independiente contribuyendo en la eficiencia global del proceso de análisis predictivo de la deserción estudiantil. Al ser un sistema web, la capa de interfaz de usuario (frontend) se implementa mediante tecnologías como HTML, CSS y Javascript optimizadas con Bootstrap para tener un formulario web intuitivo para el estudiante. Respecto al servidor (backend) se puede resaltar que en la arquitectura se especifica que este tiene la función de almacenar las respuestas en una base de datos, ejecutar el proceso de análisis predictivo y generar visualizaciones entendibles con los resultados obtenidos para esto se usa el lenguaje de programación Python las librerías Scikit-Learn y Matplotlib.

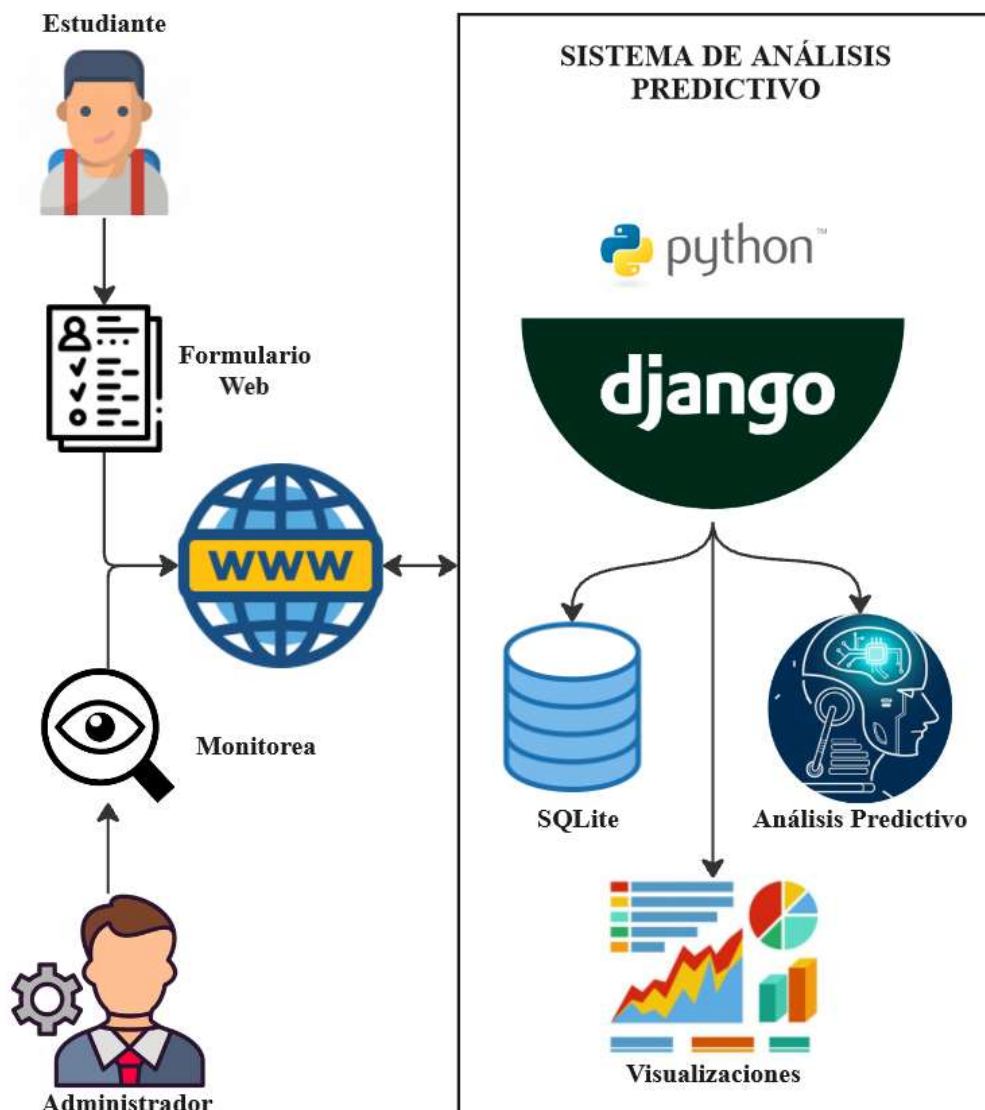


Gráfico 6. Arquitectura del sistema de análisis predictivo de la deserción estudiantil

1.25. Formato de Historias de Usuario

El uso de historias de usuario es una práctica ágil fundamental para recopilar las necesidades y expectativas de los usuarios finales, el formato diseñado en la Tabla 7 se basa en una adaptación de los esquemas comúnmente utilizados en metodologías ágiles como Scrum y Extreme Programming. Con este enfoque se busca tener una estructura clara y concisa que facilite la comprensión y comunicación efectiva entre los integrantes del equipo. El formato establece quien es el usuario que requiere una acción y porque es importante, lo cual permite mantener una perspectiva centrada en el usuario.

Tabla 7: Formato para definir las historias de usuario

Historia de Usuario No:	Número secuencial	Criterios de Aceptación:
Yo como:	Indicar la persona que usa la funcionalidad	<ul style="list-style-type: none"> • Condiciones específicas que una historia de usuario debe cumplir para ser considerada completa y aceptable.
Quiero:	Establecer la necesidad	
Para:	Definir los beneficios que se requieren	
Definición de Hecho (DoD):	<ul style="list-style-type: none"> • Detalles claros de cuándo una historia de usuario se considera completada. 	

1.26. Formato para la Pila del Producto

La gestión eficiente del desarrollo del sistema de análisis predictivo de la deserción estudiantil depende de una correcta definición de la pila del producto para esto se propone una estructura organizada que brinde una visión clara de todas las funcionalidades que deben implementarse, a la par se debe priorizar las tareas en función del valor que le aportan al usuario final. Por otro lado, el ordenar las actividades de acuerdo con su prioridad permite que el equipo inicie su trabajo con aquellas necesidades más críticas que serán revisadas y aprobadas por el dueño del producto. En la Tabla 8 se expone el formato usado para la pila del producto del proyecto.

Tabla 8: Formato para definir la pila del producto

ID	Tarea	Prioridad	Estado	Estimación	Explicación
Identificador único para cada historia de usuario o tarea.	Enunciado que describe la necesidad del usuario.	Nivel de importancia (Alta, Media, Baja)	Indica el progreso actual (Pendiente, En progreso, Completo).	Cantidad de puntos o tiempo estimado para completar la tarea.	Detallar las razones que permitieron determinar los puntos de historia

1.27. Formato para ejecutar pruebas del sistema

Es indispensable verificar que las funcionalidades implementadas tengan concordancia con los criterios de aceptación definidos para cada una de las historias de usuario, esto hace necesario proponer un formato que ayude en la organización y trazabilidad de las pruebas para determinar posibles problemas y validar que realmente el sistema cumpla con las expectativas del usuario. Las pruebas de regresión ayudan a identificar errores de manera temprana durante el desarrollo lo que reduce el costo y tiempo requeridos para solucionarlos, lo que resulta de suma importancia para que el equipo pueda corregir y ajustar las implementaciones de manera eficiente evitando inconvenientes inesperados en fases avanzadas o durante el despliegue del sistema, en la Tabla 9 se muestra el formato utilizado.

Tabla 9: Formato para ejecutar las pruebas del sistema

ID	Descripción del Caso de Prueba	Resultado Esperado	Resultado Real	Estado
Identificador único de la prueba	Explicación detallada de lo que se va a probar	Lo que se espera obtener al ejecutar la prueba)	Lo que realmente se obtuvo al ejecutar la prueba	Indicar si la prueba está "Pendiente", "Aprobada" o "Fallida".

CAPITULO III. PROPUESTA

En este capítulo se presentan los procesos metodológicos y técnicos ejecutados para el desarrollo del sistema de análisis predictivo de la deserción estudiantil en la Universidad Técnica de Cotopaxi (UTC). La propuesta se estructura en distintas etapas, comenzando con la planificación que abarca la definición de historias de usuario, la articulación de la pila del producto y la planificación de los sprints. Estos elementos han sido esenciales para coordinar y organizar todas las tareas dentro de un marco de desarrollo ágil. Contribuyen en una implementación eficiente y ajustada a las necesidades de la institución. Además, se detallan las fases de ejecución, que incluyen la creación de prototipos de interfaz gráfica de usuario y la codificación de las funcionalidades. Cada fase se enfoca hacia garantizar que el sistema propuesto sea técnicamente sólido, relevante y útil para apoyar la toma de decisiones informadas que puedan prevenir la deserción estudiantil en la UTC. La implementación de esta propuesta busca generar un impacto positivo en la retención estudiantil, contribuyendo al bienestar de los estudiantes y al fortalecimiento de la institución educativa.

3.1. Etapa de Planificación

El proceso de desarrollo del sistema de análisis predictivo de la deserción estudiantil en la Universidad Técnica de Cotopaxi parte desde la definición de historias de usuario, sprint backlog, definición de hecho (DoD), product backlog y sprint backlog. Cada uno de estos artefactos aportan en la coordinación del tiempo requerido para implementar el sistema propuesto.

3.1.1. Historias de usuario

El Development Team definió las historias de usuario (Tablas 10, 11 y 12) basándose en la realidad de la Universidad Técnica de Cotopaxi, asegurando que cada funcionalidad del sistema de análisis predictivo esté alineada con las necesidades y desafíos específicos de los estudiantes y la institución. Este artefacto es crucial en el desarrollo del sistema, dado a que proporciona una guía clara para el equipo de desarrollo, permite priorizar el trabajo en función del valor que aporta

a los usuarios y facilita la comunicación entre todas las partes interesadas, lo que contribuye a la creación de un producto efectivo y relevante.

Tabla 10: Historia de Usuario No. HU001

Historia de Usuario No:	HU001	Criterios de Aceptación:
Yo como:	Estudiante	<ul style="list-style-type: none"> • El formulario debe ser accesible en línea. • Debe permitir responder a todas las preguntas. • Debe mostrar un mensaje de confirmación al finalizar.
Quiero:	Responder el formulario de deserción estudiantil	
Para:	Proporcionar información que ayude a identificar factores de deserción.	
Definición de Hecho (DoD):	<ul style="list-style-type: none"> <input type="checkbox"/> El formulario es accesible en línea desde cualquier dispositivo y navegador compatible. <input type="checkbox"/> Permite responder todas las preguntas y valida correctamente las respuestas. <input type="checkbox"/> Muestra un mensaje de confirmación al finalizar indicando que la información fue enviada correctamente. <input type="checkbox"/> Ha sido probado por usuarios y no presenta errores técnicos. <input type="checkbox"/> Cumple con las políticas de privacidad y protección de datos. <input type="checkbox"/> La documentación de uso está disponible y clara. <input type="checkbox"/> Aprobado por el propietario del producto. 	

Tabla 11: Historia de Usuario No. HU002

Historia de Usuario No:	HU002	Criterios de Aceptación:
Yo como:	Administrador	<ul style="list-style-type: none"> • Los resultados deben ser accesibles a través de una interfaz amigable. • Debe permitir la filtración y búsqueda de datos.
Quiero:	Visualizar los resultados derivados de las respuestas dadas por los estudiantes.	
Para:	Tomar decisiones informadas relacionadas a la deserción estudiantil en la UTC.	
Definición de Hecho (DoD):	<ul style="list-style-type: none"> • La interfaz es amigable e intuitiva para los administradores. • Los resultados derivados de las respuestas de los estudiantes son accesibles en tiempo real. • Se puede filtrar y buscar datos específicos dentro de los resultados. • La visualización de los resultados es clara y comprensible (tablas, gráficos, estadísticas). 	

	<ul style="list-style-type: none"> • El sistema muestra resultados de manera rápida y eficiente, sin tiempos de carga largos. • La seguridad está garantizada, permitiendo acceso solo a usuarios autorizados. • La funcionalidad ha sido validada por el administrador y cumple con los criterios de aceptación.
--	--

Tabla 12: Historia de Usuario No. HU003

Historia de Usuario No:	HU003	Criterios de Aceptación:
Yo como:	Administrador	<ul style="list-style-type: none"> • Los informes deben estar en un formato descargable (ej. PDF, Excel). • Deben incluir todos los datos relevantes.
Quiero:	Descargar informes	
Para:	Acceder a la información para su revisión y presentación.	
Definición de Hecho (DoD):	<ul style="list-style-type: none"> • Los informes están disponibles en formatos descargables como PDF o Excel. • Los informes incluyen todos los datos relevantes de manera clara y completa. • El proceso de descarga funciona correctamente sin errores. • Los informes pueden ser descargados por el administrador de manera rápida y sin problemas de rendimiento. 	

3.1.2. Pila del Producto

El Product Backlog es una herramienta fundamental en la gestión ágil del proyecto, actuando como un resumen priorizado de todas las historias de usuario, tareas y requisitos necesarios para el desarrollo del sistema. Permite que el equipo de desarrollo tenga una visión clara de las prioridades y el progreso del proyecto, facilitando la planificación de sprints y la asignación de tareas de manera eficiente. Este artefacto se ajusta a las necesidades cambiantes de la Universidad Técnica de Cotopaxi, lo que ayuda a mantener la alineación con los objetivos del negocio y a fomentar la colaboración continua entre todos los miembros del equipo.

3.1.2.1. Proceso de Estimación Realizado

La estimación de puntos de historia en el Product Backlog (Tabla 13) se definió utilizando la técnica de estimación ágil conocida como "Planning Poker", adaptada para un entorno en el que el equipo de desarrollo está compuesto por una sola persona (el investigador). En este enfoque, el desarrollador revisa cada historia de usuario y asigna un número de puntos basado en una escala de Fibonacci, considerando el esfuerzo y la complejidad asociados con cada tarea. Este proceso de autoevaluación permite identificar y reflexionar sobre los requisitos, facilitando una estimación relativa al esfuerzo del desarrollador. A medida que avanza el proyecto, el investigador puede ajustar las estimaciones según la experiencia adquirida, mejorando la planificación y el manejo del tiempo a lo largo de la ejecución del proyecto.

Tabla 13: Pila de producto del sistema de análisis predictivo

ID	Historia de usuario	Prioridad	Estado	Estimación	Explicación de la Estimación
HU001	Como estudiante, quiero responder el formulario de deserción estudiantil, para proporcionar información que ayude a identificar factores de deserción.	Alta	Pendiente	8 puntos	El formulario requiere diseño, validación de respuestas, y mensajes de confirmación. Esto implica trabajo en la interfaz y validación.
HU002	Como administrador, quiero visualizar los resultados derivados de las respuestas dadas por los estudiantes, para tomar decisiones informadas relacionadas a la deserción estudiantil.	Alta	Pendiente	5 puntos	La interfaz debe ser intuitiva, mostrar resultados en tiempo real y permitir filtrado y búsqueda de datos. La complejidad aumenta debido a la visualización de datos y la seguridad.
HU003	Como administrador, quiero descargar informes, para acceder a la información para su revisión y presentación.	Media	Pendiente	3 puntos	La generación de informes descargables en formatos como PDF o Excel es relativamente sencilla, pero requiere trabajo en asegurar que los datos sean claros y completos

Resumen de Estimaciones:

- HU001: 8 puntos (Mayor complejidad debido al diseño y validación del formulario).
- HU002: 5 puntos (Interfaz de resultados con filtrado y seguridad).
- HU003: 3 puntos (Descarga de informes, complejidad moderada).

3.1.3. Pila del Sprint

La pila del sprint (Tabla 14) contribuye significativamente a la gestión del proyecto, proporcionando una estructura clara y organizada que facilitó la planificación efectiva de las tareas. Este artefacto contribuyó a la transparencia y la responsabilidad, permitiendo al investigador monitorear su progreso y realizar ajustes en tiempo rápido, lo que incrementó la productividad general. Además, la pila del sprint alinea el trabajo diario con los objetivos del análisis predictivo de la deserción estudiantil, garantizando que las tareas priorizadas aportaran un valor significativo a la toma de decisiones oportunas y efectivas.

Tabla 14: Planificación de la Pila del Sprint

Sprint	Inicio	Fin	Historia de Usuario	Nota/Comentarios
Sprint No. 1	01/09/2024	13/09/2024 2 semanas	HU001 - Formulario de deserción	Trabajo en la interfaz y validación de respuestas
Sprint No. 2	15/09/2024	26/09/2024 2 semanas	HU002 - Visualización de resultados	Implementación de interfaz para resultados
Sprint No. 3	27/10/2024	07/11/2024 2 semanas	HU003 - Generación de informes	Implementación de generación de informes PDF/Excel
Sprint No. 4	10/11/2024	21/11/2024 2 semanas	HU001 - Revisión y pruebas del formulario	Ajustes de diseño, validación final
Sprint No. 5	24/11/2024	05/12/2024 2 semanas	HU002 - Revisión de visualización de resultados	Mejoras de usabilidad y seguridad
Sprint No. 6	05/01/2025	16/01/2025 2 semanas	HU003 - Revisión final de los informes	Pruebas finales de la generación de informes

En la Tabla 14 se aprecia que los Sprint del 1 al 3 se enfocan en las historias de usuario más grandes y de mayor prioridad. Las historias que tienen una estimación

más alta (como HU001, HU002 y HU003) deben recibir más tiempo al principio. Luego, se ejecuta los Sprints del 4 al 6 destinados a la revisión, pruebas, y optimización de las funcionalidades implementadas en los sprints anteriores. La idea fue tener tiempo suficiente para ajustar cualquier detalle y probar que todo funcione de acuerdo con lo esperado.

3.2. Etapa de Ejecución

La etapa de ejecución consiste en tomar los elementos planificados, diseñarlos y llevarlos a código funcional iniciando con el prototipado de interfaces gráficas de usuario GUI, la codificación e incremento de funcionalidades.

3.2.1. Diseños de Interfaz Gráfica de Usuario

3.2.1.1. Interfaz 1: Pantalla de Inicio

Los elementos considerados son el logo de la Universidad Técnica de Cotopaxi, botón "Empezar Encuesta" e información sobre el proyecto y propósito del cuestionario.



Gráfico 7: Interfaz de Inicio del Sistema

3.2.1.2. Interfaz 2: Validación de estudiantes

Previo a completar el formulario de encuesta los estudiantes como medida de seguridad deben ingresar su número de cédula como filtro para permitir que las respuestas solamente sean completadas por usuarios autorizados.

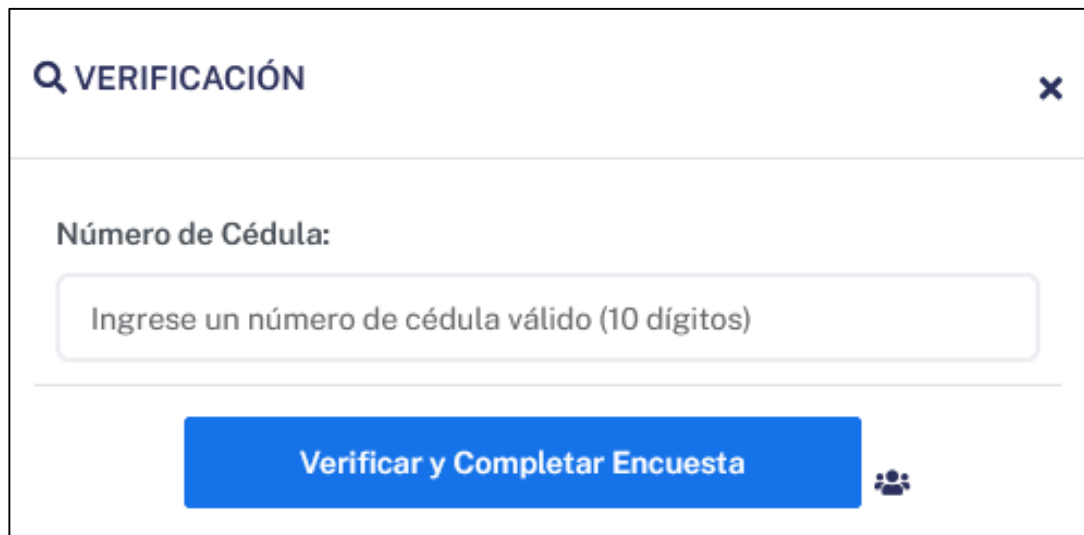


Gráfico 8: Interfaz de validación de estudiantes

3.2.1.3. Interfaz 3: Formulario de Deserción Estudiantil

Esta interfaz gráfica está compuesta por cinco secciones destinadas a recopilar información de los estudiantes sobre sus datos demográficos, socioeconómicos, preferencias de carrera, experiencia académica, bienestar y motivación.

CUESTIONARIO SOBRE DESERCIÓN ESTUDIANTIL ✕

Objetivo: Recopilar información relevante sobre las características demográficas, socioeconómicas, académicas y personales de los estudiantes de la Universidad Técnica de Cotopaxi, con el fin de identificar los factores que influyen en la deserción estudiantil.

Confidencialidad: Todas las respuestas son anónimas y se utilizarán únicamente con fines académicos. La privacidad será respetada en todo momento.

Sección

1. Edad <input type="text" value="Seleccione una Opción"/>	2. Género <input type="text" value="Seleccione una Opción"/>	3. Estado Civil <input type="text" value="Seleccione una Opción"/>
4. ¿Tienes Hijos? <input type="text" value="Seleccione una Opción"/>	5. ¿Tienes Alguna Discapacidad? <input type="text" value="Seleccione una Opción"/>	6. Etnia <input type="text" value="Seleccione una Opción"/>

Sección

7. Residencia según la zona de planificación en Ecuador. <input type="text" value="Seleccione una Opción"/>	8. Nivel de formación académica del padre. <input type="text" value="Seleccione una Opción"/>	9. Nivel de formación académica de la madre. <input type="text" value="Seleccione una Opción"/>
10. Tipo de colegio que realizó sus estudios. <input type="text" value="Seleccione una Opción"/>	11. ¿Cuál es el nivel de ingreso mensual de tu hogar? <input type="text" value="Seleccione una Opción"/>	12. ¿Con quién vives actualmente? <input type="text" value="Seleccione una Opción"/>
13. ¿Vives en un departamento alquilado? <input type="text" value="Seleccione una Opción"/>	14. ¿Cuál es la distancia promedio que recorres para llegar a la universidad? <input type="text" value="Seleccione una Opción"/>	15. ¿Trabajas mientras estudias? <input type="text" value="Seleccione una Opción"/>

19. En una escala del 1 al 5, ¿Cómo evaluarías tu rendimiento académico hasta ahora? <input type="text" value="Seleccione una Opción"/>	20. ¿Te sientes apoyado/a por tus profesores y la institución? <input type="text" value="Seleccione una Opción"/>
--	--

Sección

21. ¿Cómo calificarías tu situación alimentaria actual? <input type="text" value="Seleccione una Opción"/>	22. En una escala del 1 al 5, ¿Qué tan probable es que consideres abandonar tus estudios? <input type="text" value="Seleccione una Opción"/>	23. ¿Cuáles son las principales razones por las cuales podrías considerar dejar la universidad? (Selecciona todas las que apliquen).
---	---	--

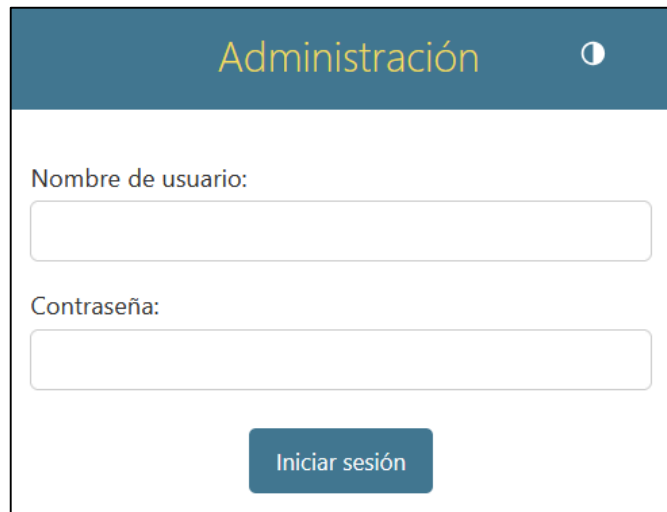
Problemas financieros
 Falta de interés en la carrera
 Problemas personales
 Dificultades académicas
 Carga laboral
 Responsabilidades familiares
 Distancia y transporte

Guardar Respuestas
o Cerrar

Gráfico 9: Interfaz del formulario de deserción estudiantil

3.2.1.4. Interfaz 4: Formulario de Autenticación

La seguridad es un factor fundamental dentro de cualquier sistema por lo cual en esta propuesta se solicita el nombre de usuario y contraseña para permitir el acceso a usuarios a las opciones de administración.



Administración

Nombre de usuario:

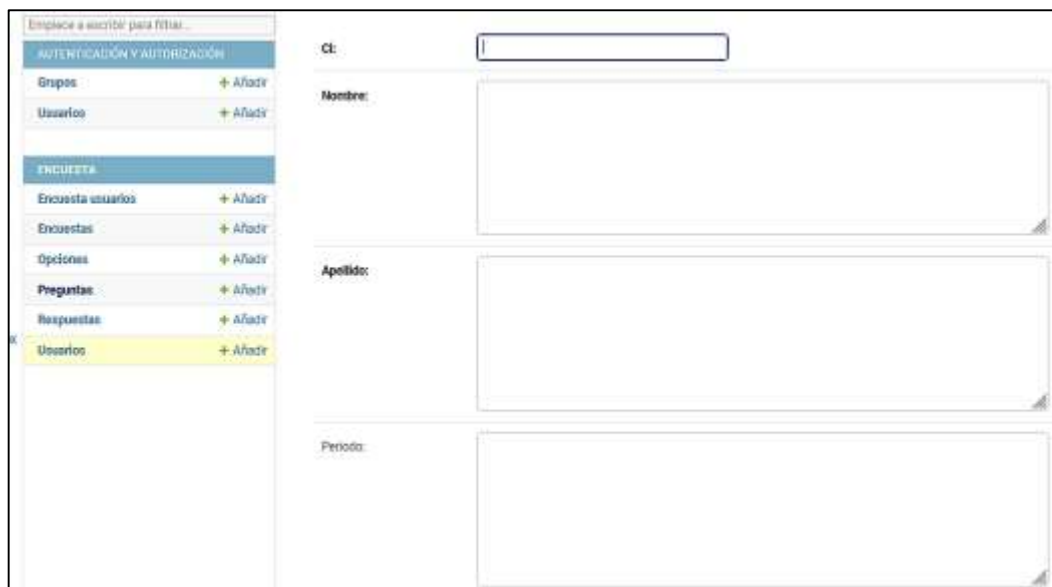
Contraseña:

Iniciar sesión

Gráfico 10: Interfaz de autenticación

3.2.1.5. Interfaz 5: Gestión de Estudiantes

Una vez que el usuario se autentica en el sistema tiene la opción de insertar, actualizar, consultar y eliminar estudiantes para esto se dispone de una interfaz de administración intuitiva.



Empiece a escribir para filtrar...

ANTERIOR

Grupos + Añadir

Usuarios + Añadir

ENCUESTA

Encuesta usuarios + Añadir

Encuestas + Añadir

Opciones + Añadir

Preguntas + Añadir

Respuestas + Añadir

Usuarios + Añadir

CI:

Nombre:

Apellido:

Periodo:

Gráfico 11: Interfaz de gestión de estudiantes

3.2.1.6. Interfaz 6: Menú de Opciones

En lugar de usar un menú tradicional en la parte superior que suele ocupar espacio valioso, en este proyecto se optó por un menú lateral (sidebar) que ayuda a que el contenido principal ocupe toda la parte superior de la pantalla, dejando el espacio lateral izquierdo para la navegación.



Gráfico 12: Menú de opciones del sistema predictivo

3.2.2. Modelo de Base de Datos

Teniendo en cuenta que el desarrollo del formulario de recolección de datos se hace mediante el framework Django con el lenguaje de programación Python, se usa SQLite como motor de base de datos para almacenar las respuestas. A continuación, se presenta el modelo de datos diseñado:

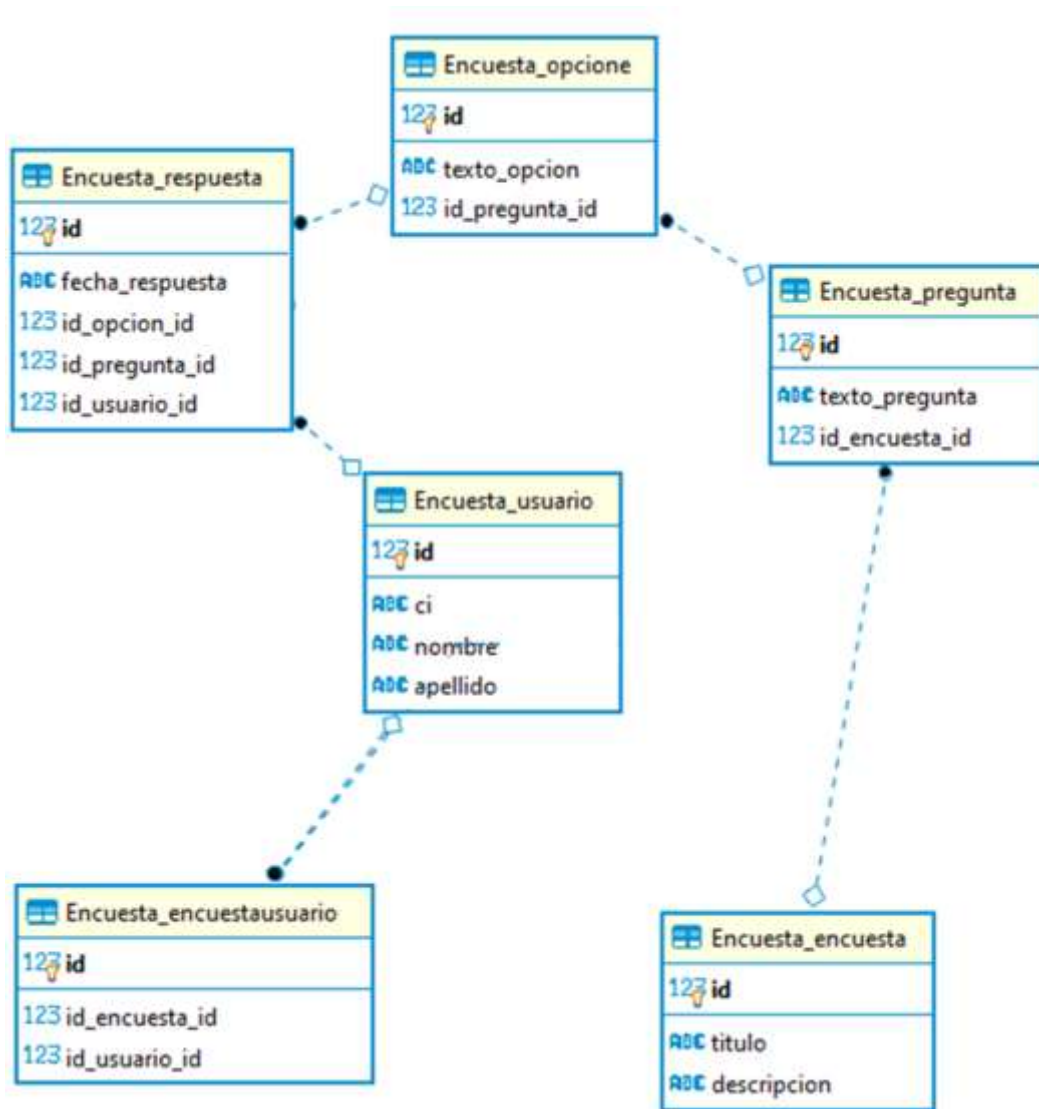


Gráfico 13: Modelo de datos del sistema de análisis predictivo

Django cuenta con un sistema de mapeo objeto relacional (ORM) que permite diseñar clases que se convierten en un modelo relacional compuesto por entidades, atributos y relaciones dentro del cual se almacenan todas las respuestas emitidas por los estudiantes que fueron recolectadas mediante el formulario web implementado.

```

Aplicaciones > Encuesta > models.py > ...
1  from django.db import models
2
3  class Usuario (models.Model):
4      id = models.AutoField(primary_key=True)
5      ci = models.CharField(max_length=10)
6      nombre = models.TextField()
7      apellido = models.TextField()
8      def __str__(self):
9          return f"{self.ci}, {self.nombre}"
10
11 class Encuesta (models.Model):
12     id = models.AutoField(primary_key=True)
13     titulo = models.TextField()
14     descripcion = models.TextField()
15     def __str__(self):
16         return f"{self.titulo}"
17
18 class EncuestaUsuario (models.Model):
19     id = models.AutoField(primary_key=True)
20     id_usuario = models.ForeignKey(Usuario, on_delete=models.CASCADE)
21     id_encuesta = models.ForeignKey(Encuesta, on_delete=models.CASCADE)
22     #para gestionar un seguimiento hay que colocar el autorizado (VARIABLE)
23     def __str__(self):
24         return f"{self.id}"
25
26 class Pregunta (models.Model):
27     id = models.AutoField(primary_key=True)
28     id_encuesta = models.ForeignKey(Encuesta, on_delete=models.CASCADE)
29     texto_pregunta = models.TextField()
30     def __str__(self):
31         return f"{self.id}"

```

Gráfico 14: Modelos diseñados en Django para el formulario de deserción estudiantil

3.2.3. Análisis de los Datos

El sistema web desarrollado con Django y SQLite resulta efectivo para recopilar datos provenientes de 533 estudiantes de la UTC en sus distintas sedes. Luego de recabar los datos demográficos, socioeconómicos, preferencias de carrera, experiencia académica, bienestar y motivación se procede a resumirlos de la siguiente manera:

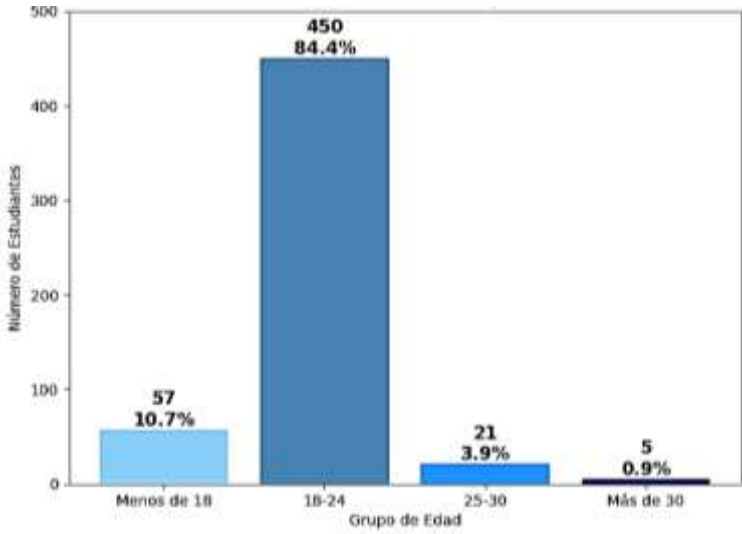
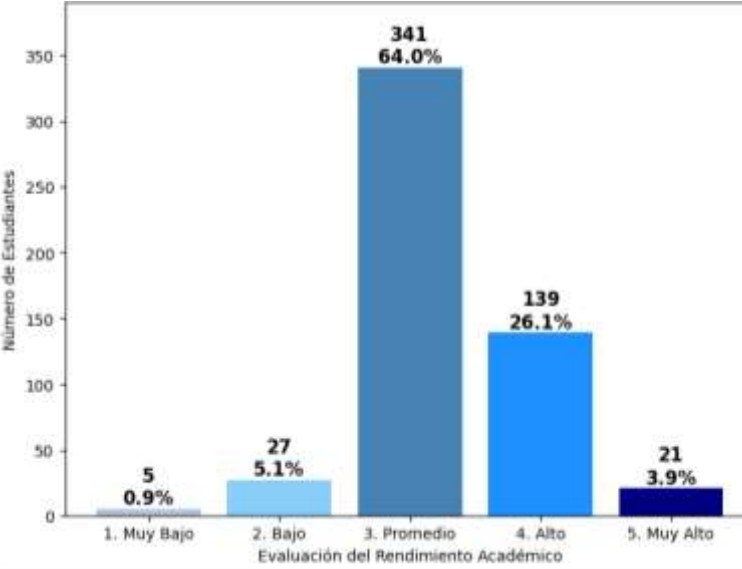
UNIVERSIDAD TÉCNICA DE COLOMBIA			
ENCUESTAS			
ADMINISTRACIÓN			
REPORTE DE RESPUESTAS			
533 INSTANCIAS RECOPIADAS			
ID	PREGUNTA	OPCIÓN	RESPUESTAS
1	Edad	Menos de 18	57
		18-24	458
		25-31	21
		Más de 30	5
		Misceláneo	171
2	Género	Femenino	382
		Soltero	158
		Conjugado	9
3	Estado Civil	Unión Libre	8
		Diversamente	1
		OT	38
4	¿Dónde vive?	Urb	435

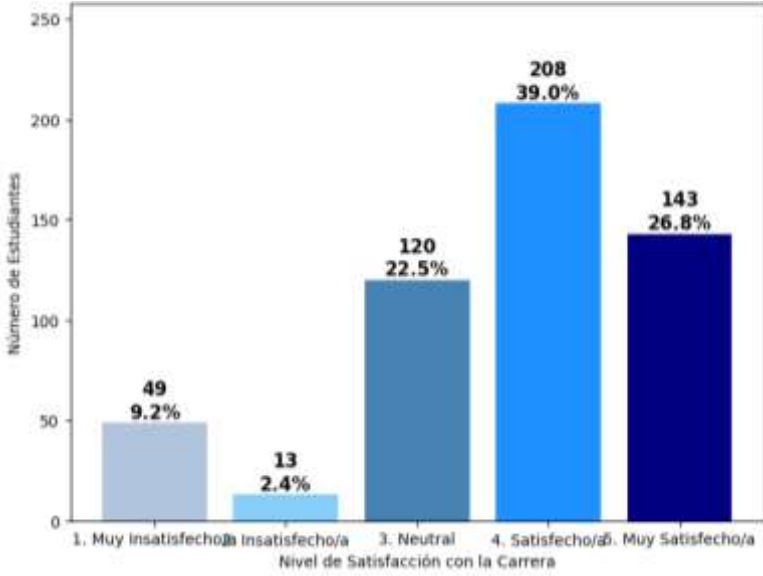
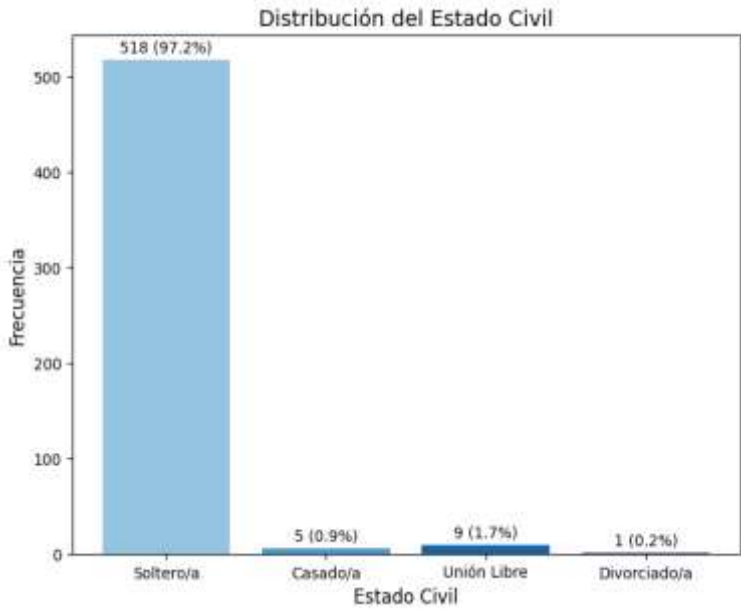
Gráfico 15: Resumen de datos recopilados mediante el cuestionario web

De acuerdo con los datos obtenidos se generan visualizaciones que permiten identificar de forma gráfica las tendencias y comportamientos de los estudiantes entorno a las preguntas recopiladas. En la Tabla 15 se describe los principales hallazgos obtenidos, cabe mencionar que la pregunta número 23 ubicada al final de dicha tabla tiene vital relevancia para el análisis predictivo de la deserción estudiantil:

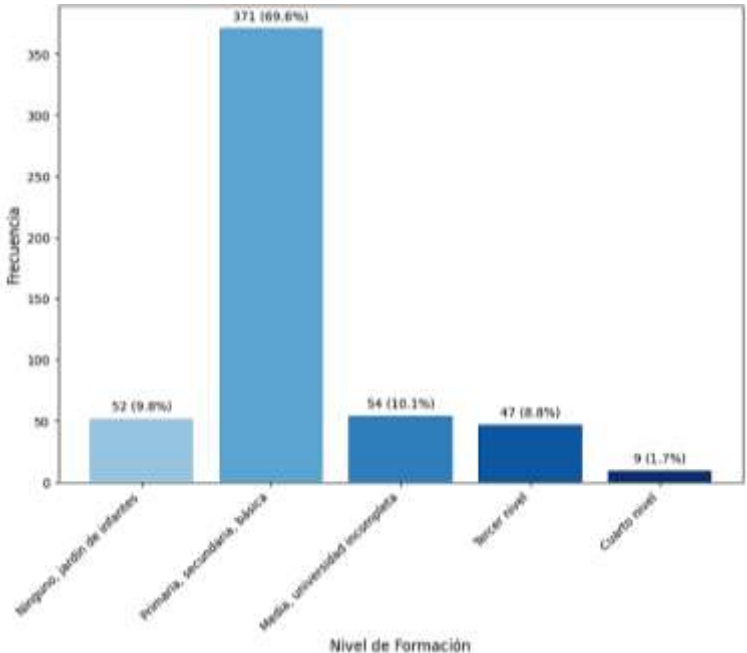
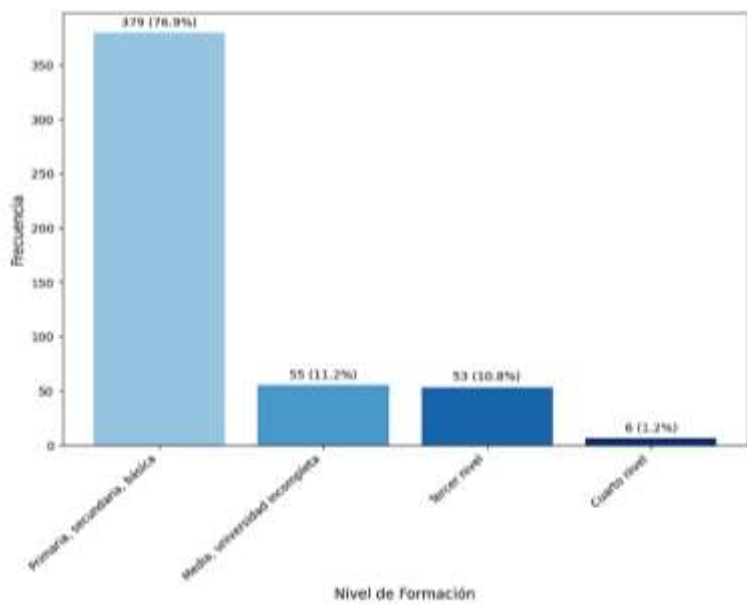
Tabla 15: Análisis e interpretación de los datos obtenidos

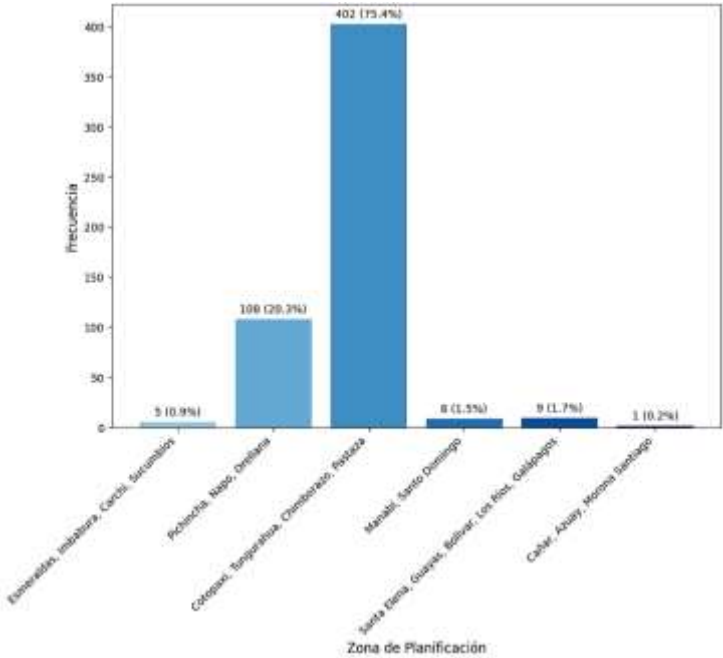
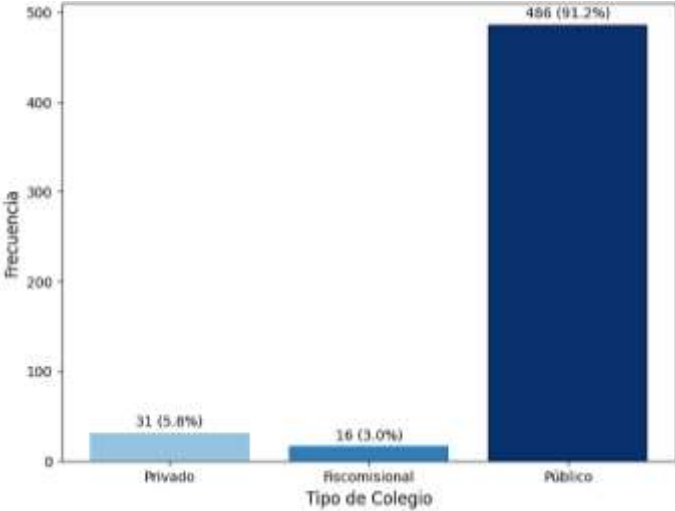
N	Pregunta	Resultados Obtenidos
1	¿Ciclo en que estas matriculado actualmente?	<p>La mayoría de los estudiantes se encuentran en primer ciclo, representando el 79.7% del total de respuestas (439 estudiantes). Le siguen los alumnos de nivelación con un 16.5% (91 estudiantes). El grupo con menor representación es el segundo ciclo, solo con 3 estudiantes (0.5%).</p>

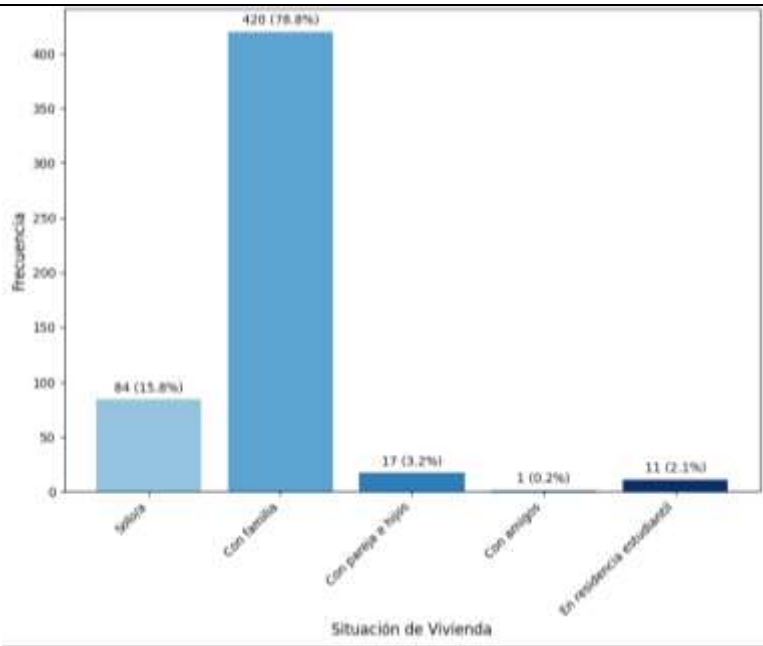
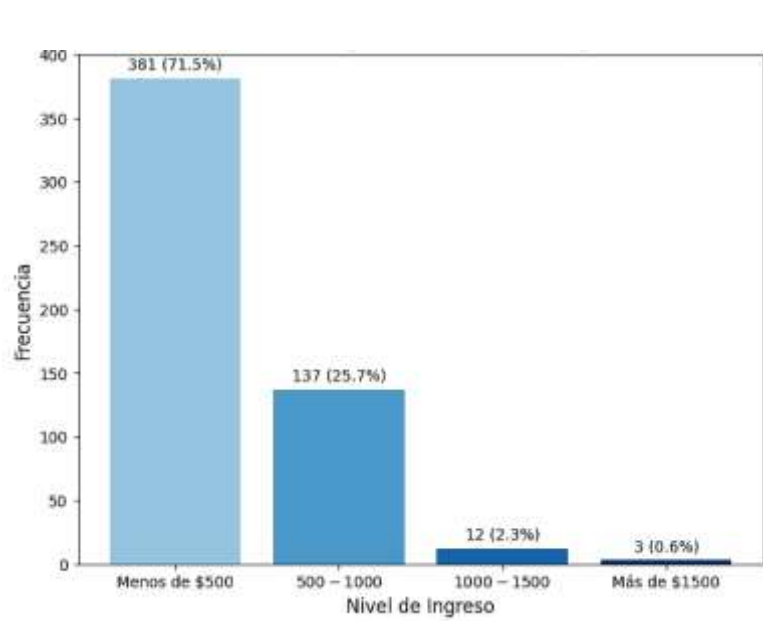
2	Edad	 <p>La mayor parte de los estudiantes pertenecen al grupo de 18-24 años, representando el 85.4% (450 estudiantes). Le sigue el grupo de Menos de 18 años con un 10.8% (57 estudiantes). Los grupos de 25-30 años y Más de 30 años tienen una representación menor, solo con 4.0% (21 estudiantes) y 0.9% (5 estudiantes) respectivamente.</p>
3	En una escala del 1 al 5, ¿cómo evaluarías tu rendimiento académico hasta ahora?	 <p>La mayor cantidad de estudiantes consideran que su rendimiento académico es "Promedio", representando el 64.0% del total de respuestas (341 estudiantes). Le sigue el grupo que se evalúa como "Alto", con un 26.1% (139 estudiantes). Las categorías de "Muy Alto" (21 estudiantes, 3.9%), "Bajo" (27 estudiantes, 5.1%) y "Muy Bajo" (5 estudiantes, 0.9%) tienen una menor representación.</p> <p>Esto indica que la percepción general del rendimiento académico se concentra en niveles medios y altos.</p>

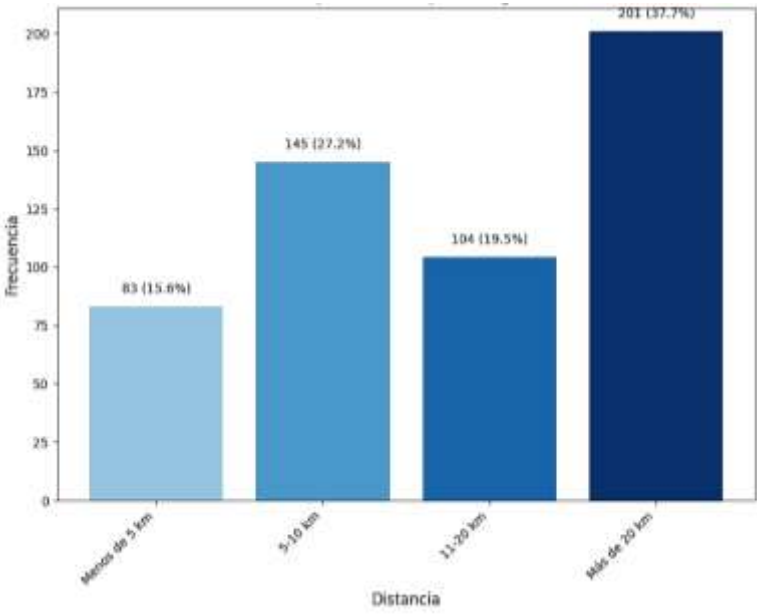
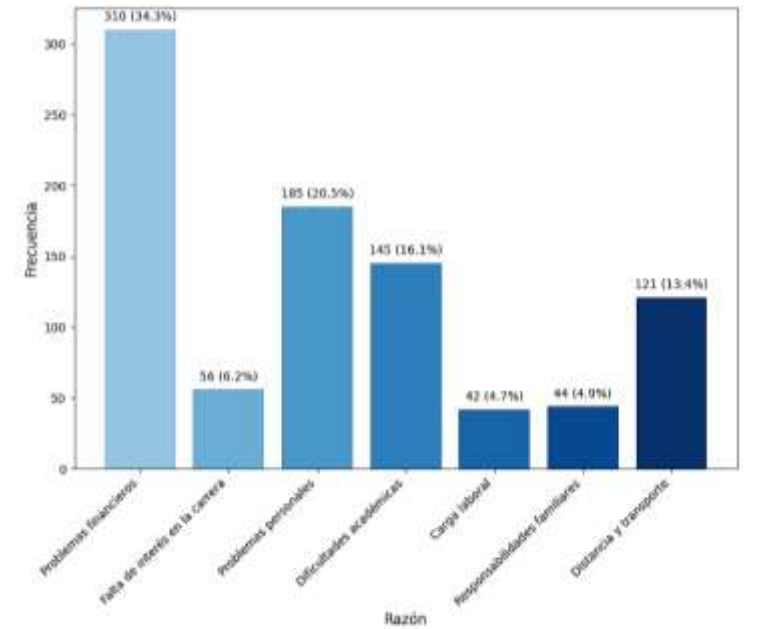
4	<p>En una escala del 1 al 5, ¿Qué tan satisfecho/a estás con tu elección de carrera?</p>	 <p>La mayoría de los estudiantes están satisfechos con su elección de carrera, con un 39.1% (208 estudiantes), seguido del grupo "Muy Satisfecho", que representa el 26.9% (143 estudiantes). Un 22.5% (120 estudiantes) se mantiene en una postura neutral, mientras que los niveles de insatisfacción son menores: "Muy Insatisfecho/a": 9.2% (49 estudiantes) e "Insatisfecho/a": 2.4% (13 estudiantes)</p>
5	<p>Estado Civil</p>	 <p>El estado civil con mayor frecuencia es "Soltero/a" (518 personas), que representa un 98.5% de la muestra. Los otros estados civiles son significativamente menos representados: "Casado/a" con 5 personas (0.9%), "Unión Libre" con 9 personas (1.7%) y "Divorciado/a" con solo 1 persona (0.2%).</p>

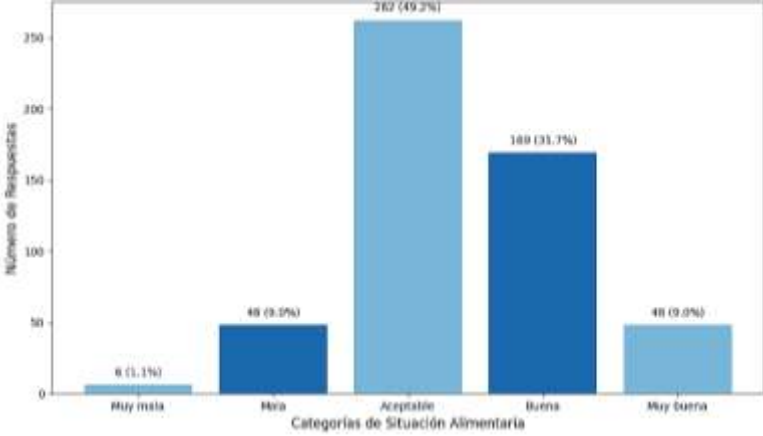
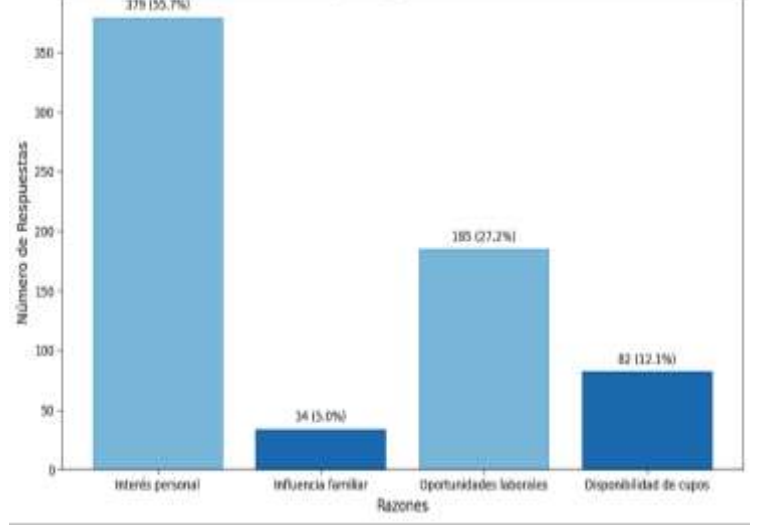
6	Etnia	<p>La etnia "Mestizo" es la más representada, con 469 personas, que corresponde al 88.6% de los estudiantes. Las otras etnias tienen una representación menor: "Indígena" con 57 personas (10.6%), "Montubio" con 6 personas (1.1%), y "Blanco" con solo 1 persona (0.2%).</p>
7	Género	<p>El género "Femenino" es el más representado con 362 personas, lo que representa el 67.9% de la muestra. El género "Masculino" tiene 171 personas, que corresponden al 32.1% de la muestra.</p>

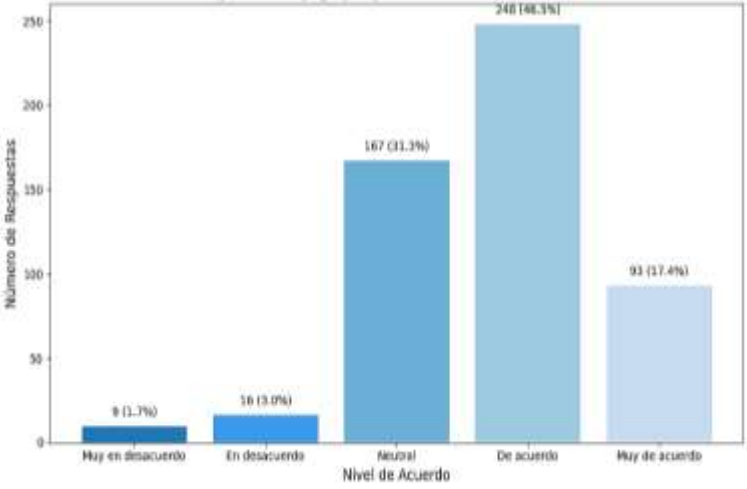
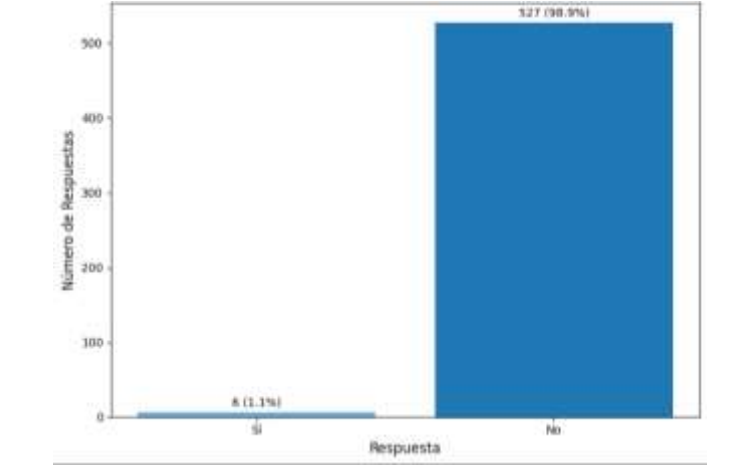
8	Nivel de formación académica de la madre	 <p>La categoría más representada es "Primaria, secundaria, básica" con 371 personas, lo que constituye el 71.2% de la muestra. Las demás categorías tienen una representación menor: "Media, universidad incompleta" con 54 personas (10.5%), "Tercer nivel" con 47 personas (9.1%), "Ninguno, jardín de infantes" con 52 personas (10.1%), y finalmente "Cuarto nivel" con 9 personas (1.7%).</p>
9	Nivel de formación académica del padre.	 <p>El nivel más representado es "Primaria, secundaria, básica", con 379 personas, lo que representa el 75.1% de la muestra. Las categorías restantes tienen una representación menor: "Media, universidad incompleta" con 55 personas (11.1%), "Tercer nivel" con 53 personas (10.6%), y "Cuarto nivel" con solo 6 personas (1.2%).</p>

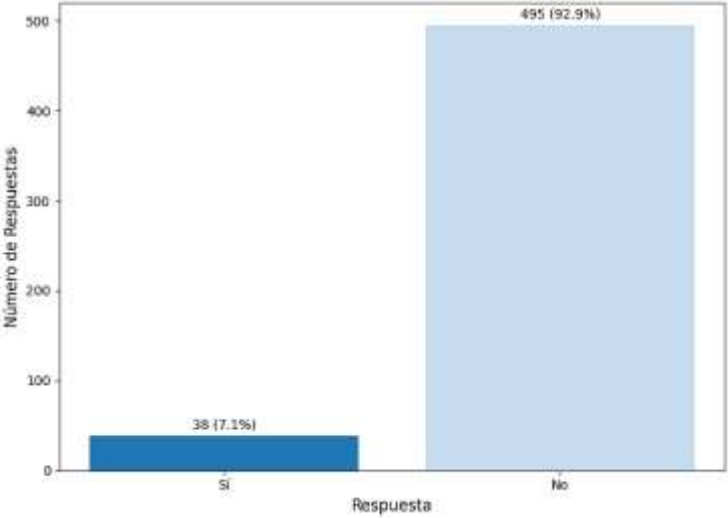
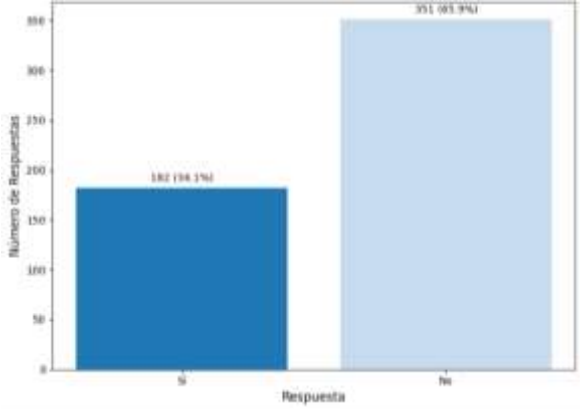
10	Residencia según la zona de planificación en Ecuador.	 <p>La zona con mayor población de estudiantes es "Cotopaxi, Tungurahua, Chimborazo, Pastaza" con 402 personas, lo que representa el 69.1%. Las otras zonas tienen una representación mucho menor: "Pichincha, Napo, Orellana" con 108 personas (18.5%), "Santa Elena, Guayas, Bolívar, Los Ríos, Galápagos" con 9 personas (1.5%), "Manabí, Santo Domingo" con 8 personas (1.4%), "Esmeraldas, Imbabura, Carchi, Sucumbios" con 5 personas (0.9%) y "Cañar, Azuay, Morona Santiago" con solo 1 persona (0.2%).</p>
11	Tipo de colegio que realizó sus estudios.	 <p>El tipo de colegio más frecuente es "Público" con 486 personas, lo que representa el 90.5%. Las otras categorías tienen una representación menor: "Privado" con 31 personas (5.7%) y "Fiscomisional" con 16 personas (3.0%).</p>

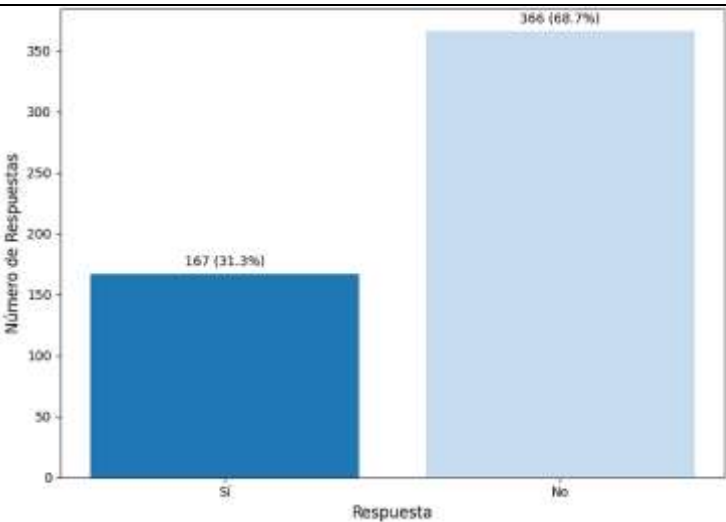
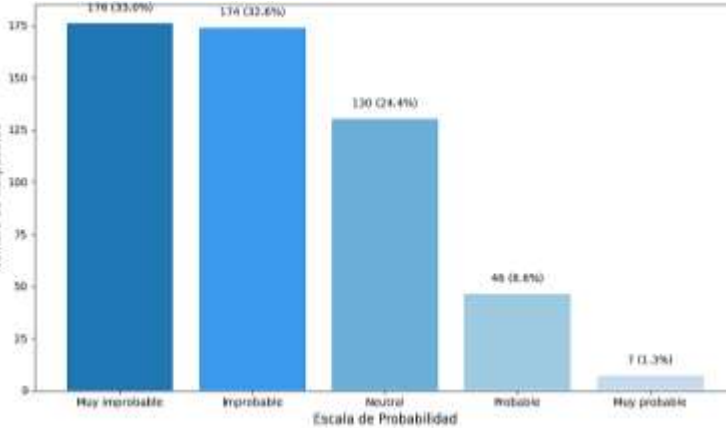
12	¿Con quién vives actualmente?	 <p>La categoría más seleccionada es "Con familia" con 420 personas, lo que representa el 75.1% de la muestra. Las otras categorías tienen una representación menor: "Solo/a" con 84 personas (14.9%), "Con pareja e hijos" con 17 personas (3.0%), "En residencia estudiantil" con 11 personas (2.0%) y "Con amigos" con solo 1 persona (0.2%).</p>
13	¿Cuál es el nivel de ingreso mensual de tu hogar?	 <p>La opción más frecuente es "Menos de \$500" con 381 personas, lo que representa el 71.7% de la muestra. Las otras categorías tienen una representación mucho menor: "\$500 - \$1000" con 137 personas (26.0%), "\$1000 - \$1500" con 12 personas (2.3%) y "Más de \$1500" con solo 3 personas (0.6%).</p>

14	¿Cuál es la distancia promedio que recorres para llegar a la universidad?	 <p>La categoría más representada es "Más de 20 km", con 201 personas, lo que representa el 37.5% de la muestra. Las otras categorías son "5-10 km" con 145 personas (27.1%), "11-20 km" con 104 personas (19.5%) y "Menos de 5 km" con 83 personas (15.5%).</p>
15	¿Cuáles son las principales razones por las cuales podrías considerar dejar la universidad?	 <p>La razón más común es "Problemas financieros" con 310 personas, lo que representa el 28.2% de la muestra. Las otras razones incluyen "Problemas personales" con 185 personas (16.8%), "Dificultades académicas" con 145 personas (13.1%), "Distancia y transporte" con 121 personas (10.9%), "Falta de interés en la carrera" con 56 personas (5.1%), "Responsabilidades familiares" con 44 personas (4.0%) y "Carga laboral" con 42 personas (3.8%).</p>

16	¿Cómo calificarías tu situación alimentaria actual?	 <p>La mayoría de los encuestados (45.4%) considera que su situación alimentaria es aceptable, lo que podría reflejar una situación intermedia, ni óptima ni completamente negativa. Solo un pequeño porcentaje percibe que su situación es muy mala o muy buena.</p>
17	¿Por qué elegiste tu carrera actual?	 <p>La mayoría de las personas (379 respuestas, 49.3%) elige su carrera basada en el interés personal, lo que indica que la motivación intrínseca es un factor fundamental para la decisión de carrera. En segundo lugar, las oportunidades laborales son una razón importante para 185 personas (24.4%), lo que sugiere que las perspectivas de empleo son un factor significativo para muchos al seleccionar su carrera.</p> <p>En comparación, las razones más externas, como la disponibilidad de cupos (82 respuestas, 10.7%) y la influencia familiar (34 respuestas, 4.4%), tienen menos peso en la toma de decisiones, lo que implica que, en su mayoría, las personas toman decisiones más autónomas o centradas en sus propios intereses y objetivos profesionales.</p>

18	<p>¿Te sientes apoyado/a por tus profesores y la institución?</p>	 <table border="1"> <caption>Distribución de Respuestas: Nivel de Acuerdo</caption> <thead> <tr> <th>Nivel de Acuerdo</th> <th>Número de Respuestas</th> <th>Porcentaje</th> </tr> </thead> <tbody> <tr> <td>Muy en desacuerdo</td> <td>9</td> <td>1.7%</td> </tr> <tr> <td>En desacuerdo</td> <td>16</td> <td>3.0%</td> </tr> <tr> <td>Neutral</td> <td>167</td> <td>31.3%</td> </tr> <tr> <td>De acuerdo</td> <td>248</td> <td>48.3%</td> </tr> <tr> <td>Muy de acuerdo</td> <td>93</td> <td>17.4%</td> </tr> </tbody> </table> <p>La mayoría de los encuestados (248 respuestas, 37.7%) se siente "De acuerdo" con el apoyo recibido por parte de sus profesores e institución, lo que sugiere que una gran parte percibe un nivel satisfactorio de respaldo. Sin embargo, también hay un grupo significativo que se siente "Neutral" (167 respuestas, 25.2%), lo que podría indicar que algunos estudiantes no tienen una percepción clara sobre el apoyo recibido o no consideran que sea ni negativo ni positivo.</p> <p>Un menor porcentaje de personas se siente "Muy de acuerdo" (93 respuestas, 14%), lo que muestra que hay estudiantes con una percepción fuerte y positiva sobre el apoyo institucional. Por otro lado, 9 personas (1.4%) se sienten "Muy en desacuerdo" y 16 personas (2.4%) "En desacuerdo", lo que indica que una pequeña minoría no recibe el apoyo que espera. En general, aunque la mayoría de los estudiantes se siente apoyada, existe un grupo considerable con una visión neutral o con cierta insatisfacción.</p>	Nivel de Acuerdo	Número de Respuestas	Porcentaje	Muy en desacuerdo	9	1.7%	En desacuerdo	16	3.0%	Neutral	167	31.3%	De acuerdo	248	48.3%	Muy de acuerdo	93	17.4%
Nivel de Acuerdo	Número de Respuestas	Porcentaje																		
Muy en desacuerdo	9	1.7%																		
En desacuerdo	16	3.0%																		
Neutral	167	31.3%																		
De acuerdo	248	48.3%																		
Muy de acuerdo	93	17.4%																		
19	<p>¿Tienes Alguna Discapacidad?</p>	 <table border="1"> <caption>Distribución de Respuestas: ¿Tienes Alguna Discapacidad?</caption> <thead> <tr> <th>Respuesta</th> <th>Número de Respuestas</th> <th>Porcentaje</th> </tr> </thead> <tbody> <tr> <td>Si</td> <td>6</td> <td>1.1%</td> </tr> <tr> <td>No</td> <td>527</td> <td>98.9%</td> </tr> </tbody> </table> <p>De las personas encuestadas, la gran mayoría (527 respuestas, 98.9%) no tiene ninguna discapacidad, lo que representa un porcentaje muy alto. Solo un pequeño grupo de personas (6 respuestas, 1.1%) sí tiene alguna discapacidad.</p>	Respuesta	Número de Respuestas	Porcentaje	Si	6	1.1%	No	527	98.9%									
Respuesta	Número de Respuestas	Porcentaje																		
Si	6	1.1%																		
No	527	98.9%																		

20	¿Tienes Hijos?	 <p>De las personas encuestadas, 495 personas (93.3%) no tienen hijos, lo que es una gran mayoría. En cambio, 38 personas (6.7%) sí tienen hijos, lo que indica que una pequeña proporción de los encuestados tiene hijos.</p>
21	¿Trabajas mientras estudias?	 <p>De las personas encuestadas, 182 personas (34.1%) trabajan mientras estudian, mientras que una mayor proporción, 351 personas (65.9%), no trabaja mientras estudia.</p>

22	¿Vives en un departament o alquilado?	 <p>De las personas encuestadas, 167 personas (31.3%) viven en un departamento alquilado, mientras que una mayoría significativa (366 personas, 68.7%) no viven en un departamento alquilado. Este patrón refleja que la mayoría de los encuestados posee una vivienda propia o no depende de un alquiler.</p>
23	Variable objetivo (abandono de estudios).	 <p>La mayoría de los encuestados considera "Muy improbable" (176 respuestas, 32.3%) o "Improbable" (174 respuestas, 31.8%) que abandonen sus estudios, lo que indica una fuerte motivación o compromiso con la educación.</p> <p>Un número considerable se muestra "Neutral" (130 respuestas, 23.7%), lo que podría reflejar dudas o incertidumbre, pero sin una inclinación clara hacia el abandono. Un porcentaje pequeño, 46 personas (8.3%), considera "Probable" abandonar los estudios, mientras que solo 7 personas (1.3%) encuentran "Muy probable" que lo hagan, lo que sugiere que, en general, la tasa de abandono percibida es baja.</p>

3.2.4. Modelado Predictivo

Para el modelo predictivo se tiene 23 preguntas con múltiples opciones de respuesta, y se inicia convirtiendo estas respuestas en un formato adecuado para el análisis predictivo, de forma que cada estudiante se represente con su respuesta a cada una de las preguntas, para esto se realiza lo siguiente:

- Se organizan las respuestas por preguntas y categorías: Se estructura la data en un formato en el que cada observación (fila) corresponda a un estudiante y sus respuestas. Se usan las frecuencias para expandir los datos, generando una fila por cada estudiante según su respuesta.
- Se codifican de las variables categóricas: Las preguntas tienen múltiples categorías (por ejemplo, género, estado civil, etc.), y se transforman en variables binarizadas usando One-Hot Encoding, en el apartado 3.2.8 y Gráfico 21 se detalla de forma minuciosa el proceso efectuado.
- Una vez que los datos estén en el formato adecuado, se procede a entrenar usando el modelo Random Forest y Máquina de Vector de Soporte para realizar predicciones y determinar el más eficiente. Estos procesos se detallan en los apartados 3.2.10 y 3.2.14 respectivamente.

3.2.5. Análisis Exploratorios Previos

Dentro de cualquier proyecto relacionado con la aplicación de aprendizaje automático es importante realizar un Análisis Exploratorio de Datos (EDA) porque permite disponer de una calidad y distribución adecuada de los datos antes de aplicar modelos de Machine Learning, esto contribuye en que los resultados sean confiables. En la deserción estudiantil de forma particular, el EDA ayuda a identificar patrones, tendencias clave, detectar sesgos en las respuestas, tratar valores atípicos o nulos, y visualizar mejor la distribución de los factores que influyen en la decisión de abandonar la universidad.

3.2.5.1. Distribución de los Datos

La distribución de los datos permite evidenciar si existe sesgo en las respuestas de la pregunta relacionada con la intención de abandonar los estudios: “En una escala

del 1 al 5, ¿Qué tan probable es que consideres abandonar tus estudios?”. Siendo las opciones de respuesta:

- Muy improbable
- Improbable
- Neutral
- Probable
- Muy probable

Con los datos tabulados en dicha pregunta se analizó mediante Python y Google Colab obteniendo los resultados presentados en los gráficos 16 y 17.

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import skew, kurtosis

# Datos de la pregunta clave con etiquetas
respuestas = [1, 2, 3, 4, 5]
etiquetas = ["Muy improbable", "Improbable", "Neutral", "Probable", "Muy probable"]
frecuencias = [176, 175, 131, 46, 7]

# Expandir los datos según sus frecuencias
datos_expand = np.repeat(respuestas, frecuencias)

# Calcular sesgo y curtosis
sesgo = skew(datos_expand)
curtosis_valor = kurtosis(datos_expand)

# Crear histograma con KDE
plt.figure(figsize=(8, 5))
sns.histplot(datos_expand, bins=5, kde=True, color="royalblue", edgecolor="black")

# Etiquetas personalizadas en el eje x
plt.xticks(respuestas, etiquetas, rotation=30)

# Líneas de media y mediana
plt.axvline(np.mean(datos_expand), color='red', linestyle='dashed', linewidth=2, label=f'Media: {np.mean(datos_expand):.2f}')
plt.axvline(np.median(datos_expand), color='green', linestyle='dashed', linewidth=2, label=f'Mediana: {np.median(datos_expand):.2f}')

# Títulos y etiquetas
plt.title("Distribución de Respuestas - Probabilidad de Abandonar Estudios")
plt.xlabel("Escala de Respuestas")
plt.ylabel("Frecuencia")
plt.legend()
plt.show()

# Imprimir análisis de sesgo
print(f"Sesgo: {sesgo:.2f} (Valores positivos indican sesgo a la derecha)")
print(f"Curtosis: {curtosis_valor:.2f} (Valores altos indican colas más pesadas)")
```

Gráfico 16: Código de análisis para determinar sesgo en la deserción estudiantil

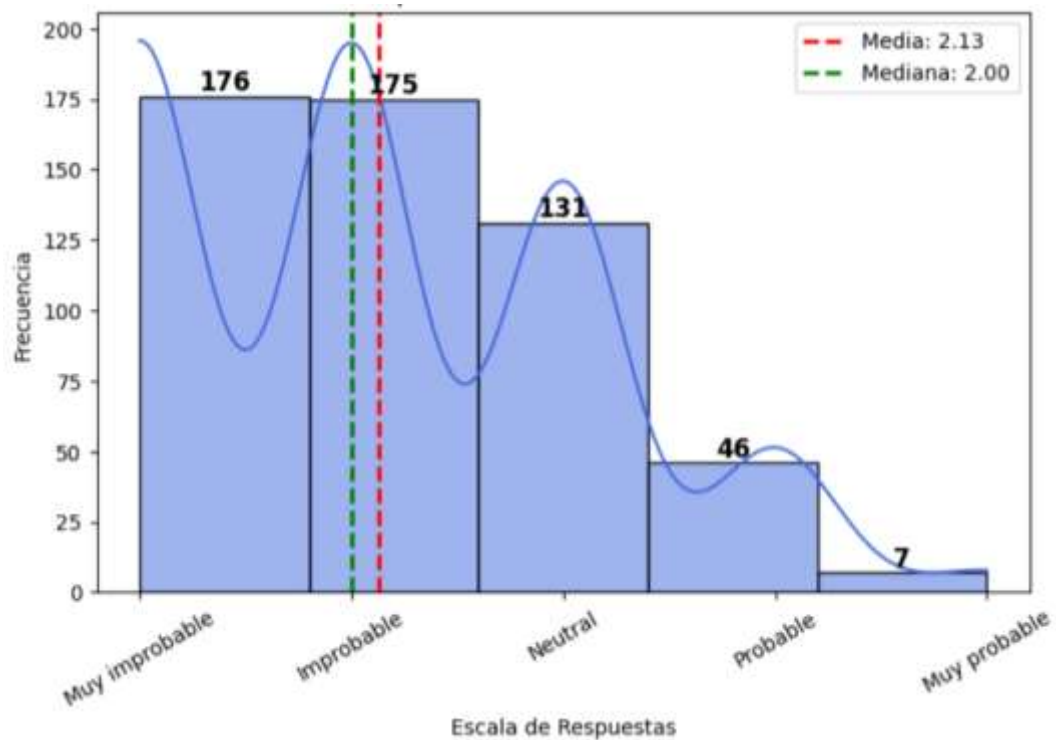


Gráfico 17: Distribución de respuestas sobre la probabilidad de deserción estudiantil

Con la Figura 17, se determina un sesgo de +0.55 que mide la asimetría de la distribución de los datos, al ser un valor positivo establece que la distribución esta ligeramente inclinada hacia la derecha. Esto indica que hay mayor cantidad de respuestas en las opciones Muy Improbable e Improbable por sobre el número de estudiantes que indicaron Probable o Muy Probable su decisión de abandonar sus estudios. Así mismo, se establece que no hay valores extremos significativos que denota que las respuestas están distribuidas sin grandes picos al no haber concentraciones extremas en una sola respuesta, esto sugiere una distribución pertinente para un análisis predictivo.

3.2.5.2. Tratamiento de Valores Nulos y Atípicos

La encuesta en línea desarrollada para recolectar datos tiene sus campos debidamente validados lo que evita datos nulos o en blanco, no obstante, en el código Python (Figura 18) usado para aplicar algoritmos de machine learning se hace una limpieza y agrupación, con el fin de que si existen múltiples respuestas para una misma pregunta se selecciona el valor más frecuente (moda). Así mismo, para manejar datos atípicos se efectúa una estandarización con *StandardScaler*

mediante el código del Gráfico 19 en el cual se convierten las variables a una distribución con media 0 y desviación estándar 1.

```
df_clean = df.groupby(["usuario_id", "pregunta_texto"])['opcion_texto']
            .agg(lambda x: x.mode()[0]).reset_index()
```

Gráfico 18: Manejo de valores nulos y/o atípicos

```
# Estandarización
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

Gráfico 19: Estandarización de datos

3.2.6. Procesamiento del DataSet

Se estructura un DataFrame con las respuestas y frecuencias de las preguntas, luego se expanden los datos para tener una fila por cada estudiante usando Google Colab y la librería Pandas, cuya salida tiene el siguiente formato:

	usuario_id	pregunta_texto	opcion_texto
0	1	Ciclo en que estás matriculado actualmente.	Primero
1	1	Edad	18-24
2	1	En una escala del 1 al 5, ¿Cómo evaluarías tu ...	3. Promedio
3	1	En una escala del 1 al 5, ¿Qué tan probable es...	2. Improbable
4	1	En una escala del 1 al 5, ¿Qué tan satisfecho/...	2. Insatisfecho/a

Gráfico 20: Cabecera del dataframe expandido

3.2.7. Definición de la Variable Objetivo

Dentro de cualquier modelo de aprendizaje supervisado se debe definir la variable objetivo que en esta investigación es la deserción estudiantil. Esto se plantea como un problema de clasificación binaria donde el modelo intenta predecir si un estudiante será desertor (1) o no desertor (0). Además, si esta variable no se escoge de forma pertinente, el modelo podría aprender valores irrelevantes. Es así, que la variable objetivo se basa en la pregunta No. 22: “En una escala del 1 al 5, ¿Qué tan

probable es que consideres abandonar tus estudios?” asignando las escalas descritas en la Tabla 16:

Tabla 16: Clasificación binaria para la variable objetivo

Variable Objetivo	0 (No desertor)	1 (Desertor)
Deserción Estudiantil	Si la respuesta es: 1. Muy improbable 2. Improbable 3. Neutral	Si la respuesta es: 4. Probable 5. Muy probable

3.2.8. Conversión de Variables Categóricas en Numéricas

Los algoritmos de machine learning tienen la particularidad de no poder procesar datos categóricos, es decir que estén en formato de texto como: masculino, femenino, privado, público, etc. Por lo tanto, es necesario convertir dichas categorías en números para que el modelo pueda interpretar y procesar los datos correctamente siendo capaz de calcular distancias, probabilidades y correlaciones entre variables. En la investigación se usa *LabelEncoding* para hacer dicha transformación como se evidencia en el Gráfico 21.

```
# Convertir variables categóricas en numéricas
encoder = LabelEncoder()
for col in df_pivot.select_dtypes(include=['object']).columns:
    df_pivot[col] = encoder.fit_transform(df_pivot[col].astype(str))
```

Gráfico 21: Conversión de variables categóricas a numéricas

3.2.8.1. Metodología de Conversión Utilizada

En el análisis predictivo se usa la técnica de mapeo manual de categorías a valores binarios para la variable objetivo “Deserción” como se detalla en el Gráfico 16, se usa la función *.map()* para convertir respuestas en una variable binaria donde 0 representa que el estudiante no considera desertar y 1 indica que si lo haría.

```
df_pivot["desercion"] = df_pivot["En una escala del 1 al 5, ¿Qué tan "
"probable es que consideres abandonar tus estudios?"].map(
    {"1. Muy improbable": 0, "2. Improbable": 0, "3. Neutral": 0,
    "4. Probable": 1, "5. Muy probable": 1}
)
```

Gráfico 22: Mapeo Manual de categorías a valores binarios

3.2.8.2. Matriz de Correlación de Variables

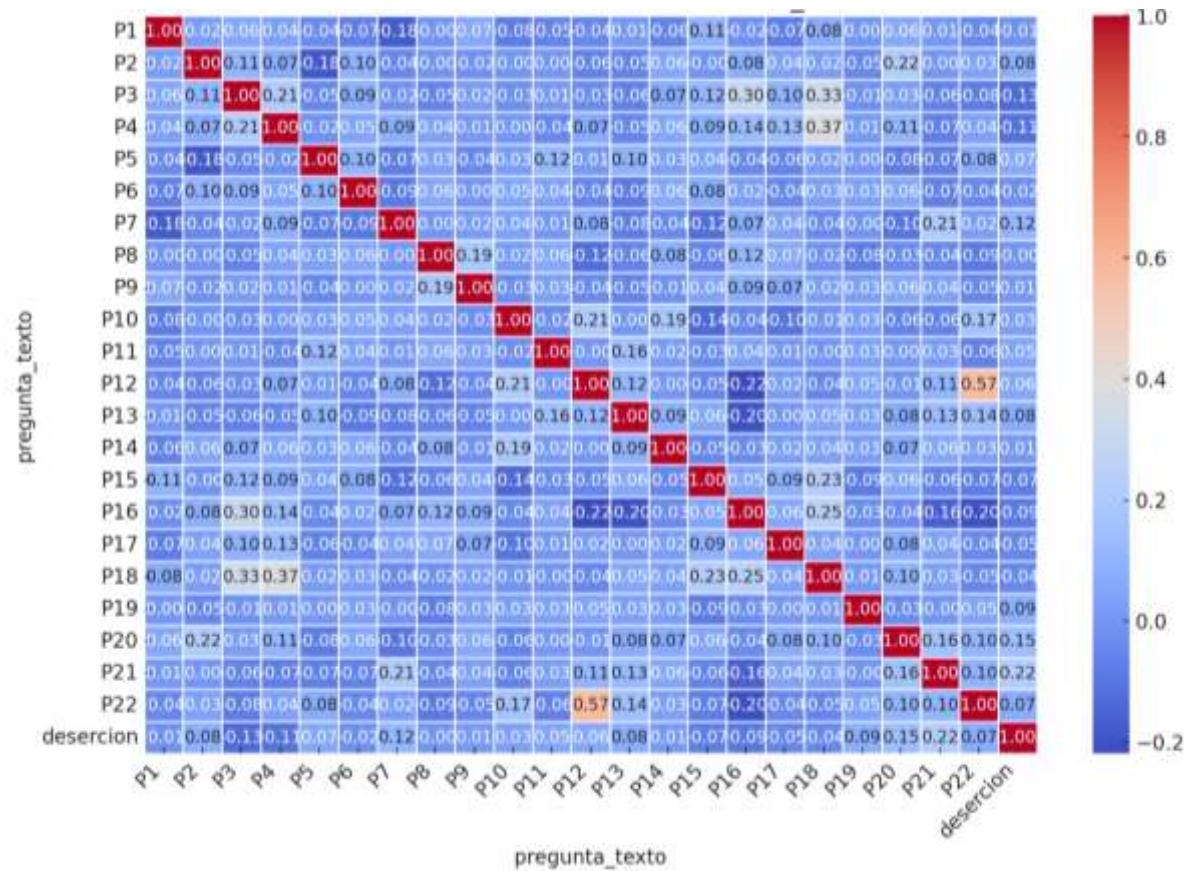


Gráfico 23: Matriz de correlación de variables

En el Gráfico 23 se presenta la matriz de correlación de variables donde para facilitar su visualización se asignó siglas a las variables según el número de la pregunta correspondiente en la encuesta completada por los estudiantes, la nomenclatura se describe en la Tabla 17.

Tabla 17: Nomenclatura usada en la matriz de correlación de variables

Código	Pregunta
P1	Ciclo en que estás matriculado actualmente.
P2	Edad
P3	En una escala del 1 al 5, ¿Cómo evaluarías tu rendimiento académico hasta ahora?
P4	En una escala del 1 al 5, ¿Qué tan satisfecho/a estás con tu elección de carrera?
P5	Estado Civil
P6	Etnia
P7	Género
P8	Nivel de formación académica de la madre.
P9	Nivel de formación académica del padre.
P10	Residencia según la zona de planificación en Ecuador.
P11	Tipo de colegio que realizó sus estudios.
P12	¿Con quién vives actualmente?
P13	¿Cuál es el nivel de ingreso mensual de tu hogar?
P14	¿Cuál es la distancia promedio que recorres para llegar a la universidad?
P15	¿Cuáles son las principales razones por las cuales podrías considerar dejar la universidad?
P16	¿Cómo calificarías tu situación alimentaria actual?
P17	¿Por qué elegiste tu carrera actual?
P18	¿Te sientes apoyado/a por tus profesores y la institución?
P19	¿Tienes Alguna Discapacidad?
P20	¿Tienes Hijos?
P21	¿Trabajas mientras estudias?
P22	¿Vives en un departamento alquilado?
P23	Variable objetivo (abandono de estudios).

La matriz de correlación muestra la correspondencia entre las variables estudiadas, dicha correlación varía entre -1 y 1. Los valores cercanos a 1 indican una fuerte relación positiva, es decir si una variable aumenta la otra también lo hace, si el valor

tiende a -1 la relación es negativa que indica que cuando una variable aumenta, la otra disminuye, mientras que los valores cercanos a 0 establecen una relación débil o inexistente entre variables. De acuerdo a esto, se evidencia que factores como el tener hijos, rendimiento académico, satisfacción con la carrera e ingresos económicos influyen en la deserción estudiantil.

3.2.9. Selección de Técnicas de Machine Learning Supervisado

En [52] se realiza un análisis comparativo de distintos modelos de machine learning aplicados al estudio de deserción estudiantil, se determina que el análisis predictivo requiere seleccionar un modelo que equilibre precisión, interpretabilidad y eficiencia computacional siendo Random Forest y Support Vector Machine las opciones más adecuadas en comparación con otras técnicas como redes neuronales o Gradient Boosting. Debido a las razones detalladas en la Tabla 18 se selecciona RF y SVM para aplicarlos en la presente investigación.

Tabla 18: Comparativa de Técnicas de Machine Learning [52]

	Random Forest (RF)	Máquina de Vectores de Soporte SVM	Redes Neuronales (ANN)	Gradient Boosting (GB)
Precisión en datos medianos	Alta	Alta	Necesita más datos	Alta
Computacionalmente eficiente	Rápido	Moderado	Costoso	Costoso
Robustez ante datos ruidosos	Alta	Alta	Sensible	Alta
Manejo de desbalance de clases	Bueno	Bueno	Bueno con técnicas avanzadas	Bueno
Facilidad de interpretación	Alta	Media	Baja	Baja
Requerimiento de hiperparámetros	Bajo	Medio	Alto	Alto

3.2.10. Entrenamiento del Modelo con RandomForest

Para entrenar el modelo de análisis predictivo de la deserción estudiantil en la UTC, se usa la técnica de aprendizaje supervisado llamada *RandomForest* debido a que tiene la capacidad de manejar tanto datos numéricos como categóricos de forma robusta y eficaz en la clasificación de datos con múltiples variables predictoras como es el caso de la encuesta que se compone de 23 preguntas en total. Previo al entrenamiento, los datos fueron limpiados y transformados mediante *LabelEncoding*, de igual modo se empleó *One-Hot Encoding* para evitar la generación de relaciones falsas entre categorías. Finalmente se eliminaron los valores nulos y se equilibró las clases para evitar sesgos en la predicción.

Para evaluar el desempeño del modelo se divide el conjunto de datos en dos subconjuntos:

- 80% para entrenamiento (X_{train}, y_{train})
- 20% para pruebas (X_{test}, y_{test})

Dicha división garantiza que el modelo predictivo pueda aprender patrones sin la necesidad de memorizar datos, con esto se tiene una evaluación objetiva. Al usar *Random Forest* se crea múltiples árboles de decisión y se combina sus predicciones para mejorar la precisión, la configuración del modelo considera los aspectos descritos en la Tabla 19.

Tabla 19: Configuración del modelo de análisis predictivo *Random Forest*

No	Descripción	Detalle
1	100 árboles de decisión (n_estimators=100)	Para tener un balance entre el rendimiento del algoritmo y el tiempo de cómputo.
2	Criterio de Gini	Para medir la calidad de la división en los nodos.
3	Profundidad máxima (max_depth=None)	Se permite que los árboles crezcan hasta que no se pueda hacer más divisiones significativas.
4	Selección aleatoria de características (max_features='sqrt')	Esto mejora la generalización y evita sobreajuste.

Una vez definido el modelo, se procede a entrenarlo con los datos de (X_train, y_train), haciendo un ajuste de los pesos de las variables para optimizar la predicción de la deserción estudiantil. En el Gráfico 24 se muestra el código generado:

```
# Separar variables predictoras y objetivo
X = df_pivot.drop(columns=["desercion", "usuario_id"])
y = df_pivot["desercion"]

# División del conjunto de datos en entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify=y)

# Estandarización de datos
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Entrenar un modelo de Random Forest
clf = RandomForestClassifier(n_estimators=100, random_state=42)
clf.fit(X_train, y_train)
```

Gráfico 24: Entrenamiento del modelo de análisis de la deserción estudiantil de la UTC con RF

3.2.11. Predicción y Evaluación del Modelo *RandomForest*

Después del entrenamiento del modelo se efectuó una evaluación usando métricas de clasificación como:

- Para medir la cantidad de predicciones correctas se usa Precisión (*Accuracy*).
- Con el fin de visualizar aciertos y errores en la clasificación de estudiantes desertores y no desertores se gráfica la matriz de confusión.
- La evaluación de la capacidad del modelo para identificar a los estudiantes con alto riesgo de deserción se hace con métricas clave como *Recall* y *F1-Score*.

En el Gráfico 25 se evidencia los resultados obtenidos de la predicción y evaluación del modelo, donde se observa una buena precisión global (*Accuracy*) de 0.896 (**89.6%**) para clasificar correctamente la mayoría de los casos, lo que demuestra su fiabilidad. Respecto a la precisión, la clase de no desertores (0) corresponde al 90% de los estudiantes lo que significa que el modelo acierta en la mayoría de los casos cuando no habrá deserción, mientras que el *Recall*, *F1-Score* y *Support* también tienen valores aceptables.

Accuracy: 0.8958333333333334				
	precision	recall	f1-score	support
0	0.90	1.00	0.95	86
1	0.00	0.00	0.00	10
accuracy			0.90	96
macro avg	0.45	0.50	0.47	96
weighted avg	0.80	0.90	0.85	96

Gráfico 25: Predicción y evaluación del modelo *Random Forest* en la deserción estudiantil de la UTC

3.2.12. Matriz de Confusión del Modelo *RandomForest*

La matriz de confusión se presenta en el Gráfico 26, donde la clase 0 corresponde a los no desertores mientras que la clase 1 son los desertores. Para la clase 0, se tiene 86 verdaderos negativos (TN) que significa que el modelo predijo correctamente que no desertan y 0 falsos positivos (FP) lo que indica que el modelo no clasificó erróneamente a alguien que no deserta como desertor. Respecto a la clase 1, se tiene 10 falsos negativos (FN) que denota que el modelo falló al identificar 10 casos de deserción y 0 verdaderos positivos (TP) evidenciando que el modelo nunca detectó correctamente un desertor.

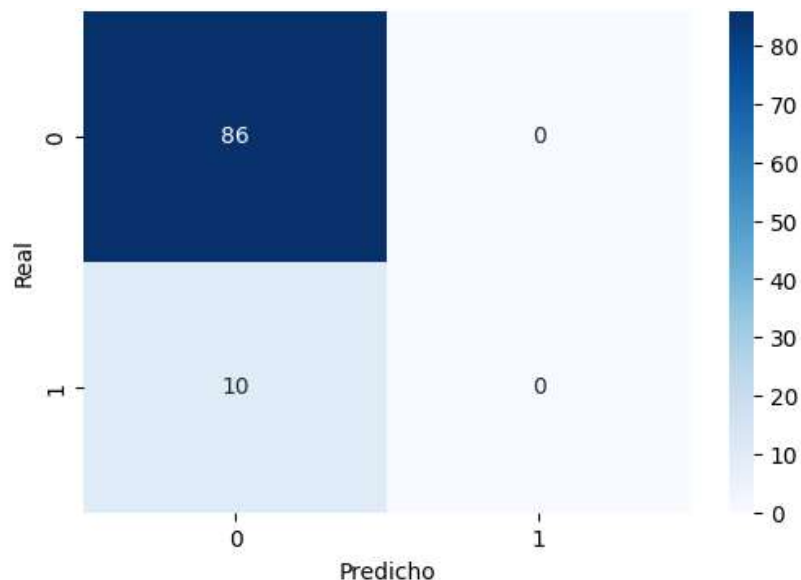


Gráfico 26: Matriz de confusión del modelo predictivo RandomForest

3.2.13. Eficiencia del Modelo *RandomForest*

El modelo *RandomForest* presenta una falencia dado a que esta sesgado hacia la clase 0 (No desertores) y no detecta con precisión la deserción (clase 1). Esto se debe a que hay un desbalance de clases, es decir hay más estudiantes que no desertarán en relación de los que sí. Esto hace que sea necesario probar soluciones como balanceo de clases para lo cual se aplica el modelo predictivo de máquina de soporte de vectores (SVM).

3.2.14. Entrenamiento del Modelo con Máquina de Soporte de Vectores

Dado a que *RandomForest* presenta falencias en la predicción, se recurre como alternativa adicional al entrenamiento del algoritmo predictivo de deserción estudiantil con el modelo de Máquinas de Soporte Vectorial (SVM), mismo que alcanzó una precisión global (accuracy) del 93.06% sobre el conjunto de prueba. Esto evidencia que el modelo clasifica correctamente el **93.06%** de los casos en la muestra de validación. Teniendo en cuenta que la precisión es alta, se puede indicar que el modelo es robusto para distinguir entre estudiantes con alta y baja probabilidad de deserción. En el Gráfico 27, se evidencia la codificación del modelo SVM:

```
# Entrenar SVM con calibración para mejorar la confianza en probabilidades
svm_model = SVC(class_weight="balanced", kernel="rbf", C=10, gamma="scale", probability=True, random_state=42)
calibrated_svm = CalibratedClassifierCV(svm_model, method='sigmoid', cv=5)
calibrated_svm.fit(X_train, y_train)

# Obtener probabilidades y encontrar el umbral óptimo basado en F1-score
y_scores = calibrated_svm.predict_proba(X_test)[:,-1]
thresholds = np.linspace(0.1, 0.9, 50)
f1_scores = [f1_score(y_test, (y_scores >= t).astype(int)) for t in thresholds]
best_threshold = thresholds[np.argmax(f1_scores)]

# Aplicar el nuevo threshold
y_pred_adjusted = (y_scores >= best_threshold).astype(int)

# Evaluar el modelo
accuracy = accuracy_score(y_test, y_pred_adjusted)
print("Accuracy:", accuracy)
print(classification_report(y_test, y_pred_adjusted))
```

Gráfico 27: Codificación del modelado predictivo con SVM

De igual manera, las métricas adicionales como: precisión, recall y F1-score para ambas clases (desertor y no desertor) refuerzan la propuesta de que el modelo es confiable, esto se presenta en el Gráfico 28.

Accuracy: 0.930635838150289				
	precision	recall	f1-score	support
0	0.91	0.95	0.93	85
1	0.95	0.91	0.93	88
accuracy			0.93	173
macro avg	0.93	0.93	0.93	173
weighted avg	0.93	0.93	0.93	173

Gráfico 28: Métricas de evaluación del modelo SVM

A continuación, se realiza un análisis de la evaluación efectuada en cada una de las clases:

Tabla 20: Análisis del modelo predictivo SVM

CLASE	PRECISIÓN	RECALL	F1-SCORE
	0.91	0.95	0.93
Clase 0 (Estudiantes que no desertan)	91% de los estudiantes clasificados como "no desertores" realmente no desertan. Un valor alto indica que el modelo comete pocos falsos positivos, es decir, clasifica correctamente a los estudiantes que continúan con sus estudios	El 95% de los estudiantes que realmente no desertan fueron correctamente identificados por el modelo. Un recall alto representa que el modelo detecta la mayoría de los casos verdaderos de estudiantes que permanecen en la UTC.	Esta métrica combina precisión y recall, mostrando que hay un equilibrio adecuado entre ambos.
	0.95	0.91	0.93
Clase 1 (Estudiantes con alta probabilidad de deserción)	El 95% de los estudiantes clasificados como desertores realmente están en riesgo de abandonar sus estudios. Un valor alto evidencia que el modelo rara vez clasifica erróneamente a estudiantes que en realidad continuarían.	El modelo identifica correctamente el 91% de los estudiantes que efectivamente tienen una alta probabilidad de deserción. Esto demuestra que el modelo omite algunos casos de deserción (falsos negativos), pero en una proporción aceptable.	Indica que la capacidad predictiva del modelo en esta clase es muy fuerte y balanceada.

A diferencia de *RandomForest*, el modelo SVM ha demostrado ser altamente eficaz para predecir la deserción estudiantil, con una precisión del **93.06%** y métricas de rendimiento equilibradas. Su capacidad para identificar tanto estudiantes en riesgo de deserción como aquellos que continuarán sus estudios lo hace una herramienta útil para la intervención temprana esto se refleja en la matriz de confusión.

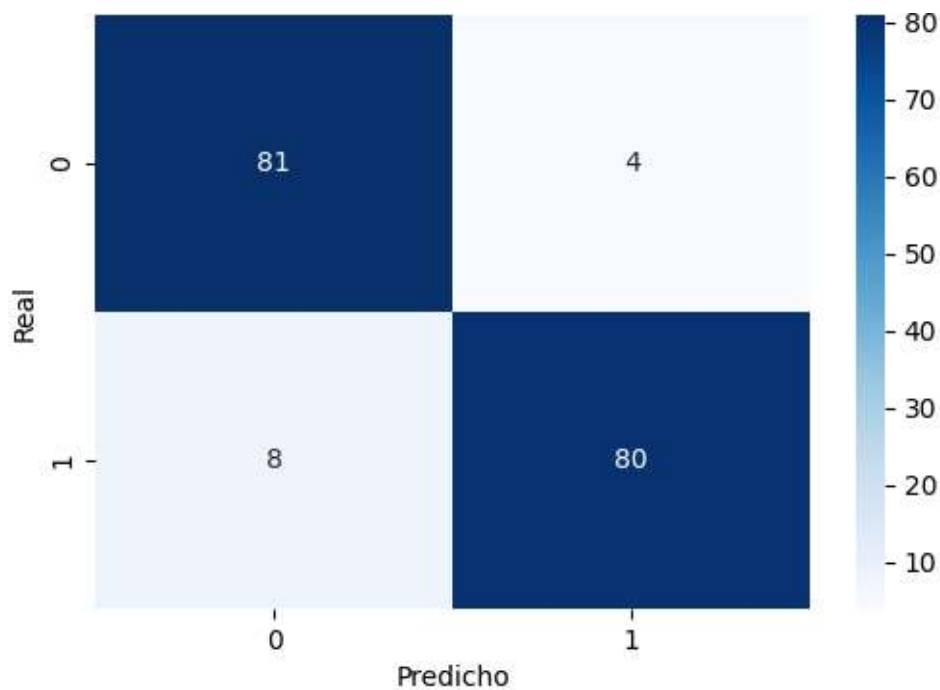


Gráfico 29: Matriz de confusión del modelo predictivo SVM

En el Gráfico 29, se observa 81 verdaderos negativos (TN) que representan casos correctamente clasificados como no desertores, 4 Falsos positivos (FP) que son casos clasificados erróneamente como desertores cuando en realidad no lo son y 8 Falsos negativos (FN) que indican casos clasificados erróneamente como no desertores, cuando en realidad sí tienen alta probabilidad de deserción. Finalmente se tiene 80, verdaderos positivos (TP) que indican casos correctamente clasificados como desertores. La matriz de confusión confirma que el modelo tiene un rendimiento eficiente, con una alta capacidad para predecir tanto estudiantes que continuarán como aquellos que tienen una alta probabilidad de abandonar sus

estudios. El número de errores (FN y FP) es bajo, lo que indica un equilibrio óptimo entre precisión y recall.

La precisión de **93.06%** se obtiene de forma automática gracias al uso de *sklearn.metrics* disponible en la biblioteca *Scikit-Learn* usada en para entrenar el modelo SVM con base a los datos recopilados en la encuesta llenada por los estudiantes, cabe mencionar que estos cálculos se hacen considerando la siguiente formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Donde:

- TP (*True Positives*/Verdaderos Positivos): Casos correctamente clasificados como deserción.
- TN (*True Negatives*/Verdaderos Negativos): Casos correctamente clasificados como permanencia.
- FP (*False Positives*/Falsos Positivos): Casos incorrectamente clasificados como deserción.
- FN (*False Negatives*/Falsos Negativos): Casos incorrectamente clasificados como permanencia.

3.2.15. Influencia de las Variables para el Modelo Predictivo

Con base al análisis realizado las variables más influyentes para predecir la deserción estudiantil de acuerdo con el instrumento diseñado para el efecto son:

- ¿Trabajas mientras estudias?
- ¿Te sientes apoyado/a por tus profesores y la institución?
- ¿Tienes hijos?
- Satisfacción con la carrera (escala del 1 al 5).
- Residencia según la zona de planificación en Ecuador.
- Género.
- Edad.
- Estado civil.

- Autoevaluación del rendimiento académico (escala del 1 al 5).

Sin embargo, la variable más relevante es **¿Trabajas mientras estudias?**, lo que indica que la carga laboral podría influir fuertemente en la deserción ante la necesidad económica de algunos estudiantes. También destaca **la situación familiar** (hijos, estado civil y/o ingresos familiares, etc.), esto se ha comprobado mediante la determinación de la matriz de correlación de variables.

3.2.16. Comprobación de la Hipótesis

La hipótesis propuesta en la investigación establece que, “Si se aplican técnicas de Machine Learning sobre datos académicos, socioeconómicos y demográficos, entonces se podrá predecir la deserción estudiantil con una precisión superior al 90%”, misma que ha sido comprobada mediante la implementación de dos modelos de aprendizaje automático. El primer modelo fue Random Forest que alcanzó una precisión del 89.6%, cercana al umbral establecido, mientras que el modelo de Support Vector Machine (SVM) superó este umbral con una precisión del 93.06%. Estos resultados comprueban que, sí es posible predecir la deserción estudiantil con una precisión superior al 90% utilizando técnicas de Machine Learning supervisado, validando de este modo la hipótesis planteada.

3.2.17. Importancia de Dividir los Datos para Entrenamiento y Pruebas

Los resultados han sido satisfactorios en gran medida a la división de los datos, el 80% se usó para entrenamiento y el 20% para pruebas al momento de aplicar los algoritmos de aprendizaje de máquina utilizados, como se evidencia en el Gráfico 30.

```
# División en entrenamiento y prueba
X_train, Test, y_train, y_test = train_test_split(X_resampled, y_resampled, test_size=0.2,
random_state=42)
```

Gráfico 30: División de datos en entrenamiento y pruebas

3.2.18. Estrategias de Mejora con Base a los Hallazgos Obtenidos

El análisis predictivo realizado permitió identificar patrones en los factores que influyen en la deserción estudiantil. Sin embargo, estos hallazgos solo son valiosos si pueden traducirse en estrategias y acciones concretas para reducir la deserción en la Universidad Técnica de Cotopaxi, por lo que en la Tabla 21 se resume los hallazgos clave y se propone recomendaciones basadas en los resultados.

Tabla 21: Estrategias de mejora para evitar la deserción estudiantil.

Hallazgos	Acciones Recomendadas
<p>Fuerte asociación entre la deserción y respuestas relacionadas con problemas económicos.</p>	<p>Continuar y fortalecer el programa de becas y apoyos económicos:</p> <ul style="list-style-type: none"> • Incrementar los programas de becas parciales o completas dirigidos a estudiantes en riesgo de deserción. • Agilizar el proceso de solicitud de becas y mejorar la difusión de estos programas. <p>Facilidades de Pago:</p> <ul style="list-style-type: none"> • Permitir opciones de pago fraccionado o diferido para segundas y terceras matriculas de estudiantes en dificultades económicas. • Evaluar la posibilidad de subsidios o préstamo de materiales de estudio costosos como computadores, osciloscopios, analizadores de red eléctrica, etc. <p>Programas de Trabajo-Estudio:</p> <ul style="list-style-type: none"> • Firmar convenios con empresas locales para ofrecer oportunidades de empleo con horarios flexibles para estudiantes. • Implementar programas de prácticas preprofesionales remuneradas. <p>Asesoría Financiera:</p> <ul style="list-style-type: none"> • Ofrecer sesiones de educación financiera para que los estudiantes puedan administrar mejor sus recursos y evitar la deserción por falta de planificación económica.

Hallazgos	Acciones Recomendadas
<p>Algunos estudiantes presentan una posible deserción debido a dificultades académicas, desmotivación o falta de conexión con la carrera seleccionada.</p>	<p>Fortalecimiento del Acompañamiento Académico:</p> <ul style="list-style-type: none"> • Mejorar e implementar programas de tutorías personalizadas y mentorías entre estudiantes avanzados y nuevos. • Facilitar recursos de autoaprendizaje y nivelación en materias críticas como cursos MOOC (Curso en Línea Masivo). <p>Flexibilización y Modernización del Currículo:</p> <ul style="list-style-type: none"> • Revisar y actualizar de ser necesario el plan de estudios de las carreras con mayor deserción. • Implementar metodologías de enseñanza más dinámicas, como aprendizaje basado en proyectos y casos reales con el apoyo de inteligencia artificial. <p>Vinculación con el Mercado Laboral y Comunidad:</p> <ul style="list-style-type: none"> • Generar alianzas con instituciones y comunidades locales para que los estudiantes realicen proyectos aplicados y prácticas comunitarias. • Fomentar la participación en voluntariados y proyectos sociales, lo que refuerza la motivación y el sentido de pertenencia.

CONCLUSIONES GENERALES

- La revisión bibliográfica permitió establecer que el Machine Learning es una herramienta efectiva para la predicción de la deserción estudiantil mediante la aplicación de modelos de aprendizaje supervisado se confirmó que algoritmos como Máquinas de Soporte Vectorial (SVM) y Random Forest permiten identificar patrones en variables académicas, socioeconómicas y personales que influyen en la deserción en la Universidad Técnica de Cotopaxi.
- La recolección de datos mediante un formulario de encuesta en ambiente web viabilizó la implementación de técnicas de preprocesamiento de datos, balanceo y selección de características relevantes, junto con el uso de herramientas como Python, Scikit-Learn y Google Colab, se consiguió desarrollar un sistema capaz de predecir la deserción con una confiabilidad del 93.06%, lo que se verificó mediante métricas como precisión, F1-score y curva ROC.
- El Machine Learning aplicado a la educación superior permite tomar decisiones informadas y diseñar estrategias preventivas, los modelos empleados fueron *Random Forest* y SVM siendo este último el más eficiente para predecir la deserción estudiantil, sin embargo, la correcta interpretación de los resultados y sobre todo su integración con los sistemas de gestión académica podrán contribuir significativamente a la reducción de la deserción estudiantil en la UTC.
- La evaluación del desempeño determinó que el modelo predictivo desarrollado con el algoritmo de Máquina de Vector de Soporte demuestra ser más eficiente por sobre Random Forest, esto se comprobó mediante la matriz de confusión, cálculo de Accuracy y la validación de los modelos con datos de entrenamiento (80%) y prueba (20%) consiguiendo así un apoyo tecnológico importante para la toma de decisiones relacionadas con la deserción estudiantil.

RECOMENDACIONES

- Revisar nueva literatura científica relacionada con la integración de modelos de IA en la gestión académica universitaria para proponer nuevos enfoques que permitan monitorear en tiempo real a los estudiantes en riesgo de deserción. Esto permitirá una intervención temprana mediante tutorías, asesoramiento académico y apoyo socioeconómico, reduciendo así la tasa de abandono.
- Fortalecer la recolección y almacenamiento de datos mediante la creación de una base de datos estructurada y actualizada que permita la implementación de sistemas automatizados de seguimiento estudiantil para garantizar información precisa y completa, permitiendo mejorar la capacidad predictiva del modelo.
- Explorar la aplicación de Machine Learning en otras áreas de la universidad, como la predicción del rendimiento académico, la personalización del aprendizaje y la optimización de recursos institucionales, así como también probar con la implementación de nuevas técnicas de inteligencia artificial.
- Considerar los resultados obtenidos para diseñar estrategias institucionales que permitan intervenir a tiempo y reducir la deserción estudiantil teniendo en cuenta los hallazgos de la presente investigación para mejorar las políticas de apoyo académico, financiero y psicológico que beneficien a los estudiantes de la UTC en riesgo de abandono escolar.

BIBLIOGRAFÍA

- [1] D. Gutiérrez, J. Vélez y J. López, «Indicadores de deserción universitaria y factores asociados,» *EducaT*, 2021.
- [2] X. Erazo y E. Rosero, «Orientación vocacional y su influencia en la deserción universitaria,» *Horizontes*, 2021.
- [3] N. De la Cruz, G. Santana, G. Arteaga y K. Toala, «El factor socio económico y su repercusión en el sistema de ingreso en la Universidad Técnica de Manabí, Ecuador,» *Investigar*, 2024.
- [4] A. Chancusi, R. Logroño y L. López, «La intención de abandono universitario en los estudiantes de la Universidad Técnica de Cotopaxi,» *PentaCiencias*, vol. 4, nº 1, 2022.
- [5] F. G. Olaniyi, «Predicting Student Performance in Higher Educational Institutions Using Video Learning Analytics and Data Mining Techniques,» *Oluwaseun Oladeji Olaniyi, Chinasa Susan Adigwe, Anthony Abalaka, Nishant Hemantkumar Shah*, p. 9, 2024.
- [6] O. O. Olaniyi, O. J. Olalekan y S. Oladiipo Olabanji, «Advancing Data-Driven Decision Making in Smart Cities through Big Data Analytics: A Comprehensive Review of Existing Literature,» *.Current Journal of Applied Science and Techn*, vol. 42, nº 25, pp. 10-18, 2023.
- [7] W. R. Balladares Jalinás, L. R. Bermúdez Jalinás y L. G. C. Manuel, «Repositorio Institucional de la UNAN-Managua,» 10 Febrero 2021. [En línea]. Available: <https://repositorio.unan.edu.ni/id/eprint/14794/1/14794.pdf>. [Último acceso: 20 Noviembre 2024].
- [8] M. M. Banda Casa, A. Serrano Bonifaz y F. S. Mendoza Moreira, «Estrategias institucionales frente a la deserción escolar del estudiantado: una revisión sistemática,» *Revista Minerva*, vol. 5, nº 8, pp. 124-149, 22 Mayo 2024.
- [9] R. H. Alejandro, «Factores asociados a la prevención del abandono escolar: una mirada desde la implementación del programa tutorías pedagógicas,» *Sophia Austral*, vol. 28:8, nº 21, pp. 1-21, 2022.
- [10] F. Ruffinatto, L. G. Pereno y L. Juaneu, «Factores Institucionales: Deserción Estudiantil en Psicología de la Universidad Nacional de Córdoba,» *The International Education and Learning Review / Revista Internacional de Educación y Aprendizaje*, vol. 10, nº 2, pp. 137-152, 2022.
- [11] E. A. Cedeño Castillo, V. I. Eras Briones, D. M. Meza Arguello y M. V. Sigcho Ocampo, «Factores que inciden en la deserción escolar en tiempos de pandemia por el Covid-19: Perspectivas de docentes de instituciones

- públicas y privadas,» *Mundo Recursivo Revista Científica*, vol. 6, nº 1, pp. 84-98, Enero 2023.
- [12] J. A. González García, S. L. Canchola Magdaleno y R. L. Ulloa Cazarez, «Enfoques, estudios y perspectivas teóricas sobre la deserción escolar en la educación superior,» *Revista de Investigación Educativa de la REDIECH*, vol. 15, p. e1959, 2024.
- [13] S. Chávez, F. Hernández Reyes y M. E. Piña Gutiérrez, « Causas e implicaciones psicológicas, sociales y educativas de la deserción escolar,» *PsicoEducativa: reflexiones y propuestas*, vol. 6, nº 11, 2020.
- [14] K. Leon Cortes, *Prácticas de integración universitaria para la reducción del abandono (Las tutorías-mentorías)*, COLOMBIA: Universidad del Rosario,, 2020.
- [15] E. M. Rojas, «Tipos de aprendizaje de Machine Learning,» de *Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo*, México, Revista Ibérica de Sistemas e Tecnologías de Informação, 2020, p. 586–599.
- [16] A. Osorio y N. Enerieth, «El derecho de autor en la Inteligencia Artificial de machine learning (Copyright Law in the Artificial Intelligence of Machine Learning),» nº 30, 2020.
- [17] J. Bobadilla, *Machine Learning y Deep Learning: Usando Python, Scikit y Keras*, Ediciones de la U, 2021.
- [18] E. E. Pina, *Aprendizaje por Refuerzo mediante Deep Learning para las Ciudades Inteligentes*, Escuela Técnica Superior de Ingenieros Informáticos, 2021.
- [19] Y. H. Xiaonan Zou y K. S. Zhewen Tian, 7th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China: Wuhan University of Technology, Hubei Key Laboratory of Modern Auto Parts Technology, Wuhan, China, 2020.
- [20] H. Garzón Saénz, P. G. Ortiz y J. Garzón, *Características del hogar y pobreza: una aplicación de las máquinas de soporte vectorial*, Colombia: Fundación Universitaria Tecnológico Comfenalco (Colombia)*, 2023.
- [21] S. Guerra, J. Rave, M. Correa y M. Toro, *APLICACIONES CON ÁRBOLES DE DECISIÓN Y BOSQUES ALEATORIOS EN LA PREDICCIÓN DEL*, Colombia: Universidad Eafit, 2019.
- [22] A. F. Agundis Martínez, *Evaluación de algoritmo de Aprendizaje Automático mediante Bosques Aleatorios para el seguimiento de variables críticas en un proceso de manufactura*, NUEVO LEÓN: Tecnológico Nacional de México, 2023.
- [23] A. Burkov, *The Hundred-Page Machine Learning Book*, Leanpub, 2019.

- [24] A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow*, O'Reilly Media, Inc., 2019.
- [25] T. Fernandez Martín, M. Solis Salazar, M. T. Hernandez Jimenez y T. E. Moreira Mora, «Un análisis multinomial y predictivo de los factores asociados a la deserción universitaria.,» *Educare*, vol. 23, n° 1, pp. 73-97, 2019.
- [26] M. Cruz, M. González y J. C. Rangel, «Técnicas de machine learning aplicadas a la evaluación del rendimiento y a la predicción de la deserción de estudiantes universitarios, una revisión.,» *Prisma Tecnológico*, vol. 1, n° 13, pp. 77-88, 2022.
- [27] H. E. Caselli Gismondi, «Repositorio UNS,» 2021. [En línea]. Available: <https://repositorio.uns.edu.pe/handle/20.500.14278/3804>. [Último acceso: 2024 11 20].
- [28] A. Valera Blanco, «uvadoc,» 2023. [En línea]. Available: <https://uvadoc.uva.es/handle/10324/63431>. [Último acceso: 2024 11 21].
- [29] M. Cobo Carrillo, «Ucrea,» 2023 09 07. [En línea]. Available: <https://repositorio.unican.es/xmlui/bitstream/handle/10902/29819/CoboCarrilloMarcos.pdf?sequence=1&isAllowed=y>. [Último acceso: 2024 11 20].
- [30] D. Lopez Ramos y L. Arco Garcia, «Aprendizaje profundo para la extracción de aspectos en opiniones textuales,» *Revista Cubana de Ciencias Informáticas*, vol. 2, n° 13, pp. 105-145, 2019.
- [31] P. M. Geovanna López, «datalat,» 12 2023. [En línea]. Available: <https://al-dato.datalat.org/wp-content/uploads/2024/01/2.-Geovanna-Lopez-Pedro-Manosalva-Criminalidad-AI-Dato.pdf>. [Último acceso: 2024 11 20].
- [32] C. A. Ponte Ramirez, «Repositorio de la Universidad César Vallejo,» 2024. [En línea]. Available: <https://repositorio.ucv.edu.pe/handle/20.500.12692/150161>. [Último acceso: 21 11 2024].
- [33] M. L. González Acosta y D. I. Rodríguez Rodríguez, «Redaly,» 05 Febrero 2020. [En línea]. Available: <https://www.redalyc.org/articulo.oa?id=563662155013>. [Último acceso: 20 Noviembre 2024].
- [34] J. M. Norambuena, M. G. Badilla Quintana y Y. Angulo Lopez, «Redaly,» 02 Febrero 2022. [En línea]. Available: <https://www.redalyc.org/journal/5771/577170677011/html/>. [Último acceso: 20 Noviembre 2024].
- [35] L. Cedeño Valarezo, J. Morales Carillo, C. P. Quijije Vera y S. A. Palau Delgado, «ResearchGate,» Julio 2022. [En línea]. Available: https://www.researchgate.net/publication/362297220_Modelos_predicti

vos_aplicados_en_la_educacion_Casos_abandono_de_estudio. [Último acceso: 20 Noviembre 2024].

- [36] E. Bautista Villegas, «Metodologías ágiles XP y Scrum, empleadas para el desarrollo de páginas web, bajo MVC, con lenguaje PHP y framework Laravel,» *Revista Amazonía Digital 1.1*, vol. 1, n° 1, 2022.
- [37] L. M. Armijos Ortega y E. L. Lojan Cueva, «Estudio de la adopción de metodologías ágiles en proyectos de desarrollo de software en la región 7 del Ecuador,» *Revista Espacios*, vol. 45, n° 4, 2024.
- [38] E. M. Rojas, «Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo - ProQuest,» *Revista Ibérica de Sistemas e Tecnologías de Informação*, n° E28, pp. 586-599, 2020.
- [39] D. F. Barrera, C. D. Duran y S. Palmet, «Aplicación de árboles de decisión para la identificación de adaptabilidad de estudiantes en educación online,» *Revista Innovación y Software*, vol. 4, n° 2, pp. 166-181, 2023.
- [40] D. González Rey, «Análisis de datos y desarrollo de modelo predictivo en el sector energético,» 2024. [En línea]. Available: https://dspace.unia.es/bitstream/handle/10334/8734/1610_González.pdf?sequence=1&isAllowed=y. [Último acceso: 20 Noviembre 2024].
- [41] L. F. Osorio Muñoz y M. N. Suntasig Jaen, «Universidad Técnica de Cotopaxi - Repositorio,» Agosto 2024. [En línea]. Available: <https://repositorio.utc.edu.ec/server/api/core/bitstreams/c4f4536d-edda-4909-bf4e-bdeca82b425a/content>. [Último acceso: 20 Noviembre 2024].
- [42] E. Martín Calvo, «Universidad de Valladolid - Repositorio Documental,» 2024. [En línea]. Available: <https://uvadoc.uva.es/handle/10324/71428>. [Último acceso: 20 Noviembre 2024].
- [43] Universidad Técnica de Cotopaxi, «utc.edu.ec,» 2024. [En línea]. Available: <http://www.utc.edu.ec/GRADO/oferta-grado1>.
- [44] A. Haro, E. Chisag y J. Ruiz, «Tipos y clasificación de las investigaciones,» *Revista Latino Americana de Ciencias Sociales y Humanidades*, 2024.
- [45] C. Arbulú, «Definición de método de investigación inductivo.,» 2023.
- [46] T. Blanchar y N. Martínez, «¿Entrevista o encuesta? Una diferencia necesaria,» *Revista Latina de Comunicación Social*, 2024.
- [47] J. Valencia y O. Carmenates, «Validación de un instrumento para su implementación en el proceso de capacitación comunitaria,» *REMCA*, 2023.
- [48] B. Cuji, W. Gavilanes y R. Sanchez, «Modelo predictivo de deserción estudiantil basado en arboles de decisión,» *Espacios*, 2017.

- [49] F. Montalvo y N. Sánchez, «Modelo de predicción de deserción: Un estudio de caso de estudiantes de la Universidad Politécnica Estatal del Carchi,» *Prometeo*, 2023.
- [50] A. Aco, B. Hanco y Y. Pérez, «Análisis comparativo de Técnicas de Machine Learning para la predicción de casos de deserción universitaria,» *RISTI*, vol. 51, 2023.

ANEXOS

Anexo A: Cuestionario sobre la Deserción Estudiantil

Objetivo: Recopilar información relevante sobre las características demográficas, socioeconómicas, académicas y personales de los estudiantes de la Universidad Técnica de Cotopaxi, con el fin de identificar los factores que influyen en la deserción estudiantil.

Confidencialidad: Todas las respuestas son anónimas y se utilizarán únicamente con fines académicos. La privacidad será respetada en todo momento.

Instrucciones: Lea cada pregunta cuidadosamente y seleccione la respuesta que mejor refleje su situación actual.

Sección 1: Datos Demográficos

1. Edad:

() Menos de 18

() 18- 24

() 25 – 30

() Más de 30

2. Género

() Masculino

() Femenino

3. Estado Civil:

() Soltero/a

() Casado/a

() Unión Libre

() Divorciado/a

() Viudo/a

4. ¿Tienes hijos?

() Sí

() No

5. ¿Tienes alguna discapacidad?

() Sí

() No

6. Etnia

- () Afroecuatoriana
- () Indígena
- () Mestizo
- () Blanco
- () Montubio

Sección 2: Datos Socioeconómicos

7. Residencia según la zona de planificación en Ecuador

- () Esmeraldas, Imbabura, Carchi, Sucumbíos
- () Pichincha, Napo, Orellana
- () Cotopaxi, Tungurahua, Chimborazo, Pastaza
- () Manabí, Santo Domingo
- () Santa Elena, Guayas, Bolívar, Los Ríos, Galápagos
- () Cañar, Azuay, Morona Santiago
- () El Oro, Loja, Zamora Chinchipe
- () Extranjeros

8. Nivel de formación académica del padre

- () Ninguno, jardín de infantes
- () Primaria, secundaria, básica
- () Media, universidad incompleta
- () Tercer nivel
- () Cuarto nivel

9. Nivel de formación académica de la madre

- () Ninguno, jardín de infantes
- () Primaria, secundaria, básica
- () Media, universidad incompleta
- () Tercer nivel
- () Cuarto nivel

10. Tipo de colegio que realizó sus estudios

- () Privado
- () Fiscomisional
- () Público

11. ¿Cuál es el nivel de ingreso mensual de tu hogar?

- Menos de \$500
- \$500 - \$1000
- \$1000 - \$1500
- Más de \$1500

12. ¿Con quién vives actualmente?

- Solo/a
- Con familia
- Con pareja e hijos
- Con amigos
- En residencia estudiantil

13. ¿Vives en un departamento alquilado?

- Sí
- No

14. ¿Cuál es la distancia promedio que recorres para llegar a la universidad?

- Menos de 5 km
- 5-10 km
- 11-20 km
- Más de 20 km

15. ¿Trabajas mientras estudias?

- Sí
- No

Sección 3: Preferencias de Carrera

16. Ciclo en que estas matriculado actualmente

- Nivelación
- Primero
- Segundo
- Tercero
- Cuarto
- Quinto
- Sexto
- Séptimo

() Octavo

17. ¿Por qué elegiste tu carrera actual? (Selecciona todas las que apliquen)

() Interés personal

() Influencia familiar

() Oportunidades laborales

() Disponibilidad de cupos

18. En una escala del 1 al 5, ¿qué tan satisfecho/a estás con tu elección de carrera?

1 - Muy insatisfecho/a

2 - Insatisfecho/a

3 - Neutral

4 - Satisfecho/a

5 - Muy satisfecho/a

Sección 4: Experiencia Académica

19. En una escala del 1 al 5, ¿cómo evaluarías tu rendimiento académico hasta ahora?

1 - Muy bajo

2 - Bajo

3 - Promedio

4 - Alto

5 - Muy alto

20. ¿Te sientes apoyado/a por tus profesores y la institución?

1 - Muy en desacuerdo

2 - En desacuerdo

3 - Neutral

4 - De acuerdo

5 - Muy de acuerdo

Sección 5: Bienestar y Motivación

21. ¿Cómo calificarías tu situación alimentaria actual?

1 - Muy mala

2 - Mala

3 - Aceptable

4 - Buena

5 - Muy buena

22. En una escala del 1 al 5, ¿qué tan probable es que consideres abandonar tus estudios en el próximo año?

1 - Muy improbable

2 - Improbable

3 - Neutral

4 - Probable

5 - Muy probable

23. ¿Cuáles son las principales razones por las que podrías considerar dejar la universidad? (Selecciona todas las que apliquen)

() Problemas financieros

() Falta de interés en la carrera

() Problemas personales

() Dificultades académicas

() Carga laboral

() Responsabilidades familiares

() Distancia y transporte

Anexo B: Codificación del Análisis Predictivo

```
# Cargar librerías
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix,
roc_curve, f1_score
from sklearn.utils import resample
from imblearn.over_sampling import SMOTE

# Cargar archivo CSV en Google Colab
file_path = "reporte1.csv" # Ajustar la ruta en Colab
df = pd.read_csv(file_path)

# Manejo de duplicados
df_clean = df.groupby(["usuario_id", "pregunta_texto"])["opcion_texto"].agg(lambda x:
x.mode()[0]).reset_index()

# Pivotear la tabla
df_pivot = df_clean.pivot(index="usuario_id", columns="pregunta_texto",
values="opcion_texto").reset_index()

# Crear la variable objetivo
df_pivot["desercion"] = df_pivot["En una escala del 1 al 5, ¿Qué tan probable es que
consideres abandonar tus estudios?"].map(
    {"1. Muy improbable": 0, "2. Improbable": 0, "3. Neutral": 0, "4. Probable": 1, "5. Muy
probable": 1}
)
df_pivot.drop(columns=["En una escala del 1 al 5, ¿Qué tan probable es que consideres
abandonar tus estudios?"], inplace=True)

# Convertir variables categóricas en numéricas
encoder = LabelEncoder()
for col in df_pivot.select_dtypes(include=['object']).columns:
    df_pivot[col] = encoder.fit_transform(df_pivot[col].astype(str))

# Balancear clases con SMOTE
X = df_pivot.drop(columns=["desercion", "usuario_id"])
y = df_pivot["desercion"]
smote = SMOTE(random_state=42)
X_resampled, y_resampled = smote.fit_resample(X, y)
```

```

# División en entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X_resampled, y_resampled, test_size=0.2,
random_state=42)

# Estandarización
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Entrenar SVM con probabilidades habilitadas
svm_model = SVC(class_weight="balanced", kernel="rbf", C=10, gamma="scale",
probability=True, random_state=42)
svm_model.fit(X_train, y_train)

# Obtener probabilidades de predicción
y_scores = svm_model.predict_proba(X_test)[:, 1]

# Determinar el umbral óptimo usando F1-score
thresholds = np.linspace(0.1, 0.9, 50)
best_f1 = 0
optimal_threshold = 0.5
for threshold in thresholds:
    y_pred_temp = (y_scores >= threshold).astype(int)
    f1 = f1_score(y_test, y_pred_temp)
    if f1 > best_f1:
        best_f1 = f1
        optimal_threshold = threshold

# Aplicar el nuevo threshold
y_pred_adjusted = (y_scores >= optimal_threshold).astype(int)

# Evaluar el modelo con el nuevo threshold
accuracy = accuracy_score(y_test, y_pred_adjusted)
print("Accuracy:", accuracy)
print(classification_report(y_test, y_pred_adjusted))

# Matriz de confusión ajustada
plt.figure(figsize=(6,4))
sns.heatmap(confusion_matrix(y_test, y_pred_adjusted), annot=True, fmt="d",
cmap="Blues")
plt.xlabel("Predicho")
plt.ylabel("Real")
plt.title("Matriz de Confusión Ajustada")
plt.show()

```

```
# Graficar la curva ROC
fpr, tpr, _ = roc_curve(y_test, y_scores)
plt.figure(figsize=(8,6))
plt.plot(fpr, tpr, label=f'ROC curve (AUC = {np.trapz(tpr, fpr):.2f})')
plt.scatter(fpr[np.argmax(tpr - fpr)], tpr[np.argmax(tpr - fpr)], marker='o', color='red',
label=f'Optimal Threshold = {optimal_threshold:.2f}')
plt.xlabel("False Positive Rate")
plt.ylabel("True Positive Rate")
plt.title("Curva ROC y Threshold Óptimo")
plt.legend()
plt.show()
```

Anexo C: Código de la Matriz de Confusión

```
# Cargar librerías
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder

# Cargar archivo CSV
file_path = "reporte1.csv" # Ajustar la ruta en Colab si es necesario
df = pd.read_csv(file_path)

# Manejo de duplicados
df_clean = df.groupby(["usuario_id", "pregunta_texto"])[ "opcion_texto"].agg(lambda x:
x.mode()[0]).reset_index()

# Pivotear la tabla
df_pivot = df_clean.pivot(index="usuario_id", columns="pregunta_texto",
values="opcion_texto").reset_index()

# Crear la variable objetivo
df_pivot["desercion"] = df_pivot["En una escala del 1 al 5, ¿Qué tan probable es que
consideres abandonar tus estudios?"].map(
    {"1. Muy improbable": 0, "2. Improbable": 0, "3. Neutral": 0, "4. Probable": 1, "5. Muy
probable": 1}
)
df_pivot.drop(columns=["En una escala del 1 al 5, ¿Qué tan probable es que consideres
abandonar tus estudios?"], inplace=True)

# Convertir variables categóricas en numéricas
encoder = LabelEncoder()
for col in df_pivot.select_dtypes(include=['object']).columns:
    df_pivot[col] = encoder.fit_transform(df_pivot[col].astype(str))

# Generar un diccionario con nomenclatura tipo "P1", "P2", etc.
question_columns = [col for col in df_pivot.columns if col not in ["usuario_id", "desercion"]]
question_abbreviations = {col: f"P{i+1}" for i, col in enumerate(question_columns)}
question_abbreviations["desercion"] = "desercion"

# Renombrar las columnas en el DataFrame
df_pivot_renamed =
df_pivot.rename(columns=question_abbreviations).drop(columns=["usuario_id"])

# Calcular la matriz de correlación
```

```
correlation_matrix_renamed = df_pivot_renamed.corr()

# Graficar la matriz de correlación
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix_renamed, annot=True, cmap="coolwarm", fmt=".2f",
            linewidths=0.5)
plt.title("Matriz de Correlación de Variables")
plt.xticks(rotation=45, ha='right')
plt.yticks(rotation=0)
plt.show()

# Mostrar la nomenclatura utilizada
question_abbreviations
```

Anexo D: Hoja de Vida del Investigador



VERA MACHUCA CARLOS ALBERTO

DATOS PERSONALES.

PROFESIÓN : Ingeniero en Informática y Sistemas Computacionales

FECHA DE NACIMIENTO : 22 – Agosto - 1992

EDAD : 32 años

ESTADO CIVIL : Unión Libre

CEDULA DE IDENTIDAD : 120548683-8

NACIONALIDAD : Ecuatoriana

ETNIA : Mestiza

TIPO DE SANGRE : RH+

E-mail : carlos.vera8@utc.edu.ec

DIRECCIÓN : Av. San Pablo y Chile

TELÉFONOS : 0988021083

ESTUDIOS REALIZADOS.

INSTRUCCIÓN PRIMARIA: Unidad Educativa Narciso Cerda Maldonado

INSTRUCCIÓN SECUNDARIA: Instituto Tecnológico Superior La Maná

INSTRUCCIÓN SUPERIOR: Universidad Técnica de Cotopaxi

TÍTULOS OBTENIDOS.

- Técnico En Comercio y Administración Especialización Aplicaciones Informáticas
- Ingeniero En Informática y Sistemas Computacionales