



**UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE GUAYAQUIL
CARRERA DE COMPUTACIÓN**

**ANÁLISIS DE CAUSAS DEL ABANDONO ESTUDIANTIL A TRAVÉS DE
TÉCNICAS DE MACHINE LEARNING EN LA UPS SEDE GUAYAQUIL**

Trabajo de titulación previo a la obtención del
Título de Ingeniería en Ciencias de la Computación

AUTOR: VICENTE ANTONIO CHIRIGUAYA MALDONADO

TUTOR: GALO ENRIQUE VALVERDE LANDIVAR

Guayaquil – Ecuador

2024

**CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE
TITULACIÓN**

Yo, Vicente Antonio Chiriguaya Maldonado con documento de identificación N° 0928947670 manifiesto que:

Soy el autor y responsable del presente trabajo; y, autorizo a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Guayaquil, 2 de mayo de 2024

Atentamente,



Vicente Antonio Chiriguaya Maldonado

0928947670

**CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE
TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA**

Yo, Vicente Antonio Chiriguaya Maldonado con documento de identificación No. 0928947670, expreso mi voluntad y por medio del presente documento cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor(a) del Artículo Académico: “Análisis de causas de abandono estudiantil a través de técnicas de Machine Learning en la UPS sede Guayaquil”, el cual ha sido desarrollado para optar por el título de: Ingeniero en Computación, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribo este documento en el momento que hago la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Guayaquil, 2 de mayo de 2024

Atentamente,



Vicente Antonio Chiriguaya Maldonado

0928947670

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Galo Enrique Valverde Landivar con documento de identificación N° 0912511532, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: ANÁLISIS DE CAUSAS DE ABANDONO ESTUDIANTIL A TRAVÉS DE TÉCNICAS DE MACHINE LEARNING EN LA UPS SEDE GUAYAQUIL, realizado por Vicente Antonio Chiriguaya Maldonado con documento de identificación N° 0928947670, obteniendo como resultado final el trabajo de titulación bajo la opción Artículo Académico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Guayaquil, 2 de mayo de 2024

Atentamente,

A handwritten signature in black ink, appearing to read 'Galo Valverde Landivar', is written over a horizontal line.

Msc. Galo Enrique Valverde Landivar

0912511532

DEDICATORIA

Dedico este trabajo a mis padres Andrea Maldonado y Pedro Chiriguaya, sin ustedes no hubiera alcanzado este logro; a mi mejor amigo Rolando Lema, quien hoy descansa en paz, pero sé que celebra junto a mí este gran triunfo. A todas las personas que me han inspirado, mis maestros, mentores y amigos.

Dedico este trabajo también a aquella persona que hace trece años dudaba de sus capacidades, pero hoy lo ha conseguido, me dedico este trabajo a mí... gracias por no haberte rendido a pesar de que el mundo se viniera encima.

AGRADECIMIENTO

Agradezco a Dios sobre todas las cosas, por guiarme y brindarme fortaleza. Agradezco a mis padres, Andrea Maldonado y Pedro Chiriguaya, por su constante paciencia y amor, sin ellos no podría haber llegado tan lejos. Agradezco a mi mejor amigo Rolando Lema, que hoy descansa en santa paz, por sus constantes consejos y apoyo. A mi tío Vicente Chiriguaya Guevara, por su apoyo constante. A mis amigos y compañeros de curso Arianna Sánchez, Alejandro Chucuyan, Shirley Santorum, Ramiro Oñate y Daniel Valencia, por su infinito soporte, a mis amigas Denisse Sánchez y Gabriela Noroña, por nunca dejarme caer e impulsarme a seguir incluso cuando he querido tirar la toalla. Agradezco a mi amigo e investigador Msc. Jeffrey Vargas por haberme inspirado a continuar mis estudios y brindarme sus consejos para este trabajo, a mi tutor Msc. Galo Valverde por su paciencia y apoyo, y a la Universidad Politécnica Salesiana por permitirme presentar este trabajo de investigación.

RESUMEN

Este estudio investigó las causas del abandono estudiantil en la Universidad Politécnica Salesiana sede Guayaquil, utilizando técnicas de aprendizaje automático para analizar datos recopilados a través de encuestas. Se identificaron factores claves como el género, el nivel académico, la percepción de la carrera, la motivación, las dificultades económicas y la salud mental, mismos significativos a la hora de predecir el abandono estudiantil. El algoritmo de Bosques Aleatorios mostró ser muy efectivo en este análisis. A pesar de existir limitaciones que se relacionan con el tamaño de la muestra, este estudio aporta información valiosa para el desarrollo de políticas de intervención enfocadas en mejorar la retención estudiantil. Se hace énfasis en la importancia de abordar el bienestar económico y psicológico de los estudiantes y se sugiere la necesidad de futuras investigaciones para explorar más a fondo los factores que se han identificado, y con ello desarrollar estrategias efectivas para prevenir el abandono estudiantil.

Palabras claves: Abandono estudiantil, aprendizaje automático, Bosques Aleatorios, rendimiento académico, salud mental y educación.

ABSTRACT

This study investigated the causes of student dropout at the Universidad Politécnica Salesiana in Guayaquil, using machine learning techniques to analyze survey data. Key factors such as gender, academic level, career perception, motivation, economic difficulties, and mental health were identified as significant in predicting student dropout. The Random Forest algorithm proved to be very effective in this analysis. Despite limitations related to the sample size, this study provides valuable information for developing intervention policies focused on improving student retention. The importance of addressing the economic and psychological well-being of students is emphasized, and future research is suggested to further explore the identified factors and develop effective strategies to prevent student dropout.

Keywords: Student dropout, machine learning, Random Forest, academic performance, mental health and education.

ÍNDICE DE CONTENIDO

1. INTRODUCCIÓN	11
2. REVISIÓN DE LITERATURA	12
2.1. Abandono estudiantil.....	12
2.2. Machine Learning	13
3. METODOLOGÍA	15
6.1 Métodos y técnicas de Recopilación de datos empleadas	15
6.2 Métodos y técnicas de Análisis de datos	17
4. RESULTADOS.....	19
7.1 Resultados de la encuesta.....	19
7.2 Resultado del análisis	27
5. DISCUSIÓN	32
6. CONCLUSIÓN.....	34
7. REFERENCIAS.....	35

ÍNDICE DE ILUSTRACIONES

Ilustración 1. Metodología propuesta.....	15
Ilustración 2. Resultados de la pregunta ¿Cómo te identificas?.....	19
Ilustración 3. Resultados de la pregunta ¿Cuál es tu sede?	19
Ilustración 4. Resultados de la pregunta ¿Cuál es tu carrera?.....	20
Ilustración 5. Resultados de la pregunta ¿En qué semestres estás?	20
Ilustración 6. Resultados de la pregunta ¿Dónde vives?	21
Ilustración 7. Resultados Nivel Económico.	21
Ilustración 8. Respuestas de la pregunta ¿Cuál es tu estado civil?	22
Ilustración 9. Número de hijos	22
Ilustración 10. Respuestas de la pregunta ¿Actualmente te encuentras trabajando?	23

Ilustración 11. Respuestas de la pregunta ¿Quién paga tus estudios?.....	23
Ilustración 12. Respuestas a la pregunta ¿Cómo consideras tu promedio?.....	24
Ilustración 13. Respuestas a la pregunta ¿Consideras que tienes dificultades para pagar tus estudios?.....	24
Ilustración 14. Nivel económico vs Dificultad de Pago.....	25
Ilustración 15. Respuestas a la pregunta ¿Consideras mejor la presencialidad o la virtualidad?	25
Ilustración 16. Resultado de Preguntas	26
Ilustración 17. Respuestas a la pregunta ¿Qué tan probable es que abandones tus estudios? .	27
Ilustración 18. Matriz de correlación - Causas Académicas	28
Ilustración 19. Matriz de Correlación - Causas Personales.....	29
Ilustración 20. Matriz Correlación - Causas Socioeconómicas.....	29
Ilustración 21. Matriz de Correlación - Causas Psicológicas.....	30

ÍNDICE DE TABLAS

Tabla 1. Factores seleccionados	15
Tabla 2. Preguntas de la encuesta.....	16
Tabla 3. Definición de variables y medidas	17
Tabla 4. Tabla de resultados de preguntas	26
Tabla 5. Rendimiento de Algoritmos de Machine Learning.....	31

1. INTRODUCCIÓN

El abandono estudiantil representa uno de los principales problemas a los que enfrenta la educación en todo el mundo. Sus causas varían y dependen de factores particulares derivados, incluso, de la realidad social por la que atraviesa una zona específica de un país, siendo tal que puede variar de institución en institución, por ello la importancia de focalizar y estudiar casos específicos.

Las consecuencias del abandono estudiantil representan un problema no solo para las personas afectadas, sino para la sociedad en general. En el caso de un estudiante que abandona su carrera, este se ve perjudicado por la pérdida de competitividad en un mercado cada vez más cualificado donde ya es casi una norma el tener por lo menos un título de tercer nivel. Para la sociedad también representa una pérdida ya que esta persona podría llegar a agrandar las cifras de desempleo o subempleo al no tener acceso a mejores oportunidades propias de un grado.

Para conocer y estudiar estas causas se puede hacer uso de técnicas de aprendizaje automático (*Machine Learning*) el cual se ha convertido en una herramienta indispensable no sólo para crear predicciones certeras, sino también para identificar patrones complejos.

El presente estudio buscará brindar un panorama más claro respecto a las causas de abandono estudiantil haciendo uso de algoritmos de aprendizaje automático, con el fin de aportar un modelo que ayude a predecir este fenómeno, así como brindar un panorama más claro respecto a la situación basado en los datos obtenidos de una encuesta realizada a 100 estudiantes de la Universidad Politécnica Salesiana (UPS) sede Guayaquil.

2. REVISIÓN DE LITERATURA

2.1. Abandono estudiantil

A nivel mundial el abandono estudiantil es abordado como un problema para distintos centros educativos. En un estudio realizado en la Facultad de Ingeniería Eléctrica y de Computadoras de la Universidad de Pristina, se destaca el impacto negativo en la sociedad del abandono estudiantil. (Kabashi et al., 2022)

Por otro lado, Casanova et al. (2018) indica que el abandono estudiantil es multifacético y puede tener consecuencias no solo de forma individual, sino también a nivel social, esto dado que de forma individual los estudiantes pueden enfrentarse a falta de oportunidades a la hora de conseguir empleo, así como el estar limitados a desarrollarse profesionalmente, lo cual deriva en una calidad de vida inferior a la de una persona con un título de grado. Mientras que a nivel social limita la cantidad de individuos calificados en la fuerza laboral, lo que puede llevar a tener un impacto en la competitividad e innovación de un país. También destaca que este impacto puede ser incluso económico al no poder recuperar la inversión inicial que una persona realiza en sus estudios. El mismo autor señala que las principales causas de abandono estudiantil son: el rendimiento académico, su historial académico, variables socioeconómicas y elección de la carrera y la universidad.

Acorde a esto la mayoría de los autores señalan el factor socioeconómico como una de las causas principales; hay casos particulares como Singh & Alhulail (2022) quien considera también factores personales y un mayor riesgo en casos dónde existe un contexto social desfavorecido.

También como un factor recurrente se encuentra problemas relacionados al estudiante; en el caso de Maphosa et al. (2023) atribuye aproximadamente el 18% a problemas personales, el 16% a cambio de calificaciones y 14% a problemas administrativos.

Casos como la motivación y la autodisciplina también son mencionados en varios estudios (Alvarado-Uribe et al., 2022; Coussement et al., 2020; Dass et al., 2021; Yaibuates et al., 2022), adicional a esto, uno de los factores a considerar es la salud mental ya que ciertos autores mencionan temas relacionados con el estrés, la ideación suicida, el abuso sexual o físico (Logoz

et al., 2023) o los trastornos mentales (Tong et al., 2023) como una de las principales causas de abandono estudiantil.

Se debe hacer énfasis en que la decisión de un estudiante de abandonar sus estudios puede estar dada por una combinación de una o varias de las causas antes mencionadas (Mayra & Mauricio, 2018).

2.2. Machine Learning

Villarreal-Torres et al. (2023) argumenta que el aprendizaje automático (Machine Learning) es una rama de la inteligencia artificial y se define como el estudio y construcción de sistemas que aprenden a partir de un gran volumen de datos.

Hegde & Prageeth (2018) por su parte nos dice que se pueden utilizar técnicas de aprendizaje automático para desarrollar algoritmos que permitan a las computadoras inferir y predecir datos en base a la información recolectada, así como mejorar su rendimiento a medida que se le proporciona más datos.

En el campo de la educación, los algoritmos de Machine Learning son de gran ayuda a la hora de analizar tanto datos demográficos como académicos, y encontrar patrones útiles (Hegde & Prageeth, 2018; Ujkani et al., 2022; Won et al., 2023).

El estudio de Fernandez-Garcia et al. (2021) destaca que el aprendizaje automático se utiliza para construir modelos capaces de predecir y anticipar el abandono estudiantil

Singh & Alhulail (2022) menciona que su estudio utiliza el aprendizaje automático para crear un modelo para predecir el abandono universitario en diferentes etapas utilizando datos académicos, para ello se utiliza la regresión logística con el objetivo de analizar los factores que influyen en la decisión de los estudiantes de continuar o no con sus estudios.

Alruwais (2023) discute diversos algoritmos de aprendizaje automático para predecir el abandono estudiantil en cursos en línea; se mencionan modelos de factorización como las máquinas de factorización, mismas que son excelentes a la hora de trabajar con datos muy dispersos y de grandes dimensiones. También destaca el uso de modelos de aprendizaje profundo como redes neuronales profundas, mismas que demostraron resultados muy

prometedores a la hora de extraer aspectos significativos de los datos educativos, como patrones de aprendizaje secuenciales, dinámicas temporales e interacciones complejas. Este estudio señala que, a pesar de las ventajas potenciales de las técnicas de factorización, muchos modelos solo se centran en el uso de redes neuronales recurrentes (RNN) o redes de memoria a corto y largo plazo (LSTM).

En el estudio de Revathy & Kamalakkannan (2021) se enfatiza el uso de la Reducción de la Dimensionalidad a través del Análisis de Componentes Principales, como mecanismo de ayuda a la hora de analizar datos con técnicas de Machine Learning; este se utiliza para transformar un conjunto de variables en uno más pequeño que aún conserva la mayor parte de su información.

En otro punto de vista, Nabil et al. (2021) argumenta el uso de varios algoritmos de aprendizaje automático como redes neuronales profundas (DNN), árboles de decisión (DT), regresión logística (LR), clasificadores de vectores de soporte (SVC), vecinos más cercanos (KNN), bosques aleatorios (RF) y potenciación del gradiente (GB), para analizar y evaluar los datos de su investigación, destacando que las DNN lograron mejores resultados en términos de precisión, recuperación, puntuación F1 y métricas de error de clasificación.

Solis et al. (2018) menciona en su artículo que el Algoritmo de Bosques Aleatorios (RF) aplicado con 10 variables aleatoriamente seleccionadas, predijo correctamente el 91% de las deserciones.

Por su parte Ujkani et al. (2022) destaca la Regresión Logística como una de las técnicas de aprendizaje automático más utilizadas para la construcción de modelos. En su caso el uso del lenguaje Python con bibliotecas como NumPy y Scikit-Learn, permitió construir un modelo con una alta precisión (90%) y una puntuación F1 de 0.85.

3. METODOLOGÍA

Esta sección describe la metodología utilizada en esta investigación. Inicialmente se detallan las causas más comunes de abandono estudiantil para luego elaborar una encuesta y obtener datos dónde se utilizará un análisis exploratorio y técnicas de aprendizaje automático para obtener resultados para luego entrenar un modelo y probarlo para comparar la precisión de las técnicas seleccionadas.

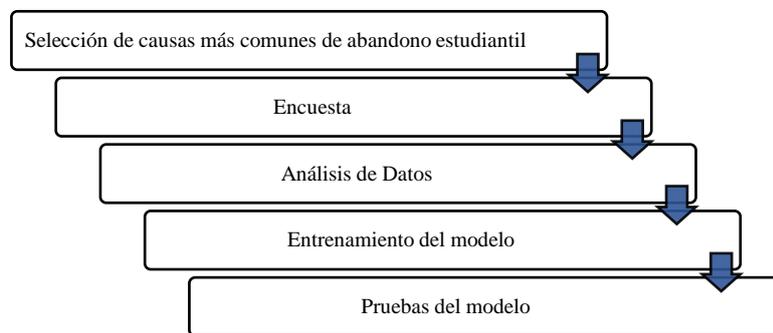


Ilustración 1. Metodología propuesta.

6.1 Métodos y técnicas de Recopilación de datos empleadas

Para recopilar información se hizo uso de encuestas a través de la plataforma Google Forms. Para ello se definieron preguntas en base a una selección de factores comunes en casos de abandono estudiantil.

6.1.1 Selección de factores

Se establecieron 4 causas basadas en la revisión de la literatura, de cada dimensión se establecieron factores en base al razonamiento lógico y las preguntas de investigación. La siguiente tabla detalla lo seleccionado:

Tabla 1. Factores seleccionados

Causa	Factores
Académica	Académico, Virtualidad
Personal	Personal, Familiar
Socioeconómica	Económico, Social
Psicológico	Adaptación, Motivación, Salud Mental

Cada uno de estos factores representa una característica importante basada en la revisión de literatura, la causa académica busca evaluar factores como el rendimiento académico, la ubicación del centro universitario, incluso la virtualidad, misma que fue implementada de forma recurrente durante la pandemia, y, por la situación actual del Ecuador, ha sido implementada en diversas ocasiones. En el caso de la causa personal esta busca evaluar factores como el género, la ubicación del estudiante, su estado civil, si tiene cargas familiares. La causa Socioeconómica por su lado busca estudiar la relevancia de factores económicos y sociales como el impacto de la situación actual del país y las dificultades económicas que el estudiante pueda presentar. También se agregó una causa psicológica en el que se busca medir el impacto de factores como adaptación, motivación y salud mental, con el fin de medir la importancia de este dentro del abandono estudiantil.

6.1.2 Preguntas de la encuesta

En base al uso del razonamiento lógico, una vez identificado los factores se procedió a elaborar la encuesta, la cual tenía las siguientes preguntas y opciones:

Tabla 2. Preguntas de la encuesta

Factor	Pregunta	Tipo
Académico	¿Cuál es tu sede?	Opción Múltiple
Académico	¿Cuál es tu carrera?	Opción Múltiple
Académico	¿En qué semestre estás?	Opción Múltiple
Académico	¿Cómo consideras tu promedio?	Opción Múltiple
Académico	¿Consideras que no tienes acceso a suficientes recursos académicos?	Opción Múltiple
Académico	¿Consideras que no hay compromiso de parte de tus profesores?	Opción Múltiple
Virtualidad	¿Consideras mejor la presencialidad o la virtualidad?	Opción Múltiple
Personal	¿Cómo te identificas?	
Personal	¿Dónde vives?	
Personal	¿Quién se encarga de pagar tus estudios?	Selección Múltiple
Familiar	¿Cuál es tu estado civil?	Opción Múltiple
Familiar	¿Cuántos hijos tienes?	Opción Múltiple
Económico	¿Cuál es tu rango de pago en la universidad?	Opción Múltiple
Económico	¿Consideras que tienes dificultades para pagar tus estudios?	Opción Múltiple
Económico	¿Actualmente estás trabajando?	Opción Múltiple
Económico	¿Te sientes agobiado de no poder pagar tus estudios?	Opción Múltiple
Social	¿Te es difícil asistir a la universidad?	Opción Múltiple

Social	¿La situación del país te agobia?	Opción Múltiple
Social	¿Consideras que ir a la universidad representa un peligro para tu vida?	Opción Múltiple
Social	¿Piensas constantemente en emigrar?	Opción Múltiple
Adaptación	¿Tienes dificultades para adaptarte a la universidad?	Opción Múltiple
Adaptación	¿No comprendes a tus profesores?	Opción Múltiple
Adaptación	¿Consideras que no tienes un grupo de estudio o amigos?	Opción Múltiple
Motivación	¿Crees que te equivocaste de carrera?	Opción Múltiple
Motivación	¿Consideras que no tienes interés en lo que estudias?	Opción Múltiple
Motivación	¿Consideras que no tienes motivación?	Opción Múltiple
Motivación	¿Consideras que tu carrera no tiene salida profesional?	Opción Múltiple
Motivación	¿Consideras que no tienes claras tus metas profesionales?	Opción Múltiple
Motivación	¿Tienes muchas responsabilidades?	Opción Múltiple
Salud Mental	¿Consideras que sufres de depresión o has pensado en quitarte la vida?	Opción Múltiple
Salud Mental	¿Consideras que tienes muchos vicios o uno muy dañino?	Opción Múltiple

Adicional a las preguntas antes mencionadas se incluyó la pregunta: ¿Qué tan probable es que abandones tus estudios? Con el fin de obtener información relevante para el análisis y la predicción de la posibilidad del estudiante de realizar esta acción.

6.2 Métodos y técnicas de Análisis de datos

Con la obtención de los datos de 100 estudiantes de la Universidad Politécnica Salesiana sede Guayaquil se procedió a realizar un análisis exploratorio de los datos. Como primer paso se realizó una limpieza de datos con el lenguaje Python, utilizando la librería Pandas, a continuación, se utilizó estadística descriptiva para notar la relación de ciertas variables para, posteriormente, proceder a renombrar las variables (ver Tabla 3), luego de esto se aplicaron varios algoritmos de Machine Learning para obtener resultados y comparar su efectividad al predecir datos.

Tabla 3. Definición de variables y medidas

Variable	Pregunta
SEDE	¿Cuál es tu sede?
CARRERA	¿Cuál es tu carrera?
SEMESTRE	¿En qué semestre estás?
PROMEDIO	¿Cómo consideras tu promedio?

Q15_REC_ACAD	¿Consideras que no tienes acceso a suficientes recursos académicos?
Q16_COMP_PROF	¿Consideras que no hay compromiso de parte de tus profesores?
PREF_VIRTU	¿Consideras mejor la presencialidad o la virtualidad?
GENERO	¿Cómo te identificas?
ZONA	¿Dónde vives?
RESP_ESTUD	¿Quién se encarga de pagar tus estudios?
EST_CIVIL	¿Cuál es tu estado civil?
NUM_HIJOS	¿Cuántos hijos tienes?
QUINTIL	¿Cuál es tu rango de pago en la universidad?
DIF_PAGO	¿Consideras que tienes dificultades para pagar tus estudios?
TIPO_TRABAJ	¿Actualmente estás trabajando?
Q05_AGOB_PAG	¿Te sientes agobiado de no poder pagar tus estudios?
Q13_DIF_AS_UNI	¿Te es difícil asistir a la universidad?
Q12_SIT_PAIS	¿La situación del país te agobia?
Q14_PELIG_UNI	¿Consideras que ir a la universidad representa un peligro para tu vida?
Q18_EMIGRAR	¿Piensas constantemente en emigrar?
Q01_ADAPT_UNIV	¿Tienes dificultades para adaptarte a la universidad?
Q02_COMP_PROF	¿No comprendes a tus profesores?
Q09_GRUP_EST	¿Consideras que no tienes un grupo de estudio o amigos?
Q03_CARR_EQUI	¿Crees que te equivocaste de carrera?
Q04_INT_EST	¿Consideras que no tienes interés en lo que estudias?
Q06_SIN_MOTIV	¿Consideras que no tienes motivación?
Q07_SIN_S_PROF	¿Consideras que tu carrera no tiene salida profesional?
Q17_META_PROF	¿Consideras que no tienes claras tus metas profesionales?
Q10_MUCH_RESP	¿Tienes muchas responsabilidades?
Q08_DEPRESION	¿Consideras que sufres de depresión o has pensado en quitarte la vida?
Q11_VICIOS	¿Consideras que tienes muchos vicios o uno muy dañino?
PROB_ABAND_EST	¿Qué tan probable es que abandones tus estudios?

4. RESULTADOS

7.1 Resultados de la encuesta

Se procede a mostrar los resultados más relevantes de la encuesta. Como primer punto se obtuvo que el 47% de las personas encuestadas se identifican como mujer, mientras que el 53% se identifican como hombres, revelando que la participación fue casi equitativa (Ilustración 2).

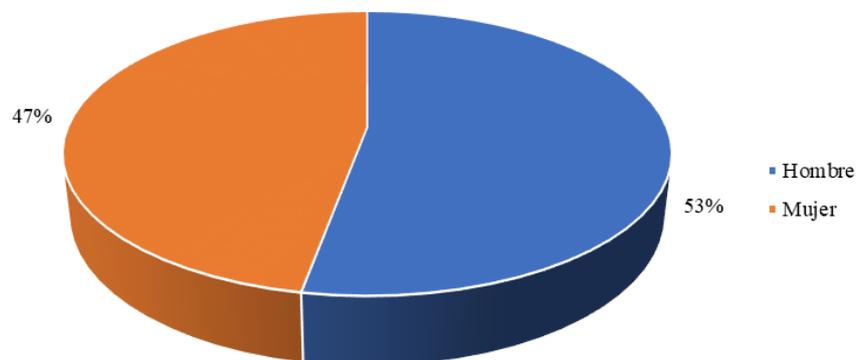


Ilustración 2. Resultados de la pregunta ¿Cómo te identificas?

Los datos también reflejan una mayor participación de estudiantes del campus Centenario, siendo esta del 85%, mientras que la participación del campus María Auxiliadora fue del 15% (Ilustración 3).

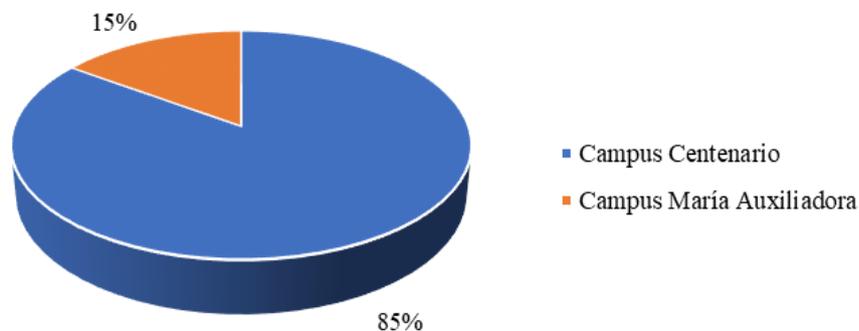


Ilustración 3. Resultados de la pregunta ¿Cuál es tu sede?

Los datos también revelan que la carrera con mayor participación fue la carrera de Computación, con cerca del 33% de los encuestados. Luego le siguen la carrera de Administración de Empresas con el 10%, Contabilidad y Auditoría con el 9%, Ingeniería Industrial con el 7% e Ingeniería Automotriz con el 5%; la participación de otras carreras cuyo porcentaje sea menor o igual al 4% de participación, se resume en el 36% (Ilustración 4).

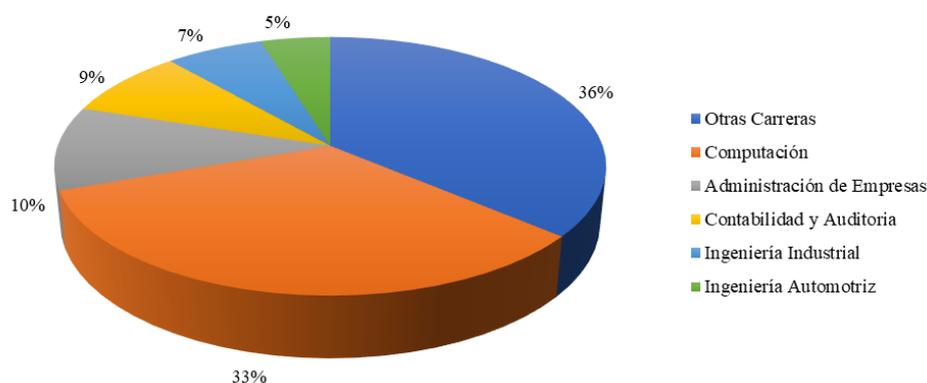


Ilustración 4. Resultados de la pregunta ¿Cuál es tu carrera?

De acuerdo con los datos la participación tuvo mayor participación en estudiantes de segundo semestre (12 personas), quinto semestre (15 personas), sexto semestre (22 personas), séptimo semestre (12 personas) y octavo semestre (26 personas), siendo este último el que mayor participación tuvo (Ilustración 5).

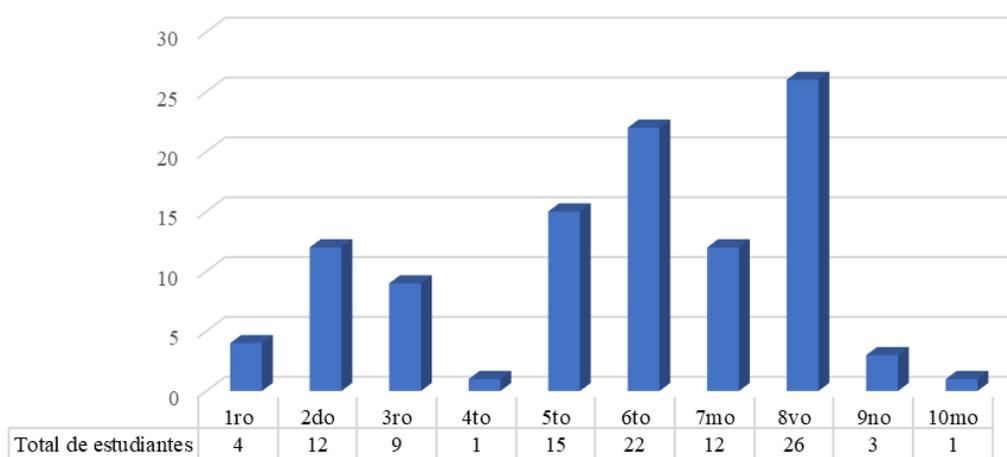


Ilustración 5. Resultados de la pregunta ¿En qué semestres estás?

También la encuesta tuvo una mayoría de participantes del sur de la ciudad de Guayaquil (48%) seguido por el norte de la ciudad (29%) y el centro de la ciudad (11%), los datos revelan también que hubo un 12% de participación de personas que viven fuera de la ciudad (Ilustración 6).

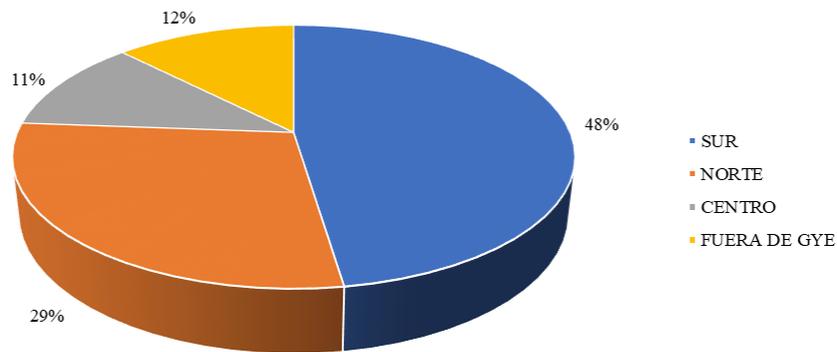


Ilustración 6. Resultados de la pregunta ¿Dónde vives?

Otro dato importante para conocer a los participantes de la encuesta es su nivel económico, para ello se tomaron en cuenta los resultados de la pregunta: “¿Cuál es tu rango de pago mensual en la universidad?”, dividiendo las respuestas en 3 categorías: Ingresos bajos para todos los estudiantes cuyos pagos mensuales a la universidad sean menores o iguales a 200, ingresos medios para todos los estudiantes cuyos pagos mensuales sean mayores a 200 pero menores o iguales a 450, e ingresos altos para todos los estudiantes cuyos pagos mensuales a la universidad sean mayores a 450 dólares o paguen al contado o con tarjeta de crédito.

Se obtuvo de este análisis que el 52% de los encuestados cuenta con un nivel económico medio, el 36% con un nivel económico bajo y el 12% con un nivel económico alto (Ilustración 7).

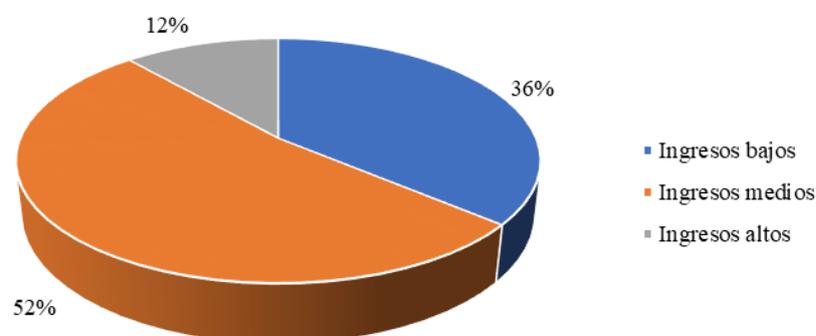


Ilustración 7. Resultados Nivel Económico.

Otro dato importante para conocer a la población es saber el estado civil de los encuestados, los estudiantes solteros representan el 93.33% del total de la población, mientras que los casados al 4.76% y aquellos con unión libre al 1.90% (Ilustración 8); además, de los estudiantes solteros el 98.96% no tiene hijos mientras que el 1.04% tiene un hijo, en cuanto a los estudiantes casados el 50% de ellos no tiene hijos, el otro 50% tiene solo 1 hijo; de los estudiantes en unión libre el 20% no tiene hijos, el 40% tiene 1 hijo y el otro 40% tiene 2 hijos (Ilustración 9).

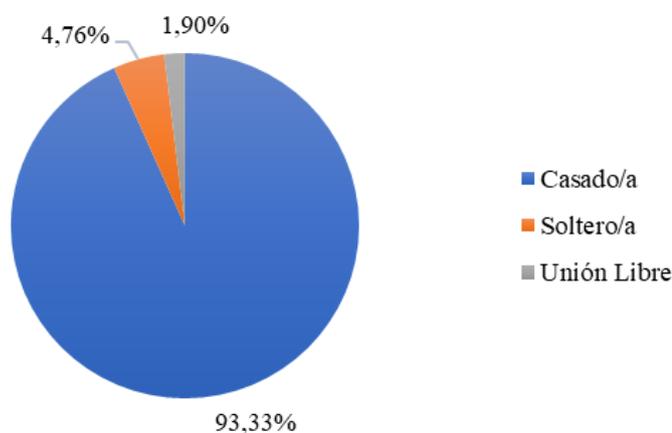


Ilustración 8. Respuestas de la pregunta ¿Cuál es tu estado civil?

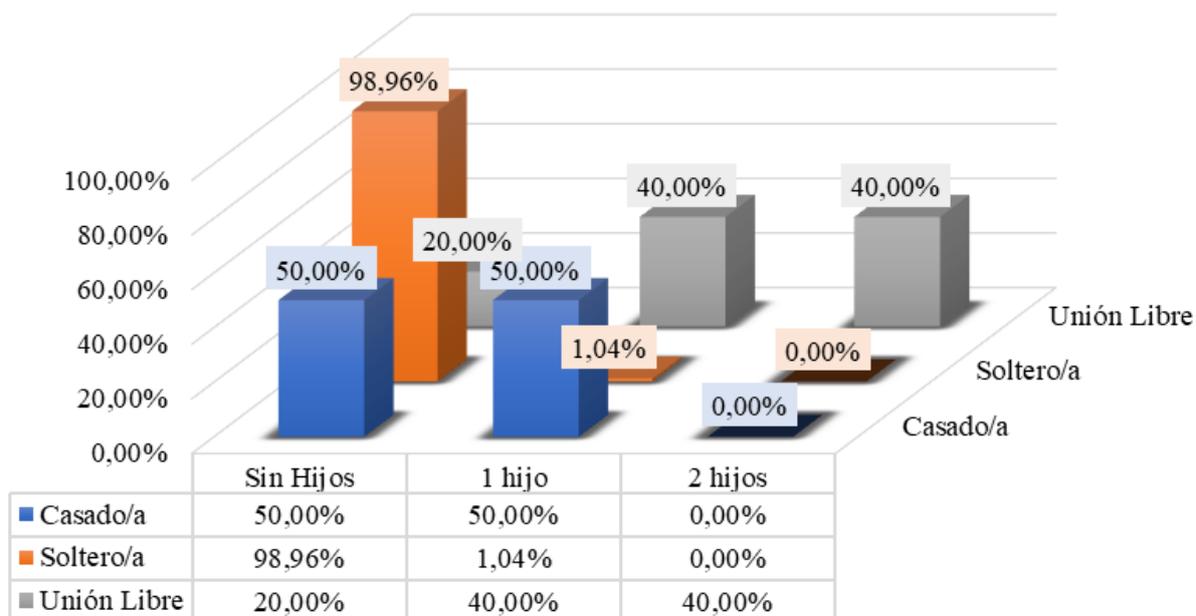


Ilustración 9. Número de hijos

En cuanto a la situación laboral de las personas encuestadas tenemos que el 50% de las personas encuestadas no trabaja, mientras que el 19% trabaja a tiempo completo, el 13% a medio tiempo, el 10% trabaja a tiempo parcial (jornada menor a 20 horas semanales) y el 8% tiene un emprendimiento (Ilustración 10).

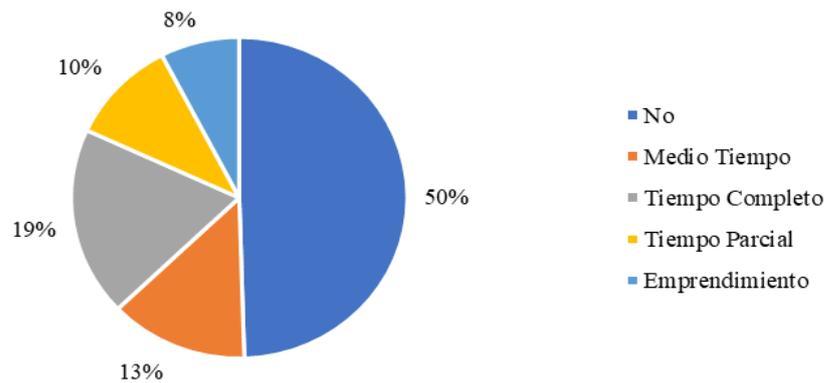


Ilustración 10. Respuestas de la pregunta ¿Actualmente te encuentras trabajando?

Además, del total de encuestados, 67 de ellos respondieron que Un familiar o cercano paga sus estudios, lo que correspondería al 64% del total encuestado. También 39 de ellos respondieron que pagan sus estudios, de estos 3 también seleccionaron que un familiar o cercano ayuda a pagar sus estudios, el total correspondería al 37% del universo de encuestados. Las personas con beca corresponden únicamente a 4 encuestados, de ellos dos también mencionan que un familiar cercano paga sus estudios y de estos dos uno indica que el también aporta a pagar sus estudios, representando, el total, al 4% del universo de personas encuestadas (Ilustración 11).

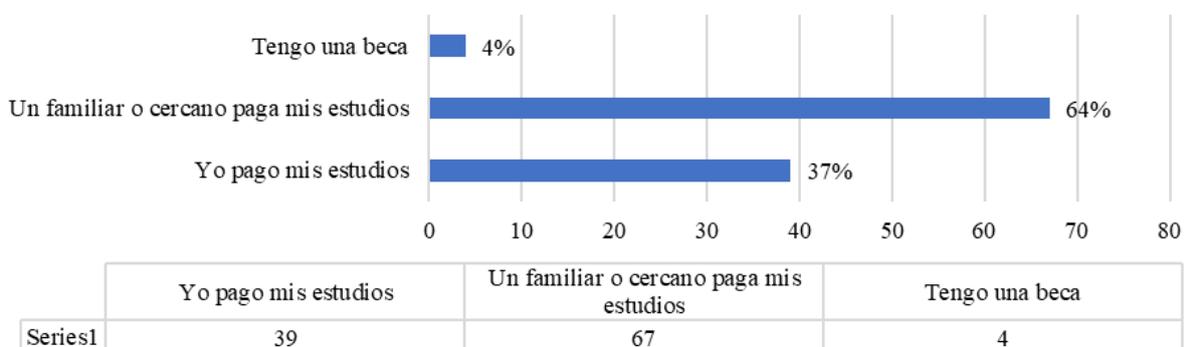


Ilustración 11. Respuestas de la pregunta ¿Quién paga tus estudios?

La encuesta también nos indica que el 45% de las personas considera su promedio estudiantil bueno, el 38% lo considera muy bueno, el 9% sobresaliente y el 8% regular (Ilustración 12).

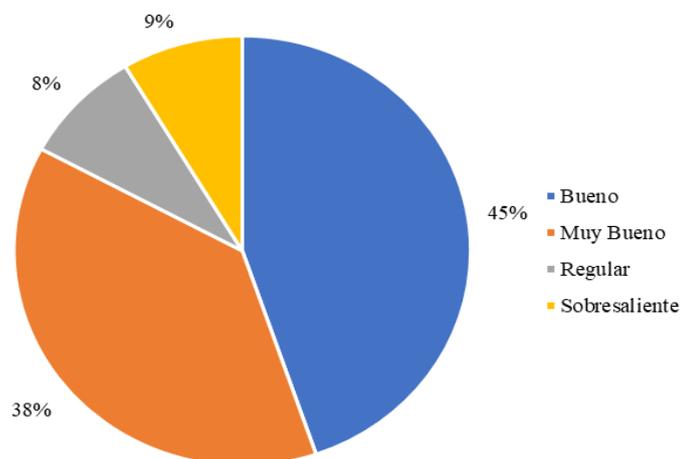


Ilustración 12. Respuestas a la pregunta ¿Cómo consideras tu promedio?

Un dato importante a revisar es la relación entre la dificultad para pagar los estudios y el nivel económicos. Dentro de la encuesta se preguntó: “¿Consideras que tienes dificultades para pagar tus estudios?” y se obtuvieron los siguientes resultados: El 32% considera que suele tener dificultades, el 27% considera que no mientras que el 41% considera que tiene dificultades para pagar sus estudios (Ilustración 13).

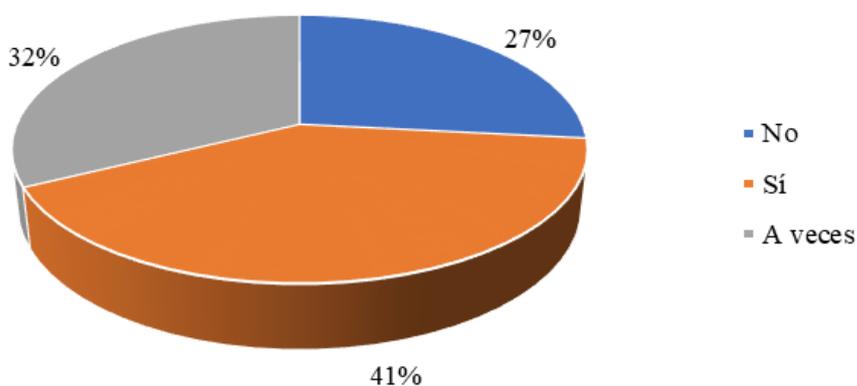


Ilustración 13. Respuestas a la pregunta ¿Consideras que tienes dificultades para pagar tus estudios?

Estos datos se cruzaron con el nivel económico y se obtuvo que de las personas con ingresos altos el 25% no considera que tenga dificultades para pagar la universidad, mientras que el

33.33% considera que sí y el 41.67% considera que a veces suele tener dificultades. Por otro lado, en las personas de ingresos bajos el 21.05% considera que no tiene dificultades, el 42.11% considera que si y el 36.84% considera que a veces tiene dificultades. En el caso de ingresos medios el 30.91% de las personas considera que no tiene dificultades, mientras que el 41.82% considera que si y el 27.27% considera que a veces suele tener dificultades (Ilustración 14).

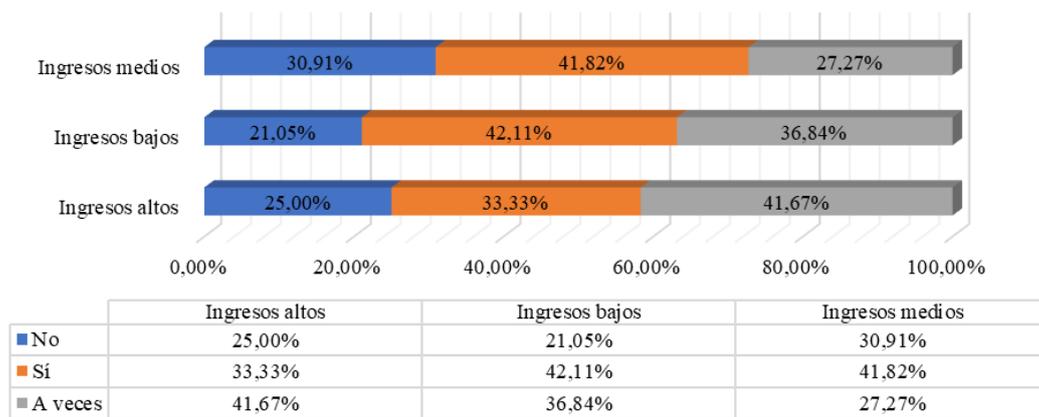


Ilustración 14. Nivel económico vs Dificultad de Pago

Un factor que se tomó en cuenta y que muestra resultados interesantes es el de la virtualidad, los resultados de la pregunta: “¿Consideras mejor la presencialidad o la virtualidad?” nos detallan que, según los encuestados, el 43% prefiere las clases presenciales, el 37% prefiere una combinación de ambas y apenas el 20% prefiere las clases virtuales (Ilustración 15).

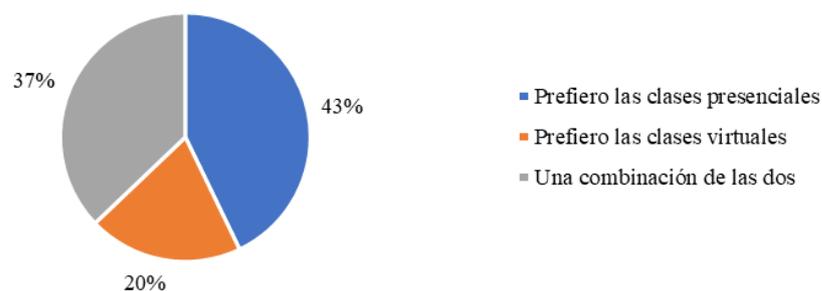


Ilustración 15. Respuestas a la pregunta ¿Consideras mejor la presencialidad o la virtualidad?

A continuación, se muestran los resultados que obtuvieron las preguntas que se clasifican en 5 valores: Muy de acuerdo, De acuerdo, Indiferente, En desacuerdo y Muy en desacuerdo, en base al 100% de los encuestados (Ilustración 16) (Tabla 4).

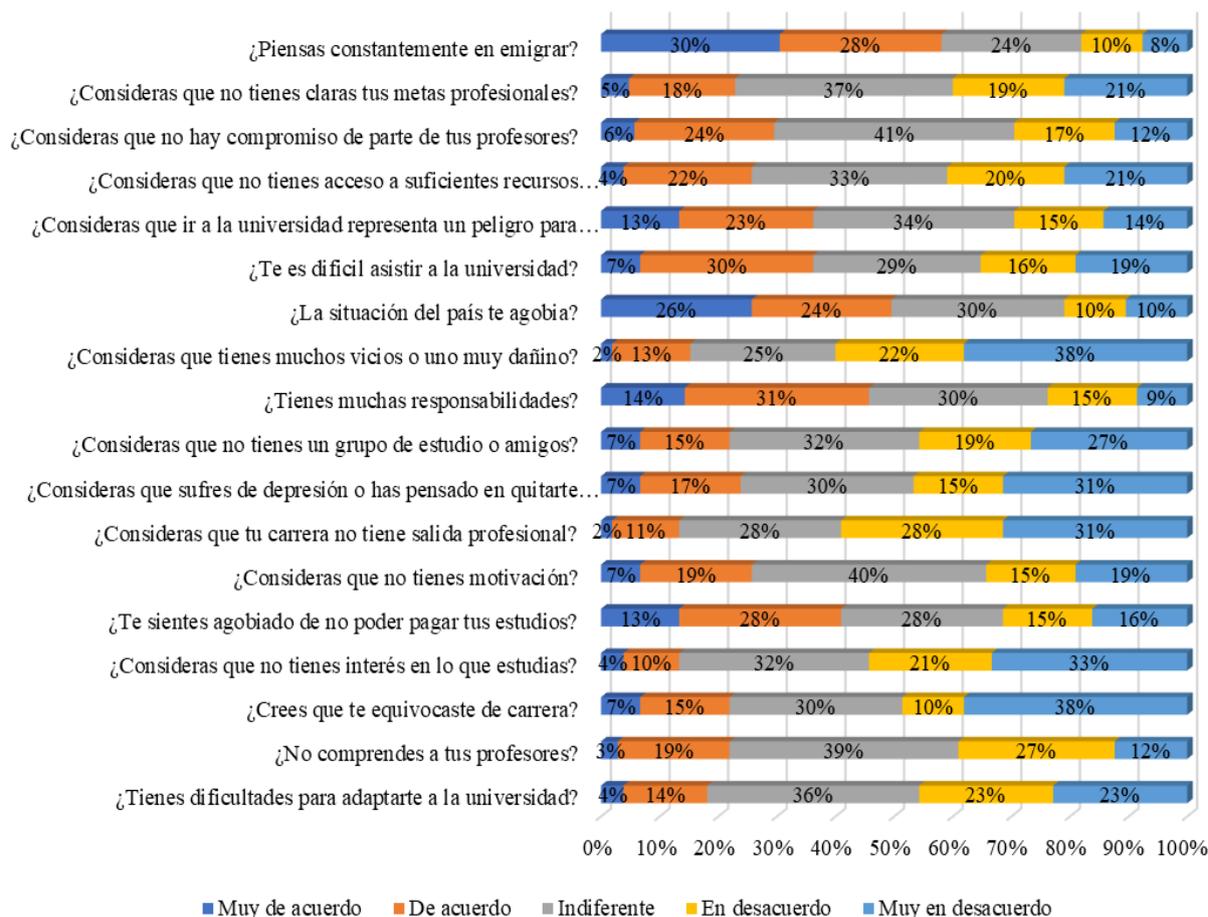


Ilustración 16. Resultado de Preguntas

Tabla 4. Tabla de resultados de preguntas

Pregunta	Muy de acuerdo	De acuerdo	Indiferente	En desacuerdo	Muy en desacuerdo
¿Tienes dificultades para adaptarte a la universidad?	4%	14%	36%	23%	23%
¿No comprendes a tus profesores?	3%	19%	39%	27%	12%
¿Crees que te equivocaste de carrera?	7%	15%	30%	10%	38%
¿Consideras que no tienes interés en lo que estudias?	4%	10%	32%	21%	33%
¿Te sientes agobiado de no poder pagar tus estudios?	13%	28%	28%	15%	16%
¿Consideras que no tienes motivación?	7%	19%	40%	15%	19%
¿Consideras que tu carrera no tiene salida profesional?	2%	11%	28%	28%	31%
¿Consideras que sufres de depresión o has pensado en quitarte la vida?	7%	17%	30%	15%	31%
¿Consideras que no tienes un grupo de estudio o amigos?	7%	15%	32%	19%	27%
¿Tienes muchas responsabilidades?	14%	31%	30%	15%	9%
¿Consideras que tienes muchos vicios o uno muy dañino?	2%	13%	25%	22%	38%

¿La situación del país te agobia?	26%	24%	30%	10%	10%
¿Te es difícil asistir a la universidad?	7%	30%	29%	16%	19%
¿Consideras que ir a la universidad representa un peligro para tu vida?	13%	23%	34%	15%	14%
¿Consideras que no tienes acceso a suficientes recursos académicos?	4%	22%	33%	20%	21%
¿Consideras que no hay compromiso de parte de tus profesores?	6%	24%	41%	17%	12%
¿Consideras que no tienes claras tus metas profesionales?	5%	18%	37%	19%	21%
¿Piensas constantemente en emigrar?	30%	28%	24%	10%	8%

Una última pregunta de la encuesta se centra en saber la probabilidad de que un estudiante abandone sus estudios, del total general tenemos que el 40% respondió que no los abandonaría (probabilidad de abandonarlos 0%), el 27% respondió que hay poca probabilidad (probabilidad de abandonarlos 50%), el 24% respondió que no ha pensado en ello (probabilidad de abandonarlos 25%), el 8% respondió que ha estado pensándolo (probabilidad de abandonarlos 75%), y el 1% respondió que los abandonará (100%) (Ilustración 17. Respuestas a la pregunta ¿Qué tan probable es que abandones tus estudios?).

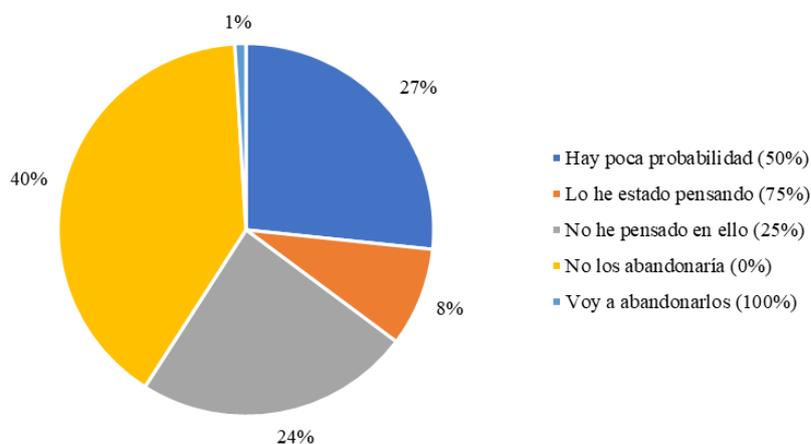


Ilustración 17. Respuestas a la pregunta ¿Qué tan probable es que abandones tus estudios?

7.2 Resultado del análisis

En el análisis de datos se limpió el Data Frame obtenido de la encuesta para luego renombrar las columnas (Tabla 3) y transformar los datos en tipos numéricos a través del uso de la librería LabelEncoder de Sklearn.

Luego, para una mejor visualización, se obtuvieron cuatro matrices de correlación basándose en cada una de las causas seleccionadas (Tabla 1). La matriz de correlación permite saber que tan relacionada están dos variables, entre más cercana sea la relación a 1 más correlacionada están las variables y viceversa.

Para el análisis de las matrices de correlación se toman en cuenta las variables cuyo valor, en valor absoluto, de su correlación superen el umbral de 0.1, luego de 0.15 y luego de 0.2, con el fin de obtener diferentes grupos de datos con los cuales medir el rendimiento de 4 algoritmos de aprendizaje automático como son Árboles de decisión (AD), Bosques Aleatorios (AF), K-Vecinos más cercanos (KNN) y Regresión Lineal (RL). Se utiliza el valor absoluto ya que para el modelado de datos las variables con correlación negativa llegan a tomar importancia gracias a que su relación aporta una relación inversa, además estas variables pueden llegar a tener mayor impacto a la hora de predecir, así como permitir la reducción del sesgo.

Se puede notar que, en la matriz de correlación de causas académicas, las variables más significativas (umbral en valor absoluto mayor a 0.1) serían: SEDE, CARRERA, SEMESTRE, PROMEDIO, Q15_REC_ACAD y Q16_COMP_PROF (Ilustración 18).

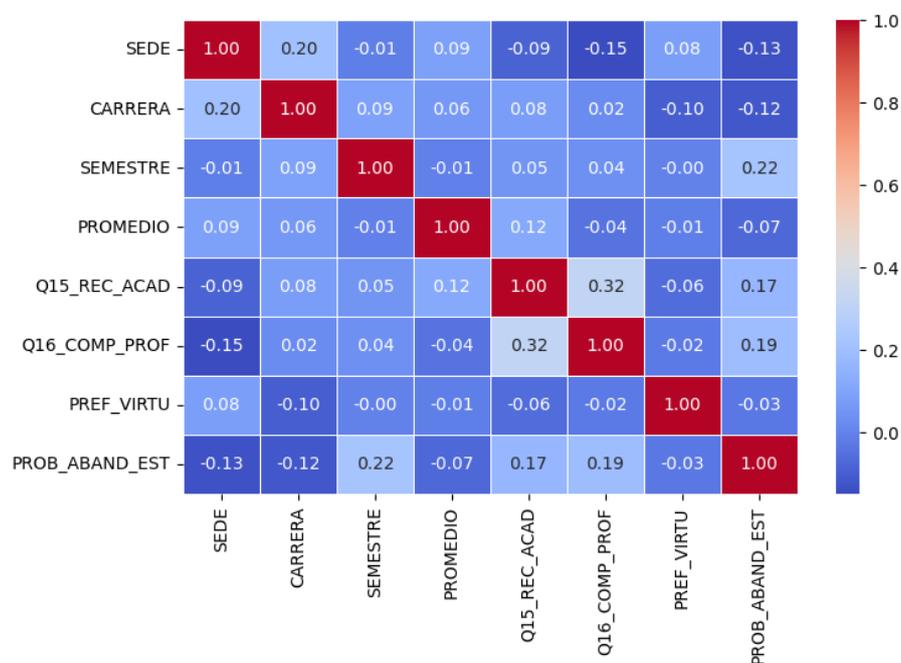


Ilustración 18. Matriz de correlación - Causas Académicas

Para las causas personales la matriz de correlación, las variables más significativas son: GENERO y NUM_HIJOS (Ilustración 19).

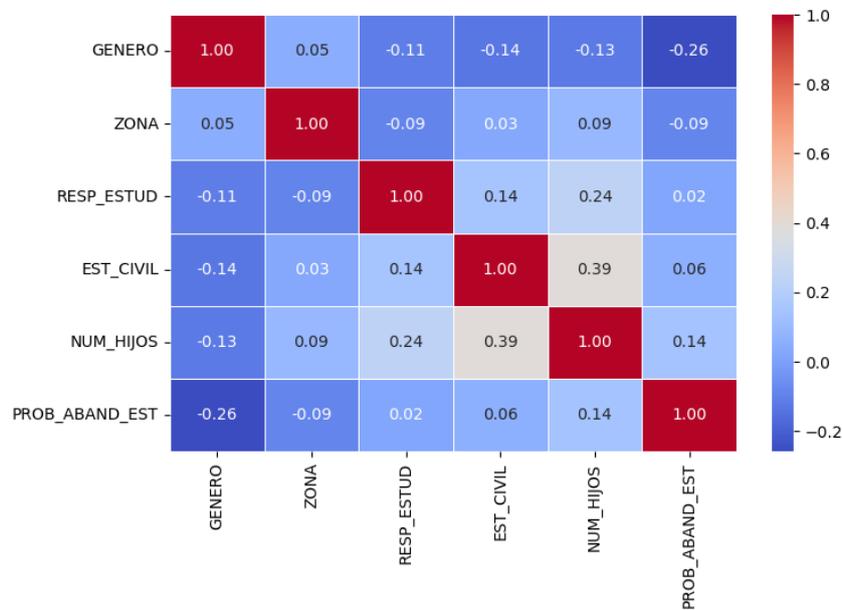


Ilustración 19. Matriz de Correlación - Causas Personales

Para el caso de la matriz socioeconómica los valores más significativos son: DIF_PAGO, TIPO_TRABAJ, Q05_AGOB_PAG, Q12_SIT_PAIS y Q14_PELIG_UNIV, (Ilustración 20).

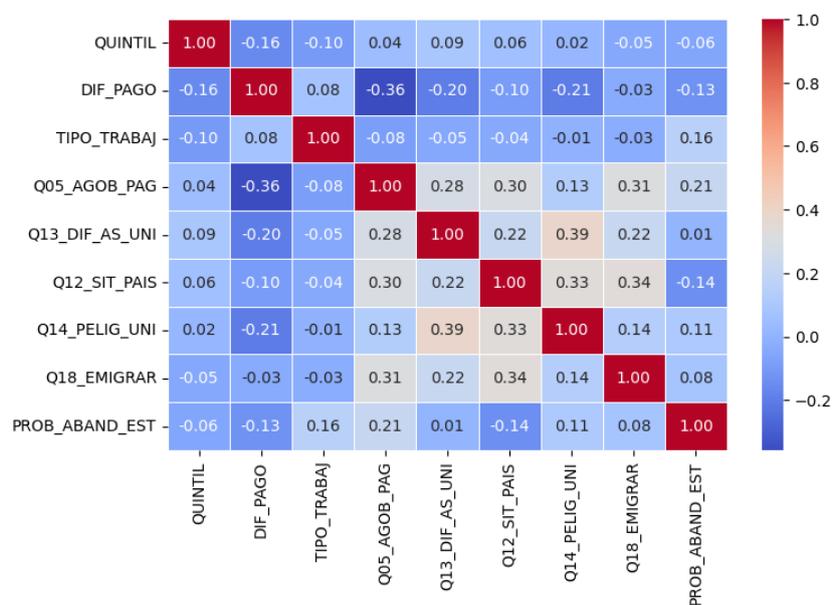


Ilustración 20. Matriz Correlación - Causas Socioeconómicas

Las causas psicológicas, por otro lado, aportan varios datos significativos al análisis. Estas son las variables que superan el umbral: Q01_ADAPT_UNIV, Q02_COMP_PROF, Q03_CARR_EQUI, Q04_INT_EST, Q06_SIN_MOTIV, Q07_SIN_S_PROF, Q17_META_PROF, Q10_MUCH_RESP y Q08_DEPRESION (Ilustración 21).

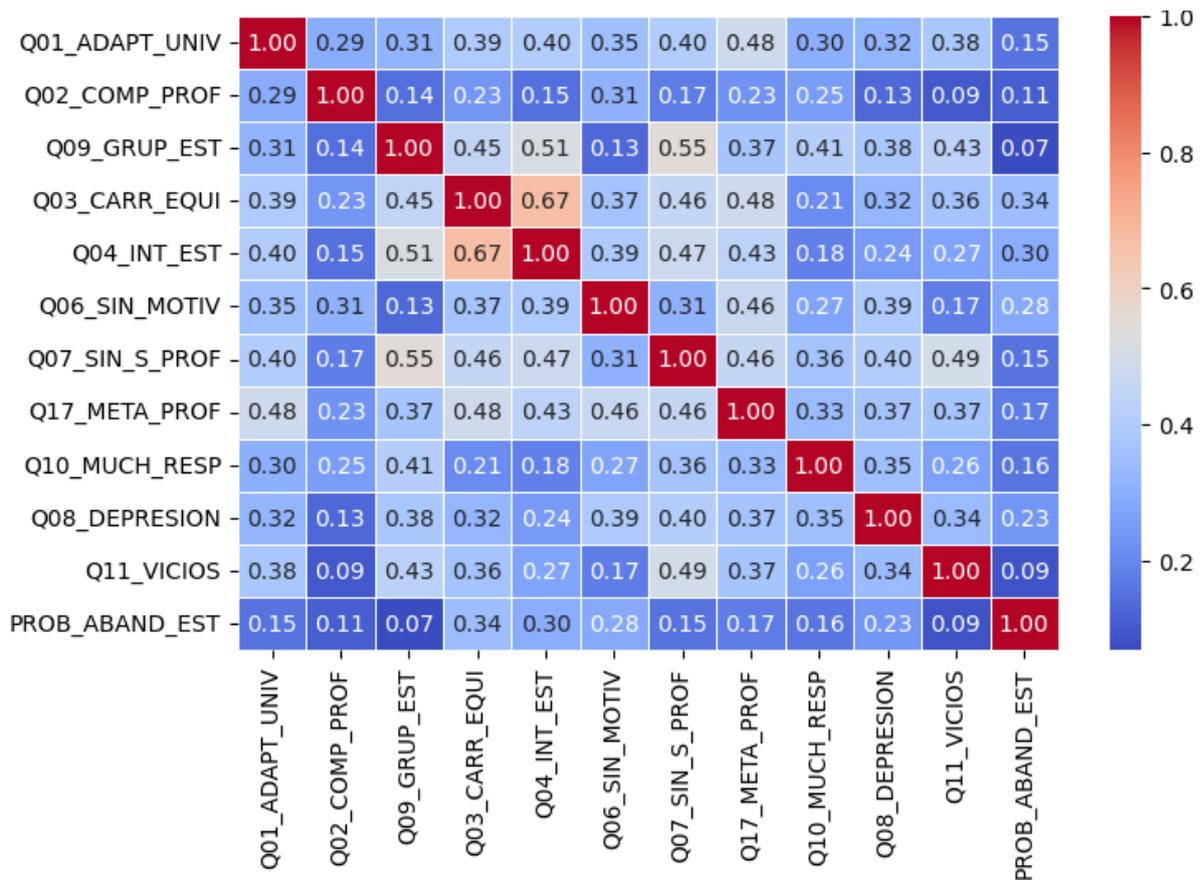


Ilustración 21. Matriz de Correlación - Causas Psicológicas

Luego de seleccionar las causas se procedió a testear diferentes algoritmos de aprendizaje automático; los resultados (Tabla 5) muestran que la Relación Lineal (RL) no es una buena técnica para aplicar con nuestro modelo de datos, en este caso el Error Cuadrado (R^2) debe ser cercano a 1 para tener un buen rendimiento, y los valores se acercan más a cero o incluso llegan a ser negativos; en el caso de los árboles de decisión su mejor rendimiento fue con el umbral de 0.25 los datos, mientras que los K-Vecinos más cercano (KNN) tuvo un rendimiento bajo con un umbral superior a 0.20. El mejor rendimiento lo obtuvimos con un umbral de 0.20 usando Bosques Aleatorios (RF), lo cual coincide con la teoría ya que en modelos con pocos

datos suele ser una de las técnicas más efectiva, en este caso nos dio un rendimiento de 0.82, lo cual se considera aceptable para un modelo de predicción.

Tabla 5. Rendimiento de Algoritmos de Machine Learning

Umbral	<u>Rendimiento</u>			
	F1-Score			R ²
	DT	RF	KNN	RL
0,1	0,64	0,73	0,64	-0,34
0,15	0,55	0,73	0,64	0,15
0,2	0,64	0,82	0,64	0,09
0,25	0,73	0,73	0,55	-0,16

5. DISCUSIÓN

Los datos han arrojado información clave respecto a las causas de abandono estudiantil; de acuerdo con las matrices de correlación presentadas y el algoritmo de aprendizaje automático con mejor rendimiento (algoritmo de Bosques Aleatorios - RF), los factores más importantes a la hora de predecir el abandono estudiantil en la Universidad Politécnica Salesiana sede Guayaquil serían: el género, el semestre en el cual se encuentra un estudiante, su pensamiento respecto a si escogió o no la carrera correcta, el interés que este tiene respecto a su carrera, que tan agobiado se siente por sus pagos, si está motivado y si tiene problemas relacionados con la depresión.

Sin embargo, dadas las limitaciones de este estudio en cuanto al volumen de datos, al no tener acceso a una fuente de datos oficial de la universidad, las condiciones descritas podrían variar con un universo superior de personas, aun así, dentro de los datos se puede encontrar información valiosa que puede ayudar a identificar problemáticas a estudiar a futuro.

Si analizamos las matrices de correlaciones podremos encontrar temas que pueden ser interesantes de abordar en futuros trabajos; existe una correlación importante en la matriz correspondiente a las causas académicas, entre como un estudiante considera el compromiso del docente y el tener o no acceso a suficientes recursos académicos (Ilustración 18).

Otra correlación importante identificada dentro de la matriz de causas socioeconómicas es la que existe entre la dificultad para asistir a la universidad y la sensación de peligro (Ilustración 20). Por otro lado, la mayor correlación la podemos notar a en la matriz de causas psicológicas; hay una correlación muy fuerte entre la sensación de haber seleccionado mal la carrera y el interés de un estudiante en la misma, también entre el no tener amistades o grupo de estudios y la sensación de que la carrera no tiene salida profesional, también entre no tener un grupo de estudios y el interés por la carrera (Ilustración 21).

Otros datos interesantes también los revela el análisis de la encuesta. El tema de la migración está muy presente en los estudiantes, habiendo una mayoría que por lo menos lo está pensando, quizá esto tenga una relación muy fuerte con la situación actual del país (Ilustración 16).

También hay que mencionar que la mayoría de los encuestados no está dispuesto a dejar sus estudios, pero si hay un rango del 25% que se ha planteado hacerlo (Ilustración 17).

Un punto que no se incluyó en el análisis fue una sección con una pregunta abierta, donde los encuestados podían dejar una opinión respecto a que factor sería decisivo a la hora de abandonar sus estudios, entre las respuestas dadas las que más resaltaron fueron el no poder seguir pagando los estudios, la inseguridad, la falta de empleo, el estado de ánimo y el acompañamiento docente, lo cual coincide con los resultados de este estudio.

6. CONCLUSIÓN

Los resultados van acorde a los estudios revisados, el algoritmo que brindó excelentes resultados fue Bosques Aleatorios (RF), lo que respalda que sea una de las técnicas más utilizadas a la hora de analizar y predecir el abandono estudiantil, tal como se analizó, siendo factores importantes, hablando específicamente del entorno de la UPS sede Guayaquil, la economía, la motivación y la salud mental.

Este estudio pudo revelar factores críticos que son relevantes a la hora de predecir el abandono estudiantil en la Universidad Politécnica Salesiana sede Guayaquil, destacando la importancia del género, el nivel académico, la percepción de la carrera, la motivación, las dificultades económicas y la salud mental.

Los resultados, que son respaldados por la aplicación del algoritmo de Bosques Aleatorios (RF), hacen énfasis en la relación entre factores personales, académicos y socioeconómicos. Es imperativo que futuras investigaciones busquen expandir este estudio aportando más datos y explorando a más profundidad los factores que se han identificado, especialmente aquellos que se relacionan con el compromiso docente y el acceso a recursos académicos, para poder desarrollar estrategias más efectivas que eviten a un estudiante abandonar sus estudios.

A pesar de contar con limitaciones que se relacionan con el tamaño de la muestra, este estudio aporta información valiosa para proponer políticas de intervención enfocadas en la retención estudiantil, haciendo énfasis en la necesidad de abordar aspectos relacionados con el bienestar económico y psicológico. La colaboración entre la administración de la universidad, docentes y estudiantes es un factor clave para crear un entorno educativo más equitativo y apoyar al éxito de sus estudiantes.

7. REFERENCIAS

- Alruwais, N. M. (2023). Deep FM-Based Predictive Model for Student Dropout in Online Classes. *IEEE Access*, *11*, 96954–96970. <https://doi.org/10.1109/ACCESS.2023.3312150>
- Alvarado-Uribe, J., Mejía-Almada, P., Masetto Herrera, A. L., Molontay, R., Hilliger, I., Hegde, V., Montemayor Gallegos, J. E., Ramírez Díaz, R. A., & Ceballos, H. G. (2022). Student Dataset from Tecnológico de Monterrey in Mexico to Predict Dropout in Higher Education. *Data* 2022, Vol. 7, Page 119, 7(9), 119. <https://doi.org/10.3390/DATA7090119>
- Casanova, J. R., Cervero, A., Núñez, J. C., Almeida, L. S., & Bernardo, A. (2018). Factors that determine the persistence and dropout of university students. *Scopus*, *30*(4), 408–414. <https://doi.org/10.7334/PSICOTHEMA2018.155>
- Coussement, K., Phan, M., De Caigny, A., Benoit, D. F., & Raes, A. (2020). Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model. *Decision Support Systems*, *135*, 113325. <https://doi.org/10.1016/J.DSS.2020.113325>
- Dass, S., Gary, K., & Cunningham, J. (2021). Predicting Student Dropout in Self-Paced MOOC Course Using Random Forest Model. *Information 2021*, Vol. 12, Page 476, 12(11), 476. <https://doi.org/10.3390/INFO12110476>
- Fernandez-Garcia, A. J., Preciado, J. C., Melchor, F., Rodriguez-Echeverria, R., Conejero, J. M., & Sanchez-Figueroa, F. (2021). A real-life machine learning experience for predicting university dropout at different stages using academic data. *IEEE Access*, *9*, 133076–133090. <https://doi.org/10.1109/ACCESS.2021.3115851>
- Hegde, V., & Prageeth, P. P. (2018). Higher education student dropout prediction and analysis through educational data mining. *Proceedings of the 2nd International Conference on Inventive Systems and Control, ICISC 2018*, 694–699. <https://doi.org/10.1109/ICISC.2018.8398887>

- Kabashi, Q., Shabani, I., & Caka, N. (2022). Analysis of the Student Dropout Rate at the Faculty of Electrical and Computer Engineering of the University of Prishtina, Kosovo, from 2001 to 2015. *IEEE Access*, *10*, 68126–68137. <https://doi.org/10.1109/ACCESS.2022.3185620>
- Logoz, F., Weber, R., Eggenberger, L., & Walther, A. (2023). Thoughts about dropping out of studies as warning sign for suicidal ideation and mental health problems in male university students. *Journal of Men's Health*, *19*(8), 60–69. <https://doi.org/10.22514/jomh.2023.072>
- Maphosa, M., Doorsamy, W., & Paul, B. S. (2023). Student Performance Patterns in Engineering at the University of Johannesburg: An Exploratory Data Analysis. *IEEE Access*, *11*, 48977–48987. <https://doi.org/10.1109/ACCESS.2023.3277225>
- Mayra, A., & Mauricio, D. (2018). Factors to predict dropout at the universities: A case of study in Ecuador. *IEEE Global Engineering Education Conference, EDUCON, 2018-April*, 1238–1242. <https://doi.org/10.1109/EDUCON.2018.8363371>
- Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2021). Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks. *IEEE Access*, *9*, 140731–140746. <https://doi.org/10.1109/ACCESS.2021.3119596>
- Revathy, M., & Kamalakkannan, S. (2021). Collaborative learning for improving intellectual skills of dropout students using datamining techniques. *Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021*, 236–240. <https://doi.org/10.1109/ICAIS50930.2021.9395912>
- Singh, H. P., & Alhulail, H. N. (2022). Predicting Student-Teachers Dropout Risk and Early Identification: A Four-Step Logistic Regression Approach. *IEEE Access*, *10*, 6470–6482. <https://doi.org/10.1109/ACCESS.2022.3141992>
- Solis, M., Moreira, T., Gonzalez, R., Fernandez, T., & Hernandez, M. (2018). Perspectives to Predict Dropout in University Students with Machine Learning. *2018 IEEE International Work Conference on Bioinspired Intelligence, IWOB 2018 - Proceedings*. <https://doi.org/10.1109/IWOB.2018.8464191>

- Tong, Y., Wang, S., Cao, L., Zhu, D., Wang, F., Xie, F., Zhang, X., Wang, G., & Su, P. (2023). School dropouts related to mental disorders: A systematic review and meta-analysis. In *Asian Journal of Psychiatry* (Vol. 85). Elsevier B.V. <https://doi.org/10.1016/j.ajp.2023.103622>
- Ujkani, B., Minkovska, D., & Stoyanova, L. (2022). Application of Logistic Regression Technique for Predicting Student Dropout. *2022 31st International Scientific Conference Electronics, ET 2022 - Proceedings*. <https://doi.org/10.1109/ET55967.2022.9920280>
- Villarreal-Torres, H., Ángeles-Morales, J., Marín-Rodríguez, W., Andrade-Girón, D., Cano-Mejía, J., Mejía-Murillo, C., Flores-Reyes, G., & Palomino-Márquez, M. (2023). Classification model for student dropouts using machine learning: A case study. *EAI Endorsed Transactions on Scalable Information Systems*, 10(5). <https://doi.org/10.4108/EETSIS.VI.3455>
- Won, H. S., Kim, M. J., Kim, D., Kim, H. S., & Kim, K. M. (2023). University Student Dropout Prediction Using Pretrained Language Models. *Applied Sciences (Switzerland)*, 13(12). <https://doi.org/10.3390/app13127073>
- Yaibuates, M., Kantawong, K., Pravesjit, S., Yoddumnern, A., & Chaichumpa, S. (2022). The Classification Method for Risks of Dropout Students: A Case Study of Vocational Students. *7th International Conference on Digital Arts, Media and Technology, DAMT 2022 and 5th ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering, NCON 2022*, 50–53. <https://doi.org/10.1109/ECTIDAMTNCON53731.2022.9720395>