UNIVERSIDAD NACIONAL DE CAÑETE

FACULTAD DE INGENIERÍA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Modelo predictivo basado en machine learning para el diagnóstico temprano en la deserción universitaria de la UNDC, periodo 2020 – 2022

TESIS PARA OBTAR EL TÍTULO PROFESIONAL DE INGENIERO DE SISTEMAS

AUTOR:

Hugo Jesús Sulca Martínez

https://orcid.org/0009-0003-3177-7605

ASESORA:

Dra. Amanda Durán Carhuamaca

https://orcid.org/0000-0001-8183-5891

CO-ASESOR:

Ph. D. Dulio Oseda Gago

https://orcid.org/0000-0002-3136-6094

LÍNEA DE INVESTIGACIÓN:

Ciencia de datos e inteligencia artificial

San Vicente de Cañete, Perú, 2024

FIRMA DIGITAL
Firmado digitalmente por: DURAN CARHUAMACA AMANDA FIR 20114878 hard
Motivo: En señal de conformidad
Fecha: 27/05/2025 15:19:13-0500

REPÚBLICA DEL PERÚ Firma Digital
Firmado digitalmente por: OSEDA GAGO Dulio FAU 20491363402 soft
Motivo: Soy el autor del documento
Fecha: 26/05/2025 14:00:48-0500

Firma de los Jurados



Firmado digitalmente por:
LARICO UCHAMACO GUIDO
RAUL FIR 02443179 hard
Motivo: Soy el autor del
documento
Fecha: 29/05/2025 13:47:18-0500

Dr. Guido Raúl Larico Uchamaco
Presidente
Jurado Evaluador
Facultad de Ingeniería



Firmado digitalmente por:
ROQUE TITO EDWIN
Motivo: En señal de
conformidad
Fecha: 29/05/2025 11:39:14-0500

Dr. Edwin Roque Tito
Secretario
Jurado Evaluador
Facultad de Ingeniería



Firmado digitalmente por:
ASTO HUAMAN Leonidas FAU
20491363402 soft
Motivo: En señal de
conformidad
Fecha: 29/05/2025 10:33:29-0500

M. Sc. Leonidas Asto Huamán
Vocal
Jurado Evaluador
Facultad de Ingeniería

Acta de Sustentación

UNIVERSIDAD NACIONAL DE CAÑETE

Código: F-M01.03.04-GT-003 Fecha de Aprobación: 16-10-2023	Versión: 04

FACULTAD DE INGENIERIA

ACTA DE SUSTENTACIÓN DE TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO DE SISTEMAS

En la ciudad de Cañete, en la sala de sustentaciones N° 01 de la Facultad de Ingeniería, siendo las 10:00 horas, del día 18 de diciembre del 2024, se reunió el Jurado Evaluador y el asesor, designados por DECRETO No 013-2023-UIIS/FI/UNDC.
Presidente(a): Dr. Guido Raúl Larico Uchamaco
Secretario (a): Dr. Edwin Roque Tito
Vocal : M.Sc. Leonidas Asto Huaman

Asesor : Mg. Amanda Durán Carhuamaca

Para el acto de sustentación de la tesis titulada: **Modelo predictivo basado en machine learning para el diagnóstico temprano en la deserción universitaria de la UNDC, periodo 2020 – 2022**; presentada por el Bachiller (es) **Hugo Jesús Sulca Martínez** para optar al Título Profesional de Ingeniero de Sistemas.

Después de concluido el acto de sustentación y luego de que el sustentante dio respuesta a las preguntas respectivas, el Jurado Evaluador calificó la sustentación o defensa de la investigación según los siguientes criterios, y promediando las notas del jurado se obtuvo:

Tesista: Hugo Jesús Sulca Martínez						
Criterio	**Desaprobado (menor de 14)**	**Regular (14 -15)**	**Bueno (16-17)**	**Muy bueno (18-19)**	**Excelente (20)**	**Puntaje**
Claridad de la exposición		15				
Dominio del tema		15				
Visión holística del trabajo		15				
Promedio						15

Para obtener la nota final de todo el proceso de investigación, conforme al artículo 100° del Reglamento General de Investigación de la Universidad Nacional de Cañete aprobado mediante Resolución N° 180-2023-UNDC/CO, se procedió a calcular la nota final:

Tesista	**Nota del Informe (valor 40%)**	**Nota de Sustentación (valor 60%)**	**Nota Final**
Bach. Hugo Jesús Sulca Martínez	15	15	15

Y, establece como calificación FINAL: 15 (Quince)

- Aprobado, con nota Regular (X), Bueno (…), Muy Bueno (…), Excelente (…)
- Desaprobado (…)

F-M01.03.04-GT-003Toda copia de este documento, sea del entorno virtual o del documento original en físico es considerada "copia no controlada"

Por tanto, el graduado se encuentra expedito(s) (X), impedido(s) ( ) para realizar los trámites que corresponden a la obtención del Título Profesional de Ingeniero de Sistemas.

Siendo las 20 : 50 Horas se dio por terminado el acto de sustentación.

Dr. Edwin Roque Tito **Secretario**	Dr. Guido Raúl Larico Uchamaco **Presidente**	M.Sc. Leonidas Asto Huaman **Vocal**	Mg. Amanda Durán Carhuamaca **Asesor**

F-M01.03.04-GT-003Toda copia de este documento, sea del entorno virtual o del documento original en físico es considerada "copia no controlada"

Dedicatoria

El presente trabajo está dedicado a mi padre;

Bernardo Sulca y a mi madre Ana Martínez.

Así como también a mis hermanos mayores;

Percy y Franco.

Por siempre apoyarme y aconsejarme,

a pesar de mis desobediencias, nunca

me abandonaron y siempre confiaron en mí.

Agradecimientos

Agradecer a Dios por iluminarme y darme la fuerza de continuar cumpliendo mis metas, a mis padres por sus consejos, paciencia y apoyo incondicional, a mis hermanos mayores, por ser un ejemplo de profesionales. A mis docentes universitarios por las enseñanzas a lo largo de mi carrera profesional, en especial a mi asesora Amanda Duran por guiarme en la elaboración de mi tesis y al ingeniero Edwin Roque por su apoyo. También agradezco a todas las personas que me apoyaron de alguna manera en el proceso de mi tesis y formación universitaria, amigos y compañeros.

Índice de Contenido

Firma de los Jurados 2
Acta de Sustentación 2
Dedicatoria 5
Agradecimientos 6
Índice General 7
Índice de Tablas 10
Índice de Figuras 11
Resumen 12
Abstract 13
1. Capítulo I: Introducción 14
1.1. Planteamiento del Problema 14
1.2. Formulación del Problema 17
1.2.1. Problema General 17
1.2.2. Problemas Específicos 17
1.3. Objetivos de Investigación 17
1.3.1. Objetivo General 17
1.3.2. Objetivos Específicos 18
1.4. Justificación de la Investigación 18
1.4.1. Justificación Teórica 18
1.4.2. Justificación Práctica 18
1.4.3. Justificación Metodológica 19
1.5. Delimitación de la Investigación 20
1.5.1. Delimitación Teórica o Conceptual 20
1.5.2. Delimitación Temporal 20
1.5.3. Delimitación Espacial 20
2. Capítulo II: Marco Teórico 21
2.1. Antecedentes de Investigación 21
2.1.1. Artículos Científicos 21
2.1.2. Tesis Internacionales y Nacionales 24
2.2. Bases teóricas 30
2.2.1. Variable Independiente: Modelo Predictivo Basado en Machine Learning 30
2.2.2. Variable Dependiente: Diagnóstico Temprano de la Deserción Universitaria 35
2.3. Definición de Términos Básicos 40
2.3.1. Machine Learning 40

2.3.2. Algoritmo de Machine Learning....40
2.3.3. Modelo Predictivo....40
2.3.4. Regresión Logística....40
2.3.5. Diagnóstico Temprano....41
2.3.6. Deserción Universitaria....41
2.3.7. Entrenamiento....41
2.3.8. Datos de Entrenamiento....41
2.3.9. Datos de Prueba....41
2.3.10. Predicción....41
3. Capítulo III: Hipótesis y Variables....42
3.1. Hipótesis....42
3.1.1. Hipótesis General....42
3.1.2. Hipótesis Especificas....42
3.2. Definición Conceptual de las Variables....43
3.3. Operacionalización de las Variables....44
4. Capítulo IV: Metodología....45
4.1. Enfoque de la Investigación....45
4.2. Tipo de Investigación....45
4.3. Nivel de investigación....45
4.4. Método de Investigación....45
4.5. Diseño de Investigación....46
4.6. Población, Muestra y Muestreo....46
4.6.1. Población....46
4.6.2. Muestra....46
4.6.3. Muestreo....47
4.7. Técnicas e Instrumentos de Recolección de Datos....47
4.7.1. Técnicas....47
4.7.2. Instrumentos....47
4.7.3. Validez....48
4.7.4. Confiabilidad....48
4.8. Técnicas Estadísticas de Análisis de Datos....49
4.9. Consideraciones éticas....49
4.9.1. Consentimiento informado....49
4.9.2. Protección de datos personales....49
4.9.3. Respeto a la dignidad y a la diversidad....50
4.9.4. Honestidad y rigor científico....50
5. Capítulo V: Resultados....51

5.1. Resultados Descriptivos 51
5.2. Resultados Inferenciales 54
5.2.1. Prueba de Normalidad 54
5.2.2. Prueba de Hipótesis Especificas 56
5.2.3. Prueba de Hipótesis General 68
6. Capítulo VI: Discusión 71
7. Capítulo VII: Conclusiones 75
8. Capítulo VIII: Recomendaciones 76
Referencias 77
ANEXO 01 *Matriz de Consistencia* 86
ANEXO 02 *Instrumento de recolección de datos* 87
ANEXO 03 *Validación de los instrumentos de investigación* 88
ANEXO 04 *Base de datos de los indicadores* 97
ANEXO 05 *Autorización de la institución para realizar la investigación* 106
ANEXO 06 *Desarrollo de la solución tecnológica* 107
ANEXO 07 *Autorización de Inclusión de Tesis en el Repositorio de la UNDC* 111
ANEXO 08 *Constancia de similitud* 113
ANEXO 09 *Artículo científico* 115

Índice de Tablas

Tabla 1 Definición Conceptual de las Variables....43
Tabla 2 *Operacionalización de las Variables*....44
Tabla 3 *Estadística de Fiabilidad*....49
Tabla 4 *Tabla de Kolmogorov Smirnov*....55
Tabla 5 *Pruebas de Chi Cuadrado para P1*....57
Tabla 6 *Pruebas de Chi Cuadrado para P2*....58
Tabla 7 *Pruebas de Chi Cuadrado para P3*....58
Tabla 8 *Pruebas de Chi Cuadrado para P4*....59
Tabla 9 *Pruebas de Chi Cuadrado para P5*....60
Tabla 10 *Pruebas de Chi Cuadrado para P6*....60
Tabla 11 *Pruebas de Chi Cuadrado para P7*....61
Tabla 12 *Pruebas de Chi Cuadrado para P8*....61
Tabla 13 *Pruebas de Chi Cuadrado para P9*....62
Tabla 14 *Pruebas de Chi Cuadrado para P10*....63
Tabla 15 *Pruebas de Chi Cuadrado para P11*....63
Tabla 16 *Pruebas de Chi Cuadrado para P12*....64
Tabla 17 *Pruebas de Chi Cuadrado para P13*....64
Tabla 18 *Pruebas de Chi Cuadrado para P14*....65
Tabla 19 *Pruebas de Chi Cuadrado para P15*....66
Tabla 20 *Pruebas de Chi Cuadrado para P16*....66
Tabla 21 *Pruebas de Chi Cuadrado para P17*....67
Tabla 22 *Pruebas de Chi Cuadrado para P18*....67
Tabla 23 *Pruebas de Chi Cuadrado para P19*....68

Índice de Figuras

Figura 1 *Distribución de los encuestados según edad* 97
Figura 2 *Distribución por sexo de los estudiantes encuestados* 97
Figura 3 *Distribución de los estudiantes encuestados según ciclo académico* 97
Figura 4 *Distribución de los estudiantes encuestados según carrera profesional* 98
Figura 5 *Distribución de respuestas a la pregunta P1* 98
Figura 6 *Distribución de respuestas a la pregunta P2* 98
Figura 7 *Distribución de respuestas a la pregunta P3* 99
Figura 8 *Distribución de respuestas a la pregunta P4* 99
Figura 9 *Distribución de respuestas a la pregunta P5* 99
Figura 10 *Distribución de respuestas a la pregunta P6* 100
Figura 11 *Distribución de respuestas a la pregunta P7* 100
Figura 12 *Distribución de respuestas a la pregunta P8* 100
Figura 13 *Distribución de respuestas a la pregunta P9* 101
Figura 14 *Distribución de respuestas a la pregunta P10* 101
Figura 15 *Distribución de respuestas a la pregunta P11* 101
Figura 16 *Distribución de respuestas a la pregunta P12* 102
Figura 17 *Distribución de respuestas a la pregunta P13* 102
Figura 18 *Distribución de respuestas a la pregunta P14* 102
Figura 19 *Distribución de respuestas a la pregunta P15* 103
Figura 20 *Distribución de respuestas a la pregunta P16* 103
Figura 21 *Distribución de respuestas a la pregunta P17* 103
Figura 22 *Distribución de respuestas a la pregunta P18* 104
Figura 23 *Distribución de respuestas a la pregunta P19* 104
Figura 24 *Distribución de respuestas a la pregunta R1* 104
Figura 25 *Capturas del formulario de encuesta aplicado mediante Google Forms* 107
Figura 26 *Captura del total de respuestas* 107
Figura 27 *Resultados de la encuesta en Google Sheets* 108
Figura 28 *Código escrito en Python* 108
Figura 29 *Captura del código aplicado en el entorno Google Colab* 109
Figura 30 *Código documentado* 109
Figura 31 *Resultados del modelo de regresión logística* 110

Resumen

La presente investigación tuvo como objetivo implementar un modelo predictivo para el diagnóstico temprano en la deserción universitaria de la Universidad Nacional de Cañete. Para ello, se usó el modelo de regresión logística, junto al lenguaje de programación Python y el entorno de programación Google Colaboratory, se creó un código capaz de predecir si un alumno podía desertar o no su carrera universitaria, tomando como base de datos una encuesta realizada a los estudiantes universitarios por Google Forms. El modelo clasifica 80,64% correctamente de las observaciones en el conjunto de pruebas, el 82% de las veces que el modelo predice que un estudiante no va a desertar, está en lo correcto, el 71% de las veces que el modelo predice que un estudiante va a desertar, está en lo correcto. Finalmente se concluye, aunque el modelo mostró una alta precisión en la identificación de estudiantes que no van a desertar, su rendimiento fue moderado en la predicción de aquellos que sí lo harán, lo que sugiere la necesidad de continuar mejorando el modelo para incrementar su sensibilidad y precisión en ambos casos.

Palabras claves: *Inteligencia artificial, aprendizaje automático, regresión logística, deserción universitaria*

Abstract

The objective of this research was to implement a predictive model for early diagnosis of university dropout at the National University of Cañete. To achieve this, a logistic regression model was used, along with the Python programming language and the Google Colaboratory environment. A code was created capable of predicting whether a student was likely to drop out of their university career, based on a dataset from a survey conducted among university students via Google Forms. The model correctly classified 80.64% of the observations in the test set. In 82% of the cases when the model predicted that a student would not drop out, it was correct, and in 71% of the cases when the model predicted that a student would drop out, it was also correct. In conclusion, although the model demonstrated high accuracy in identifying students who would not drop out, its performance was moderate in predicting those who would, suggesting the need for further improvements to increase its sensitivity and accuracy in both cases.

Keywords*: Artificial intelligence, machine learning, logistic regression, university dropout*

Capítulo I: Introducción

1.1. Planteamiento del Problema

La deserción universitaria es un problema que afecta a muchos jóvenes y tiene consecuencias negativas para su desarrollo personal, profesional y social. Según el Ministerio de Educación, la deserción universitaria en el Perú alcanza el 30%, siendo motivada especialmente por la falta de una buena orientación vocacional y por razones económicas [1]. Además, la COVID-19 ha agravado esta situación, al generar dificultades de conectividad, bienestar y rendimiento académico.

Según el estudio: "Deserción universitaria de los estudiantes de una universidad peruana", los autores identificaron varias dimensiones que contribuyen a la deserción universitaria: factores personales, factores educativos, factores institucionales y factores financieros. Los factores personales incluyen problemas personales o familiares que pueden afectar la capacidad del estudiante para continuar sus estudios. Los factores educativos se refieren a las dificultades académicas, como el rendimiento académico bajo o la falta de interés en el curso. Los factores institucionales pueden incluir la falta de apoyo de la universidad o la insatisfacción con los servicios proporcionados. Los factores financieros se refieren a las dificultades económicas que pueden impedir que los estudiantes continúen sus estudios [2]. Estos factores pueden generar frustración, estrés y baja autoestima en los estudiantes, lo que afecta su motivación y compromiso con el aprendizaje.

Ante este escenario, se hace necesario identificar los factores que influyen en la deserción universitaria, así como predecir y prevenir los casos de riesgo. Para ello, se puede recurrir al uso de machine learning, una rama de la inteligencia artificial que permite analizar grandes volúmenes de datos y extraer patrones y conocimiento útil para la toma de decisiones.

El machine learning ha sido aplicado para el estudio de la deserción universitaria en diferentes contextos y países, obteniendo resultados prometedores. Por ejemplo, en Venezuela se evaluaron diferentes algoritmos de machine learning para predecir la deserción universitaria con variables académicas y socioeconómicas de los estudiantes [3]. En Chile se aplicaron técnicas de machine learning para analizar los factores que influyen en la deserción y el retraso académico, así como para identificar patrones de comportamiento de los estudiantes [4]. En España se utilizó inteligencia artificial y machine learning para diseñar un sistema de alerta temprana que permita detectar y prevenir la deserción universitaria [5].

En el Perú también se han realizado algunos estudios que aplican machine learning para el problema de la deserción universitaria. Por ejemplo, se diseñó un sistema para predecir la deserción de los alumnos de la Universidad Tecnológica del Perú usando machine learning y los patrones de comportamiento de los alumnos desertores [6]. También, se evaluaron diferentes algoritmos de machine learning para predecir la deserción universitaria con variables académicas y socioeconómicas de los estudiantes [3].

El problema que se plantea en esta tesis es el siguiente: ¿De qué manera la implementación de un modelo predictivo basado en machine learning permite el diagnóstico temprano en la deserción universitaria de la UNDC en periodo 2020 – 2022? Este problema surge a partir del conocimiento de la tasa de deserción universitaria en el Perú siendo de 16.2% en 2022 [7] y sus implicancias negativas para los estudiantes, las instituciones y el país. Además, se considera que el uso de machine learning puede ofrecer una solución innovadora y eficaz para abordar este problema, al aprovechar las ventajas que ofrece esta tecnología para el análisis de datos y la generación de conocimiento.

Factores Económicos

Es un factor crucial que puede influir en la decisión de un estudiante de continuar o abandonar sus estudios universitarios. Según un estudio realizado en la Universidad

Interamericana para el Desarrollo (UNID) [8], los aspectos económicos que influyen en el abandono escolar de los estudiantes incluyen la necesidad de trabajar, la pérdida de porcentaje de beca, el rezago en pagos y la generación de intereses.

Estos factores económicos pueden ser una carga significativa para los estudiantes y sus familias. Si un estudiante tiene dificultades para pagar la matrícula, es posible que decida abandonar sus estudios. Además, los costos de vida, como alojamiento, alimentación y transporte, pueden poner presión financiera sobre el estudiante y hacer que sea más difícil continuar con sus estudios.

Situación Académica

La "situación académica" es otro factor crucial que puede influir en la decisión de un estudiante de continuar o abandonar sus estudios universitarios. Según un estudio realizado en Colombia [9], la deserción universitaria se centra en la primera mitad de la carrera: al finalizar el cuarto semestre, el 39% de los estudiantes ya había abandonado sus estudios.

Esto puede deberse a varios factores relacionados con la situación académica del estudiante. Por ejemplo, las dificultades académicas, como un rendimiento académico bajo o la incapacidad para mantenerse al día con el material del curso, pueden hacer que un estudiante considere abandonar sus estudios. Además, la carga de trabajo del curso puede ser abrumadora para algunos estudiantes, lo que puede contribuir a su decisión de abandonar.

Vocación Profesional

La "vocación profesional" es un factor crucial que puede influir en la decisión de un estudiante de continuar o abandonar sus estudios universitarios. Según un estudio realizado por Xavier Erazo Guerra y Elena Rosero Morales [10], una adecuada orientación vocacional facilitará en los estudiantes el autoconocimiento de sus actitudes y aptitudes que les guiará a desempeñarse en la vida universitaria y laboral.

El objeto de la investigación fue determinar la orientación vocacional y su influencia en la deserción universitaria para la mejora de la retención estudiantil, e identificar las causas y factores para que los estudiantes abandonen sus estudios en una carrera universitaria.

Los autores encontraron que los factores externos e internos inciden en la decisión del ingreso a la universidad y pueden determinar la motivación y orientación en los bachilleres para cumplir con su proyecto de vida.

1.2. Formulación del Problema

1.2.1. Problema General

¿De qué manera la implementación de un modelo predictivo basado en machine learning influye en el diagnóstico temprano la deserción universitaria de la UNDC en periodo 2020 – 2022?

1.2.2. Problemas Específicos

¿De qué manera la implementación de un modelo predictivo basado en machine learning que incorpora los factores económicos influye en el diagnóstico temprano la deserción universitaria de la UNDC en periodo 2020 – 2022?

¿De qué manera la implementación de un modelo predictivo basado en machine learning que incorpora la situación académica influye en el diagnóstico temprano la deserción universitaria de la UNDC en periodo 2020 – 2022?

¿De qué manera la implementación de un modelo predictivo basado en machine learning que incorpora la vocación profesional influye en el diagnóstico temprano la deserción universitaria de la UNDC en periodo 2020 – 2022?

1.3. Objetivos de Investigación

1.3.1. Objetivo General

Implementar un modelo predictivo basado en machine learning que influye en el diagnóstico temprano en la deserción universitaria de la UNDC en periodo 2020 – 2022.

1.3.2. Objetivos Específicos

Implementar un modelo predictivo basado en machine learning, que incorpora los factores económicos, que influye en el diagnóstico temprano en la deserción universitaria de la UNDC en periodo 2020 – 2022.

Implementar un modelo predictivo basado en machine learning, que incorpora la situación académica, que influye en el diagnóstico temprano en la deserción universitaria de la UNDC en periodo 2020 – 2022.

Implementar un modelo predictivo basado en machine learning, que incorpora la vocación profesional, que influye en el diagnóstico temprano en la deserción universitaria de la UNDC en periodo 2020 – 2022.

1.4. Justificación de la Investigación

1.4.1. Justificación Teórica

La justificación teórica radica en la necesidad de abordar de manera más eficaz y proactiva el problema de la deserción universitaria. Este fenómeno, que afecta a las instituciones de educación superior en todo el mundo, tiene implicaciones profundas tanto para los estudiantes como para las instituciones. Al aplicar técnicas de machine learning a los datos del estudiante, podemos desarrollar modelos que pueden predecir qué estudiantes tienen más probabilidades de abandonar. Esto permite a las instituciones identificar a los estudiantes en riesgo temprano y ofrecer intervenciones dirigidas para apoyar a estos estudiantes.

1.4.2. Justificación Práctica

La justificación práctica radica en su potencial para mejorar la retención de estudiantes en las universidades. Al identificar a los estudiantes en riesgo, las universidades pueden dirigir sus recursos y esfuerzos hacia estos estudiantes, lo que puede resultar en una mayor eficiencia y efectividad en sus esfuerzos de retención. También puede tener implicaciones significativas para los propios estudiantes, al recibir intervenciones tempranas y apoyo dirigido, los estudiantes en riesgo pueden tener una mayor probabilidad de continuar con sus estudios, lo que puede resultar en mejores resultados educativos y oportunidades de carrera.

1.4.3. Justificación Metodológica

La justificación metodológica radica en la capacidad de las técnicas de machine learning para analizar grandes cantidades de datos y detectar patrones complejos. El machine learning es una rama de la inteligencia artificial que se basa en algoritmos y modelos estadísticos para realizar tareas. En el contexto de la deserción universitaria, los modelos de machine learning pueden ser entrenados en datos históricos de estudiantes para aprender los patrones y factores que contribuyen a la deserción. Una vez entrenados, estos modelos pueden ser utilizados para predecir la probabilidad de deserción de los estudiantes actuales basándose en sus características individuales. Además, el uso de machine learning permite que el modelo se actualice y mejore con el tiempo a medida que se disponga de más datos. Esto significa que el modelo puede adaptarse a cambios en los patrones de deserción y seguir siendo efectivo a medida que cambian las circunstancias.

Los resultados de la investigación servirán para la Universidad Nacional de Cañete en:

- Identificar los factores que influyen en la deserción de sus estudiantes y su peso relativo en el modelo predictivo.

- Proponer recomendaciones para prevenir y reducir la deserción de sus estudiantes, basadas en el análisis de los datos y el conocimiento generado por el modelo predictivo.
- Aportar al avance del conocimiento científico sobre la deserción universitaria y el uso de machine learning como herramienta de investigación.

1.5. Delimitación de la Investigación

La investigación se delimita de la siguiente manera:

1.5.1. Delimitación Teórica o Conceptual

Se basa en los conceptos y técnicas de machine learning, así como en los estudios previos sobre la deserción universitaria y sus factores asociados.

1.5.2. Delimitación Temporal

Se considera el periodo 2020 - 2022, que abarca desde el inicio de la emergencia sanitaria por el COVID-19 hasta el año 2022.

1.5.3. Delimitación Espacial

Se enfoca en la Universidad Nacional de Cañete, en la que empecé mis estudios universitarios el año 2017 y egresé el 2022.

Capítulo II: Marco Teórico

2.1. Antecedentes de Investigación

2.1.1. Artículos Científicos

La deserción universitaria es un problema que afecta tanto a los individuos como a la sociedad, ya que implica una pérdida de oportunidades educativas, laborales y sociales para los estudiantes que abandonan sus estudios, así como un desperdicio de recursos y talentos para el sistema educativo y el desarrollo nacional. Por ello, es importante identificar a los estudiantes que tienen un mayor riesgo de deserción y brindarles apoyo y orientación para que puedan continuar y completar su formación académica.

El machine learning ha sido aplicado para el estudio de la deserción universitaria en diferentes contextos y países, obteniendo resultados prometedores. A continuación, se presentan algunos ejemplos han utilizado esta tecnología para abordar el problema de la deserción universitaria:

Villa et al. [11] en su artículo titulado: “A methodological proposal to address the academic dropout phenomenon based on an intelligent prediction model: a case study” publicado en la revista Millenium, tuvo como objetivo diagnosticar tempranamente los posibles factores que causan la deserción en los estudiantes universitarios, usando modelos de predicción que combina el Análisis de Supervivencia, Árboles de Decisión y Random Forest, dividido en tres fases. En la primera fase se determina la probabilidad de que un estudiante permanezca en la universidad, en la segunda fase se identifica los patrones de supervivencia en relación a las variables estudiadas (calificaciones, edad, género, etc.), en la tercera fase se identifica las variables más importantes para predecir la deserción. El estudio fue aplicado en una universidad chilena donde se identificó que las puntuaciones en

lenguaje del examen de ingreso (PSU) son la variable más importante para predecir la deserción universitaria, también se observó que aproximadamente el 25% de los estudiantes abandonaron sus estudios durante el primer año. En conclusión, la combinación de Análisis de Supervivencia, Árboles de Decisión y Bosque Aleatorio en un modelo de predicción inteligente, ofrece una herramienta poderosa para abordar el problema de la deserción académica en las universidades. A pesar de que el estudio se desarrolló en el contexto de una universidad chilena, la metodología propuesta es flexible y adaptable a diferentes realidades académicas.

Valero et al. [3] en su artículo titulado: "Deserción universitaria: Evaluación de diferentes algoritmos de Machine Learning para su predicción" publicado en la revista de Ciencias Sociales (Ve), cuyo objetivo es determinar que algoritmo de Machine Learning tiene mejor desempeño para detectar la deserción universitaria. Este artículo presenta una metodología basada en machine learning para predecir la probabilidad de deserción de los estudiantes de pregrado. Se utilizaron datos históricos de 652 estudiantes y se aplicaron diferentes técnicas de preprocesamiento, selección de variables y modelado. Se compararon los resultados de varios algoritmos de machine learning y se encontró que el mejor modelo fue el de bosques aleatorios, con una precisión del 86% y una curva ROC del 0.93. En conclusión, el modelo obtenido puede ayudar a predecir en los primeros ciclos de estudios, que alumnos tienen mayor probabilidad de abandonar sus estudios, así como, alertar a la oficina de bienestar par que identifique las necesidades de los estudiantes y planifique tutorías individuales y grupales.

Shuo-chang et al. [12] en el artículo titulado: "Precision education with statistical learning and deep learning: a case study in Taiwan" en la revista "International Journal of Educational Technology in Higher Education", que tiene

como objetivo determinar cómo los datos masivos y la inteligencia artificial pueden ayudar a las universidades a comprender más precisamente los antecedentes de los estudiantes, según los cuales se pueden proporcionar tomas de decisiones adecuadas para prevenir la deserción. Para este estudio, se tomaron muestras de 3552 estudiantes de una universidad en Taiwán. Se utilizó un método de aprendizaje estadístico y un método de aprendizaje automático basado en redes neuronales profundas para predecir la probabilidad de abandono de las aulas universitarias. Los resultados revelaron que el rendimiento académico del estudiante (con respecto a la dinámica del porcentaje de clasificación de clase), las solicitudes de préstamos estudiantiles, el número de ausencias de la escuela y el número de materias alertadas predijeron con éxito si los estudiantes abandonarían la universidad o no con una tasa de precisión del 68% cuando se empleó el método de aprendizaje estadístico, y del 77% para el método de aprendizaje profundo. En esta investigación se concluye que estos resultados pueden permitir que la universidad proporcione intervenciones a los estudiantes para ayudar en la selección del curso y mejorar sus competencias según sus aptitudes, lo que podría reducir la tasa de abandono y facilitar el aprendizaje adaptativo.

Rodríguez et al. [13] presentó el artículo titulado “Deserción Universitaria: Nuevo Análisis Metodológico”. Este artículo presenta una aplicación de machine learning para predecir la deserción escolar en el sistema de educación pública de Chile, utilizando bases de datos públicas del Ministerio de Educación. Los autores compararon diferentes modelos predictivos, como árboles de decisión, redes neuronales, máquinas de vectores de soporte y bosques aleatorios, con una regresión logística tradicional. Los resultados muestran que el modelo de redes neuronales tiene el mejor desempeño, con una precisión del 86% y una sensibilidad del 82%. Los autores también ofrecen algunas recomendaciones prácticas para otros investigadores y formuladores de políticas

educativas que buscan implementar sistemas de alerta temprana para la deserción escolar. Entre ellas, destacan la importancia de tener datos actualizados y confiables, de definir claramente el concepto de deserción escolar, de seleccionar las variables más relevantes para el análisis, de evaluar los modelos predictivos con criterios objetivos y de diseñar intervenciones personalizadas para los estudiantes en riesgo.

2.1.2. Tesis Internacionales y Nacionales

Hecho la revisión de literatura científica, se tiene los siguientes antecedentes en los componentes:

Tesis Internacionales. Cuzme [14] en su tesis de maestría titulada "Modelo de aprendizaje automático para la predicción del riesgo de deserción en estudiantes de la Universidad Técnica de Manabí", cuyo objetivo es elaborar un modelo predictivo para la estimación del riesgo de deserción en estudiantes de la Universidad Técnica de Manabí mediante la utilización de técnicas de aprendizaje automático y la selección de las características de mayor relevancia en la base de datos académica de la institución. En este estudio se utilizó como metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), siguiendo sus 6 fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación. A partir de las métricas de evaluación los resultados esperados fueron calculados con el algoritmo de clasificación Random Forest que realizó una mejor predicción sobre la deserción estudiantil con un margen de error de 0.05, valorado por las métricas del área bajo la curva AUC ROC de 0.95 y el Puntaje F1 de 0.95. El autor constató la irregularidad de los estudiantes en los primeros niveles académicos, por lo cual se verificaron que muchos estudiantes desertan en estos niveles durante su carrera universitaria.

Quintero [15] presentó la tesis de maestría titulada “Diseño de un modelo predictivo para generar alertas tempranas de deserción universitaria en los programas de pregrado presenciales de la Facultad de Ingeniería de la Universidad de Antioquia”, cuyo objetivo fue diseñar un modelo predictivo que pueda identificar que estudiantes de la Facultad de Ingeniería de la Universidad de Antioquia están en riesgo de desertar, para ello usó técnicas de learning analytics y machine learning. Se emplearon los datos históricos contenidos en las bases de datos de la universidad y la información disponible en la plataforma del ICFES, y se aplicaron dos técnicas de machine learning, redes neuronales artificiales (RNA) y Xtreme gradient boosting (XG Boost). Los resultados mostraron que XG Boost tuvo el mejor desempeño con un accuracy de 74,91%, también se construyó un prototipo de software que usa el modelo entrenado y que recibe los datos de un nuevo estudiante y lo identifica como un posible desertor o no. El autor concluye que aunque el modelo predictivo desarrollado ha demostrado ser útil, aun se puede mejorar su precisión incluyendo más variables como por ejemplo la vocación profesional.

En su tesis doctoral “Minería de datos aplicada a estrategias para minimizar el rezago académico y la deserción universitaria en carreras de informática de la UNNOBA” Russo [16], busca definir y desarrollar un modelo de intervención áulica virtual que permita detectar patrones de rendimiento académico y predecir, de forma temprana, potenciales casos de deserción. La principal fuente de información fue el sistema SIU-Guaraní de la UNNOBA. Además, se emplearon informes de ingreso de todas las carreras correspondientes al período 2011-2016, analizándolos mediante la técnica de clustering. También se utilizó técnicas de ARIMA para estimar predicciones futuras basándose en datos pasados, herramientas de Business Intelligence (BI) y la Suite Pentaho para el desarrollo y la gestión de la información. La investigación reveló

que, durante el período 2010-2015, solo el 40% de los alumnos aprobó las tres materias del primer cuatrimestre requeridas para inscribirse en cuatro de las cinco asignaturas del segundo cuatrimestre. A partir de estos datos, el autor deduce que los índices de deserción y desgranamiento en estas carreras son significativamente altos durante los primeros dos años.

En su tesis "Aplicativo web para el análisis, selección y admisión de aspirantes al programa de ingeniería de sistemas de la Universidad De Cundinamarca en la extensión Facatativá utilizando modelos predictivos de minería de datos", Gómez [17] se propone crear una plataforma web que facilite el análisis y evaluación de patrones de conocimiento relacionados con los perfiles de los aspirantes al programa de Ingeniería de Sistemas. Esta herramienta tiene como finalidad respaldar las políticas de selección y admisión de la Universidad de Cundinamarca, extensión Facatativá. La investigación, de carácter aplicado y enfoque cuantitativo, estableció que el modelo de regresión logística es el método óptimo debido a su superior rendimiento en las pruebas comparativas. La esencia del proyecto radicó en la combinación de la minería de datos y el diseño de un aplicativo web, que no solo sirve para analizar las razones de deserción en la mencionada universidad, sino que también integra y visualiza métricas como el accuracy score, la curva de aprendizaje, la matriz de confusión y un árbol de decisión basado en los datos recopilados.

Tesis Nacionales

Cevallos y Barahoma [18] presentaron una investigación rigurosa en su tesis denominada “Modelo para automatizar el proceso de predicción de la deserción en estudiantes universitarios en el primer año de estudio”. En este trabajo, los autores se embarcaron en la misión de diseñar e implementar un modelo tecnológico avanzado que, mediante técnicas de análisis predictivo, busca automatizar y mejorar la precisión

en la identificación temprana de estudiantes universitarios con un alto riesgo de desertar durante su primer año académico.

Para lograr esto, adoptaron la metodología Ad-Hoc, garantizando la calidad y estructura del proceso investigativo al adherirse a las buenas prácticas recomendadas en la reconocida guía PMBOK. Un aspecto crucial de su investigación fue la identificación y definición de 20 variables críticas que sirven como indicadores representativos de los factores que más influyen en la deserción universitaria. Estas variables abarcan una gama amplia de enfoques, desde aspectos personales y socioeconómicos hasta dimensiones institucionales e interpersonales.

Tras un exhaustivo análisis y comparativa de técnicas, Cevallos y Barahoma llegaron a la conclusión de que las redes bayesianas ofrecen una precisión superior en este contexto, logrando una tasa de predicción del 67.10%. Además, identificaron que la herramienta más eficaz y adecuada para llevar a cabo este tipo de análisis predictivo es el software IBM SPSS Modeler, dada su robustez y capacidades avanzadas en el manejo y modelado de datos.

De la Cruz [19], en su investigación presentada en la tesis titulada "Diseño de un modelo predictivo basado en Machine Learning para el control de la deserción de estudiantes en la universidad Ricardo Palma", abordó un problema persistente en la educación superior: la deserción estudiantil. Con el objetivo principal de crear y afinar un modelo predictivo utilizando técnicas avanzadas de Machine Learning, la investigación se centró específicamente en la situación de los estudiantes de la Universidad Ricardo Palma.

Para llevar a cabo este ambicioso proyecto, De la Cruz optó por el algoritmo de Regresión Logística, una herramienta poderosa que no solo arroja luz sobre las causas

subyacentes de la deserción, sino que también puede predecir, con un grado de precisión, qué estudiantes tienen mayor probabilidad de abandonar sus estudios.

Sin embargo, antes de poder implementar y entrenar el algoritmo, fue esencial preparar y estructurar adecuadamente los datos disponibles. De la Cruz accedió a las bases de datos de la Universidad Ricardo Palma y, tras un proceso meticuloso, construyó un Datamart específico para consolidar la información relevante. Una vez que los datos estuvieron organizados, se almacenaron en una base de datos no relacional, específicamente Hive. Pero el proceso no terminó ahí; para garantizar un procesamiento eficiente y escalable de la información, los datos se integraron en una infraestructura de Big Data utilizando la tecnología Hadoop Distributed File System (HDFS). Este enfoque permitió no solo manejar grandes volúmenes de datos, sino también garantizar que el análisis y las predicciones se realizaran de manera oportuna y precisa.

En su detallada investigación, Albán [20] abordó un tema de creciente preocupación en el ámbito académico con su tesis: “Contribuciones a la predicción de la deserción universitaria a través de minería de datos”. El núcleo de su estudio se centra en cómo las técnicas avanzadas de minería de datos pueden enriquecer y mejorar la capacidad de prever la deserción estudiantil en el nivel universitario.

Mediante un análisis exhaustivo, Albán identificó siete factores tecnológicos determinantes que tienen un impacto negativo en la permanencia de los estudiantes en la universidad. Estos factores, que van desde el acceso y uso de tecnologías hasta el nivel de adaptación a las herramientas digitales educativas, se convirtieron en variables esenciales en los modelos predictivos que diseñó. Para tal fin, utilizó técnicas como la regresión logística y el árbol de decisión, herramientas poderosas en el ámbito de la ciencia de datos.

Adicionalmente, profundizó en cómo ciertos comportamientos y hábitos de los universitarios, en particular las adicciones relacionadas con la tecnología y medios digitales, pueden actuar como predictores significativos de la deserción. Este enfoque innovador le permitió obtener resultados notables. Utilizando el algoritmo CHAID, logró alcanzar un asombroso 97.95% de precisión en su modelo predictivo.

La obra de Albán no solo arroja luz sobre la complejidad del fenómeno de la deserción universitaria, sino que también proporciona herramientas valiosas y efectivas para abordar el problema de manera proactiva y basada en datos.

En su investigación, titulada "Business Analytics para el pronóstico de deserciones de estudiantes en el Colegio Héroes de la Breña", Lozano [21] profundiza en la importancia y el impacto del Business Analytics en la anticipación y gestión de deserciones estudiantiles en el mencionado colegio. Su objetivo principal radica en discernir cómo esta herramienta tecnológica puede influir positivamente en la predicción de tales eventos.

El enfoque de la investigación es netamente cuantitativo, con una naturaleza aplicada y un diseño pre experimental. Lozano basó su estudio en una muestra compuesta por 152 registros, de los cuales solo 106 se consideraron válidos para el análisis. Estos datos se dividieron de manera que el 30% se destinó para el entrenamiento del modelo y el 70% restante para su evaluación y prueba.

El desarrollo de los modelos de clasificación se llevó a cabo utilizando el versátil lenguaje de programación Python, específicamente bajo el entorno Anaconda. Este entorno proporciona múltiples librerías especializadas que facilitan la implementación y ajuste de algoritmos de Machine Learning. Entre los algoritmos evaluados, el Random Forest destacó particularmente, logrando una exactitud del 59% en general y un impresionante 79% en la predicción específica de posibles desertores.

Al finalizar su estudio, Lozano llegó a la conclusión de que el Business Analytics, al ofrecer resultados tan prometedores en la predicción temprana de deserciones, se presenta como una herramienta valiosa para instituciones educativas. Su implementación no solo permite identificar a tiempo a los estudiantes en riesgo, sino que también brinda la oportunidad de tomar medidas proactivas para garantizar su continuidad y éxito académico.

2.2. Bases teóricas

2.2.1. Variable Independiente: Modelo Predictivo Basado en Machine Learning

Para la presente investigación la variable independiente es: *"Modelo predictivo basado en Machine learning"*.

Definición. Un "Modelo predictivo basado en Machine Learning" es un tipo de modelo que utiliza algoritmos de aprendizaje automático para predecir resultados futuros basándose en patrones históricos y datos actuales. Estos modelos pueden aprender y mejorar a medida que se les proporciona más datos, lo que les permite hacer predicciones cada vez más precisas. Suelen seguir un proceso general que incluye la recopilación de datos, la limpieza y preparación de los datos, la selección de características, el entrenamiento del modelo, la validación del modelo y finalmente la implementación del modelo [22].

Dimensiones.

- **Tiempo**. De entrenamiento de un modelo, que es el tiempo que se necesita para entrenar un modelo en un conjunto de datos específicos [23].
- **Precisión.** es una métrica para evaluar los modelos de clasificación. Informalmente, la precisión es la fracción de predicciones que nuestro modelo acertó. Formalmente, la precisión tiene la siguiente definición: Precisión = Número de predicciones correctas / Número total de predicciones [24].

- **Pronóstico.** Se refiere a la salida de un algoritmo que ha sido entrenado en un conjunto de datos históricos. El algoritmo genera valores probables para variables desconocidas en cada registro de los nuevos datos [25].

Indicadores.

- ***Tiempo de Procesamiento del Algoritmo.*** Es el tiempo que tarda un algoritmo en completar su ejecución. Este tiempo puede variar dependiendo de varios factores, como la eficiencia del algoritmo, la velocidad del procesador y la cantidad de datos a procesar [26].
- ***Tiempo de Respuesta de una Consulta.*** Es el tiempo que tarda un sistema en responder a una consulta específica. En el contexto de las bases de datos, se refiere al tiempo que tarda en recuperar los datos solicitados [27].
- ***Porcentaje de Predicción Acertada.*** Este indicador, también conocido como precisión o exactitud, es una medida de cuántas predicciones realizadas por el modelo son correctas. Se calcula como el número total de predicciones correctas dividido por el número total de predicciones [28].
- ***Puntaje F1.*** Es una medida que combina la precisión y la exhaustividad (recall) para proporcionar una única métrica de rendimiento. Se calcula como la media armónica de la precisión y la exhaustividad [28].
- ***Accuracy (Exactitud).*** Es una métrica que mide la proporción de predicciones correctas (tanto positivas como negativas) realizadas por el modelo en relación con el total de predicciones [29].
- ***Recall (Exhaustividad).*** Es una métrica que mide la proporción de verdaderos positivos que fueron correctamente identificados por el modelo [29].

Teorías.

- ***Inteligencia Artificial.*** La inteligencia es la capacidad de las computadoras para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones, como si fuera una persona [30].

  Una IA tiene la capacidad de imitar funciones de la mente humana, como: aprendizaje, entendimiento, uso del lenguaje, etc.

  El británico Alan Turing es considerado como pionero y fundador de la rama de la Inteligencia Artificial, quien desarrolló el modelo matemático de la máquina de Turing [31].

  La inteligencia artificial ya está siendo usada en diferentes situaciones como, por ejemplo:

  - ✓ Protección contra ciber ataques
  - ✓ Reconocimiento visual
  - ✓ Mejoras del desempeño en la estrategia comercial
  - ✓ Distribución de contenido en las redes sociales
  - ✓ Traducción automática de textos
  - ✓ Ayuda en la toma de decisiones empresariales

- ***Machine Learning.*** El machine learning pertenece al sub campo de las ciencias de la computación, se encarga de construir algoritmos que funcionan a base de una colección de ejemplos de algún fenómeno. [32]

  Tiene por objetivo, realizar una aproximación a la realidad, optimizando el error y la precisión de nuestro modelo.

  En base a un registro histórico, para poder automatizarlo y relacionar los inputs y outputs en un lenguaje que podamos entender. [33]

  El termino Machine Learning fue acuñado por Arthur L. Samuel, con su investigación en la creación de una IA en torno a un juego de damas, utilizando

dos métodos: aprendizaje de memoria y aprendizaje por generalización. La computadora IBM 7094 pudo vencer al maestro de las damas Robert Nealey en 1962. [34]

Tipos de Machine Learning, Según la APD [35], los tipos de machine learning son:

- ✓ **Aprendizaje Supervisado:** La máquina aprende en base de ejemplos que le brinda el operador, con un conjunto de datos que incluye entradas y salidas deseadas. El algoritmo identifica un patrón en los datos, realiza predicciones y es corregido por el operador, hasta que se consigue un alto nivel de rendimiento y precisión.
- ✓ **Aprendizaje Sin Supervisión:**El algoritmo estudia los datos para identificar patrones, sin la intervención de un operador. El algoritmo organiza los datos a través de grupos para describir su estructura. Mientras más datos reciba, mayor será su capacidad para tomar decisiones.
- ✓ **Aprendizaje Por Refuerzo:** El algoritmo trabaja en base a procesos reglamentados. Intenta explorar diferentes posibilidades y evaluando cada resultado hasta llegar al más óptimo. Aprende a través de ensayo y error.

Tipos de algoritmos de machine learning , Según APD los tipos de algoritmos de machine learning son:

- ✓ **Algoritmos de Regresión:** El algoritmo relaciona distintas variables, enfocándose en una variable dependiente. Se utiliza principalmente en análisis estadísticos, predicción y pronóstico.
- ✓ **Algoritmos Bayesianos:** Están basados en el “teorema de Bayes”, cada valor es independiente de cualquier otro. Lo que permite predecir una clase en función de un conjunto de dado de características.

- ✓ **Algoritmos de Agrupación:** Sirven para agrupar una serie de datos no etiquetados. Los grupos están representados por la variable K. Cada dato es enviado a un K grupo según sus características.
- ✓ **Algoritmos de Árbol de Decisión:** Es un algoritmo de clasificación y regresión, similar a un diagrama de flujo. Donde en cada nodo del árbol representa una prueba en una variable específica, y cada rama es el resultado de esa prueba.
- ✓ **Algoritmos de Redes Neuronales:** Están inspirado en el cerebro y como procesa la información. Son un gran número de procesos interconectados, que trabajan al unísono para resolver problemas específicos. Se usa para analizar datos de entrada complejos, como procesos comerciales, de fabricación o problemas empresariales.
- ✓ **Algoritmos de Reducción de Dimensión:** Son algoritmos que reducen el número de variables para encontrar la información exacta.
- ✓ **Algoritmos de Aprendizaje Profundo:** Son algoritmos que automatizan el análisis predictivo. Se ejecutan a través de redes neuronales, donde pasan una representación simplificada de los datos de capa a capa.

- ***Regresión Logística.*** Este método de análisis estadístico, se emplea frecuentemente en modelado y análisis predictivo, y se aplica también en el campo del aprendizaje automático. En este tipo de análisis, la variable dependiente es finita o categórica, pudiendo ser A o B (en el caso de la regresión binaria) o una serie de opciones finitas como A, B, C o D (en el caso de la regresión multinomial). Este método se utiliza en programas de estadística para entender la relación entre la variable dependiente y una o más variables independientes, estimando probabilidades mediante una ecuación de regresión logística [36].

Tipos de Regresión Logística: Según IBM hay tres modelos de regresión logística, que se definen en función de una .respuesta categórica.

- ✓ **Binario:** La respuesta o variable dependiente es de naturaleza dicotómica, es decir. sólo tiene dos resultados posibles.
- ✓ **Multinomial:** La variable dependiente tiene tres o más resultados posibles; sin embargo, estos valores no tienen un orden específico.
- ✓ **Ordinal:** Cuando la variable de respuesta tiene tres o más resultados posibles, pero en este caso, estos valores sí tienen un orden definido.

2.2.2. Variable Dependiente: Diagnóstico Temprano de la Deserción Universitaria

Para la presente investigación la variable dependiente es: *"Diagnóstico temprano de la deserción universitaria"*

Definición. El "Diagnóstico temprano de la deserción universitaria" se refiere a la identificación anticipada de los estudiantes que están en riesgo de abandonar sus estudios. Este diagnóstico se realiza a través del análisis de diversos factores como el rendimiento académico, la asistencia a clases, la participación en actividades extracurriculares, entre otros. El objetivo es intervenir a tiempo y proporcionar el apoyo necesario para prevenir la deserción. Este diagnóstico temprano puede ser crucial para mejorar las tasas de retención y graduación en las instituciones de educación superior [37].

Dimensiones.

Factores Económicos. Son las condiciones y recursos financieros que pueden influir en su capacidad para mantenerse y completar su educación superior.

Situación Académica. El rendimiento académico es un nivel de conocimientos demostrado en un área o materia comparado con la norma de edad y nivel académico.

Vocación Profesional. Se refiere a esa inclinación natural o llamado interior que una persona experimenta hacia un campo específico de trabajo o actividad. Es más que simplemente elegir una profesión; implica descubrir aquello que nos interesa, en lo que somos buenos y que nos permite sentirnos realizados y plenos en nuestra vida laboral [38].

Indicadores

Carga Financiera. Es la suma de los distintos gastos financieros y la amortización del capital, referida a los préstamos formalizados y pendientes de reembolso [39].

Falta de Recursos. Se refiere a la escasez o insuficiencia de recursos básicos como agua, alimentos, energía, vivienda, etc., que se consideran fundamentales para satisfacer la supervivencia o de recursos no básicos que satisfacen distintas necesidades en las sociedades humanas en varios aspectos [40].

Financiamiento de Estudios Universitarios. Se refiere a los recursos financieros que se utilizan para cubrir los costos asociados con la educación universitaria. Esto puede incluir matrícula, alojamiento, libros, suministros y otros gastos relacionados [41].

Rendimiento Académico. Es una evaluación del conocimiento adquirido en el ámbito escolar, terciario o universitario. Un estudiante con buen rendimiento académico es aquel que obtiene calificaciones positivas en los exámenes que debe rendir a lo largo de una cursada [42].

Dificultad de los Programas de Estudio. Se refiere a los desafíos y obstáculos que pueden surgir al tratar de cumplir con los requisitos y expectativas de un programa de estudio específico [43].

Apoyo Académico. Es el soporte proporcionado a los estudiantes para ayudarles a tener éxito en sus estudios. Esto puede incluir tutorías, asesoramiento académico, servicios de orientación y otros recursos disponibles para ayudar a los estudiantes a alcanzar sus objetivos académicos [44].

Tipo de Preparación. Este término se refiere a la formación previa que un estudiante ha recibido antes de ingresar a un programa universitario específico. Esto puede incluir la educación secundaria, cursos preuniversitarios o cualquier otra forma de preparación académica [45].

Cursos Matriculados. Se refiere a los cursos en los que un estudiante se ha inscrito durante un semestre o año académico específico [46].

Teorías

Deserción Estudiantil. La deserción estudiantil es un fenómeno social y educativo que se refiere al abandono definitivo o temporal de los estudios por parte de los estudiantes, antes de obtener el título académico correspondiente, debido a diversos factores de índole personal, familiar, económica, académica o institucional. La deserción estudiantil tiene consecuencias negativas tanto para los individuos como para la sociedad, ya que limita las oportunidades de desarrollo personal y profesional, y afecta el capital humano y el crecimiento económico de un país [47].

Causas de la Deserción Estudiantil: La deserción estudiantil puede ocurrir en cualquier nivel educativo, pero se ha prestado especial atención al nivel universitario, debido a su importancia para la formación de profesionales calificados y competentes que puedan responder a las demandas del mercado laboral y a los desafíos del siglo XXI. Según datos de la UNESCO, la tasa promedio de deserción universitaria en América Latina y el Caribe es del 50%, lo que significa que la mitad de los estudiantes que ingresan a la educación superior no la concluyen.

Existen múltiples causas que explican la deserción universitaria, las cuales pueden agruparse en cuatro categorías: causas individuales, familiares, académicas e institucionales [47].

- ✓ **Causas Individuales:** Se refieren a las características personales, psicológicas, motivacionales y vocacionales de los estudiantes, que influyen en su decisión de continuar o abandonar sus estudios. Algunos ejemplos son: falta de interés, bajo rendimiento, problemas de salud, embarazo, consumo de drogas, dificultades de adaptación o integración social, falta de habilidades o competencias para el estudio, falta de claridad sobre el proyecto de vida o la elección profesional, etc.
- ✓ **Causas Familiares:** Se refieren al contexto familiar y socioeconómico de los estudiantes, que condiciona sus posibilidades y expectativas educativas. Algunos ejemplos son: bajo nivel educativo o cultural de los padres, falta de apoyo o supervisión familiar, problemas económicos o laborales, conflictos o violencia familiar, migración o movilidad geográfica, etc.
- ✓ **Causas Académicas:** Se refieren al currículo, la metodología, la evaluación y el clima educativo que se desarrolla en la institución universitaria, que afecta el proceso de enseñanza-aprendizaje y la satisfacción estudiantil. Algunos ejemplos son: baja calidad académica o pertinencia del programa de estudios, dificultad o exigencia de las asignaturas, falta de orientación o asesoría académica, métodos de enseñanza tradicionales o poco innovadores, sistemas de evaluación poco flexibles o transparentes, falta de comunicación o retroalimentación entre docentes y estudiantes, etc.
- ✓ **Causas Institucionales:** Se refieren a la organización, la gestión, la infraestructura y los servicios que ofrece la institución universitaria, que

inciden en el bienestar y el desarrollo integral de los estudiantes. Algunos ejemplos son: falta de recursos financieros o materiales para el funcionamiento adecuado de la universidad, falta de políticas o programas para prevenir o atender la deserción estudiantil, falta de becas o ayudas económicas para los estudiantes con necesidades especiales, falta de espacios físicos o virtuales para el estudio o la recreación, falta de servicios complementarios como biblioteca, transporte, salud, deporte, cultura, etc.

Deserción Universitaria: Es el abandono de la carrera universitaria elegida por el estudiante. Como resultado de un conflicto interno o externo que termina en abandonar sus estudios universitarios. [48]

Existen diferentes factores que interactúan con los estudiantes para que tomen la decisión de desertar, que son cambiantes según la región y momento económico. [49]

Los enfoques que se destacan en el estudio de la deserción estudiantil universitaria, son:

- ✓ **Enfoques Psicológicos:** Fijan su atención en condiciones personales internas que actúan en el comportamiento y lleva a que los individuos eviten o propicien situaciones o eventos.
- ✓ **Enfoques Sociológicos o Ambientales:** Marcan el papel que tienen las fuerzas externas (estatus social, raza, prestigio institucional) para la persistencia estudiantil
- ✓ **Enfoques Económicos:** La incapacidad o capacidad que tiene el estudiante para solventar los gastos de sus estudios universitarios.
- ✓ **Enfoques Organizacionales:** Consideran la deserción como el reflejo del impacto que tiene la organización sobre la socialización y satisfacción del estudiante.

- ✓ **Enfoques Interaccional:** La deserción depende de la manera en que cada estudiante interpreta las experiencias vividas en la universidad, así como el grado que se da entra la institución y los individuos.

2.3. Definición de Términos Básicos

Machine Learning

Es una rama de la inteligencia artificial que se encarga de crear sistemas capaces de aprender de forma automática a partir de los datos, sin necesidad de una programación explícita. El aprendizaje automático se puede clasificar en dos tipos: supervisado y no supervisado. El aprendizaje supervisado consiste en entrenar al sistema con datos etiquetados, es decir, con la respuesta correcta para cada caso. El aprendizaje no supervisado consiste en dejar que el sistema encuentre patrones o estructuras ocultas en los datos sin etiquetar [22].

Algoritmo de Machine Learning

Es una secuencia de instrucciones o reglas que le indican al sistema cómo procesar los datos y extraer patrones o conocimiento de ellos [50].

Modelo Predictivo

Es un modelo matemático que permite estimar la probabilidad de que ocurra un evento futuro, basándose en datos históricos y actuales, y aplicando técnicas de análisis estadístico o de aprendizaje automático [51]. En este estudio se utilizó el modelo de Regresión Logística.

Regresión Logística

La regresión logística calcula la probabilidad de que un evento ocurra, como participar o no en una votación, considerando un conjunto específico de variables independientes. En el ámbito del aprendizaje automático, la regresión logística forma parte de los modelos de aprendizaje supervisado [36].

Diagnóstico Temprano

Es el proceso de identificar y reconocer los signos o síntomas de una condición o problema antes de que se agrave o se vuelva crónico. El diagnóstico temprano permite intervenir de forma oportuna y efectiva para prevenir o mitigar las consecuencias negativas del problema [52].

Deserción Universitaria

Es el fenómeno social y educativo que se refiere al abandono definitivo o temporal de los estudios por parte de los estudiantes universitarios, antes de obtener el título académico correspondiente, debido a diversos factores de índole personal, familiar, económica, académica o institucional. La deserción universitaria tiene implicaciones negativas tanto para los individuos como para la sociedad, ya que limita las oportunidades de desarrollo personal y profesional, y afecta el capital humano y el crecimiento económico de un país [47].

Entrenamiento

Es el proceso por el que se ajustan los parámetros del algoritmo de machine learning para optimizar su rendimiento y precisión [22].

Datos de Entrenamiento

Son los datos que se usan para entrenar al algoritmo de machine learning. Deben ser representativos, variados y de calidad [22].

Datos de Prueba

Son los datos que se usan para evaluar el rendimiento y la generalización del algoritmo de machine learning una vez entrenado [22].

Predicción

Es la salida que genera el modelo predictivo basado en machine learning al recibir una entrada nueva o desconocida [22].

Capítulo III: Hipótesis y Variables

3.1. Hipótesis

3.1.1. Hipótesis General

La implementación de un modelo predictivo basado en machine learning influye en el diagnóstico temprano de la deserción universitaria de la UNDC en periodo 2020 – 2022.

3.1.2. Hipótesis Especificas

La implementación de un modelo predictivo basado en machine learning que incorpora los factores económicos influye en el diagnóstico temprano de la deserción universitaria de la UNDC en periodo 2020 – 2022.

La implementación de un modelo predictivo basado en machine learning que incorpora la situación académica influye en el diagnóstico temprano de la deserción universitaria de la UNDC en periodo 2020 – 2022.

La implementación de un modelo predictivo basado en machine learning que incorpora la vocación profesional influye en el diagnóstico temprano de la deserción universitaria de la UNDC en periodo 2020 – 2022.

3.2. Definición Conceptual de las Variables

Tabla 1

Definición Conceptual de las Variables

Variables	Definición conceptual	Definición operacional
Modelo Predictivo basado en Machine Learning	Es un tipo de modelo que utiliza algoritmos de aprendizaje automático para predecir resultados futuros basándose en patrones históricos y datos actuales. Estos modelos pueden aprender y mejorar a medida que se les proporciona más datos, lo que les permite hacer predicciones cada vez más precisas [22].	Conjunto de algoritmos matemáticos implementados en Python, que procesan datos históricos y actuales de los estudiantes para identificar patrones y predecir con un 80.64% de precisión el riesgo de deserción universitaria. Este modelo utiliza Regresión Logística como técnica principal, evaluando factores económicos, académicos y vocacionales.
Diagnóstico temprano de la deserción universitaria	Se refiere a la identificación anticipada de los estudiantes que están en riesgo de abandonar sus estudios. Este diagnóstico se realiza a través del análisis de diversos factores como el rendimiento académico, la asistencia a clases, la participación en actividades extracurriculares, entre otros. El objetivo es intervenir a tiempo y proporcionar el apoyo necesario para prevenir la deserción [37].	Probabilidad de que un estudiante abandone sus estudios en función de factores específicos. Se mide mediante un cuestionario estructurado aplicado a una muestra representativa, analizando las respuestas relacionadas con las dimensiones económicas, académicas y vocacionales.

3.3. Operacionalización de las Variables

Tabla 2

Operacionalización de las Variables

Variables	Dimensiones	Indicadores	Items o fórmulas	Instrumento y escala de medición
VI: Modelo Predictivo basado en Machine Learning	Tiempo	- Tiempo de procesamiento del algoritmo	$T_{procesamiento} = T_{final} - T_{inicio}$	Software de análisis y métricas de machine learning
		- Tiempo de respuesta de una consulta	$T_{respuesta} = T_{final} - T_{inicio}$	
	Precisión	- Porcentaje de predicción acertada	$\text{Porcentaje de predicción acertada} = \left(\frac{\text{Número de predicciones correctas}}{\text{Total de predicciones}}\right) \times 100$	
		- Puntaje F1	$\text{Puntaje F1} = 2 \times \frac{\text{Precisión} \times \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}}$	
		- Accuracy (Exactitud)	$\text{Accuracy} = \frac{\text{Verdaderos Positivos(TP)} + \text{Verdaderos Negativos(TN)}}{\text{Total predicciones}}$	
	Pronostico	- Recall (Exhaustividad)	$\text{Sensibilidad(Recall)} = \frac{\text{Verdaderos Positivos(TP)}}{\text{Verdaderos Positivos(TP)} + \text{Falsos Negativos(FN)}}$	
VD: Diagnóstico temprano de la deserción universitaria	Factores económicos	- Carga financiera - Falta de recursos - Financiamiento de estudios universitarios	1 – 2 – 3 – 4 – 5 – 6	Cuestionario (Escala de Likert) 1: Nada 2: Poco 3: Bastante 4: Mucho
	Situación académica	- Rendimiento académico - Dificultad de los programas de estudio - Apoyo académico	7 – 8 – 9 – 10 – 11 – 12 – 13 – 14 – 15	
	Vocación profesional	- Tipo de preparación - Cursos matriculados	16 – 17 – 18 – 19 – 20	

Capítulo IV: Metodología

4.1. Enfoque de la Investigación

La presente investigación se basa en un enfoque cuantitativo, el enfoque cuantitativo se basa en la recopilación y análisis de datos empíricos utilizando métodos estandarizados y estructurados. Según el Creswell destaca que la investigación cuantitativa se centra en la medición de variables y la comprobación de hipótesis mediante la aplicación de técnicas estadísticas [53].

4.2. Tipo de Investigación

La naturaleza de la investigación es de tipo aplicada, la investigación aplicada se refiere a la realización de estudios que buscan resolver problemas específicos en el mundo real o aplicar conocimientos teóricos a situaciones prácticas. La investigación aplicada tiene como objetivo la solución de problemas concretos mediante la aplicación de métodos y técnicas de investigación científica [54].

4.3. Nivel de Investigación

El nivel de investigación planteado será el nivel explicativo. Según [55], las investigaciones explicativas consideran la manipulación de la variable independiente (VI) que es la causa sobre las variables dependientes (VD) que es el efecto, en una realidad especifica determinada.

4.4. Método de Investigación

En el marco de la presente investigación, centrada en el diseño de un Modelo Predictivo basado en Machine Learning para el diagnóstico temprano en la deserción universitaria, se ha adoptado el método hipotético deductivo. Es un proceso científico que comienza con una hipótesis, una explicación tentativa de un fenómeno, la cual debe ser comprobable y falsable. [56].

Grove y Menton [57], lo describen como un enfoque formalizado para probar teorías científicas dentro de la filosofía de la ciencia. Este método inicia con la formulación de hipótesis derivadas deductivamente a partir de principios o teorías generales, las cuales luego se someten a contrastación empírica mediante observaciones o experimentos. El objetivo es evaluar la capacidad de las hipótesis para predecir y explicar fenómenos dentro de un marco controlado.

4.5. Diseño de Investigación

El diseño de investigación utilizado fue Pre experimental. El diseño pre experimental generalmente utilizado en estudios exploratorios o evaluaciones iniciales donde los recursos para un control riguroso son limitados. Es útil en contextos donde se necesita una evaluación preliminar de un programa o intervención, aunque no ofrecen la misma robustez metodológica que los diseños experimentales o cuasi experimentales [58].

A diferencia de los diseños experimentales y cuasi experimentales, los diseños pre experimentales carecen de aleatorización y de un grupo de control sólido, lo que reduce su validez interna y dificulta descartar la influencia de variables externas. Son adecuados principalmente para evaluaciones exploratorias o preliminares [54].

4.6. Población, Muestra y Muestreo

4.6.1. Población

Registros socio académico vocacionales de todos los estudiantes de la UNDC. Representados por 1 764.

4.6.2. Muestra

Para la muestra se usa la siguiente formula:

$$n = \frac{Z^2 \times p \times q \times N}{e^2 \times (N-1) + Z^2 \times p \times q}$$

Donde:

N es el tamaño de la población finita (1764)

Z es el valor correspondiente al nivel de confianza (para un 95%, Z = 1,96)

p es la proporción esperada (0,5)

q es 1 – p (también 0,5 en este caso)

e es el margen de error (5%)

El tamaño de la muestra para la población de 1764 registros socio académicos, con un nivel de confianza del 95% y un margen de error del 5%, es **316 estudiantes.**

4.6.3. Muestreo

En esta investigación se utilizará un muestreo no probabilístico por conveniencia, ya que se incluyeron todos los integrantes de la muestra.

4.7. Técnicas e Instrumentos de Recolección de Datos

4.7.1. Técnicas

Técnica de encuesta. En la presente investigación se utilizará la técnica de la encuesta donde se determinará las características de los estudiantes que desertaron en la universidad. Para [59], la técnica de encuesta es un método de investigación que utiliza un conjunto de procedimientos estandarizados para recoger y analizar una serie de datos de una muestra de casos representativa de una población o universo más amplio. Esta técnica permite explorar, describir, predecir y/o explicar una serie de características.

Validación Cruzada (Cross-validation). Es una herramienta computacional que facilita la evaluación cuantitativa de modelos de Machine Learning en función de diversos criterios y métricas, permitiendo a los investigadores y científicos de datos comparar y optimizar sus modelos de forma sistemática.

4.7.2. Instrumentos

Instrumentos de Encuesta: Cuestionario. Para la presente investigación se usará el cuestionario, el cuál contendrá preguntas relacionadas a la variable dependiente

y dimensiones. Un “Cuestionario” es una herramienta de investigación que consta de una serie de preguntas diseñadas para recopilar información de los encuestados. Los cuestionarios pueden ser utilizados para recopilar datos cuantitativos y cualitativos, y son comúnmente utilizados en la investigación social y de mercado [60].

4.7.3. Validez

La validez de los instrumentos de recolección de datos se refiere a la capacidad de un instrumento de medición o recolección de datos para medir o recopilar lo que se desea medir. En otras palabras, la validez es la precisión con la que un instrumento mide lo que se supone que debe medir. Si un instrumento no es válido, los datos que se recolectan no serán precisos o representativos [61].

Es por ello que la validez de los instrumentos de recolección de datos se hizo mediante el juicio de expertos como se muestra en el ANEXO 3, se solicitó la opinión de individuos que poseen conocimientos y experiencia en el tema en cuestión, docentes con el grado de Doctor y/o Magister de la Universidad Nacional de Cañete.

4.7.4. Confiabilidad

La confiabilidad de un instrumento se refiere al grado en el que un instrumento de medición proporciona resultados consistentes [62].

En este estudio, se evaluó la confiabilidad mediante el cálculo del coeficiente Alfa de Cronbach, ampliamente reconocido en la investigación científica para medir la consistencia interna de los ítems que conforman un instrumento de medición.

La confiabilidad del instrumento fue evaluada mediante el cálculo del coeficiente **Alfa de Cronbach**, obteniendo un valor de **0.704**, lo cual indica un nivel aceptable de consistencia interna entre los ítems que componen el cuestionario. Este resultado respalda la fiabilidad del instrumento para medir de manera consistente las dimensiones evaluadas: económica, académica y vocación profesional. El valor

obtenido está en el rango recomendado para instrumentos en investigaciones sociales y educativas, validando su capacidad para generar resultados consistentes y replicables en contextos similares.

Tabla 3

Estadística de Fiabilidad

Alfa de Cronbach	Número de elementos
0,704	20

4.8. Técnicas Estadísticas de Análisis de Datos

Pruebas de hipótesis: Para contrastar la hipótesis nula de que no existe una relación significativa entre las variables académicas y socioeconómicas de los estudiantes y la probabilidad de deserción, y la hipótesis alternativa de que sí existe una relación significativa. Se utilizará la Prueba T, para comparar la media de dos grupos de datos y determinar si son significativamente diferentes entre sí.

4.9. Consideraciones éticas

4.9.1. Consentimiento informado

Se solicitará el consentimiento informado de los estudiantes que participen en la investigación, explicándoles el propósito, los objetivos, los procedimientos, los beneficios, los riesgos y las garantías de confidencialidad y anonimato de la investigación. Se les informará que su participación es voluntaria y que pueden retirarse en cualquier momento sin consecuencias negativas.

4.9.2. Protección de datos personales

Se protegerán los datos personales de los estudiantes, utilizando códigos numéricos o alfanuméricos para identificarlos, y almacenando los datos en medios seguros y con acceso restringido. Se evitará el uso de datos sensibles o innecesarios

para la investigación, y se respetará el derecho de los estudiantes a acceder, rectificar, cancelar u oponerse al tratamiento de sus datos.

4.9.3. Respeto a la dignidad y a la diversidad

Se respetará la dignidad y la diversidad de los estudiantes, evitando cualquier forma de discriminación, estigmatización, coacción o daño físico o psicológico. Se reconocerá la autonomía y la libertad de los estudiantes para tomar sus propias decisiones sobre su trayectoria académica, y se les brindará apoyo y orientación si lo requieren.

4.9.4. Honestidad y rigor científico

Se actuará con honestidad y rigor científico en todas las fases de la investigación, evitando el plagio, la falsificación, la fabricación o la manipulación de los datos. Se citarán adecuadamente las fuentes utilizadas, siguiendo las normas IEEE. Se reportarán los resultados con transparencia y objetividad, reconociendo las limitaciones y las implicaciones éticas del estudio.

Capítulo V: Resultados

Se diseñó una encuesta como instrumento de recolección de datos y fue validada por juicio de expertos para asegurar su validez y precisión. Posteriormente, la encuesta fue implementada en Google Forms y, con la colaboración de la unidad de TI de la Universidad Nacional de Cañete, se publicó en el sistema académico SIVIRENO. A los estudiantes que accedían al SIVIRENO para consultar sus notas se les solicitaba como paso previo, completar la encuesta.

Una vez finalizado el periodo de recolección de datos, las respuestas se exportaron a una hoja de cálculo para realizar una limpieza inicial de datos, eliminando entradas incompletas o inconsistentes. Finalmente, los datos limpios fueron preparados para su procesamiento y análisis mediante el algoritmo de regresión logística.

5.1. Resultados Descriptivos

Como se observa en el ANEXO E, la muestra del estudio incluye 316 participantes, con una distribución de género de 43,5% mujeres y 56,5% hombres. Participan todos los ciclos y carreras universitarias.

La pregunta N° 1: ¿En qué medida dependes económicamente de alguien más?, los estudiantes han respondido: “No dependo de otros” un 16,3%, “Dependo parcialmente de otras personas” un 35,8%, “Dependo bastante” un 21,5%, “Dependo totalmente de otras personas” un 26,5%.

Pregunta N° 3: Si tienes un empleo, ¿cómo influye en tu rendimiento académico?, los estudiantes han respondido: “No influye en mi rendimiento académico” un 5%, “Influye poco en mi rendimiento académico” un 30,8%, “Influye bastante en mi rendimiento académico” un 20,5%, “Influye mucho en mi rendimiento académico” un 7%, “No tengo empleo” un 36,8%.

Pregunta N° 3: Si tienes responsabilidades familiares, ¿cómo afecta esto a tus estudios?, los estudiantes han respondido: “No afecta en mis estudios” un 22,3%, “Afecta poco en mis estudios” un 46,3%, “Afecta poco en mis estudios” un 25,5%, “Afecta mucho en mis estudios” un 6%.

Pregunta N° 4: ¿Cuánto impacta tu situación económica en la posibilidad de continuar con tus estudios?, los estudiantes han respondido: “No impacta en nada mi situación económica” un 12,8%, “Impacta poco” un 39,3%, “Impacta bastante” un 33%, “Impacta mucho mi situación económica para seguir estudiando” un 15%.

Pregunta N° 5: ¿En qué medida influye tu salud en tu capacidad para continuar con tus estudios?, los estudiantes han respondido: “No influye en nada” un 25%, “Influye poco” un 34,3%, “Influye bastante” un 27%, “Influye mucho mi estado de salud para seguir estudiando” un 13,7%.

Pregunta N° 6: Si tienes una pareja, ¿cómo afecta esto a tu desempeño académico?, los estudiantes han respondido: “No afecta en mi desempeño académico” un 50,5%, “Afecta poco” un 17,3% “Afecta bastante” un 5,2%”, “Afecta mucho en mi desempeño académico” un 3,5%, “No tengo pareja” un 50,5%.

Pregunta N° 7: ¿Las otras actividades que desempeñas afectan tu tiempo de estudio?, los estudiantes han respondido: “No afectan en nada a mis estudios” un 19,5%; “Afectan poco” un 55,7%; “Afectan poco” un 20,8%; “Afectan mucho” un 4%.

Pregunta N° 8: Si experimentas violencia familiar, ¿cómo afecta esto a tu rendimiento académico?, los estudiantes han respondido: “No afecta en mi rendimiento académico” un 42,8%; “Afecta poco” un 21,3%; “Afecta bastante” un 24,7%; “Afectan mucho” un 11,3%.

Pregunta N° 9: ¿En qué medida la interacción con tus profesores afecta tu desempeño académico?, los estudiantes han respondido: “No afecta en mi desempeño

académico" un 27,8%; "Afecta poco" un 31,5%; "Afecta bastante" un 30,3%; "Afecta mucho" un 10.5%.

Pregunta N°10: ¿Cómo afecta el acoso de tus compañeros a tu rendimiento académico? los estudiantes han respondido: "No afecta en mi rendimiento académico" un 49,5%; "Afecta poco" un 25%, "Afecta bastante" un 17,8%; "Afecta mucho" un 7,8%.

Pregunta N° 11: ¿La distancia entre tu hogar y la universidad afecta tu asistencia a clases?, los estudiantes han respondido: "No afecta a mi asistencia a clases" un 21%, "Afecta poco" un 32,5%, "Afecta bastante" un 26,5%, "Afecta mucho" un 20%.

Pregunta N° 12: ¿Te sientes cómodo con tus compañeros de clase?, los estudiantes han respondido: "No me siento cómodo" un 6%, "Me siento un poco cómodo" un 24,8%, "Me siento un poco cómodo" un 42%, "Me siento muy cómodo con mis compañeros de clase" un 27,2%.

Pregunta N° 13: ¿Has experimentado discriminación en la universidad?, los estudiantes han respondido: "No he experimentado discriminación en la universidad" un 76,3%; "He experimentado poca discriminación en la universidad" un 20,8%; "He experimentado bastante discriminación en la universidad" un 2%; "He experimentado mucha discriminación en la universidad" un 1%.

Pregunta N° 14: ¿Crees que dedicas suficiente tiempo al estudio de tus materias?, los estudiantes han respondido: "No he dedicado tiempo a mis estudios" un 3%, "He dedicado poco tiempo a mis estudios" un 48,2%; "He dedicado bastante tiempo a mis estudios" un 41,8%; "He dedicado mucho tiempo a mis estudios" un 7%.

Pregunta N° 15: ¿Consideras que tu carrera te facilitará encontrar empleo?, los estudiantes han respondido: "No me facilitará encontrar empleo" un 5,8%; "Me facilitará un poco encontrar empleo" un 34%; "Me facilitará bastante encontrar empleo" un 44,5%; "Me facilitará mucho encontrar empleo" un 15,7%.

Pregunta N° 16: ¿Tus expectativas académicas se cumplieron después de salir del colegio?, los estudiantes han respondido: “No se cumplieron mis expectativas académicas” un 20,5%; “Se cumplieron un poco” un 46%; “Se cumplieron bastante” un 26,8%; “Se cumplieron totalmente mis expectativas académicas” un 6,8%.

Pregunta N° 17: ¿Los conocimientos adquiridos en la universidad te son útiles en otros aspectos de tu vida?, los estudiantes han respondido**,** los estudiantes han respondido: “No me son útiles en otros aspectos de mi vida” un 2,8%; “Son poco útiles en otros aspectos de mi vida” un 27%; “Son bastante útiles en otros aspectos de mi vida” un 51,5%; “Son muy útiles en otros aspectos de mi vida” un 18,8%.

Pregunta N° 18: ¿La influencia de tu familia afectó tu elección de carrera?, los estudiantes han respondido: “No influencio mi familia en la elección de mi carrera” un 30,5%; “Influencio un poco” un 36,3%; “Influencio bastante” un 21%; “Influencio mucho mi familia en la elección de mi carrera”.

Pregunta N° 19: ¿Tenías en mente estudiar una carrera diferente a la que finalmente elegiste?, los estudiantes han respondido: “No tenía pensado en otra carrera” un 18.5%; “Tenía pensado pocas opciones” un 44,8%; “Tenía pensado bastantes opciones” un 16,8%; “Tenía pensado muchas otras opciones” un 20%.

La última pregunta: “¿Has pensado en desertar de tu carrera universitaria?”, los estudiantes han respondido “No lo tengo pensado” un 92%; mientras que un 8% ha respondido “Estoy decidido en desertar mi actual carrera universitaria".

Los datos muestran que en general, los estudiantes no desean desertar de su actual carrera.

5.2. Resultados Inferenciales

5.2.1. Prueba de Normalidad

Para determinar si los datos siguen una distribución normal, es fundamental realizar una prueba de normalidad.

Si n >= 50, se emplea el método de Kolmogorov – Smirnov

Si n <= 50, se emplea el método de Shapiro – Wilk

La muestra tiene más de cincuenta datos, el método de Kolmogorov – Smirnov es la herramienta ideal para analizar su distribución.

Se plantea los siguientes criterios de decisión:

Hipótesis nula (H_0): Los datos siguen una distribución normal.

Hipótesis alternativa (H_1): Los datos no siguen una distribución normal.

Si el *p* – valor es menor que 0,05 se rechaza H_0 (los datos no son normales)

Si el *p* – valor es mayor o igual a 0,05 se rechaza H_0 (los datos son normales)

Tabla 4

Tabla de Kolmogorov Smirnov

Números de variables	Valor Estadístico	Grados de Libertad	Significancia (*p*)
P1	0,221	464	< 0,001
P2	0,243	464	< 0,001
P3	0,244	464	< 0,001
P4	0,226	464	< 0,001
P5	0,213	464	< 0,001
P6	0,328	464	< 0,001
P7	0,304	464	< 0,001
P8	0,250	464	< 0,001

Números de variables	Valor Estadístico	Grados de Libertad	Significancia (*p*)
P9	0,195	464	< 0,001
P10	0,294	464	< 0,001
P11	0,193	464	< 0,001
P12	0,233	464	< 0,001
P13	0,446	464	< 0,001
P14	0,310	464	< 0,001
P15	0,228	464	< 0,001
P16	0,257	464	< 0,001
P17	0,265	464	< 0,001
P18	0,223	464	< 0,001
P19	0,265	464	< 0,001
R1	0,488	464	< 0,001

Un valor $p < 0,001$ significa que se rechaza la hipótesis nula (H_0), quiere decir que los datos no siguen una distribución normal.

Las respuestas de las variables analizadas no se distribuyen normalmente, por lo tanto, se debe utilizar pruebas estadísticas no paramétricas para analizar estas variables.

5.2.2. Prueba de Hipótesis Especificas

Se aplica Chi cuadrado a todas las variables.

P1: ¿En qué medida dependes económicamente de alguien más?

Se plantea los siguientes criterios de decisión:

H_0: No existe una relación significativa entre las respuestas a P1 y la deserción (R1).

H_1: Existe una relación significativa entre las respuestas a P1 y la deserción (R1).

Tabla 5

Pruebas de Chi Cuadrado para P1

	Valor	**Grado de Libertad**	**Significancia asintótica (bilateral)**
Chi cuadrado de Pearson	20,643	3	< 0,001
Razón de verosimilitud	20,117	3	< 0,001
Asociación lineal por lineal	18,463	1	< 0,001
N de casos válidos	464		

La interpretación de resultados es la siguiente:

Si $p < 0,05$: Rechaza H_0 y concluye que existe una relación significativa entre P1 y R1.

Si $p \geq 0,05$: No rechaza H_0, lo que indica que no hay evidencia suficiente de una relación significativa.

Los resultados de la prueba Chi-cuadrado (X^2) para P1 muestran un $p < 0,001$, lo cual es menor que el nivel de significancia ($\alpha = 0,05$). Por lo tanto, se rechaza la

hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P1 y la deserción universitaria (R1).

P2: Si tienes un empleo, ¿cómo influye en tu rendimiento académico?

Tabla 6

Pruebas de Chi Cuadrado para P2

	Valor	**Grado de Libertad**	**Significancia asintótica (bilateral)**
Chi cuadrado de Pearson	37,454	4	< 0,001
Razón de verosimilitud	39,811	4	< 0,001
Asociación lineal por lineal	2,490	1	0,115
N de casos válidos	464		

Los resultados de la prueba Chi-cuadrado (X^2) para P2 muestran un $p < 0,001$, lo cual es menor que el nivel de significancia ($\alpha = 0,05$). Por lo tanto, se rechaza la hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P1 y la deserción universitaria (R1).

P3: Si tienes responsabilidades familiares, ¿cómo afecta esto a tus estudios?

Tabla 7

Pruebas de Chi Cuadrado para P3

	Valor	**Grado de Libertad**	**Significancia asintótica(bilateral)**
Chi cuadrado de Pearson	22,884	3	< 0,001
Razón de verosimilitud	18,133	3	< 0,001

Asociación lineal por lineal	14,304	1	< 0,001
N de casos válidos	464		

Los resultados de la prueba Chi-cuadrado (X^2) para P3 muestran un $p < 0,001$, lo cual es menor que el nivel de significancia ($\alpha = 0,05$). Por lo tanto, se rechaza la hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P3 y la deserción universitaria (R1).

P4: ¿Cuánto impacta tu situación económica en la posibilidad de continuar con tus estudios?

Tabla 8

Pruebas de Chi Cuadrado para P4

	Valor	**Grado de Libertad**	**Significancia asintótica (bilateral)**
Chi cuadrado de Pearson	17,647	3	< 0,001
Razón de verosimilitud	16,005	3	0,001
Asociación lineal por lineal	12,473	1	< 0,001
N de casos válidos	464		

Los resultados de la prueba Chi-cuadrado (X^2) para P3 muestran un $p < 0,001$, lo cual es menor que el nivel de significancia ($\alpha = 0,05$). Por lo tanto, se rechaza la hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P4 y la deserción universitaria (R1).

P5: ¿En qué medida influye tu salud en tu capacidad para continuar con tus estudios?

Tabla 9

Pruebas de Chi Cuadrado para P5

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
Chi cuadrado de Pearson	33,571	3	< 0,001
Razón de verosimilitud	31,880	3	< 0,001
Asociación lineal por lineal	23,546	1	< 0,001
N de casos válidos	464		

Los resultados de la prueba Chi-cuadrado (X^2) para P5 muestran un $p < 0,001$, lo cual es menor que el nivel de significancia ($\alpha = 0,05$). Por lo tanto, se rechaza la hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P5 y la deserción universitaria (R1).

P6: Si tienes una pareja, ¿cómo afecta esto a tu desempeño académico?

Tabla 10

Pruebas de Chi Cuadrado para P6

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
Chi cuadrado de Pearson	24,756	4	< 0,001
Razón de verosimilitud	24,061	4	< 0,001
Asociación lineal por lineal	5,957	1	0,015
N de casos válidos	464		

Los resultados de la prueba Chi-cuadrado (X^2) para P6 muestran un $p < 0,001$, lo cual es menor que el nivel de significancia ($\alpha = 0,05$). Por lo tanto, se rechaza la

hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P6 y la deserción universitaria (R1).

P7: Si tienes una pareja, ¿cómo afecta esto a tu desempeño académico?

Tabla 11

Pruebas de Chi Cuadrado para P7

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
Chi cuadrado de Pearson	39,587	3	< 0,001
Razón de verosimilitud	32,863	3	< 0,001
Asociación lineal por lineal	28,217	1	< 0,001
N de casos válidos	464		

Los resultados de la prueba Chi-cuadrado (X^2) para P7 muestran un $p < 0,001$, lo cual es menor que el nivel de significancia ($\alpha = 0,05$). Por lo tanto, se rechaza la hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P7 y la deserción universitaria (R1).

P8: Si experimentas violencia familiar, ¿cómo afecta esto a tu rendimiento académico?

Tabla 12

Pruebas de Chi Cuadrado para P8

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
Chi cuadrado de Pearson	26,785	3	< 0,001
Razón de verosimilitud	24,465	3	< 0,001
Asociación lineal por lineal	18,030	1	< 0,001

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
N de casos válidos	464		

Los resultados de la prueba Chi-cuadrado (X^2) para P8 muestran un $p < 0,001$, lo cual es menor que el nivel de significancia ($\alpha = 0,05$). Por lo tanto, se rechaza la hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P8 y la deserción universitaria (R1).

P9: ¿En qué medida la interacción con tus profesores afecta tu desempeño académico?

Tabla 13

Pruebas de Chi Cuadrado para P9

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
Chi cuadrado de Pearson	21,138	3	< 0,001
Razón de verosimilitud	19,167	3	< 0,001
Asociación lineal por lineal	16,132	1	< 0,001
N de casos válidos	464		

Los resultados de la prueba Chi-cuadrado (X^2) para P9 muestran un $p < 0,001$, lo cual es menor que el nivel de significancia ($\alpha = 0,05$). Por lo tanto, se rechaza la hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P9 y la deserción universitaria (R1).

P10: ¿Cómo afecta el acoso de tus compañeros a tu rendimiento académico?

Tabla 14

Pruebas de Chi Cuadrado para P10

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
Chi cuadrado de Pearson	33,125	3	< 0,001
Razón de verosimilitud	28,802	3	< 0,001
Asociación lineal por lineal	14,409	1	< 0,001
N de casos válidos	464		

Los resultados de la prueba Chi-cuadrado (X^2) para P10 muestran un $p < 0,001$, lo cual es menor que el nivel de significancia ($\alpha = 0,05$). Por lo tanto, se rechaza la hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P10 y la deserción universitaria (R1).

P11: ¿La distancia entre tu hogar y la universidad afecta tu asistencia a clases?

Tabla 15

Pruebas de Chi Cuadrado para P11

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
Chi cuadrado de Pearson	29,266	3	< 0,001
Razón de verosimilitud	28,743	3	< 0,001
Asociación lineal por lineal	24.182	1	< 0,001
N de casos válidos	464		

Los resultados de la prueba Chi-cuadrado (X^2) para P11 muestran un $p < 0,001$, lo cual es menor que el nivel de significancia ($\alpha = 0,05$). Por lo tanto, se rechaza la

hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P11 y la deserción universitaria (R1).

P12: ¿Te sientes cómodo con tus compañeros de clase?

Tabla 16

Pruebas de Chi Cuadrado para P12

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
Chi cuadrado de Pearson	9,250	3	0,026
Razón de verosimilitud	8,515	3	0,036
Asociación lineal por lineal	7,114	1	0,008
N de casos válidos	464		

Los resultados de la prueba Chi-cuadrado (X^2) para P11 muestran un $p = 0,026$, lo cual es menor que el nivel de significancia ($\alpha = 0,05$). Por lo tanto, se rechaza la hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P12 y la deserción universitaria (R1).

P13: ¿Has experimentado discriminación en la universidad?

Tabla 17

Pruebas de Chi Cuadrado para P13

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
Chi cuadrado de Pearson	7,275	3	0,064
Razón de verosimilitud	6,517	3	0,089
Asociación lineal por lineal	6,952	1	0,008

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
N de casos válidos	464		

Los resultados de la prueba de Chi-cuadrado (X^2) para P13 muestran un p = 0.064 que es mayor que el nivel de significancia (α = 0.05). Por lo tanto, no se rechaza la hipótesis nula (H_0), lo que indica que no hay evidencia estadísticamente significativa para afirmar que las respuestas a P13 están relacionadas con la deserción universitaria (R1).

P14: ¿Crees que dedicas suficiente tiempo al estudio de tus materias?

Tabla 18

Pruebas de Chi Cuadrado para P14

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
Chi cuadrado de Pearson	26,944	3	< 0,001
Razón de verosimilitud	29,529	3	< 0,001
Asociación lineal por lineal	11,693	1	< 0,001
N de casos válidos	464		

Los resultados de la prueba Chi-cuadrado (X2) para P14 muestran un p < 0,001, lo cual es menor que el nivel de significancia (α = 0,05). Por lo tanto, se rechaza la hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P14 y la deserción universitaria (R1).

P15: ¿Consideras que tu carrera te facilitará encontrar empleo?

Tabla 19

Pruebas de Chi Cuadrado para P15

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
Chi cuadrado de Pearson	46,367	3	< 0,001
Razón de verosimilitud	47,815	3	< 0,001
Asociación lineal por lineal	31,701	1	< 0,001
N de casos válidos	464		

Los resultados de la prueba Chi-cuadrado (X2) para P15 muestran un $p < 0,001$, lo cual es menor que el nivel de significancia ($\alpha = 0,05$). Por lo tanto, se rechaza la hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P15 y la deserción universitaria (R1).

P16: ¿Tus expectativas académicas se cumplieron después de salir del colegio?

Tabla 20

Pruebas de Chi Cuadrado para P16

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
Chi cuadrado de Pearson	47,431	3	< 0,001
Razón de verosimilitud	53,201	3	< 0,001
Asociación lineal por lineal	32,098	1	< 0,001
N de casos válidos	464		

Los resultados de la prueba Chi-cuadrado (X2) para P16 muestran un $p < 0,001$, lo cual es menor que el nivel de significancia ($\alpha = 0,05$). Por lo tanto, se rechaza la

hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P16 y la deserción universitaria (R1).

P17: ¿Los conocimientos adquiridos en la universidad te son útiles en otros aspectos de tu vida?

Tabla 21

Pruebas de Chi Cuadrado para P17

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
Chi cuadrado de Pearson	46,667	3	< 0,001
Razón de verosimilitud	44,715	3	< 0,001
Asociación lineal por lineal	41,708	1	< 0,001
N de casos válidos	464		

Los resultados de la prueba Chi-cuadrado (X2) para P17 muestran un $p < 0,001$, lo cual es menor que el nivel de significancia ($\alpha = 0,05$). Por lo tanto, se rechaza la hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P17 y la deserción universitaria (R1).

P18: ¿La influencia de tu familia afectó tu elección de carrera?

Tabla 22

Pruebas de Chi Cuadrado para P18

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
Chi cuadrado de Pearson	35,653	3	< 0,001
Razón de verosimilitud	31,095	3	< 0,001

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
Asociación lineal por lineal	14,227	1	< 0,001
N de casos válidos	464		

Los resultados de la prueba Chi-cuadrado (X2) para P18 muestran un $p < 0,001$, lo cual es menor que el nivel de significancia ($\alpha = 0,05$). Por lo tanto, se rechaza la hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P18 y la deserción universitaria (R1).

P19: ¿Tenías en mente estudiar una carrera diferente a la que finalmente elegiste?

Tabla 23

Pruebas de Chi Cuadrado para P19

	Valor	Grado de Libertad	Significancia asintótica (bilateral)
Chi cuadrado de Pearson	27,441	3	< 0,001
Razón de verosimilitud	27,491	3	< 0,001
Asociación lineal por lineal	18,059	1	< 0,001
N de casos válidos	464		

Los resultados de la prueba Chi-cuadrado (X2) para P19 muestran un $p < 0,001$, lo cual es menor que el nivel de significancia ($\alpha = 0,05$). Por lo tanto, se rechaza la hipótesis nula y se concluye que existe una relación significativa entre las respuestas a P19 y la deserción universitaria (R1).

5.2.3. Prueba de Hipótesis General

Teniendo en cuenta que el planteamiento de la hipótesis fue el siguiente:

Hipótesis Nula (H_0). La implementación del modelo predictivo basado en machine learning no tiene un efecto significativo en el diagnóstico temprano de la deserción universitaria en la UNDC.

Hipótesis Alternativa (H_1). La implementación del modelo predictivo basado en machine learning tiene un efecto significativo en el diagnóstico temprano de la deserción universitaria en la UNDC.Se confirma la hipótesis alternativa (H_1), concluyendo que el modelo predictivo basado en machine learning tiene un impacto significativo en el diagnóstico temprano de la deserción universitaria, evidenciando su efectividad como herramienta para identificar factores críticos asociados al abandono estudiantil. Este impacto se evidencia al analizar los resultados obtenidos en cada una de las dimensiones evaluadas:

Resultados por Dimensión. Los resultados obtenidos en la **dimensión económica**, representada por las preguntas P1, P2, P3, P4, P5 y P6 muestran que todas las pruebas de hipótesis específicas para estas preguntas resultaron significativas ($p <$ 0.), indicando una relación estadísticamente significativa entre los factores económicos y la deserción universitaria (R1). Esto permite generalizar que la dimensión económica, en su conjunto, tiene un impacto relevante en la probabilidad de deserción. Los hallazgos sugieren que los problemas financieros enfrentados por los estudiantes contribuyen de manera importante a su decisión de abandonar los estudios, destacando la necesidad de implementar estrategias institucionales orientadas a mitigar estos factores.

Los resultados obtenidos en la **dimensión académica**, representada por las preguntas P7, P8, P9, P10, P11, P12, P13 y P14, indican que siete de las ocho preguntas resultaron estadísticamente significativas en las pruebas de hipótesis ($p < 0.05$). Esto

sugiere que los factores relacionados con el desempeño académico y la experiencia educativa tienen una influencia importante en la deserción universitaria (R1). Aunque la pregunta P13 no presentó significancia estadística ($p > 0.05$), su contribución al análisis general no invalida el impacto global de la dimensión académica, dado que el 87.5% de los factores analizados dentro de esta dimensión muestran una relación significativa con la deserción. Estos resultados destacan la necesidad de fortalecer las estrategias académicas, tales como el apoyo a estudiantes con dificultades de rendimiento, la mejora en la calidad de la enseñanza y el acompañamiento académico personalizado, como medidas clave para reducir la deserción.

Los resultados obtenidos en la **dimensión vocación profesional**, representada por las preguntas P15, P16, P17, P18 y P19, muestran que todas las pruebas de hipótesis específicas para estas preguntas resultaron estadísticamente significativas ($p < 0.05$). Esto indica que los factores relacionados con la claridad vocacional, la satisfacción con la carrera elegida y el interés por los estudios tienen una relación directa y significativa con la deserción universitaria (R1). En particular, estos hallazgos sugieren que la falta de afinidad con la carrera elegida y la incertidumbre sobre las metas profesionales son factores clave que incrementan la probabilidad de deserción. Por lo tanto, se concluye que la dimensión vocación profesional, en su conjunto, tiene un impacto importante en el fenómeno de la deserción, subrayando la importancia de implementar programas de orientación vocacional y actividades de acompañamiento desde el ingreso a la universidad para reducir el riesgo de abandono.

Capítulo VI: Discusión

Los resultados obtenidos en esta investigación confirman la influencia significativa del modelo predictivo basado en machine learning en el diagnóstico temprano de la deserción universitaria, lo que coincide con estudios previos como el realizado por Candia Oviedo [63], en el que se utilizó aprendizaje automático para predecir el rendimiento académico en estudiantes universitarios. En ambos casos, los algoritmos de clasificación demostraron ser efectivos para identificar patrones clave en los datos educativos. Sin embargo, a diferencia de Candia Oviedo, nuestro modelo se enfoca específicamente en el diagnóstico de deserción, obteniendo resultados significativos para la mayoría de las dimensiones evaluadas, lo que evidencia la adaptabilidad de los enfoques de machine learning en contextos educativos diversos y su potencial para mejorar la toma de decisiones en políticas de retención estudiantil.

En concordancia con la investigación realizada por De la Cruz [19], quien utilizó un enfoque similar al nuestro en la Universidad Ricardo Palma. En ambos casos, los modelos predictivos facilitaron la identificación de factores críticos asociados a la deserción estudiantil, como la deuda semestral, el promedio de asistencias y los ingresos familiares. Sin embargo, mientras en el presente estudio se emplearon algoritmos como la regresión logística para un análisis puntual de los factores económicos, académicos y vocacionales, De la Cruz implementó una arquitectura más amplia que integró múltiples algoritmos de machine learning en un servidor Big Data, maximizando la automatización del análisis de datos. Aunque las estrategias metodológicas presentan diferencias, ambos trabajos coinciden en la necesidad de generar reportes detallados para que las instituciones educativas implementen estrategias efectivas que aborden las causas específicas de la deserción. Estos hallazgos subrayan el valor de los modelos predictivos como herramientas clave para mejorar la toma de decisiones en el ámbito educativo.

En línea con la investigación realizada por Pérez y Rojas en la Universidad Tecnológica del Perú [6], podemos mencionar que en la presente investigación se utilizó regresión logística para analizar los factores económicos, académicos y vocacionales, mientras que Pérez y Rojas implementaron el algoritmo Support Vector Machine (SVM), destacando su capacidad para identificar patrones de comportamiento relacionados con la deserción. Ambos estudios emplearon datos históricos de estudiantes para detectar posibles deserciones estudiantiles con cierta precisión. Sin embargo, el uso de diferentes algoritmos resalta enfoques complementarios, donde el SVM, con una certeza superior al 90 %, demuestra robustez en escenarios no lineales gracias al uso de la función Kernel. A pesar de estas diferencias metodológicas, ambos trabajos subrayan la importancia de estas herramientas en la planificación institucional, permitiendo intervenciones tempranas y personalizadas que buscan reducir la tasa de deserción y mejorar la experiencia educativa.

Los resultados de este estudio, que validan la implementación de un modelo predictivo basado en machine learning para el diagnóstico temprano de la deserción universitaria, encuentran paralelismos con el trabajo de Forero Corba y Negre Bennasar [64], enfocado en la predicción de la competencia digital docente, el modelo propuesto por Forero y Negre, realizado en software Orange Data Mining, utiliza rúbricas de evaluación para medir la competencia digital docente en niveles explorador, integrador e innovador. Mientras en este trabajo se consideraron dimensiones como la económica, académica y vocacional, en el contexto docente se utilizaron rúbricas estandarizadas y variables sociodemográficas para evaluar competencias digitales. Estas similitudes sugieren que el enfoque puede adaptarse a otros sectores donde la identificación temprana de deficiencias o patrones puede mejorar la toma de decisiones, como en programas de capacitación profesional, estrategias de inclusión digital o personalización del aprendizaje.

Se encuentran analogías significativas con el trabajo de Yandar et al. [65], quienes realizaron una revisión sistemática de literatura enfocada en la predicción de energía eléctrica. Ambas investigaciones subrayan el impacto de técnicas de machine learning como las redes neuronales artificiales (RNA) para modelar y prever comportamientos complejos. Mientras que nuestro trabajo abordó dimensiones académicas, económicas y vocacionales para identificar factores críticos en la deserción, Yandar et al. destacaron el uso de RNA para la predicción precisa del consumo energético, mostrando cómo estas herramientas pueden integrarse en sectores diversos, desde la educación hasta la gestión de recursos energéticos.

También se encuentran aplicaciones complementarias en el ámbito de la salud mental, tal como se evidencia en el artículo científico de Osorio et. al [66], se destaca el uso de modelos estadísticos como la regresión logística, el SVM y los bosques aleatorios. Mientras que nuestro modelo se enfoca en factores económicos, académicos y vocacionales, Osorio et al. identifican variables predictoras clave como neuroimágenes, características sociodemográficas y síntomas clínicos. En ambos estudios subrayamos la importancia de un enfoque preventivo, en mi estudio para reducir la deserción, y en el estudio revisado, para priorizar y personalizar la gestión de la salud mental.

En este estudio, las hipótesis específicas fueron evaluadas utilizando chi cuadrado, y los resultados muestran que la mayoría fueron aceptadas, mientras que la **Pregunta 13**: "¿Has experimentado discriminación en la universidad?" dentro de la dimensión **situación académica** no tuvo significancia estadística. La **hipótesis general**, que plantea que el modelo predictivo basado en machine learning influye significativamente en el diagnóstico temprano de la deserción universitaria, fue aceptada, dado que los valores $p < 0,05$ en la mayoría de las hipótesis específicas respaldan la relación significativa entre las variables analizadas y la deserción universitaria.

A partir de los resultados obtenidos, se plantean las siguientes hipótesis que pueden guiar investigaciones futuras y refinar el modelo predictivo:

La implementación de estrategias de apoyo financiero en la universidad influye en la disminución del riesgo de deserción asociado a factores económicos.

El fortalecimiento de los programas de tutoría académica y la mejora en la percepción de calidad docente tienen un impacto significativo en la retención estudiantil.

La introducción de programas de orientación vocacional en los primeros ciclos académicos mejora la afinidad de los estudiantes con su carrera y reduce la probabilidad de deserción.

La incorporación de sistemas inteligentes basados en machine learning para el monitoreo continuo de datos académicos y socioeconómicos mejora la detección temprana de estudiantes en riesgo de deserción.

La integración de dimensiones adicionales, como la salud mental y el entorno familiar, mejora la capacidad predictiva del modelo para diagnosticar la deserción estudiantil.

Capítulo VII: Conclusiones

Se confirma que la implementación de un modelo predictivo basado en machine learning influye significativamente en el diagnóstico temprano de la deserción universitaria en la UNDC durante el periodo 2020 – 2022.

1. Los factores económicos, como la capacidad financiera y la estabilidad económica del estudiante, demostraron tener una influencia significativa en la probabilidad de deserción, con todas las preguntas de esta dimensión resultando significativas ($p < 0,05$). Este hallazgo resalta la necesidad de implementar estrategias de apoyo financiero para mitigar estos factores de riesgo.
2. La dimensión académica, si bien tuvo en su mayoría resultados significativos, mostró que una pregunta (P13) no tuvo una relación estadísticamente relevante ($p > 0.05$). Esto sugiere que la carga académica, el desempeño y la percepción de calidad educativa son predictores importantes, pero requieren una revisión exhaustiva para refinar las variables no significativas.
3. La afinidad vocacional y la claridad en los objetivos profesionales demostraron ser factores clave para el diagnóstico temprano de la deserción, con todas las preguntas de esta dimensión resultando significativas ($p < 0.05$). Esto subraya la importancia de incorporar programas de orientación vocacional en las estrategias institucionales.
4. El modelo predictivo basado en machine learning mostró un desempeño confiable en la clasificación de estudiantes en riesgo, con una precisión del 80,64% de los casos usando el modelo predictivo de regresión logística, respaldado por indicadores como un Alfa de Cronbach de 0.704 y valores $p < 0.05$ en la mayoría de los análisis. Esto valida el uso de técnicas de machine learning como herramientas efectivas para el diseño de estrategias preventivas.

Capítulo VIII: Recomendaciones

Basado en la influencia significativa de los factores económicos en la deserción universitaria, se recomienda establecer programas de becas, subsidios y opciones flexibles de pago para estudiantes con limitaciones financieras. Esto podría complementarse con talleres de gestión financiera personal para ayudar a los estudiantes a administrar mejor sus recursos.

Para abordar los problemas relacionados con la dimensión académica, se sugiere diseñar tutorías personalizadas enfocadas en mejorar el rendimiento académico y reducir la carga percibida por los estudiantes. Asimismo, la capacitación constante del cuerpo docente sobre metodologías de enseñanza innovadoras podría mejorar la percepción de calidad educativa.

Debido al impacto de los factores vocacionales en la deserción, es crucial implementar un programa robusto de orientación vocacional desde los primeros ciclos universitarios. Este sistema debería incluir evaluaciones periódicas de afinidad profesional, talleres de exploración vocacional y un acompañamiento continuo para alinear los objetivos personales con los académicos.

Se recomienda construir una plataforma digital que utilice modelos predictivos para monitorear en tiempo real el desempeño académico, la situación económica y los indicadores vocacionales de los estudiantes. Esto permitiría una intervención temprana y personalizada en función de las necesidades detectadas.

Para mejorar la precisión del diagnóstico, se sugiere incorporar variables adicionales como factores de salud mental, entorno familiar y contexto social en futuras versiones del modelo. Esto ofrecerá una visión más integral de las causas de la deserción y permitirá diseñar estrategias más efectivas.

Referencias

[1] Ministerio de Educación, «Tasa de deserción en educación universitaria se redujo a 11.5%,» 8 Noviembre 2021. [En línea]. Available: https://www.gob.pe/institucion/minedu/noticias/552273-tasa-de-desercion-en-educacion-universitaria-se-redujo-a-11-5.

[2] R. C. Dávila Morán, E. d. C. Agüero Corzo, H. Portillos Rios y O. R. Quimbita Chiluisa, «Deserción universitaria de los estudiantes de una universidad peruana,» *Revista Universidad y Sociedad,* vol. 14, nº 2, 25 08 2022.

[3] J. E. Valero Cajahuanca, A. F. Navarro Raymundo, A. C. Larios Franco y J. D. Julca Flores, «Deserción universitaria: Evaluación de diferentes algoritmos de Machine Learning para su predicción,» *Revista de Ciencias Sociales (Ve)*, vol. 28, nº 3, pp. 362-375, 2022.

[4] R. I. Castro López, «Aplicación de técnicas de machine learning para el estudio de deserción temprana y egreso oportuno en estudiantes de ingeniería de la facultad de ciencias físicas y matemáticas.,» Universidad De Chile, Chile, 2020.

[5] I. Díaz, A. Bernardo, M. Esteban y L. Rodrígues Muñiz, «Variables Influencing University Dropout: A Machine Learning-Based Study,» ICEUTE 2020, España, 2020.

[6] C. A. Perez Bedia y L. E. Rojas Segovia, «Diseño de un sistema para predecir la deserción de los alumnos mediante Machine learning en la Universidad Tecnológica del Perú.,» Universidad Tecnológica del Perú, Lima, 2020.

[7] A. M. Tataje Okamoto, K. M. Chacaltana Hernández y S. G. Varela Guevara, «Adaptación universitaria en estudiantes de primer año de una universidad pública de Ica – Perú,» Universidad Autónoma del Perú, Lima, 2024.

[8] T. M. Oca Sánchez, «Educación superior: factores económicos que inciden en la deserción escolar. Caso de las licenciaturas de la UNID Tlalnepantla,» *Revista Iberoamericana para la Investigación y el Desarrollo Educativo,* vol. 12, nº 23, 2021.

[9] N. J. Cárdenas Ballesteros, A. M. Díaz Rodriguez y A. E. Chaparro Ortiz, «Revisión sistemática de los factores que inciden en la deserción académica en la educación superior,» Universidad Cooperativa de Colombia, Colombia, 2021.

[10] X. F. Erazo Guerra y E. d. R. Rosero Morales, «Orientación vocacional y su influencia en la deserción universitaria,» *Horizontes,* vol. 5, nº 18, p. 16, 2021.

[11] A. Villa Murillo, L. Costa y C. Vásquez, «A methodological proposal to address the academic dropout phenomenon based on an intelligent prediction model: a case study,» *Millenium - Journal of Education, Technologies and Health,* vol. 2, nº 23, 2023.

[12] T. Shuo-Chang, C. Cheng-Huan, S. Yi-Tzone, C. Jin-Shuei y W. Trong-Neng, «Precision education with statistical learning and deep learning: a case study in Taiwan,» *International Journal of Educational Technology in Higher Education,* vol. 17, nº 12, 2020.

[13] A. Rodriguez, J. Espinoza, L. Ramirez y A. Ganga, «Deserción Universitaria: Nuevo Análisis Metodológico,» *Formación Universitaria,* vol. 11, nº 6, pp. 107-118, 2018.

[14] M. G. Cuzme Romero, «Modelo de aprendizaje automático para la predicción del riesgo de deserción en estudiantes de la Universidad Técnica de Manabí,» Tesis de maestría, Universidad de las Fuerzas Armadas, Ecuador, 2020.

[15] Y. A. Quintero, «Diseño de un modelo predictivo para generar alertas tempranas de deserción universitaria en los programas de pregrado presenciales de la Facultad de Ingeniería de la Universidad de Antioquia,» Tesis de maestría, Universidad de Antioquia, Colombia, 2022.

[16] C. Russo, «Minería de datos aplicada a estrategias para minimizar el rezago académico y la deserción universitaria en carreras de informática de la UNNOBA,» Tesis doctoral, Universidad Nacional de la Plata, Argentina, 2019.

[17] D. A. Gómez Pinzón, «Aplicativo web para el análisis, selección y admisión de aspirantes al programa de ingeniería de sistemas de la Universidad de Cundinamarca en la extensión facatativá utilizando modelos predictivos de minería de datos.,» Tesis de pregrado, Universidad De Cundinamarca, Colombia, 2016.

[18] E. N. Cevallos Medina y C. J. Barahona Chunga, «Modelo para automatizar el proceso de predicción de la deserción en estudiantes universitarios en el primer año de estudio,» Tesis de licenciatura, Universidad Peruana de Ciencias Aplicadas, Lima, 2021.

[19] V. R. De La Cruz Quispe, «Diseño de un modelo predictivo basado en machine learning para el control de la deserción de estudiantes en la Universidad Ricardo Palma,» Tesis de licenciatura, Universidad Nacional Tecnólogica de Lima Sur, Villa el Salvador, 2019.

[20] M. S. Albán Taipe, «Contribuciones a la predicción de la deserción universitaria a través de minería de datos,» Tesis doctoral, Universidad Nacional Mayor de San Marcos, Lima, 2019.

[21] G. M. Lozano Chamorro, «Business Analytics para el pronóstico de deserciones de estudiantes en el Colegio Héroes de la Breña,» Tesis de licenciatura, Universidad César Vallejo, Lima, 2021.

[22] M. R. Alfonso Centeno, «Big Data. Técnicas de machine learning para la creación de modelos predictivos para empresas,» Tesis de licenciatura, Comillas Universidad Pontificia, Madrid, 2020.

[23] Wikipedia contributors, «Wikipedia, The Free Encyclopedia.,» [En línea]. Available: https://en.wikipedia.org/wiki/Machine_learning. [Último acceso: 2023].

[24] Google, «Developers.google,» [En línea]. Available: https://developers.google.com/machine-learning/crash-course/classification/accuracy?hl=en. [Último acceso: 2023].

[25] H2O, «H2O.ai,» [En línea]. Available: https://h2o.ai/wiki/prediction/. [Último acceso: 2023].

[26] G. Azabache Martínez, «Brave Developer,» 5 marzo 2022. [En línea]. Available: https://bravedeveloper.com/2022/03/05/time-complexity-o-complejidad-temporal-eficiencia-de-los-algoritmos-en-terminos-de-tiempo/.

[27] O. Martínez de Quel Pérez y M. Sillero Quintana, «Sobre la expresión “respuesta de reacción” y el concepto "tiempo de respuesta",» *Apunts*, 2014.

[28] soloelectronicos, «Soloelectronicos,» 19 enero 2022. [En línea]. Available: https://soloelectronicos.com/2022/01/19/que-es-la-puntuacion-f1/.

[29] R. Díaz, «The Machine Learners,» [En línea]. Available: https://www.themachinelearners.com/metricas-de-clasificacion/#Accuracy.

[30] L. P. Rouhiainen, Inteligencia artificial 101 cosas que debes saber hoy sobre nuestro futuro, España: Planeta, 2018.

[31] «Inteligencia Artificial,» *INCyTu*, 2018.

[32] A. Burkov, The hundred-page machine learning book, 2019.

[33] J. Díez García, «Machine learning aplicado al trading,» Universidad Pontificia Comillas, España, 2021.

[34] F. Gabel, «Some studies in machine learning using,» 2019.

[35] APD, «¿Cuáles son los tipos de algoritmos del machine learning?,» APD, España, 2019.

[36] IBM, «IBM,» Julio 2016. [En línea]. Available: https://www.ibm.com/es-es/topics/logistic-regression.

[37] M. P. Seminara, «La deserción universitaria: resiliencia como posiblidad de logro,» *Universidad Nacional Autónoma de México,* vol. 21, nº 5, p. 11, 2020.

[38] UTP, «UTP,» 23 mayo 2023. [En línea]. Available: https://www.utp.edu.pe/blog/orientacion-vocacional/que-es-la-vocacion-profesional.

[39] F. Coll Morales, «Economipedia,» 1 abril 2020. [En línea]. Available: https://economipedia.com/definiciones/carga-financiera.html.

[40] Colaboradores de Wikipedia, «Wikipedia,» 21 septiembre 2023. [En línea]. Available: https://es.wikipedia.org/wiki/Escasez#Referencias.

[41] S. Donoso D., R. Ruiz V., R. Schmal S. y M. Schaffernicht Sch., «Financiamiento de los estudios universitarios: un desafío estratégico de sustentabilidad,» *Universidad de Talca,* nº 29, 2004.

[42] J. O. Moreno Treviño y S. N. Cortez Soto, «Rendimiento académico y habilidades de estudiantes en escuelas públicas y privadas: evidencia de los determinantes de las brechas en aprendizaje para México,» *Revista de economía,* vol. 37, 2020.

[43] Á. Díaz Barriga, «Construcción de programas de estudio en la perspectiva del enfoque de desarrollo de competencias,» *Universidad Nacional Autónoma de México,* 2014.

[44] Mineducacion, [En línea]. Available: https://mineducacion.gov.co/1759/articles-356821_recurso_1.pdf.

[45] Universidad Europea, 17 Mayo 2023. [En línea]. Available: https://universidadeuropea.com/blog/tipos-formacion/.

[46] J. Pérez Porto y M. Merino, «Definición de,» 17 agosto 2020. [En línea]. Available: https://definicion.de/matricula/.

[47] K. Rivera Vergaray, «Modelo predictivo para la detección temprana de estudiantes con alto riesgo de deserción académica,» Redalyc, 2021.

[48] I. M. Poveda Velasco, «Los factores que influyen sobre la deserción universitaria. Estudio en la UMRPSFXCh - Bolivia, análisis con ecuaciones estructurales,» Universidad Mayor Real y Pontificia San Francisco Xavier de Chuquisaca, Bolivia, 2019.

[49] Q. V. Ingrith, «Análisis de las causas de deserción universitaria,» Universidad Nacional Abierta y a Distancia, Colombia, 2016.

[50] ZeroBit, «Platzi,» 2020. [En línea]. Available: https://platzi.com/tutoriales/1469-algoritmos-practico/7790-que-son-los-algoritmos/.

[51] Gamco, «Gamco.es,» [En línea]. Available: https://gamco.es/glosario/modelo-predictivo/.

[52] Universidad de La Guajira, «uniguajira.edu.co,» [En línea]. Available: https://www.uniguajira.edu.co/diagnostico-e-identificacion-de-problemas.

[53] W. C. John, Research Design: Qualitative, Quantitative and Mixed Methods Approaches, Estados Unidos: SAGE, 2014.

[54] R. Hernández Sampieri, C. Fernández Collado y L. María del Pilar Baptista, Metdológia de la investigación, México: McGraw Hill, 2014.

[55] R. Hernández, C. Fernández y P. Baptista, Metodología de la Investigación, México D.F.: Mc Graw Hill, 2014.

[56] Wikipedia contributors, «Wikipedia, The Free Encyclopedia.,» [En línea]. Available: https://es.wikipedia.org/wiki/Método_hipotético-deductivo. [Último acceso: 2010].

[57] W. M. Grove y H. M. William, Hypothetico-Deductive Model, Estados Unidos: John Wiley & Sons, 2015.

[58] D. T. Campbell y C. S. Julian, Experimental and quasi-experimental designs for research, USA: Houghton Mifflin Company, 1963.

[59] C. Anguita, R. Labrador y D. Campos, «La encuesta como técnica de investigación. Elaboración de cuestionarios y tratamiento estadístico de los datos,» *Atención Primaria*, vol. 31, nº 8, pp. 527-538, 2003.

[60] C. Ortega, «QuestionPro,» [En línea]. Available: https://www.questionpro.com/blog/es/que-es-un-cuestionario/#que_es_un_cuestionario.

[61] A. M. Chaparro, «Centro Banamex,» 22 Marzo 2023. [En línea]. Available: https://www.centrobanamex.com.mx/que-es-la-validez-y-confiabilidad-de-un-instrumento-de-recoleccion-de-datos/.

[62] M. M. Terán Cázares, «Factores determinantes del Comportamiento de Ciudadania Organizacional: Su relación con el Entorno Sociocultural a través de la Cultura Organizacional y los atributos de Personalidad.,» Universidad Autónoma de Nuevo Leon, México, 2009.

[63] D. I. Candia Oviedo, «Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático,» Universidad Nacional de San Antonio ABAD del Cusco, Cusco, 2019.

[64] W. Forero Corba y N. Bennasar, «Diseño y simulación de un modelo de predicción para la evaluación de la competencia digital docente usando técnicas de Machine Learning,» *Edutec*, nº 89, pp. 18-43, 2024.

[65] K. L. Yandar, O. Revelo Sánchez y M. Bolaños Gonzáles, «Machine learning for electric power prediction: a systematic literature review,» *inycomp,* vol. 26, nº 2, 2024.

[66] S. Osorio Castrillon, L. M. Giraldo Marín, H. H. Jaramillo Villegas y C. C. Piedrahita Escobar, «Machine learning aplicado en la clasificación y predicción de la depresión: Una revisión sistemática,» *risti,* nº E47, pp. 363-375, 2022.

[67] F. Sonia, L. Laura y H. Waldo, «Caracterización de la deserción universitaria en la UNRN utilizando Minería de Datos. Un caso de estudio.,» Universidad Nacional de Río Negro, Argentina, 2013.

[68] C. León Aznar, «BigMark: análisis y predicción de calificaciones mediante Big Data.,» Universidad Complutense de Madrid, Madrid, 2018.

[69] J. C. González Sánchez y M. J. Peñaloza Pérez, «Identificación y predicción de estudiantes en riesgo de deserción académica por medio de modelos basados en machine learning,» Fundación Universitaria Los Libertadores, Colombia, 2021.

[70] E. P. Jazmine y Á. Cuervo Martínez, «Validez de contenido y juicio de expertos: Una aproximación a su utilización,» Universidad Nacional de Colombia, Colombia, 2008.

[71] M. Karimi-Haghighi, D. Hernández-Leo, C. Castillo y V. Moreno Oliver, «Predicting Early Dropout: Calibration and Algorithmic Fairness,» arXiv, 2021.

[72] Wikipedia, «wikipedia.org,» [En línea]. Available: https://es.wikipedia.org/w/index.php?title=Cuasi_experimento&oldid=149819260.

[73] P. Fernández Garcia, G. Vallejo Seco, P. Levacic Rojas y E. Tuero Herrero, «Validez Estructurada para una investigación cuasi-experimental de calidad. Se cumplen 50 años de la presentación en sociedad de los diseños cuasi-experimentales,» Scielo, 2014.

[74] B. Cabré, «DISEÑOS CUASI-EXPERIMENTALES Y LONGITUDINALES,» Universidad de Barcelona, España.

[75] R. V. Kevin, «Modelo predictivo para la detección temprana de estudiantes con alto riesgo de deserción académica,» Redalyc, 2021.

[76] «IBM,» [En línea]. Available: https://www.ibm.com/topics/machine-learning. [Último acceso: 2023].

[77] M. Rodriguez Urrego, «La investigación sobre deserción universitaria en Colombia 2006-2016. Tendencias y resultados,» *Pedagogía y Saberes,* nº 51, 2019.

[78] M. Nicuesa, «DefiniciónABC,» octubre 2022. [En línea]. Available: https://www.definicionabc.com/ciencia/metodo-empirico-analitico.php. [Último acceso: 2023].

[79] T. Romá Ferri, «RUA: Repositorio Institucional de la Universidad de Alicante,» 2014. [En línea]. Available: https://rua.ua.es/dspace/bitstream/10045/42141/1/DICS_T4_SRI_EvaluaResultados_C14_15.pdf.

[80] I. Castillo, «Lifeder,» 1 septiembre 2020. [En línea]. Available: https://www.lifeder.com/instrumentos-investigacion-documental/.

ANEXO 01: Matriz de Consistencia

Tabla 24
Matriz de Consistencia

Título: Modelo predictivo basado en machine learning para el diagnóstico temprano en la deserción universitaria de la UNDC, periodo 2020 – 2022

Problema general	Objetivo general	Hipótesis general	Variables y dimensiones	Metodología
¿De qué manera la implementación de un modelo predictivo basado en machine learning influye en el diagnóstico temprano la deserción universitaria de la UNDC en periodo 2020 – 2022?	Implementar un modelo predictivo basado en machine learning que influye en el diagnóstico temprano en la deserción universitaria de la UNDC en periodo 2020 – 2022.	La implementación de un modelo predictivo basado en machine learning influye en el diagnóstico temprano de la deserción universitaria de la UNDC en el periodo 2020 – 2022.	**Variable independiente** Modelo predictivo basado en machine learning **Dimensiones** Tiempo Precisión Pronóstico **Variable dependiente** Diagnóstico temprano de la deserción universitaria **Dimensiones** Factores económicos Situación académica Vocación profesional	**Enfoque de investigación** Cuantitativo **Método de investigación** Hipotético deductivo **Tipo de investigación** Aplicada **Nivel de investigación** Explicativo **Diseño de investigación** Pre experimental **Población** 1 764 estudiantes **Muestra** 316 estudiantes **Técnica de recolección de datos** Encuesta
Problemas específicos	**Objetivos específicos**	**Hipótesis especificas**		
¿De qué manera la implementación de un modelo predictivo basado en machine learning que incorpora los factores económicos influye en el diagnóstico temprano la deserción universitaria de la UNDC en periodo 2020 – 2022? ¿De qué manera la implementación de un modelo predictivo basado en machine learning que incorpora la situación académica influye en el diagnóstico temprano la deserción universitaria de la UNDC en periodo 2020 – 2022? ¿De qué manera la implementación de un modelo predictivo basado en machine learning que incorpora la vocación profesional influye en el diagnóstico temprano la deserción universitaria de la UNDC en periodo 2020 – 2022?	Implementar un modelo predictivo basado en machine learning, que incorpora los factores económicos, que influye en el diagnóstico temprano en la deserción universitaria de la UNDC en periodo 2020 – 2022. Implementar un modelo predictivo basado en machine learning, que incorpora la situación académica, que influye en el diagnóstico temprano en la deserción universitaria de la UNDC en periodo 2020 – 2022. Implementar un modelo predictivo basado en machine learning, que incorpora la vocación profesional, que influye en el diagnóstico temprano en la deserción universitaria de la UNDC en periodo 2020 – 2022.	La implementación de un modelo predictivo basado en machine learning que incorpora los factores económicos influye en el diagnóstico temprano de la deserción universitaria de la UNDC en periodo 2020 – 2022. La implementación de un modelo predictivo basado en machine learning que incorpora la situación académica influye en el diagnóstico temprano de la deserción universitaria de la UNDC en periodo 2020 – 2022. La implementación de un modelo predictivo basado en machine learning que incorpora la vocación profesional influye en el diagnóstico temprano de la deserción universitaria de la UNDC en periodo 2020 – 2022.		

ANEXO 02: Instrumento de Recolección de Datos

ENCUESTA UNIVERSITARIA

NADA	POCO	BASTANTE	MUCHO
1	2	3	4

Variable dependiente: Diagnóstico temprano de la deserción universitaria

N°	DIMENSIONES	AFIRMACIONES	1	2	3	4
1	Factores económicos	¿En qué medida dependes económicamente de alguien más?				
2		Si tienes un empleo, ¿cómo influye en tu rendimiento académico?				
3		Si tienes responsabilidades familiares, ¿cómo afecta esto a tus estudios?				
4		¿Cuánto impacta tu situación económica en la posibilidad de continuar con tus estudios?				
5		¿En qué medida influye tu salud en tu capacidad para continuar con tus estudios?				
6		Si tienes una pareja, ¿cómo afecta esto a tu desempeño académico?				
7	Situación académica	¿Las otras actividades que desempeñas afectan tu tiempo de estudio?				
8		Si experimentas violencia familiar, ¿cómo afecta esto a tu rendimiento académico?				
9		¿En qué medida la interacción con tus profesores afecta tu desempeño académico?				
10		¿Cómo afecta el acoso de tus compañeros a tu rendimiento académico?				
11		¿La distancia entre tu hogar y la universidad afecta tu asistencia a clases?				
12		¿Te sientes cómodo con tus compañeros de clase?				
13		¿Has experimentado discriminación en la universidad?				
14		¿Crees que dedicas suficiente tiempo al estudio de tus materias?				
15	Vocación profesional	¿Consideras que tu carrera te facilitará encontrar empleo?				
16		¿Tus expectativas académicas se cumplieron después de salir del colegio?				
17		¿Los conocimientos adquiridos en la universidad te son útiles en otros aspectos de tu vida?				
18		¿La influencia de tu familia afectó tu elección de carrera?				
19		¿Tenías en mente estudiar una carrera diferente a la que finalmente elegiste?				
20	¿Has pensado en desertar de tu carrera universitaria?		NO		SÍ	

ANEXO 03: Validación de los Instrumentos de Investigación

FICHA DE VALIDACIÓN POR JUICIO DE EXPERTO

1. **Datos Generales**

1.1. **Nombres y Apellidos del Experto** : ANGEL FERNANDO NAVARRO RAYMUNDO

1.2. **Grado académico / mención** : DOCTOR EN INGENIERÍA DE SISTEMAS

1.3. **N° DNI / doc. Identificación** : 41110448

1.4. **Teléfono/ Celular** : 990 064 322

1.5. **Cargo e institución donde labora** : DOCENTE PRINCIPAL UNTELS

1.6. **Tesista** : SULCA MARTÍNEZ, Hugo Jesús

1.7. **Nombre del Instrumento** : FICHA TÉCNICA DE RECOLECCIÓN DE DATOS

1.8. **Título de la Tesis** :"IMPLEMENTACIÓN DE UN MODELO PREDICTIVO BASADO EN MACHINE LEARNING PARA EL DIAGNÓSTICO TEMPRANO EN LA DESERCIÓN UNIVERSITARIA DE LA UNDC EN PERIODO PANDEMIA 2020 – 2022"

1.9. **Lugar y fecha** : Cañete, 17 de enero del 2023.

2. **Aspectos de la Evaluación**

Indicadores	Criterios	Deficiente	Aceptable	Bueno
		1	3	5
1. Objetividad	El instrumento está organizado y expresado en datos observables.			X
2. Organización	Presentación ordenada.			X
3. Suficiencia	Comprende aspectos de las variables en cantidad y calidad suficiente.			X
4. Pertinencia	Permite conseguir datos de acuerdo a los objetivos planteados.			X
5. Consistencia	Pretende conseguir datos basado en teorías o modelos teóricos.			X
6. Coherencia	Entre variables, dimensiones, indicadores e ítems.			X
7. Metodología	La estrategia responde al propósito de la investigación.			X
8. Aplicación	Los datos permiten un tratamiento estadístico pertinente.			X
Conteo total de marcas		A	B	C
				40

$$\text{Coeficiente de validez} = \frac{1xA + 3xB + 5xC}{40} = \text{suma}/40$$

3. **Opinión de aplicabilidad**

Intervalo	Categoría	
[0,20 – 0,40]	No válido, reformular	◯
<0,41 – 0,60]	No válido, modificar	◯
<0,61 – 0,80]	Válido, mejorar	◯
<0,81 – 1,00]	Válido, aplicar	⊗

4. **Criterio de valoración del Juez:**

- Procede su aplicación	(X)
- No procede su aplicación	( )

5. Recomendaciones

APLICAR.....



Firma del Experto
DNI/documento de identidad

FICHA DE VALIDACIÓN POR JUICIO DE EXPERTO

1. **Datos Generales**

1.1. **Nombres y Apellidos del Experto** : …DR. DULIO OSEDA GAGO……

1.2. **Grado académico / mención** : …DR. EN SISTEMAS DE INGENIERIA……..

1.3. N° DNI / doc. Identificación :….20044737……

1.4. **Teléfono/ Celular** : …964689004……..….

1.5. **Cargo e institución donde labora** : VICERRECTOR DE INVESTIGAICON - UNACH

1.6. **Tesista** : SULCA MARTÍNEZ, Hugo Jesús

1.7. **Nombre del Instrumento** : FICHA TÉCNICA DE RECOLECCIÓN DE DATOS

1.8. **Título de la Tesis** :"IMPLEMENTACIÓN DE UN MODELO PREDICTIVO BASADO EN MACHINE LEARNING PARA EL DIAGNÓSTICO TEMPRANO EN LA DESERCIÓN UNIVERSITARIA DE LA UNDC EN PERIODO PANDEMIA 2020 – 2022"

1.9. **Lugar y fecha** : Cañete, 17 de enero del 2023.

2. **Aspectos de la Evaluación**

Indicadores	Criterios	Deficiente	Aceptable	Bueno
		1	3	5
1. Objetividad	El instrumento está organizado y expresado en datos observables.			X
2. Organización	Presentación ordenada.			X
3. Suficiencia	Comprende aspectos de las variables en cantidad y calidad suficiente.			X
4. Pertinencia	Permite conseguir datos de acuerdo a los objetivos planteados.			X
5. Consistencia	Pretende conseguir datos basado en teorías o modelos teóricos.			X
6. Coherencia	Entre variables, dimensiones, indicadores e ítems.			X
7. Metodología	La estrategia responde al propósito de la investigación.			X
8. Aplicación	Los datos permiten un tratamiento estadístico pertinente.			X
Conteo total de marcas		A	B	C
				40

$$\text{Coeficiente de validez} = \frac{1xA + 3xB + 5xC}{40} = \text{suma}/40$$

3. **Opinión de aplicabilidad**

Intervalo	Categoría	
[0,20 – 0,40]	No válido, reformular	◯
<0,41 – 0,60]	No válido, modificar	◯
<0,61 – 0,80]	Válido, mejorar	◯
<0,81 – 1,00]	Válido, aplicar	Ⓧ

4. **Criterio de valoración del Juez:**

- Procede su aplicación	(X)
- No procede su aplicación	( )

5. **Recomendaciones**

.....APLICAR.....



Firma del Experto
DNI 20044737

FICHA DE VALIDACIÓN POR JUICIO DE EXPERTO

1. **Datos Generales**

1.1. **Nombres y Apellidos del Experto** : DANY DORIAN ISUIZA PEREZ

1.2. **Grado académico / mención** : INGENIERO DE SISTEMAS

1.3. N° DNI / doc. Identificación : 41639193

1.4. **Teléfono/ Celular** : 989 091 437

1.5. **Cargo e institución donde labora** : DOCENTE UNAMAD

1.6. **Tesista** : SULCA MARTÍNEZ, Hugo Jesús

1.7. **Nombre del Instrumento** : FICHA TÉCNICA DE RECOLECCIÓN DE DATOS

1.8. **Título de la Tesis** :"IMPLEMENTACIÓN DE UN MODELO PREDICTIVO BASADO EN MACHINE LEARNING PARA EL DIAGNÓSTICO TEMPRANO EN LA DESERCIÓN UNIVERSITARIA DE LA UNDC EN PERIODO PANDEMIA 2020 – 2022"

1.9. **Lugar y fecha** : Cañete, 17 de enero del 2023.

2. **Aspectos de la Evaluación**

Indicadores	Criterios	Deficiente	Aceptable	Bueno
		1	3	5
1. Objetividad	El instrumento está organizado y expresado en datos observables.			X
2. Organización	Presentación ordenada.			X
3. Suficiencia	Comprende aspectos de las variables en cantidad y calidad suficiente.			X
4. Pertinencia	Permite conseguir datos de acuerdo a los objetivos planteados.			X
5. Consistencia	Pretende conseguir datos basado en teorías o modelos teóricos.			X
6. Coherencia	Entre variables, dimensiones, indicadores e ítems.			X
7. Metodología	La estrategia responde al propósito de la investigación.			X
8. Aplicación	Los datos permiten un tratamiento estadístico pertinente.			X
Conteo total de marcas		A	B	C
				40

$$\text{Coeficiente de validez} = \frac{1xA + 3xB + 5xC}{40} = \text{suma}/40$$

3. **Opinión de aplicabilidad**

Intervalo	Categoría	
[0,20 – 0,40]	No válido, reformular	◯
<0,41 – 0,60]	No válido, modificar	◯
<0,61 – 0,80]	Válido, mejorar	◯
<0,81 – 1,00]	Válido, aplicar	Ⓧ

4. **Criterio de valoración del Juez:**

- Procede su aplicación	(X)
- No procede su aplicación	( )

5. **Recomendaciones**

APLICAR.....



Firma del Experto
DNI/documento de identidad

Constancia de Grados y Títulos de los Validadores

PERÚ | Ministerio de Educación | Superintendencia Nacional de Educación Superior Universitaria | Dirección de Documentación e Información Universitaria y Registro de Grados y Títulos

CONSTANCIA DE INSCRIPCIÓN EN EL REGISTRO NACIONAL DE GRADOS Y TÍTULOS

La Dirección de Documentación e Información Universitaria y Registro de Grados y Títulos, a través del Ejecutivo de la Unidad de Registro de Grados y Títulos, deja constancia que la información contenida en este documento se encuentra inscrita en el Registro Nacional de Grados y Títulos administrada por la Sunedu.

INFORMACIÓN DEL CIUDADANO

Apellidos	**NAVARRO RAYMUNDO**
Nombres	**ANGEL FERNANDO**
Tipo de Documento de Identidad	**DNI**
Numero de Documento de Identidad	**41110448**

INFORMACIÓN DE LA INSTITUCIÓN

Nombre	**UNIVERSIDAD ALAS PERUANAS S.A.**
Rector	**FIDEL RAMIREZ PRADO**
Secretario General (E)	**CESAR PRADO BARRANTES**
Director	**JORGE LAZO ARRASCO**

INFORMACIÓN DEL DIPLOMA

Grado Académico	**DOCTOR**
Denominación	**DOCTOR EN INGENIERIA DE SISTEMAS**
Fecha de Expedición	**27/02/15**
Resolución/Acta	**14510-2015-R-UAP**
Diploma	**A01726661**
Fecha Matrícula	**Sin información (*****)**
Fecha Egreso	**Sin información (*****)**

Fecha de emisión de la constancia:
22 de Julio de 2024

SUNEDU
SUPERINTENDENCIA NACIONAL DE EDUCACIÓN SUPERIOR UNIVERSITARIA

Firmado digitalmente por: SUPERINTENDENCIA NACIONAL DE EDUCACION SUPERIOR UNIVERSITARIA
Motivo: Servidor de Agente automatizado.
Fecha: 22/07/2024 07:20:17-0500

CÓDIGO VIRTUAL 0002004721

ROLANDO RUIZ LLATANCE
EJECUTIVO
Unidad de Registro de Grados y Títulos
Superintendencia Nacional de Educación Superior Universitaria - Sunedu

Esta constancia puede ser verificada en el sitio web de la Superintendencia Nacional de Educación Superior Universitaria - Sunedu (www.sunedu.gob.pe), utilizando lectora de códigos o teléfono celular enfocando al código QR. El celular debe poseer un software gratuito descargado desde internet.
Documento electrónico emitido en el marco de la Ley N° Ley N° 27269 – Ley de Firmas y Certificados Digitales, y su Reglamento aprobado mediante Decreto Supremo N° 052-2008-PCM.
(*) El presente documento deja constancia únicamente del registro del Grado o Título que se señala.

(*****) La falta de información de este campo, no involucra por sí misma un error o la invalidez de la inscripción del grado y/o título, puesto que, a la fecha de su registro, no era obligatorio declarar dicha información. Sin perjuicio de lo señalado, de requerir mayor detalle, puede contactarnos a nuestra central telefónica: 01 500 3930, de lunes a viernes, de 08:30 a.m. a 4:30 p.m.

PERÚ | Ministerio de Educación | Superintendencia Nacional de Educación Superior Universitaria | Dirección de Documentación e Información Universitaria y Registro de Grados y Títulos

CONSTANCIA DE INSCRIPCIÓN EN EL REGISTRO NACIONAL DE GRADOS Y TÍTULOS

La Dirección de Documentación e Información Universitaria y Registro de Grados y Títulos, a través del Ejecutivo de la Unidad de Registro de Grados y Títulos, deja constancia que la información contenida en este documento se encuentra inscrita en el Registro Nacional de Grados y Títulos administrada por la Sunedu.

INFORMACIÓN DEL CIUDADANO

Apellidos	**OSEDA GAGO**
Nombres	**DULIO**
Tipo de Documento de Identidad	**DNI**
Numero de Documento de Identidad	**20044737**

INFORMACIÓN DE LA INSTITUCIÓN

Nombre	**UNIVERSIDAD NACIONAL DEL CENTRO DEL PERÚ**
Rector (T)	**MOISES RONALD VASQUEZ CAICEDO AYRAS**
Secretario General (T)	**HUGO ROSULO LOZANO NUÑEZ**
Director (T)	**BARTOLOME SAENZ LOAYZA**

INFORMACIÓN DEL DIPLOMA

Grado Académico	**DOCTOR**
Denominación	**DOCTOR EN SISTEMAS DE INGENIERÍA**
Fecha de Expedición	**11/09/18**
Resolución/Acta	**4267-CU-2018**
Diploma	**0000353**
Fecha Matrícula	**20/04/2012**
Fecha Egreso	**22/01/2014**

CÓDIGO VIRTUAL 0002004731

Fecha de emisión de la constancia:
22 de Julio de 2024

SUNEDU
SUPERINTENDENCIA NACIONAL DE EDUCACIÓN SUPERIOR UNIVERSITARIA

Firmado digitalmente por: SUPERINTENDENCIA NACIONAL DE EDUCACION SUPERIOR UNIVERSITARIA
Motivo: Servidor de Agente automatizado.
Fecha: 22/07/2024 07:25:11-0500

ROLANDO RUIZ LLATANCE
EJECUTIVO
Unidad de Registro de Grados y Títulos
Superintendencia Nacional de Educación Superior Universitaria - Sunedu

Esta constancia puede ser verificada en el sitio web de la Superintendencia Nacional de Educación Superior Universitaria - Sunedu (www.sunedu.gob.pe), utilizando lectora de códigos o teléfono celular enfocando al código QR. El celular debe poseer un software gratuito descargado desde internet.
Documento electrónico emitido en el marco de la Ley N° Ley N° 27269 – Ley de Firmas y Certificados Digitales, y su Reglamento aprobado mediante Decreto Supremo N° 052-2008-PCM.
(*) El presente documento deja constancia únicamente del registro del Grado o Título que se señala.

PERÚ | Ministerio de Educación | Superintendencia Nacional de Educación Superior Universitaria | Dirección de Documentación e Información Universitaria y Registro de Grados y Títulos

CONSTANCIA DE INSCRIPCIÓN EN EL REGISTRO NACIONAL DE GRADOS Y TÍTULOS

La Dirección de Documentación e Información Universitaria y Registro de Grados y Títulos, a través del Ejecutivo de la Unidad de Registro de Grados y Títulos, deja constancia que la información contenida en este documento se encuentra inscrita en el Registro Nacional de Grados y Títulos administrada por la Sunedu.

INFORMACIÓN DEL CIUDADANO

Apellidos	**ISUIZA PEREZ**
Nombres	**DANY DORIAN**
Tipo de Documento de Identidad	**DNI**
Numero de Documento de Identidad	**41639193**

INFORMACIÓN DE LA INSTITUCIÓN

Nombre	**UNIVERSIDAD TECNOLÓGICA DEL PERÚ S.A.C.**
Rector	**ENRIQUE OSVALDO BEDOYA SANCHEZ**
Secretario General	**LUIS ALBERTO ZULOAGA ROTTA**
Decano	**LUIS ALBERTO ZULOAGA ROTTA**

INFORMACIÓN DEL DIPLOMA

Título profesional	**INGENIERO DE SISTEMAS**
Fecha de Expedición	**20/07/2010**
Resolución/Acta	**102-2010/R-UTP**
Diploma	**A01158880**

Fecha de emisión de la constancia:
22 de Julio de 2024

CÓDIGO VIRTUAL 0002004733

SUNEDU
SUPERINTENDENCIA NACIONAL DE EDUCACIÓN SUPERIOR UNIVERSITARIA

ROLANDO RUIZ LLATANCE
EJECUTIVO
Unidad de Registro de Grados y Títulos
Superintendencia Nacional de Educación Superior Universitaria - Sunedu

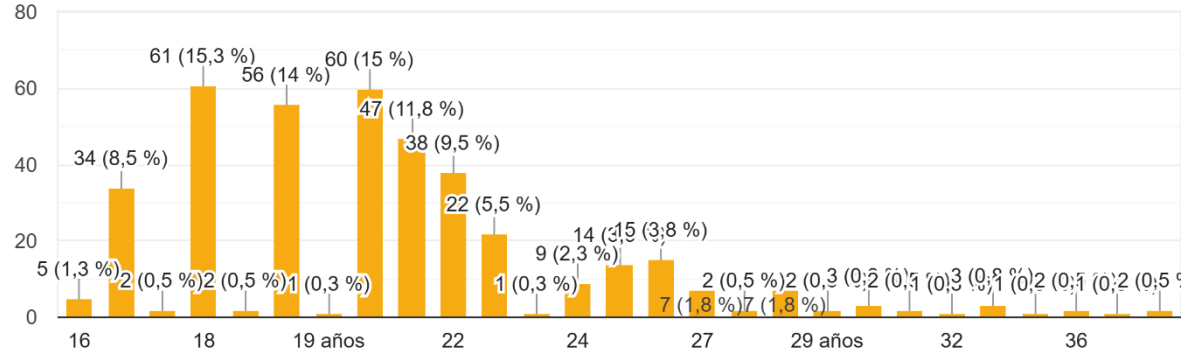
Firmado digitalmente por: SUPERINTENDENCIA NACIONAL DE EDUCACION SUPERIOR UNIVERSITARIA
Motivo: Servidor de Agente automatizado.
Fecha: 22/07/2024 07:27:51-0500

Esta constancia puede ser verificada en el sitio web de la Superintendencia Nacional de Educación Superior Universitaria - Sunedu (www.sunedu.gob.pe), utilizando lectora de códigos o teléfono celular enfocando al código QR. El celular debe poseer un software gratuito descargado desde internet.
Documento electrónico emitido en el marco de la Ley N° Ley N° 27269 – Ley de Firmas y Certificados Digitales, y su Reglamento aprobado mediante Decreto Supremo N° 052-2008-PCM.
(*) El presente documento deja constancia únicamente del registro del Grado o Título que se señala.

ANEXO 04: Base de Datos de los Indicadores

Figura 1

Distribución de los encuestados según edad



Nota: Elaboración propia basada en datos recolectados mediante Google Forms.

Figura 2

Distribución por sexo de los estudiantes encuestados

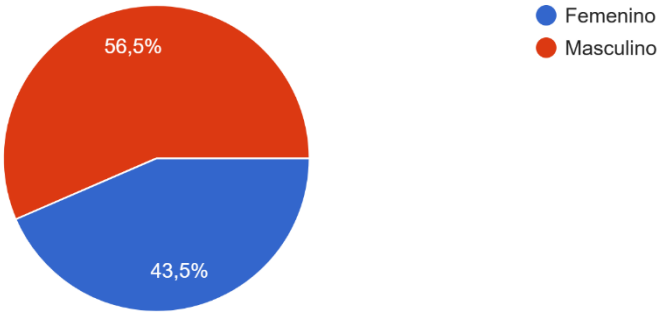


Figura 3

Distribución de los estudiantes encuestados según ciclo académico

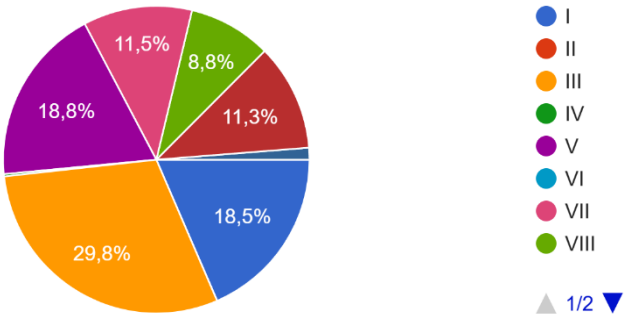


Figura 4
Distribución de los estudiantes encuestados según carrera profesional

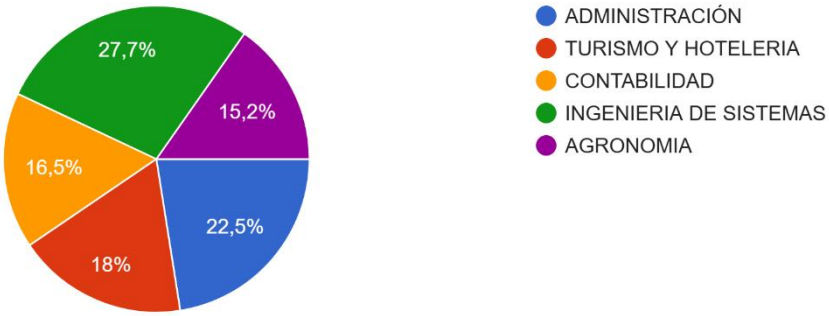


Figura 5
Distribución de respuestas a la pregunta P1: ¿En qué medida dependes económicamente de alguien más?

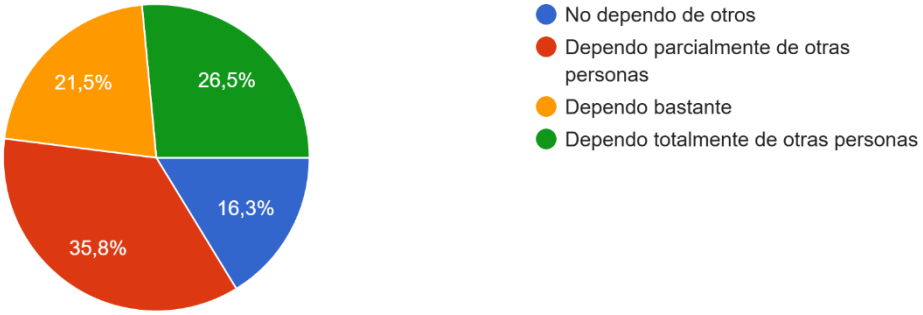


Figura 6
Distribución de respuestas a la pregunta P2: Si tienes un empleo, ¿cómo influye en tu rendimiento académico?

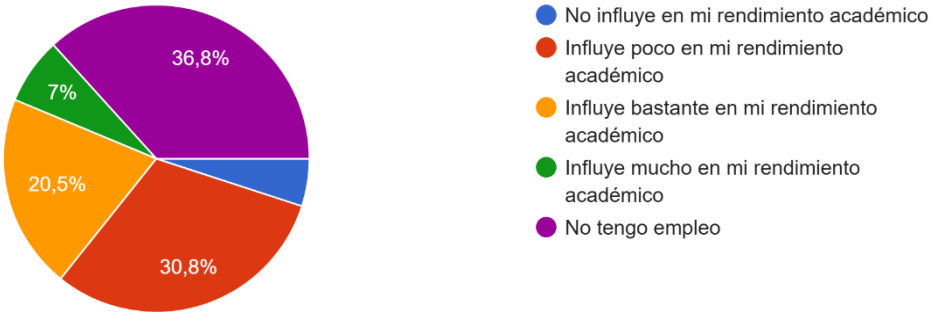


Figura 7

Distribución de respuestas a la pregunta P3: Si tienes responsabilidades familiares, ¿cómo afecta esto a tus estudios?

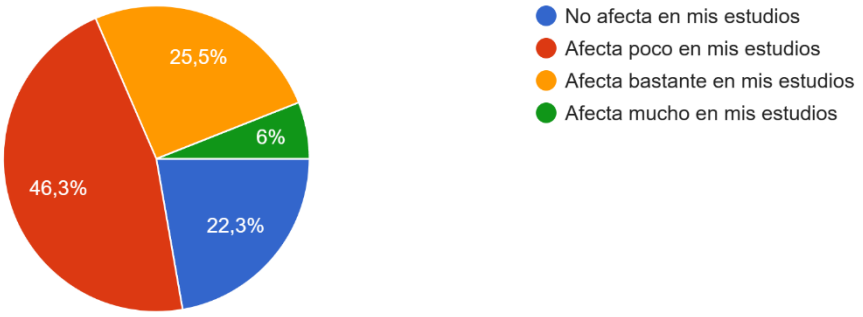


Figura 8

Distribución de respuestas a la pregunta P4: ¿Cuánto impacta tu situación económica en la posibilidad de continuar con tus estudios?

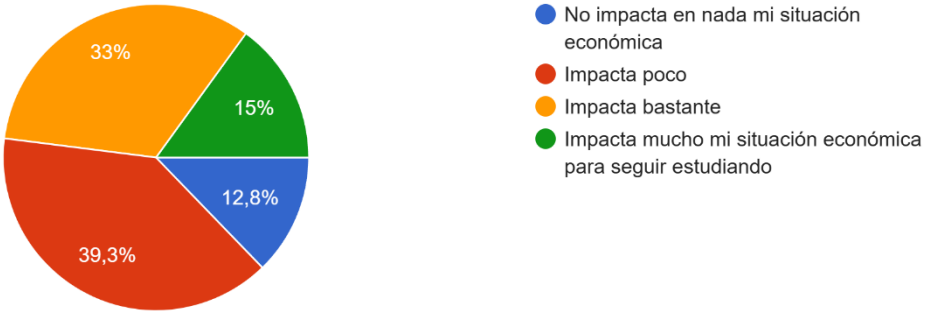


Figura 9

Distribución de respuestas a la pregunta P5: ¿En qué medida influye tu salud en tu capacidad para continuar con tus estudios?"

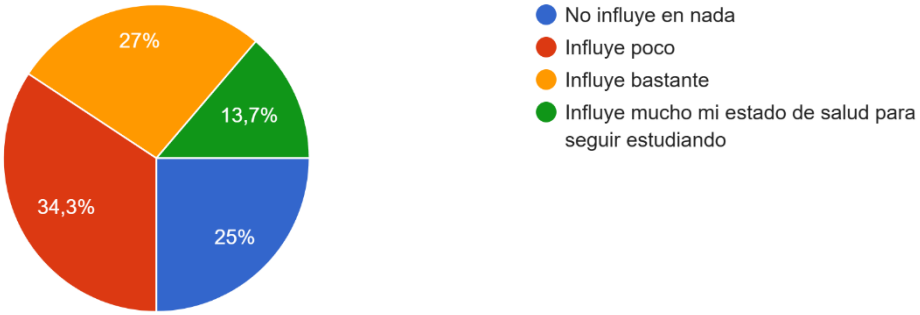


Figura 10

Distribución de respuestas a la pregunta P6: Si tienes una pareja, ¿cómo afecta esto a tu desempeño académico?

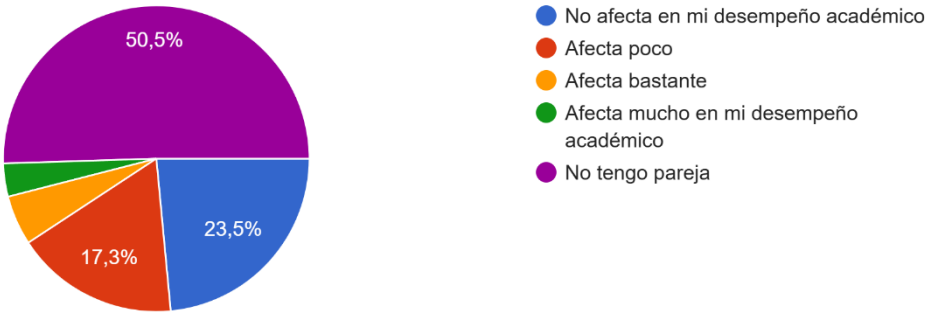


Figura 11

Distribución de respuestas a la pregunta P7: ¿Las otras actividades que desempeñas afectan tu tiempo de estudio?"

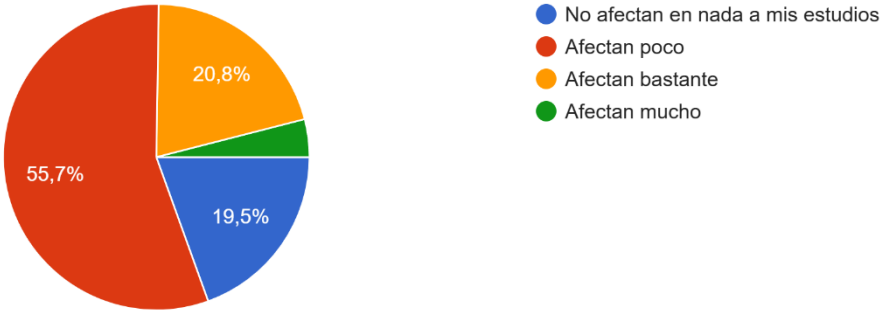


Figura 12

Distribución de respuestas a la pregunta P8: Si experimentas violencia familiar, ¿cómo afecta esto a tu rendimiento académico?

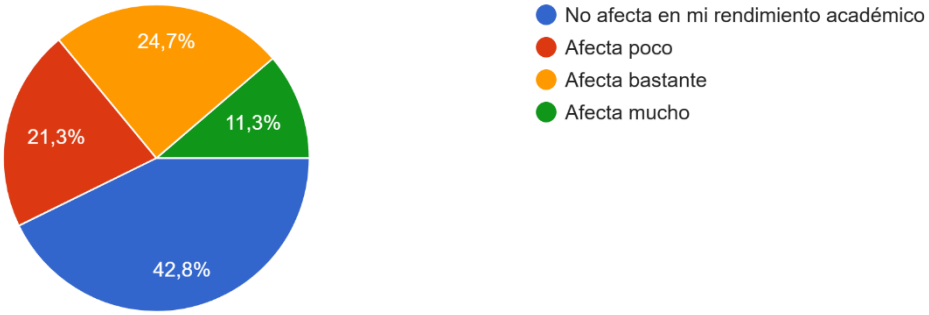


Figura 13

Distribución de respuestas a la pregunta P9: ¿En qué medida la interacción con tus profesores afecta tu desempeño académico?

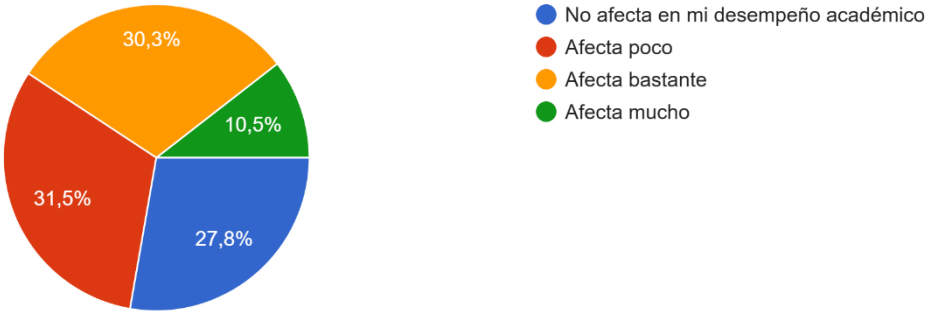


Figura 14

Distribución de respuestas a la pregunta P10: ¿Cómo afecta el acoso de tus compañeros a tu rendimiento académico?

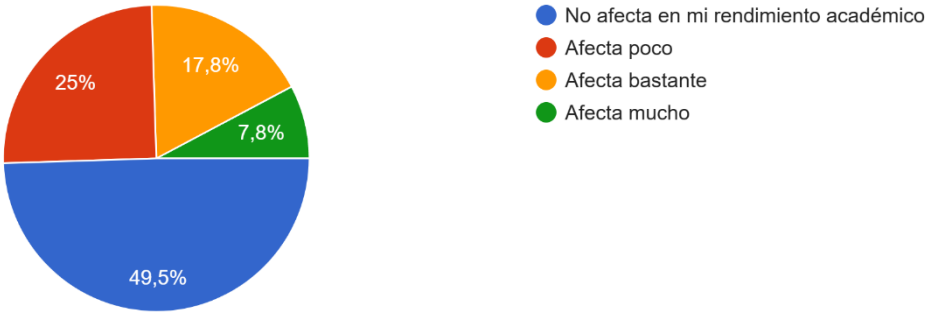


Figura 15

Distribución de respuestas a la pregunta P11: ¿La distancia entre tu hogar y la universidad afecta tu asistencia a clases?

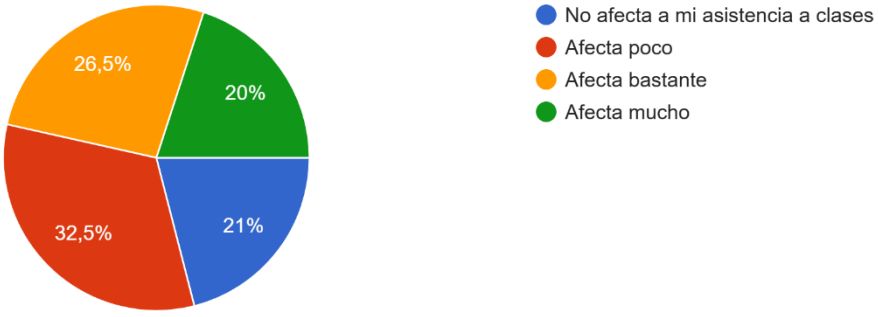


Figura 16

Distribución de respuestas a la pregunta P12: ¿Te sientes cómodo con tus compañeros de clase?

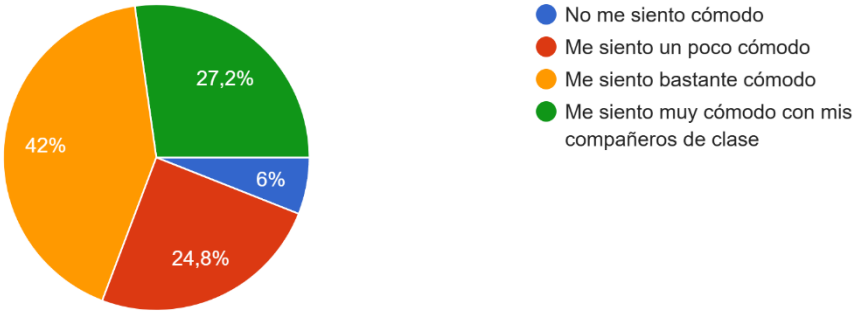


Figura 17

Distribución de respuestas a la pregunta P13: ¿Has experimentado discriminación en la universidad?



Figura 18

Distribución de respuestas a la pregunta P14: ¿Crees que dedicas suficiente tiempo al estudio de tus materias?

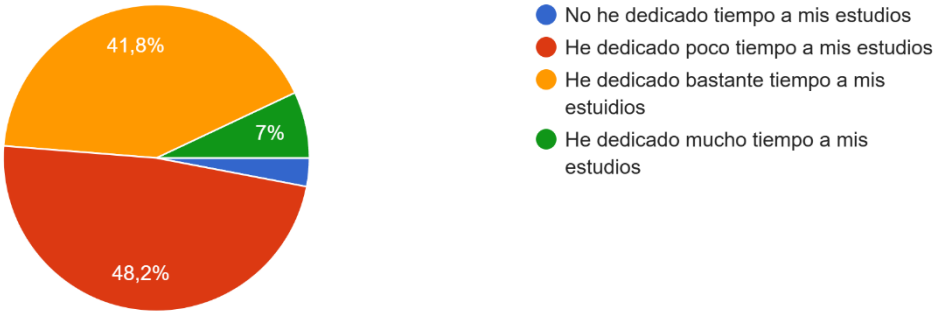


Figura 19

Distribución de respuestas a la pregunta P15: ¿Consideras que tu carrera te facilitará encontrar empleo?

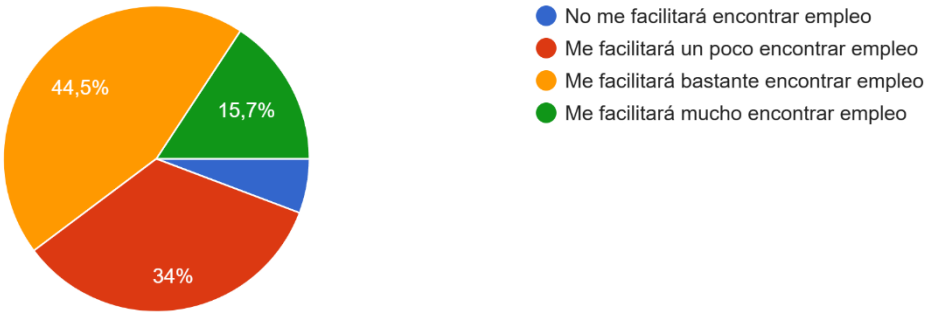


Figura 20

Distribución de respuestas a la pregunta P16: ¿Tus expectativas académicas se cumplieron después de salir del colegio?

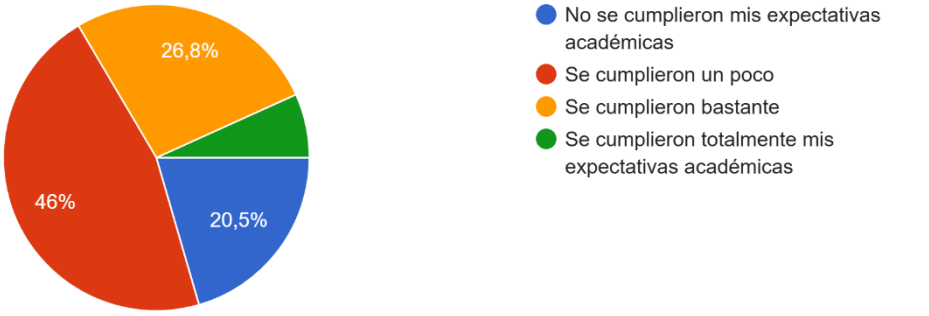


Figura 21

Distribución de respuestas a la pregunta P17: ¿Los conocimientos adquiridos en la universidad te son útiles en otros aspectos de tu vida?



Figura 22

Distribución de respuestas a la pregunta P18: ¿La influencia de tu familia afectó tu elección de carrera?

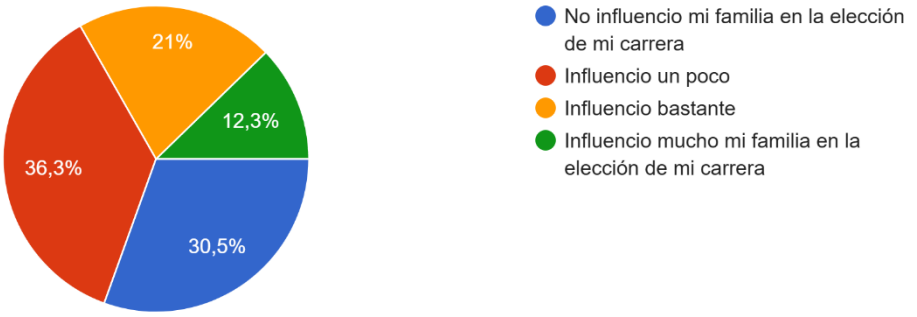


Figura 23

Distribución de respuestas a la pregunta P19: ¿Tenias en mente estudiar una carrera diferente a la que finalmente elegiste?

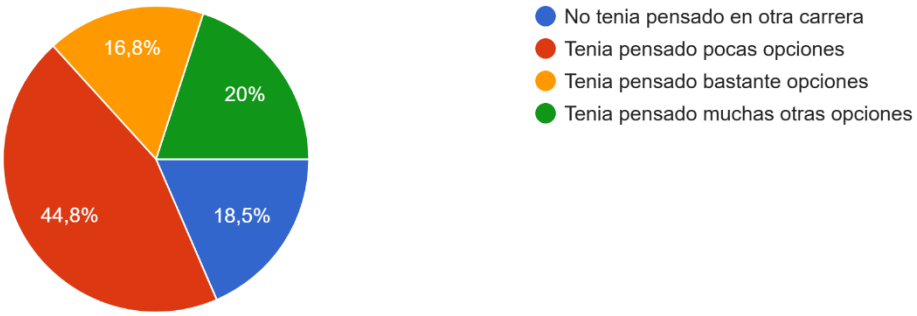
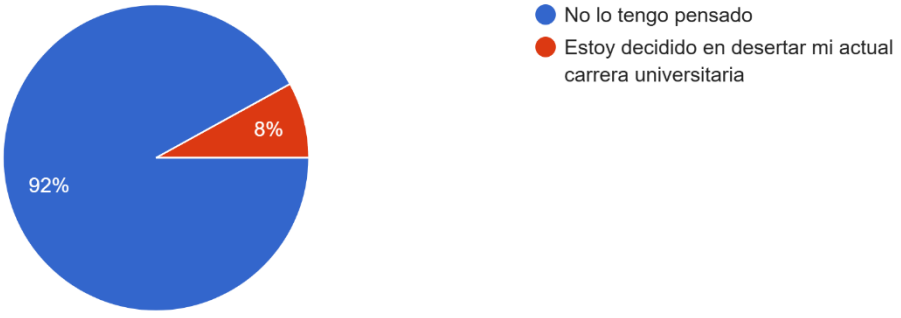


Figura 24

Distribución de respuestas a la pregunta R1: ¿Has pensado en desertar de tu carrera universitaria?



Anexo 04: Base de Datos con las Respuestas de la Encuesta

N°	Edad	Sexo	Ciclo	Carrera Profesional	1. ¿En qué medida dependes económicamente de alguien más?	2. Si tienes un empleo, ¿cómo influye en tu rendimiento académico?	3. Si tienes responsabilidades familiares, ¿Cómo afecta esto a tus estudios?	4. ¿Cuánto impacta tu situación económica en la posibilidad de continuar con tus estudios?	5. ¿En qué medida influye tu salud en tu capacidad para continuar con tus estudios?	6. Si tienes una pareja, ¿cómo afecta esto a tu desempeño académico?	7. ¿Las otras actividades que desempeñas afectan tu tiempo de estudio?	8. Si experimentas violencia familiar, ¿cómo afecta esto a tu rendimiento académico?	9. ¿En qué medida la interacción con tus profesores afecta tu desempeño académico?
1	23	M	X	Ing. De Sistemas	Dependo totalmente de otras personas	No influye en mi rendimiento académico	Afecta bastante en mis estudios	No impacta en nada mi situación económica	Influye bastante	Afecta poco	Afectan bastante	Afecta bastante	Afecta bastante
2	21	M	IX	Ing. De Sistemas	Dependo parcialmente de otras personas	Influye poco en mi rendimiento académico	Afecta poco en mis estudios	Impacta bastante	Influye poco	No tengo pareja	Afectan poco	No afecta en mi rendimiento académico	Afecta bastante
3	22	F	IX	Ing. De Sistemas	Dependo bastante	Influye poco en mi rendimiento académico	Afecta poco en mis estudios	Impacta bastante	Influye bastante	Afecta poco	Afectan poco	Afecta bastante	Afecta bastante
4	22	M	IX	Ing. De Sistemas	No dependo de otros	Influye bastante en mi rendimiento académico	No afecta en mis estudios	Impacta poco	Influye poco	No afecta en mi desempeño académico	Afectan poco	No afecta en mi rendimiento académico	No afecta en mi desempeño académico
5	24	M	IX	Ing. De Sistemas	Dependo parcialmente de otras personas	Influye bastante en mi rendimiento académico	Afecta poco en mis estudios	Impacta bastante	Influye bastante	Afecta mucho en mi desempeño académico	Afectan bastante	No afecta en mi rendimiento académico	Afecta poco
6	22	F	IX	Ing. De Sistemas	Dependo parcialmente de otras personas	Influye poco en mi rendimiento académico	No afecta en mis estudios	Impacta bastante	Influye bastante	No tengo pareja	Afectan bastante	Afecta bastante	Afecta poco
7	21	M	V	Administración	No dependo de otros	Influye poco en mi rendimiento académico	Afecta poco en mis estudios	Impacta bastante	Influye poco	No tengo pareja	No afectan en nada a mis estudios	No afecta en mi rendimiento académico	Afecta poco
8	24	F	IX	Ing. De Sistemas	No dependo de otros	Influye bastante en mi rendimiento académico	Afecta bastante en mis estudios	Impacta bastante	Influye mucho mi estado de salud para seguir estudiando	No afecta en mi desempeño académico	Afectan poco	Afecta mucho	No afecta en mi desempeño académico
9	21	F	III	Agronomía	Dependo bastante	Influye poco en mi rendimiento académico	Afecta poco en mis estudios	Impacta bastante	Influye bastante	Afecta bastante	No afectan en nada a mis estudios	Afecta bastante	Afecta poco
10	19	F	III	Ing. De Sistemas	Dependo bastante	No tengo empleo	Afecta poco en mis estudios	Impacta poco	Influye bastante	No tengo pareja	Afectan poco	No afecta en mi rendimiento académico	No afecta en mi desempeño académico
11	21	F	IX	Ing. De Sistemas	Dependo parcialmente de otras personas	No influye en mi rendimiento académico	No afecta en mis estudios	No impacta en nada mi situación económica	No influye en nada	No tengo pareja	No afectan en nada a mis estudios	No afecta en mi rendimiento académico	No afecta en mi desempeño académico

Nota: Esta tabla muestra un fragmento representativo de la base de datos. El archivo completo contiene registros de 316 estudiantes y 20 preguntas del cuestionario aplicado.

ANEXO 05: Autorización de la Institución para Realizar la Investigación

UNIVERSIDAD NACIONAL DE CAÑETE
Vicepresidencia Académica
Ley de Creación N° 29488
"Año del Bicentenario, de la consolidación de nuestra Independencia, y de la conmemoración de las heroicas batallas de Junín y Ayacucho"

San Vicente de Cañete,08 de abril de 2024

CARTA N° 002-2024-UNDC/P/VPAC

Magister:

AMANDA DURAN CARHUAMACA
Asesora de Tesis en la carrera de Ingeniería de sistemas
Docente Nombrada en la UNDC

Presente .-

ASUNTO : Autorización de ejecución de investigación.

De mi especial consideración:

Con mi cordial saludo me dirijo a usted y a la vez este despacho autoriza la ejecución del Proyecto de Investigación titulado "Modelo Predictivo basado en Machine Learning para el Diagnóstico Temprano en la Deserción Universitaria de la UNDC, periodo 2020 – 2022" en la Universidad Nacional de Cañete, a cargo del estudiante Hugo Jesús Sulca Martínez.

En ese sentido, se informa que se brindará todas las facilidades respectivas para el desarrollo de la mencionada investigación.

Sin otro en particular me suscribo de usted.

Atentamente,

FIRMA DIGITAL
Firmado digitalmente por: PEÑA HUAMAN ROGER HERNANDO FIR 25689655 hard
Motivo: Soy el autor del documento
Fecha: 08/04/2024 16:09:56-0500

Dr. ROGER HERNANDO PEÑA HUAMAN
Vicepresidente Académico de la UNDC

Cc: Archivo

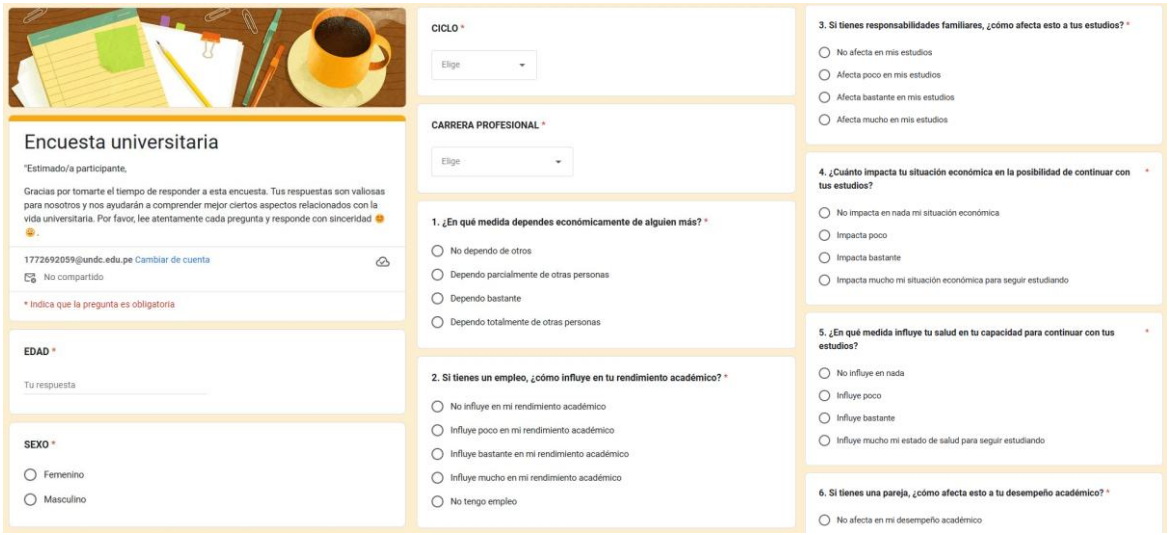
RHPH/yma

ANEXO 06: Desarrollo de la Solución Tecnológica

Se aplica la encuesta al alumnado de la Universidad Nacional de Cañete, usando Google Forms.

Figura 25

Capturas del formulario de encuesta aplicado mediante Google Forms



Se tiene un total de 400 alumnos encuestados.

Figura 26

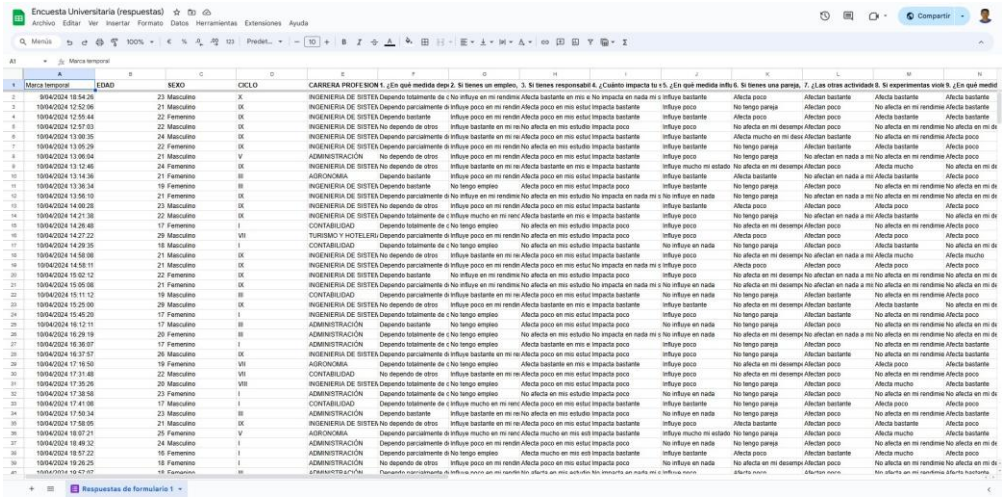
Captura del total de respuestas



Las respuestas se pasan a formato Excel.

Figura 27

Resultados de la encuesta en Google Sheets



Para aplicar la regresión logística, se utilizó código con el lenguaje de programación Python en el entorno de Google Colaboratory.

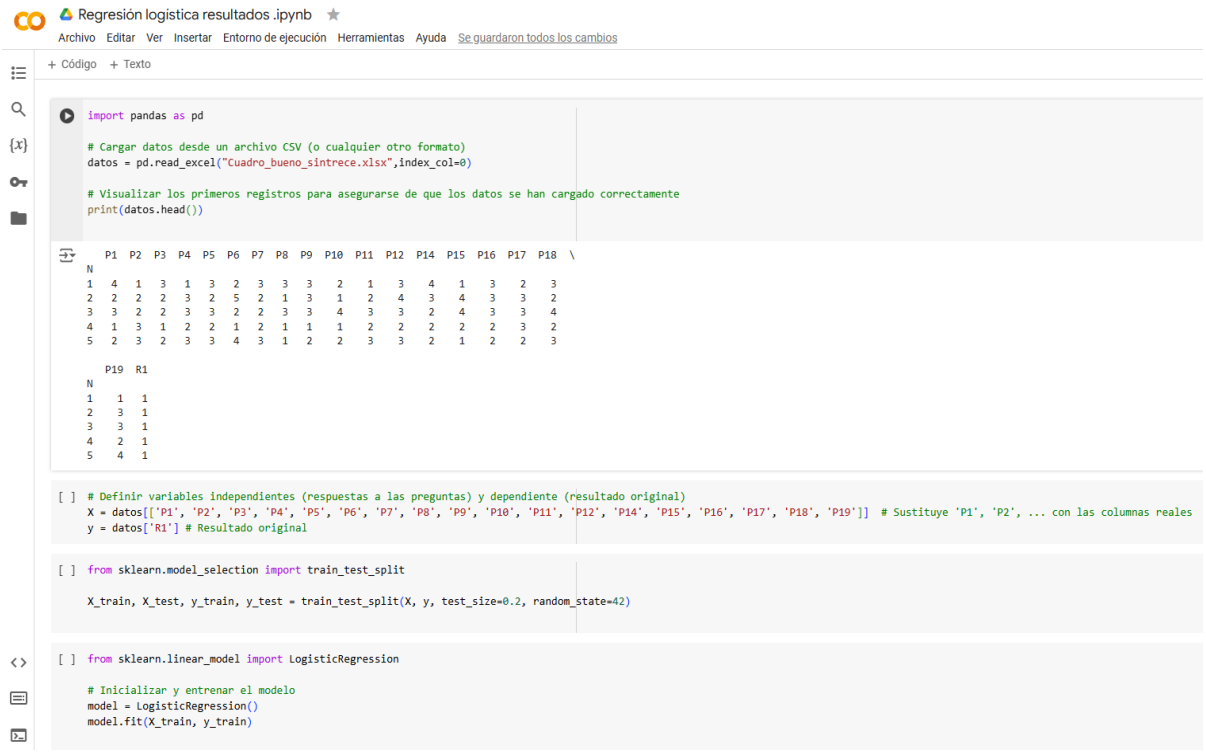
Figura 28

Código escrito en Python

```
import pandas as pd
data = pd.read_excel("Cuadro_aleatorio.xlsx",index_col=0)
print(data.head())
X = data[['P1', 'P2', 'P3', 'P4', 'P5', 'P6', 'P7', 'P8', 'P9', 'P10', 'P11', 'P12',
'P13', 'P14', 'P15', 'P16', 'P17', 'P18', 'P19']]  # Sustituye 'P1', 'P2', ... con las
columnas reales
y = data['R1']
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
model.fit(X_train, y_train)
from sklearn.metrics import accuracy_score, classification_report
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print("Precisión del modelo:", accuracy)
print(classification_report(y_test, y_pred))
data['R2'] = model.predict(X)
print(data['R2'])
```

Figura 29

Captura del código aplicado en el entorno Google Colab



Documento el código para entender el propósito y la funcionalidad del código.

Figura 30

Código documentado

```
# Importamos la biblioteca pandas
import pandas as pd
# Leemos los datos del archivo Excel con las respuestas de los alumnos
data = pd.read_excel("Cuadro_aleatorio.xlsx",index_col=0)
# Imprimimos las primeras filas de los datos para asegurarse de que los datos se han cargado correctamente
print(data.head())
# Definir variables independientes (respuestas a las preguntas) y dependiente (resultado original)
X = data[['P1', 'P2', 'P3', 'P4', 'P5', 'P6', 'P7', 'P8', 'P9', 'P10', 'P11', 'P12', 'P13', 'P14', 'P15', 'P16',
'P17', 'P18', 'P19']]
# Seleccionamos la columna objetivo para el modelo (Respuesta Original)
y = data['R1']
# Importamos la función para dividir los datos en entrenamiento y prueba
from sklearn.model_selection import train_test_split
# Dividimos los 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)
# Importamos el modelo de Regresión Logística
from sklearn.linear_model import LogisticRegression
# Creamos una instancia del modelo
model = LogisticRegression()
# Entrenamos el modelo con los datos de entrenamiento
model.fit(X_train, y_train)
# Importamos las métricas de precisión y reporte de clasificación
from sklearn.metrics import accuracy_score, classification_report
# Predecimos los valores de la variable objetivo para los datos de prueba
y_pred = model.predict(X_test)
# Calculamos la precisión del modelo
accuracy = accuracy_score(y_test, y_pred)
# Imprimimos la precisión del modelo
print("Precisión del modelo:", accuracy)
# Imprimimos el reporte de clasificación
print(classification_report(y_test, y_pred))
# Predecimos los valores de la variable objetivo para todos los datos
data['R2'] = model.predict(X)
# Imprimimos la nueva columna de predicciones
print(data['R2'])
# Imprimimos las primeras filas de los datos
data.head()
```

Como resultado de la precisión del modelo tenemos:

Figura 31
Resultados del modelo de regresión logística

```
Precisión del modelo: 0.8064516129032258
              precision    recall  f1-score   support

           1       0.82      0.94      0.88        69
           2       0.71      0.42      0.53        24

    accuracy                           0.81        93
   macro avg       0.77      0.68      0.70        93
weighted avg       0.79      0.81      0.79        93
```

La precisión del modelo (accuracy), tiene como resultado 0.8064516129032258, lo que significa que el modelo predijo correctamente el 80,64% de los casos.

Precision, para la clase 1 (Sí), es 0.82, significa que el modelo predice que un ejemplo pertenece a la clase 1, acierta el 82% de las veces. Para la clase 2 (No), la precisión es 0.71.

F1 Score, combina precision y recall en un solo número utilizando la media armónica. Un F1-score perfecto es 1, y un F1-score peor es 0. Para la clase 1 el F1-score es 0.88 y para la clase 2 el F1-score es 0.53.

Support, es la cantidad de ejemplos reales de cada clase en el conjunto de datos de prueba. Hay 69 ejemplos de la Clase 1 y 24 ejemplos de la Clase 2.

ANEXO 07: Autorización de Inclusión de Tesis en el Repositorio de la UNDC

UNIVERSIDAD NACIONAL DE CAÑETE

AUTORIZACIÓN DE PUBLICACIÓN EN EL REPOSITORIO INSTITUCIONAL DE LA UNDC

I INFORMACIÓN DEL DOCUMENTO

1.1 Título del documento:

MODELO PREDICTIVO BASADO EN MACHINE LEARNING PARA EL DIAGNÓSTICO TEMPRANO EN LA DESERCIÓN UNIVERSITARIA DE LA UNDC, PERIODO 2020 – 2022

1.2 Descripción (Seleccionar)

Tesis de Pregrado	X	Libro	
Tesis de Maestría		Capítulo de libro	
Tesis de Doctorado		Monografía	
Trabajo de Investigación		Artículo	
Tesis de Segunda Especialidad		Conferencia	
Trabajo de Suficiencia Profesional		Programa informativo	
Trabajo académico		Datos	

1.3 Autores del documento

Apellidos completos	Nombres completos	Correo electrónico	DNI	ORCID
Sulca Martínez	Hugo Jesús	1772692059@undc.edu.pe	72692059	0009-0003-3177-7605

1.4 Asesor (es)

Apellidos completos	Nombres completos	Correo electrónico	DNI	ORCID
Durán Carhuamaca	Amanda	aduran@undc.edu.pe	20114878	0000-0001-8183-5891
Oseda Gago	Dulio	doseda@undc.edu.pe	20044737	0000-0002-3136-6094

1.5 Centro de Investigación

Facultad	Ingeniería
Escuela Profesional	Ingeniería de Sistemas

I. ORIGINALIDAD DEL TRABAJO PRESENTADO

Con la presentación de esta ficha, el(los) autor(es) señalan expresamente que la obra es original, ya que sus contenidos son producto de su directa contribución intelectual. Se reconoce también que todos los datos y las referencias a materiales ya publicados están debidamente identificados con su respectivo crédito e incluidos en las notas bibliográficas y en las citas que se destacan como tal,

III. AUTORIZACIÓN DEL TRABAJO PRESENTADO

Toda copia de este documento, sea del entorno virtual o del documento original en físico es considerada "copia no controlada".

3.1 Por medio del presente documento, afirmo y garantizo(amos) ser el (los) legítimo (s), titular(es) de todos los derechos de propiedad intelectual sobre los entregables, las obras, los contenidos, los productos y/o las creaciones en general (en adelante, los "Contenidos" que serán incluidos en el Repositorio Institucional de la Universidad Nacional de Cañete (en adelante, la "Universidad").

3.2 Autorizo a la Universidad a publicar los Contenidos en el Repositorio Institucional de la Universidad y, en consecuencia, en el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, sobre la base de lo establecido en la Ley N* 30035, sus normas reglamentarias, modificatorias, sustitutorias y conexas, y de acuerdo con las políticas de acceso abierto que la Universidad aplique en relación con su Repositorio Institucional. Autorizo expresamente toda consulta y uso de los Contenidos, por parte de cualquier persona, por el tiempo de duración de los derechos patrimoniales de autor y derechos conexos, a título gratuito y a nivel mundial.

3.3 En consecuencia, la Universidad tendrá la posibilidad de divulgar y difundir los Contenidos, de manera total o parcial, sin limitación alguna y sin derecho a pago de contraprestación, remuneración ni regalía alguna a favor mío; en los medios, canales y plataformas que la Universidad y/o el Estado de la República del Perú determinen, a nivel mundial, sin restricción geográfica alguna y de manera indefinida, pudiendo crear y/o extraer los metadatos sobre los Contenidos, e incluir los Contenidos en los índices y buscadores que estimen necesarios para promover su difusión.



Firma

10/03/2025

Fecha

IV: PUBLICACIÓN DIFERIDA

Este ítem solo es de interés para los autores que han decidido la publicación diferida de sus documentos. Caso contrario, obviar este apartado.

Autorización	Motivo (marcar)		Firma
Solicito la publicación diferida del documento depositado en el repositorio, por 12 meses. (indicar el tiempo que solicita)	Exclusividad de revista, editor		
	Por patente		
	Secreto o seguridad nacional		
	Por otras razones		

Cualquier motivo utilizado, el solicitante debe acreditar con documento su pedido (subir junto con la autorización la evidencia que sustenta su solicitud).

Toda copia de este documento, sea del entorno virtual o del documento original en físico es considerada "copia no controlada".

ANEXO 08: Constancia de Similitud

UNIVERSIDAD NACIONAL DE CAÑETE
Ley de Creación N° 29488
FACULTAD DE INGENIERÍA
UNIDAD DE INVESTIGACION DE LA FACULTAD DE INGENIERIA
"Año del Bicentenario, de la Consolidación de nuestra Independencia
Y de la conmemoración de las heroicas batallas de Junín y Ayacucho"

San Vicente de Cañete, 16 de diciembre del 2024

CONSTANCIA DE SIMILITUD DE INFORME FINAL DE TESIS N° 024-2024

Título del Trabajo de Investigación: "Modelo predictivo basado en machine learning para el diagnóstico temprano en la deserción universitaria de la UNDC, periodo 2020 - 2022"

Autor(es): Hugo Jesús Sulca Martínez

HUGO JESUS SULCA MARTINEZ

INFORME DE ORIGINALIDAD

12%	11%	4%	6%
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	repositorio.ucv.edu.pe Fuente de Internet	2%
2	Submitted to Universidad Nacional de Cañete Trabajo del estudiante	2%
3	repositorio.untels.edu.pe Fuente de Internet	1%
4	clabes.uct.cl Fuente de Internet	1%
5	repositorio.espe.edu.ec Fuente de Internet	1%
6	bibliotecadigital.udea.edu.co Fuente de Internet	1%
7	repositorio.une.edu.pe Fuente de Internet	1%
8	www.apd.es Fuente de Internet	1%
9	hdl.handle.net Fuente de Internet	1%



Por lo tanto, se otorga esta constancia a solicitud del interesado(a) para los fines que estime conveniente.

FIRMA DIGITAL
Firmado digitalmente por: DURAN CARHUAMACA AMANDA FIR 20114878 hard
Motivo: Soy el autor del documento
Fecha: 16/12/2024 16:17:30-0500

Mg. Amanda Durán Carhuamaca
(e) Directora de la Unidad de Investigación
Facultad de Ingeniería – UNDC

ANEXO 09
Artículo científico

MODELO PREDICTIVO BASADO EN MACHINE LEARNING PARA EL DIAGNÓSTICO TEMPRANO EN LA DESERCIÓN UNIVERSITARIA DE LA UNDC, PERIODO 2020 – 2022

SULCA, Hugo
1772692059@undc.edu.pe
Escuela Profesional de Ingeniería de Sistemas, Facultad de Ingeniería, Universidad Nacional de Cañete

1. INTRODUCCIÓN

1.1. Planteamiento del problema

La deserción universitaria es un problema que afecta a muchos jóvenes y tiene consecuencias negativas para su desarrollo personal, profesional y social. Según el Ministerio de Educación, la deserción universitaria en el Perú alcanza el 30%, siendo motivada especialmente por la falta de una buena orientación vocacional y por razones económicas [1]. Además, la COVID-19 ha agravado esta situación, al generar dificultades de conectividad, bienestar y rendimiento académico.

Según el estudio: "Deserción universitaria de los estudiantes de una universidad peruana", los autores identificaron varias dimensiones que contribuyen a la deserción universitaria: factores personales, factores educativos, factores institucionales y factores financieros. Los factores personales incluyen problemas personales o familiares que pueden afectar la capacidad del estudiante para continuar sus estudios. Los factores educativos se refieren a las dificultades académicas, como el rendimiento académico bajo o la falta de interés en el curso. Los factores institucionales pueden incluir la falta de apoyo de la universidad o la insatisfacción con los servicios proporcionados. Los factores financieros se refieren a las dificultades económicas que pueden impedir que los estudiantes continúen sus estudios [2]. Estos factores pueden generar frustración, estrés y baja autoestima en los estudiantes, lo que afecta su motivación y compromiso con el aprendizaje.

Ante este escenario, se hace necesario identificar los factores que influyen en la deserción universitaria, así como predecir y prevenir los casos de riesgo. Para ello, se puede recurrir al uso de machine learning, una rama de la inteligencia artificial que permite analizar grandes volúmenes de datos y extraer patrones y conocimiento útil para la toma de decisiones.

El machine learning ha sido aplicado para el estudio de la deserción universitaria en diferentes contextos y países, obteniendo resultados prometedores. Por ejemplo, en Venezuela se evaluaron diferentes algoritmos de machine learning para

predecir la deserción universitaria con variables académicas y socioeconómicas de los estudiantes [3]. En Chile se aplicaron técnicas de machine learning para analizar los factores que influyen en la deserción y el retraso académico, así como para identificar patrones de comportamiento de los estudiantes [4]. En España se utilizó inteligencia artificial y machine learning para diseñar un sistema de alerta temprana que permita detectar y prevenir la deserción universitaria [5].

En el Perú también se han realizado algunos estudios que aplican machine learning para el problema de la deserción universitaria. Por ejemplo, se diseñó un sistema para predecir la deserción de los alumnos de la Universidad Tecnológica del Perú usando machine learning y los patrones de comportamiento de los alumnos desertores [6]. También, se evaluaron diferentes algoritmos de machine learning para predecir la deserción universitaria con variables académicas y socioeconómicas de los estudiantes [3].

El problema que se plantea en esta tesis es el siguiente: ¿De qué manera la implementación de un modelo predictivo basado en machine learning permite el diagnóstico temprano en la deserción universitaria de la UNDC en periodo 2020 – 2022? Este problema surge a partir del conocimiento de la tasa de deserción universitaria en el Perú siendo de 16.2% en 2022 [7] y sus implicancias negativas para los estudiantes, las instituciones y el país. Además, se considera que el uso de machine learning puede ofrecer una solución innovadora y eficaz para abordar este problema, al aprovechar las ventajas que ofrece esta tecnología para el análisis de datos y la generación de conocimiento.

Factores económicos

Es un factor crucial que puede influir en la decisión de un estudiante de continuar o abandonar sus estudios universitarios. Según un estudio realizado en la Universidad Interamericana para el Desarrollo (UNID) [8], los aspectos económicos que influyen en el abandono escolar de los estudiantes incluyen la necesidad de trabajar, la pérdida de porcentaje de beca, el rezago en pagos y la generación de intereses.

Estos factores económicos pueden ser una carga significativa para los estudiantes y sus familias. Si un estudiante tiene dificultades para pagar la matrícula, es posible que decida abandonar sus estudios. Además, los costos de vida, como alojamiento, alimentación y transporte, pueden poner presión financiera sobre el estudiante y hacer que sea más difícil continuar con sus estudios.

Situación académica

La “situación académica” es otro factor crucial que puede influir en la decisión de un estudiante de continuar o abandonar sus estudios universitarios. Según un estudio realizado en Colombia [9], la deserción universitaria se centra en la primera mitad de la carrera: al finalizar el cuarto semestre, el 39% de los estudiantes ya había abandonado sus estudios.

Esto puede deberse a varios factores relacionados con la situación académica del estudiante. Por ejemplo, las dificultades académicas, como un rendimiento académico bajo o la incapacidad para mantenerse al día con el material del curso,

pueden hacer que un estudiante considere abandonar sus estudios. Además, la carga de trabajo del curso puede ser abrumadora para algunos estudiantes, lo que puede contribuir a su decisión de abandonar.

Vocación profesional

La "vocación profesional" es un factor crucial que puede influir en la decisión de un estudiante de continuar o abandonar sus estudios universitarios. Según un estudio realizado por Xavier Erazo Guerra y Elena Rosero Morales [10], una adecuada orientación vocacional facilitará en los estudiantes el autoconocimiento de sus actitudes y aptitudes que les guiará a desempeñarse en la vida universitaria y laboral.

El objeto de la investigación fue determinar la orientación vocacional y su influencia en la deserción universitaria para la mejora de la retención estudiantil, e identificar las causas y factores para que los estudiantes abandonen sus estudios en una carrera universitaria.

Los autores encontraron que los factores externos e internos inciden en la decisión del ingreso a la universidad y pueden determinar la motivación y orientación en los bachilleres para cumplir con su proyecto de vida.

1.2. Formulación del problema

1.2.1. Problema General

¿De qué manera la implementación de un modelo predictivo basado en machine learning influye en el diagnóstico temprano la deserción universitaria de la UNDC en periodo 2020 – 2022?

1.2.2. Problemas Específicos

¿De qué manera la implementación de un modelo predictivo basado en machine learning que incorpora los factores económicos influye en el diagnóstico temprano la deserción universitaria de la UNDC en periodo 2020 – 2022?

¿De qué manera la implementación de un modelo predictivo basado en machine learning que incorpora la situación académica influye en el diagnóstico temprano la deserción universitaria de la UNDC en periodo 2020 – 2022?

¿De qué manera la implementación de un modelo predictivo basado en machine learning que incorpora la vocación profesional influye en el diagnóstico temprano la deserción universitaria de la UNDC en periodo 2020 – 2022?

1.3. Objetivos de investigación

1.3.1. Objetivo general

Implementar un modelo predictivo basado en machine learning que influye en el diagnóstico temprano en la deserción universitaria de la UNDC en periodo 2020 – 2022.

1.3.2. Objetivos específicos

Implementar un modelo predictivo basado en machine learning, que incorpora los factores económicos, que influye en el diagnóstico temprano en la deserción universitaria de la UNDC en periodo 2020 – 2022.

Implementar un modelo predictivo basado en machine learning, que incorpora la situación académica, que influye en el diagnóstico temprano en la deserción universitaria de la UNDC en periodo 2020 – 2022.

Implementar un modelo predictivo basado en machine learning, que incorpora la vocación profesional, que influye en el

diagnóstico temprano en la deserción universitaria de la UNDC en periodo 2020 – 2022.

1.4. Justificación de la investigación

1.4.1. Justificación Teórica

La justificación teórica radica en la necesidad de abordar de manera más eficaz y proactiva el problema de la deserción universitaria. Este fenómeno, que afecta a las instituciones de educación superior en todo el mundo, tiene implicaciones profundas tanto para los estudiantes como para las instituciones. Al aplicar técnicas de machine learning a los datos del estudiante, podemos desarrollar modelos que pueden predecir qué estudiantes tienen más probabilidades de abandonar. Esto permite a las instituciones identificar a los estudiantes en riesgo temprano y ofrecer intervenciones dirigidas para apoyar a estos estudiantes.

1.4.2. Justificación Práctica

La justificación práctica radica en su potencial para mejorar la retención de estudiantes en las universidades. Al identificar a los estudiantes en riesgo, las universidades pueden dirigir sus recursos y esfuerzos hacia estos estudiantes, lo que puede resultar en una mayor eficiencia y efectividad en sus esfuerzos de retención. También puede tener implicaciones significativas para los propios estudiantes, al recibir intervenciones tempranas y apoyo dirigido, los estudiantes en riesgo pueden tener una mayor probabilidad de continuar con sus estudios, lo que puede resultar en mejores resultados educativos y oportunidades de carrera.

1.4.3. Justificación Metodológica

La justificación metodológica radica en la capacidad de las técnicas de machine learning para analizar grandes cantidades de datos y detectar patrones complejos. El machine learning es una rama de la inteligencia artificial que se basa en algoritmos y modelos estadísticos para realizar tareas. En el contexto de la deserción universitaria, los modelos de machine learning pueden ser entrenados en datos históricos de estudiantes para aprender los patrones y factores que contribuyen a la deserción. Una vez entrenados, estos modelos pueden ser utilizados para predecir la probabilidad de deserción de los estudiantes actuales basándose en sus características individuales. Además, el uso de machine learning permite que el modelo se actualice y mejore con el tiempo a medida que se disponga de más datos. Esto significa que el modelo puede adaptarse a cambios en los patrones de deserción y seguir siendo efectivo a medida que cambian las circunstancias.

Los resultados de la investigación servirán para la Universidad Nacional de Cañete en:

- Identificar los factores que influyen en la deserción de sus estudiantes y su peso relativo en el modelo predictivo.
- Proponer recomendaciones para prevenir y reducir la deserción de sus estudiantes, basadas en el análisis de los datos y el conocimiento generado por el modelo predictivo.
- Aportar al avance del conocimiento científico sobre la deserción universitaria y el uso de machine learning como herramienta de investigación.

1.5. Delimitación de la investigación

La investigación se delimita de la siguiente manera:

1.5.1. Delimitación teórica o conceptual

Se basa en los conceptos y técnicas de machine learning, así como en los estudios previos sobre la deserción universitaria y sus factores asociados.

1.5.2. Delimitación temporal

Se considera el periodo 2020 - 2022, que abarca desde el inicio de la emergencia sanitaria por el COVID-19 hasta el año 2022.

1.5.3. Delimitación espacial

Se enfoca en la Universidad Nacional de Cañete, en la que empecé mis estudios universitarios el año 2017 y egresé el 2022.

2. CAPÍTULO II – MARCO TEÓRICO

2.1. Antecedentes de investigación

2.1.1. Artículos científicos

La deserción universitaria es un problema que afecta tanto a los individuos como a la sociedad, ya que implica una pérdida de oportunidades educativas, laborales y sociales para los estudiantes que abandonan sus estudios, así como un desperdicio de recursos y talentos para el sistema educativo y el desarrollo nacional. Por ello, es importante identificar a los estudiantes que tienen un mayor riesgo de deserción y brindarles apoyo y orientación para que puedan continuar y completar su formación académica.

El machine learning ha sido aplicado para el estudio de la deserción universitaria en diferentes contextos y países, obteniendo resultados prometedores. A continuación, se presentan algunos ejemplos han utilizado esta tecnología para abordar el problema de la deserción universitaria:

Villa et al. [11] en su artículo titulado: “A methodological proposal to address the academic dropout phenomenon based on an intelligent prediction model: a case study” publicado en la revista Millenium, tuvo como objetivo diagnosticar tempranamente los posibles factores que causan la deserción en los estudiantes universitarios, usando modelos de predicción que combina el Análisis de Supervivencia, Árboles de Decisión y Random Forest, dividido en tres fases. En la primera fase se determina la probabilidad de que un estudiante permanezca en la universidad, en la segunda fase se identifica los patrones de supervivencia en relación a las variables estudiadas (calificaciones, edad, género, etc.), en la tercera fase se identifica las variables más importantes para predecir la deserción. El estudio fue aplicado en una universidad chilena donde se identificó que las puntuaciones en lenguaje del examen de ingreso (PSU) son la variable más importante para predecir la deserción universitaria, también se observó que aproximadamente el 25% de los estudiantes abandonaron sus estudios durante el primer año. En conclusión, la combinación de Análisis de Supervivencia, Árboles de Decisión y Bosque Aleatorio en un modelo de predicción inteligente, ofrece una herramienta poderosa para abordar el problema de la deserción académica en las universidades. A pesar de que el estudio se desarrolló en el contexto de una universidad chilena, la metodología propuesta es flexible y adaptable a diferentes realidades académicas.

Valero et al. [3] en su artículo titulado: "Deserción universitaria: Evaluación de diferentes algoritmos de Machine Learning para su predicción" publicado en la revista de Ciencias Sociales (Ve), cuyo objetivo es determinar que algoritmo de Machine Learning tiene mejor desempeño para detectar la deserción universitaria. Este artículo presenta una metodología basada en machine learning para predecir la probabilidad de deserción de los estudiantes de pregrado. Se utilizaron datos históricos de 652 estudiantes y se aplicaron diferentes técnicas de preprocesamiento, selección de variables y modelado. Se compararon los resultados de varios algoritmos de machine learning y se encontró que el mejor modelo fue el de bosques aleatorios, con una precisión del 86% y una curva ROC del 0.93. En conclusión, el modelo obtenido puede ayudar a predecir en los primeros ciclos de estudios, que alumnos tienen mayor probabilidad de abandonar sus estudios, así como, alertar a la oficina de bienestar par que identifique las necesidades de los estudiantes y planifique tutorías individuales y grupales.

Shuo-chang et al. [12] en el artículo titulado: "Precision education with statistical learning and deep learning: a case study in Taiwan" en la revista "International Journal of Educational Technology in Higher Education", que tiene como objetivo determinar cómo los datos masivos y la inteligencia artificial pueden ayudar a las universidades a comprender más precisamente los antecedentes de los estudiantes, según los cuales se pueden proporcionar tomas de decisiones adecuadas para prevenir la deserción. Para este estudio, se tomaron muestras de 3552 estudiantes de una universidad en Taiwán. Se utilizó un método de aprendizaje estadístico y un método de aprendizaje automático basado en redes neuronales profundas para predecir la probabilidad de abandono de las aulas universitarias. Los resultados revelaron que el rendimiento académico del estudiante (con respecto a la dinámica del porcentaje de clasificación de clase), las solicitudes de préstamos estudiantiles, el número de ausencias de la escuela y el número de materias alertadas predijeron con éxito si los estudiantes abandonarían la universidad o no con una tasa de precisión del 68% cuando se empleó el método de aprendizaje estadístico, y del 77% para el método de aprendizaje profundo. En esta investigación se concluye que estos resultados pueden permitir que la universidad proporcione intervenciones a los estudiantes para ayudar en la selección del curso y mejorar sus competencias según sus aptitudes, lo que podría reducir la tasa de abandono y facilitar el aprendizaje adaptativo.

Rodríguez et al. [13] presentó el artículo titulado "Deserción Universitaria: Nuevo Análisis Metodológico". Este artículo presenta una aplicación de machine learning para predecir la deserción escolar en el sistema de educación pública de Chile, utilizando bases de datos públicas del Ministerio de Educación. Los autores compararon diferentes modelos predictivos, como árboles de decisión, redes neuronales, máquinas de vectores de soporte y bosques aleatorios, con una regresión logística tradicional. Los resultados muestran que el modelo de redes neuronales tiene el mejor desempeño, con una precisión del 86% y una sensibilidad del 82%. Los autores también ofrecen algunas recomendaciones prácticas para otros investigadores y formuladores de

políticas educativas que buscan implementar sistemas de alerta temprana para la deserción escolar. Entre ellas, destacan la importancia de tener datos actualizados y confiables, de definir claramente el concepto de deserción escolar, de seleccionar las variables más relevantes para el análisis, de evaluar los modelos predictivos con criterios objetivos y de diseñar intervenciones personalizadas para los estudiantes en riesgo.

2.1.2. Tesis nacionales e internacionales

Hecho la revisión de literatura científica, se tiene los siguientes antecedentes en los componentes:

- **Tesis internacionales**

Cuzme [14] en su tesis de maestría titulada "Modelo de aprendizaje automático para la predicción del riesgo de deserción en estudiantes de la Universidad Técnica de Manabí", cuyo objetivo es elaborar un modelo predictivo para la estimación del riesgo de deserción en estudiantes de la Universidad Técnica de Manabí mediante la utilización de técnicas de aprendizaje automático y la selección de las características de mayor relevancia en la base de datos académica de la institución. En este estudio se utilizó como metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), siguiendo sus 6 fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación. A partir de las métricas de evaluación los resultados esperados fueron calculados con el algoritmo de clasificación Random Forest que realizó una mejor predicción sobre la deserción estudiantil con un margen de error de 0.05, valorado por las métricas del área bajo la curva AUC ROC de 0.95 y el Puntaje F1 de 0.95. El autor constató la irregularidad de los estudiantes en los primeros niveles académicos, por lo cual se verificaron que muchos estudiantes desertan en estos niveles durante su carrera universitaria.

Quintero [15] presentó la tesis de maestría titulada "Diseño de un modelo predictivo para generar alertas tempranas de deserción universitaria en los programas de pregrado presenciales de la Facultad de Ingeniería de la Universidad de Antioquia", cuyo objetivo fue diseñar un modelo predictivo que pueda identificar que estudiantes de la Facultad de Ingeniería de la Universidad de Antioquia están en riesgo de desertar, para ello usó técnicas de learning analytics y machine learning. Se emplearon los datos históricos contenidos en las bases de datos de la universidad y la información disponible en la plataforma del ICFES, y se aplicaron dos técnicas de machine learning, redes neuronales artificiales (RNA) y Xtreme gradient boosting (XG Boost). Los resultados mostraron que XG Boost tuvo el mejor desempeño con un accuracy de 74,91%, también se construyó un prototipo de software que usa el modelo entrenado y que recibe los datos de un nuevo estudiante y lo identifica como un posible desertor o no. El autor concluye que aunque el modelo predictivo desarrollado ha demostrado ser útil, aun se puede mejorar su precisión incluyendo más variables como por ejemplo la vocación profesional.

En su tesis doctoral "Minería de datos aplicada a estrategias para minimizar el rezago académico y la deserción universitaria en carreras de informática de la UNNOBA" Russo [16], busca definir y

desarrollar un modelo de intervención áulica virtual que permita detectar patrones de rendimiento académico y predecir, de forma temprana, potenciales casos de deserción. La principal fuente de información fue el sistema SIU-Guaraní de la UNNOBA. Además, se emplearon informes de ingreso de todas las carreras correspondientes al período 2011-2016, analizándolos mediante la técnica de clustering. También se utilizó técnicas de ARIMA para estimar predicciones futuras basándose en datos pasados, herramientas de Business Intelligence (BI) y la Suite Pentaho para el desarrollo y la gestión de la información. La investigación reveló que, durante el período 2010-2015, solo el 40% de los alumnos aprobó las tres materias del primer cuatrimestre requeridas para inscribirse en cuatro de las cinco asignaturas del segundo cuatrimestre. A partir de estos datos, el autor deduce que los índices de deserción y desgranamiento en estas carreras son significativamente altos durante los primeros dos años.

En su tesis "Aplicativo web para el análisis, selección y admisión de aspirantes al programa de ingeniería de sistemas de la Universidad De Cundinamarca en la extensión Facatativá utilizando modelos predictivos de minería de datos", Gómez [17] se propone crear una plataforma web que facilite el análisis y evaluación de patrones de conocimiento relacionados con los perfiles de los aspirantes al programa de Ingeniería de Sistemas. Esta herramienta tiene como finalidad respaldar las políticas de selección y admisión de la Universidad de Cundinamarca, extensión Facatativá. La investigación, de carácter aplicado y enfoque cuantitativo, estableció que el modelo de regresión logística es el método óptimo debido a su superior rendimiento en las pruebas comparativas. La esencia del proyecto radicó en la combinación de la minería de datos y el diseño de un aplicativo web, que no solo sirve para analizar las razones de deserción en la mencionada universidad, sino que también integra y visualiza métricas como el accuracy score, la curva de aprendizaje, la matriz de confusión y un árbol de decisión basado en los datos recopilados.

- **Tesis Nacionales**

Cevallos y Barahoma [18] presentaron una investigación rigurosa en su tesis denominada “Modelo para automatizar el proceso de predicción de la deserción en estudiantes universitarios en el primer año de estudio”. En este trabajo, los autores se embarcaron en la misión de diseñar e implementar un modelo tecnológico avanzado que, mediante técnicas de análisis predictivo, busca automatizar y mejorar la precisión en la identificación temprana de estudiantes universitarios con un alto riesgo de desertar durante su primer año académico.

Para lograr esto, adoptaron la metodología Ad-Hoc, garantizando la calidad y estructura del proceso investigativo al adherirse a las buenas prácticas recomendadas en la reconocida guía PMBOK. Un aspecto crucial de su investigación fue la identificación y definición de 20 variables críticas que sirven como indicadores representativos de los factores que más influyen en la deserción universitaria. Estas variables abarcan una gama amplia de enfoques, desde aspectos personales y socioeconómicos hasta dimensiones institucionales e interpersonales.

Tras un exhaustivo análisis y comparativa de técnicas, Cevallos y Barahoma llegaron

a la conclusión de que las redes bayesianas ofrecen una precisión superior en este contexto, logrando una tasa de predicción del 67.10%. Además, identificaron que la herramienta más eficaz y adecuada para llevar a cabo este tipo de análisis predictivo es el software IBM SPSS Modeler, dada su robustez y capacidades avanzadas en el manejo y modelado de datos.

De la Cruz [19], en su investigación presentada en la tesis titulada "Diseño de un modelo predictivo basado en Machine Learning para el control de la deserción de estudiantes en la universidad Ricardo Palma", abordó un problema persistente en la educación superior: la deserción estudiantil. Con el objetivo principal de crear y afinar un modelo predictivo utilizando técnicas avanzadas de Machine Learning, la investigación se centró específicamente en la situación de los estudiantes de la Universidad Ricardo Palma.

Para llevar a cabo este ambicioso proyecto, De la Cruz optó por el algoritmo de Regresión Logística, una herramienta poderosa que no solo arroja luz sobre las causas subyacentes de la deserción, sino que también puede predecir, con un grado de precisión, qué estudiantes tienen mayor probabilidad de abandonar sus estudios.

Sin embargo, antes de poder implementar y entrenar el algoritmo, fue esencial preparar y estructurar adecuadamente los datos disponibles. De la Cruz accedió a las bases de datos de la Universidad Ricardo Palma y, tras un proceso meticuloso, construyó un Datamart específico para consolidar la información relevante. Una vez que los datos estuvieron organizados, se almacenaron en una base de datos no relacional, específicamente Hive. Pero el proceso no terminó ahí; para garantizar un procesamiento eficiente y escalable de la información, los datos se integraron en una infraestructura de Big Data utilizando la tecnología Hadoop Distributed File System (HDFS). Este enfoque permitió no solo manejar grandes volúmenes de datos, sino también garantizar que el análisis y las predicciones se realizaran de manera oportuna y precisa.

En su detallada investigación, Albán [20] abordó un tema de creciente preocupación en el ámbito académico con su tesis: “Contribuciones a la predicción de la deserción universitaria a través de minería de datos”. El núcleo de su estudio se centra en cómo las técnicas avanzadas de minería de datos pueden enriquecer y mejorar la capacidad de prever la deserción estudiantil en el nivel universitario.

Mediante un análisis exhaustivo, Albán identificó siete factores tecnológicos determinantes que tienen un impacto negativo en la permanencia de los estudiantes en la universidad. Estos factores, que van desde el acceso y uso de tecnologías hasta el nivel de adaptación a las herramientas digitales educativas, se convirtieron en variables esenciales en los modelos predictivos que diseñó. Para tal fin, utilizó técnicas como la regresión logística y el árbol de decisión, herramientas poderosas en el ámbito de la ciencia de datos.

Adicionalmente, profundizó en cómo ciertos comportamientos y hábitos de los universitarios, en particular las adicciones relacionadas con la tecnología y medios digitales, pueden actuar como predictores significativos de la deserción. Este enfoque innovador le permitió obtener resultados notables. Utilizando el algoritmo CHAID,

logró alcanzar un asombroso 97.95% de precisión en su modelo predictivo.

La obra de Albán no solo arroja luz sobre la complejidad del fenómeno de la deserción universitaria, sino que también proporciona herramientas valiosas y efectivas para abordar el problema de manera proactiva y basada en datos.

En su investigación, titulada "Business Analytics para el pronóstico de deserciones de estudiantes en el Colegio Héroes de la Breña", Lozano [21] profundiza en la importancia y el impacto del Business Analytics en la anticipación y gestión de deserciones estudiantiles en el mencionado colegio. Su objetivo principal radica en discernir cómo esta herramienta tecnológica puede influir positivamente en la predicción de tales eventos.

El enfoque de la investigación es netamente cuantitativo, con una naturaleza aplicada y un diseño pre experimental. Lozano basó su estudio en una muestra compuesta por 152 registros, de los cuales solo 106 se consideraron válidos para el análisis. Estos datos se dividieron de manera que el 30% se destinó para el entrenamiento del modelo y el 70% restante para su evaluación y prueba.

El desarrollo de los modelos de clasificación se llevó a cabo utilizando el versátil lenguaje de programación Python, específicamente bajo el entorno Anaconda. Este entorno proporciona múltiples librerías especializadas que facilitan la implementación y ajuste de algoritmos de Machine Learning. Entre los algoritmos evaluados, el Random Forest destacó particularmente, logrando una exactitud del 59% en general y un impresionante 79% en la predicción específica de posibles desertores.

Al finalizar su estudio, Lozano llegó a la conclusión de que el Business Analytics, al ofrecer resultados tan prometedores en la predicción temprana de deserciones, se presenta como una herramienta valiosa para instituciones educativas. Su implementación no solo permite identificar a tiempo a los estudiantes en riesgo, sino que también brinda la oportunidad de tomar medidas proactivas para garantizar su continuidad y éxito académico.

2.2. Bases teóricas

2.2.1. Variable independiente

Para la presente investigación la variable independiente es: *"Modelo predictivo basado en Machine learning"*.

A. Definición

Un "Modelo predictivo basado en Machine Learning" es un tipo de modelo que utiliza algoritmos de aprendizaje automático para predecir resultados futuros basándose en patrones históricos y datos actuales. Estos modelos pueden aprender y mejorar a medida que se les proporciona más datos, lo que les permite hacer predicciones cada vez más precisas. Suelen seguir un proceso general que incluye la recopilación de datos, la limpieza y preparación de los datos, la selección de características, el entrenamiento del modelo, la validación del modelo y finalmente la implementación del modelo [22].

B. Dimensiones

i. **Tiempo**. De entrenamiento de un modelo, que es el tiempo que se necesita para entrenar un modelo en un conjunto de datos específicos [23].
ii. **Precisión.** es una métrica para evaluar los modelos de clasificación. Informalmente, la precisión es la

fracción de predicciones que nuestro modelo acertó. Formalmente, la precisión tiene la siguiente definición: Precisión = Número de predicciones correctas / Número total de predicciones [24].

iii. **Pronóstico.** Se refiere a la salida de un algoritmo que ha sido entrenado en un conjunto de datos históricos. El algoritmo genera valores probables para variables desconocidas en cada registro de los nuevos datos [25].

C. Indicadores

i. **Tiempo de procesamiento del algoritmo.** Es el tiempo que tarda un algoritmo en completar su ejecución. Este tiempo puede variar dependiendo de varios factores, como la eficiencia del algoritmo, la velocidad del procesador y la cantidad de datos a procesar [26].

ii. **Tiempo de respuesta de una consulta.** Es el tiempo que tarda un sistema en responder a una consulta específica. En el contexto de las bases de datos, se refiere al tiempo que tarda en recuperar los datos solicitados [27].

iii. **Porcentaje de predicción acertada.** Este indicador, también conocido como precisión o exactitud, es una medida de cuántas predicciones realizadas por el modelo son correctas. Se calcula como el número total de predicciones correctas dividido por el número total de predicciones [28].

iv. **Puntaje F1.** Es una medida que combina la precisión y la exhaustividad (recall) para proporcionar una única métrica de rendimiento. Se calcula como la media armónica de la precisión y la exhaustividad [28].

v. **Accuracy (Exactitud).** Es una métrica que mide la proporción de predicciones correctas (tanto positivas como negativas) realizadas por el modelo en relación con el total de predicciones [29].

vi. **Recall (Exhaustividad).** Es una métrica que mide la proporción de verdaderos positivos que fueron correctamente identificados por el modelo [29].

D. Teorías

i. Inteligencia Artificial

La inteligencia es la capacidad de las computadoras para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones, como si fuera una persona [30].

Una IA tiene la capacidad de imitar funciones de la mente humana, como: aprendizaje, entendimiento, uso del lenguaje, etc.

El británico Alan Turing es considerado como pionero y fundador de la rama de la Inteligencia Artificial, quien desarrolló el modelo matemático de la máquina de Turing [31].

La inteligencia artificial ya está siendo usada en diferentes situaciones como, por ejemplo:

- Protección contra ciber ataques
- Reconocimiento visual
- Mejoras del desempeño en la estrategia comercial
- Distribución de contenido en las redes sociales
- Traducción automática de textos

- Ayuda en la toma de decisiones empresariales

ii. Machine Learning

El machine learning pertenece al sub campo de las ciencias de la computación, se encarga de construir algoritmos que funcionan a base de una colección de ejemplos de algún fenómeno. [32]
Tiene por objetivo, realizar una aproximación a la realidad, optimizando el error y la precisión de nuestro modelo.
En base a un registro histórico, para poder automatizarlo y relacionar los inputs y outputs en un lenguaje que podamos entender. [33]
El termino Machine Learning fue acuñado por Arthur L. Samuel, con su investigación en la creación de una IA en torno a un juego de damas, utilizando dos métodos: aprendizaje de memoria y aprendizaje por generalización. La computadora IBM 7094 pudo vencer al maestro de las damas Robert Nealey en 1962. [34]

iii. Tipos de Machine Learning

Según la APD [35], los tipos de machine learning son:

- **Aprendizaje Supervisado**

La máquina aprende en base de ejemplos que le brinda el operador, con un conjunto de datos que incluye entradas y salidas deseadas. El algoritmo identifica un patrón en los datos, realiza predicciones y es corregido por el operador, hasta que se consigue un alto nivel de rendimiento y precisión.

- **Aprendizaje sin supervisión**

El algoritmo estudia los datos para identificar patrones, sin la intervención de un operador. El algoritmo organiza los datos a través de grupos para describir su estructura. Mientras más datos reciba, mayor será su capacidad para tomar decisiones.

- **Aprendizaje por refuerzo**

El algoritmo trabaja en base a procesos reglamentados. Intenta explorar diferentes posibilidades y evaluando cada resultado hasta llegar al más óptimo. Aprende a través de ensayo y error.

iv. Tipos de algoritmos de machine learning

Según APD los tipos de algoritmos de machine learning son:

- **Algoritmos de regresión**

El algoritmo relaciona distintas variables, enfocándose en una variable dependiente. Se utiliza principalmente en análisis estadísticos, predicción y pronóstico.

- **Algoritmos bayesianos**

Están basados en el “teorema de Bayes”, cada valor es independiente de cualquier otro. Lo que permite predecir una clase en función de un conjunto de dado de características.

- **Algoritmos de agrupación**

Sirven para agrupar una serie de datos no etiquetados. Los grupos están representados por la variable K. Cada dato es enviado a un K grupo según sus características.

- **Algoritmos de árbol de decisión**

Es un algoritmo de clasificación y regresión, similar a un diagrama de flujo. Donde en cada nodo del árbol representa una prueba en una variable específica, y cada rama es el resultado de esa prueba.

- **Algoritmos de redes neuronales**

Están inspirado en el cerebro y como procesa la información. Son un gran número de procesos interconectados, que trabajan al unísono para resolver problemas específicos. Se usa para analizar datos de entrada complejos, como procesos comerciales, de fabricación o problemas empresariales.

- **Algoritmos de reducción de dimensión**

Son algoritmos que reducen el número de variables para encontrar la información exacta.

- **Algoritmos de aprendizaje profundo**

Son algoritmos que automatizan el análisis predictivo. Se ejecutan a través de redes neuronales, donde pasan una representación simplificada de los datos de capa a capa.

v. Regresión logística

Este método de análisis estadístico, se emplea frecuentemente en modelado y análisis predictivo, y se aplica también en el campo del aprendizaje automático. En este tipo de análisis, la variable dependiente es finita o categórica, pudiendo ser A o B (en el caso de la regresión binaria) o una serie de opciones finitas como A, B, C o D (en el caso de la regresión multinomial). Este método se utiliza en programas de estadística para entender la relación entre la variable dependiente y una o más variables independientes, estimando probabilidades mediante una ecuación de regresión logística [36].

vi. Tipos de regresión logística

Según IBM hay tres modelos de regresión logística, que se definen en función de una respuesta categórica.

- **Binario**

La respuesta o variable dependiente es de naturaleza dicotómica, es decir. sólo tiene dos resultados posibles.

- **Multinomial**

La variable dependiente tiene tres o más resultados posibles; sin embargo, estos valores no tienen un orden específico.

- **Ordinal**

Cuando la variable de respuesta tiene tres o más resultados posibles, pero en este caso, estos valores sí tienen un orden definido.

2.2.2. Variable dependiente

Para la presente investigación la variable dependiente es: *"Diagnóstico temprano de la deserción universitaria"*

A. Definición

El "Diagnóstico temprano de la deserción universitaria" se refiere a la identificación anticipada de los estudiantes que están en riesgo de abandonar sus estudios. Este diagnóstico se realiza a través del análisis de diversos factores como el rendimiento académico, la asistencia a clases, la participación en actividades extracurriculares, entre otros. El objetivo es intervenir a tiempo y proporcionar el apoyo necesario para prevenir la deserción. Este diagnóstico temprano puede ser crucial para mejorar las tasas de retención y graduación en las instituciones de educación superior [37].

B. Dimensiones

i. **Factores económicos.** Son las condiciones y recursos financieros que pueden influir en su capacidad para mantenerse y completar su educación superior.

ii. **Situación académica.** El rendimiento académico es un nivel de conocimientos demostrado en un área o materia comparado con la norma de edad y nivel académico.

iii. **Vocación profesional.** Se refiere a esa inclinación natural o llamado interior que una persona experimenta hacia un campo específico de trabajo o actividad. Es más que simplemente elegir una profesión; implica descubrir aquello que nos interesa, en lo que somos buenos y que nos permite sentirnos realizados y plenos en nuestra vida laboral [38].

C. Indicadores

i. **Carga financiera.** Es la suma de los distintos gastos financieros y la amortización del capital, referida a los préstamos formalizados y pendientes de reembolso [39].

ii. **Falta de recursos.** Se refiere a la escasez o insuficiencia de recursos básicos como agua, alimentos, energía, vivienda, etc., que se consideran fundamentales para satisfacer la supervivencia o de recursos no básicos que satisfacen distintas necesidades en las sociedades humanas en varios aspectos [40].

iii. **Financiamiento de estudios universitarios.** Se refiere a los recursos financieros que se utilizan para cubrir los costos asociados con la educación universitaria. Esto puede incluir matrícula, alojamiento, libros, suministros y otros gastos relacionados [41].

iv. **Rendimiento académico.** Es una evaluación del conocimiento adquirido en el ámbito escolar, terciario o universitario. Un estudiante con buen rendimiento académico es aquel que obtiene calificaciones positivas en los exámenes que debe rendir a lo largo de una cursada [42].

v. **Dificultad de los programas de estudio.** Se refiere a los desafíos y obstáculos que pueden surgir al tratar de cumplir con los requisitos y expectativas de un programa de estudio específico [43].

vi. **Apoyo académico.** Es el soporte proporcionado a los estudiantes para ayudarles a tener éxito en sus estudios. Esto puede incluir tutorías, asesoramiento académico, servicios de orientación y otros recursos disponibles para ayudar a los estudiantes a alcanzar sus objetivos académicos [44].

vii. **Tipo de Preparación.** Este término se refiere a la formación previa que un estudiante ha recibido antes de ingresar a un programa universitario específico. Esto puede incluir la educación secundaria, cursos preuniversitarios o cualquier otra forma de preparación académica [45].

viii. **Cursos matriculados.** Se refiere a los cursos en los que un estudiante se ha inscrito durante un semestre o año académico específico [46].

D. Teorías

i. Deserción Estudiantil

La deserción estudiantil es un fenómeno social y educativo que se refiere al abandono definitivo o temporal de los estudios por parte de los estudiantes, antes de obtener el título académico correspondiente, debido a diversos factores de índole personal, familiar, económica, académica o institucional. La deserción

estudiantil tiene consecuencias negativas tanto para los individuos como para la sociedad, ya que limita las oportunidades de desarrollo personal y profesional, y afecta el capital humano y el crecimiento económico de un país [47].

ii. Causas de la deserción estudiantil

La deserción estudiantil puede ocurrir en cualquier nivel educativo, pero se ha prestado especial atención al nivel universitario, debido a su importancia para la formación de profesionales calificados y competentes que puedan responder a las demandas del mercado laboral y a los desafíos del siglo XXI. Según datos de la UNESCO, la tasa promedio de deserción universitaria en América Latina y el Caribe es del 50%, lo que significa que la mitad de los estudiantes que ingresan a la educación superior no la concluyen.

Existen múltiples causas que explican la deserción universitaria, las cuales pueden agruparse en cuatro categorías: causas individuales, familiares, académicas e institucionales [47].

- **Causas individuales**

Se refieren a las características personales, psicológicas, motivacionales y vocacionales de los estudiantes, que influyen en su decisión de continuar o abandonar sus estudios. Algunos ejemplos son: falta de interés, bajo rendimiento, problemas de salud, embarazo, consumo de drogas, dificultades de adaptación o integración social, falta de habilidades o competencias para el estudio, falta de claridad sobre el proyecto de vida o la elección profesional, etc.

- **Causas familiares**

Se refieren al contexto familiar y socioeconómico de los estudiantes, que condiciona sus posibilidades y expectativas educativas. Algunos ejemplos son: bajo nivel educativo o cultural de los padres, falta de apoyo o supervisión familiar, problemas económicos o laborales, conflictos o violencia familiar, migración o movilidad geográfica, etc.

- **Causas académicas**

Se refieren al currículo, la metodología, la evaluación y el clima educativo que se desarrolla en la institución universitaria, que afecta el proceso de enseñanza-aprendizaje y la satisfacción estudiantil. Algunos ejemplos son: baja calidad académica o pertinencia del programa de estudios, dificultad o exigencia de las asignaturas, falta de orientación o asesoría académica, métodos de enseñanza tradicionales o poco innovadores, sistemas de evaluación poco flexibles o transparentes, falta de comunicación o retroalimentación entre docentes y estudiantes, etc.

- **Causas institucionales**

Se refieren a la organización, la gestión, la infraestructura y los servicios que ofrece la institución universitaria, que inciden en el bienestar y el desarrollo integral de los estudiantes. Algunos ejemplos son: falta de recursos financieros o materiales para el funcionamiento adecuado de la universidad, falta de políticas o programas para prevenir o atender la deserción estudiantil, falta de becas o ayudas económicas para los estudiantes con necesidades especiales, falta de espacios físicos o virtuales para el estudio o la recreación, falta de servicios

complementarios como biblioteca, transporte, salud, deporte, cultura, etc.

iii. Deserción Universitaria

Es el abandono de la carrera universitaria elegida por el estudiante. Como resultado de un conflicto interno o externo que termina en abandonar sus estudios universitarios. [48]

Existen diferentes factores que interactúan con los estudiantes para que tomen la decisión de desertar, que son cambiantes según la región y momento económico. [49]

Los enfoques que se destacan en el estudio de la deserción estudiantil universitaria, son:

- **Enfoques Psicológicos**

Fijan su atención en condiciones personales internas que actúan en el comportamiento y lleva a que los individuos eviten o propicien situaciones o eventos.

- **Enfoques Sociológicos o Ambientales**

Marcan el papel que tienen las fuerzas externas (estatus social, raza, prestigio institucional) para la persistencia estudiantil

- **Enfoques Económicos**

La incapacidad o capacidad que tiene el estudiante para solventar los gastos de sus estudios universitarios.

- **Enfoques Organizacionales**

Consideran la deserción como el reflejo del impacto que tiene la organización sobre la socialización y satisfacción del estudiante.

- **Enfoques Interaccional**

La deserción depende de la manera en que cada estudiante interpreta las experiencias vividas en la universidad, así como el grado que se da entra la institución y los individuos.

2.3. Definición de términos básicos

2.3.1. Machine Learning

Es una rama de la inteligencia artificial que se encarga de crear sistemas capaces de aprender de forma automática a partir de los datos, sin necesidad de una programación explícita. El aprendizaje automático se puede clasificar en dos tipos: supervisado y no supervisado. El aprendizaje supervisado consiste en entrenar al sistema con datos etiquetados, es decir, con la respuesta correcta para cada caso. El aprendizaje no supervisado consiste en dejar que el sistema encuentre patrones o estructuras ocultas en los datos sin etiquetar [22].

2.3.2. Algoritmo de machine learning

Es una secuencia de instrucciones o reglas que le indican al sistema cómo procesar los datos y extraer patrones o conocimiento de ellos [50].

2.3.3. Modelo Predictivo

Es un modelo matemático que permite estimar la probabilidad de que ocurra un evento futuro, basándose en datos históricos y actuales, y aplicando técnicas de análisis estadístico o de aprendizaje automático [51]. En este estudio se utilizó el modelo de Regresión Logística.

2.3.4. Regresión Logística

La regresión logística calcula la probabilidad de que un evento ocurra, como participar o no en una votación, considerando un conjunto específico de variables independientes. En el ámbito del aprendizaje automático, la regresión

logística forma parte de los modelos de aprendizaje supervisado [36].

2.3.5. Diagnóstico temprano

Es el proceso de identificar y reconocer los signos o síntomas de una condición o problema antes de que se agrave o se vuelva crónico. El diagnóstico temprano permite intervenir de forma oportuna y efectiva para prevenir o mitigar las consecuencias negativas del problema [52].

2.3.6. Deserción Universitaria

Es el fenómeno social y educativo que se refiere al abandono definitivo o temporal de los estudios por parte de los estudiantes universitarios, antes de obtener el título académico correspondiente, debido a diversos factores de índole personal, familiar, económica, académica o institucional. La deserción universitaria tiene implicaciones negativas tanto para los individuos como para la sociedad, ya que limita las oportunidades de desarrollo personal y profesional, y afecta el capital humano y el crecimiento económico de un país [47].

2.3.7. Entrenamiento

Es el proceso por el que se ajustan los parámetros del algoritmo de machine learning para optimizar su rendimiento y precisión [22].

2.3.8. Datos de entrenamiento

Son los datos que se usan para entrenar al algoritmo de machine learning. Deben ser representativos, variados y de calidad [22].

2.3.9. Datos de prueba

Son los datos que se usan para evaluar el rendimiento y la generalización del algoritmo de machine learning una vez entrenado [22].

2.3.10. Predicción

Es la salida que genera el modelo predictivo basado en machine learning al recibir una entrada nueva o desconocida [22].

3. CAPÍTULO III – METODOLOGÍA

3.1. Enfoque de la investigación

La presente investigación se basa en un enfoque cuantitativo.
El enfoque cuantitativo se basa en la recopilación y análisis de datos empíricos utilizando métodos estandarizados y estructurados. Según el Creswell destaca que la investigación cuantitativa se centra en la medición de variables y la comprobación de hipótesis mediante la aplicación de técnicas estadísticas [53].

3.2. Tipo de Investigación

La naturaleza de la investigación es de tipo aplicada.
La investigación aplicada se refiere a la realización de estudios que buscan resolver problemas específicos en el mundo real o aplicar conocimientos teóricos a situaciones prácticas. La investigación aplicada tiene como objetivo la solución de problemas concretos mediante la aplicación de métodos y técnicas de investigación científica [54].

3.3. Nivel de investigación

El nivel de investigación planteado será el nivel explicativo. Según [55], las investigaciones explicativas consideran la manipulación de la variable independiente (VI) que es la causa sobre las variables dependientes (VD) que es el efecto, en una realidad especifica determinada.

3.4. Método de Investigación

En el marco de la presente investigación, centrada en el diseño de un Modelo Predictivo basado en Machine Learning para el diagnóstico temprano en la deserción universitaria, se ha adoptado el método hipotético deductivo. Es un proceso científico que comienza con una hipótesis, una explicación tentativa de un fenómeno, la cual debe ser comprobable y falsable. [56].

Grove y Menton [57], lo describen como un enfoque formalizado para probar teorías científicas dentro de la filosofía de la ciencia. Este método inicia con la formulación de hipótesis derivadas deductivamente a partir de principios o teorías generales, las cuales luego se someten a contrastación empírica mediante observaciones o experimentos. El objetivo es evaluar la capacidad de las hipótesis para predecir y explicar fenómenos dentro de un marco controlado.

3.5. Diseño de Investigación

El diseño de investigación utilizado fue Pre experimental. El diseño pre experimental generalmente utilizado en estudios exploratorios o evaluaciones iniciales donde los recursos para un control riguroso son limitados. Es útil en contextos donde se necesita una evaluación preliminar de un programa o intervención, aunque no ofrecen la misma robustez metodológica que los diseños experimentales o cuasi experimentales [58].

A diferencia de los diseños experimentales y cuasi experimentales, los diseños pre experimentales carecen de aleatorización y de un grupo de control sólido, lo que reduce su validez interna y dificulta descartar la influencia de variables externas. Son adecuados principalmente para evaluaciones exploratorias o preliminares [54].

3.6. Población, muestra y muestreo

3.6.1. Población

Registros socio académico vocacionales de todos los estudiantes de la UNDC. Representados por 1 764.

3.6.2. Muestra

Para la muestra se usa la siguiente formula:

$$n = \frac{Z^2 \times p \times q \times N}{e^2 \times (N-1) + Z^2 \times p \times q}$$

Donde:

N es el tamaño de la población finita (1764)
Z es el valor correspondiente al nivel de confianza (para un 95%, Z = 1,96)
p es la proporción esperada (0,5)
q es 1 – p (también 0,5 en este caso)
e es el margen de error (5%)

El tamaño de la muestra para la población de 1764 registros socio académicos, con un nivel de confianza del 95% y un margen de error del 5%, es **316 estudiantes.**

3.6.3. Muestreo

En esta investigación se utilizará un muestreo no probabilístico por conveniencia, ya que se incluyeron todos los integrantes de la muestra.

3.7. Técnicas e Instrumentos de recolección de Datos

3.7.1. Técnicas

A. Técnica de encuesta

En la presente investigación se utilizará la técnica de la encuesta donde se determinará las características de los estudiantes que desertaron en la universidad. Para [59], la técnica de encuesta es un método de investigación que utiliza un conjunto de

procedimientos estandarizados para recoger y analizar una serie de datos de una muestra de casos representativa de una población o universo más amplio. Esta técnica permite explorar, describir, predecir y/o explicar una serie de características.

B. Validación Cruzada (Cross-validation)

Es una herramienta computacional que facilita la evaluación cuantitativa de modelos de Machine Learning en función de diversos criterios y métricas, permitiendo a los investigadores y científicos de datos comparar y optimizar sus modelos de forma sistemática.

3.7.2. Instrumentos

A. Instrumentos de encuesta: cuestionario

Para la presente investigación se usará el cuestionario, el cuál contendrá preguntas relacionadas a la variable dependiente y dimensiones. Un "Cuestionario" es una herramienta de investigación que consta de una serie de preguntas diseñadas para recopilar información de los encuestados. Los cuestionarios pueden ser utilizados para recopilar datos cuantitativos y cualitativos, y son comúnmente utilizados en la investigación social y de mercado [60].

3.7.3. Validez

La validez de los instrumentos de recolección de datos se refiere a la capacidad de un instrumento de medición o recolección de datos para medir o recopilar lo que se desea medir. En otras palabras, la validez es la precisión con la que un instrumento mide lo que se supone que debe medir. Si un instrumento no es válido, los datos que se recolectan no serán precisos o representativos [61].

Es por ello que la validez de los instrumentos de recolección de datos se hizo mediante el juicio de expertos como se muestra en el ANEXO 3, se solicitó la opinión de individuos que poseen conocimientos y experiencia en el tema en cuestión, docentes con el grado de Doctor y/o Magister de la Universidad Nacional de Cañete.

3.7.4. Confiabilidad

La confiabilidad de un instrumento se refiere al grado en el que un instrumento de medición proporciona resultados consistentes [62].

En este estudio, se evaluó la confiabilidad mediante el cálculo del coeficiente Alfa de Cronbach, ampliamente reconocido en la investigación científica para medir la consistencia interna de los ítems que conforman un instrumento de medición.

La confiabilidad del instrumento fue evaluada mediante el cálculo del coeficiente **Alfa de Cronbach**, obteniendo un valor de **0.704**, lo cual indica un nivel aceptable de consistencia interna entre los ítems que componen el cuestionario. Este resultado respalda la fiabilidad del instrumento para medir de manera consistente las dimensiones evaluadas: económica, académica y vocación profesional. El valor obtenido está en el rango recomendado para instrumentos en investigaciones sociales y educativas, validando su capacidad para generar resultados consistentes y replicables en contextos similares.

TABLA II. ALFA DE CRONBACH

Estadísticas de fiabilidad

Alfa de Cronbach	N de elementos
.704	20

3.8. Técnicas estadísticas de análisis de datos

Pruebas de hipótesis: Para contrastar la hipótesis nula de que no existe una relación significativa entre las variables académicas y socioeconómicas de los estudiantes y la probabilidad de deserción, y la hipótesis alternativa de que sí existe una relación significativa. Se utilizará la Prueba T, para comparar la media de dos grupos de datos y determinar si son significativamente diferentes entre sí.

3.9. Consideraciones éticas

- **Consentimiento informado:** Se solicitará el consentimiento informado de los estudiantes que participen en la investigación, explicándoles el propósito, los objetivos, los procedimientos, los beneficios, los riesgos y las garantías de confidencialidad y anonimato de la investigación. Se les informará que su participación es voluntaria y que pueden retirarse en cualquier momento sin consecuencias negativas.
- **Protección de datos personales:** Se protegerán los datos personales de los estudiantes, utilizando códigos numéricos o alfanuméricos para identificarlos, y almacenando los datos en medios seguros y con acceso restringido. Se evitará el uso de datos sensibles o innecesarios para la investigación, y se respetará el derecho de los estudiantes a acceder, rectificar, cancelar u oponerse al tratamiento de sus datos.
- **Respeto a la dignidad y a la diversidad:** Se respetará la dignidad y la diversidad de los estudiantes, evitando cualquier forma de discriminación, estigmatización, coacción o daño físico o psicológico. Se reconocerá la autonomía y la libertad de los estudiantes para tomar sus propias decisiones sobre su trayectoria académica, y se les brindará apoyo y orientación si lo requieren.
- **Honestidad y rigor científico:** Se actuará con honestidad y rigor científico en todas las fases de la investigación, evitando el plagio, la falsificación, la fabricación o la manipulación de los datos. Se citarán adecuadamente las fuentes utilizadas, siguiendo las normas IEEE. Se reportarán los resultados con transparencia y objetividad, reconociendo las limitaciones y las implicaciones éticas del estudio.

4. CAPÍTULO IV – RESULTADOS

Se diseñó una encuesta como instrumento de recolección de datos y fue validada por juicio de expertos para asegurar su validez y precisión. Posteriormente, la encuesta fue implementada en Google Forms y, con la colaboración de la unidad de TI de la Universidad Nacional de Cañete, se publicó en el sistema académico SIVIRENO. A los estudiantes que accedían al SIVIRENO para consultar sus notas se les solicitaba como paso previo, completar la encuesta.

Una vez finalizado el periodo de recolección de datos, las respuestas se exportaron a una hoja de cálculo para realizar una limpieza inicial de datos, eliminando entradas incompletas o inconsistentes. Finalmente, los datos limpios fueron preparados para su procesamiento y análisis mediante el algoritmo de regresión logística.

4.1. Resultados descriptivos

Como se observa en el ANEXO 5, la muestra del estudio incluye 316 participantes, con una distribución de género de 43,5% mujeres y 56,5% hombres. Participan todos los ciclos y carreras universitarias.

i. **Pregunta N° 1:** ¿En qué medida dependes económicamente de alguien más?, los estudiantes han respondido: “No dependo de otros” un 16,3%, “Dependo parcialmente de otras personas” un 35,8%, “Dependo bastante” un 21,5%, “Dependo totalmente de otras personas” un 26,5%.

ii. **Pregunta N° 2:** Si tienes un empleo, ¿cómo influye en tu rendimiento académico?, los estudiantes han respondido: “No influye en mi rendimiento académico” un 5%, “Influye poco en mi rendimiento académico” un 30,8%, “Influye bastante en mi rendimiento académico” un 20,5%, “Influye mucho en mi rendimiento académico” un 7%, “No tengo empleo” un 36,8%.

iii. **Pregunta N° 3:** Si tienes responsabilidades familiares, ¿cómo afecta esto a tus estudios?, los estudiantes han respondido: “No afecta en mis estudios” un 22,3%, “Afecta poco en mis estudios” un 46,3%, “Afecta poco en mis estudios” un 25,5%, “Afecta mucho en mis estudios” un 6%.

iv. **Pregunta N° 4:** ¿Cuánto impacta tu situación económica en la posibilidad de continuar con tus estudios?, los estudiantes han respondido: “No impacta en nada mi situación económica” un 12,8%, “Impacta poco” un 39,3%, “Impacta bastante” un 33%, “Impacta mucho mi situación económica para seguir estudiando” un 15%.

v. **Pregunta N° 5:** ¿En qué medida influye tu salud en tu capacidad para continuar con tus estudios?, los estudiantes han respondido: “No influye en nada” un 25%, “Influye poco” un 34,3%, “Influye bastante” un 27%, “Influye mucho mi estado de salud para seguir estudiando” un 13,7%.

vi. **Pregunta N° 6:** Si tienes una pareja, ¿cómo afecta esto a tu desempeño académico?, los estudiantes han respondido: “No afecta en mi desempeño académico” un 50,5%, “Afecta poco” un 17,3% “Afecta bastante” un 5,2%”, “Afecta mucho en mi desempeño académico” un 3,5%, “No tengo pareja” un 50,5%.

vii. **Pregunta N° 7:** ¿Las otras actividades que desempeñas afectan tu tiempo de estudio?, los estudiantes han respondido: “No afectan en nada a mis estudios” un 19,5%; “Afectan poco” un 55,7%; “Afectan poco” un 20,8%; “Afectan mucho” un 4%.

viii. **Pregunta N° 8:** Si experimentas violencia familiar, ¿cómo afecta esto a tu rendimiento académico?, los estudiantes han respondido: “No afecta en mi rendimiento académico” un 42,8%; “Afecta poco” un 21,3%; “Afecta bastante” un 24,7%; “Afectan mucho” un 11,3%.

ix. **Pregunta N° 9:** ¿En qué medida la interacción con tus profesores afecta tu desempeño académico?, los estudiantes han respondido: “No afecta en mi desempeño académico” un 27,8%; “Afecta poco” un 31,5%; “Afecta bastante” un 30,3%; “Afecta mucho” un 10.5%.

x. **Pregunta N°10:** ¿Cómo afecta el acoso de tus compañeros a tu rendimiento académico? los estudiantes han respondido: “No afecta en mi rendimiento académico” un 49,5%; “Afecta poco” un 25%, “Afecta bastante” un 17,8%; “Afecta mucho” un **7,8%.**

xi. **Pregunta N° 11:** ¿La distancia entre tu hogar y la universidad afecta tu asistencia a clases?, los estudiantes han respondido: “No afecta a mi asistencia a clases” un 21%, “Afecta poco” un 32,5%, “Afecta bastante” un 26,5%, “Afecta mucho” un 20%.

xii. **Pregunta N° 12:** ¿Te sientes cómodo con tus compañeros de clase?, los estudiantes han respondido: “No me siento cómodo” un 6%, “Me siento un poco cómodo” un 24,8%, “Me siento un poco cómodo” un 42%, “Me siento muy cómodo con mis compañeros de clase” un 27,2%.

xiii. **Pregunta N° 13:** ¿Has experimentado discriminación en la universidad?, los estudiantes han respondido: “No he experimentado discriminación en la universidad” un 76,3%; “He experimentado poca discriminación en la universidad” un 20,8%; “He experimentado bastante discriminación en la universidad” un 2%; “He experimentado mucha discriminación en la universidad” un 1%.

xiv. **Pregunta N° 14:** ¿Crees que dedicas suficiente tiempo al estudio de tus materias?, los estudiantes han respondido: “No he dedicado tiempo a mis estudios” un 3%, “He dedicado poco tiempo a mis estudios” un 48,2%; “He dedicado bastante tiempo a mis estudios” un 41,8%; “He dedicado mucho tiempo a mis estudios” un 7%.

xv. **Pregunta N° 15:** ¿Consideras que tu carrera te facilitará encontrar empleo?, los estudiantes han respondido: “No me facilitará encontrar empleo” un 5,8%; “Me facilitará un poco encontrar empleo” un 34%; “Me facilitará bastante encontrar empleo” un 44,5%; “Me facilitará mucho encontrar empleo” un 15,7%.

xvi. **Pregunta N° 16:** ¿Tus expectativas académicas se cumplieron después de salir del colegio?, los estudiantes han respondido: “No se cumplieron mis expectativas académicas” un 20,5%; “Se cumplieron un poco” un

46%; “Se cumplieron bastante” un 26,8%; “Se cumplieron totalmente mis expectativas académicas” un 6,8%.

xvii. **Pregunta N° 17:** ¿Los conocimientos adquiridos en la universidad te son útiles en otros aspectos de tu vida?, los estudiantes han respondido, los estudiantes han respondido: “No me son útiles en otros aspectos de mi vida” un 2,8%; “Son poco útiles en otros aspectos de mi vida” un 27%; “Son bastante útiles en otros aspectos de mi vida” un 51,5%; “Son muy útiles en otros aspectos de mi vida” un 18,8%.

xviii. **Pregunta N° 18:** ¿La influencia de tu familia afectó tu elección de carrera?, los estudiantes han respondido: “No influencio mi familia en la elección de mi carrera” un 30,5%; “Influencio un poco” un 36,3%; “Influencio bastante” un 21%; “Influencio mucho mi familia en la elección de mi carrera”.

xix. **Pregunta N° 19:** ¿Tenías en mente estudiar una carrera diferente a la que finalmente elegiste?, los estudiantes han respondido: “No tenía pensado en otra carrera” un 18.5%; “Tenía pensado pocas opciones” un 44,8%; “Tenía pensado bastantes opciones” un 16,8%; “Tenía pensado muchas otras opciones” un 20%.

xx. **La última pregunta:** “¿Has pensado en desertar de tu carrera universitaria?”, los estudiantes han respondido “No lo tengo pensado” un 92%; mientras que un 8% ha respondido “Estoy decidido en desertar mi actual carrera universitaria".

Los datos muestran que en general, los estudiantes no desean desertar de su actual carrera.

4.2. Resultados inferenciales

4.2.1. Prueba de normalidad

Para determinar si los datos siguen una distribución normal, es fundamental realizar una prueba de normalidad.

Si n >= 50, se emplea el método de Kolmogorov – Smirnov

Si n <= 50, se emplea el método de Shapiro – Wilk

La muestra tiene más de cincuenta datos, el método de Kolmogorov – Smirnov es la herramienta ideal para analizar su distribución.

Se plantea los siguientes criterios de decisión:

Hipótesis nula (H_0): Los datos siguen una distribución normal.

Hipótesis alternativa (H_1): Los datos no siguen una distribución normal.

Si el p – valor es menor que 0,05 se rechaza H_0 (los datos no son normales)

Si el p – valor es mayor o igual a 0,05 se rechaza H_0 (los datos son normales)

TABLA III. Prueba de normalidad

	Kolmogorov-Smirnov[a]		
	Estadístico	gl	Sig.
P1	.221	464	<.001
P2	.243	464	<.001
P3	.244	464	<.001
P4	.226	464	<.001
P5	.213	464	<.001
P6	.328	464	<.001
P7	.304	464	<.001
P8	.250	464	<.001
P9	.195	464	<.001
P10	.294	464	<.001
P11	.193	464	<.001
P12	.233	464	<.001
P13	.446	464	<.001
P14	.310	464	<.001
P15	.228	464	<.001
P16	.257	464	<.001
P17	.265	464	<.001
P18	.223	464	<.001
P19	.265	464	<.001
R1	.488	464	<.001

a. Corrección de significación de Lilliefors

Un valor $p < 0{,}001$ significa que se rechaza la hipótesis nula (H_0), quiere decir que los datos no siguen una distribución normal.
Las respuestas de las variables analizadas no se distribuyen normalmente, por lo tanto, se debe utilizar pruebas estadísticas no paramétricas para analizar estas variables.

4.2.2. Prueba de hipótesis general

Teniendo en cuenta que el planteamiento de la hipótesis fue el siguiente:

Hipótesis Nula (H_0): La implementación del modelo predictivo basado en machine learning no tiene un efecto significativo en el diagnóstico temprano de la deserción universitaria en la UNDC.

Hipótesis Alternativa (H_1): La implementación del modelo predictivo basado en machine learning tiene un efecto significativo en el diagnóstico temprano de la deserción universitaria en la UNDC.

Se confirma la hipótesis alternativa (H_1), concluyendo que el modelo predictivo basado en machine learning tiene un impacto significativo en el diagnóstico temprano de la deserción universitaria, evidenciando su efectividad como herramienta para identificar factores críticos asociados al abandono estudiantil. Este impacto se evidencia al analizar los resultados obtenidos en cada una de las dimensiones evaluadas:

Resultados por Dimensión:

Los resultados obtenidos en la **dimensión económica**, representada por las preguntas P1, P2, P3, P4, P5 y P6 muestran que todas las pruebas de hipótesis específicas para estas preguntas resultaron significativas ($p < 0.$), indicando una relación estadísticamente significativa entre los factores económicos y la deserción universitaria (R1). Esto permite generalizar que la dimensión económica, en su conjunto, tiene un impacto relevante en la probabilidad de deserción. Los hallazgos sugieren que los problemas financieros enfrentados por los estudiantes contribuyen de manera importante a su decisión de abandonar los estudios, destacando la necesidad de implementar estrategias institucionales orientadas a mitigar estos factores.

Los resultados obtenidos en la **dimensión académica**, representada por las preguntas P7, P8, P9, P10, P11, P12, P13 y P14, indican que siete de las ocho preguntas resultaron estadísticamente significativas en las pruebas de hipótesis ($p < 0.05$). Esto sugiere que los factores relacionados con el desempeño académico y la experiencia educativa tienen una influencia importante en la deserción universitaria (R1). Aunque la pregunta P13 no presentó significancia estadística ($p > 0.05$), su contribución al análisis general no invalida el impacto global de la dimensión académica, dado que el 87.5% de los factores analizados dentro de esta dimensión muestran una relación significativa con la deserción. Estos resultados destacan la necesidad de fortalecer las estrategias académicas, tales como el apoyo a estudiantes con dificultades de rendimiento, la mejora en la calidad de la enseñanza y el acompañamiento académico personalizado, como medidas clave para reducir la deserción.

Los resultados obtenidos en la **dimensión vocación profesional**, representada por las preguntas P15, P16, P17, P18 y P19,

muestran que todas las pruebas de hipótesis específicas para estas preguntas resultaron estadísticamente significativas ($p < 0.05$). Esto indica que los factores relacionados con la claridad vocacional, la satisfacción con la carrera elegida y el interés por los estudios tienen una relación directa y significativa con la deserción universitaria (R1). En particular, estos hallazgos sugieren que la falta de afinidad con la carrera elegida y la incertidumbre sobre las metas profesionales son factores clave que incrementan la probabilidad de deserción. Por lo tanto, se concluye que la dimensión vocación profesional, en su conjunto, tiene un impacto importante en el fenómeno de la deserción, subrayando la importancia de implementar programas de orientación vocacional y actividades de acompañamiento desde el ingreso a la universidad para reducir el riesgo de abandono.

5. CAPÍTULO V – DISCUSIÓN

Los resultados obtenidos en esta investigación confirman la influencia significativa del modelo predictivo basado en machine learning en el diagnóstico temprano de la deserción universitaria, lo que coincide con estudios previos como el realizado por Candia Oviedo [63], en el que se utilizó aprendizaje automático para predecir el rendimiento académico en estudiantes universitarios. En ambos casos, los algoritmos de clasificación demostraron ser efectivos para identificar patrones clave en los datos educativos. Sin embargo, a diferencia de Candia Oviedo, nuestro modelo se enfoca específicamente en el diagnóstico de deserción, obteniendo resultados significativos para la mayoría de las dimensiones evaluadas, lo que evidencia la adaptabilidad de los enfoques de machine learning en contextos educativos diversos y su potencial para mejorar la toma de decisiones en políticas de retención estudiantil.

En concordancia con la investigación realizada por De la Cruz [19], quien utilizó un enfoque similar al nuestro en la Universidad Ricardo Palma. En ambos casos, los modelos predictivos facilitaron la identificación de factores críticos asociados a la deserción estudiantil, como la deuda semestral, el promedio de asistencias y los ingresos familiares. Sin embargo, mientras en el presente estudio se emplearon algoritmos como la regresión logística para un análisis puntual de los factores económicos, académicos y vocacionales, De la Cruz implementó una arquitectura más amplia que integró múltiples algoritmos de machine learning en un servidor Big Data, maximizando la automatización del análisis de datos. Aunque las estrategias metodológicas presentan diferencias, ambos trabajos coinciden en la necesidad de generar reportes detallados para que las instituciones educativas implementen estrategias efectivas que aborden las causas específicas de la deserción. Estos hallazgos subrayan el valor de los modelos predictivos como herramientas clave para mejorar la toma de decisiones en el ámbito educativo.

En línea con la investigación realizada por Pérez y Rojas en la Universidad Tecnológica del Perú [6], podemos mencionar que en la presente investigación se utilizó regresión logística para analizar los factores económicos, académicos y vocacionales, mientras que Pérez y Rojas

implementaron el algoritmo Support Vector Machine (SVM), destacando su capacidad para identificar patrones de comportamiento relacionados con la deserción. Ambos estudios emplearon datos históricos de estudiantes para detectar posibles deserciones estudiantiles con cierta precisión. Sin embargo, el uso de diferentes algoritmos resalta enfoques complementarios, donde el SVM, con una certeza superior al 90 %, demuestra robustez en escenarios no lineales gracias al uso de la función Kernel. A pesar de estas diferencias metodológicas, ambos trabajos subrayan la importancia de estas herramientas en la planificación institucional, permitiendo intervenciones tempranas y personalizadas que buscan reducir la tasa de deserción y mejorar la experiencia educativa.

Los resultados de este estudio, que validan la implementación de un modelo predictivo basado en machine learning para el diagnóstico temprano de la deserción universitaria, encuentran paralelismos con el trabajo de Forero Corba y Negre Bennasar [64], enfocado en la predicción de la competencia digital docente, el modelo propuesto por Forero y Negre, realizado en software Orange Data Mining, utiliza rúbricas de evaluación para medir la competencia digital docente en niveles explorador, integrador e innovador. Mientras en este trabajo se consideraron dimensiones como la económica, académica y vocacional, en el contexto docente se utilizaron rúbricas estandarizadas y variables sociodemográficas para evaluar competencias digitales. Estas similitudes sugieren que el enfoque puede adaptarse a otros sectores donde la identificación temprana de deficiencias o patrones puede mejorar la toma de decisiones, como en programas de capacitación profesional, estrategias de inclusión digital o personalización del aprendizaje.

Se encuentran analogías significativas con el trabajo de Yandar et al. [65], quienes realizaron una revisión sistemática de literatura enfocada en la predicción de energía eléctrica. Ambas investigaciones subrayan el impacto de técnicas de machine learning como las redes neuronales artificiales (RNA) para modelar y prever comportamientos complejos. Mientras que nuestro trabajo abordó dimensiones académicas, económicas y vocacionales para identificar factores críticos en la deserción, Yandar et al. destacaron el uso de RNA para la predicción precisa del consumo energético, mostrando cómo estas herramientas pueden integrarse en sectores diversos, desde la educación hasta la gestión de recursos energéticos.

También se encuentran aplicaciones complementarias en el ámbito de la salud mental, tal como se evidencia en el artículo científico de Osorio et. al [66], se destaca el uso de modelos estadísticos como la regresión logística, el SVM y los bosques aleatorios. Mientras que nuestro modelo se enfoca en factores económicos, académicos y vocacionales, Osorio et al. identifican variables predictoras clave como neuroimágenes, características sociodemográficas y síntomas clínicos. En ambos estudios subrayamos la importancia de un enfoque preventivo, en mi estudio para reducir la deserción, y en el estudio revisado, para priorizar y personalizar la gestión de la salud mental.

En este estudio, las hipótesis específicas fueron evaluadas utilizando chi cuadrado, y los resultados muestran que la mayoría fueron aceptadas, mientras que la **Pregunta 13**: "¿Has experimentado discriminación en la universidad?" dentro de la dimensión **situación académica** no tuvo significancia estadística. La **hipótesis general**, que plantea que el modelo predictivo basado en machine learning influye significativamente en el diagnóstico temprano de la deserción universitaria, fue aceptada, dado que los valores $p < 0,05$ en la mayoría de las hipótesis específicas respaldan la relación significativa entre las variables analizadas y la deserción universitaria.

A partir de los resultados obtenidos, se plantean las siguientes hipótesis que pueden guiar investigaciones futuras y refinar el modelo predictivo:

- La implementación de estrategias de apoyo financiero en la universidad influye en la disminución del riesgo de deserción asociado a factores económicos.
- El fortalecimiento de los programas de tutoría académica y la mejora en la percepción de calidad docente tienen un impacto significativo en la retención estudiantil.
- La introducción de programas de orientación vocacional en los primeros ciclos académicos mejora la afinidad de los estudiantes con su carrera y reduce la probabilidad de deserción.
- La incorporación de sistemas inteligentes basados en machine learning para el monitoreo continuo de datos académicos y socioeconómicos mejora la detección temprana de estudiantes en riesgo de deserción.
- La integración de dimensiones adicionales, como la salud mental y el entorno familiar, mejora la capacidad predictiva del modelo para diagnosticar la deserción estudiantil.

6. CAPÍTULO VI – CONCLUSIONES

Se confirma que la implementación de un modelo predictivo basado en machine learning influye significativamente en el diagnóstico temprano de la deserción universitaria en la UNDC durante el periodo 2020 – 2022.

1. Los factores económicos, como la capacidad financiera y la estabilidad económica del estudiante, demostraron tener una influencia significativa en la probabilidad de deserción, con todas las preguntas de esta dimensión resultando significativas ($p < 0,05$). Este hallazgo resalta la necesidad de implementar estrategias de apoyo financiero para mitigar estos factores de riesgo.

2. La dimensión académica, si bien tuvo en su mayoría resultados significativos, mostró que una pregunta (P13) no tuvo una relación estadísticamente relevante ($p > 0.05$). Esto sugiere que la carga académica, el desempeño y la percepción de calidad educativa son predictores importantes, pero requieren una revisión exhaustiva para refinar las variables no significativas.

3. La afinidad vocacional y la claridad en los objetivos profesionales demostraron ser factores clave para el diagnóstico temprano de la deserción, con todas las preguntas de esta dimensión resultando significativas ($p < 0.05$). Esto subraya la importancia de incorporar programas de orientación

vocacional en las estrategias institucionales.

4. El modelo predictivo basado en machine learning mostró un desempeño confiable en la clasificación de estudiantes en riesgo, con una precisión del 80,64% de los casos usando el modelo predictivo de regresión logística, respaldado por indicadores como un Alfa de Cronbach de 0.704 y valores $p < 0.05$ en la mayoría de los análisis. Esto valida el uso de técnicas de machine learning como herramientas efectivas para el diseño de estrategias preventivas.

7. CAPÍTULO VII – RECOMENDACIONES

Basado en la influencia significativa de los factores económicos en la deserción universitaria, se recomienda establecer programas de becas, subsidios y opciones flexibles de pago para estudiantes con limitaciones financieras. Esto podría complementarse con talleres de gestión financiera personal para ayudar a los estudiantes a administrar mejor sus recursos.

Para abordar los problemas relacionados con la dimensión académica, se sugiere diseñar tutorías personalizadas enfocadas en mejorar el rendimiento académico y reducir la carga percibida por los estudiantes. Asimismo, la capacitación constante del cuerpo docente sobre metodologías de enseñanza innovadoras podría mejorar la percepción de calidad educativa.

Debido al impacto de los factores vocacionales en la deserción, es crucial implementar un programa robusto de orientación vocacional desde los primeros ciclos universitarios. Este sistema debería incluir evaluaciones periódicas de afinidad profesional, talleres de exploración vocacional y un acompañamiento continuo para alinear los objetivos personales con los académicos.

Se recomienda construir una plataforma digital que utilice modelos predictivos para monitorear en tiempo real el desempeño académico, la situación económica y los indicadores vocacionales de los estudiantes. Esto permitiría una intervención temprana y personalizada en función de las necesidades detectadas.

Para mejorar la precisión del diagnóstico, se sugiere incorporar variables adicionales como factores de salud mental, entorno familiar y contexto social en futuras versiones del modelo. Esto ofrecerá una visión más integral de las causas de la deserción y permitirá diseñar estrategias más efectivas.

REFERENCIAS

[1] Ministerio de Educación, «Tasa de deserción en educación universitaria se redujo a 11.5%,» 8 Noviembre 2021. [En línea]. Available: https://www.gob.pe/institucion/minedu/noticias/552273-tasa-de-desercion-en-educacion-universitaria-se-redujo-a-11-5.

[2] R. C. Dávila Morán, E. d. C. Agüero Corzo, H. Portillos Rios y O. R. Quimbita Chiluisa, «Deserción universitaria de los estudiantes de una universidad peruana,» *Revista Universidad y Sociedad,* vol. 14, nº 2, 25 08 2022.

[3] J. E. Valero Cajahuanca, A. F. Navarro Raymundo, A. C. Larios Franco y J. D. Julca Flores, «Deserción universitaria: Evaluación de diferentes algoritmos de Machine Learning para su predicción,» *Revista de Ciencias Sociales (Ve),* vol. 28, nº 3, pp. 362-375, 2022.

[4] R. I. Castro López, «Aplicación de técnicas de machine learning para el estudio de deserción temprana y egreso oportuno en estudiantes de ingeniería de la facultad de ciencias físicas y matemáticas.,» Universidad De Chile, Chile, 2020.

[5] I. Díaz, A. Bernardo, M. Esteban y L. Rodrígues Muñiz, «Variables Influencing University Dropout: A Machine Learning-Based Study,» ICEUTE 2020, España, 2020.

[6] C. A. Perez Bedia y L. E. Rojas Segovia, «Diseño de un sistema para predecir la deserción de los alumnos mediante Machine learning en la Universidad Tecnológica del Perú.,» Universidad Tecnológica del Perú, Lima, 2020.

[7] A. M. Tataje Okamoto, K. M. Chacaltana Hernández y S. G. Varela Guevara, «Adaptación universitaria en estudiantes de primer año de una universidad pública de Ica – Perú,» Universidad Autónoma del Perú, Lima, 2024.

[8] T. M. Oca Sánchez, «Educación superior: factores económicos que inciden en la deserción escolar. Caso de las licenciaturas de la UNID Tlalnepantla,» *Revista Iberoamericana para la Investigación y el Desarrollo Educativo,* vol. 12, nº 23, 2021.

[9] N. J. Cárdenas Ballesteros, A. M. Díaz Rodriguez y A. E. Chaparro Ortiz, «Revisión sistemática de los factores que inciden en la deserción académica en la educación superior,» Universidad Cooperativa de Colombia, Colombia, 2021.

[10] X. F. Erazo Guerra y E. d. R. Rosero Morales, «Orientación vocacional y su influencia en la deserción universitaria,» *Horizontes,* vol. 5, nº 18, p. 16, 2021.

[11] A. Villa Murillo, L. Costa y C. Vásquez, «A methodological proposal to address the academic dropout phenomenon based on an intelligent prediction model: a case study,» *Millenium - Journal of Education, Technologies and Health,* vol. 2, nº 23, 2023.

[12] T. Shuo-Chang, C. Cheng-Huan, S. Yi-Tzone, C. Jin-Shuei y W. Trong-Neng, «Precision education with statistical learning and deep learning: a case study in Taiwan,» *International Journal of Educational Technology in Higher Education,* vol. 17, nº 12, 2020.

[13] A. Rodriguez, J. Espinoza, L. Ramirez y A. Ganga, «Deserción Universitaria: Nuevo Análisis Metodológico,» *Formación Universitaria,* vol. 11, nº 6, pp. 107-118, 2018.

[14] M. G. Cuzme Romero, «Modelo de aprendizaje automático para la predicción del riesgo de deserción en estudiantes de la Universidad Técnica de Manabí,» Tesis de maestría, Universidad de las Fuerzas Armadas, Ecuador, 2020.

[15] Y. A. Quintero, «Diseño de un modelo predictivo para generar alertas tempranas de deserción universitaria en los programas de pregrado presenciales de la Facultad de Ingeniería de la Universidad de Antioquia,» Tesis de maestría, Universidad de Antioquia, Colombia, 2022.

[16] C. Russo, «Minería de datos aplicada a estrategias para minimizar el rezago académico y la deserción universitaria en carreras de informática de la UNNOBA,» Tesis doctoral, Universidad Nacional de la Plata, Argentina, 2019.

[17] D. A. Gómez Pinzón, «Aplicativo web para el análisis, selección y admisión de aspirantes al programa de ingeniería de sistemas de la Universidad de Cundinamarca en la extensión facatativá utilizando modelos predictivos de minería de datos.,» Tesis de pregrado, Universidad De Cundinamarca, Colombia, 2016.

[18] E. N. Cevallos Medina y C. J. Barahona Chunga, «Modelo para automatizar el proceso de predicción de la deserción en estudiantes universitarios en el primer año de estudio,» Tesis de licenciatura, Universidad Peruana de Ciencias Aplicadas, Lima, 2021.

[19] V. R. De La Cruz Quispe, «Diseño de un modelo predictivo basado en machine learning para el control de la deserción de estudiantes en la Universidad Ricardo Palma,» Tesis de licenciatura, Universidad Nacional Tecnólogica de Lima Sur, Villa el Salvador, 2019.

[20] M. S. Albán Taipe, «Contribuciones a la predicción de la deserción universitaria a través de minería de datos,» Tesis doctoral, Universidad Nacional Mayor de San Marcos, Lima, 2019.

[21] G. M. Lozano Chamorro, «Business Analytics para el pronóstico de deserciones de estudiantes en el Colegio Héroes de la Breña,» Tesis de licenciatura, Universidad César Vallejo, Lima, 2021.

[22] M. R. Alfonso Centeno, «Big Data. Técnicas de machine learning para la creación de modelos predictivos para empresas,» Tesis de licenciatura, Comillas Universidad Pontificia, Madrid, 2020.

[23] Wikipedia contributors, «Wikipedia, The Free Encyclopedia.,» [En línea]. Available: https://en.wikipedia.org/wiki/Machine_learning. [Último acceso: 2023].

[24] Google, «Developers.google,» [En línea]. Available: https://developers.google.com/machine-learning/crash-course/classification/accuracy?hl=en. [Último acceso: 2023].

[25] H2O, «H2O.ai,» [En línea]. Available: https://h2o.ai/wiki/prediction/. [Último acceso: 2023].

[26] G. Azabache Martínez, «Brave Developer,» 5 marzo 2022. [En línea]. Available: https://bravedeveloper.com/2022/03/05/time-complexity-o-complejidad-temporal-eficiencia-de-los-algoritmos-en-terminos-de-tiempo/.

[27] O. Martínez de Quel Pérez y M. Sillero Quintana, «Sobre la expresión “respuesta de reacción” y el concepto "tiempo de respuesta",» *Apunts*, 2014.

[28] soloelectronicos, «Soloelectronicos,» 19 enero 2022. [En línea]. Available:

https://soloelectronicos.com/2022/01/19/que-es-la-puntuacion-f1/.

[29] R. Díaz, «The Machine Learners,» [En línea]. Available: https://www.themachinelearners.com/metricas-de-clasificacion/#Accuracy.

[30] L. P. Rouhiainen, Inteligencia artificial 101 cosas que debes saber hoy sobre nuestro futuro, España: Planeta, 2018.

[31] «Inteligencia Artificial,» *INCyTu,* 2018.

[32] A. Burkov, The hundred-page machine learning book, 2019.

[33] J. Díez García, «Machine learning aplicado al trading,» Universidad Pontificia Comillas, España, 2021.

[34] F. Gabel, «Some studies in machine learning using,» 2019.

[35] APD, «¿Cuáles son los tipos de algoritmos del machine learning?,» APD, España, 2019.

[36] IBM, «IBM,» Julio 2016. [En línea]. Available: https://www.ibm.com/es-es/topics/logistic-regression.

[37] M. P. Seminara, «La deserción universitaria: resiliencia como posiblidad de logro,» *Universidad Nacional Autónoma de México,* vol. 21, nº 5, p. 11, 2020.

[38] UTP, «UTP,» 23 mayo 2023. [En línea]. Available: https://www.utp.edu.pe/blog/orientacion-vocacional/que-es-la-vocacion-profesional.

[39] F. Coll Morales, «Economipedia,» 1 abril 2020. [En línea]. Available: https://economipedia.com/definiciones/carga-financiera.html.

[40] Colaboradores de Wikipedia, «Wikipedia,» 21 septiembre 2023. [En línea]. Available: https://es.wikipedia.org/wiki/Escasez#Referencias.

[41] S. Donoso D., R. Ruiz V., R. Schmal S. y M. Schaffernicht Sch., «Financiamiento de los estudios universitarios: un desafío estratégico de sustentabilidad,» *Universidad de Talca,* nº 29, 2004.

[42] J. O. Moreno Treviño y S. N. Cortez Soto, «Rendimiento académico y habilidades de estudiantes en escuelas públicas y privadas: evidencia de los determinantes de las brechas en aprendizaje para México,» *Revista de economía,* vol. 37, 2020.

[43] Á. Díaz Barriga, «Construcción de programas de estudio en la perspectiva del enfoque de desarrollo de competencias,» *Universidad Nacional Autónoma de México,* 2014.

[44] Mineducacion, [En línea]. Available: https://mineducacion.gov.co/1759/articles-356821_recurso_1.pdf.

[45] Universidad Europea, 17 Mayo 2023. [En línea]. Available: https://universidadeuropea.com/blog/tipos-formacion/.

[46] J. Pérez Porto y M. Merino, «Definición de,» 17 agosto 2020. [En línea]. Available: https://definicion.de/matricula/.

[47] K. Rivera Vergaray, «Modelo predictivo para la detección temprana de estudiantes con alto riesgo de deserción académica,» Redalyc, 2021.

[48] I. M. Poveda Velasco, «Los factores que influyen sobre la deserción universitaria. Estudio en la UMRPSFXCh - Bolivia, análisis con ecuaciones estructurales,» Universidad Mayor Real y Pontificia San Francisco Xavier de Chuquisaca, Bolivia, 2019.

[49] Q. V. Ingrith, «Análisis de las causas de deserción universitaria,» Universidad Nacional Abierta y a Distancia, Colombia, 2016.

[50] ZeroBit, «Platzi,» 2020. [En línea]. Available:

https://platzi.com/tutoriales/1469-algoritmos-practico/7790-que-son-los-algoritmos/.

[51] Gamco, «Gamco.es,» [En línea]. Available: https://gamco.es/glosario/modelo-predictivo/.

[52] Universidad de La Guajira, «uniguajira.edu.co,» [En línea]. Available: https://www.uniguajira.edu.co/diagnostico-e-identificacion-de-problemas.

[53] W. C. John, Research Design: Qualitative, Quantitative and Mixed Methods Approaches, Estados Unidos: SAGE, 2014.

[54] R. Hernández Sampieri, C. Fernández Collado y L. María del Pilar Baptista, Metdológia de la investigación, México: McGraw Hill, 2014.

[55] R. Hernández, C. Fernández y P. Baptista, Metodología de la Investigación, México D.F.: Mc Graw Hill, 2014.

[56] Wikipedia contributors, «Wikipedia, The Free Encyclopedia.,» [En línea]. Available: https://es.wikipedia.org/wiki/Metodo_hipotético-deductivo. [Último acceso: 2010].

[57] W. M. Grove y H. M. William, Hypothetico-Deductive Model, Estados Unidos: John Wiley & Sons, 2015.

[58] D. T. Campbell y C. S. Julian, Experimental and quasi-experimental designs for research, USA: Houghton Mifflin Company, 1963.

[59] C. Anguita, R. Labrador y D. Campos, «La encuesta como técnica de investigación. Elaboración de cuestionarios y tratamiento estadístico de los datos,» *Atención Primaria*, vol. 31, nº 8, pp. 527-538, 2003.

[60] C. Ortega, «QuestionPro,» [En línea]. Available: https://www.questionpro.com/blog/es/que-es-un-cuestionario/#que_es_un_cuestionario.

[61] A. M. Chaparro, «Centro Banamex,» 22 Marzo 2023. [En línea]. Available: https://www.centrobanamex.com.mx/que-es-la-validez-y-confiabilidad-de-un-instrumento-de-recoleccion-de-datos/.

[62] M. M. Terán Cázares, «Factores determinantes del Comportamiento de Ciudadania Organizacional: Su relación con el Entorno Sociocultural a través de la Cultura Organizacional y los atributos de Personalidad.,» Universidad Autónoma de Nuevo Leon, México, 2009.

[63] D. I. Candia Oviedo, «Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático,» Universidad Nacional de San Antonio ABAD del Cusco, Cusco, 2019.

[64] W. Forero Corba y N. Bennasar, «Diseño y simulación de un modelo de predicción para la evaluación de la competencia digital docente usando técnicas de Machine Learning,» *Edutec*, nº 89, pp. 18-43, 2024.

[65] K. L. Yandar, O. Revelo Sánchez y M. Bolaños Gonzáles, «Machine learning for electric power prediction: a systematic literature review,» *inycomp*, vol. 26, nº 2, 2024.

[66] S. Osorio Castrillon, L. M. Giraldo Marín, H. H. Jaramillo Villegas y C. C. Piedrahita Escobar, «Machine learning aplicado en la clasificación y predicción de la depresión: Una revisión sistemática,» *risti*, nº E47, pp. 363-375, 2022.

[67] F. Sonia, L. Laura y H. Waldo, «Caracterización de la deserción universitaria en la UNRN utilizando Minería de Datos. Un caso de estudio.,» Universidad Nacional de Río Negro, Argentina, 2013.

[68] C. León Aznar, «BigMark: análisis y predicción de calificaciones mediante Big Data.,» Universidad Complutense de Madrid, Madrid, 2018.

[69] J. C. González Sánchez y M. J. Peñaloza Pérez, «Identificación y predicción de estudiantes en riesgo de deserción académica por medio de modelos basados en machine learning,» Fundación Universitaria Los Libertadores, Colombia, 2021.

[70] E. P. Jazmine y Á. Cuervo Martínez, «Validez de contenido y juicio de expertos: Una aproximación a su utilización,» Universidad Nacional de Colombia, Colombia, 2008.

[71] M. Karimi-Haghighi, D. Hernández-Leo, C. Castillo y V. Moreno Oliver, «Predicting Early Dropout: Calibration and Algorithmic Fairness,» arXiv, 2021.

[72] Wikipedia, «wikipedia.org,» [En línea]. Available: https://es.wikipedia.org/w/index.php?title=Cuasi_experimento&oldid=149819260.

[73] P. Fernández Garcia, G. Vallejo Seco, P. Levacic Rojas y E. Tuero Herrero, «Validez Estructurada para una investigación cuasi-experimental de calidad. Se cumplen 50 años de la presentación en sociedad de los diseños cuasi-experimentales,» Scielo, 2014.

[74] B. Cabré, «DISEÑOS CUASI-EXPERIMENTALES Y LONGITUDINALES,» Universidad de Barcelona, España.

[75] R. V. Kevin, «Modelo predictivo para la detección temprana de estudiantes con alto riesgo de deserción académica,» Redalyc, 2021.

[76] «IBM,» [En línea]. Available: https://www.ibm.com/topics/machine-learning. [Último acceso: 2023].

[77] M. Rodriguez Urrego, «La investigación sobre deserción universitaria en Colombia 2006-2016. Tendencias y resultados,» *Pedagogía y Saberes*, nº 51, 2019.

[78] M. Nicuesa, «DefiniciónABC,» octubre 2022. [En línea]. Available: https://www.definicionabc.com/ciencia/metodo-empirico-analitico.php. [Último acceso: 2023].

[79] T. Romá Ferri, «RUA: Repositorio Institucional de la Universidad de Alicante,» 2014. [En línea]. Available: https://rua.ua.es/dspace/bitstream/10045/42141/1/DICS_T4_SRI_EvaluaResultados_C14_15.pdf.

[80] I. Castillo, «Lifeder,» 1 septiembre 2020. [En línea]. Available: https://www.lifeder.com/instrumentos-investigacion-documental/.