

UNIVERSIDAD NACIONAL DEL CALLAO

FACULTAD DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



“PROPUESTA DE MODELO PREDICTIVO A TRAVÉS DE MINERÍA DE DATOS Y DESERCIÓN ESTUDIANTIL UNIVERSITARIA DE LA ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD NACIONAL DEL CALLAO”

TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO DE SISTEMAS

AUTORES

CABALLERO CUSTODIO CHRISTIAN JEAN CLAUDE

HIDALGO BELLIDO ANGGIE BETSABE

LEÓN SAUCEDO JAVIER EDUARDO

ASESOR

Mag. CASAZOLA CRUZ OSWALDO DANIEL

LÍNEA DE INVESTIGACIÓN: INGENIERÍA Y TECNOLOGÍA










Callao, 2024

PERÚ

Document Information

Analyzed document	TSIS - CABALLERO_HIDALGO_LEON.docx (D181864836)
Submitted	2023-12-14 21:00:00 UTC+01:00
Submitted by	Unidad FIIS
Submitter email	fiis.investigacion@unac.edu.pe
Similarity	8%
Analysis address	fiis.investigacion.unac@analysis.arkund.com

Sources included in the report

W	URL: http://repositorio.undac.edu.pe/bitstream/undac/829/1/T026_40573846_M.pdf Fetched: 2020-11-30 01:58:24	 7
SA	Tesis ZAMBRANO JESUS Y ZAVALA WASHINGTON Modelo predictivo de deserción estudiantil para la Facultad de Ciencias Informáticas de la Universidad Técnica de Manabí..pdf Document Tesis ZAMBRANO JESUS Y ZAVALA WASHINGTON Modelo predictivo de deserción estudiantil para la Facultad de Ciencias Informáticas de la Universidad Técnica de Manabí..pdf (D129598588)	 2
SA	Universidad Nacional del Callao / METODOLOGÍAS Y TÉCNICAS DE MINERÍA DE DATOS PARA PREDECIR LA DESERCIÓN DE ESTUDIANTES ANTES UNIVERSITARIOS.docx Document METODOLOGÍAS Y TÉCNICAS DE MINERÍA DE DATOS PARA PREDECIR LA DESERCIÓN DE ESTUDIANTES ANTES UNIVERSITARIOS.docx (D109581857) Submitted by: icicyt@unac.edu.pe Receiver: icicyt.vri.unac@analysis.arkund.com	 8
SA	TESIS+V3-3.pdf Document TESIS+V3-3.pdf (D142484576)	 2
SA	Trabajo TSP_VF.docx Document Trabajo TSP_VF.docx (D142756086)	 1
SA	LEYTON TARQUI - PLAN DE TESIS.pdf Document LEYTON TARQUI - PLAN DE TESIS.pdf (D131947132)	 4
SA	TESIS_MANTILLA_ANTONIO_VFINAL.docx Document TESIS_MANTILLA_ANTONIO_VFINAL.docx (D171142637)	 1
W	URL: https://link.gale.com/apps/doc/A366458506/IFME?u=googlescholar&sid=IFME&xid=170505ba Fetched: 2021-12-19 04:31:11	 2
SA	M2.878_20222_PEC4.2 - Redacción de la memoria (Entrega Final)_20336348.txt Document M2.878_20222_PEC4.2 - Redacción de la memoria (Entrega Final)_20336348.txt (D171251468)	 1

INFORMACIÓN BÁSICA

FACULTAD: FACULTAD DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS

ESCUELA: ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS.

TÍTULO: “PROPUESTA DE MODELO PREDICTIVO A TRAVÉS DE MINERÍA DE DATOS Y DESERCIÓN ESTUDIANTIL UNIVERSITARIA DE LA ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD NACIONAL DEL CALLAO”.

**AUTOR(ES): CABALLERO CUSTODIO CHRISTIAN JEAN CLAUDE.
CÓDIGO ORCID: 0009-0008-9844-8343
DNI: 77920962**

**HIDALGO BELLIDO ANGGIE BETSABE
CÓDIGO ORCID: 0009-0006-2057-250X
DNI: 48653230**

**LEÓN SAUCEDO JAVIER EDUARDO
CÓDIGO ORCID: 0009-0005-3632-7907
DNI: 73964838**

**ASESOR: CASAZOLA CRUZ OSWALDO DANIEL
CÓDIGO ORCID: 0000-0003-2521-530X
DNI: 40081695**

**LUGAR DE EJECUCIÓN: FACULTAD DE INGENIERÍA INDUSTRIAL
Y DE SISTEMAS**

UNIDAD DE ANÁLISIS: ESTUDIANTES

TIPO DE INVESTIGACIÓN: APLICADA

ENFOQUE DE INVESTIGACIÓN: CUANTITATIVO

DISEÑO DE INVESTIGACIÓN: CORRELACIONAL

TEMA OCDE: OTRAS INGENIERÍAS Y TECNOLOGÍAS

HOJA DE REFERENCIA DEL JURADO Y APROBACIÓN

MIEMBROS DEL JURADO EVALUADOR

MG. MANUEL ABELARDO ALCÁNTARA RAMÍREZ	PRESIDENTE
MG. ANGELINO ABAD RAMOS CHOQUEHUANCA	SECRETARIO
MG. JOSÉ JESÚS BRINGAS ZÚÑIGA	VOCAL
MG. YESMI KATIA ORTEGA ROJAS	SUPLENTE

ASESOR: MG. DANIEL OSWALDO CASAZOLA CRUZ

N° de libro: 01

N° de folio: 08

N° de acta: 002-2024-I-CTT-IS

Fecha de aprobación de la tesis: 06 DE ENERO DEL 2024



ACTA DE SUSTENTACIÓN



ACTA DE SUSTENTACION POR MODALIDAD DE CICLO TALLER DE TESIS PARA LA OBTENCIÓN DEL TÍTULO PROFESIONAL DE INGENIERO DE SISTEMAS

ACTA N° 002-2024-I-CTT-IS

Siendo las 08.20 horas del día 06 de Enero del año 2024, encontrándose reunidos en el Auditorium de la FIIS, el **Dr. ENRIQUE GARCÍA TALLEDO**, en representación de la Rectora de la UNAC; el **JURADO DE SUSTENTACIÓN DE TESIS** (designado por resolución **002-2024-CF-FIIS**) de la Facultad Ingeniería Industrial y de Sistemas de la Universidad Nacional del Callao, para la evaluación de las Tesis que conllevan a la obtención del Título Profesional de **INGENIERO DE SISTEMAS**, el que se encuentra conformado por los siguientes docentes ordinarios:

PRESIDENTE	MG. MANUEL ABELARDO ALCÁNTARA RAMÍREZ
SECRETARIO	MG. ANGELINO ABAD RAMOS CHOQUEHUANCA
VOCAL	MG. JESÚS JOSÉ BRINGAS ZÚÑIGA
SUPLENTE	MG. YESMI KATIA ORTEGA ROJAS

Con el quórum reglamentario de ley y de conformidad con lo establecido por el Reglamento de Grados y Títulos vigente se dio inicio al Acto de Sustentación de la Tesis de los Bachilleres: **HIDALGO BELLIDO ANGGIE BETSABE, LEÓN SAUCEDO JAVIER EDUARDO, CABALLERO CUSTODIO CHRISTIAN JEAN CLAUDE**, quienes, habiendo cumplido con los requisitos para optar el Título Profesional de **INGENIERO DE SISTEMAS**, sustentan la tesis titulada **"PROPUESTA DE MODELO PREDICTIVO A TRAVÉS DE MINERÍA DE DATOS Y DESERCIÓN ESTUDIANTIL UNIVERSITARIA DE LA ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD NACIONAL DEL CALLAO"**, cumpliendo con la sustentación en acto público, de manera presencial.

Luego de la exposición, y de la absolución de las preguntas formuladas por el Jurado de Sustentación y efectuadas las deliberaciones pertinentes, **SE ACORDÓ:** Dar por **APROBADO** con la escala de calificación cuantitativa **(16)** y calificación cualitativa **(Muy Bueno)** a la presente tesis, conforme a lo dispuesto en el Art. 24 del Reglamento de Grados y Títulos de la UNAC, aprobado por Resolución de Consejo Universitario N° 150-2023-CU del 15 de junio del 2023.

Se dio por concluida la Sesión a las 08:50 horas del día 06 de enero del 2024.

MG. MANUEL ABELARDO ALCÁNTARA RAMÍREZ
Presidente

MG. ANGELINO ABAD RAMOS CHOQUEHUANCA
Secretario

MG. JESÚS JOSÉ BRINGAS ZÚÑIGA
Vocal

MG. YESMI KATIA ORTEGA ROJAS
Suplente



UNIVERSIDAD NACIONAL DEL CALLAO

Facultad de Ingeniería Industrial y de Sistemas

Comisión de Grados y Títulos

"Año del Bicentenario, de la consolidación de nuestra Independencia, y de la conmemoración de las heroicas batallas de Junín y Ayacucho"

DICTAMEN FAVORABLE

N° 009-2024-IS-TESIS-CC-CGT-FIIS

La Comisión de Grados y Títulos, habiendo visto y revisado el Expediente N° E2040955-SGD de fecha **18 de marzo de 2024**, del bachiller **León Saucedo Javier Eduardo** con código **1615215132** cumple con lo establecido para la obtención del Título Profesional de Ingeniero de Sistemas señalado en los Art. 80 y 81° que indica en la Resolución N° 150-2023-CU del 15 de Junio del 2023.

Que, en el Acta de Sustentación de Tesis N° **002-2024-I-CTT-IS** de fecha **6 de enero de 2024** con Ciclo Taller de Tesis para la obtención de Título Profesional de Ingeniero de Sistemas, sustentó la tesis titulada **"PROPUESTA DE MODELO PREDICTIVO A TRAVÉS DE MINERÍA DE DATOS Y DESERCIÓN ESTUDIANTIL UNIVERSITARIA DE LA ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD NACIONAL DEL CALLAO"**. lo cual el Jurado de Sustentación de Tesis acordó conceder el calificativo cualitativo **MUY BUENO** y calificativo cuantitativo de **16 (Dieciseis)**.

La **Comisión de Grados y Títulos** en su sesión del 27 de marzo de 2024 da la conformidad del cumplimiento de los requisitos para obtener **el Título Profesional de Ingeniero de Sistemas** por la Modalidad de Tesis con Ciclo Taller de Tesis.

Bellavista 27 de marzo de 2024


UNIVERSIDAD NACIONAL DEL CALLAO
FACULTAD DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS
COMISIÓN DE GRADOS Y TÍTULOS

MG. VÍCTOR EDGARDO ROCHA FERNÁNDEZ
PRESIDENTE


UNIVERSIDAD NACIONAL DEL CALLAO
FACULTAD DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS
COMISIÓN DE GRADOS Y TÍTULOS FIIS

MG. ROMEL DARIÓ BAZÁN ROBLES
SECRETARIO


UNIVERSIDAD NACIONAL DEL CALLAO
FACULTAD DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS
COMISIÓN DE GRADOS Y TÍTULOS FIIS

ING. OMAR TUPAC AMARU CASTILLO PAREDES
VOCAL

Silvia.
Archivo.

Dedicatoria

Esta tesis va dedicada a nuestras familias, por ser siempre nuestro soporte, nuestro motivo de salir adelante y por su amor incondicional que nos alumbra en nuestros caminos.

Agradecimiento

Nuestro más sincero agradecimiento a la universidad Nacional del Callao, que desde pregrado nos ha forjado académicamente. Agradezco mucho la ayuda de nuestros maestros, nuestros compañeros y a toda la universidad en general. A su vez agradezco mucho nuestra familia que con su abnegado apoyo incondicional ha sido posible que podamos terminar con gran satisfacción nuestra presente investigación.

Índice

RESUMEN.....	12
Abstract	13
INTRODUCCIÓN.....	14
I. PLANTEAMIENTO DEL PROBLEMA.....	16
1.1 Descripción de la realidad problemática.....	16
1.2. Formulación del problema.....	19
1.2.1. Problema general	19
1.2.2. Problemas Específicos	19
1.3. Objetivos	20
1.3.1. Objetivo general	20
1.3.2. Objetivos específicos.....	20
1.4. Justificación	20
1.5. Limitaciones.....	22
II. MARCO TEÓRICO	23
2.1 Antecedentes.....	23
2.1.1 Antecedente internacionales	23
2.1.2 Antecedentes nacionales	26
2.2 Bases teóricas	28
2.2.1 Modelos Predictivo	28
2.2.2 Minería De Datos.....	31
2.2.3 Deserción Estudiantil.....	32
2.2.4 Metodología CRISP-DM	34

2.3 Marco conceptual	37
Conceptualización de modelo	37
Conceptualización de predicción	38
2.4 Definición de términos básicos	38
III. HIPOTESIS Y VARIABLES	39
3.1 Hipótesis	39
3.1.1 Operacionalización de la variable	42
IV. DISEÑO METODOLÓGICO	44
4.1. Tipo y diseño de investigación.	44
4.1.1 Diseño de investigación.....	44
4.1.2 Tipo de investigación	44
4.2. Método de investigación.	45
4.3. Población y muestra.....	45
4.3.1 Población.....	45
4.3.2 Muestra	45
4.4. Lugar de estudio.	46
4.5. Técnicas e instrumentos para la recolección de la información.	46
4.5.1 Técnica.....	46
4.5.2 Instrumentos.....	46
V. RESULTADOS.....	47
5.1. Resultados descriptivos	47
5.2. Resultados inferenciales	58

VI. DISCUSIÓN DE RESULTADOS	66
6.1. Contrastación y demostración de la hipótesis con los resultados.	66
6.2. Contrastación de los resultados con otros estudios similares	74
6.3. Ética	78
VII. CONCLUSIONES.....	79
VIII. RECOMENDACIONES	80
IX. REFERENCIAS.....	82

Índice de Tablas

Tabla 1: Deserción universitaria internacional	16
Tabla 2: Deserción Universitaria por Etapa (Internacional)	18
Tabla 3: Deserción Universitaria Nacional (Perú).....	18
Tabla 4: Comprensión e identificación de los datos	36
Tabla 5: Operacionalización de variables	43
Tabla 6: ¿Con qué frecuencia su familia lo apoya para continuar sus estudios universitarios?	47
Tabla 7: ¿Con qué frecuencia recibe apoyo de su familia para sus estudios universitarios?	48
Tabla 8: ¿Cómo calificaría el nivel de apoyo que recibe de su familia para sus estudios universitarios?.....	49
Tabla 9: ¿En qué escala la familia lo motivan a continuar sus estudios universitarios?	50
Tabla 10: ¿Con qué frecuencia su familia lo apoya para continuar sus estudios universitarios?	51
Tabla 11: ¿Con qué frecuencia recibe apoyo de su familia para sus estudios universitarios?	52
Tabla 12: ¿Cómo calificaría el nivel de apoyo que recibe de su familia para sus estudios universitarios?	53
Tabla 13: ¿En qué escala la familia lo motivan a continuar sus estudios universitarios?	54
Tabla 14: Resumen del modelo para la hipótesis general	56
Tabla 15: Resumen del modelo para la hipótesis específica 1	56

Tabla 16: Resumen del modelo para la hipótesis específica 2.....	57
Tabla 17: Resumen del modelo para la hipótesis específica 3.....	58
Tabla 18: Matriz de confusión para la dimensión social	60
Tabla 19: Reporte de Clasificación para la dimensión social.....	60
Tabla 20: Importancia de características para la dimensión social.....	61
Tabla 21: Matriz de confusión para la dimensión Económica.....	62
Tabla 22: Reporte de clasificación para la dimensión económica	62
Tabla 23: Importancia de las características para la dimensión económica	62
Tabla 24: Matriz de confusión para la dimensión académica	63
Tabla 25: Reporte de Clasificación para la dimensión académica	63
Tabla 26: Importancia de características para la dimensión académica	64
Tabla 27: Matriz de confusión totalizada	65
Tabla 28: Prueba de muestras emparejadas - Hipótesis General	67
Tabla 29: Prueba de muestras emparejadas - Hipótesis Específica 1 ..	69
Tabla 30: Prueba de muestras emparejadas - Hipótesis Específica 2 ..	71
Tabla 31: Prueba de muestras emparejadas - Hipótesis Específica 3 ..	73

Índice de ilustraciones

Ilustración 1: Deserción universitaria internacional Fuente: elaboración propia.....	17
Ilustración 2: Deserción Universitaria Nacional (Perú).....	19
Ilustración 3: Minería de datos Educativos	21
Ilustración 4: Metodología CRISP - DM.....	22
Ilustración 5: ¿Con qué frecuencia su familia lo apoya para continuar sus estudios universitarios?	47
Ilustración 6: ¿Con qué frecuencia recibe apoyo de su familia para sus estudios universitarios?	48
Ilustración 7: ¿Cómo calificaría el nivel de apoyo que recibe de su familia para sus estudios universitarios?	49
Ilustración 8: ¿En qué escala la familia lo motivan a continuar sus estudios universitarios?	50
Ilustración 9: ¿Con qué frecuencia su familia lo apoya para continuar sus estudios universitarios?	51
Ilustración 10: ¿Con qué frecuencia recibe apoyo de su familia para sus estudios universitarios?	52
Ilustración 11: ¿Cómo calificaría el nivel de apoyo que recibe de su familia para sus estudios universitarios?	53
Ilustración 12: ¿En qué escala la familia lo motivan a continuar sus estudios universitarios?	54
Ilustración 13: R Pearson	55

RESUMEN

La deserción estudiantil es un problema muy común en todas las universidades y trae consigo diversos problemas que no solo afecta al desertor sino también la imagen de la universidad, es por ello que se realizó una propuesta de un modelo predictivo aplicando minería de datos para la deserción universitaria en la escuela profesional de Ingeniería de Sistemas, utilizando árbol de decisión como técnica principal por su alto grado de precisión en pronósticos de deserción universitaria y CRISP-DM como metodología por sus 6 fases que la hacen completa para este tipo de modelamiento predictivo. Para la adquisición de información, se realizó una encuesta a 65 alumnos de la facultad de Ingeniería de Sistemas, donde cada uno mencionó los diversos factores del motivo de su salida definitiva de la carrera, Después de la evaluación con los datos originales se introduce un conjunto de datos de prueba en el sistema para analizar los resultados. El análisis comparativo de los resultados indica que la predicción ha ayudado determinar con mayor precisión el mejoramiento en el resultado.

Palabras clave: Deserción estudiantil, Predicción, Minería de Datos, CRISP-DM, Árbol de decisión

Abstract

Student dropout is a very common problem in all universities and brings with it various problems that not only affects the dropout but also the image of the university, which is why a proposal for a predictive model was made applying data mining for dropout. university in the professional school of Systems Engineering, using decision tree as the main technique due to its high degree of precision in university dropout forecasts and CRISP-DM as a methodology due to its 6 phases that make it complete for this type of predictive modeling. To acquire information, a survey was carried out with 65 students from the Faculty of Systems Engineering, where each one mentioned the various factors of the reason for their definitive departure from the degree. After the evaluation with the original data, a set of test data into the system to analyze the results. The comparative analysis of the results indicates that the prediction has helped determine with greater precision the improvement in the result.

Keywords: Student dropout, Prediction, Data Mining, CRISP-DM, Decision tree

INTRODUCCIÓN

Según (Alania 2018), el objetivo de esta investigación es anticipar la manera en que el fenómeno de la deserción estudiantil ha evolucionado para convertirse en un desafío social que impacta a numerosas instituciones universitarias en todo el territorio peruano. La disminución de la cantidad de estudiantes que abandonan sus estudios se ha convertido en una problemática relevante en la actualidad, siendo una preocupación constante para cada una de las universidades. Estas instituciones tienen en mente implementar estratégicamente un plan con el fin de reducir la tasa de estudiantes que optan por dejar sus estudios.

En la investigación de (Luna Puicon et al. 2022), ha brindado una comprensión más profunda del problema de la deserción, revelando que no se trata únicamente de una cuestión institucional. Se ha reconocido que todos los participantes en el proceso de enseñanza-aprendizaje contribuyen de alguna manera a este fenómeno. Por lo tanto, abordar y resolver este problema implica la colaboración y el compromiso de todos los involucrados en el sistema educativo. Desde el momento inicial en que una persona se acerca a la institución con la intención de formar parte de ella, se espera que se comprometa con el proceso, facilitando así una inserción más fluida y efectiva. Esto, a su vez, favorecerá la retención a lo largo de la formación académica hasta su conclusión.

Por otra parte, la minería de datos es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto. Los algoritmos de uso común de minería de datos: Regresión lineal, regresión logística, árbol de decisión, máquinas de vectores de soporte, red Neuronales, para la construcción del modelo de minería de datos basado en series de tiempo logrando realizar las predicciones de deserción escolar y el uso de la metodología XP para el desarrollo del sistema. Sobre la base de lo anterior y considerando los

estudios previos realizados, se concluye que mediante técnicas de minería de datos es posible determinar modelos que estimen la deserción estudiantil universitaria, analizando la historia académica del estudiante junto a los factores socio económicos y otros, que determinan su condición de potencial desertor (Alvarado et al. 2020).

I. PLANTEAMIENTO DEL PROBLEMA

1.1 Descripción de la realidad problemática

A nivel Internacional la deserción universitaria constituye un desafío que impacta a todas las instituciones de educación superior. Se estima que alrededor del 50% de los estudiantes no logran completar sus estudios hasta obtener el título universitario (Dojchinovska, 2021). En Estados Unidos, aproximadamente el 40% de los estudiantes universitarios de pregrado abandonan sus estudios, mientras que, en la Unión Europea, países como el Reino Unido, Noruega y Francia presentan tasas de deserción universitaria en torno al 16%, 17% y 24%, respectivamente (Rudin, 2019). No obstante, en Italia se registra la tasa de abandono más elevada, alcanzando el 33%, seguido de cerca por los Países Bajos con un 31%. En el caso de América Latina, se observa que el 50% de las personas de 25 a 29 años que comenzaron estudios universitarios no lograron concluirlos (Behr, y otros, 2020).

País/Región	Tasa de Deserción (%)
Estados Unidos	40
Unión Europea	
- Reino Unido	16
- Noruega	17
- Francia	24
Italia	33
Países Bajos	31
América Latina	50

Tabla 1: Deserción universitaria internacional

Fuente: elaboración propia

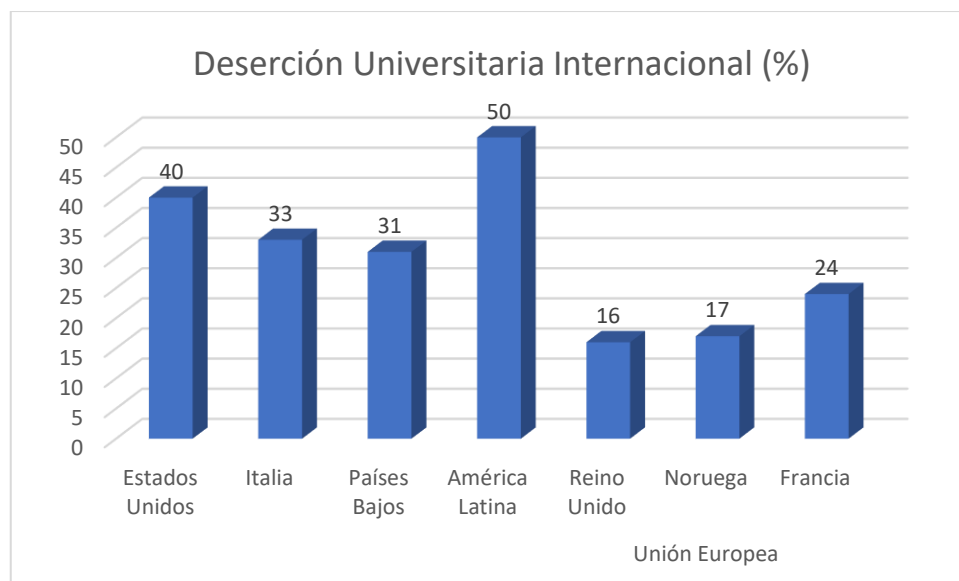


Ilustración 1: Deserción universitaria internacional

Fuente: elaboración propia

Es importante señalar que hay diversos tipos de abandono de la educación universitaria, y cada uno puede ser examinado desde perspectivas distintas. A pesar de que la deserción puede ocurrir en cualquier etapa de los estudios, la mayoría de los casos se registran durante los primeros años, con una proporción significativa. En los Estados Unidos, alrededor del 30% de los estudiantes universitarios de primer año dejaron la educación antes de llegar al segundo año, según datos hasta aproximadamente el año 2018. (Rudin, 2019), En el Reino Unido, el 6,3% de los estudiantes a tiempo completo abandonaron sus estudios durante el primer año en el período 2016-2017. En América Latina, Colombia experimentó una tasa de deserción del 31% en el primer año, mientras que Chile registró un 21%, y Perú alcanzó un 33% en el año 2019, previo al inicio de la pandemia. (Behr, y otros, 2020).

País/Región	2016-2017	2017-2018	2018-2019	2019	2020	2021
Estados Unidos	-	-	-	30	28	25
Colombia	-	-	-	31	28	26
Chile	-	-	-	21	19	17
Perú	-	-	-	33	30	28
Reino Unido	6.3	5.8	6.5	-	-	-

Tabla 2: Deserción Universitaria por Etapa (Internacional)
Fuente: elaboración propia

A nivel nacional de acuerdo con (Sociedad et al. 2022), basándose en informes estadísticos, el II Informe Bienal sobre la realidad universitaria en el Perú, elaborado por SUNEDU en el año 2020, señala que entre los años 2012 y 2018, se produjo una deserción de las aulas universitarias a nivel nacional, oscilando entre el 15.8% y el 17.6%, dependiendo de la región geográfica. Además, proporciona datos más detallados, indicando que en la selva fue del 24.6%, en la costa del 24%, y en la sierra del 18.2%. Frente a la situación de la pandemia de Covid-19, el 42.6% de los estudiantes universitarios clasificados en el nivel no extremo de pobreza dejaron sus estudios superiores. En contraste, en el nivel no pobre, la tasa de abandono fue del 18.1% (Benites 2021).

Región	2012	2013	2014	2015	2016	2017	2018
Nacional	16	16.5	16.2	16.8	17.1	17.3	17.6
Selva	20.5	21	20.8	21.2	22	23	24.6
Costa	16.8	17	16.5	17.2	17.5	18	24
Sierra	14.5	15	14.8	15.5	16	16.5	18.2

Tabla 3: Deserción Universitaria Nacional (Perú)
Fuente: elaboración propia

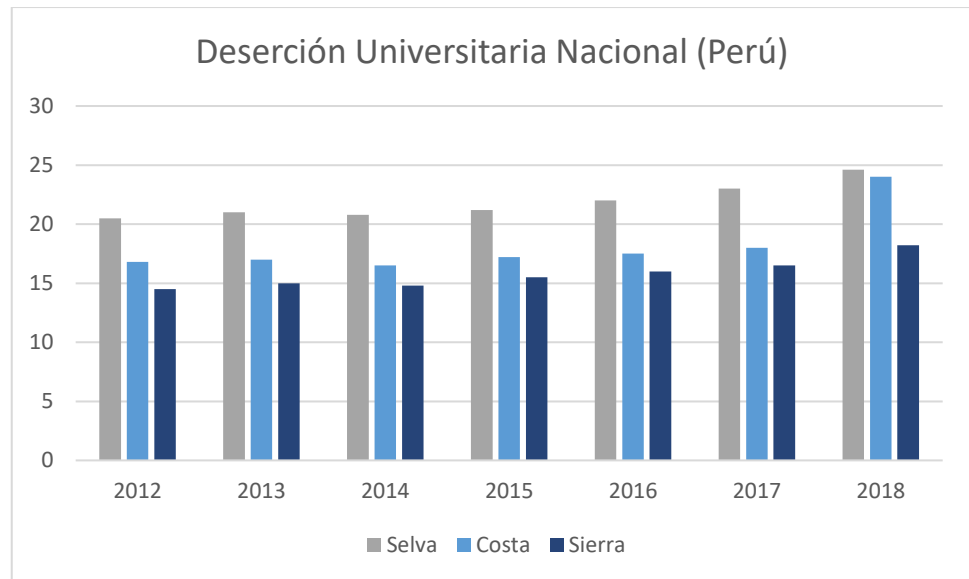


Ilustración 2: Deserción Universitaria Nacional (Perú)

Fuente: elaboración propia

1.2. Formulación del problema

1.2.1. Problema general

¿De qué manera un modelo predictivo aplicando minería de datos determina la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao?

1.2.2. Problemas Específicos

PE1: ¿Se podrá comprobar si el factor social influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao aplicando la minería de datos?

PE2: ¿Se podrá comprobar si el factor económico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao aplicando la minería de datos?

PE3: ¿Se podrá comprobar si el factor académico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao aplicando la minería de datos?

1.3. Objetivos

1.3.1. Objetivo general

Proponer un modelo predictivo que determine la deserción estudiantil universitaria a través de la minería de datos de la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao.

1.3.2. Objetivos específicos

OE1: Comprobar si el factor social influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

OE2: Comprobar si el factor económico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

OE3: Comprobar si el factor académico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

1.4. Justificación

La Minería de Datos Educativos (Educational Data Mining - EDM) como exponen, es una disciplina relacionada con el desarrollo de métodos para extraer información útil a partir de los datos que se generan en los entornos educativos, y utilizarla para mejorar dicho

entorno. La información así obtenida se convierte en el insumo indispensable para la toma de decisiones (Romero, Ventura 2013).

EDM con las diversas técnicas de clasificación, la predicción comienza con la recopilación de datos actuales para pronosticar un futuro alternativo, con el fin de monitorear los cambios en los datos para detectar los patrones que generan dicho cambio, es imprescindible realizar un análisis temporal de las BD. Dentro de las metodologías para la extracción y manipulación de información CRISP-DM como metodología proporciona no solamente las fases generales de construcción y evaluación de un modelo, también provee una guía detallada sobre las tareas correspondientes en cada fase para la aplicación de la misma, teniendo así un ciclo de vida iterativo (Romero, Ventura 2013).

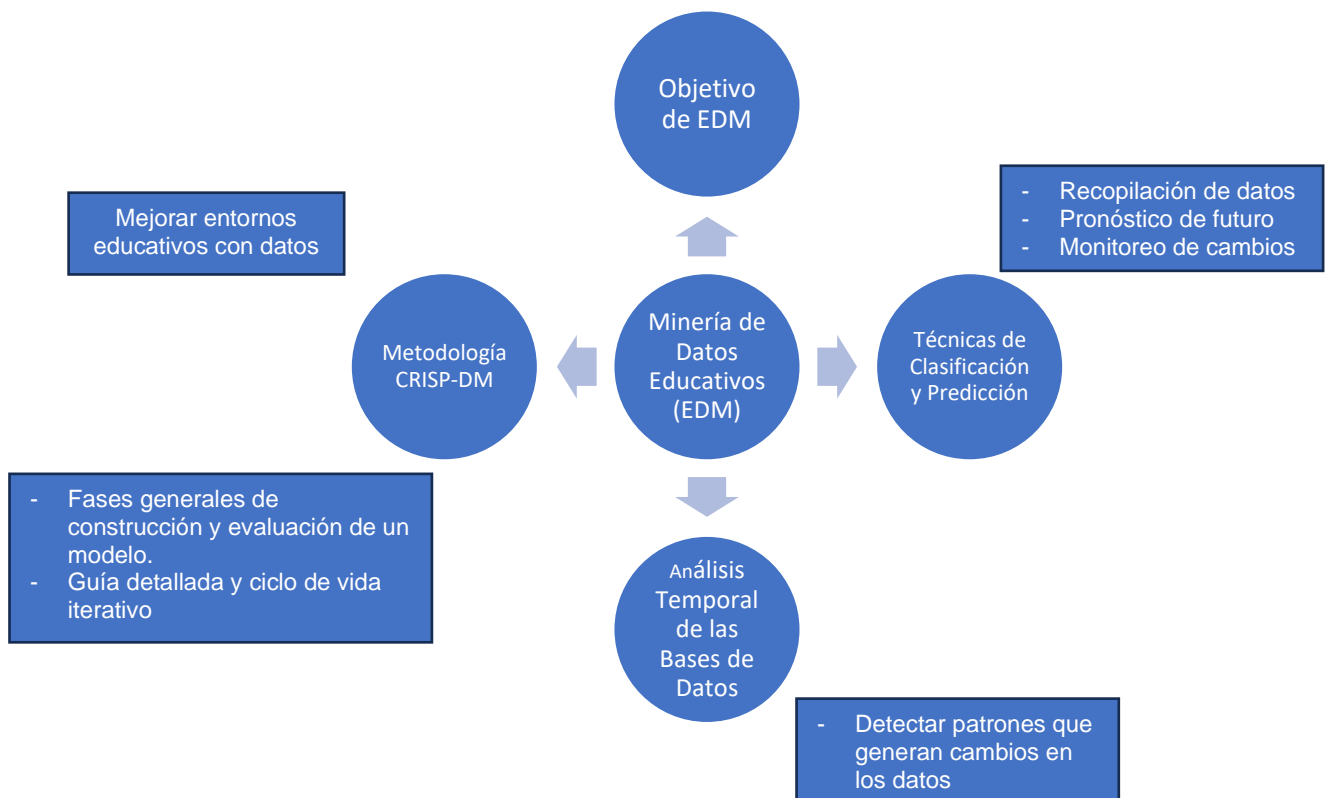


Ilustración 3: Minería de datos Educativos

Fuente: Elaboración Propia

METODOLOGÍA CRISP-DM

Se empleó la metodología Cross-Industry Standard Process for Data Mining (CRISP-DM) y la herramienta Clementine para realizar la extracción automatizada de conocimiento. Utilizando una red neuronal perceptrón multicapa, se logró identificar las variables que tienen un mayor impacto en el éxito y abandono de los estudios universitarios. Para generar el árbol de decisión, se aplicó el algoritmo C5.0, mientras que para el agrupamiento en tres grupos se utilizaron los algoritmos k-means y TwoStep. Estos resultados fueron comparados mediante una matriz de confusión (Gamarra, Matos, Yupanqui 2018).

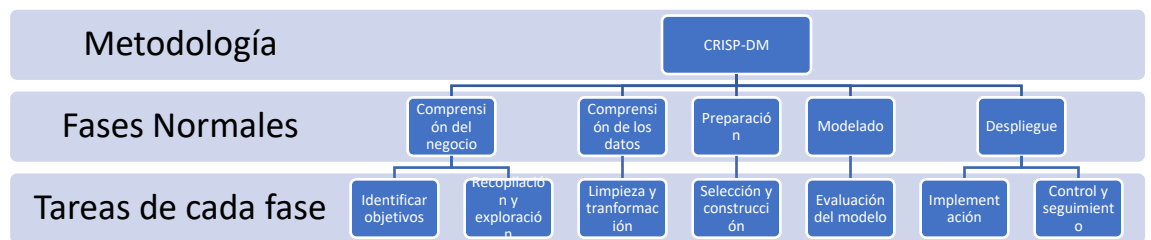


Ilustración 4: Metodología CRISP - DM

Fuente: Elaboración Propia

1.5. Limitaciones

Con el propósito de hacer factible el proyecto de investigación se establecieron criterios que limiten el alcance del proyecto, estableciendo las restricciones teóricas, espaciales y temporales.

Teórico

Existen diferentes tipos de metodologías dominantes para el proceso de la minería de datos dentro de las cuales se mencionan las metodologías KDD, CRISP-DM y SEMMA. En este proyecto se ha creído conveniente solo utilizar la metodología CRISP-DM, para

predecir la deserción universitaria de la carrera de ingeniería de sistemas de la Universidad Nacional del Callao.

Espacial

Para hacer válido el uso del proyecto y los resultados, se consideró conveniente desarrollar un modelo predictivo para determinar la deserción universitaria aplicando minería de datos a través de casos de estudio: identificación de patrones para la predicción de deserción universitaria.

Temporal

La verificación de los resultados se restringe a los semestres correspondientes entre los períodos de 2021 a 2023, períodos en que se hizo factible la recolección de información.

II. MARCO TEÓRICO

2.1 Antecedentes

Se presenta a continuación el contexto histórico del trabajo de investigación realizado. En esa parte se describen y analizan las investigaciones previas relacionadas con el tema de investigación, realizadas tanto en el país como en el extranjero.

2.1.1 Antecedente internacionales

En la tesis titulada “Minería de datos aplicada a estrategias para minimizar el rezago académico y la deserción universitaria en carreras de informática de la UNNOBA”. Para obtener el título profesional de Doctor en Ciencias Informáticas en la Universidad de Nacional de la Plata. Tuvo como objetivo crear un modelo de intervención mediante el uso de aulas virtuales con el fin de concebir e instaurar tácticas tecnológicas que simplifiquen el monitoreo del rendimiento académico de aquellos estudiantes que enfrentan

desafíos educativos o corren el riesgo de dejar sus estudios. En lo que concierne a la metodología utilizada, se llevó a cabo a través de un análisis retrospectivo de carácter descriptivo y correlacional, caracterizado por la adopción de un enfoque integrado multimodal. Este enfoque se fundamentó en la fusión y cotejo de métodos cuantitativos y cualitativos. En este estudio, se seleccionó una muestra de 63 estudiantes de la universidad en cuestión. En relación a las herramientas utilizadas, se emplearon dos: una encuesta y una entrevista. La encuesta, llevada a cabo mediante un cuestionario semidirigido y de autogestión, permitió la recopilación de datos estandarizados de un grupo específico de individuos con el objetivo de obtener una comprensión más profunda de la población. En relación con la investigación, se llega a la conclusión de que se puede anticipar posibles casos de abandono, lo que contribuirá a mejorar las políticas institucionales destinadas al apoyo y retención de los estudiantes (Russo 2019).

En el artículo científico denominado “Minería de datos aplicada a la deserción estudiantil. Caso: Licencia en Computación – Universidad del Zulia - NPF” publicado en la Revista EDUCARE. Tienen como objetivo obtener información acerca de los factores que afectan la capacidad de los estudiantes para concluir sus estudios universitarios en la Licenciatura en Computación de LUZ-NPF. Para llevar a cabo la ejecución de la labor, se empleó una investigación de carácter descriptivo, siguiendo la metodología computacional CRISP-DM y utilizando la herramienta WEKA. Para lograrlo, la muestra seleccionada se derivó de tres grupos: estudiantes en los primeros tres semestres, profesores en los primeros cinco semestres del periodo I-2012 y los informes de matrícula del período 2008-2011 proporcionados por Control

de Estudios. Los instrumentos empleados incluyeron Árboles de decisión C4.5 y la técnica de los vecinos más cercanos. Finalmente, la investigación tuvo como resultado, que entre las tendencias identificadas se destacaban la limitada experiencia previa en lógica y matemáticas, la falta de recursos económicos para adquirir equipos de computación, la falta de enfoque en los estudios y una baja cantidad de horas dedicadas al estudio. En relación con la conclusión, se determina que, gracias a la precisión del modelo, logró identificar las tres causas principales por las cuales los estudiantes de la Licenciatura tienden a abandonar. Estas son: insuficiencia de conocimientos adquiridos para su retención, carencia de recursos académicos y falta de concentración en los estudios con escasa dedicación de horas al mismo (Marcano, Rodríguez 2015).

Como otro antecedente internacional titulada “Modelo para la predicción de la deserción de estudiantes de pregrado, basado en técnicas de minería de datos”, tiene como objetivo principal de este proyecto de investigación es desarrollar un modelo para anticipar la deserción de estudiantes de pregrado en la Universidad de la Costa - CUC, mediante el análisis de diversos factores socioeconómicos y académicos. Utilizando una metodología aplicada y experimental, se creó un conjunto de datos (dataset) en la fase de caracterización, recopilando información demográfica, cultural, social, familiar, educativa, estatus socioeconómico y perfil psicológico de cada estudiante. Esta recopilación abarcó los periodos desde 2013-1 hasta 2018-2. La obtención de dicha información se llevó a cabo mediante los formularios de inscripción completados por los estudiantes al ingresar a la institución, recopilando un total de 1.606 registros individuales de estudiantes. En conclusión, se observó que el algoritmo que

logró la mayor tasa de aciertos fue Random Forest, perteneciente a la categoría de árboles de decisión, alcanzando una precisión del 84.8%. Durante la etapa de desarrollo, se incorporó el modelo a una aplicación que posibilita la predicción de la probabilidad de deserción de un estudiante o un conjunto de ellos (Camargo 2020).

2.1.2 Antecedentes nacionales

En la tesis titulada “Modelo de minería de datos basado en factores asociados para la predicción de deserción estudiantil universitaria”. Para obtener el título profesional de Ingeniero de Sistemas en la Universidad Nacional de Moquegua. Planteó como objetivo desarrollar un modelo de minería de datos que examinara el paradigma de aprendizaje con el fin de abordar la problemática de la deserción estudiantil que afecta a las universidades. Se utilizó un enfoque experimental en la metodología del proyecto. El estudio tiene como objetivo analizar las características de las variables del modelo predictivo en relación con la variable dependiente, con el propósito de identificar o describir las características más relevantes para abordar el problema planteado en la investigación. Se empleó como muestra a los estudiantes matriculados en el primer ciclo académico durante el periodo 2017-2 en la Universidad Nacional de Moquegua para desarrollar un modelo predictivo. Los instrumentos utilizados para la recopilación de datos fueron aplicados a los estudiantes que ingresaron a la filial Ilo de la Universidad Nacional de Moquegua durante el semestre académico 2017-2, y la información se obtuvo a través de la Oficina de Actividades y Servicios Académicos. En conclusión, el modelo predictivo fue capaz de identificar y categorizar a los estudiantes propensos a abandonar. Se implementó una

metodología directa para construir los modelos, siguiendo el paradigma de aprendizaje, con el fin de comprender las fases críticas del proyecto durante el proceso de retroalimentación (Mamani 2019).

Como otro antecedente nacional titulada “Aplicación de técnicas de minería de datos para predecir la deserción estudiantil en la educación básica regular en la región de Lambayeque”. Planteó como objetivo proponer una herramienta mediante el empleo de técnicas de minería de datos, con el propósito de brindar al usuario acceso a información precisa para realizar predicciones sobre los estudiantes que se inscribirán en los años venideros. Esto se traducirá en resultados a corto plazo que garantizarán la confiabilidad de dichas predicciones, proporcionando así apoyo a la institución en la toma de decisiones futuras. Referente a la metodología y herramientas que utilizó, se clasificó como tecnológica y de diseño experimental. Se considera tecnológica debido a que, mediante un enfoque científico, busca desarrollar aplicaciones prácticas (investigación aplicada) para la utilización de un producto o para mejorar dicho producto. El elemento de estudio determinado como población es el Elemento de Registro en los períodos 2006 – 2015 en la Región Lambayeque donde está conformado por tres ugeles: Chiclayo, Lambayeque y Ferreñafe. Se concluyó que la metodología CRISP DM fue impotante para desarrollar el modelo de minería de datos centrado en series temporales con el objetivo de prever la deserción escolar en la región de Lambayeque (Piscoya 2016)

En la tesis titulada “Aplicación de un Modelo de Minería de Datos para Identificación de Patrones que Influyen en la Deserción Académica en el Instituto Superior Leonardo

Davinci Usando IBM SPSS MODELLER y la Metodología CRISP-DM". Para obtener el título profesional de Ingeniero de Computación y Sistemas en la Universidad Privada Antenor Orrego. Se planteó como objetivo desarrollar un modelo de minería de datos con la finalidad de identificar las posibles razones o patrones que conducen a la deserción estudiantil. Esto se basó en el análisis de las características de los datos de los alumnos. Teniendo como metodología de investigación método hipotético – deductivo y diseño Pre-experimental con pre-prueba y post-prueba. Utilizando como población Todos los estudiantes del Instituto Superior Leonardo Davinci y como muestra los registros de estudiantes de los años 2018 -2019 almacenados en la base de datos del Instituto Superior Leonardo Davinci. En relación con las conclusiones del estudio, se determina que, tras la aplicación de cuatro modelos de minería de datos, a saber, Árbol C&R, Árbol C5.0, Árbol AS y Red Bayesiana, el modelo de Árbol C5.0 demostró un análisis de datos más efectivo con una precisión del 94.203%. Este modelo fue seleccionado como el preferido para examinar los patrones relacionados con la deserción académica de los estudiantes. Como recomendación, se sugiere la utilización de la herramienta IBM SPSS Modeler para implementar soluciones de minería de datos debido a su facilidad de uso y interfaces comprensibles (Pando, Zarate 2020).

2.2 Bases teóricas

2.2.1 Modelos Predictivo

Concepto de modelo predictivo

Un modelo predictivo es un mecanismo que predice el comportamiento de un individuo. Utiliza las características

del individuo como entrada y proporciona una calificación predictiva como salida (Espino 2017).

Para qué sirve el Modelo Predictivo

Un modelo predictivo responde preguntas sobre datos futuros (Beltran 2018).

Los modelos predictivos sirven para evaluar situaciones de tipo “qué pasaría si” y servir de guía para la toma de decisiones y prevenir situaciones adversas (Beltran 2018).

Los modelos predictivos ayudan a diversos sectores como el Marketing, servicios financieros, administraciones públicas, empresariales, etc. En el marketing, por ejemplo, permite predecir el comportamiento de los clientes y evaluar determinadas acciones (Beltran 2018).

Por tanto, un Modelo Predictivo sirve para analizar qué puede pasar, y se puede desarrollar mediante el uso de diversas técnicas de minería de datos que combinan las matemáticas e inteligencia artificial, a partir de variables específicas (Beltran 2018).

Técnicas de análisis predictivos

A pesar de las diferencias metodológicas y matemáticas entre los tipos de modelos, el objetivo general de todos ellos es similar: predecir resultados futuros basándose en datos pasados.

Aunque hay algunas técnicas que son específicas de clasificación y otras de regresión, la mayoría de las técnicas funcionan con ambos. Un motivo de confusión frecuente es la técnica de regresión logística, que solo funciona para problemas de clasificación y no de regresión.

Árboles de decisión

Son herramientas analíticas empleadas para el descubrimiento de reglas y relaciones mediante la ruptura y subdivisión sistemática de la información contenida en el conjunto de datos. El árbol de decisión se construye partiendo el conjunto de datos en dos (CART) o más (CHAID) subconjuntos de observaciones a partir de los valores que toman las variables predictoras (Beltran 2018).

Redes neuronales

Las redes neuronales son una nueva forma de analizar la información con una diferencia fundamental con respecto a las técnicas tradicionales: son capaces de detectar y aprender patrones y características dentro de los datos (Beltran 2018).

Análisis bayesiano

Según Beltran, las redes bayesianas son una alternativa para minería de datos, la cual tiene varias ventajas (Beltran 2018).

- Permiten aprender sobre relaciones de dependencia y causalidad.
- Permiten combinar conocimiento con datos.
- Evitan el sobre-ajuste de los datos.
- Pueden manejar bases de datos incompletas.

Regresión logística

Beltrán, lo describe como: “Una técnica básica de Data Mining, pueden trabajar con variables discretas y también requiere que todas las variables sean lineales”. La Regresión Logística o Logistic Regression (LR) en inglés, es un caso especial de regresiones cuyo uso es para predecir el resultado de una variable dependiente categórica, tiene

bastante uso en los cálculos de probabilidades, donde se predice la ocurrencia de un evento en función de otros factores (Beltran 2018).

Regresión lineal

“Una técnica básica de Data Mining, se implementa identificando una variable dependiente (y) y todas las variables independientes (X1, X2 ...)” (Beltran 2018).

2.2.2 Minería De Datos

La minería de datos puede definirse inicialmente como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos mediante el uso de métodos automatizados (Morales A 2018).

La técnica denominada como Data Mining puede ser definida como el proceso de extracción de información y patrones de comportamiento que permanecen ocultos entre grandes cantidades de información: Es un proceso iterativo en el que a los avances que se van produciendo en cada paso se les denomina descubrimientos (KDD – Knowledge Discovery in Database) (Beltran 2018).

La Minería de Datos Educativa permite modelar la deserción de estudiantes considerando sus datos socioeconómicos, académicos y personales (Alvarado et al. 2020).

Se define a la minería de datos educativa, resaltando su capacidad para identificar patrones nuevos y no triviales para resolver y mejorar procesos educativos (Camino-Hampshire, Cruz Barbosa, Urbina-Nájera 2020).

Metodologías

Las tres metodologías dominantes para el proceso de la minería de datos son: KDD por sus siglas en inglés “”, CRISP-DM y SEMMA.

KDD (KDD – Knowledge Discovery in Database)

El término Descubrimiento de Conocimiento en Bases de Datos

(Knowledge Discovery in Databases, o KDD para abreviar) empezó a utilizarse en 1989 para referirse al amplio proceso de búsqueda de conocimiento en bases de datos, y para enfatizar la aplicación a "alto nivel" de métodos específicos de minería de datos (Beltran 2018).

CRISP-DM

Cross Industry Standard Process for Data Mining (CRISP-DM)

Es una metodología muy usada por los investigadores y la comunidad académica por el nivel detalle con el que describen las tareas en cada fase del proceso, y además también porque incorporan actividades de gestión proyecto (Moine 2017).

SEMMA

SEMMA es un acrónimo de Sample, Explore, Modify, Model, and Assess. De acuerdo con (Moine 2017), esta metodología está enfocada al modelado de la minería de datos en otros términos en sus procesos de selección, exploración y modelado de grandes volúmenes de datos para descubrir patrones de negocio desconocidos.

2.2.3 Deserción Estudiantil

La deserción estudiantil es un problema complejo y crucial en el campo de la educación, que está presente en todos los

niveles y modalidades del sistema educativo, por tanto, la detección temprana es una estrategia clave para las instituciones académicas (Alvarado et al. 2020).

Para (Vries et al. 2011) la deserción es considerada como un acto de abandono de la universidad y como el fracaso para completar un programa de estudios.

Según (García, 2019) la deserción es un número de estudiantes matriculados que no siguen la trayectoria normal del programa académico, ya sea por retirarse de ella o por demorar más tiempo de lo previsto en finalizar, por repetir cursos o retiros temporales.

Naturaleza de la deserción

El tema de la deserción estudiantil es un tema que ha sido examinado por algunas fundaciones mundiales, por ejemplo, UNESCO, OCDE, IELSAC y CINDA, en las que existe una preocupación increíble por los marcadores locales (García, 2019).

Es una obligación de las entidades educativas, especialmente las universitarias, establecer mecanismos académicos, administrativos y de ajuste a la vida universitaria de sus estudiantes para que superen en debida forma las dificultades de los programas académicos y culminen con éxito la carrera en la cual cada estudiante ha cifrado su meta.

Clases de deserción

Las clases de deserción que se presentan se basan en el estudio de (Páramo, Correa 2012).

- Deserción total: abandono definitivo de la formación académica individual.

- Deserción discriminada por causas: según la causa de la decisión.
- Deserción por facultad: cambió facultad-facultad.
- Deserción por carrera profesional: cambio de carrera en una misma facultad.
- Deserción en los primeros ciclos de carrera: por inadecuada adaptación a la vida universitaria.
- Deserción acumulada: sumatoria de deserciones en una institución.

Características del desertor

Las características del desertor que se presentan se basan en el estudio de (Páramo, Correa 2012).

- Problemas de disciplina
- Carecen de motivación e interés para realizar su labor educativa.
- Nivel socio-económico bajo o sin opción económica.
- Ausentismo a clases.
- Problemas de salud.
- Proviene de ambientes familiares y sociales violentos.
- Desmotivación hacia la carrera y a la universidad.

2.2.4 Metodología CRISP-DM

Se utilizó esta metodología debido a sus 6 fases que se acoplan a este tipo de problema. Se hizo un estudio exhaustivo sobre qué tipo de metodología y técnica son los más acertados a la predicción de la deserción estudiantil, encontrándonos con la metodología Crisp-dm y la técnica de árbol de decisión

Según (Espinosa Zúñiga 2020) las fases de la metodología CRISP – DM son las siguientes:

1. Comprensión del problema o negocio: Esta es la etapa más importante, ya que, si no se tiene una correcta comprensión del problema, o negocio, de nada servirán las etapas siguientes.

- Identificación del problema: Consiste en entender y delimitar la problemática, así como identificar los requisitos, supuestos, restricciones y beneficios del proyecto.
- Determinación de objetivos: Puntualiza las metas a lograr al proponer una solución basada en un modelo de minería de datos. En el presente estudio el objetivo es obtener un patrón de comportamiento de las entidades de la República Mexicana, de acuerdo con las unidades económicas ubicadas en cada entidad.
- Evaluación de la situación actual: Especifica el estado actual antes de implementar la solución de minería de datos propuesta, a fin de tener un punto de comparación que permita medir el grado de éxito del proyecto.

2. Comprensión de los datos: Se selecciona y adapta los datos, para poder identificar los problemas de calidad de datos y así obtener datos potenciales para poder analizar.

- Recolección de datos: Consiste en obtener los datos a utilizar en el proyecto identificando las fuentes, las técnicas empleadas en su recolección, los problemas encontrados en su obtención y la forma como se resolvieron los mismos.
- Descripción de datos: Identifica el tipo, formato, volumetría y significado de cada dato.
- Exploración de datos: Radica en aplicar pruebas estadísticas básicas que permitan conocer las

propiedades de los datos a fin de entenderlos lo mejor posible.

3. Preparación de datos: Generalmente esta es la etapa que consume más tiempo en el proyecto, y es donde se seleccionan los datos que se transforman de acuerdo con los resultados de la etapa anterior a fin de utilizarlos en la etapa de modelado.

- Limpieza de datos: Aplicación de diferentes técnicas, por ejemplo, normalización de datos, discretización de campos numéricos, tratamiento de valores ausentes, tratamiento de duplicados e imputación de datos.
- Creación de indicadores: Genera indicadores que potencien la capacidad predictiva de los datos a partir de los datos existentes y ayudan a detectar comportamientos interesantes para modela
- Transformación de datos: Cambia el formato o estructura de ciertos datos sin modificar su significado, a fin de aplicarles alguna técnica particular en la etapa de modelado.

Item	Periodo	CICLO						Total
		2	3	4	5	6	7	
1	2020-A					10	12	22
2	2020-B	5					3	8
3	2021-A						1	1
4	2021-B			10	3	2	2	17
5	2022-A	10	1		1	1	3	16
6	2022-B			1		1	2	4
7	2023-A	15		10			1	26
8	2023-B				3	1	2	6
								100

Tabla 4: Comprensión e identificación de los datos

Fuente: elaboración propia

4. Modelado: Se selecciona la técnica a utilizar para construir el modelo.

- Selección de técnica de modelado: Elige la técnica apropiada de acuerdo con el problema a resolver, los datos disponibles, las herramientas de minería de datos disponibles, así como el dominio de la técnica elegida.
- Selección de datos de prueba: En algunos tipos de modelos se requiere dividir la muestra en datos de entrenamiento y de validación.
- Obtención del modelo: Genera el mejor modelo mediante un proceso iterativo de modificación de parámetros del mismo.

5. Evaluación del modelo: En esta etapa se determina la calidad del modelo con base en el análisis de ciertas métricas estadísticas del mismo, comparando los resultados con resultados previos.

6. Implementación del modelo: Esta etapa explota, mediante acciones concretas, el conocimiento adquirido mediante el modelo. Aquí también es importante documentar los resultados de manera clara para el usuario final y asegurarse de que todas las etapas de la metodología se documenten debidamente para hacer una revisión del proyecto a fin de obtener lecciones aprendidas durante el proceso.

2.3 Marco conceptual

Conceptualización de modelo

Según (Alban 2019) son representaciones idealizadas de la estructura real y su medio ambiente, está fuertemente influenciada por los objetivos de la computación. Los modelos representan a la

realidad sólo si se toman en cuenta todos los factores que afectan a las conclusiones de ellos.

Conceptualización de predicción

Según (Alban 2019) predicción / pronóstico de un modelo se refiere a su capacidad para predecir la mejor respuesta o salida más cercana a la realidad, sobre la base de los datos de entrada.

2.4 Definición de términos básicos

Método: Modo ordenado y sistemático de proceder para lograr un fin/conjunto de reglas.

Metodología: Conjunto de métodos que se siguen en una disciplina científica/ciencia del método y de la sistematización científica.

Predicción: Es la acción de aquello que supuestamente va ocurrir. Donde se puede predecir partiendo de conocimientos científicos, revelaciones o de algún tipo de indicios.

Deserción Estudiantil: Refieren que la deserción institucional se origina siempre y cuando aquellos estudiantes irrumpen su asistencia a la universidad por varias semanas.

Minería de Datos: La minería de datos es un proceso automático en el que se combinan descubrimiento y análisis. El proceso consiste en extraer patrones en forma de reglas o funciones, a partir de los datos, para que el usuario los analice. Esta tarea implica generalmente preprocesar los datos, hacer minería de datos y evaluar e interpretar los resultados.

Python: Python es un versátil lenguaje de programación de propósito general que ocupa un lugar destacado entre los ocho idiomas de programación más utilizados globalmente. Su utilidad abarca la creación de diversas aplicaciones, desde aplicaciones web hasta aplicaciones de escritorio, incluyendo intérpretes de scripts, entre muchas otras posibilidades.

Librería pandas y scikit-learn: Pandas se destaca por brindarnos una solución eficaz para cargar y manipular los datos, mientras que, por su parte, scikit-learn nos proporciona las herramientas esenciales para construir y evaluar el modelo de árbol de decisiones.

III. HIPOTESIS Y VARIABLES

El presente trabajo de investigación es de tipo cuantitativo, esto significa que nuestro análisis se ha basado en el tratamiento y procesamiento de los datos asociados a las variables de estudio que se han establecido en los supuestos de la investigación.

3.1 Hipótesis

En este apartado se presentan las hipótesis de trabajo que se formularon para guiar el desarrollo del proyecto de investigación.

3.1.1 Hipótesis General

Existe una relación significativa entre los factores identificados a través de la minería de datos y la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao.

3.1.2 Hipótesis Específicos

HE1: La minería de datos comprueba que el factor social influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

HE2: La minería de datos comprueba que el factor económico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

HE3: La minería de datos comprueba que el factor académico influye en la deserción estudiantil universitaria en la Escuela

Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

Definición conceptual de variables

Variable Dependiente: Deserción Estudiantil

La deserción estudiantil es uno de los fenómenos sociales y educativos que se observa en las diversas sociedades, en menor o mayor grado, ante ello, se han realizado diversos estudios que intentan responder qué factores influyen para que el estudiante deje de asistir o formar parte de una institución educativa (Allpas Ponce 2018).

Variable Independiente: Minería de Datos

El datamining (minería de datos), es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto (Sinnexus 2023).

Operacionalización de variables

Variable dependiente: deserción estudiantil

El artículo de Tinto (2012) nos proporciona una base para abordar la deserción universitaria y nos brinda un marco conceptual para la operacionalización. A continuación, se describe la operacionalización de la variable en nuestro estudio siguiendo las pautas del autor.

Definición de Deserción Universitaria (Tinto, 2012): De acuerdo con la definición propuesta por Tinto, la deserción universitaria es considerada como la interrupción de los estudios durante dos semestres académicos consecutivos, lo que se comprende como la no inscripción en ningún curso durante ese período. Esta definición nos establece un marco para determinar cuándo un estudiante se considerará desertor en nuestra investigación.

Indicador Temporal (Angulo, Pergelova 2013): Siguiendo el enfoque de Tinto, se registrará el estado de matrícula de los estudiantes al final de cada semestre académico, identificando si el estudiante se encuentra inscrito en cursos o si ha interrumpido su matrícula durante dos semestres consecutivos.

Causas de Deserción (Angulo, Pergelova 2013): Al igual que en el artículo de Tinto, recopilaremos datos sobre las razones de la deserción mediante encuestas o entrevistas con los estudiantes desertores. Esto nos permitirá identificar motivos como problemas académicos, personales, económicos y otras causas que pueden estar relacionadas con la deserción universitaria.

Características del Estudiante (Angulo, Pergelova 2013): Seguiremos la recomendación de Tinto al registrar datos demográficos y académicos de los estudiantes, como edad, género, GPA, situación económica y su historial académico. Estas variables nos ayudarán a comprender mejor quiénes son los estudiantes en riesgo de deserción.

VARIABLES CONTEXTUALES (Angulo, Pergelova 2013): Al igual que en el artículo de Tinto, recopilaremos datos sobre factores contextuales, como el acceso a servicios de apoyo académico, el nivel de participación en actividades extracurriculares, el acceso a asesoramiento académico y otros servicios institucionales. Estos datos nos permitirán evaluar el impacto de las variables contextuales en la deserción.

3.1.1 Operacionalización de la variable

Variables	Definición	Dimensiones	Indicadores	Técnica e instrumento de recolección	Ítems
Deserción Estudiantil	Es considerada como la interrupción de los estudios durante dos semestres académicos consecutivos, lo que se comprende como	Sociales	<ul style="list-style-type: none"> ● Apoyo familiar ● Motivación externa ● Autopercepción 	Cuestionario validado por juicio de expertos	Pregunta 1 Pregunta 2 Pregunta 3 Pregunta 4 Pregunta 5 Pregunta 6 Pregunta 7 Pregunta 8 Pregunta 9 Pregunta 10 Pregunta 11 Pregunta 12

	la no inscripción en ningún curso durante ese período	Económicas	<ul style="list-style-type: none"> ● Situación Laboral ● Financiación de los estudios ● Economía familiar ● Dependencia económica 		Pregunta 13 Pregunta 14 Pregunta 15 Pregunta 16 Pregunta 17 Pregunta 18 Pregunta 19 Pregunta 20 Pregunta 21
		Académicas	<ul style="list-style-type: none"> ● Rendimiento ● Desempeño 		Pregunta 22 Pregunta 23 Pregunta 24 Pregunta 25 Pregunta 26 Pregunta 27

Tabla 5: Operacionalización de variables

Fuente: elaboración propia

IV. DISEÑO METODOLÓGICO

4.1. Tipo y diseño de investigación.

4.1.1 Diseño de investigación

El diseño de la investigación se configuró como un estudio predictivo mediante el uso de técnicas de minería de datos. Adoptando un enfoque transeccional, se buscó establecer correlaciones causales para describir la relación entre la deserción estudiantil y las variables identificadas a través de la minería de datos (Alania 2018).

4.1.2 Tipo de investigación

El tipo de investigación será aplicada, puesto que la finalidad del estudio consiste en encontrar una solución al problema planteado, se busca transformar o producir cambios en un determinado evento o acontecimiento que sea de preocupación para el investigador (Palomino Orizano, y otros, 2015).

Es de enfoque cuantitativo, ya que, tal como indica el libro de (Rodríguez Araínga, 2011) se utilizará la recolección de datos para probar la hipótesis sobre la base de la medición y el análisis estadístico para la prueba de teorías.

Investigación de tipo no experimental, transversal y correlacional; (Fernandez, Baptista 2014) las investigaciones no experimentales son aquellas

que, durante el proceso de su investigación, el investigador solo se limita a mirar como los factores o elementos se van desarrollando sin ningún agente externo que pueda

alterar o adulterar el proceso natural de relación entre ambos.

Así mismo esta investigación puede ser también transversal o transversales, porque junta la información de un único tiempo

o circunstancia y tiene como finalidad detallar, investigar o mencionar como es la relación entre los factores en ese único momento.

4.2. Método de investigación.

El método de investigación en el presente trabajo es el método deductivo ya que se recolectan datos de estudiantes para llegar a una hipótesis o teoría sobre la deserción de su carrera profesional.

Como indica (Viñán 2015), el método deductivo nos ayudará a conocer los factores predominantes y a través de ellos las universidades puedan encontrar la solución para la mejora.

4.3. Población y muestra.

La población es el conjunto de valores que cada variable toma en las unidades que conforman el universo. Por ello, se puede decir, cuando el universo tiene N elementos, que la población estadística es de tamaño Z (Escalante, Valderrama, Vásquez 2023).

4.3.1 Población

La población estuvo constituida por los estudiantes matriculados en los semestres correspondientes entre los períodos de 2021 a 2023 de la facultad de Ingeniería Industrial y de Sistemas de la Universidad Nacional del Callao.

4.3.2 Muestra

La muestra utilizada fue no aleatoria, por muestreo de conveniencia o selección intencional, conformada por 100 estudiantes identificados en los semestres correspondientes entre los períodos de 2021 a 2023 de la facultad de Ingeniería Industrial y de Sistemas de la Universidad Nacional del Callao.

4.4. Lugar de estudio.

El lugar de estudio fue la Escuela de Ingeniería Sistemas de la Facultad de Ingeniería Industrial y de Sistemas de la Universidad Nacional del Callao.

4.5. Técnicas e instrumentos para la recolección de la información.

4.5.1 Técnica

Se utilizó información histórica de los alumnos matriculados obtenida de las bases de datos de la Escuela de Ingeniería de Sistemas de la Facultad de Ingeniería Industrial y de Sistemas de la Universidad Nacional del Callao.

Análisis de documentos

Consiste en el estudio exhaustivo de documentos, artículos, revistas, páginas web referente a las variables de la investigación los cuales presentan una serie de teorías, técnicas, métodos que pretenden dar solución al problema que se estudia en la investigación.

4.5.2 Instrumentos

Se utilizó una encuesta a 50 estudiantes de la escuela profesional de Ingeniería de Sistemas para la recolección de datos de los posibles desertores.

V. RESULTADOS

5.1. Resultados descriptivos

¿Con qué frecuencia su familia lo apoya para continuar sus estudios universitarios?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Nada satisfecho	18	36,0	36,0	36,0
	Poco satisfecho	26	52,0	52,0	88,0
	Regular	3	6,0	6,0	94,0
	Satisfecho	3	6,0	6,0	100,0
	Total	50	100,0	100,0	

Tabla 6: ¿Con qué frecuencia su familia lo apoya para continuar sus estudios universitarios?

Fuente: elaboración propia

La ilustración N°5 muestra los datos señalados:

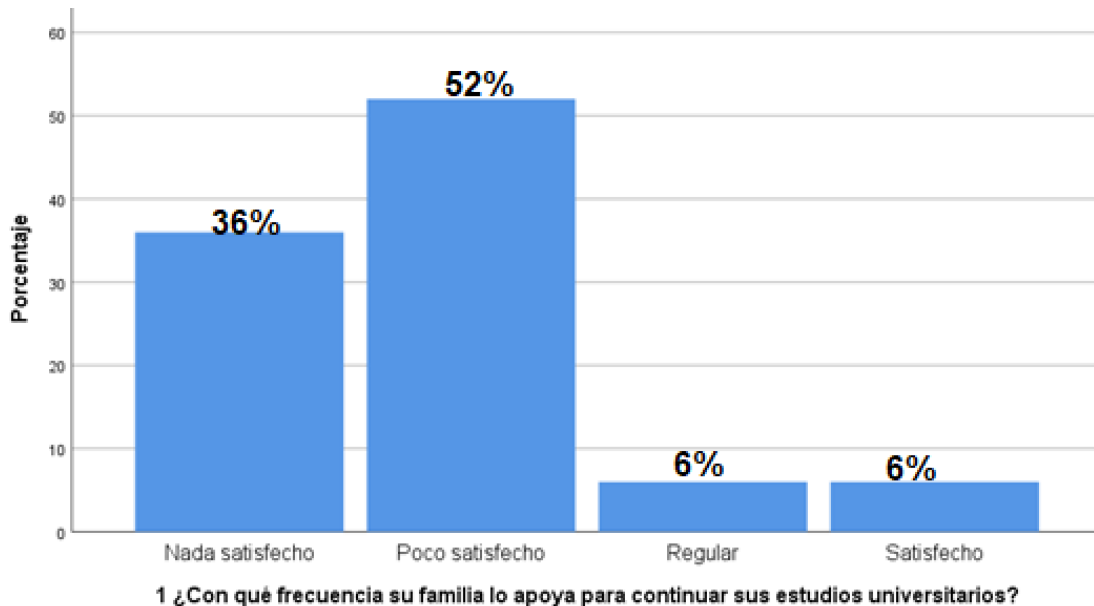


Ilustración 5: ¿Con qué frecuencia su familia lo apoya para continuar sus estudios universitarios?

Fuente: elaboración propia

En la tabla N°6 se puede apreciar que más del 50% está poco satisfecho con la frecuencia que la familia apoya sus estudios, y el 36% está nada satisfecho, siendo un 6% regular y satisfecho.

¿Con qué frecuencia recibe apoyo de su familia para sus estudios universitarios?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Nada satisfecho	21	42,0	42,0	42,0
	Poco satisfecho	23	46,0	46,0	88,0
	Regular	2	4,0	4,0	92,0
	Satisfecho	2	4,0	4,0	96,0
	Muy satisfecho	2	4,0	4,0	100,0
	Total	50	100,0	100,0	

Tabla 7: ¿Con qué frecuencia recibe apoyo de su familia para sus estudios universitarios?

Fuente: elaboración propia

La ilustración N°6 muestra los datos señalados:

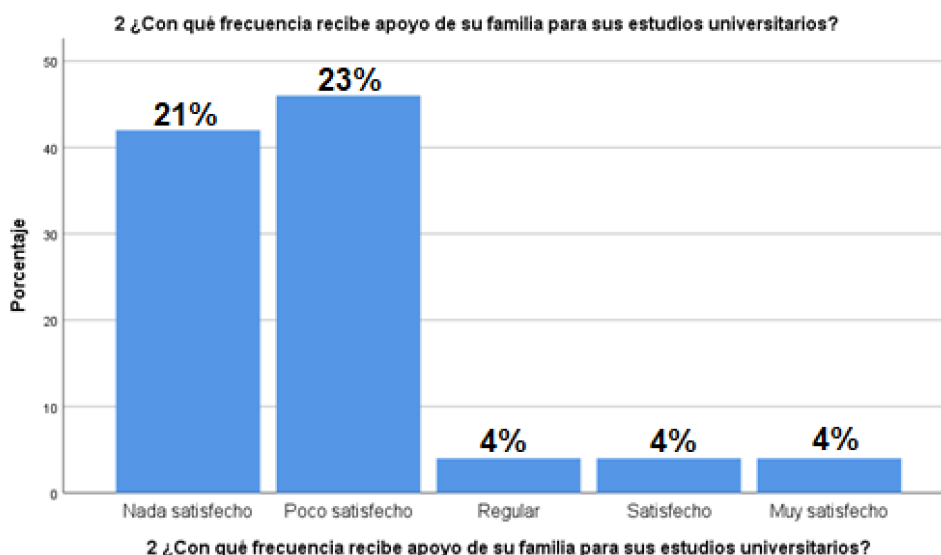


Ilustración 6: ¿Con qué frecuencia recibe apoyo de su familia para sus estudios universitarios?

Fuente: elaboración propia

En la tabla N°7 se puede apreciar que el 42% no está nada satisfecho con la frecuencia en que recibe apoyo de su familia para sus estudios universitario y 46% poco satisfecho, además como dato especial, las respuestas, regular, satisfecho y muy satisfecho están al 4%.

¿Cómo calificaría el nivel de apoyo que recibe de su familia para sus estudios universitarios?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Muy bajo	26	52,0	52,0	52,0
	Bajo	18	36,0	36,0	88,0
	Regular	2	4,0	4,0	92,0
	Alto	2	4,0	4,0	96,0
	Muy alto	2	4,0	4,0	100,0
	Total	50	100,0	100,0	

Tabla 8: ¿Cómo calificaría el nivel de apoyo que recibe de su familia para sus estudios universitarios?

Fuente: elaboración propia

La ilustración N°7 muestra los datos señalados:

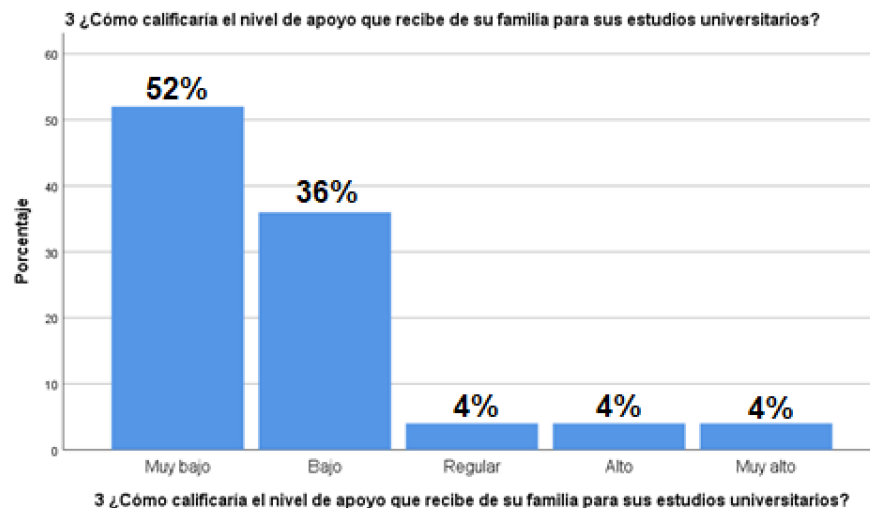


Ilustración 7: ¿Cómo calificaría el nivel de apoyo que recibe de su familia para sus estudios universitarios?

Fuente: elaboración propia

En la tabla N°8 se puede apreciar que el 52% califica como muy bajo el nivel de apoyo que recibe de su familia para sus estudios universitarios, mientras que el 36% califica como bajo. Además, al nivel de 4% lo califican como regular, alto y muy alto.

¿En qué escala la familia lo motivan a continuar sus estudios universitarios?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Muy bajo	15	30,0	30,0	30,0
	Bajo	23	46,0	46,0	76,0
	Regular	8	16,0	16,0	92,0
	Alto	2	4,0	4,0	96,0
	Muy alto	2	4,0	4,0	100,0
	Total	50	100,0	100,0	

Tabla 9: ¿En qué escala la familia lo motivan a continuar sus estudios universitarios?

Fuente: elaboración propia

La ilustración N°8 muestra los datos señalados

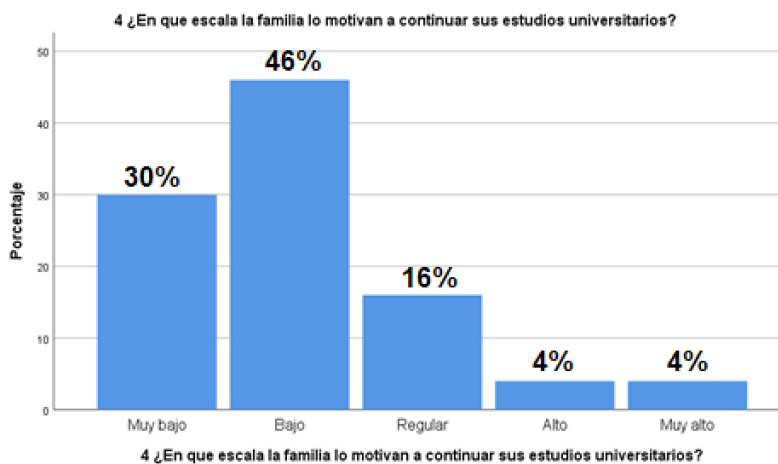


Ilustración 8: ¿En qué escala la familia lo motivan a continuar sus estudios universitarios?

Fuente: elaboración propia

En la tabla N°9 se puede apreciar que el 46% considera como baja la motivación que reciben de su familia a continuar con sus estudios universitarios, mientras que el 30% lo considera como muy bajo. Además, se aprecia que el 16% lo considera la motivación como regular y, como dato adicional, las respuestas de alto y muy alto están al mismo nivel con 4%.

Post Test

¿Con qué frecuencia su familia lo apoya para continuar sus estudios universitarios?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Regular	17	34,0	34,0	34,0
	Satisfecho	11	22,0	22,0	56,0
	Muy satisfecho	22	44,0	44,0	100,0
	Total	50	100,0	100,0	

Tabla 10: ¿Con qué frecuencia su familia lo apoya para continuar sus estudios universitarios?

Fuente: elaboración propia

La ilustración N°9 muestra los datos señalados:

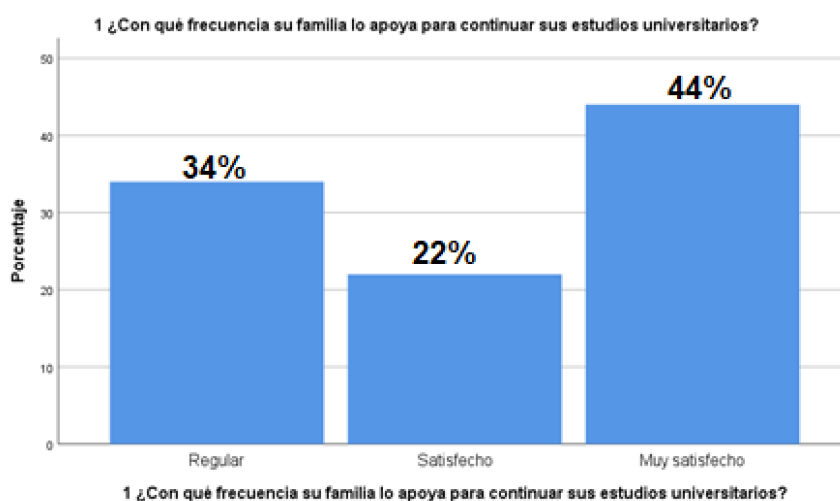


Ilustración 9: ¿Con qué frecuencia su familia lo apoya para continuar sus estudios universitarios?

Fuente: elaboración propia

En la tabla N°10 se puede apreciar que el 44% se encuentra muy satisfecho con el apoyo de su familia para continuar con sus estudios universitarios, mientras que un 34% lo considera como regular y el 22% se encuentra satisfecho.

¿Con qué frecuencia recibe apoyo de su familia para sus estudios universitarios?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Poco satisfecho	2	4,0	4,0	4,0
	Regular	14	28,0	28,0	32,0
	Satisfecho	16	32,0	32,0	64,0
	Muy satisfecho	18	36,0	36,0	100,0
	Total	50	100,0	100,0	

Tabla 11: ¿Con qué frecuencia recibe apoyo de su familia para sus estudios universitarios?

Fuente: elaboración propia

La ilustración N°10 muestra los datos señalados:

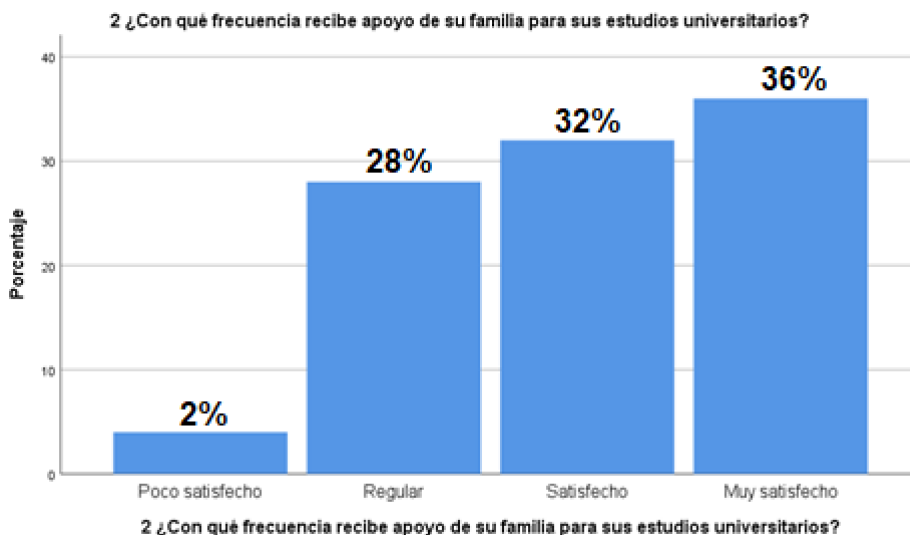


Ilustración 10: ¿Con qué frecuencia recibe apoyo de su familia para sus estudios universitarios?

Fuente: elaboración propia

En la tabla N°11 se puede apreciar 36% se encuentra muy satisfecho de recibir apoyo de su familia para sus estudios universitarios, un 32% se encuentra satisfecho, un 28% considera regular el apoyo que recibe de su familia. Además, un 4% considera estar poco satisfecho del apoyo que recibe de su familia.

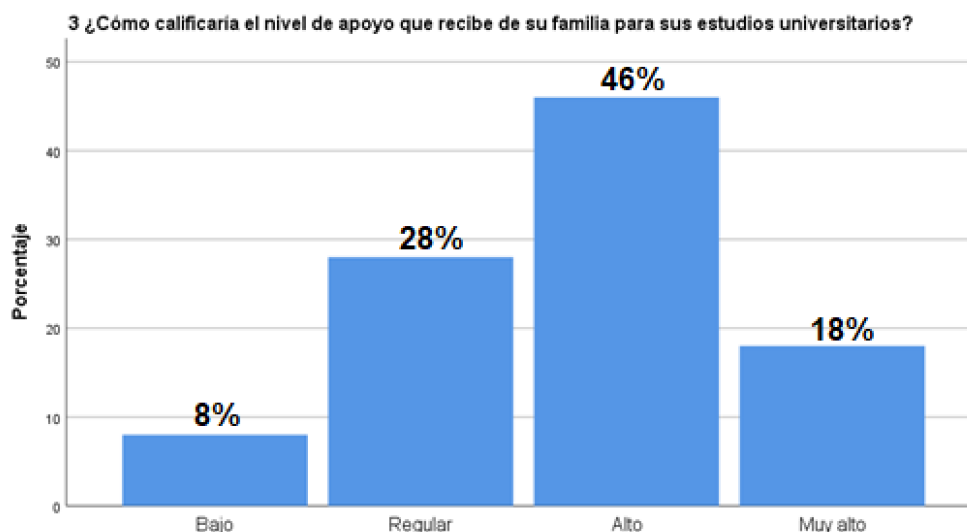
¿Cómo calificaría el nivel de apoyo que recibe de su familia para sus estudios universitarios?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Bajo	4	8,0	8,0	8,0
	Regular	14	28,0	28,0	36,0
	Alto	23	46,0	46,0	82,0
	Muy alto	9	18,0	18,0	100,0
	Total	50	100,0	100,0	

Tabla 12: ¿Cómo calificaría el nivel de apoyo que recibe de su familia para sus estudios universitarios?

Fuente: elaboración propia

La ilustración N°11 muestra los datos señalados:



3 ¿Cómo calificaría el nivel de apoyo que recibe de su familia para sus estudios universitarios?

Ilustración 11: ¿Cómo calificaría el nivel de apoyo que recibe de su familia para sus estudios universitarios?

Fuente: elaboración propia

En la tabla N°12 se puede apreciar que un 46% considera alto el nivel de apoyo que recibe de su familia para sus estudios universitarios, un 28% considera como regular, mientras que en la calificación de muy alto y bajo tenemos un 18% y 8% respectivamente.

¿En qué escala la familia lo motivan a continuar sus estudios universitarios?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Bajo	2	4,0	4,0	4,0
	Regular	16	32,0	32,0	36,0
	Alto	15	30,0	30,0	66,0
	Muy alto	17	34,0	34,0	100,0
	Total	50	100,0	100,0	

Tabla 13: ¿En qué escala la familia lo motivan a continuar sus estudios universitarios?

Fuente: elaboración propia

La ilustración N°12 muestra los datos señalados:

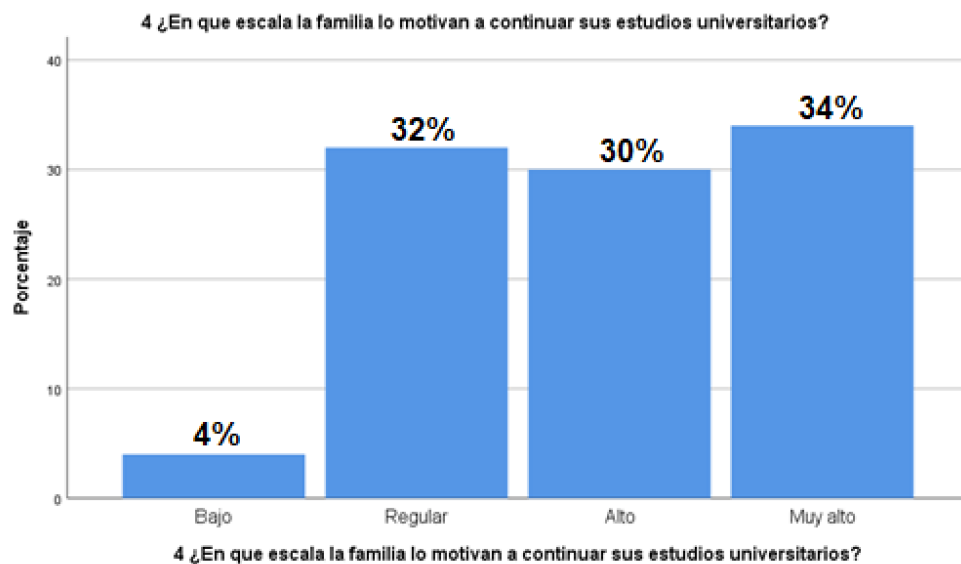


Ilustración 12: ¿En qué escala la familia lo motivan a continuar sus estudios universitarios?

Fuente: elaboración propia

En la tabla N°13 se puede apreciar que un 34% considera que la familia lo motiva a continuar con sus estudios universitarios, un 32% lo considera como regular y un 30% lo considera como alto. Mientras tanto un 4% considera baja la motivación que recibe.

Resultados Inferenciales

Usamos R Pearson para hallar el grado de relación entre las variables.

Recordemos que el R de Pearson es un número que se encuentra entre -1 y +1 por lo que tendrá dos interpretaciones



Ilustración 13: R Pearson

Fuente: elaboración propia

1. Según el signo

- a. Si el R de Pearson es positivo se interpreta que las variables son directamente proporcionales, o sea, si una aumenta la otra también lo hará, y viceversa.
- b. Si el R de Pearson es negativo, se interpreta que las variables son inversamente proporcionales, si una aumenta, la otra disminuye.

2. Según la proximidad a 1

- a. si el R de Pearson calculado se aproxima a +1 ó -1 quiere decir que la relación entre las variables es estrecha o fuerte.
- b. si el R de Pearson calculado se aproxima a 0, quiere decir que la relación entre las variables es débil.

HIPÓTESIS GENERAL

Existe una relación significativa entre los factores identificados a través de la minería de datos y la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao.

Resumen del modelo				
Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación
1	,701 ^a	,491	,40	3,08554

a. Predictores: (Constante), detección de deserción

Tabla 14: Resumen del modelo para la hipótesis general

Fuente: elaboración propia

Interpretación:

1. Como el R de Pearson 0.701 es positivo se interpreta que las variables son directamente proporcionales. A mayor implementación de minería de datos, mayor detección de deserción estudiantil.
2. Como el R de Pearson calculado 0.701 se aproxima a 1 quiere decir que la relación entre las variables es fuerte

HIPÓTESIS ESPECÍFICA 1

La minería de datos comprueba que el factor social influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

Resumen del modelo				
Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación
1	,693 ^a	,480	,401	4,021

a. Predictores: (Constante), factor social

Tabla 15: Resumen del modelo para la hipótesis específica 1

Fuente: elaboración propia

Interpretación:

1. Como el R de Pearson 0.693 es positivo se interpreta que las variables son directamente proporcionales, o sea, si se implementa la minería de datos se comprueba el factor social.
2. Como el R de Pearson calculado 0.693 se aproxima a 1 quiere decir que la relación entre las variables es fuerte.

HIPÓTESIS ESPECÍFICA 2

La minería de datos comprueba que el factor económico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

Resumen del modelo				
Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación
1	,566a	,320	,307	2,9857

a. Predictores: (Constante), factor económico

Tabla 16: Resumen del modelo para la hipótesis específica 2

Fuente: elaboración propia

Interpretación:

1. Como el R de Pearson 0.566 es positivo se interpreta que las variables son directamente proporcionales, o sea, si se implementa la minería de datos se comprueba el factor económico.
2. Como el R de Pearson calculado 0.566 se aproxima a 1 quiere decir que la relación entre las variables es fuerte.

HIPÓTESIS ESPECÍFICA 3

La minería de datos comprueba que el factor académico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

Resumen del modelo				
Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación
1	,771 ^a	,594	,507	3,2274

a. Predictores: (Constante), factor académico

Tabla 17: Resumen del modelo para la hipótesis específica 3

Fuente: elaboración propia

Interpretación:

1. Como el R de Pearson 0.566 es positivo se interpreta que las variables son directamente proporcionales, o sea, si se implementa la minería de datos se comprueba el factor económico.
2. Como el R de Pearson calculado 0.566 se aproxima a 1 quiere decir que la relación entre las variables es fuerte.

5.2. Resultados inferenciales

Resultados Inferenciales

En el contexto de la deserción estudiantil en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao, la aplicación de CRISP-DM nos ha proporcionado valiosos insights sobre la efectividad de los modelos predictivos desarrollados.

Estos modelos, basamos en minería de datos, se centraron en las dimensiones social, económica y académica para predecir la probabilidad de deserción estudiantil.

Fases del modelo de referencia CRISP-DM

1. Fase de Comprensión del negocio (Business Understanding)

Comprender los objetivos del negocio: La Universidad Nacional del Callao pertenece a una de las universidades nacionales de renombre que tiene como objetivo el de formar profesionales, generando y promoviendo la investigación científica, tecnológica y humanística, en los estudiantes universitarios con calidad, competitividad y responsabilidad social para el desarrollo sostenible del país

2. Fase de comprensión de los datos (Data Understanding):

Recolección de los datos: Para la recolección de los datos se realizaron encuestas validadas por expertos para recopilar información de 186 participantes, de los cuales 50 eran desertores y 136 eran no desertores.

3. Fase de preparación de los datos (Data preparation):

Limpieza de datos: Se realizaron procesos de limpieza para poder abordar datos faltantes o inconsistentes en las encuestas.

Transformación de los datos: Se asignaron etiquetas a los datos, agruparon por dimensiones (social, económica y académica) y, además, se utilizaron algoritmos en Python para preparar los conjuntos de datos y prueba.

4. Fase del modelado (Data Modelling)

Selección de modelos: Se implementó un modelo de clasificación utilizando Random Forest para cada dimensión (social, económica y académica) y, además, el totalizado de las mismas.

Entrenamiento del modelo: Se entrenaron los modelos con los datos de las encuestas y se evaluaron utilizando métricas como precisión, matriz de confusión y métricas específicas para cada dimensión.

5. Fase de evaluación (Deployment)

Resultados del modelo:

- **Dimensión Social:**

Accuracy: La precisión es del 100%, lo que indica que el modelo predijo correctamente todos los casos (24 no desertores y 4 desertores).

Matriz de Confusión:

		Predicción			
		NO		SI	
Realidad	NO	VN	24	FP	0
	SI	FN	0	VP	4

Tabla 18: Matriz de confusión para la dimensión social

Fuente: elaboración propia

La matriz de confusión muestra que no hubo falsos positivos ni falsos negativos.

Reporte de Clasificación

	precision	recall	F1-score	support
0	1.00	1.00	1.00	24
1	1.00	1.00	1.00	4
accuracy			1.00	28
macro avg	1.00	1.00	1.00	28
weighted avg	1.00	1.00	1.00	28

Tabla 19: Reporte de Clasificación para la dimensión social

Fuente: elaboración propia

La métrica F1-score es 1.00, lo que indica un rendimiento perfecto.

Importancia de características:

Pregunta	% importancia
pregunta 1	0.5068
pregunta 2	0.1315
pregunta 3	0.0842
pregunta 4	0.0407
pregunta 5	0.0182
pregunta 6	0.021
pregunta 7	0.0316
pregunta 8	0.0217
pregunta 9	0.0527
pregunta 10	0.0357
pregunta 11	0.0335
pregunta 12	0.0224

Tabla 20: Importancia de características para la dimensión social

Fuente: elaboración propia

Las preguntas 1, 2 y 9 son las más influyentes según el modelo.

Resumen de resultados del modelo para la dimensión SOCIAL:

El modelo aplicado a la dimensión social ha demostrado un rendimiento excepcional, logrando una precisión del 100%. La matriz de confusión reveló que no se produjeron falsos positivos ni falsos negativos. La métrica F1-score alcanzó el valor máximo de 1.00, indicando un equilibrio perfecto entre precisión y recall. Las características más influyentes identificadas por el modelo se centran en las preguntas 1, 2 y 9 de la encuesta. En resumen, el modelo predice con confianza un 13.24% de potenciales desertores en esta dimensión.

- Dimensión Económica:

Accuracy: La precisión es del 82.14%, lo que indica un buen rendimiento, pero con algunas predicciones incorrectas.

Matriz de Confusión:

		Predicción			
		NO		SI	
Realidad	NO	VN	21	FP	3
	SI	FN	2	VP	2

Tabla 21: Matriz de confusión para la dimensión Económica

Fuente: elaboración propia

Hay algunos falsos positivos y falsos negativos.

Reporte de Clasificación

	precision	recall	fi-score	support
0	0.91	0.88	0.89	24
1	0.40	0.50	0.44	4
accuracy			0.82	28
macro avg	0.66	0.69	0.67	28
weighted avg	0.84	0.82	0.83	28

Tabla 22: Reporte de clasificación para la dimensión económica

Fuente: elaboración propia

La métrica F1-score es más baja en comparación con la dimensión social.

Importancia de características:

Pregunta	% importancia
pregunta 13	0.1004
pregunta 14	0.0856
pregunta 15	0.0856
pregunta 16	0.0706
pregunta 17	0.116
pregunta 18	0.187
pregunta 19	0.1802
pregunta 20	0.0847
pregunta 21	0.0899

Tabla 23: Importancia de las características para la dimensión económica

Fuente: elaboración propia

Las preguntas 18, 19 y 20 son las más influyentes según el modelo.

Resumen de resultados del modelo para la dimensión ECONÓMICA:

El rendimiento del modelo en la dimensión económica es sólido, con una precisión del 82.14%. Sin embargo, la matriz de confusión revela la presencia de algunos falsos positivos y falsos negativos. La métrica F1-score alcanzó 0.83, destacando una eficacia considerable. Las preguntas 18, 19 y 20 se identificaron como las más influyentes en esta dimensión. El modelo predice un 14.71% de potenciales desertores en el aspecto económico.

- Dimensión Académica:

Accuracy: La precisión es del 75%, indicando un rendimiento aceptable, pero con margen de mejora.

Matriz de Confusión:

		Predicción			
		NO		SI	
Realidad	NO	VN	19	FP	5
	SI	FN	2	VP	2

Tabla 24: Matriz de confusión para la dimensión académica

Fuente: elaboración propia

Hay algunos falsos positivos y falsos negativos.

Reporte de Clasificación

	precision	recall	fi-score	support
0	0.90	0.79	0.84	24
1	0.29	0.50	0.36	4
accuracy			0.82	28
macro avg	0.60	0.65	0.60	28
weighted avg	0.82	0.75	0.78	28

Tabla 25: Reporte de Clasificación para la dimensión académica

Fuente: elaboración propia

La métrica F1-score es más baja en comparación con las dos dimensiones anteriores.

Importancia de características:

Pregunta	% importancia
pregunta 22	0.1383
pregunta 23	0.2389
pregunta 24	0.1347
pregunta 25	0.1521
pregunta 26	0.142
pregunta 27	0.1939

Tabla 26: Importancia de características para la dimensión académica

Fuente: elaboración propia

La pregunta 23 es la más influyente según el modelo.

Resumen de resultados del modelo para la dimensión ACADÉMICA:

La dimensión académica presenta un rendimiento aceptable, con una precisión del 75%. La matriz de confusión indica la presencia de algunos errores de predicción, pero la métrica F1-score alcanza 0.78, señalando un rendimiento sólido. La pregunta 23 se destaca como la característica más influyente. El modelo prevé un 13.97% de potenciales desertores en esta dimensión.

RESUMEN TOTALIZADO

Potenciales Desertores según el modelo en todas las dimensiones: 57 (41.91%)

Matriz de Confusión Totalizada

		Predicción			
		NO		SI	
Realidad	NO	VN	64	FP	8
	SI	FN	4	VP	8

Tabla 27: Matriz de confusión totalizada

Fuente: elaboración propia

El modelo identificó correctamente a 8 desertores y 64 no desertores, pero cometió errores en 8 casos.

En resumen, considerando todas las dimensiones, el modelo predice un 41.91% de potenciales desertores en la población estudiantil. La matriz de confusión totalizada muestra 8 verdaderos positivos, 64 verdaderos negativos, 8 falsos positivos y 4 falsos negativos. Estos resultados indican que el modelo identificó correctamente a la mayoría de los no desertores, pero existe margen de mejora en la identificación de desertores potenciales.

6. Fase de Monitoreo (Monitoring):

Evaluación Continua: Se considera realizar monitoreo constante al rendimiento del modelo y considerar actualizaciones en el comportamiento de los patrones basados en nuevos datos o cambios en el entorno educativo.

VI. DISCUSIÓN DE RESULTADOS

6.1. Contrastación y demostración de la hipótesis con los resultados.

HIPÓTESIS GENERAL

Existe una relación significativa entre los factores identificados a través de la minería de datos y la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao. H_a : Hipótesis alterna (es la hipótesis que planteaste)

H_0 : No existe una relación significativa entre los factores identificados a través de la minería de datos y la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao.

H_a : Existe una relación significativa entre los factores identificados a través de la minería de datos y la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao.

$E = 0.05$ (nivel de significancia SIG) 5% (por usar muestra)

Entonces estadísticamente vamos a hallar el p (nivel de significancia o SIG)

Si $p < 0.05$, se rechaza la hipótesis nula, se acepta la hipótesis alterna

Si $p > 0.05$ se acepta la hipótesis nula, se rechaza la hipótesis alterna

Prueba de muestras emparejadas

		Diferencias emparejadas							
		Media	Desv. Desviación	Desv. Error promedio	95% de intervalo de confianza de la diferencia		t	gl	Sig. (bilateral)
					Inferior	Superior			
Par 1	pre – pos	27,16000	13,92745	1,96964	31,11814	23,20186	-13,789	49	,000

Tabla 28: Prueba de muestras emparejadas - Hipótesis General

Fuente: elaboración propia

Como $p < 0.05$, se rechaza la hipótesis nula, se considera aceptada la hipótesis alterna.

Ha: Existe una relación significativa entre los factores identificados a través de la minería de datos y la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao.

HIPÓTESIS ESPECÍFICA 1

La minería de datos comprueba que el factor social influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

Ho: La minería de datos comprueba que el factor social no influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

Ha: La minería de datos comprueba que el factor social influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

$E = 0.05$ (nivel de significancia SIG) 5% (por usar muestra)

Entonces estadísticamente vamos a hallar el p (nivel de significancia o SIG)

Si $p < 0.05$, se rechaza la hipótesis nula, se acepta la hipótesis alterna

Si $p > 0.05$ se acepta la hipótesis nula, se rechaza la hipótesis alterna

Prueba de muestras emparejadas								
	Media	Desv. Desviación	Desv. Error promedio	95% de intervalo de confianza de la diferencia		t	gl	Sig. (bilateral)
				Inferior	Superior			
RC_PRETEST - RC_POSTEST	8,530000	6,58774	1,44711	9,61142	3,967611	5,2278	49	,000

Tabla 29: Prueba de muestras emparejadas - Hipótesis Específica 1

Fuente: elaboración propia

Como $p = 0.00 < 0.05$, se rechaza la hipótesis nula, se acepta la hipótesis alterna.

Ha: La minería de datos comprueba que el factor social influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

HIPÓTESIS ESPECÍFICA 2

La minería de datos comprueba que el factor económico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

Ho: La minería de datos comprueba que el factor económico no influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

Ha: La minería de datos comprueba que el factor económico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

$E = 0.05$ (nivel de significancia SIG) 5% (por usar muestra)

Entonces estadísticamente vamos a hallar el p (nivel de significancia o SIG)

Si $p < 0.05$, se rechaza la hipótesis nula, se acepta la hipótesis alterna

Si $p > 0.05$ se acepta la hipótesis nula, se rechaza la hipótesis alterna

Prueba de muestras emparejadas

	Diferencias emparejadas					t	gl	Sig. (bilateral)
	Media	Desv. Desviación	Desv. Error promedio	95% de intervalo de confianza de la diferencia				
				Inferior	Superior			
GC_PRETEST - GC_POSTEST	,824615	6,11544	1,221442	8,654744	4,4212117	,141	49	,000

Tabla 30: Prueba de muestras emparejadas - Hipótesis Específica 2

Fuente: elaboración propia

Como $p = 0.00 < 0.05$, se rechaza la hipótesis nula, se acepta la hipótesis alterna

Ha: La minería de datos comprueba que el factor económico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

HIPÓTESIS ESPECÍFICA 3

La minería de datos comprueba que el factor académico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

Ho: La minería de datos comprueba que el factor académico no influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

Ha: La minería de datos comprueba que el factor académico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.

$E = 0.05$ (nivel de significancia SIG) 5% (por usar muestra)

Entonces estadísticamente vamos a hallar el p (nivel de significancia o SIG)

Si $p < 0.05$, se rechaza la hipótesis nula, se acepta la hipótesis alterna

Si $p > 0.05$ se acepta la hipótesis nula, se rechaza la hipótesis alterna

Prueba de muestras emparejadas								
Diferencias emparejadas								
	Media	Desv. Desviación	Desv. Error promedio	95% de intervalo de confianza de la diferencia		t	gl	Sig. (bilateral)
				Inferior	Superior			
GC_PRETEST - GC_POSTEST	8,6554	7,21222	1,01201	7,21127	4,84574	5,847	49	,000

Tabla 31: Prueba de muestras emparejadas - Hipótesis Específica 3

Fuente: elaboración propia

Como $p\ 0.00 < 0.05$, se rechaza la hipótesis nula, se acepta la hipótesis alterna.

Ha: La minería de datos comprueba que el factor académico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de dato

6.2. Contrastación de los resultados con otros estudios

similares

Contrastación con (Russo 2019)

(Russo 2019) en su investigación “Minería de datos aplicada a estrategias para minimizar el rezago académico y la deserción universitaria en carreras de informática de la UNNOBA” se centra en anticipar los casos de abandono estudiantil contribuyendo a mejorar las políticas institucionales de apoyo y retención mientras que en esta investigación se centra en evaluar la relación entre factores identificados a través de la minería de datos y la deserción estudiantil. En ambas investigaciones se comparte un interés común en abordar la deserción estudiantil en el ámbito de carreras informáticas de las instituciones universitarias. Se utilizaron cuestionarios en ambas investigaciones como fuente de recopilación de datos.

En esta investigación se utilizan pruebas de muestras emparejadas a través de herramientas de minería de datos y análisis estadísticos, mientras que (Russo 2019) integra un enfoque multimodal comparando métodos cuantitativos y cualitativos.

Nuestra investigación se basa en la muestra de 50 estudiantes a quienes se les realizó encuestas para determinar una relación significativa entre los factores identificados y la deserción estudiantil, mientras que la tesis de (Russo 2019) trabajó con una muestra de 63 estudiantes de quienes se concluye que puede anticipar posibles casos de deserción y abandono contribuyendo a mejorar políticas institucionales.

Contrastación con (Marcano, Rodríguez 2015)

Ambos estudios comparten el objetivo común de explorar los factores que influyen en la deserción estudiantil universitaria mediante la aplicación de técnicas de minería de datos.

Ambos casos utilizan técnicas de minería de datos para analizar la deserción estudiantil y tienen el objetivo de identificar factores clave que afectan la capacidad de los estudiantes para completar sus estudios.

Para la recolección de datos la investigación de (Marcano, Rodríguez 2015) se basó en la muestra de estudiantes y profesores de un período específico utilizando informes de matrículas, mientras que esta investigación se centró en recolectar datos e información de los estudiantes a través de encuestas.

Además de compartir la utilización de técnicas cuantitativas para la medición y análisis de los datos, en la implementación se evidencian ligeras diferencias. Mientras que en esta investigación se utilizó la metodología CRISP-DM y Random Forest, la tesis de (Marcano, Rodríguez 2015) se basó en la metodología CRISP-DM, árboles de decisión C4.5 y la técnica de los vecinos más cercanos.

Ambos estudios identificaron factores contribuyentes a la deserción estudiantil, tales como experiencia previa, falta de recursos económicos y falta de enfoque en los estudios. Sin embargo, se notó que las tendencias identificadas podían variar en algunos detalles.

Finalmente, se resalta la complejidad de la deserción estudiantil y destaca la importancia de considerar factores diversos en los esfuerzos por mejorar las tasas de retención.

Contrastación con (Camargo 2020)

Ambos estudios comparten el objetivo fundamental de desarrollar un modelo predictivo para anticipar la deserción estudiantil mediante el uso de técnicas de minerías de datos.

Ambos casos utilizan la minería de datos para identificar patrones y tendencias que pueden prever la probabilidad de deserción estudiantil.

Mientras que en esta investigación se hace uso de la metodología CRISP-DM, algoritmos de minería de datos logrando una precisión promedio de 85.7% además de utilizar análisis estadísticos, mientras que en la tesis de (Camargo 2020) utiliza la metodología CRISP-DM y la herramienta Random Forest, alcanzando una precisión del 84.8% para predecir la deserción estudiantil.

La combinación de estos enfoques contribuye a una comprensión más holística y efectiva de la problemática de la deserción estudiantil en entornos académicos de informática.

Contrastación con (Mamani 2019)

Ambas investigaciones comparten el objetivo de desarrollar modelos de minería de datos para abordar la problemática de la deserción estudiantil en el ámbito universitario.

En ambos casos se adopta un enfoque experimental y emplean la minería de datos para comprender y detectar patrones que ayuden a prever abordando la problemática de la deserción estudiantil utilizando como muestra estudiantes universitarios para desarrollar los modelos predictivos.

Mientras que esta investigación se basó en la recopilación de datos a través de encuestas, basándose en el modelo CRISP-DM y la minería de datos para la predicción, además de hacer uso de métodos estadísticos, la tesis de (Mamani 2019) utilizó la información obtenida a través de la oficina de actividades y servicios académicos durante un solo semestre.

Finalmente, se obtiene que el modelo predictivo aplicado en ambos casos es capaz de identificar y categorizar a los estudiantes propensos a abandonar.

Contrastación con (Piscoya 2016)

Ambas investigaciones comparten el objetivo central de utilizar técnicas de minería de datos para predecir la deserción estudiantil, con el propósito de proporcionar información precisa para la toma de decisiones en instituciones educativas.

En ambos casos se utilizan técnicas de minería de datos para desarrollar modelos predictivos que aborden el problema de la deserción estudiantil, aplicándose a estudiantes en períodos específicos y de cierta región geográfica.

En ambos casos se evidencia el uso de la metodología CRISP-DM para poder desarrollar el modelo de minería de datos centrado en la deserción estudiantil.

Contrastación con (Pando, Zarate 2020)

Ambas investigaciones comparten el objetivo de desarrollar modelos de minería de datos para identificar patrones que influyen en la deserción estudiantil en instituciones educativas específicas.

En ambos casos se hace uso de la metodología CRISP-DM para el desarrollo de un modelo de minería de datos y, además, se hace uso de herramientas estadísticas para el tratamiento de los datos.

El diseño es experimental lo que implicó un método hipotético deductivo y realización de pruebas pre y post test.

Mientras que en esta investigación se hace uso de algoritmos de árboles de decisión a través de Random Forest, la tesis de (Pando, Zarate 2020) hace uso de Árbol C&R, Árbol C5.0, Árbol AS, y Red Bayesiana, con el modelo de Árbol C5.0 demostrando mayor efectividad.

La tesis de (Pando, Zarate 2020) logra una precisión del 94.20%, mientras que en la investigación se logra una precisión promedio del 85.7%.

6.3. Ética

Esta investigación fue elaborada de acuerdo con los lineamientos y reglamentos de la Universidad Nacional del Callao.

Los datos mostrados en esta investigación fueron recogidos y procesados de una manera adecuada sin distorsionar ni adulterar, los datos están fundamentados en el instrumento aplicado al Pre-Test y Post-Test de estudio.

Se respetó a los integrantes, no se hizo ninguna discriminación, de sexo, raza o religión. Para ello se solicitó autorización de la documentación a utilizar a las personas correspondientes e involucradas en esta investigación.

VII. CONCLUSIONES

- 1- Se ha conseguido crear un modelo predictivo que presenta una clasificación y predicción más efectiva de los alumnos propensos a la deserción. Los modelos de ensamble basados en árboles de decisión demostraron ser particularmente óptimos, cumpliendo así con el objetivo principal de esta investigación.
- 2- Mediante la aplicación de técnicas avanzadas de minería de datos, se evidencia la capacidad de alcanzar una eficacia superior al 85% en la predicción de la deserción estudiantil, basada en la identificación de patrones significativos. Estos resultados destacan la notable aplicabilidad y utilidad práctica de los modelos predictivos, ofreciendo una herramienta eficaz para anticipar y abordar la problemática de la deserción estudiantil con precisión. La capacidad de superar este umbral de eficacia refuerza la confianza en la validez y relevancia de estos modelos, proporcionando así una base sólida para la toma de decisiones y la implementación de estrategias efectivas de retención estudiantil.
- 3- Se determina que la cuestión de la deserción universitaria no se debe a un solo elemento; es un fenómeno que puede explicarse mediante múltiples factores, siendo los más influyentes aquellos de índole social, económica y académica.
- 4- Se concluye con contundencia que la aplicación de la metodología CRISP-DM en el desarrollo de modelos de minería de datos ha demostrado ser un enfoque efectivo para determinar y detectar patrones esenciales, contribuyendo de manera significativa a prever y abordar la problemática crítica de la deserción estudiantil. Este enfoque metodológico ha facilitado una comprensión profunda de los factores influyentes y ha permitido la identificación precisa de patrones de deserción.

VIII. RECOMENDACIONES

- 1- Se recomienda considerar la implementación práctica de este modelo en entornos académicos. Sería beneficioso integrar los modelos de ensamble basados en árboles de decisión en los sistemas de seguimiento estudiantil, permitiendo una identificación temprana de los alumnos con riesgo de deserción. Además, se sugiere realizar un monitoreo continuo y ajustes periódicos en el modelo para garantizar su relevancia y precisión a lo largo del tiempo. La colaboración con expertos en educación y la retroalimentación constante del personal docente podrían enriquecer aún más la aplicación y efectividad del modelo predictivo en la práctica educativa.
- 2- Se recomienda integrar estos modelos de manera sistemática en los procesos institucionales de gestión educativa. Además, se sugiere la realización de talleres y capacitaciones para el personal educativo y administrativo, a fin de comprender la aplicación y la interpretación de estos modelos en el contexto educativo específico. Es crucial establecer protocolos para el monitoreo continuo de la efectividad de los modelos, permitiendo ajustes y mejoras proactivas. La colaboración con expertos en educación y análisis de datos podría contribuir a una implementación más efectiva y a la adaptación continua de estos modelos a las necesidades cambiantes del entorno educativo.
- 3- Se recomienda adoptar un enfoque integral en la implementación de estrategias para abordar este problema. Se sugiere la creación de programas de apoyo que aborden no solo las dificultades académicas, sino también aquellas relacionadas con aspectos sociales y económicos. La colaboración con expertos en psicología, trabajo social y asesoramiento académico puede enriquecer estas iniciativas. Además, se insta a la recopilación continua de datos sobre los motivos de deserción para adaptar y personalizar las estrategias según las necesidades cambiantes de los estudiantes. Un enfoque colaborativo y multidisciplinario será esencial para desarrollar intervenciones efectivas y sostenibles.

- 4- Se sugiere la capacitación del personal involucrado en la aplicación de la metodología y la incorporación de expertos en minería de datos en el proceso. Además, se insta a mantener un enfoque flexible y adaptativo, realizando evaluaciones periódicas de la metodología para asegurar su relevancia y eficacia a lo largo del tiempo. La colaboración interdisciplinaria, con la participación de educadores y analistas de datos, puede enriquecer aún más la aplicación práctica de la metodología y su capacidad para abordar la deserción estudiantil de manera efectiva.

IX. REFERENCIAS

ALANIA, Pit, 2018. Aplicación de técnicas de minería de datos para predecir la deserción estudiantil de la facultad de ingeniería de la Universidad Nacional Daniel Alcides Carrión. *Universidad Nacional “Daniel Alcides Carrión.”*

ALBAN, Mayra, 2019. *Contribuciones a la Predicción de la Deserción Universitaria a través de Minería de Datos* [online]. Retrieved from : http://cybertesis.unmsm.edu.pe/bitstream/handle/cybertesis/10776/Alban_tm.pdf?sequence=1&isAllowed=y

ALLPAS PONCE, Victoria, 2018. Factores que influyen en el nivel de deserción de los estudiantes del programa de actualización y perfeccionamiento de la Academia de la Magistratura. *Repositorio Institucional Universidad San Ignacio de Loyola* [online]. p. 33. Retrieved from : <http://repositorio.usil.edu.pe/handle/USIL/3817>

ALVARADO, Nicanor et al., 2020. Modelo para la estimación de la deserción estudiantil Awajún y Wampis empleando minería de datos. *Revista de Ciencia y Tecnología*. No. 34, pp. 45–50. DOI 10.36995/j.recyt.2020.34.006.

ANGULO, Luis and PERGELOVA, Albena, 2013. The Student Retention Puzzle Revisited: The Role of Institutional Image. *Journal of Nonprofit and Public Sector Marketing*. Vol. 25, no. 4, pp. 334–353. DOI 10.1080/10495142.2013.830545.

BELTRAN, Beatriz, 2018. MINERÍA DE DATOS. *Planetary and Space Science*. Vol. 30, no. 1. DOI 10.1016/0032-0633(82)90071-X.

BENITES, Rodolfo, 2021. La Educación Superior Universitaria en el Perú post-pandemia VF. .

CAMARGO, Anibal, 2020. MODELO PARA LA PREDICCIÓN DE LA DESERCIÓN DE ESTUDIANTES DE PREGRADO, BASADO EN

TÉCNICAS DE MINERÍA DE DATOS. *UNIVERSIDAD DE LA COSTA - CUC*. pp. 1–121.

CAMINO-HAMPSHIRE, José Carlos, CRUZ BARBOSA, Raúl and URBINA-NÁJERA, Argelia Berenice, 2020. Deserción escolar universitaria: Patrones para prevenirla aplicando minería de datos educativa. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*. Vol. 26, no. 1. DOI 10.7203/relieve.26.1.16061.

ESCALANTE, Jaime, VALDERRAMA, Carlos and VÁSQUEZ, Antenor, 2023. La deserción universitaria: un problema no resuelto en el Perú. *Hacedor - AIAPÆC*. Vol. 7, no. 1, pp. 60–72. DOI 10.26495/rch.v7i1.2421.

ESPINO, Carlos, 2017. “Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso. 26/27 [online]. Vol. I, no. Principio activo y prestación ortoprotésica, p. 67. Retrieved from : <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/59565/6/caresptimTFG0117memòria.pdf>

ESPINOSA ZÚÑIGA, Javier Jesús, 2020. Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública. *Ingeniería Investigación y Tecnología*. Vol. 21, no. 1, pp. 1–13. DOI 10.22201/fi.25940732e.2020.21n1.008.

FERNANDEZ, Carlos and BAPTISTA, Pilar, 2014. *Metodología de la Investigación*. ISBN 9781456223960.

GAMARRA, Daniel, MATOS, Rocio and YUPANQUI, Miguel, 2018. Detección de patrones de éxito en estudios universitarios de la Universidad Continental. *Apuntes de Ciencia & Sociedad*. Vol. 08, no. 01. DOI 10.18259/acs.2018005.

GARCIA, J., 2019. Implementación de un Modelo Computacional basado en Reglas de Clasificación Supervisadas para la Predicción de la Deserción Estudiantil en la Universidad Peruana Unión Filial Juliaca. . p. 12.

LUNA PUICON, Nancy Milagros et al., 2022. Factors and conditions of school dropout in Latin America. *Universidad Ciencia y Tecnología*. Vol. 26, no. 114, pp. 108–117. DOI 10.47460/uct.v26i114.595.

MAMANI, Diego, 2019. Modelo de minería de datos basados en factores asociados para la predicción de deserción estudiantil universitaria. . No. 052, pp. 1–18.

MARCANO, Yelitza and RODRÍGUEZ, Rodney, 2015. 2. Minería de datos aplicada a la deserción estudiantil. Caso: Licenciatura en Computación-Universidad del Zulia-NPF. *Revista EDUCARE - UPEL-IPB - Segunda Nueva Etapa 2.0*. Vol. 18, no. 2, pp. 31–51. DOI 10.46498/reduipb.v18i2.131.

MOINE, Juan Miguel, 2017. Minería de datos aplicada a la educación: modelo de deserción universitaria en la Universidad Tecnológica Nacional, Facultad Regional Rosario. *XIX Workshop de Investigadores en Ciencias de la Computación* [online]. pp. 301–304. Retrieved from : <http://sedici.unlp.edu.ar/handle/10915/61720>

MORALES A, 2018. Aplicación de la minería de datos a los registros académicos de los estudiantes de la Universidad Nacional Santiago Antúnez de Mayolo. [online]. Retrieved from : <http://repositorio.unasam.edu.pe/handle/UNASAM/2115>

PANDO, Aurea and ZARATE, Winny, 2020. APLICACIÓN DE UN MODELO DE MINERÍA DE DATOS PARA IDENTIFICACIÓN DE PATRONES QUE INFLUYEN EN LA DESERCIÓN ACADEMICA EN EL INSTITUTO SUPERIOR LEONARDO DAVINCI USANDO IBM SPSS MODELER Y LA METODOLOGÍA CRISP-DM. [online]. Retrieved from : <http://200.62.226.186/handle/upaorep/6663>

PÁRAMO, Gabriel and CORREA, Carlos, 2012. Deserción Estudiantil Universitaria. Conceptualización. .

PISCOYA, Luis, 2016. APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA PREDECIR LA DESERCIÓN ESTUDIANTIL EN LA

EDUCACION BASICA REGULAR EN LA REGION DE LAMBAYEQUE.
Universidad Señor de Sipán.

ROMERO, Cristobal and VENTURA, Sebastian, 2013. Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. Vol. 3, no. 1, pp. 12–27. DOI 10.1002/widm.1075.

RUSSO, Claudia, 2019. Deserción Universitaria En Carreras De Informática De La Unnoba. *Universidad Nacional de la Plata* [online]. Retrieved from : <http://sedici.unlp.edu.ar/handle/10915/79958>

SINNEXUS, 2023. Datamining (Minería de datos). *Sinnexus*. . 2023.

SOCIEDAD, Universidad Y et al., 2022. Volumen 14| Número 2 | Marzo-Abril. [online]. Retrieved from : <https://orcid.org/0000-0003-3181-8801>

VIÑÁN, Maria, 2015. Estudio del Rendimiento Académico Aplicando Técnicas de Minería de Datos. . p. 167.

VRIES, Wietse et al., 2011. *¿Desertores o decepcionados? Distintas causas para abandonar los estudios universitarios.* .

ANEXO N° 1: Matriz de consistencia

PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES E INDICADORES	METODOLOGÍA
<p>PROBLEMÁTICA GENERAL</p> <p>¿De qué manera un modelo predictivo aplicando minería de datos determina la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao?</p> <p>PROBLEMAS ESPECÍFICOS</p> <p>PE1: ¿Se podrá comprobar si el factor social influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao aplicando la minería de datos?</p>	<p>OBJETIVO GENERAL</p> <p>Proponer un modelo predictivo que determine la deserción estudiantil universitaria a través de la minería de datos de la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao.</p> <p>OBJETIVOS ESPECÍFICOS</p> <p>OE1: Comprobar si el factor social influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.</p> <p>OE2: Comprobar si el factor económico influye en la deserción estudiantil</p>	<p>HIPÓTESIS GENERAL</p> <p>Existe una relación significativa entre los factores identificados a través de la minería de datos y la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao.</p> <p>HIPÓTESIS ESPECÍFICAS</p> <p>HE1: La minería de datos comprueba que el factor social influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.</p>	<p>VARIABLE INDEPENDIENTE</p> <p>Minería de Datos</p> <p>VARIABLE DEPENDIENTE</p> <p>Deserción estudiantil</p> <p>Indicadores:</p> <ul style="list-style-type: none"> ●Apoyo familiar ●Motivación externa ●Autopercepción ●Situación Laboral ●Financiación de los estudios ●Economía familiar ●Dependencia económica ●Rendimiento ●Desempeño 	<p>TIPO DE ESTUDIO</p> <p>Aplicada</p> <p>Diseño de estudio:</p> <p>Correlacional</p> <p>POBLACIÓN Y MUESTRA</p> <p>La población estuvo constituida por los alumnos de la escuela profesional de Ingeniería de sistemas de la Universidad Nacional del Callao.</p> <p>La muestra utilizada fue no aleatoria, por muestreo de conveniencia o selección intencional, conformada por los alumnos matriculados en los semestres correspondientes entre los períodos de 2021 a 2023</p>

<p>PE2: ¿Se podrá comprobar si el factor económico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao aplicando la minería de datos?</p> <p>PE3: ¿Se podrá comprobar si el factor académico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao aplicando la minería de datos?</p>	<p>universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.</p> <p>OE3: Comprobar si el factor académico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.</p>	<p>HE2: La minería de datos comprueba que el factor económico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.</p> <p>HE3: La minería de datos comprueba que el factor académico influye en la deserción estudiantil universitaria en la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao a través de la minería de datos.</p>		<p>de la facultad de Ingeniería Industrial y de Sistemas de la Universidad Nacional del Callao.</p> <p>Lugar de estudio</p> <p>Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Callao.</p> <p>Técnica: Encuesta</p> <p>Instrumentos: Cuestionario. Escala de Likert.</p>
---	--	---	--	--

ANEXO N° 2: Instrumentos validados

FICHA DE VALIDEZ POR JUECES EXPERTOS (II)

ESCALA DE CALIFICACIÓN

Estimado (a): Dr. Casazola Cruz Osvaldo Daniel

Teniendo como base los criterios que a continuación se presenta, se le solicita dar su opinión sobre el instrumento de recolección de datos que se adjunta:

Marque con una (X) en SI o NO, en cada criterio según su opinión.

CRITERIOS	SI	NO	OBSERVACIÓN
1. El instrumento recoge información que permite dar respuesta al problema de investigación.	X		
2. El instrumento propuesto responde a los objetivos del estudio.	X		
3. La estructura del instrumento es adecuada.	X		
4. Los ítems del instrumento responden a la operacionalización de las variables.	X		
5. La secuencia presentada facilita el desarrollo del instrumento.	X		
6. Los ítems son claros y entendibles.	X		
7. El número de ítems es adecuado para su aplicación.	X		

Opinión de aplicabilidad: Aplicable Aplicable después de corregir []

No aplicable []

SUGERENCIAS:

.....

Apellidos y nombres del juez validador: Casazola Cruz Osvaldo Daniel

DNI: 40081695

Especialidad del validador: metodólogo [] temático [] estadístico

2023

FICHA DE VALIDEZ POR JUECES EXPERTOS (II)

ESCALA DE CALIFICACIÓN

Estimado (a): GUEVARA VALDIVIEZO YERENIA

Teniendo como base los criterios que a continuación se presenta, se le solicita dar su opinión sobre el instrumento de recolección de datos que se adjunta:

Marque con una (X) en SI o NO, en cada criterio según su opinión.

CRITERIOS	SI	NO	OBSERVACIÓN
1. El instrumento recoge información que permite dar respuesta al problema de investigación.	X		
2. El instrumento propuesto responde a los objetivos del estudio.	X		
3. La estructura del instrumento es adecuada.	X		
4. Los ítems del instrumento responden a la operacionalización de las variables.	X		
5. La secuencia presentada facilita el desarrollo del instrumento.	X		
6. Los ítems son claros y entendibles.	X		
7. El número de ítems es adecuado para su aplicación.	X		

Opinión de aplicabilidad: Aplicable Aplicable después de corregir

No aplicable

SUGERENCIAS:

.....

Apellidos y nombres del juez validador: Guevara Valdiviezo Yerenia

DNI: 27736581

Especialidad del validador: metodólogo temático estadístico



FICHA DE VALIDEZ POR JUECES EXPERTOS (II)
ESCALA DE CALIFICACIÓN

Estimado (a):

Teniendo como base los criterios que a continuación se presenta, se le solicita dar su opinión sobre el instrumento de recolección de datos que se adjunta:

Marque con una (X) en SI o NO, en cada criterio según su opinión.

CRITERIOS	SI	NO	OBSERVACIÓN
1. El instrumento recoge información que permite dar respuesta al problema de investigación.	X		
2. El instrumento propuesto responde a los objetivos del estudio.	X		
3. La estructura del instrumento es adecuada.	X		
4. Los ítems del instrumento responden a la operacionalización de las variables.	X		
5. La secuencia presentada facilita el desarrollo del instrumento.	X		
6. Los ítems son claros y entendibles.	X		
7. El número de ítems es adecuado para su aplicación.	X		

Opinión de aplicabilidad: Aplicable Aplicable después de corregir

No aplicable

SUGERENCIAS:

.....
.....

Apellidos y nombres del juez validador: *Bazan Poles Ronel David*

DNI: *41091024*

Especialidad del validador: metodólogo temático estadístico



FICHA DE VALIDEZ POR JUECES EXPERTOS (II)

ESCALA DE CALIFICACIÓN

Estimado (a): ARACIEL CASTAÑEDA HILARIO

Teniendo como base los criterios que a continuación se presenta, se le solicita dar su opinión sobre el instrumento de recolección de datos que se adjunta:

Marque con una (X) en SI o NO, en cada criterio según su opinión.

CRITERIOS	SI	NO	OBSERVACIÓN
1. El instrumento recoge información que permite dar respuesta al problema de investigación.	X		
2. El instrumento propuesto responde a los objetivos del estudio.	X		
3. La estructura del instrumento es adecuada.	X		
4. Los ítems del instrumento responden a la operacionalización de las variables.	X		
5. La secuencia presentada facilita el desarrollo del instrumento.	X		
6. Los ítems son claros y entendibles.	X		
7. El número de ítems es adecuado para su aplicación.	X		

Opinión de aplicabilidad: Aplicable [] Aplicable después de corregir []

No aplicable []

SUGERENCIAS:

.....

Apellidos y nombres del juez validador:

DNI: 08576566

Hilario Araciel Castañeda

Especialidad del validador: metodólogo [] temático [] estadístico []

ANEXO N° 3: Base de datos

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	p1	Numérico	8	2	1 ¿Con qué frecuencia su familia lo apoya para continuar sus estudios universita...	{1,00, Nada satisfech...	Ninguno	8	Derecha	Nominal	Entrada
2	p2	Numérico	8	2	2 ¿Con qué frecuencia recibe apoyo de su familia para sus estudios universitarios?	{1,00, Nada satisfech...	Ninguno	8	Derecha	Nominal	Entrada
3	p3	Numérico	8	2	3 ¿Cómo calificaría el nivel de apoyo que recibe de su familia para sus estudios ...	{1,00, Muy bajo}...	Ninguno	8	Derecha	Nominal	Entrada
4	p4	Numérico	8	2	4 ¿En que escala la familia lo motivan a continuar sus estudios universitarios?	{1,00, Muy bajo}...	Ninguno	8	Derecha	Nominal	Entrada
5	p5	Numérico	8	2	5 ¿Cómo calificaría la influencia de los factores externos en su motivación para c...	{1,00, Muy bajo}...	Ninguno	8	Derecha	Nominal	Entrada
6	p6	Numérico	8	2	6 ¿Qué tan importante es para usted la motivación externa para continuar sus es...	{1,00, Nada important...	Ninguno	8	Derecha	Nominal	Entrada
7	p7	Numérico	8	2	7 ¿Cómo se percibe a sí mismo en relación a su desempeño académico?	{1,00, Muy mal}...	Ninguno	8	Derecha	Nominal	Entrada
8	p8	Numérico	8	2	8 ¿Cómo calificaría su nivel de confianza en sus habilidades académicas?	{1,00, Muy bajo}...	Ninguno	8	Derecha	Nominal	Entrada
9	p9	Numérico	8	2	9 ¿Qué tan satisfecho está con su desempeño académico en la universidad?	{1,00, Nada satisfech...	Ninguno	8	Derecha	Nominal	Entrada
10	p10	Numérico	8	2	10 ¿En que escala de satisfacción laboral se encuentra actualmente mientras es...	{1,00, Nada satisfech...	Ninguno	8	Derecha	Nominal	Entrada
11	p11	Numérico	8	2	11 ¿Qué impacto tiene su trabajo a sus estudios universitarios?	{1,00, Muy bajo}...	Ninguno	8	Derecha	Nominal	Entrada
12	p12	Numérico	8	2	12 ¿Qué tan satisfecho está con su situación laboral mientras estudia en la univer...	{1,00, Nada satisfech...	Ninguno	8	Derecha	Nominal	Entrada
13	p13	Numérico	8	2	13 ¿Cómo calificaría el impacto de sus finanzas en sus estudios?	{1,00, Muy bajo}...	Ninguno	8	Derecha	Nominal	Entrada
14	p14	Numérico	8	2	14 ¿Cómo calificaría su nivel de satisfacción con la financiación de sus estudios ...	{1,00, Nada satisfech...	Ninguno	8	Derecha	Nominal	Entrada
15	p15	Numérico	8	2	15 ¿Qué tan difícil le resulta financiar sus estudios universitarios?	{1,00, Muy difícil}...	Ninguno	8	Derecha	Nominal	Entrada
16	p16	Numérico	8	2	16 ¿Cómo afecta la situación económica de su familia a sus estudios universitari...	{1,00, Muy bajo}...	Ninguno	8	Derecha	Nominal	Entrada
17	p17	Numérico	8	2	17 ¿Cómo calificaría el nivel de apoyo económico que recibe de su familia para s...	{1,00, Muy bajo}...	Ninguno	8	Derecha	Nominal	Entrada
18	p18	Numérico	8	2	18 ¿Qué tan satisfecho está con la situación económica de su familia mientras e...	{1,00, Nada satisfech...	Ninguno	8	Derecha	Nominal	Entrada
19	p19	Numérico	8	2	19 ¿Cuánto depende económicamente de alguien para continuar sus estudios un...	{1,00, Muy poco}...	Ninguno	8	Derecha	Nominal	Entrada
20	p20	Numérico	8	2	20 ¿Cómo afecta su dependencia económica a sus estudios universitarios?	{1,00, Muy poco}...	Ninguno	8	Derecha	Nominal	Entrada
21	p21	Numérico	8	2	21 ¿Qué tan satisfecho está con su nivel de independencia económica mientras ...	{1,00, Nada satisfech...	Ninguno	8	Derecha	Nominal	Entrada
22	p22	Numérico	8	2	22 ¿Cómo calificaría su rendimiento académico en la universidad?	{1,00, Muy mal}...	Ninguno	8	Derecha	Nominal	Entrada
23	p23	Numérico	8	2	23 ¿Cómo calificaría su rendimiento en comparación con sus expectativas?	{1,00, Muy por debajo}...	Ninguno	8	Derecha	Nominal	Entrada
24	p24	Numérico	8	2	24 ¿Qué tan satisfecho está con su rendimiento académico en la universidad?	{1,00, Nada satisfech...	Ninguno	8	Derecha	Nominal	Entrada
25	p25	Numérico	8	2	25 ¿Cómo calificaría su desempeño en las actividades académicas y extracurríc...	{1,00, Muy mal}...	Ninguno	8	Derecha	Nominal	Entrada
26	p26	Numérico	8	2	26 ¿Cómo calificaría su nivel de participación en las actividades académicas y ex...	{1,00, Muy bajo}...	Ninguno	8	Derecha	Nominal	Entrada
27	p27	Numérico	8	2	27 ¿Qué tan satisfecho está con su desempeño en las actividades académicas y...	{1,00, Nada satisfech...	Ninguno	8	Derecha	Nominal	Entrada
28	pp1	Numérico	8	2	1 ¿Con qué frecuencia su familia lo apoya para continuar sus estudios universita...	{1,00, Nada satisfech...	Ninguno	8	Derecha	Nominal	Entrada

	pp15	pp16	pp17	pp18	pp19	pp20	pp21	pp22	pp23	pp24	pp25	pp26	pp27	pre	pos	
1	00	3,00	2,00	4,00	4,00	1,00	3,00	5,00	5,00	4,00	3,00	4,00	4,00	3,00	59,00	97,00
2	00	4,00	3,00	3,00	5,00	4,00	1,00	3,00	2,00	5,00	3,00	3,00	4,00	3,00	58,00	91,00
3	00	5,00	2,00	2,00	3,00	1,00	3,00	3,00	3,00	3,00	3,00	3,00	3,00	4,00	56,00	82,00
4	00	5,00	3,00	3,00	3,00	2,00	3,00	4,00	3,00	4,00	5,00	3,00	4,00	3,00	65,00	97,00
5	00	5,00	3,00	3,00	5,00	4,00	2,00	2,00	1,00	4,00	5,00	4,00	4,00	4,00	63,00	95,00
6	00	3,00	3,00	3,00	3,00	2,00	1,00	5,00	3,00	3,00	3,00	3,00	3,00	5,00	64,00	91,00
7	00	4,00	3,00	4,00	3,00	2,00	2,00	4,00	4,00	3,00	4,00	3,00	4,00	3,00	58,00	94,00
8	00	4,00	3,00	3,00	4,00	1,00	2,00	3,00	5,00	3,00	4,00	4,00	4,00	4,00	63,00	95,00
9	00	3,00	2,00	4,00	4,00	2,00	2,00	2,00	4,00	3,00	3,00	4,00	4,00	5,00	66,00	88,00
10	00	4,00	1,00	5,00	3,00	1,00	1,00	5,00	5,00	4,00	4,00	3,00	5,00	3,00	65,00	96,00
11	00	5,00	2,00	4,00	4,00	2,00	2,00	4,00	3,00	3,00	5,00	4,00	3,00	3,00	69,00	101,00
12	00	5,00	2,00	5,00	4,00	1,00	2,00	5,00	2,00	5,00	5,00	4,00	3,00	3,00	65,00	94,00
13	00	5,00	2,00	4,00	4,00	2,00	1,00	3,00	3,00	4,00	5,00	4,00	4,00	3,00	64,00	92,00
14	00	5,00	1,00	5,00	5,00	1,00	2,00	3,00	3,00	5,00	5,00	5,00	3,00	4,00	69,00	102,00
15	00	5,00	2,00	3,00	3,00	1,00	3,00	3,00	4,00	3,00	5,00	3,00	4,00	3,00	63,00	91,00
16	00	3,00	1,00	5,00	3,00	1,00	3,00	3,00	5,00	3,00	5,00	3,00	3,00	5,00	81,00	90,00
17	00	4,00	2,00	5,00	4,00	2,00	3,00	4,00	4,00	4,00	5,00	4,00	5,00	4,00	71,00	103,00
18	00	5,00	3,00	4,00	3,00	2,00	3,00	3,00	3,00	3,00	4,00	3,00	2,00	5,00	60,00	88,00
19	00	2,00	3,00	4,00	4,00	2,00	3,00	4,00	5,00	4,00	4,00	4,00	3,00	3,00	58,00	92,00
20	00	3,00	3,00	4,00	3,00	2,00	3,00	3,00	5,00	5,00	4,00	5,00	4,00	3,00	66,00	103,00
21	00	5,00	3,00	4,00	5,00	1,00	3,00	3,00	4,00	5,00	4,00	4,00	1,00	4,00	65,00	93,00
22	00	4,00	3,00	3,00	4,00	2,00	3,00	2,00	4,00	3,00	3,00	3,00	5,00	3,00	58,00	94,00
23	00	5,00	3,00	2,00	5,00	1,00	3,00	3,00	4,00	5,00	3,00	5,00	4,00	4,00	56,00	91,00
24	00	3,00	3,00	3,00	4,00	2,00	3,00	3,00	4,00	4,00	3,00	3,00	3,00	5,00	83,00	94,00
25	00	5,00	3,00	5,00	5,00	1,00	2,00	4,00	3,00	5,00	3,00	4,00	2,00	5,00	79,00	93,00
26	00	4,00	2,00	4,00	3,00	2,00	1,00	5,00	3,00	5,00	3,00	1,00	5,00	3,00	96,00	93,00