

UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN

FACULTAD DE INGENIERÍA

ESCUELA DE FORMACIÓN PROFESIONAL DE SISTEMAS Y

COMPUTACIÓN



T E S I S

Técnicas de predicción y su efecto en la detección de patrones de deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023

Para optar el título profesional de:

Ingeniero de Sistemas y Computación

Autor:

Bach. Geraldine Betzabe ROJAS ROBLES

Bach. Deysy Milagros MENDOZA CABELLO

Asesor:

Mg. Pit Frank ALANIA RICALDI

Cerro de Pasco – Perú – 2024

UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN

FACULTAD DE INGENIERÍA

ESCUELA DE FORMACIÓN PROFESIONAL DE SISTEMAS Y

COMPUTACIÓN



T E S I S

**Técnicas de predicción y su efecto en la detección de patrones de
deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023**

Sustentada y aprobada ante los miembros del jurado:

**Msc. Melquiades Arturo TRINIDAD MALPARTIDA
PRESIDENTE**

**Mg. Lisbeth Gisela NEGRETE CARHUARICRA
MIEMBRO**

**Mg. Jose Luis SANCHEZ SOSA
MIEMBRO**



Universidad Nacional Daniel Alcides Carrión

Facultad de Ingeniería

Unidad de Investigación

INFORME DE ORIGINALIDAD N° 088-2024-UNDAC/UIFI

La Unidad de Investigación de la Facultad de Ingeniería de la Universidad Nacional Daniel Alcides Carrión en mérito al artículo 23° del Reglamento General de Grados Académicos y Títulos Profesionales aprobado en Consejo Universitario del 21 de abril del 2022, La Tesis ha sido evaluado por el software antiplagio Turnitin Similarity, que a continuación se detalla:

Tesis:

Técnicas de predicción y su efecto en la detección de patrones de deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023

Apellidos y nombres de los tesistas:

Bach. ROJAS ROBLES, Geraldine Betzabe

Bach. MENDOZA CABELLO, Deysy Milagros

Apellidos y nombres del Asesor:

Mg, ALANIA RICALDI, Pit Frank

Escuela de Formación Profesional

Ingeniería de Sistemas y Computación

Índice de Similitud

27 %

APROBADO

Se informa el Reporte de evaluación del software similitud para los fines pertinentes:

Cerro de Pasco, 05 de marzo del 2024

UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN
FACULTAD DE INGENIERÍA
UNIDAD DE INVESTIGACIÓN
Luis Villar Reguis Carbajal
DOCTOR EN CIENCIAS - DIRECTOR

DEDICATORIA

Dedico con todo mi corazón mi tesis a mis padres y hermanos, ya que sin ellos no lo habría logrado, con sus consejos, apoyo y protección, hoy en día muchos de mis logros se los debo a ustedes.

Geraldine Rojas Robles

A mi familia, que con su esfuerzo me impulsaron a nunca rendirme, por enseñarme a creer en mí y luchar por mis sueños.

Deisy Mendoza Cabello

AGRADECIMIENTO

A DIOS por darnos salud, fuerzas y sabiduría para continuar en nuestros proyectos profesionales.

Al mismo tiempo queremos agradecer a nuestros docentes, quienes con su paciencia y dedicación nos orientaron a ser buenos profesionales.

A nuestro asesor de tesis, Mg. Pit Frank Alania Ricaldi que con sus conocimientos y consejos nos ha orientado en cada proceso de nuestro proyecto.

RESUMEN

El trabajo de investigación que realice se titula: “Técnicas de predicción y su efecto en la detección de patrones de deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023” tuvo como objetivo principal determinar las técnicas de predicción tienen efecto en la detección de patrones de deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023, el diseño de investigación es descriptivo. La población Estudiante de la Facultad de Ingeniería Matriculados 2023-A 1274, Como muestra se tomará a los Estudiante de la Facultad de Ingeniería de la Escuela de Ingeniería de Sistemas y Computación Matriculados 2023-A 215. Con los resultados Debido a que el Chi cuadrado calculado es mayor que el Chi cuadrado tabular se comprueba la hipótesis Alterna que dice H1: En qué manera las técnicas de predicción tienen efecto en la detección de patrones de deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023, y se rechaza la hipótesis nula (H0), en conclusión las dificultades económicas y familiares, así como el bajo rendimiento son factores predominantes en la Deserción Estudiantil, poca incidencia tiene indicadores como: sanción disciplinaria, cambio de ciudad, ubicación laboral, cambio de carrera, espacio físico para los estudios, nota obtenida en el examen de admisión, influencia de los padres al momento de elegir la Carrera y enfermedad

Palabras Clave. Técnicas de predicción, Deserción estudiantil.

ABSTRACT

The research work I carried out is titled: "Prediction techniques and their effect on the detection of student dropout patterns at the UNDAC 2023 Faculty of Engineering" had the main objective of determining the prediction techniques that have an effect on the detection of patterns. of student dropout in the UNDAC Faculty of Engineering 2023, the research design is descriptive. The population Student of the Faculty of Engineering Enrolled 2023-A 1274, As a sample, the Students of the Faculty of Engineering of the School of Systems and Computing Engineering Enrolled 2023-A 215 will be taken. With the results Because the Chi square calculated is greater than the tabular Chi square, the Alternate hypothesis is tested, which says H1: In what way do prediction techniques have an effect on the detection of student dropout patterns in the UNDAC 2023 Faculty of Engineering, and the null hypothesis is rejected. (H0), in conclusion, economic and family difficulties, as well as low performance, are predominant factors in Student Dropout; there is little incidence of indicators such as: disciplinary sanction, change of city, job location, change of career, physical space for studies, grade obtained in the admission exam, influence of parents when choosing a career and illness.

Keywords. Prediction techniques, Student dropout.

INTRODUCCIÓN

En todas las instituciones educativas, especialmente en el nivel universitario se encuentran registrados estudiantes quienes se matriculan en las diferentes escuelas que estas ofrecen, pero a medida que se avanza en dichos estudios, también se observa que muchos de estos estudiantes dejan las aulas y por ende sus estudios, esta situación se debe a muchas causas o situaciones problemáticas, entre los que se pueden mencionar los factores sociales, económicos, académicos, personales entre otros. Y es así como el abandono estudiantil llega a convertirse en una problemática de tipo social que afecta a todas las Universidades a nivel mundial, por lo que reducir estos índices es la consigna que se ponen como meta muchas de estas, implementando para ello, dentro de sus planes, estrategias con el firme propósito de alcanzarlo.

Consta de los siguientes capítulos:

Capítulo I: Se empieza con la Identificación y determinación del problema, delimitación de investigación, formulación del problema, formulación de objetivos, justificaciones de la investigación y limitaciones de la investigación”.

Capitulo II: Se define las Bases teóricas – científicas, definición de términos, identificación de hipótesis y variables y finalmente la definición operacional.

Capitulo III: Se muestra el Tipo de investigación, métodos de investigación, diseño de investigación, población y muestra, técnicas e instrumentos de recolección de datos, técnicas de procesamiento y análisis de datos, tratamiento estadístico de datos, selección, validación y confiabilidad de los instrumentos de investigación y orientación ética.

Capitulo IV: Se explicará los resultados y discusión de los resultados con los de otros autores.

Finalmente, se detallará las conclusiones, recomendaciones y las referencias bibliográficas que se usaron de los distintos autores.

Los autores.

ÍNDICE

DEDICATORIA	
AGRADECIMIENTO	
RESUMEN	
ABSTRACT	
INTRODUCCIÓN	
ÍNDICE	
ÍNDICE DE TABLAS	
ÍNDICE DE FIGURAS	

CAPITULO I

PROBLEMA DE INVESTIGACIÓN

1.1.	Identificación y determinación del problema	1
1.2.	Delimitación de la investigación.....	2
1.2.1.	Delimitación espacial	2
1.2.2.	Delimitación temporal	3
1.2.3.	Delimitación conceptual	3
1.3.	Formulación del problema	3
1.3.1.	Problema general	3
1.3.2.	Problemas específicos.....	3
1.4.	Formulación de objetivos.....	3
1.4.1.	Objetivo General.....	3
1.4.2.	Objetivos específicos	3
1.5.	Justificación de la investigación.....	4
1.6.	Limitaciones de la investigación	4

CAPITULO II

MARCO TEÓRICO

2.1.	Antecedentes de estudio.....	5
2.1.1.	A nivel Internacional.....	5
2.1.2.	A nivel Nacional	8
2.1.3.	A nivel Local	12
2.2.	Bases teóricas – científicas.....	13
2.2.1.	Técnicas de predicción.....	13
2.2.2.	Minería de datos	14
2.2.3.	Trayectoria académica.....	18
2.2.4.	Deserción Estudiantil	18
2.3.	Definición de términos básicos.....	18
2.4.	Formulación de Hipótesis	19
2.4.1.	Hipótesis General	19
2.4.2.	Hipótesis Específicas.....	19
2.5.	Identificación de Variables.....	19
2.5.1.	Variables independientes.....	19
2.5.2.	Variables dependientes	19
2.6.	Definición operacional de variables e indicadores	20

CAPITULO III

METODOLOGÍA Y TECNICAS DE INVESTIGACIÓN

3.1.	Tipo de investigación.....	21
3.2.	Nivel de investigación.....	21
3.3.	Métodos de investigación	21
3.4.	Diseño de investigación.....	21
3.5.	Población y muestra	22
3.5.1.	Población.....	22
3.5.2.	Muestra.....	22

3.6.	Técnicas e instrumentos de recolección de datos	22
3.7.	Selección, validación y confiabilidad de los instrumentos de investigación.....	23
3.8.	Técnicas de procesamiento y análisis de datos.....	23
3.9.	Tratamiento Estadístico.....	24
3.10.	Orientación ética filosófica y epistémica.	24

CAPITULO IV

RESULTADOS Y DISCUSIÓN

4.1.	Descripción del trabajo de campo	25
4.2.	Presentación, análisis e interpretación de resultados	26
4.3.	Prueba de Hipótesis	34
4.4.	Discusión de resultados	37

CONCLUSIONES

RECOMENDACIONES

REFERENCIAS BIBLIOGRÁFICAS

ANEXOS

ÍNDICE DE TABLAS

Tabla 1 Definición operacional de variables.....	20
Tabla 2 Tabla de opciones.....	26
Tabla 3 Automatización de procesos	27
Tabla 4 Técnicas de predicción.....	28
Tabla 5 Dedicación de estudio.....	29
Tabla 6 Decisión de quien influyo	29
Tabla 7 Nivel económico.....	30
Tabla 8 Recurso económico	31
Tabla 9 Beca.....	31
Tabla 10 Asignatura.....	32
Tabla 11 Ingreso a la universidad	33
Tabla 12 Periodo académico	33

ÍNDICE DE FIGURAS

Figura 1 Categorías fundamentales	14
Figura 2 División de modelos	16
Figura 3 Gráfico de opciones	26
Figura 4 Gráfico de automatización de procesos	27
Figura 5 Gráfico de técnicas de predicción	28
Figura 6 Gráfico de dedicación de estudio	29
Figura 7 Gráfico de decisión de quien influyo.....	30
Figura 8 Nivel económico.....	30
Figura 9 Gráfico de recurso económico	31
Figura 10 Gráfico de beca.....	32
Figura 11 Gráfico de asignatura.....	32
Figura 12 Gráfico de ingreso a la universidad	33
Figura 13 Gráfico de periodo académico	34
Figura 14 Valor Chi - cuadrado	36

CAPITULO I

PROBLEMA DE INVESTIGACIÓN

1.1. Identificación y determinación del problema

La conceptualización de deserción escolar difiere entre varios autores, sin embargo, si la universidad pierde a sus alumnos se evidenciará en sus índices de retención estudiantil. El abandono o deserción de un estudiante en un determinado nivel puede basarse en la no matriculación de dichos estudiantes al siguiente nivel.

En todas las instituciones educativas, especialmente en el nivel universitario se encuentran registrados estudiantes quienes se matriculan en las diferentes escuelas que estas ofrecen, pero a medida que se avanza en dichos estudios, también se observa que muchos de estos estudiantes dejan las aulas y por ende sus estudios, esta situación se debe a muchas causas o situaciones problemáticas, entre los que se pueden mencionar los factores sociales, económicos, académicos, personales entre otros. Y es así como el abandono estudiantil llega a convertirse en una problemática de tipo social que afecta a todas las Universidades a nivel mundial, por lo que reducir estos índices es la consigna que se ponen como meta muchas de estas, implementando para ello, dentro de sus planes, estrategias con el firme propósito de alcanzarlo.

Según el diario El Comercio (2018), la deserción en el Perú en la zona rural es 13.4% y en la zona urbana 16%, por lo que se hace necesario no solo mirar estos valores desde una perspectiva estadística o estudiar su casuística, sino establecer modelos que ayuden en su predicción para establecer e implementar estrategias con el fin de prevenirlo (García Lázaro, 2015).

En este contexto con el proceso de descubrimiento de conocimiento en bases de datos KDD (Knowledge Discovery in Databases), la finalidad es encontrar dentro de los datos de los estudiantes de los diferentes Programas de Estudios tendencias que se encuentran ocultas, posibilitando la manera de asociar la deserción estudiantil a ciertas variables predictivas. La investigación tiene como fundamento técnico mediante los lineamientos del proceso KDD, encontrar patrones teniendo como insumo las bases de datos.

Considerando lo antes mencionado, con la finalidad de descubrir patrones o reglas que logren asociar los atributos de los datos que se tienen de los estudiantes que son parte de la deserción estudiantil en el Instituto, se plantea el análisis de los datos recolectados a partir de las actas oficiales de secretaría académica, así como de las fichas personales de ingreso de los postulantes y certificados de estudios secundarios que presenta como requisito de postulación en el examen de admisión institucional, aplicando las técnicas que se utiliza en la minería de datos, las cuales son algoritmos validados que se aplican sobre una data con el fin de descubrir tendencias entre estos y de esta forma establecer modelos predictivos. Una técnica utilizada dentro de la minería de datos son los árboles de decisión.

1.2. Delimitación de la investigación.

1.2.1. Delimitación espacial

Con la presente investigación se diagnosticó, analizó, selecciono y utilizo la minería de datos como técnica de predicción para detectar patrones de Deserción Estudiantil en la Facultad de Ingeniería de la UNDAC 2023.

1.2.2. Delimitación temporal

La investigación se llevó a cabo en el año 2023.

1.2.3. Delimitación conceptual

Concepto de Base de Datos que están relacionados con las técnicas de predicción-Minería de Datos.

1.3. Formulación del problema

1.3.1. Problema general

¿En qué manera las técnicas de predicción tienen efecto en la detección de patrones de deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023?

1.3.2. Problemas específicos

¿Cuál es la situación actual en la Facultad de ingeniería de la UNDAC 2023, con respecto a la deserción estudiantil?

¿Qué técnicas de predicción se aplica actualmente en la Facultad de ingeniería de la UNDAC 2023, con respecto a la deserción estudiantil para detectar la posible deserción estudiantil?

¿Es necesario establecer las características del modelo que permitirá detectar patrones de deserción Estudiantil en la Facultad de Ingeniería de la UNDAC 2023?

1.4. Formulación de objetivos

1.4.1. Objetivo General

Determinar las técnicas de predicción tienen efecto en la detección de patrones de deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023.

1.4.2. Objetivos específicos

Estudiar la situación actual en la Facultad de ingeniería de la UNDAC 2023, con respecto a la deserción estudiantil.

Analizar las técnicas de predicción se aplica actualmente en la Facultad de ingeniería de la UNDAC 2023, con respecto a la deserción estudiantil para detectar la posible deserción estudiantil.

Establecer las características del modelo que permitirá detectar patrones de deserción Estudiantil en la Facultad de Ingeniería de la UNDAC 2023.

1.5. Justificación de la investigación

La aplicación de una técnica de predicción como la minería de datos permitió visualizar las causas e identificar patrones por las cuales los estudiantes abandonan sus estudios superiores, expresados mediante estadísticas, los porcentajes de deserción por curso y carrera, utilizando cohortes para su estudio, esto fue posible mediante la automatización de procesos. Mediante un análisis de las gráficas estadísticas que muestren los diferentes factores económicos, sociales, etc., las autoridades de la institución podrán tomar decisiones para prever posibles abandonos estudiantiles.

1.6. Limitaciones de la investigación

Recursos económicos limitados para la recopilación de requerimientos, información y validación de la investigación.

CAPITULO II

MARCO TEÓRICO

2.1. Antecedentes de estudio

2.1.1. A nivel Internacional

Según Peña, Sola y López (2017) en su trabajo de investigación titulado “Análisis de los factores de la deserción estudiantil en una facultad de las unidades tecnológicas de Santander, Bucaramanga, Colombia” Las Unidades Tecnológicas de Santander, es una institución de Educación Superior, líder local dentro de las de su misma naturaleza, e igual que ellas presenta altos niveles de deserción, factor prioritario dentro de los estudios a nivel nacional y que se ha convertido en requisito de verificación en procesos de acreditación, de implementación obligatoria. Las estadísticas muestran niveles de deserción que superan el 50%. De esta situación surge la inquietud de cuáles son las causas que generan el abandono de los programas de educación superior y las posibles alternativas para aumentar los niveles de retención. Partiendo de referentes teóricos que abordan el fenómeno de la deserción, el proceso de investigación siguió los lineamientos de una metodología mixta de corte descriptivo. Los instrumentos de recogida de datos fueron de tipo cualitativo y cuantitativo, permitiendo obtener resultados válidos esperados; que en primer momento fueron los datos sistematizados para establecer y priorizar las causas de la

deserción; en segundo momento el análisis de los datos a través del cruce la información; y en tercer lugar proposición de estrategias que contribuyan a fijar metas y políticas institucionales.

Según Timarán, Calcedo, Hidalgo (2019) en su trabajo de investigación titulado “Árboles de decisión para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas Saber 11°” se presentan los resultados obtenidos al aplicar el modelo de clasificación basado en árboles de decisión, con el fin de detectar factores asociados al desempeño académico de los estudiantes colombianos de grado undécimo de educación media, que presentaron las pruebas Saber 11° en los años 2015 y 2016. La investigación fue de tipo descriptivo bajo el enfoque cuantitativo, aplicando un diseño no experimental. Siguiendo la metodología CRISP-DM, se seleccionó, de las bases de datos del ICFES, la información socioeconómica, académica e institucional de estos estudiantes. Se construyó, limpió y transformó un repositorio de datos y utilizando la herramienta de minería de datos WEKA, se generaron árboles de decisión que permitieron identificar patrones asociados al buen o mal desempeño académico de los estudiantes en las pruebas Saber 11°. Los patrones descubiertos ayudarán en los procesos de toma de decisiones del Ministerio de Educación Nacional, junto con las instituciones que velan por la calidad de la educación en Colombia.

Según Cid (2015) en su trabajo de investigación titulado “Propuesta de un modelo de estimación del abandono universitario” se presenta una nueva metodología para el diseño de herramientas educativas para la enseñanza de la programación orientada a objetos (POO) basadas en la taxonomía de Bloom, de modo que se logre un desarrollo sistemático del software educativo. Aunque existen muchas aplicaciones de ayuda a la enseñanza- aprendizaje de la POO, no hemos hallado métodos de diseño de dichas herramientas que contengan un objetivo pedagógico claro. La taxonomía de Bloom es una jerarquía de seis

niveles con grado creciente de aprendizaje del alumno. Cada nivel presupone la capacitación del alumno en los niveles precedentes. Según ascendemos por la jerarquía nos encontramos un mayor grado de aprendizaje. Al ser una taxonomía ampliamente aceptada, la usamos como un marco adecuado para fijar la meta educativa que debe poseer cualquier software educativo. La obtención de este objetivo hace que el diseño nos lleve a plantear una serie de características para cada uno de los tres primeros niveles de Bloom. La limitación a los niveles iniciales es debida a dos factores. El primero es la forma de aprendizaje de los discentes. Los alumnos primero captan la información (primer nivel o de conocimiento), la interiorizan (segundo nivel o de comprensión) y después la aplican (tercer nivel o de aplicación) ante nuevas situaciones. Si alguna de estas fases no es convenientemente adquirida, el aprendizaje no es duradero. El segundo factor es el software existente. Pensamos que para los niveles superiores ya hay herramientas educativas (por ejemplo, los entornos de programación) que les ofrecen soporte. Cualquier propuesta debe probar su validez, así, en el marco de este trabajo de investigación, también hemos desarrollado, utilizando esta metodología, varias herramientas educativas para la enseñanza de distintos conceptos fundamentales y distintivos de la POO con el resto de paradigmas. Este software ha sido puesto a disposición de los alumnos, que han sido los encargados de valorar su eficacia docente. Algunos han participado en distintos experimentos cuidadosamente diseñados, siguiendo una metodología de investigación general. Los resultados han demostrado que las aplicaciones educativas realizadas cumplen con su principal objetivo: ayudar al alumno en un determinado nivel de la taxonomía de Bloom. También se ha recabado la opinión de los alumnos, juzgando estas aplicaciones como útiles para su proceso de aprendizaje.

Según Cuji (2016) en su trabajo de investigación titulado “Las técnicas de predicción y su incidencia en la detección de patrones de deserción estudiantil en la carrera de docencia en informática de la Facultad DE Ciencias Humanas y de la Educación de la Universidad Técnica de Ambato” El estudio sobre Deserción Estudiantil tuvo como objetivo primario detectar el porcentaje de abandono escolar que presenta la Carrera de Docencia en Informática, a partir del año 2006 hasta el año 2015, con base en ésta información se aplicó el algoritmo de Arboles de decisión para diseñar un prototipo de modelo predictivo de Deserción Estudiantil, la metodología usada se basa en el método KDD(Knowledge Discovery in Database), detallado en cinco etapas , selección, procesamiento, transformación, minería de datos, e interpretación de la información. Posterior a la aplicación del algoritmo se obtuvo un árbol de decisión de cuatro niveles de profundidad, evidenciando que las variables nivel y notas tienen mayor influencia en la Deserción Estudiantil dentro de la Carrera. Finalmente se obtuvieron cuatro reglas que fueron programadas y visualizadas en una interfaz web, que evalúa a los nuevos posibles desertores de la Carrera de Docencia en Informática.

2.1.2. A nivel Nacional

Según Gamarra, Matos y Yupanqui (2018) en su trabajo de investigación titulado “Detección de patrones de éxito en estudios universitarios de la Universidad Continental” El objetivo del estudio fue detectar los patrones de éxito en los estudios universitarios de los estudiantes de la Universidad Continental a partir de la información académica y socio-demográfica recopilado en el primer ciclo de estudios. La investigación tuvo un diseño descriptivo transversal, los estudiantes fueron de diversas carreras profesionales que desarrollaron sus estudios entre el 2012 y 2017. Para la extracción automatizada de conocimiento se utilizó la metodología de proyectos de minería de datos Cross-Industry Standard Process for Data Mining (CRISP-DM) y el software

Clementine, mediante la red neuronal perceptrón multicapa se logró identificar las variables que más impactan en el éxito y abandono de los estudios universitarios, para obtener el árbol de decisión se aplicó el algoritmo C5.0 y para el agrupamiento (tres grupos) los algoritmos k-means y TwoStep, y estos resultados se compararon mediante una matriz de confusión. Los resultados muestran que los estudiantes que abandonaron sus estudios universitarios no pasaron del quinto ciclo; las variables que más influyen en el éxito de los estudios universitarios son: estado civil del estudiante, monto de la pensión, estado civil de los padres y con quien vive el estudiante en Huancayo. Por otro lado, las variables que más influyen en el abandono de los estudios son, con quien vive el estudiante, estado civil del estudiante, satisfacción del desempeño del docente y estado civil de los padres. En conclusión, las variables socio demográficas son las que tienen mayor predominancia sobre el éxito y fracaso de los estudios universitarios.

Según García (2020) en su trabajo de investigación titulado “Detección de patrones de deserción estudiantil mediante aplicación de Árboles de Decisión C4.5 en el IESTP “Señor de Chocán” de Querecotillo” Para ello se utilizó la metodología KDD para la selección, preprocesamiento y transformación de los datos antes de ser analizados con el algoritmo C4.5 para su interpretación y evaluación, con lo que se logró identificar que los factores que afectan la permanencia del estudiante en dicha IE son las variables predictoras: Año de ingreso, Programa de Estudios, Promedio del último semestre en el que se retiró y si Trabaja el estudiante, con porcentaje de precisión de 100%, 90.7%, 76.8% y 76.6%, respectivamente. Además, se estimó un modelo predictor de riesgo de deserción estudiantil con la aplicación de árboles de decisión C4.5 en el software Weka mediante la validación cruzada, obteniéndose una tasa de precisión del 100% y tiempo de entrenamiento de 0.02 segundos para un total de 915 instancias, con lo que el rendimiento del árbol de decisión para la validación del

modelo de predicción fue de 81.53%. Por lo que se concluye que el modelo validado logró estimar variables predictoras que influyen en la deserción estudiantil lo que servirá de apoyo para su análisis y posterior toma de decisiones en dicha institución educativa.

Según Zarate (2019) en su trabajo de investigación titulado “Reconocimiento De Patrones De Deserción De Estudiantes Universitarios Basados En Modelos De Clasificación” El trabajo estudia los factores que afectan el rendimiento académico de los estudiantes de la Universidad Nacional de San Agustín, las cuales incluyen características personales (género, edad), información previa a su ingreso a la universidad (tipo de colegio de procedencia, puntuación en examen de ingreso a la universidad) y el desempeño que posee en las asignaturas a lo largo de uno o más semestres académicos (promedio de notas por semestre, relación de créditos, nro. de asignaturas abandonadas, etc.). Para esto se ha creado una aplicación que utiliza información proporcionada por la universidad acerca de los estudiantes y que permite a los directores de las escuelas profesionales de la universidad el poder detectar de manera anticipada los estudiantes que presenten algún tipo de riesgo de abandonar la carrera. El reconocimiento de los patrones de deserción de los estudiantes es posible mediante el uso de modelos de clasificación generados a partir de diferentes algoritmos (redes neuronales, ID3, C4.5), utilizando los atributos más significantes dentro de la información que se tenía a la mano. Se ha creado modelos de clasificación por cada algoritmo que ha sido implementado y se realizaron pruebas utilizando diferentes variables de entrada por modelo. Cada modelo intenta detectar la deserción de un estudiante de su carrera profesional por lo que las variables de salida de cada uno serán: el estudiante ABANDONA y el estudiante NO ABANDONA la carrera. Se realizó también una comparación de estos modelos mediante el cálculo de medidas de rendimiento (precisión, exactitud y sensibilidad), y aquellos que ofrezcan

mejores resultados serán aplicados cuando se realice la clasificación de estudiantes en las distintas escuelas profesionales de la universidad.

Según Camborda (2014) en su trabajo de investigación titulado “Aplicación de árboles de decisión para la predicción del rendimiento académico de los estudiantes de los primeros ciclos de la Carrera de Ingeniería Civil de la Universidad Continental” el objetivo de esta investigación fue predecir el rendimiento académico identificando las variables de los factores que más influyen en el estudiante en su rendimiento académico. La herramienta utilizada en esta investigación es Árboles de decisión como modelo de predicción basado en reglas utilizado en el ámbito de la inteligencia artificial que sirve para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema. Utilizando datos sobre factores demográficos, académicos, institucionales actitudinales de los estudiantes se aplicó la herramienta de árboles de decisión, específicamente el algoritmo j48 de weka que es un software para minería de datos. Sin embargo, es preciso mencionar que cuando se utilizaban las variables en conjunto correspondientes a un solo aspecto, se tuvo menor exactitud que cuando se realizó combinando variables de diversos aspectos. Los resultados que se obtuvieron aplicando esta herramienta fueron las variables más influyentes en el estudiante que permite predecir su rendimiento académico según intervalos de notas que se consideró de acuerdo a su promedio del estudiante. Las variables académicas fueron las que resultaron con más ganancia de información por lo que con estas variables la predicción tiene mayor exactitud y definen el rendimiento académico del estudiante. Se recomienda clasificar los datos de los estudiantes para no solo predecir el rendimiento académico ayudado de la minera de datos, sino otros indicadores que son necesarios para poder predecir el rendimiento académico del estudiante, siendo la minería de datos una herramienta con un potencial increíble y aplicable en un sinnúmero de proyectos, circunstancias y finalidades.

Según Piscoya (2016) en su trabajo de investigación titulada “Aplicación de técnicas de minería de datos para predecir la deserción estudiantil en la educación básica regular en la Región de Lambayeque” tiene como objetivo proponer una herramienta utilizando las técnicas de minería de datos, donde permita al usuario tener acceso a la información precisa donde se realicen predicciones sobre los alumnos que se matriculen en los próximos años, obteniendo resultados a corto plazo, que permitirá asegurar la confiabilidad de éstos, sirviendo de apoyo a la institución para las decisiones futuras que se puedan tomar. Dentro de las técnicas predictivas se determinó utilizar los algoritmos de ETS y Redes Neuronales, al realizar el análisis se descartaron algunas técnicas adicionales por no tener los criterios necesarios para su implementación en el modelo a desarrollar. Además, se presentan los antecedentes de estudio a nivel de base teórica, tomando como fuentes libros, publicaciones, entre otros, los cuales permiten justificar muchos de los conceptos abarcados durante el proceso de investigación. Se presenta el desarrollo de las metodologías empleadas para la solución del problema planteado; se utilizó la Metodología CRISP DM, como guía para la construcción del modelo de minería de datos basado en series de tiempo logrando realizar las predicciones de deserción escolar en la región de Lambayeque donde solo se tomó como muestra la Ugel de Chiclayo en periodos anuales de manera automatizada dejando de lado el uso de herramientas ofimáticas que retrasan el proceso de los resultados; y el uso de la metodología XP para el desarrollo del sistema como solución a la optimización de los procesos mostrando los resultados.

2.1.3. A nivel Local

Según Alania (2018) en su trabajo de investigación titulado “Aplicación de técnicas de minería de datos para predecir la deserción estudiantil de la facultad de ingeniería de la Universidad Nacional Daniel Alcides Carrión” Los

métodos de clasificación como árboles de decisión, las reglas, red bayesiana, etc. se pueden aplicar a los datos académicos de una universidad para predecir el comportamiento de los estudiantes, el rendimiento en los exámenes, deserción estudiantil, etc. Esta predicción ayudará a las autoridades para identificar la deserción estudiantil y poder determinar la proyección de secciones y otras acciones. El algoritmo de árbol de decisión C4.5 (J48) se aplica en los datos de las notas finales semestrales de los estudiantes para predecir si abandona o no los estudios. El resultado del árbol de decisión predijo el número de estudiantes que son propensos a abandonar la carrera profesional. El resultado lo pueden utilizar a las autoridades para que puedan tomar las medidas para mejorar la toma de decisiones. Después de la evaluación con los datos originales se introducen 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. Para analizar la exactitud del algoritmo, se compara con el algoritmo Random Tree y se encontró que es tan eficiente en términos de precisión de los resultados académicos del estudiante y el tiempo tomado para crear el árbol.

2.2. Bases teóricas – científicas

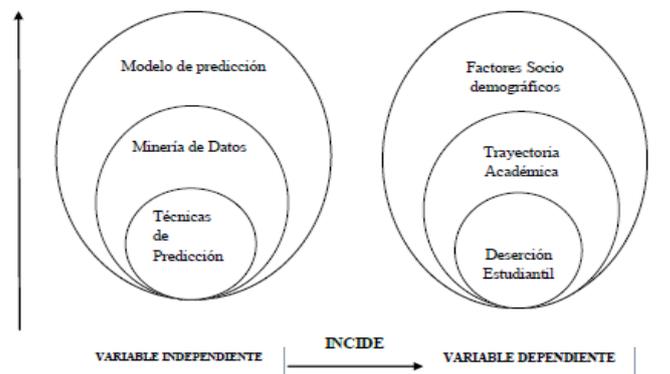
2.2.1. Técnicas de predicción

Existen dos enfoques o áreas dentro del Data Mining: las técnicas clásicas de Estadística y las de minería de datos propiamente dicha que derivan, en su mayoría, de las aportaciones de la Inteligencia Artificial como las redes neuronales, lógica borrosa, mecanismos de aprendizaje automático.

Las funciones estadísticas ofrecen diversos métodos de pronóstico para dar apoyo a la toma de decisiones empresariales. Aunque no son propias de minería de datos (se utilizaban mucho antes de que surgiera este concepto), resultan de gran utilidad a la hora de descubrir patrones o elaborar modelos de predicción. Pueden emplearse para obtener más información sobre los datos, lo

que permitirá tomar decisiones más acertadas cuando se apliquen los procesos de minería. Algunos ejemplos de estas técnicas son la regresión lineal, el Análisis Factorial, el Análisis de Componentes Principales. Mientras que entre las técnicas de minería destacan el Clustering, la Clasificación, la predicción.

Figura 1 *Categorías fundamentales*



2.2.2. Minería de datos

La minería de datos es entendida como el proceso de descubrir conocimientos interesantes, como patrones, asociaciones, cambios, anomalías y estructuras significativas a partir de grandes cantidades de datos almacenadas en bases de datos, datawarehouses, o cualquier otro medio de almacenamiento de información. (Britos & Hossian, 2005).

La aplicación de algoritmos de minería de datos requiere de actividades previas destinadas a preparar los datos de manera homogénea. Esta primera etapa es también conocida como ETL (Extract, Transform and Load) (Kimball, 2002).

Un proceso completo de aplicación de técnicas de minería, mejor conocido como proceso de descubrimiento del conocimiento en bases de datos establece a la minería de datos como una etapa del mismo. (Valero, Salvador, & García, 2008).

Dentro de ésta se pueden utilizar diversos algoritmos predictivos como: Árboles de decisión C4.5: categorizado como aprendizaje basado en similitudes, los árboles de decisión son uno de los algoritmos más sencillos y fáciles de implementar y a su vez de los más poderosos. Este algoritmo genera un árbol de decisión de forma recursiva al considerar el criterio de la mayor proporción de ganancia de información (gain ratio) (Britos & Hossian, 2005) es decir, elige al atributo que mejor clasifica a los datos. Técnica de los k vecinos más cercanos: conocido como algoritmo de aprendizaje basado en instancias, su funcionamiento es muy simple: se almacenan los ejemplos de entrenamiento de datos históricos y cuando se requiere clasificar a un nuevo objeto, se extraen los objetos más parecidos y se usa su clasificación para clasificar al nuevo objeto. (Morales, 2009) Los vecinos más cercanos a una instancia se obtienen, para el caso de los atributos continuos, utilizando la distancia Euclidiana sobre los n posibles atributos. El resultado de la clasificación por medio de este algoritmo puede ser discreto o continuo. En el caso discreto, el resultado de la clasificación es la clase más común de los k vecinos. (Valero, Salvador, & García, 2008).

El término 'KDD' es empleado para describir el proceso total de descubrimiento y extracción de conocimiento nuevo, no obvio a partir de un conjunto de datos, el cual está conformado por relaciones y patrones entre los elementos que conforman los datos.

El proceso de KDD abarca varias etapas en su realización, desde la selección de datos que pueden ser necesarios para descubrir conocimiento, hasta visualizar los resultados de dicho descubrimiento. El principal proceso dentro del KDD es la minería de datos 'Data Mining', que es la responsable de buscar, descubrir y extraer el conocimiento desde los datos. (Santamaria, 2015)

Aplicaciones de la Minería de Datos

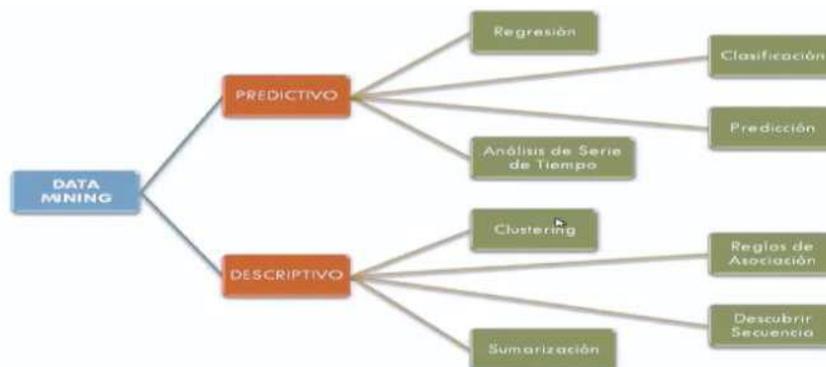
- Retención de clientes y Manejo de riesgo (score de crédito – puntuación de crédito)
- Patrones de Compra (algoritmos de recomendación)
- Detección de fraude (detección de transacciones fraudulentas en bancos, tarjetas de crédito, etc.)
- Segmentación de clientes (categorizar los clientes, por ejemplo: dirigir campañas publicitarias por segmentos)

Predicción de Ventas (Ventas que tendrá un supermercado para evitar el exceso de productos en inventario y no se pierdan los productos)

División de los modelos que usa la Minería de Datos

- Descriptivo
- Predictivo

Figura 2 División de modelos



Análisis Descriptivo de Minería de Datos (Modelos de Descripción)

Búsqueda de patrones en grandes volúmenes de información; son modelos no supervisados. Entre las técnicas más usadas en este método se tiene:

- OLAP (Visualización)
- Clustering

- Métodos factoriales – Métodos de reducción de la dimensión como el Análisis en componentes principales (ACP), etc.

Clustering

Su función principal es encontrar grupos de datos.

Análisis de Predicción de Minería de Datos (Modelo de Predicción)

Los modelos de predicción son abstracciones de representaciones del mundo real aplicadas al tratamiento predictivo. (Tutiempo Network, 2015). Los modelos de predicción se aplican generalmente cuando la variable a estudiada es cualitativa, se usa generalmente para pronosticar futuros fraudes, deserciones, score crediticio, etc.

Son modelos supervisados, las técnicas más usadas para el análisis predictivo en:

Minería de datos son:

- Series de tiempo
- Análisis Discriminante
- Regresión
- Árboles de decisión
- Métodos de potenciación, etc.

Datos sobre los que trabaja la minería de datos

Hoy en día los datos se encuentran almacenados en:

- Base de datos relacionales
- Bodegas de datos
- Base de datos transaccionales
- Bases de datos orientada a objetos y simbólicas
- Bases de datos espaciales sistemas de información geográfica
- Series cronológicas de datos y datos temporales
- Bases de datos de texto

- Bases de datos multimedia
- www (Web Mining)

La Minería de Datos se trabaja sobre todo tipo de información, sin embargo, es conveniente concentrar los datos en una matriz, para facilidad.

2.2.3. Trayectoria académica

Indudablemente la educación superior constituye la puerta de acceso más importante a la sociedad del conocimiento, toda vez que representa el medio ideal para el acrecentamiento del capital humano y de su inteligencia individual y colectiva; así fue reconocido en la Conferencia Mundial sobre Educación Superior organizada por la UNESCO en 1998: "...la educación superior y la investigación forman hoy en día la parte fundamental del desarrollo cultural, socioeconómico y ecológicamente sostenible de los individuos, las comunidades y las naciones" (UNESCO, 1998).

2.2.4. Deserción Estudiantil

Abandono de la institución educativa. Se trata de aquella situación en la que el alumno después de un proceso acumulativo de separación o retiro, finalmente, comienza a retirarse antes de la edad establecida por el sistema educativo sin obtener un certificado.

2.3. Definición de términos básicos.

- **Confiabilidad**

Los datos utilizados para la construcción del indicador deben ser fidedignos (fuentes de información satisfactorias)

- **Disponibilidad:**

Los datos básicos para la construcción del indicador deben ser de fácil obtención sin restricciones de ningún tipo.

- **Especificidad:**

Si un indicador no mide realmente lo que se desea medir, su valor es limitado, pues no permite la verdadera evaluación de la situación, al reflejar características que pertenecen a otro fenómeno paralelo.

- **Sensibilidad**

El indicador debe ser capaz de poder identificar las distintas situaciones aún en áreas con distintas particularidades, independientemente de la magnitud que ellas tengan en la comunidad

2.4. Formulación de Hipótesis

2.4.1. Hipótesis General

Las técnicas de predicción tendrán efecto en la detección de patrones de deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023.

2.4.2. Hipótesis Específicas

El estudio actual en la Facultad de ingeniería de la UNDAC 2023, ayudara a ver la deserción estudiantil.

Las técnicas de predicción que se aplican actualmente en la Facultad de ingeniería de la UNDAC 2023, con respecto a la deserción estudiantil detecta la posible deserción estudiantil.

Las características del modelo que permitirá detectar patrones de deserción Estudiantil en la Facultad de Ingeniería de la UNDAC 2023.

2.5. Identificación de Variables

2.5.1. Variables independientes

Técnicas de predicción.

2.5.2. Variables dependientes

Deserción estudiantil.

2.6. Definición operacional de variables e indicadores

Tabla 1 *Definición Operacional de Variables*

VARIABLE	DIMENSIONES
INDEPENDIENTE	- Situación actual.
Técnicas de predicción	- Análisis - Establecer características del modelo
DEPENDIENTE	- Abandono universitario.
Deserción estudiantil	- Factores

Nota. La tabla muestra la definición de variables

CAPITULO III

METODOLOGÍA Y TECNICAS DE INVESTIGACIÓN

3.1. Tipo de investigación

La presente investigación es de tipo básica pues se utilizaron conocimientos de las ciencias aplicadas a fin de detectar patrones de deserción estudiantil mediante aplicación de árboles de decisión

3.2. Nivel de investigación

El presente trabajo de investigación utilizó el nivel cuantitativo, porque se basa en el estudio y análisis de la realidad a través de diferentes procedimientos basados en la medición.

3.3. Métodos de investigación

El método que se usara es deductivo e inductivo se llevó a cabo mediante la observación directa de los hechos en lugares como: aula de clases, secretarias, y diferentes contextos con los estudiantes activos y desertores de la Carrera de Docencia en Informática, por ello se obtuvo información de la fuente, en su contexto original(Hernández Sampieri & Mendoza Torres, 2018).

3.4. Diseño de investigación

El diseño de investigación es descriptivo Se describieron las causas que propician la Deserción Estudiantil se procedió a la identificación de las características socios demográficas, formas de conducta, actitudes de los

estudiantes, para establecer patrones de comportamientos concretos sobre una posible deserción.

3.5. Población y muestra

3.5.1. Población

Según Hernández Sampieri (2018) menciona que: “es un conjunto de individuos que se hallan en un determinado sector y que nos apoya para adquirir la muestra y los resultados”

Estudiante de la Facultad de Ingeniería Matriculados 2023-A 1274.

3.5.2. Muestra

Para Arias, Villasis y Miranda (2016) comento que “Una muestra es parte de una población que tiene características comunes y se selecciona para obtener información sobre variables”.

Como muestra se tomará a los Estudiante de la Facultad de Ingeniería de la Escuela de Ingeniería de Sistemas y Computación Matriculados 2023-A 215.

3.6. Técnicas e instrumentos de recolección de datos

La técnica usada para la recolección de información fue la encuesta diseñada para recolectar información y opiniones sobre la Deserción Estudiantil y las técnicas predicción, los datos recopilados contribuyeron a la determinación de indicadores sobre las variables planteadas. El instrumento utilizado fue el cuestionario las preguntas se construyeron en base a las dimensiones e indicadores obtenidos en la operacionalización de variables, se diseñaron además encuestas online para la toma de datos de los estudiantes desertores y egresados.

3.7. Selección, validación y confiabilidad de los instrumentos de investigación.

Validez

La validez del cuestionario se realizó mediante el análisis bibliográfico de las variables de la hipótesis, y el aporte del tutor de la investigación, experto en la materia.

Para la validez de la información se realizó un análisis estadístico de datos cuantitativos como la edad, el rendimiento académico etc., mediante la aplicación del coeficiente de correlación, cruzando diferentes variables para medir su grado de correlación.

Confiabilidad

Para medir la confiabilidad del cuestionario se administró el mismo instrumento de medición, a distintos grupos (Estudiantes, Egresados, Estudiantes desertores, Docentes), esto permitió medir la consistencia y coherencia de la información mediante la visualización de la escasa variación de los patrones de respuesta entre los grupos aplicados.

3.8. Técnicas de procesamiento y análisis de datos.

El procesamiento de datos diversas funciones, que se detallan a continuación:

- Se elaboró una validación de los datos por el investigador para asegurar que la información suministrada sea correcta y útil para la investigación
- Se clasificó la información de acuerdo a los conglomerados (Estudiantes, egresados, estudiantes desertores, y docentes).
- Se tabularon los datos, presentándolos en tablas con porcentajes y gráficos
- Se analizó e interpretó la información recolectada
- Se presentó la información de manera detallada y resumida

Todo este proceso sirvió para la producción de información significativa para la investigación.

3.9. Tratamiento Estadístico.

- Tabulación de respuestas: Google Forms automáticamente recuperará y organizará las respuestas de la encuesta en una hoja de cálculo de Google Sheets. Puedes utilizar las funciones de hoja de cálculo para tabular los datos y obtener recuentos y porcentajes de las respuestas para cada pregunta.
- Estadísticas descriptivas: Utilizando las funciones de hoja de cálculo de Google Sheets, puedes calcular estadísticas descriptivas básicas, como promedios, medianas y desviaciones estándar, para las preguntas que involucren respuestas numéricas.
- Gráficos y visualizaciones: Google Sheets ofrece diversas opciones para crear gráficos y visualizaciones a partir de los datos recopilados. Puedes utilizar gráficos de barras, gráficos de sectores o gráficos de líneas para representar visualmente las respuestas y facilitar la comprensión de los resultados.

3.10. Orientación ética filosófica y epistémica.

A partir de nuestro proyecto, se ha establecido un sólido marco ético para nuestra investigación. Estamos comprometidos con los principios éticos básicos y el respeto por los derechos y la dignidad de nuestros miembros. Tomaremos las siguientes medidas para garantizar las pautas éticas para la investigación.

CAPITULO IV

RESULTADOS Y DISCUSIÓN

4.1. Descripción del trabajo de campo

El Decanato de la Facultad de Ingeniería nace en el año de 1963 como sucursal de la Universidad Comunal del Centro con el programa académico de Minas, en el año de 1965, el congreso da la ley de creación de la Universidad Nacional Daniel Alcides Carrión con su programa académico de minas, desde esa fecha nuestra facultad ha formado profesionales que en la actualidad ocupan altos cargos en las empresas mineras, conforme iba avanzando el tiempo se fue creando nuevas escuelas de formación profesional llegando hoy en día las escuelas de Ingeniería Metalúrgica, Ingeniería Geológica, Ingeniería Sistemas y Computación, Ingeniería Ambiental e Ingeniería Civil, seguimos dando la oportunidad de buena calidad de profesionales al mercado nacional e internacional.

La Facultad de Ingeniería cuenta con 95 docentes y 1680 estudiantes y 10 Asistentes Administrativos para atender a la comunidad universitaria y al pueblo de Pasco. Fue y es la organizadora de grandes eventos con la finalidad de reforzar la parte académica, en proyección social damos apoyo a las comunidades campesinas, con los municipios se ha firmado convenios con la finalidad de dar posibilidades de ingreso a las jóvenes del 5to. de Secundaria.

Al pueblo de Pasco le decimos que nuestras escuelas son de calidad para formar profesionales de calidad.

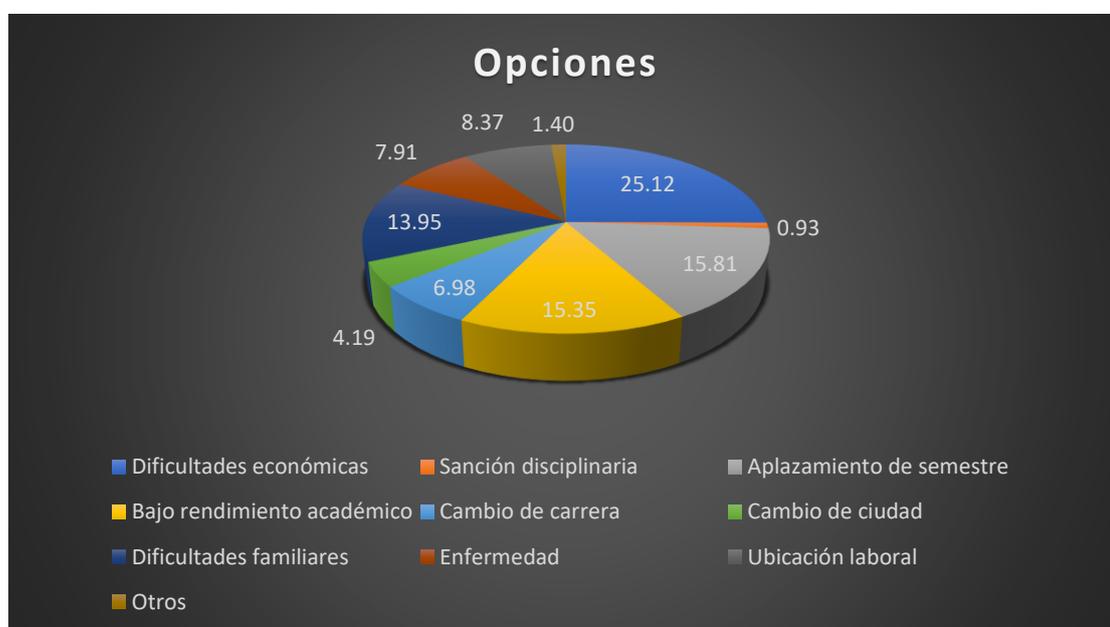
4.2. Presentación, análisis e interpretación de resultados

1. ¿Cuál de las siguientes razones propician la Deserción Estudiantil?

Tabla 2 *Tabla de opciones*

Opciones	Frecuencia	Porcentaje
Dificultades económicas	54	25.12
Sanción disciplinaria	2	0.93
Aplazamiento de semestre	34	15.81
Bajo rendimiento académico	33	15.35
Cambio de carrera	15	6.98
Cambio de ciudad	9	4.19
Dificultades familiares	30	13.95
Enfermedad	17	7.91
Ubicación laboral	18	8.37
Otros	3	1.40
Total	215	100.00

Figura 3 *Gráfico de Opciones*



Análisis e interpretación: Según el punto de vista de los estudiantes las dificultades económicas y familiares, así como el bajo rendimiento propician

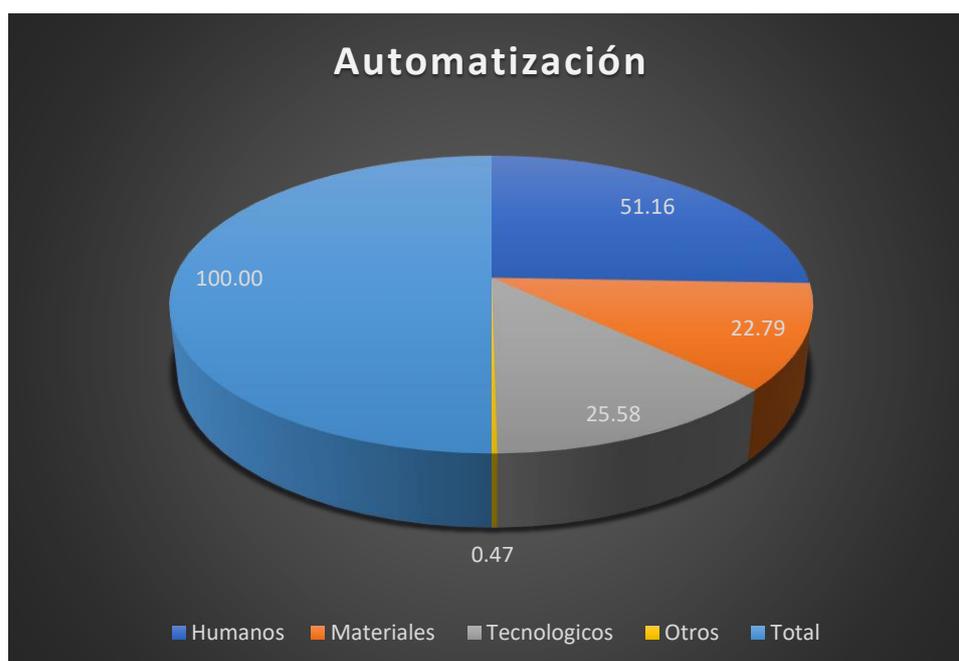
la Deserción Estudiantil, por otra parte, factores como sanción disciplinaria y cambio de ciudad tienen menor incidencia en el abandono estudiantil

2. La automatización de procesos en los estudios sobre Deserción Estudiantil optimizará el uso de recursos

Tabla 3 *Automatización de procesos*

Opciones	Frecuencia	Porcentaje
Humanos	110	51.16
Materiales	49	22.79
Tecnológicos	55	25.58
Otros	1	0.47
Total	215	100.00

Figura 4 *Gráfico de Automatización de procesos*



Análisis e interpretación: Mediante la automatización de procesos sobre Deserción Estudiantil se optimizará el uso de recurso humanos según la moda encontrada en el procesamiento de los datos, en menor cantidad se considera a los recursos tecnológicos y materiales.

3. ¿La técnica de predicción de minería de datos ayudará a detectar patrones de deserción escolar?

Tabla 4 Técnicas de predicción

Opciones	Frecuencia	Porcentaje
Si	200	93.02
No	15	6.98
Total	215	100.00

Figura 5 Gráfico de técnicas de predicción



Análisis e interpretación: El 93.02% de estudiantes considera que las técnicas de predicción, basada en minería de datos ayudarán a detectar patrones de Deserción Estudiantil en la Carrera de la Ingeniería.

4. ¿Cuánto tiempo dedica a sus estudios como parte de su trabajo autónomo (fuera de clases)?

Tabla 5 Dedicación de estudio

Opciones	Frecuencia	Porcentaje
Menos de 1 hora	22	10.23
De 1 a 2 horas	15	6.98
Más de 2 horas	178	82.79
Total	215	100.00

Figura 6 Gráfico de dedicación de estudio



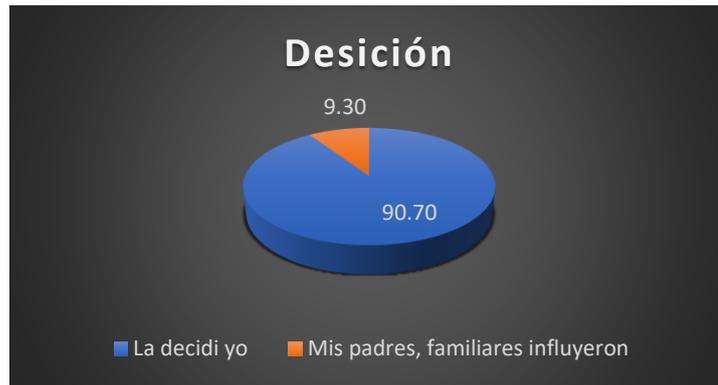
Análisis e interpretación: Los resultados muestran que el 82.79% de la población dedica más de dos horas a su trabajo autónomo, el 10.23% de una a dos horas y el 6.98% de estudiantes dedica menos de una hora.

5. La decisión de estudiar la carrera universitaria en su momento fue:

Tabla 6 Decisión de quien influyo

Opciones	Frecuencia	Porcentaje
La decidí yo	195	90.70
Mis padres, familiares influyeron	20	9.30
Total	215	100.00

Figura 7 Gráfico de decisión de quien influyo



Análisis e interpretación: La decisión de estudiar la Carrera de Ingeniería fue tomada en su gran mayoría por el propio estudiante así lo indica la estadística en donde el 90.70% de los estudiantes expresa su decisión.

6. En cuanto al nivel económico de ingreso en su familia es:

Tabla 7 Nivel económico

Opciones	Frecuencia	Porcentaje
Alto	40	18.60
Medio	130	60.47
Bajo	25	11.63
Muy bajo	20	9.30
Total	215	100.00

Figura 8 Nivel económico



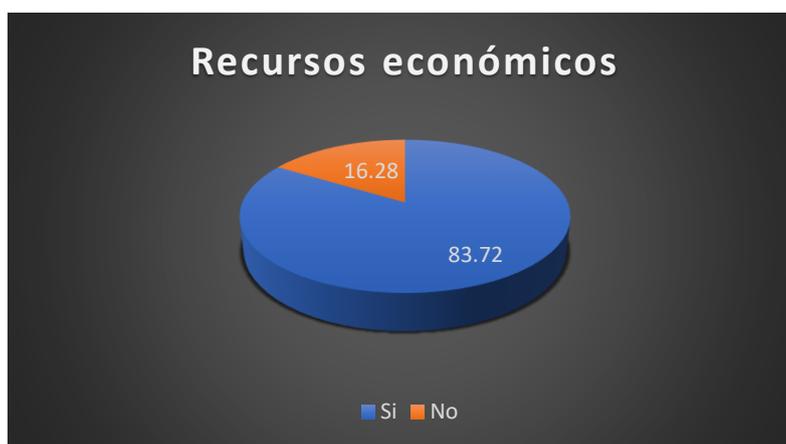
Análisis e interpretación: La mayor parte de población estudiantil de la Carrera de Ingeniería tiene un nivel económico medio, seguido del 11.63% con un nivel bajo y el 9.30% con muy bajo.

7. ¿Dispone de los recursos económicos para costear todos los gastos que se necesita para estudiar?

Tabla 8 *Recurso económico*

Opciones	Frecuencia	Porcentaje
Si	180	83.72
No	35	16.28
Total	215	100.00

Figura 9 *Gráfico de Recurso económico*



Análisis e interpretación: El 16.28% de la población que corresponde a 35 estudiantes no disponen de recursos económicos para costear los gastos de sus estudios

8. Cuenta con algún tipo de ayuda o beca para realizar sus estudios.

Tabla 9 *Beca*

Opciones	Frecuencia	Porcentaje
Si	40	18.60
No	175	81.40
Total	215	100.00

Figura 10 Gráfico de beca



Análisis e interpretación: El 81.40% de los estudiantes, solventa sus estudios con recursos propios, solo el 18.60% recibe ayuda económica para sus estudios.

9. ¿Es alguna asignatura especialmente difícil para Usted?

Tabla 10 Asignatura

Opciones	Frecuencia	Porcentaje
Si	75	34.88
No	140	65.12
Total	215	100.00

Figura 11 Gráfico de asignatura



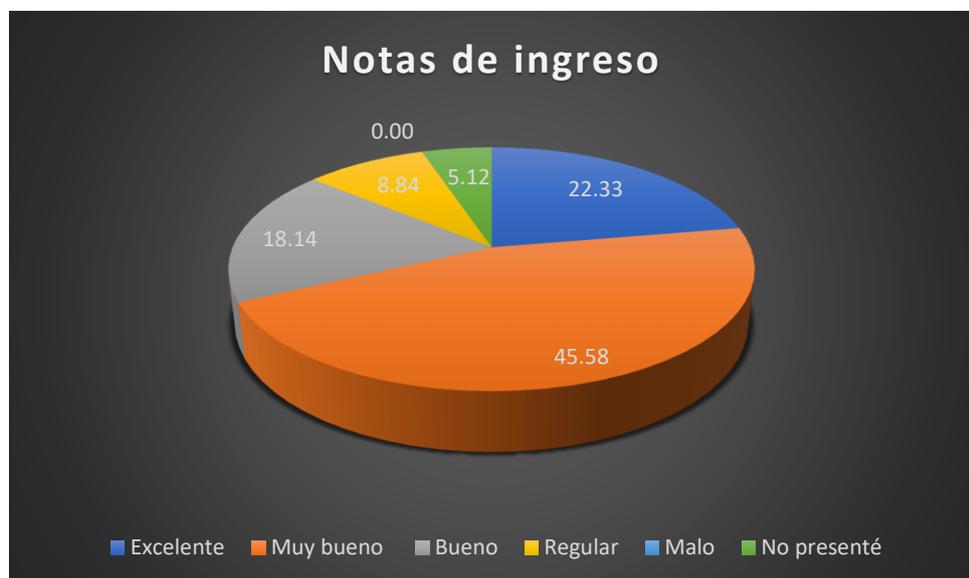
Análisis e interpretación: El 34.88% de la población tiene dificultades en materias como programación, base de datos y física, éstas son consideradas difíciles para los estudiantes.

10. En el examen de admisión de ingreso a la universidad que nota obtuvo.

Tabla 11 *Ingreso a la Universidad*

Opciones	Frecuencia	Porcentaje
Excelente	48	22.33
Muy bueno	98	45.58
Bueno	39	18.14
Regular	19	8.84
Malo	0	0.00
No presenté	11	5.12
Total	215	100.00

Figura 12 *Gráfico de Ingreso a la Universidad*



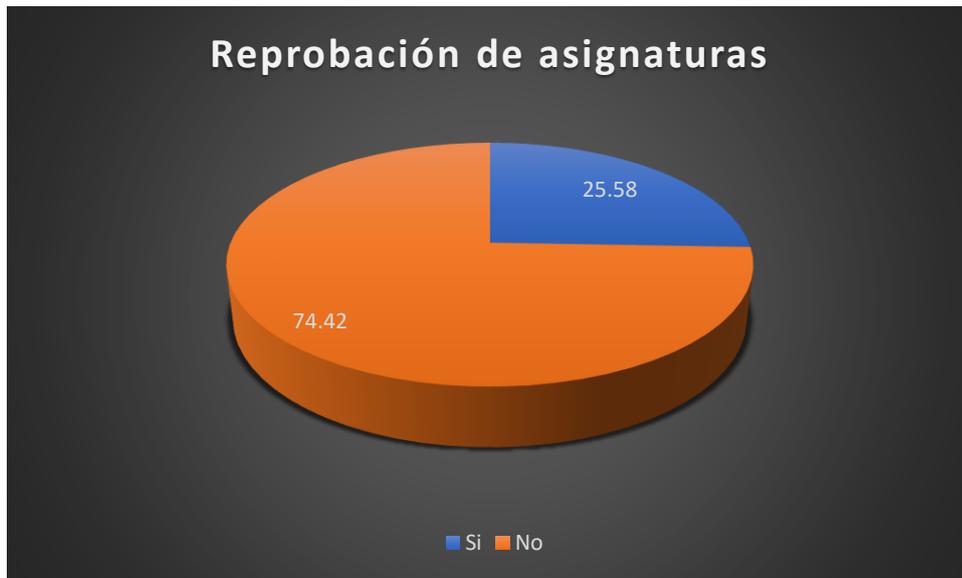
Análisis e interpretación: El 45.58% de los estudiantes obtuvieron “muy buena” calificación en el examen de ingreso a la Universidad; el 18.14% una calificación “buena”, el 8.84% lograron una calificación regular.

11. ¿Ha reprobado alguna asignatura de manera que se haya retrasado en algún periodo académico?

Tabla 12 *Periodo académico*

Opciones	Frecuencia	Porcentaje
Si	55	25.58
No	160	74.42
Total	215	100.00

Figura 13 Gráfico de Periodo académico



Análisis e interpretación: El 74.42% de los estudiantes ha tenido por lo menos una asignatura retrasada en un periodo académico, solo el 25.58% del total de la población no ha tenido retrasos en sus estudios.

4.3. Prueba de Hipótesis

H0: En qué manera las técnicas de predicción no tienen efecto en la detección de patrones de deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023.

H1: En qué manera las técnicas de predicción tienen efecto en la detección de patrones de deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023.

Se utilizó la prueba estadística Chi - Cuadrado, para determinar si existe o no, la relación entre las dos variables. Para lo cual se utilizó la siguiente fórmula:

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Donde:

x2 = Chi – Cuadrado

O = Frecuencias Observadas

E = Frecuencias Esperadas

Nivel de Significancia $\alpha = 0,05$

Nivel de confianza = 95 %

Para la verificación de la hipótesis, se eligió dos preguntas por cada variable para determinar la relación que existe entre: Las Técnicas de Predicción y la Deserción Estudiantil, las preguntas y respuestas seleccionadas fueron:

Pregunta 3:

Opciones	Frecuencia	Porcentaje
Si	200	93.02
No	15	6.98
Total	215	100.00

Pregunta 7:

Opciones	Frecuencia	Porcentaje
Si	180	83.72
No	35	16.28
Total	215	100.00

Pregunta 9:

Opciones	Frecuencia	Porcentaje
Si	75	34.88
No	140	65.12
Total	215	100.00

Pregunta 11:

Opciones	Frecuencia	Porcentaje
Si	55	25.58
No	160	74.42
Total	215	100.00

Prueba Chi Cuadrado

Variable	Alternativas	O	E	(O-E)	(O-E) ²	(O-E) ² /E
VI	Si	200	127.5	72.5	5256.25	41.23
	Si	180	127.5	52.5	2756.25	21.62
	No	15	87.5	-72.5	5256.25	60.07
	No	35	87.5	-52.5	2756.25	31.5
VD	Si	75	127.5	-52.5	2756.25	21.62
	Si	55	127.5	-72.5	5256.25	41.23
	No	140	87.5	52.5	2756.25	31.5
	No	160	87.5	72.5	5256.25	60.07
X²						308.83

El Chi - Cuadrado Calculado es 308.83.

Grados de Libertad

$$gl = (f - 1) (c - 1)$$

$$gl = (4 - 1) (2 - 1)$$

$$gl = 3 * 1$$

$$gl = 3$$

Con tres grados de libertad y 95% de nivel de confianza el valor Chi – Cuadrado teórico o tabular según la tabla de distribución es: 7,81 según se muestra en la siguiente Figura.

Figura 14 Valor Chi - Cuadrado

v/p	0,001	0,0025	0,005	0,01	0,025	0,05
1	10,8274	9,1404	7,8794	6,6349	5,0239	3,8415
2	13,8150	11,9827	10,5965	9,2104	7,3778	5,9915
3	16,2660	14,3202	12,8381	11,3449	9,3484	7,8147

Por tanto:

$$X^2 \text{ calculado} > X^2 \text{ tabular}$$

$$308.83 > 7,8147$$

Debido a que el Chi cuadrado calculado es mayor que el Chi cuadrado tabular se comprueba la hipótesis Alterna que dice H1: En qué manera las

técnicas de predicción tienen efecto en la detección de patrones de deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023, y se rechaza la hipótesis nula (H_0).

4.4. Discusión de resultados

Como resultado calculando la Chi cuadrado calculado es mayor que el Chi cuadrado tabular se comprueba la hipótesis Alternativa que dice H_1 : En qué manera las técnicas de predicción tienen efecto en la detección de patrones de deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023, y se rechaza la hipótesis nula (H_0).

Según Cuji (2016) existe un porcentaje elevado de desconocimiento por parte de la población encuestada sobre el porcentaje de Deserción Estudiantil que se da en la Carrera, así como del número de créditos que logran alcanzar los estudiantes antes de abandonar sus estudios.

CONCLUSIONES

- Las dificultades económicas y familiares, así como el bajo rendimiento son factores predominantes en la Deserción Estudiantil, poca incidencia tiene indicadores como: sanción disciplinaria, cambio de ciudad, ubicación laboral, cambio de carrera, espacio físico para los estudios, nota obtenida en el examen de admisión, influencia de los padres al momento de elegir la Carrera y enfermedad.
- La deserción de los estudiantes es el resultado de la combinación y efectos de distintas variables, las cuales muchas veces son independientes y no se pueden controlar como el nivel económico.
- Aproximadamente la mayor parte de población estudiantil de la Carrera de Ingeniería tiene un nivel económico medio, seguido del 11.63% con un nivel bajo y el 9.30% de los estudiantes tienen un nivel económico entre bajo y muy bajo, solventan sus estudios con recursos propios, conforme a este indicador probablemente 17 de los 70 estudiantes que cursan la carrera de Docencia en Informática serían posibles desertores.

RECOMENDACIONES

- Socialización de las tasas de Deserción Estudiantil entre docentes, estudiantes y autoridades de la Facultad.
- Diseñar un Modelo de predicción basado en técnicas de minería de datos(clasificación), para predecir una posible Deserción Estudiantil.
- Automatización de procesos en los estudios sobre Deserción Estudiantil para optimizar recursos humanos, materiales y tecnológicos.

REFERENCIAS BIBLIOGRÁFICAS

- Alania, P. (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."*
- Camborda, M. (2014). Aplicación de árboles de decisión para la predicción del rendimiento académico de los estudiantes de los primeros ciclos de la Carrera de Ingeniería Civil de la Universidad Continental [Universidad Nacional del Centro del Perú]. In *Universidad Nacional del Centro del Perú*. <http://repositorio.uncp.edu.pe/handle/20.500.12894/1477>
- Cid, A. (2015). Propuesta de un modelo de estimación del abandono universitario [UNIVERSIDAD REY JUAN CARLOS]. In *Cccom.Urjc.Es*. http://www.cccom.urjc.es/informacion/estatutos/archivos/estatutos_urjc.pdf
- Cuji, B. (2016). Las técnicas de predicción y su incidencia en la detección de patrones de deserción estudiantil en la carrera de docencia en informática de la Facultad DE Ciencias Humanas y de la Educación de la Universidad Técnica de Ambato [UNIVERSIDAD TÉCNICA DE AMBATO]. In *"Las Técnicas De Predicción Y Su Incidencia En La Detección De Patrones De Deserción Estudiantil En La Carrera De Docencia En Informática De La Facultad De Ciencias Humanas Y De La Educación De La Universidad Técnica De Ambato"*. <http://repositorio.uta.edu.ec/jspui/handle/123456789/3112>
- Gamarra, D., Matos, R., & Yupanqui, M. (2018). Detección de patrones de éxito en estudios universitarios de la Universidad Continental. *Apuntes de Ciencia & Sociedad*, 08(01). <https://doi.org/10.18259/acs.2018005>
- García, E. (2020). Detección de patrones de deserción estudiantil mediante aplicación de Árboles de Decisión C4.5 en el IESTP "Señor de Chocán" de Querecotillo [UNIVERSIDAD CÉSAR VALLEJO]. In *Universidad Andina del Cusco*. http://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/47102/Gutierrez_RS-SD.pdf?sequence=1&isAllowed=y

- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2018). *Metodología de la Investigación* (S. A. D. C. . McGRAW-HILL / INTERAMERICANA EDITORES (ed.); Sexta).
- Hernández Sampieri, R., & Mendoza Torres, C. (2018). *Metodología de la Investigación*.
- Peña Torres, C., Sola Martínez, T., & Lopez Nuñez, J. A. L. (2017). Análisis de los factores de la deserción estudiantil en una facultad de las unidades tecnológicas de Santander, Bucaramanga, Colombia. *Notandum*, 141–154. <https://doi.org/10.4025/notandum.44.12>
- Piscoya Ordoñez, L. (2016). “*Aplicación de técnicas de minería de datos para predecir la deserción estudiantil en la educación básica regular en la Región de Lambayeque.*” UNIVERSIDAD SEÑOR DE SIPÁN.
- Timarán Pereira, R., Caicedo Zambrano, J., & Hidalgo Troya, A. (2019). Árboles de decisión para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas Saber 11°. *Revista de Investigación, Desarrollo e Innovación*, 9(2), 363–378. <https://doi.org/10.19053/20278306.v9.n2.2019.9184>
- Zarate, A. J. (2019). Reconocimiento De Patrones De Deserción De Estudiantes Universitarios Basados En Modelos De Clasificación [Universidad Nacional San Agustín de Arequipa]. In *Universidad Nacional San Agustín de Arequipa, Perú*. <http://repositorio.unsa.edu.pe/handle/UNSA/9419>

ANEXOS

INSTRUMENTOS DE RECOLECCIÓN DE DATOS

Instrucción: Responder sinceramente a las preguntas con alternativas.

1. ¿Cuál de las siguientes razones propician la Deserción Estudiantil?
 - a. Dificultades económicas
 - b. Sanción disciplinaria
 - c. Aplazamiento de semestre
 - d. Bajo rendimiento académico
 - e. Cambio de carrera
 - f. Cambio de ciudad
 - g. Dificultades familiares
 - h. Enfermedad
 - i. Ubicación laboral
 - j. Otros
2. La automatización de procesos en los estudios sobre Deserción Estudiantil optimizará el uso de recursos
 - a. Humanos
 - b. Materiales
 - c. Tecnológicos
 - d. Otros
3. ¿La técnica de predicción de minería de datos ayudará a detectar patrones de deserción escolar?
 - a. Si
 - b. No
4. ¿Cuánto tiempo dedica a sus estudios como parte de su trabajo autónomo (fuera de clases)?
 - a. Menos de 1 hora
 - b. De 1 a 2 horas
 - c. Más de 2 horas

5. La decisión de estudiar la carrera universitaria en su momento fue académico
- a. La decidí yo
 - b. Mis padres, familiares influyeron
6. En cuanto al nivel económico de ingreso en su familia es:
- a. Alto
 - b. Medio
 - c. Bajo
 - d. Muy bajo
7. ¿Dispone de los recursos económicos para costear todos los gastos que se necesita para estudiar?
- a. Si
 - b. No
8. Cuenta con algún tipo de ayuda o beca para realizar sus estudios.
- a. Si
 - b. No
9. ¿Es alguna asignatura especialmente difícil para Usted?
- a. Si
 - b. No
10. En el examen de admisión de ingreso a la universidad que nota obtuvo
- a. Excelente
 - b. Muy bueno
 - c. Bueno
 - d. Regular
 - e. Malo
 - f. No presenté
11. ¿Ha reprobado alguna asignatura de manera que se haya retrasado en algún periodo académico?
- a. Si
 - b. No

PROPUESTA

1. Comprensión de Negocio

Esta fase inicial se enfoca en entender los objetivos y requerimiento del proyecto convirtiendo esto en la definición del problema de minería de datos.

a. Determinación de objetivos de negocio

La Universidad Nacional Daniel Alcides Carrión es una comunidad integrada por profesores, estudiantes, y graduados. Se dedica al estudio, la investigación, la educación, la difusión del saber, la cultura, y a la extensión y proyección social.

Tiene como objetivos, entre otros, de docencia y de gestión integral.

b. Evaluación de la situación

En el ámbito universitario, la UNDAC adolece de la problemática de una tasa significativa de deserción estudiantil por diferentes motivos, no todos conocidos, lo que ocasiona que se enfrente a esa situación cada vez que se presente, afectando la planificación realizada por cada escuela profesional.

Los recursos inicialmente disponibles son los siguientes:

- Materiales: Weka versión 3.8.3 como herramienta de Minería de datos y SQL Server Business Intelligence para el ETL.
- Humanos: El autor de la investigación, El Jefe de la Oficina de Informática
- Objetivo de minería de datos: Determinar las técnicas de predicción tienen efecto en la detección de patrones de deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023
- Objetivos específicos en la investigación:

Estudiar la situación actual en la Facultad de ingeniería de la UNDAC 2023, con respecto a la deserción estudiantil.

Analizar las técnicas de predicción se aplica actualmente en la Facultad de ingeniería de la UNDAC 2023, con respecto a la deserción estudiantil para detectar la posible deserción estudiantil.

Establecer las características del modelo que permitirá detectar patrones de deserción Estudiantil en la Facultad de Ingeniería de la UNDAC 2023.

Conocer estos objetivos permitirá determinar de una mejor manera la planificación en la apertura del año académico y distribución de aulas y secciones de la escuela profesional.

2. Comprensión de data

a. Recolección de data y descripción

La primera etapa de metodología CRISP-DM describe la alineación de los objetivos del proyecto con la investigación, por ende, el levantamiento de la información fue recolectada mediante la base de datos de la UNDAC 2023 que comprende del periodo 2023 periodo A, dicha información se alinea con los objetivos del proyecto y la información adquirida servirá para analizarla y poder interpretar el modelo que se desea diseñar.

Variable	Descripción
Semestre	Periodo de estudio
Asignaturas	Cursos
Notas_aprobadas	En qué mes fue atendido
Notas_desaprobados	Código de consultorio de rehabilitación
Retirados	Identificación del paciente

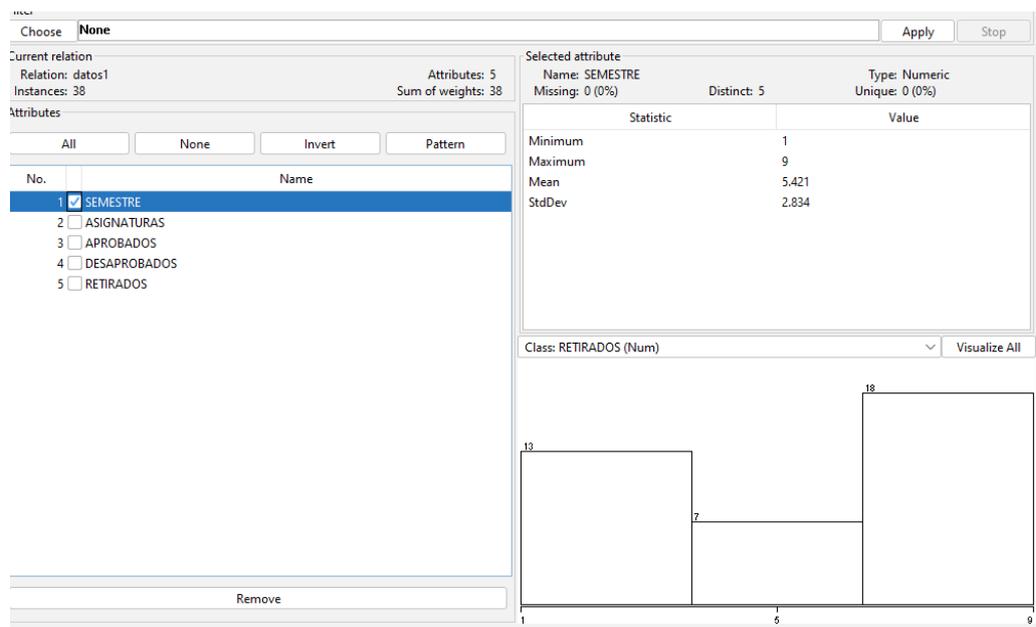
3. Preparación de data

Para preparar la data se lleva una descripción estadística de todas las etiquetas para determinarla estadísticas descriptivas, también se muestran los gráficos de la data para verificar la igualdad de partición en la clase a predecir,

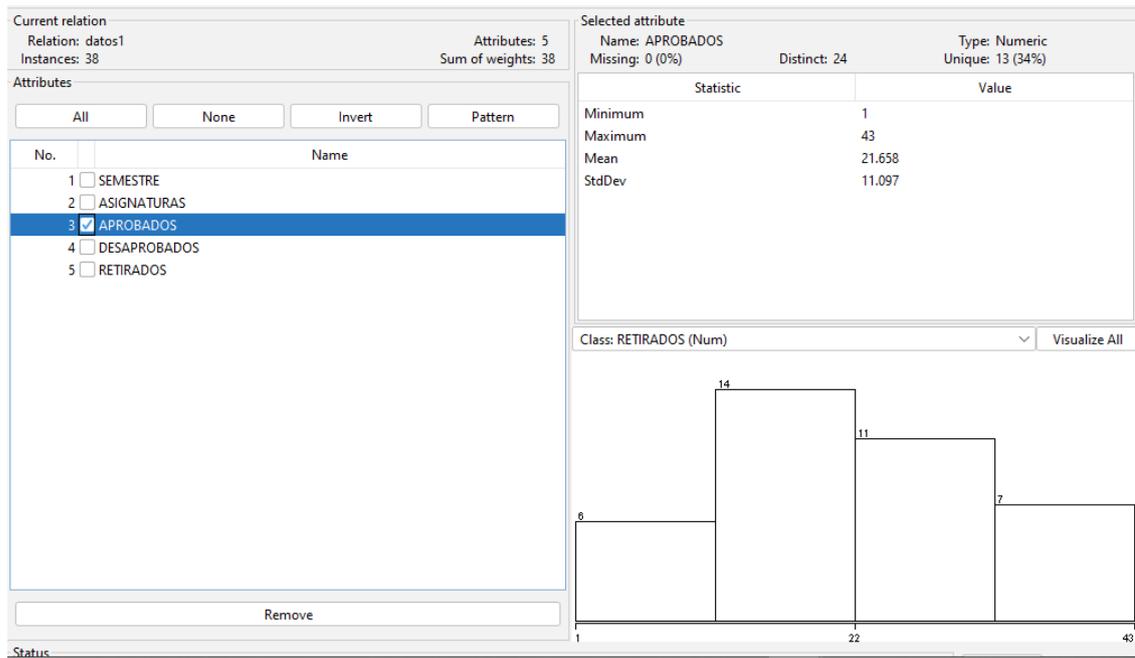
un factor clave en esta fase es el sobre muestreo en la clase minoritaria y se identifica en la observación de la data.

SEMESTRE	ASIGNATURAS	APROBADOS		DESAPROBADOS		RETIRADOS		TOTAL	
		Nº	%	Nº	%	Nº	%	Nº	%
1	Matemática Básica	34	79.07	4	9.30	5	11.63	43	100.00
1	Comunicación Oral y Escrita	41	97.62	0	0.00	1	2.38	42	100.00
1	Matemática Superior	21	40.38	20	38.46	11	21.15	52	100.00
1	Física General	27	60.00	14	31.11	4	8.89	45	100.00
1	Cálculo Diferencial	12	25.00	30	62.50	6	12.50	48	100.00
1	Algoritmos	10	18.52	37	68.52	7	12.96	54	100.00
3	Ecología	41	93.18	0	0.00	3	6.82	44	100.00
3	Idioma: Inglés Básico	39	86.67	6	13.33	0	0.00	45	100.00
3	Educación Física	43	97.73	0	0.00	1	2.27	44	100.00
3	Matemática Básica II	27	54.00	21	42.00	2	4.00	50	100.00
3	Cálculo Multivariable	22	57.89	12	31.58	4	10.53	38	100.00
3	Física I	23	38.33	34	56.67	3	5.00	60	100.00
3	Lenguaje de Programación II	10	34.48	19	65.52	0	0.00	29	100.00
5	Teoría de Sistemas	24	85.71	0	0.00	4	14.29	28	100.00
5	Investigación de Operaciones II	15	75.00	3	15.00	2	10.00	20	100.00
5	Electrónica II	19	82.61	1	4.35	3	13.04	23	100.00
5	Contabilidad y Costos	14	100.00	0	0.00	0	0.00	14	100.00
5	Administración y Calidad Total	34	94.44	1	2.78	1	2.78	36	100.00
5	Arquitectura del Computador	35	97.22	0	0.00	1	2.78	36	100.00
5	Sistemas de Bases I	20	40.82	29	59.18	0	0.00	49	100.00
7	Planeamiento y Dirección Estratégica	16	53.33	10	33.33	4	13.33	30	100.00

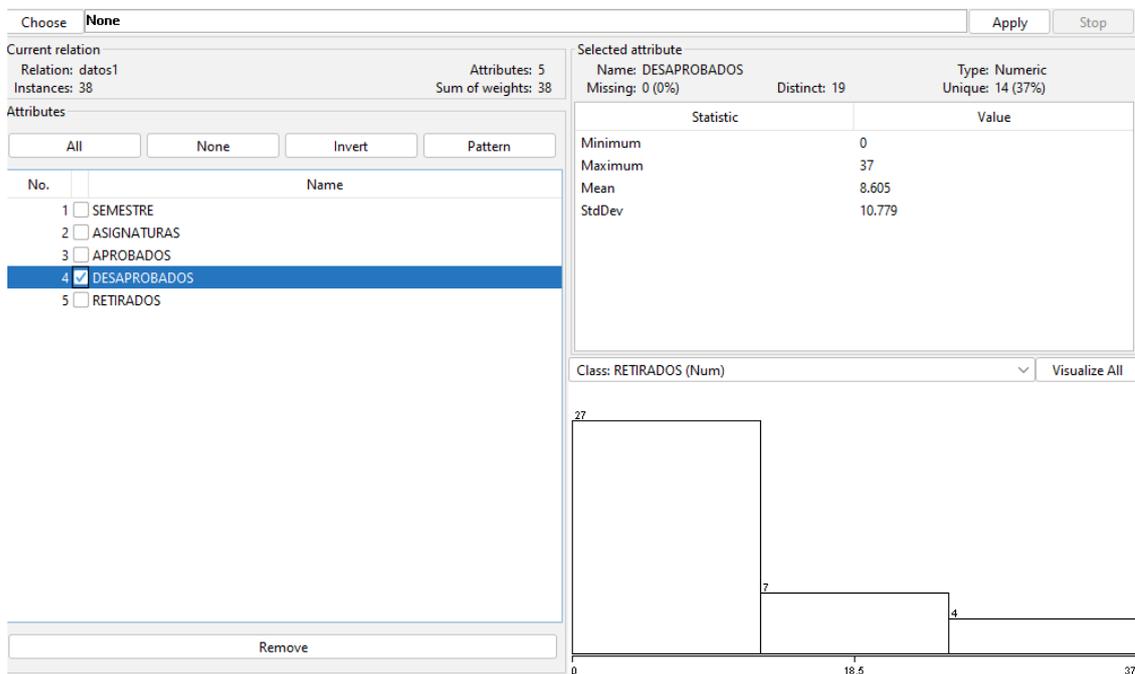
Por semestre:



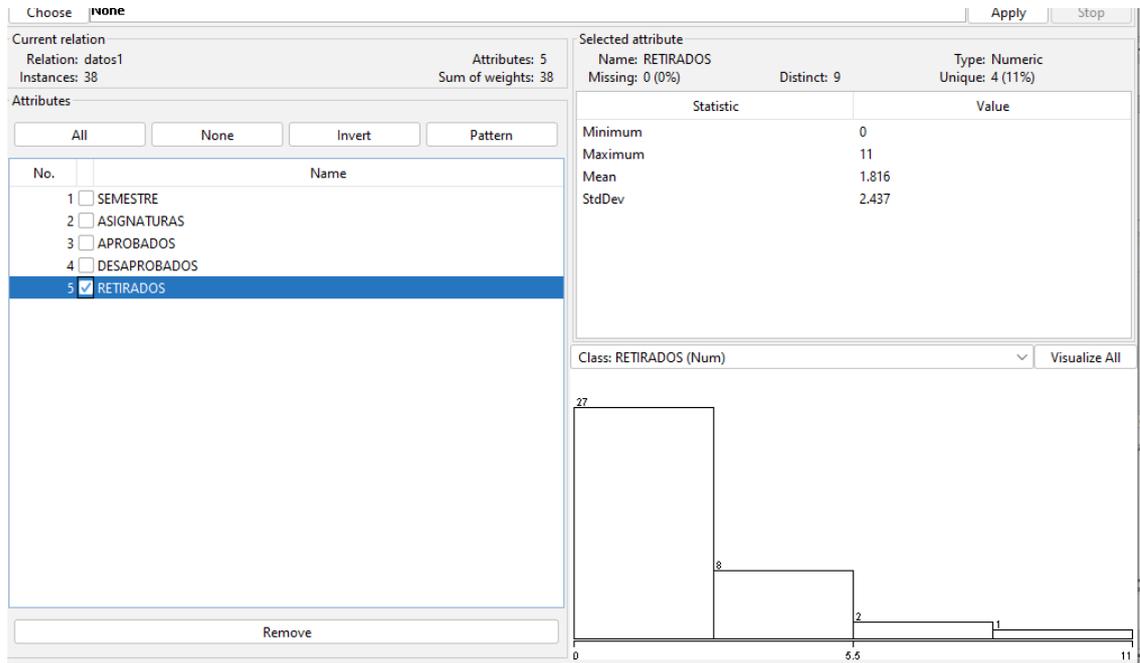
Aprobados



Desaprobados



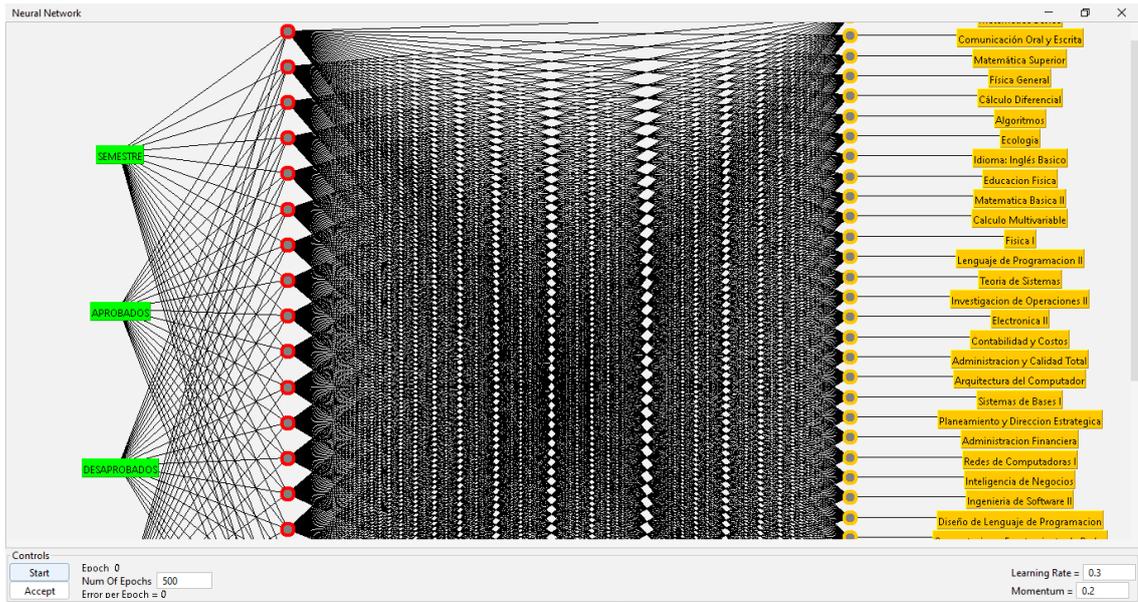
Desertaron



4. Modelado

Construcción del Modelo Perceptrón Multicapa, se grafica con 3 capas.

- La primera capa corresponde a los valores de entradas que son las 5 variables analizadas y procesadas en las fases de la metodología CRISP DM.
- La segunda capa corresponde a la capa oculta, que es parte del proceso en las redes neuronales para realizar una retropropagación y el modelo sea óptimo, esta capa consta de 2 neuronas.
- Por último, la capa deseada a predecir es de 1 neurona cuyo valor es 0,1.



La imagen describe, un modelo secuencial de un clasificador por capa, la función Dense describe las variables de 1 capa, en la primera capa describe el valor 'units' como las neuronas de la primera capa, la capa de entrada consta de 4 etiquetas, la inicialización de pesos mediante la derivada 'kernel' es de rango de -0.3 a +0.2 debido que presenta mayor valor de error cuadrático medio y es óptimo para realizar una convergencia

5. Evaluación

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer
Choose SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode
 Use training set
 Supplied test set Set...
 Percentage split % 66
 Classes to clusters evaluation (Num) RETIRADOS
 Store clusters for visualization

Ignore attributes
Start Stop

Result list (right-click for options)
23:06:24 - SimpleKMeans

Clusterer output

Initial starting points (random):
Cluster 0: 3, 'Matemática Básica II', 27, 21, 2
Cluster 1: 9, 'Prácticas Pre Profesionales I', 28, 9, 0

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data (38.0)	Cluster# 0 (14.0)	Cluster# 1 (24.0)
SEMESTRE	5.4211	2.2857	7.25
ASIGNATURAS	Matemática Básica	Matemática Básica	Teoría de Sistemas
APROBADOS	21.6579	26.4286	18.875
DESAPROBADOS	8.6053	16.1429	4.2083
RETIRADOS	1.8158	3.3571	0.9167

Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	14 (37%)
1	24 (63%)

MATRIZ DE CONSISTENCIA

Tema: “Técnicas de predicción y su efecto en la detección de patrones de deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023”

PROBLEMA GENERAL	OBJETIVO GENERAL	HIPÓTESIS GENERAL	VARIABLE INDEPENDIENTE	DIMENSIÓN	DISEÑO	POBLACIÓN Y MUESTRA
¿En qué manera las técnicas de predicción tienen efecto en la detección de patrones de deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023?	Determinar las técnicas de predicción tienen efecto en la detección de patrones de deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023.	Las técnicas de predicción tendrán efecto en la detección de patrones de deserción estudiantil en la Facultad de Ingeniería de la UNDAC 2023.	Técnicas de predicción	- Situación actual - Análisis - Establecer características del modelo	Diseño: Descriptivo Tipo de Investigación Básica	POBLACIÓN Estudiante de la Facultad de Ingeniería Matriculados 2023-A 1274. MUESTRA Se trabajará con 215 alumnos.
PROBLEMA ESPECÍFICO	OBJETIVO ESPECÍFICO	HIPÓTESIS ESPECÍFICA	VARIABLE DEPENDIENTE	DIMENSIÓN	MÉTODO DE INVESTIGACIÓN	TÉCNICAS - INSTRUMENTOS
¿Cuál es la situación actual en la Facultad de Ingeniería de la UNDAC 2023, con respecto a la deserción estudiantil?	Estudiar la situación actual en la Facultad de Ingeniería de la UNDAC 2023, con respecto a la deserción estudiantil.	El estudio actual en la Facultad de Ingeniería de la UNDAC 2023, ayudara a ver la deserción estudiantil	Deserción estudiantil.	- Abandono universitario - Factores	Método Análítica, inductiva Enfoque Cuantitativo	Técnicas: - Cuestionario.
¿Qué técnicas de predicción se aplica actualmente en la Facultad de Ingeniería de la UNDAC 2023, con respecto a la deserción estudiantil para detectar la posible deserción estudiantil?	Analizar las técnicas de predicción se aplica actualmente en la Facultad de Ingeniería de la UNDAC 2023, con respecto a la deserción estudiantil para detectar la posible deserción estudiantil.	Las técnicas de predicción que se aplican actualmente en la Facultad de Ingeniería de la UNDAC 2023, con respecto a la deserción estudiantil detecta la posible deserción estudiantil.				
¿Es necesario establecer las características del modelo que permitirá detectar patrones de	Establecer las características del modelo que permitirá detectar patrones de deserción Estudiantil	Las características del modelo que				

deserción
Estudiantil
en la
Facultad de
Ingeniería
de la
UNDAC
2023?

en la
Facultad de
Ingeniería
de la
UNDAC
2023.

permitirá
detectar
patrones de
deserción
Estudiantil
en la
Facultad de
Ingeniería
de la
UNDAC
2023.
