



Universidad
Señor de Sipán

**FACULTAD DE INGENIERÍA, ARQUITECTURA Y
URBANISMO**

**ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS
TESIS**

**Implementación de un método de clasificación para
detectar la deserción de estudiantes de la carrera de
Ingeniería de Industrias Alimentarias de una
universidad nacional peruana basado en aprendizaje de
maquina**

**PARA OPTAR EL TÍTULO PROFESIONAL
DE INGENIERO DE SISTEMAS**

Autor(a) (es):

**Bach. Campos Barrera Sandro Paul
ORCID: <https://orcid.org/0000-0001-7555-5105>**

**Bach. Pastor Oliva Cesar Augusto
ORCID: <https://orcid.org/0000-0002-3386-7466>**

Asesor(a):

**Mg. Mejia Cabrera Heber Ivan
ORCID: <https://orcid.org/0000-0002-0007-0928>**

Línea de Investigación

**Ciencias de la información como herramientas multidisciplinares y
estratégicas en el contexto industrial y de organizaciones**

Sub Línea de Investigación

**Informática y transformación digital en el contexto industrial y
organizacional**

Pimentel – Perú 2023

**IMPLEMENTACIÓN DE UN MÉTODO DE CLASIFICACIÓN PARA DETECTAR
LA DESERCIÓN DE ESTUDIANTES DE LA CARRERA DE INGENIERÍA DE
INDUSTRIAS ALIMENTARIAS DE UNA UNIVERSIDAD NACIONAL PERUANA
BASADO EN APRENDIZAJE DE MAQUINA**

Aprobación del jurado

ING. BRAVO RUIZ JAIME ARTURO
Presidente de Jurado de tesis

ING. BANCES SAAVEDRA DAVID ENRIQUE
Secretario del Jurado de Tesis

ING. ATALAYA URRUTIA CARLOS WILLIAM
Vocal del Jurado de Tesis







DECLARACIÓN JURADA DE ORIGINALIDAD

Quienes suscribimos la **DECLARACIÓN JURADA**, somos del Programa de Estudios de Ingeniería de Sistemas de la Universidad Señor de Sipán S.A.C, declaramos bajo juramento que somos autores del trabajo titulado:

IMPLEMENTACIÓN DE UN MÉTODO DE CLASIFICACIÓN PARA DETECTAR LA DESERCIÓN DE ESTUDIANTES DE LA CARRERA DE INGENIERÍA DE INDUSTRIAS ALIMENTARIAS DE UNA UNIVERSIDAD NACIONAL PERUANA BASADO EN APRENDIZAJE DE MAQUINA

El texto de mi trabajo de investigación responde y respeta lo indicado en el Código de Ética del Comité Institucional de Ética en Investigación de la Universidad Señor de Sipán (CIEI USS) conforme a los principios y lineamientos detallados en dicho documento, en relación a las citas y referencias bibliográficas, respetando al derecho de propiedad intelectual, por lo cual informo que la investigación cumple con ser inédito, original y autentico.

En virtud de lo antes mencionado, firman:

Campos Barrera Sandro Paul	DNI: 44997496	 
Pastor Oliva Cesar Augusto	DNI: 42364577	 

Pimentel, 19 de octubre de 2023

Dedicatorias

En primer lugar, a Dios por brindarme la vida y la salud para seguir cumpliendo mis objetivos trazados. A mis padres por compartir sus esfuerzos y entusiasmo conmigo para recordarme que aun, el fracaso es parte del éxito. A mi esposa y mis hijos que son el motivo para culminar con éxito mi carrera profesional.

Sandro Paul Campos Barrera

La presente investigación está dedicada a Dios, por ser mi guía y fortaleza para continuar con la culminación del presente proyecto. A mi compañera de vida Sarita, quien con su paciencia y cariño permanentemente ha sido de vital importancia en todo el proceso de mi formación profesional.

César Augusto Pastor Oliva

Agradecimientos

A mis padres y hermanos, esposa e hijos y amistades que con su apoyo estoy logrando un objetivo más, parte de mi desarrollo humanístico y profesional. A mis asesores por sus conocimientos brindados, tiempo y paciencia que han sido de gran ayuda en este proyecto.

Sandro Paul Campos Barrera

Mi gratitud a Dios por ser mi fortaleza en momentos de debilidad, mi guía y gran compañía a lo largo de mi carrera personal y profesional.

César Augusto Pastor Oliva

Resumen

La deserción estudiantil es un problema creciente en Latinoamérica, con un aumento considerable en los últimos años. Esto ha tenido un impacto económico significativo, con pérdidas que alcanzan el 26% del gasto público en educación. En Perú, entre 40,000 y 50,000 universitarios abandonan sus estudios anualmente. Asimismo, la pandemia de COVID-19 agravó la deserción, pasando del 12% en 2019 al 18.6% en 2020. Para abordar este problema, se han realizado estudios que utilizan algoritmos de clasificación, como Máquina de Vectores de Soporte, Naive Bayes, Perceptrón Multicapa y Árboles de Decisión, para predecir la deserción estudiantil. Aún ante la efectividad de los métodos, la creciente deserción requería una mayor precisión. Siendo así, esta investigación propone implementar un método de clasificación mejorado utilizando algoritmos de Aprendizaje de Máquina como Random Forest, Naive Bayes, J48, RandomTree y Support Vector Machine. Se mejoró la calidad de los datos mediante filtros supervisados y no supervisados, como ReplaceMissingValues para completar datos faltantes, SpreadSubSample, Resample y Class Balancer para equilibrar clases y validación cruzada para evaluar el desempeño de cada algoritmo propuesto. El método propuesto, junto con el algoritmo Support Vector Machine, logró una precisión del 98.88% al procesar una muestra de 358 instancias. Se demostró que eliminar datos faltantes puede reducir el rendimiento de los algoritmos clasificadores, y se usó el filtro ReplaceMissingValues para llenar los valores faltantes con la media aritmética. Este enfoque muestra un prometedor avance en la predicción de la deserción estudiantil y puede ser una herramienta valiosa para las instituciones educativas en la lucha contra este problema

Palabras Clave: Deserción estudiantil, algoritmos de predicción, métodos de clasificación, aprendizaje de máquina, filtros supervisados, filtros de balanceo, RandomForest.

Abstract

Student dropout is a growing problem in Latin America, with a considerable increase in recent years. This has had a significant economic impact, with losses reaching 26% of public spending on education. In Peru, between 40,000 and 50,000 university students drop out of school annually. Likewise, the COVID-19 pandemic exacerbated dropout, rising from 12% in 2019 to 18.6% in 2020. To address this problem, studies using classification algorithms, such as Support Vector Machine, Naive Bayes, Multilayer Perceptron, and Decision Trees, have been conducted to predict student dropout. Even in the face of the effectiveness of the methods, the increasing attrition required higher accuracy. Being so, this research proposes to implement an improved classification method using Machine Learning algorithms such as Random Forest, Naive Bayes, J48, RandomTree and Support Vector Machine. Data quality was improved using supervised and unsupervised filters, such as ReplaceMissingValues to fill in missing data, SpreadSubSample, Resample and Class Balancer to balance classes and cross validation to evaluate the model. The proposed method, together with the Support Vector Machine algorithm, achieved an accuracy of 98.88% when processing a sample of 358 instances. It was shown that removing missing data can reduce the performance of classifier algorithms, and the ReplaceMissingValues filter was used to fill the missing values with the arithmetic mean. This approach shows a promising advance in student dropout prediction and can be a valuable tool for educational institutions in combating this problem.

Keywords: Student dropout, prediction algorithms, classification methods, machine Learning, supervised filters, rolling filters, RandomForest.

Indice

I. INTRODUCCIÓN	11
1.1 Realidad Problemática.	11
1.2 Formulación del Problema.	27
1.3 Hipótesis.	27
1.4 Objetivos.	27
1.4.1 Objetivo general.	27
1.4.2 Objetivos específicos.	27
1.5 Teorías relacionadas al tema.	28
II. MATERIAL Y MÉTODO	50
II.1. Tipo y Diseño de Investigación.	50
II.2. Población y muestra.	50
II.3. Variables, Operacionalización.	52
II.4. Técnicas e instrumentos de recolección de datos, validez y confiabilidad.	54
II.5. Procedimiento de análisis de datos.	54
II.6. Criterios éticos.	57
III. RESULTADOS.	59
III.1. Resultados en Tablas y Figuras.	59
III.2. Discusión de resultados.	66
III.3. Aporte práctico.	68

IV. CONCLUSIONES Y RECOMENDACIONES	130
IV.1. Conclusiones.	130
IV.2. Recomendaciones.	131
REFERENCIAS.	132
• ANEXOS.	137

I. INTRODUCCIÓN

1.1 Realidad Problemática.

La deserción estudiantil se define como un hecho de que los estudiantes no sigan la trayectoria de un programa académico, esto se puede dar por el retiro de la carrera profesional o por demorar en terminarla, debido a los cursos desaprobados o retiros temporales, en algunos de los casos también el cambio de carrera profesional, debido a que el estudiante no tiene una idea madura sobre lo que quieren estudiar o porque el servicio que brinda la institución no es la adecuada [1].

El abandono de clases de estudiantes universitarios hoy en día es una problemática que enfrentan las universidades, ciudades y países, quienes sufren las consecuencias sociales de este fenómeno [2]. El nivel universitario en países desarrollados se ha visto gravemente afectados a causa de la pandemia, más aún en los países en desarrollo donde muchas escuelas superiores han sentido notablemente la ausencia de estudiantes por no contar con recursos digitales, siendo estos una amenaza a su economía familiar [3].

Asimismo, en el Caribe y Latino América, la deserción de estudiantes se ha ido incrementando, debido a factores que juegan un rol importante en el proceso activo del estudiante para concluir satisfactoriamente su carrera profesional[4].

Los países latinoamericanos están buscando políticas que logren identificar los factores puntuales que motivan la interrupción educativa, de tal manera que se desarrollen gestiones propias de solución para cada región, economía social y políticas de cada país[5].

En ciertos países de Latinoamérica, el porcentaje de deserción de estudiantes universitarios al año es más del 53%; asimismo, se ve reflejado el impacto económico anual en pérdidas aproximadas al 26% del gasto público en el sector educación [6].

En la actualidad el abandono académico es un problema múltiple que acarrear la mayoría de las instituciones de educación superior tanto en el Perú como a nivel mundial.

La deserción académica tiene muchas causas y direccionan a resultados como el fracaso de los estudiantes y expectativas no alcanzadas con niveles de logro muy bajos afectando a la familia y a sociedad[2].

En Perú, según el portal Logros, la tasa de deserción universitaria es del 17%. Cerca de 40 a 50 mil estudiantes universitarios abandonan la universidad cada año, esto conlleva a pérdidas económicas para las familias, seguido de la desilusión que conlleva tanto para los estudiantes y sus familias. Del porcentaje total de la deserción estudiantil universitaria, el setenta por ciento comprende a alumnos de universidades particulares, y el treinta por ciento restantes a universidades públicas.

Es importante acotar, que aproximadamente 174.000 estudiantes en Perú desertaron de sus estudios universitarios en el año 2020 (ciclo académico 2020 I), estas cifras se les atribuyen a causas de la pandemia COVID-19. Por lo cual, los resultados ascendieron a un 18,6% de 955.000 universitarios en territorio nacional, seis puntos porcentuales más que los registrados en el año 2019 que fue de 12%. Esta situación varía según la gestión de cada universidad: en las nacionales, la tasa de deserción llegó en ese año (ciclo 2020-I) al 9,85%; y en las particulares fue del 22,5%. El funcionario Jorge Mori, quien se desempeña como director general de la Dirección de Educación Superior Universitaria del Minedu (Ministerio de Educación), declaró que la deserción que se viene dando en las universidades nacionales y privadas no es un problema solo económico sino también de índole multicausal, en vista de que muchos de estos problemas provienen del núcleo familiar, orientación vocacional y académico[7].

Asimismo, el Minedu designó un presupuesto en materia de conectividad de aproximadamente 30,8 millones de soles en la entrega de chips con datos de internet para las universidades nacionales, con esta acción fueron 91.012 los beneficiarios, distribuidos por regiones, entre las cuales tenemos: Cajamarca 4.522 beneficiarios (U. N. de Jaén, U. N. de Cajamarca, U. N. Autónoma de Chota), Lambayeque 4.954 beneficiarios (U. N. Pedro Ruíz Gallo) Lima 15.158 beneficiarios (U. N. de Barranca, U. N. Agraria La Molina, U. N.

de Cañete, U. N. de Ingeniería, U. N. Tecnológica de Lima Sur, U. N. Tecnológica de Lima Sur, U. N. José Faustino Sánchez Carrión, U. N. Federico Villarreal, U. N. de Educación Enrique Guzmán y Valle, U. N. Mayor de San Marcos[7].

Del mismo modo, las universidades particulares han doblado esfuerzos para asistir a sus estudiantes para la continuidad de sus estudios, mediante estrategias como otorgamiento de becas, descuentos en matrículas y pensiones, extensión de plazo para pago de deudas, entre otras medidas. Al igual que las universidades nacionales, las particulares también se embarcaron en el camino de facilitar a sus estudiantes herramientas tecnológicas como la entrega de planes de internet, compra de software para las clases virtuales, etc.

Por su parte, el Minedu en el año 2020 de la mano con las universidades nacionales y particulares, promovieron la implementación de la educación remota con el propósito de garantizar la continuidad del servicio educativo. En ese mismo año de las 144 universidades habilitadas en Perú 137 (86 particulares y 51 nacionales) comenzaron los estudios para el ciclo académico 2020-I [7].

Los algoritmos de árboles de decisión C4.5 (J48) y RandomTree tiene una alta precisión para la identificación de patrones de comportamiento en la deserción de estudiantes, así como la predicción de estos.

Por otro lado [8] utilizó las técnicas predictivas de minería de datos tales como series de tiempo y redes neuronales para estos tipos de proyectos donde se requiere identificar con exactitud los posibles factores de deserción de los estudiantes.

CRISP-DM es una metodología que se utiliza para la creación de modelos de análisis de datos y proyecto de minería de datos, actualmente es una de las metodologías más usadas. Esta metodología con ayuda de algoritmos de inteligencia artificial ayuda a identificar patrones de comportamiento en la deserción de los estudios de estudiantes. [9]

En el proyecto de los autores [10] en relación con la estimación del rendimiento académico de los alumnos utilizaron la técnica: Árbol de decisiones (j48) y el algoritmo FT. Al comparar los resultados de ambos algoritmos pudieron concluir que FT es un algoritmo que tiene una alta tasa de precisión en cuanto a la evaluación del rendimiento escolar.

Los autores [11] en su investigación sobre la deserción de estudiantes, de estudios superiores utilizaron como técnica los algoritmos C4.5 y el algoritmo de los K vecinos más cercanos dando como factores principales de la deserción escolar son la edad, ingreso familiar, el nivel de inglés y otro pudiendo concluir que con su propuesta se puede determinar de manera oportuna los factores de riesgo.

Por otro lado, los autores [12] tuvo como objetivo en su estudio la detección de patrones de deserción escolar, teniendo como punto de partida información académica, disciplinares y socioeconómicos de los alumnos de los programas de Pregrado de la Universidad de Nariño y la Organización Universitaria IUCESMAG, donde se basó en el algoritmo Árboles de elección (j48) seleccionando la información académica, disciplinares y socioeconómicos de los alumnos.

En la universidad pública de Chile, [13], utilizaron técnicas de clasificación basadas en árbol de decisión, implementando el algoritmo C4.5, con el objetivo de optimizar los parámetros con los que serían evaluados 5288 casos de estudiantes universitarios. Como resultado, luego de la comparación con los métodos utilizados en investigaciones anteriores usando árbol de clasificación y el algoritmo J48, se obtuvo que el algoritmo C4.5 alcanzó un 87.27% de precisión de la predicción de deserción, a diferencia de investigaciones anteriores que obtuvieron con el algoritmo árbol de decisión un 60.5% y con J48 un 81%.

La solución de ingeniería planteada para este problema de deserción de estudiantes de la carrera de ingeniería de industrias alimentarias de una universidad nacional peruana, es la utilización de un método de clasificación basado en aprendizaje de máquina.

[14] realizaron la investigación, Prediction of Student Dropout in a Chilean Public University through Classification based on Decision Trees with Optimized Parameters, en una universidad pública de Chile. Las técnicas de clasificación basadas en árboles de decisión necesitan de parámetros mejorados para perfeccionar la exactitud de los resultados que se quieren alcanzar. Por esta razón se implementó el algoritmo C4.5 para optimizar los parámetros con los que serían evaluados 5288 casos de estudiantes. Como resultado luego de la comparación con los métodos utilizados en investigaciones anteriores usando árbol de clasificación y J48, se obtuvo que el algoritmo C4.5 en su implementación J48 alcanzó un 87.27% de precisión de la predicción de deserción a diferencia de investigaciones anteriores que obtuvieron con el algoritmo árbol de decisión un 60.5% y con J48 un 81%. La presente investigación ha limitado usar el conocimiento que obtienen los estudiantes de pregrado durante sus años de estudio como atributo que podría ser una alternativa para seguir mejorando la precisión en la predicción de deserción estudiantil, mencionan también que el uso de las redes neuronales serían una gran alternativa para complementar la eficiencia de este tipo de procesos. El algoritmo C4.5 ha demostrado tener un eficiente método de parametrización para divulgar información detallada de la problemática de deserción.

[15], realizaron la investigación, Analysis of classification methods for the fertility diagnosis, trabajando con la base de datos fertility diagnosis extraídas del repositorio UCI para analizar grandes volúmenes de datos suele tardar demasiado tiempo cuando este trabajo es realizado mecánicamente por profesionales e incluso en muchos de los casos llega a ser imposible. En el sector salud se necesita analizar clínicamente a los pacientes para saber las causas que rodean su consulta y a partir del diagnóstico brindar un tratamiento o hallar la causa del mal que éste padece. Por esta razón, se analizó 4 métodos de clasificación: NNge, Random forest, Kstar, y Regresión Lógica Bayesiana para procesar un historial de 100 pacientes tomando en cuenta 9 atributos que pudieran detectar la infertilidad en varones. Los resultados obtenidos evidenciaron que el método de

clasificación Kstar tuvo un tiempo de ejecución de 0 segundos y NNge un 0.03 segundos, demostrando Kstar ser el más rápido en el procesamiento de los datos. Sin embargo, en el estadístico de kappa, que es el índice donde se evalúa el mayor grado de conformidad, NNge obtuvo 0.9543 aproximándose más al valor unitario a diferencia de Kstar que obtuvo un 0.9509. Como resultado se definió que el algoritmo de clasificación NNge es el más eficaz para evaluar estos tipos de base de datos por ser el método que obtiene el mayor grado de concordancia y añadiendo que durante el proceso obtuvo un menor porcentaje de error.

[16] en su investigación Academic performance prediction by machine learning as a success/failure indicator for engineering students en la Universidad Distrital de Colombia, carrera de Ingeniería Industrial. Los factores o variables para medir la deserción en los estudiantes se vuelven difícil. Esto sucede porque en el aspecto educativo estos factores suelen ser multidimensionales y no existe una teoría que defina puntualmente una metodología para medir estos parámetros ya que estos dependen de múltiples aspectos. En la presente investigación se hace uso de la Analítica de Aprendizaje, Esta metodología permite personalizar búsquedas de los datos del campo de la educación sin importar que estos estén anidados. Con el uso de Árbol de decisión, máquinas de vectores de soporte (SVM), Red Neuronal (Perceptrón) y K- Vecinos más cercanos (KNN) examinaron los datos de 1571 registros de estudiantes y se utilizaron 30 variables más influyentes en búsqueda de datos de deserción estudiantil. Los análisis de estos datos multidimensionales obtuvieron un 66.00% con el algoritmo SVC y con Perceptrón un 66.24 % este último fue más favorable por su precisión y por obtener un mejor resultado en cuanto a métricas de evaluación. La Analítica de Aprendizaje propone un mejor tratamiento de los datos aun siendo estos multidimensionales. El modelo propuesto puede ser mejorado por encima de un 90 % si se añade a los parámetros de evaluación factores de rendimiento académico tales como gestión académica de la institución, equipamiento tecnológico sin dejar de lado los factores demográficos y socio-económicos, añadieron los autores.

[17] en su investigación Awajún and Wampis Student Dropout Estimation Model Using Data Mining, en la Universidad de Jaén. El lugar de origen del estudiante según el estudio realizado ha provocado el 45% de deserción entre los años 2012 y 2019 en la universidad de Jaén, y analizar esta variable sería una estrategia clave para detectar y resolver de forma temprana este problema. Con el uso de la metodología CRISP-DM y los algoritmos J48, Ridor y PART se procesaron los registros de 49 estudiantes provenientes de pueblos originarios (Wampis y Awajún) con 17 variables por cada instancia. De acuerdo a las reglas de clasificación adoptadas para cada algoritmo se pudo establecer 3 modelos basado en 2 reglas y se obtuvo que el 45% de la población estudiada abandona la universidad. Analizar el lugar de proveniencia del estudiante junto al historial académico y factores socio-económicos ayudarán a mejorar la temprana detección de deserción y esta permitirá la toma de decisiones asertivas en beneficio de los estudiantes.

[18], realizaron la investigación, Review of techniques, tools, algorithms and attributes for data mining used in student desertion, en la universidad distrital Francisco José de Caldas – Colombia (UDFJC). Los factores involucrados en la deserción de estudiantes abordan desde aspectos personales hasta sociológicos, organizacionales e internacionales, analizar estos tipos de factores requiere de estrategias que permitan seleccionar el algoritmo apropiado para los datos que se está tratando. Para este tipo de investigación, se utilizó algoritmos que permitan clasificaciones binarias tales como (Perceptrón Promediado, Máquina de puntos Bayes, Árbol de Decisión Potenciado, Bosque de Decisión, Selva de Decisión, Regresión Logística y Red Neuronal de Dos Clases), debido a que se busca como resultados valores binarios como deserción o no deserción. Los autores obtuvieron una base de datos del programa de Ingeniería Industrial del año 2003 al 2018 de la UDFJC, tomando como primera acción examinar la base de datos para eliminar datos redundantes o inconsistentes. Los algoritmos seleccionados se medirán por su rendimiento y es necesario utilizar la herramienta de validación cruzada para generar las métricas de evaluación (puntuación F1, exactitud, precisión y recuerdo).

Como resultado de la evaluación se determinó que el algoritmo Árbol de Decisión Potenciado de Dos Clases es el adecuado para procesar los datos, obteniendo una precisión del 90.3% y una exactitud de 93.6%. Con el uso de técnicas de aprendizaje automatizado de Azure Machine Learning Studio se ha podido detectar que la principal causa de deserción es el rendimiento académico, además que estas técnicas son entrenadas para adecuarse al tipo de datos que se desea procesar.

[19] , en su investigación Analysis of Dropouts of University Students using Data Mining Techniques en la Universidad de Católica del Norte en Antofagasta y Coquimbo – Chile. Predecir la deserción estudiantil no es una tarea fácil, y para poder elegir algoritmos que nos ayuden a descubrir las razones de deserción se debe tener en cuenta que los algoritmos que se utilicen deben destacar por su calidad de clasificación. Por esta razón se utilizaron los algoritmos de Redes Bayesianas, Redes Neuronales y Árbol de Decisión que cuentan con calidad de clasificación y fácil detección de variables o dimensiones que permiten obtener mejores resultados de deserción. Se obtuvo una base de 73.958 estudiantes pertenecientes al primer semestre del año 2000 hasta el segundo semestre del año 2013. Se construyeron 3 clasificadores, redes bayesianas, árboles de decisión y redes neuronales. Se obtuvo una precisión del 73 % para red neuronal con una clasificación correcta de 80%, 72% para Árbol de decisión con una clasificación correcta de 82% y Red bayesiana con el 76% de precisión y una clasificación correcta de 76%. Estos resultados tienden a mejorar cuando insertamos más factores como por ejemplo la información psicosocial del estudiante.

[20], en su investigación Nodes: Platform for the prediction of school dropout using artificial intelligence techniques, para el Sistema Educativo de la zona centro del Estado de Veracruz– México. La deserción estudiantil es una problemática que está afectando gran parte del mundo, y solucionar esta tarea requiere de nuevas experiencias de aprendizaje automático con herramientas tecnológicas que implemente técnicas de inteligencia artificial y permitan la rápida detección de estudiantes desertores. A través del algoritmo de

clasificación “Two Class Boosted Decision Tree” (Árbol de decisión impulsado de dos clases) con el servicio de Azure Machine Learning se implementa un modelo para obtener predicciones ejecutadas en la nube con la ayuda de APIS y base de datos. Este algoritmo permite obtener dos tipos de respuestas; un SÍ o un NO tomando como media el valor 0.55 entre 0 y 1. Implementado ya el modelo probabilístico, las instituciones pueden consultar que porcentaje de la cantidad total de sus estudiantes pueden presentar el peligro de deserción, y estas consultas las harán a través de archivos CSV (comma - separated values, Valores separados por comas) para su aplicación de forma masiva. Para poder utilizar estas técnicas de inteligencia artificial es necesario construir y entrenar modelos que permitan aprender a las computadoras a través de la inserción de datos, y que estos puedan actuar sin ser programadas.

[21] en su investigación Analysis of Attrition-Retention of College Students Using Classification Technique in Data Mining en la Universidad Gastón Dachary en Argentina. Las bases de datos que tienen las universidades deben de ser aprovechadas para tener oportunidad de contar con información valiosa y evitar la deserción de los estudiantes. A través del uso de minería de datos con los algoritmos C4.5, Naive Bayes aumentado a árbol y OneR, se pretende encontrar patrones de datos que simplifiquen el trabajo de búsqueda de estudiantes con problemas de deserción. Se creó condiciones mediante un conjunto de reglas para las clasificaciones y medir la robustez de los algoritmos puestos a prueba. Se obtuvo 855 casos para analizar de los estudiantes de la carrera de Ingeniería Informática. Para la medición de las instancias clasificadas correctamente (ICC) el algoritmo J48 tuvo una gran ventaja obteniendo un 80.23%, en diferencia al algoritmo OneR con un 76.61%. Para la medición de precisión el algoritmo OneR obtuvo un porcentaje superior de 82.3% a diferencia de J48 con un 79.7%, por otro lado, Naive Bayes con un 81.1%. En el proceso los 3 algoritmos tuvieron aciertos similares sin embargo cada uno tiene una forma diferente de identificar los atributos aun perteneciendo estos a un mismo método, es decir un atributo tiene mayor o menor importancia en el algoritmo J48, que en

Naive Bayes o OneR, esto indica que cada algoritmo tiene su forma de analizar los atributos, y se recomienda que en el análisis de la problemática se obtenga la mayor cantidad de variables para mejorar el resultado de precisión de los algoritmos.

[22], en su investigación Machine Learning to improve the MOOC experience: the case of the Universitat Politècnica de València en la universidad de Politècnica de Valencia. MOOC (Massive Online Open Courses), no es una plataforma consolidada que asegure el aprendizaje a distancia, en los últimos tiempos la Universidad Politécnica de Valencia ha sufrido altas tasas de abandono de estudiantes desconociendo los factores que influyen en esta deserción. Los mecanismos automatizados de Machine Learning permiten crear modelos efectivos que generen patrones de comportamiento a partir de los datos no utilizados por las bases de datos tradicionales. Para poner a prueba el método de automatización se tuvo el conjunto de acciones que se realizaron entre los años 2015 al 2019 de 260 cursos en las que se matricularon 700.000 participantes, esto se hace posible con una investigación basada en el diseño (IBD) con las que se construirán 03 iteraciones que permitirán conocer los indicadores de abandono, desarrollo de scripts para el filtro de valores y la creación de procedimientos automatizados que permitan consolidar propuestas de mejora. Este proceso debido a la cantidad de sus datos se propuso utilizar los modelos más utilizados de machine learning, Logistic Regression, Decision Tree classification, KNeighbors classification, Support vector machine, Naive Bayes y Random Forest con las métricas de clasificación Precisión, Recall y F1 score. Diseñar modelos automatizados con Machine Learning reutilizando los grandes volúmenes de datos mejorará la experiencia del usuario que recibe cursos en línea, de la misma forma permitirá a la universidad perfeccionar su servicio.

[23] en su investigación Perspectives to Predict Dropout in University Students with Machine Learning, en el Instituto Tecnológico de Costa Rica. Preparar de forma eficiente el archivo de datos para el entrenamiento y validación de los algoritmos, ayudará a mejorar la predicción de deserción de los estudiantes del nivel superior incluso pronosticar quienes

dejarán de matricularse en un próximo semestre académico. Con el entrenamiento de los algoritmos de aprendizaje automático, Random Forest, Redes Neuronales, Support Vector Machines y Regresión Logística se pretende identificar cuál de estos algoritmos es el indicado para utilizarse cada fin de semestre y pronosticar la deserción y utilizar programas de ayuda para la retención de estudiantes. Para el entrenamiento de los algoritmos se obtuvo registros de estudiantes de los años 2011 al 2016, 90.067 registros de 16.807 matrículas de los estudiantes. Evitar la existencia de ruido en el archivo de datos que será procesado, es una muy buena práctica para mejorar los resultados, es por ello que se eliminaron registros incompletos o con información incorrecta, teniendo como base de entrenamiento 80.527 registros de 15.720 matrículas. Para la división de los datos, el entrenamiento y la validación, los algoritmos fueron entrenados con las siguientes variables: sociodemográfico, programa de estudio, historial académico entre otras, teniendo como resultado al algoritmo Random Forest como el algoritmo que mejor se adapta y proporciona mayores resultados con mtry (Número de variables muestreadas aleatoriamente como candidatas en cada división.) de 10 y una sensibilidad del 93% y un 94% de positivos verdaderos. Los entrenamientos de estos algoritmos mejoran sus resultados cuando existe una buena especificación de las variables a usar y el archivo de datos a procesar carezca de datos que no sirven para la búsqueda de la información que se necesita.

[24], en su investigación A predictive model for identifying students with dropout profiles in online courses, en la Universidad Federal de Alagoas - Brasil. La gran cantidad de datos históricos que se tienen de los cursos en línea deben de ser aprovechados para crear modelos predictivos que prevengan la deserción estudiantil. El estudio analizó que esta modalidad ha presentado altas tasas de abandono. Para la creación de los modelos predictivos se puso a prueba 4 algoritmos de predicción, Clasificador Probabilístico simple basado en la aplicación del teorema de Bayes, Árbol de Decisión, Máquina de Vectores de Soporte y Red Neuronal Multicapa y se seleccionaron datos anónimos de 162 estudiantes

del curso de Sistemas de Información. Normalmente los datos suelen tener información no relevante, incompletos e inconsistentes, por ello se aconseja utilizar mejores técnicas de procesamiento de datos para obtener mejores resultados en la precisión. Para la presente investigación se tuvo por objetivo considerar el algoritmo que obtenga el mayor porcentaje en precisión y la menor tasa de falsos positivos, teniendo como resultado que el algoritmo Máquina de Vectores de Soporte alcanzó un precisión de 92.3% con una tasa mínima de falsos positivos de 0.06% a diferencia de Árbol de Decisión con una precisión de 86.46% y tasa de falsos positivos de 0.09%, Naive Bayes con una precisión de 85.50% y una tasa de falsos positivos de 0.11% y Red Neuronal con una precisión de 90.86% y una tasa de falsos positivos de 0.07%. La creación de tuplas debe de considerarse a partir de clasificadores con alta precisión y de esta forma clasificar datos que carezcan de etiquetas de clase, esta sería una forma de poder contar con los datos que normalmente quedan fuera del procesamiento por ser incompletos o inconsistentes.

[25] en su investigación Predicting Student Retention Using Support Vector Machines, en una universidad comunitaria del Medio Oeste, específicamente en las carreras de STEM. Pronosticar la finalización exitosa de una carrera en los estudiantes mejora la reputación de las universidades y prevé desarrollar estrategias para el asesoramiento a los estudiantes con peligro de deserción; la preparación de los datos es un proceso muy importante en la aplicación de técnicas de aprendizaje para grupos de datos no equilibrados. La necesidad de la presente investigación repercute en aumentar la tasa de retención de los estudiantes ya que esta es parte de la identificación institucional en la sociedad. Para la presente investigación se utilizaron los datos de 182 estudiantes y 09 variables que serían procesados por el algoritmo Máquina de Vectores de Soporte (SVM), teniendo como resultado cifras mayores al 70% en tasa de recuerdo y tasas de pruebas en un 78%. Se puntualiza que los resultados podrían mejorar si se toma como variables el aspecto económico y demográfico.

[26] en su investigación Bayesian Classifier Applied to Higher Education Dropout, en la universidad de Mumbai – India. La predicción de deserción de estudiantes puestos a prueba con algoritmos populares como Redes Neuronales de Avance, Maquina de Soporte Vectorial, Regresión Logística arrojan un porcentaje poco confiable para minimizar este fenómeno. El diseñar un clasificador bayesiano simple que trabaja en conjunto con las variables de clase y frontera de Markov, proporciona mejores porcentajes para detener la deserción. Con datos de 12 274 estudiantes matriculados en el periodo 2017 – 2018 en una la escuela de Ciencias de la Ingeniería, se realizó un método de procesamiento con una validación cruzada de 10. Para esta investigación se puso a prueba el desempeño de los algoritmos J48 que obtuvo una clasificación correcta de 87.8862%, Bosque Aleatorio con un resultado de 89.7785% y BayesNet con k2 y un máximo de 5 padres que mejoró el resultado con una clasificación correctamente de 91.3658%. Relacionar las variables de clase puede mejorar el desempeño de los algoritmos predictores.

[27] en su investigación Predicting of School Failure using data mining, en la universidad Autónoma de Zacateca México. El fracaso estudiantil en la actualidad es conocido como “El problema de las mil causas” esto debido a que influyen muchos factores que acrecientan el fenómeno de deserción educativa. El determinar las causas de deserción utilizando técnicas estadísticas no son procedimientos confiables, la minería de datos a través de sus técnicas predictivas ha incursionado un nuevo concepto llamado Minería de Datos Educativo (EDM). Un buen procesamiento de los datos permite mejorar notablemente la precisión de un algoritmo. Se utilizó la información de 670 alumnos con 77 atributos divididas en 10 particiones, los autores indican que para alcanzar un mejor rendimiento de precisión debe haber una limpieza de atributos, y es así que de los 77 atributos se pudieron obtener solo 15 que son según mencionan los más relevantes para un proceso eficiente. Para manipular la información con respecto a las calificaciones de los estudiantes se debe cumplir con la etapa de la discretización, proceso en el cual las notas obtenidas de los estudiantes de cambian a un formato nominal o categórico. Un buen

balanceo de los atributos para mantener el equilibrio entre la información que está tomada para la etapa de entrenamiento y la etapa de prueba mejora los resultados. se compararon 10 algoritmos que fueron sometidos a prueba teniendo como resultado que los algoritmos OneR, Prism y AdTree obtuvieron los mejores rendimientos. AdTree obtuvo 97.2% de asertividad, mientras que Prism obtuvo un 94.7% y OneR un 88.8%. intensificar la seleccionar de atributos relevantes para un proceso y equilibrar los datos de entrenamiento y prueba ayudaran a mejorar la precisión de los resultados de cualquier técnica de minería de datos.

[28] en su investigación Review of techniques, tools, algorithms and attributes for data mining used in student desertion, Colombia, el rendimiento académico en los estudiantes es una de las principales causas del abandono de estudios. Los autores refieren que tratar los datos que se almacenan en los sistemas que utilizan las escuelas es una buena alternativa para analizar y a través de técnicas de minería de datos se pueda detener este fenómeno. Para tratar esta problemática se debe de transformar la información que se adquiere, esta transformación se debe hacer a través de minería de datos en la fase de descubrimiento de datos dentro del proceso de KDD. Se evaluaron 4 técnicas de minería de datos como Clasificación, Regresión, Agrupación y Normas de Asociación y utilizaron 03 herramientas de minería de datos, Weka, Naranja y Minero Rápido. Para este caso de investigación se utilizaron variables Situación académica, Media acumulada, Media por semestre, Calificaciones por semestre para cada asignatura, Valoración en las pruebas de estado y Colegio de origen en el bachillerato, también se tomó en cuenta analizar atributos relacionados con la familia y la ubicación del estudiante. Para el caso de investigación y el análisis de que algoritmos de utilizan con más frecuencia se tiene que el algoritmo J48 es el más usado en los casos de investigación sobre deserción estudiantil.

[29] en su investigación Applying Data Mining Techniques to Predict Student Dropout: A Case Study en una universidad de Colombia, La calidad de educación que

brinda una universidad se mide por la cantidad de estudiantes graduados cada fin de año. Predecir las causas de abandono de estudio es una de las principales preocupaciones de las universidades, debido a ello es que se genera pérdidas económicas para las escuelas como para la sociedad. Es de suma importancia conocer cuáles son las características por las que los estudiantes desertan de los estudios, de esta forma se puede formar variables que ayuden a predecir que estudiantes pueden abandonar sus estudios. Se utilizó una base de datos del año 2004 al año 2010 pertenecientes a 802 estudiantes, se analizaron los datos bajo dos enfoques, el primero utilizando CRISP-DM con árboles de decisión, Regresión logística y Naive Bayes y el segundo con Watson Analytics. Se realizó una clasificación binaria en la que el abandono (0,1) es la variable que se tomó como principal objetivo. Las variables que se tomaron en cuenta para este caso fueron admisión, demografía, fechas de graduación, expediente académico, media de las calificaciones y ayudas financieras. En el proceso de entrenamiento de los datos se tuvo que la frecuencia de abandono era del 57,87% para el SI y del 47,13% para el NO. Se obtuvo que el algoritmo Árbol de decisión fue el mejor predictor con un 94% de precisión a diferencia de Naive Bayes con 92%. Los autores mencionan que relacionar la totalidad de los cursos mejorará el impacto de predicción.

[30] es su investigación Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout en sistemas E-learning. Los datos históricos de los estudiantes que almacenan los sistemas E-learning pueden llegar a ser desalentadores al momento de analizarlos, la solución para este caso es utilizar técnicas de minería predictiva y herramientas de análisis computacional. Para el caso de investigación se tomaron datos históricos de estudiantes a distancia en los años 2013 al 2015. Para detectar la deserción se utilizó como variable las calificaciones de los estudiantes, de igual forma los autores refieren que la regresión logística es una técnica muy eficiente para tratar estos tipos de datos cuantitativos. Se utilizaron los algoritmos de Red Neuronal, Máquina de

Vectores de Soporte, clasificador ARTMAP y un sistema de minería de datos educativos (SEDM). Como resultado se obtuvo que el mejor desempeño lo tuvo SEDM con un 85.71% de precisión. Crear un mecanismo que actúe automáticamente alertando la posible deserción de un estudiante sería una buena práctica que deberían tener los cursos E-learning.

La importancia de la investigación es multifacética y se justifica por varias razones fundamentales:

Tiene impacto en la educación debido a que la deserción estudiantil es un problema que afecta negativamente la calidad de la educación y el desarrollo de recursos humanos en cualquier país. Detectar y prevenir la deserción es fundamental para garantizar que los estudiantes tengan la oportunidad de completar sus estudios y adquirir las habilidades necesarias para el mercado laboral.

Aborda lo económico y financiero debido a que la deserción estudiantil conlleva costos significativos para las instituciones educativas y el Estado. Esto incluye la pérdida de matrícula y recursos invertidos en estudiantes que abandonan sus estudios, así como la disminución de futuros ingresos fiscales de graduados. Por lo tanto, la reducción de la deserción puede tener un impacto positivo en las finanzas públicas y en la eficiencia del sistema educativo.

Mejora de la planificación educativa pues el uso de algoritmos de aprendizaje automático para predecir la deserción estudiantil permite a las universidades identificar a los estudiantes en riesgo antes de que abandonen sus estudios. Esto permite implementar intervenciones tempranas, como programas de tutoría o apoyo académico, para ayudar a estos estudiantes a tener éxito. Asimismo, considera la eficiencia en la asignación de recursos al identificar a los estudiantes en riesgo, las instituciones educativas pueden asignar recursos de manera más eficiente. Esto implica concentrar los esfuerzos en aquellos que más lo necesitan, lo que puede llevar a una mejora en el rendimiento estudiantil y ahorros de costos.

De idéntica manera contribuye a la toma de decisiones basada en datos debido a que la investigación promueve la adopción de la toma de decisiones basada en datos en el ámbito educativo. Al utilizar algoritmos de aprendizaje automático, se fomenta una cultura de análisis de datos y se proporciona a las instituciones una herramienta valiosa para tomar decisiones informadas y estratégicas.

1.2 Formulación del Problema.

¿Cómo detectar la deserción de estudiantes de la carrera profesional de ingeniería de industrias alimentarias de una universidad nacional peruana?

1.3 Hipótesis.

Con la implementación de un método de clasificación basado en aprendizaje de máquina se podrá detectar la deserción de estudiantes de la carrera profesional de ingeniería de industrias alimentarias de una universidad nacional peruana.

1.4 Objetivos.

1.4.1 Objetivo general.

Implementar un método de clasificación para detectar la deserción de estudiantes de la carrera de ingeniería de industrias alimentarias de una universidad nacional peruana basado en aprendizaje de máquina.

1.4.2 Objetivos específicos.

- A) Elaborar un conjunto de datos con información relevante de los estudiantes de la carrera de ingeniería de industrias alimentarias de una universidad nacional peruana previamente seleccionada.
- B) Seleccionar algoritmos de clasificación de Machine Learning con mejor desempeño en casos similares.

- C) Desarrollar el método de clasificación basado en los algoritmos previamente seleccionados.
- D) Realizar pruebas para verificar el desempeño del método de clasificación propuesto.

1.5 Teorías relacionadas al tema.

1.5.1. Deserción Estudiantil.

La deserción estudiantil se define como un hecho de que los estudiantes no sigan la trayectoria de un programa académico, esto se puede dar por el retiro de la carrera profesional o por demorar en terminarla, debido a los cursos desaprobados o retiros temporales, en algunos de los casos también el cambio de carrera debido a que el estudiante no tiene una idea madura sobre lo que quiere estudiar o porque el servicio que brinda la institución no es la adecuada. El autor menciona que existe un proceso llamado "Proceso del Conocimiento" donde intervienen un conjunto de herramientas de minería de datos, algoritmo escogido y resultados que se obtienen según el contexto evaluado. Para ello refiere que la minería de datos es una de las herramientas mejores usadas para resolver el tema de deserción, ya que se constituye de varias etapas e incluye acciones de mucha complejidad que involucran la búsqueda de estructura, modelos y parámetros dentro de una base de datos[31].

Los primeros estudios acerca de la deserción en el nivel universitario, se remontan a los años 1970 y 1980, la inquietud por el análisis de este tema ha aumentado desde los años 1990, cuando con el incremento de las matrículas universitarias, el aumento con respecto al índice de deserción se fue haciendo notorio [32], [33].

1.5.2. Inteligencia Artificial.

Es considerada una de las ramas de las ciencias de la computación, la cual ha despertado gran expectativa en los últimos tiempos debido a su amplio campo de aplicación [33]

La IA es el resultado de una amplia y apasionada búsqueda emprendida por la humanidad para desarrollar seres artificiales a su imagen u otras formas de vida, que sean capaz de realizar tareas inteligentes. Por tal razón la IA está rodeada de misticismos, creencias, teorías, promesas y esperanzas que, de la mano con la ciencia y la tecnología, los seres humanos podrían lograr un nuevo génesis de clones y autómatas artificiales [34]

En los últimos tiempos, los objetivos de la ciencia con respecto a la IA es la simulación de la inteligencia humana replicada en una máquina que sea capaz de ser conscientes con sentimientos reales. Siendo la conciencia uno de los problemas más difíciles de simular [35] .

En la década de los ochenta, con el avance del software y hardware de aquel entonces, se pretendía a través de la IA simular capacidades humanas tales como la visión o el razonamiento, los países desarrollados destinaron grandes recursos para tal simular tal fin [35].

En la década de los noventa, se obtuvieron avances significativos en todas las áreas de la IA, tales como: aprendizaje automático, tutoría inteligente, programación, data mining, razonamiento basado en casos, razonamiento incierto, planificación multi agente, comprensión del lenguaje natural, traducción, realidad virtual, juegos, etc.



Fig. 1 Línea de tiempo de la Inteligencia Artificial

Fuente: [36]

1.5.3. Sistemas Expertos

Se puede definir como sistemas basados en conocimientos, cuyos procedimientos dan como resultados la emulación que se obtendría de una persona experta, teniendo como principales características:

- Amplio dominio de su conocimiento.
- Que esté apto a resolver problemas.
- Que los resultados sean confiables.
- Capacidad de aprender nuevos conocimientos.

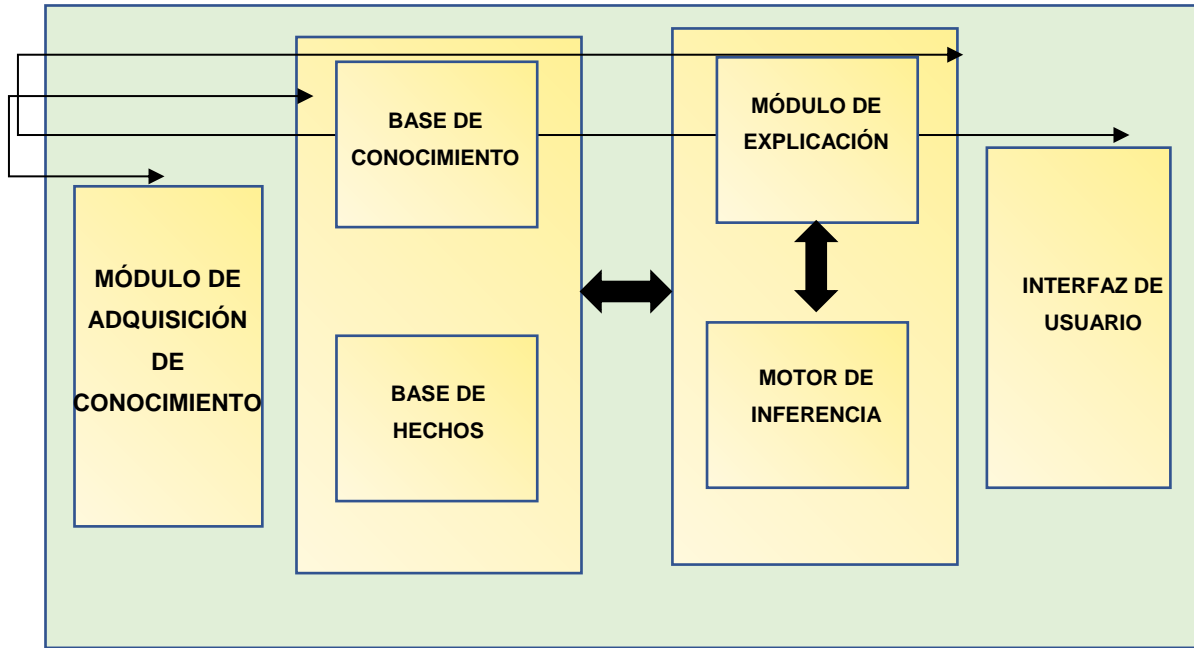


Fig. 2 Características Sistema Experto.

Fuente: [37]

Los primeros SE datan del año 1965, entre los que tenemos:

- **Dendral**, desarrollado por la universidad de Stanford en el año 1965, cuya aplicación estaba orientada a la deducción de datos acerca de estructuras químicas.
- **Macsyma**, desarrollado por el MIT (Instituto Tecnológico de Massachusett en 1965, donde desarrollaron análisis matemáticos complejos.
- **HearSay**, desarrollado por la universidad Carnegie Mellon en 1965, cuya aplicación fue Interpretar en lenguaje natural un subconjunto del idioma.

1.5.4. Redes Neuronales

Existen dos tipos de aprendizaje según las redes neuronales, el primero es el aprendizaje supervisado, que facilita un grupo de datos de entrada y el resultado es provechoso en tareas de regresión y organización, ya que estos en tienen una alta probabilidad de asertividad. También tenemos el aprendizaje no supervisado donde a la

red se le da un conjunto de datos de entrada y esta debe auto-enseñarse para brindar respuestas. Esta técnica es útil para tareas de agrupamiento. [38].

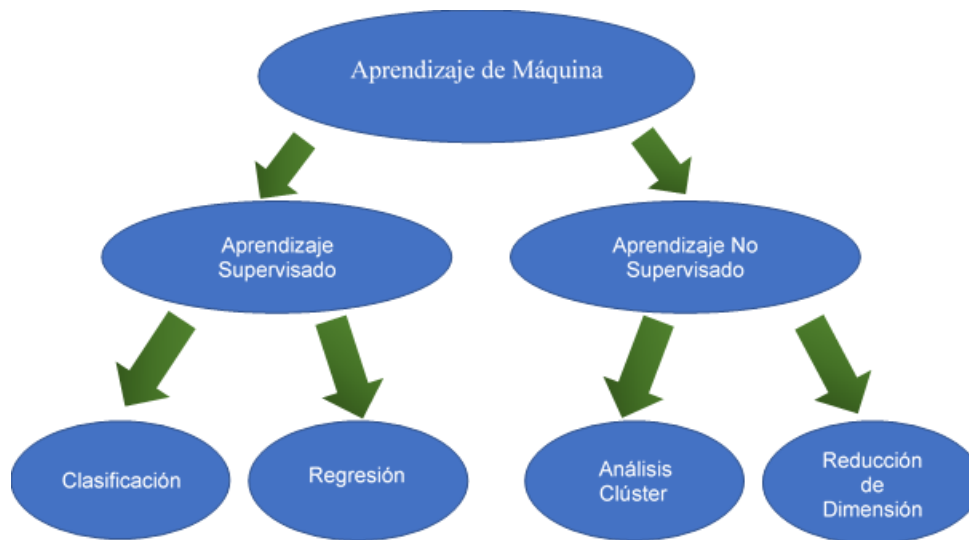


Fig.3 Redes Neuronales

Fuente: [39]

1.5.5. Minería de Datos

La minería de datos o conocida también como DM "*Data Mining*", es el término acuñado en el entorno del comercio, para la aplicación del aprendizaje automático en extensos volúmenes de datos para sustraer información de estos. Algunas aplicaciones de DM se han ampliado en distintas áreas, tales como:

- El análisis del comportamiento de consumo de clientes en centros comerciales.
- En los bancos, donde los datos históricos se usan para crear modelos de riesgo crediticio, detección de fraudes, etc.
- En manufactura, donde se utilizan los modelos de aprendizaje para la optimización, control y solución de problemas.
- En la medicina, se utilizan aplicaciones para los diagnósticos médicos.
- En las telecomunicaciones, para maximizar la óptima calidad de los servicios, se utilizan el análisis de patrones de llamadas [40].

La DM, consta de tres fundamentos genéricos[41]; la agrupación de dichos fundamentos y tras un extenso proceso de investigación y desarrollo de productos brindaron como consecuencia diversas técnicas usadas en la manipulación masiva de datos.

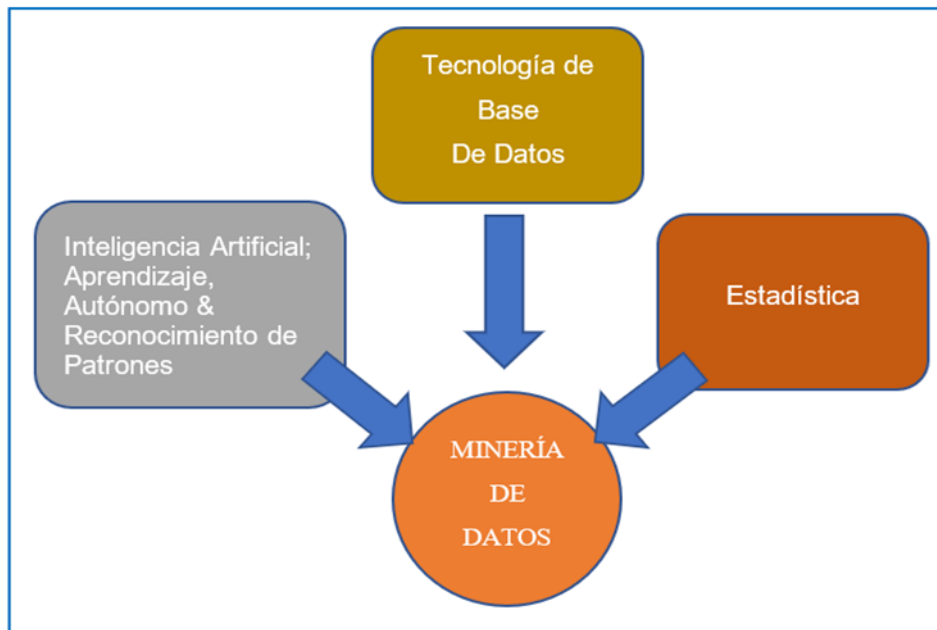


Fig. 4 Data Mining

Fuente: [41]

El proceso de DM es secuencial e iterativo; los pasos a seguir, según [42] son los siguientes:

Recolección de datos

En este proceso se puede necesitar de algún hardware o software concreto, esto según el tipo de problema o solo una conexión fiable a una base de datos, en varios escenarios este es un ejemplo común; no obstante, se puede necesitar la ayuda de sensores o inclusive del trabajo manual para desplegar el acopio de datos; parte de este proceso además considera el razonamiento del comercio; debido a que en este periodo se necesita recolectar los datos requeridos para la resolución del problema.

Preprocesamiento de datos (Limpieza).

Una vez reunidos los datos; tienen que pasar por un proceso de limpieza y transformación; ya que si se usa no solo una sino numerosas fuentes los datos tienen la posibilidad de solicitar inclusive hasta un proceso de asociación de datos. La transformación de datos es forzosa debido a que de esta es dependiente poner la información en un formato legible para los diferentes algoritmos; ejemplificando, el algoritmo entenderá datos categóricos en la edad como “De 18 a 25 años” que los datos numéricos como tales. Es viable que el algoritmo se porte de una u otra forma dependiendo de datos nulos o vacíos, por lo cual es fundamental mantener el control de que los datos sean lo más completos posibles.

Procesamiento analítico (Modelamiento)

En este proceso el analista diseña un modelo de análisis con las distintas herramientas estadísticas, las cuales les van a permitir contestar al problema propuesto o por el contrario descubrir las tendencias y patrones contemplados como parte del problema. Se necesita una evaluación del modelo donde se pueda medir la eficiencia de este; si el modelo no cumple con el inicio deseado o si los resultados logrados del proceso no son consistentes se necesita hacer una iteración a pasos anteriores para ajustar el modelo [43].

KDD (Proceso de Extracción del Conocimiento)

En la actualidad aún se confunde al KDD con minería de datos; debemos de saber que la minería de datos es una de las partes del KDD como lo podemos notar en la siguiente figura. [44].

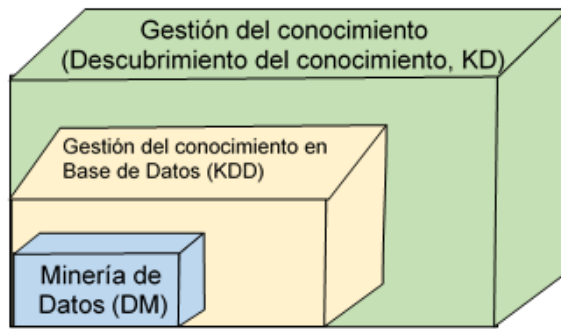


Fig. 5 Proceso de Extracción del Conocimiento

Fuente [44]

KDD se contextualiza como el proceso de análisis, indagación y selección de datos, basado en una base de datos que existe en este proceso y utiliza tecnología de inteligencia artificial para obtener muestras o patrones.

1.5.6. Sistemas Difusos

Estos sistemas permiten que los datos del mundo real sean analizados por un computador, en una escala que determina si es falso o verdadero para lo cual emplean conceptos vagos como pueden ser “húmedo” o “caliente”, lo cual les permite a los ingenieros fabricar aires acondicionados, tv, lavadoras, etc. que estiman datos complejos de definir.

Dichos sistemas datan de los años 1970, tuvieron su origen en Inglaterra (Queen Mary Collage - Londres) por el matemático Ebrahim H. Mamdani, quien fabricó un controlador difuso que se utilizó en el motor de una máquina a vapor y desde entonces el término lógica difusa es sinónimo de sistemas matemáticos o computacionales que razonen con lógica difusa.

Los sistemas difusos constan de valores de pertenencia (conjuntos difusos) o valores reales (lógica difusa) representados por números en el rango [0.0, 1.0], donde: 0.0 simboliza falso y 1.0 representa verdadero. Veamos un ejemplo: en la siguiente frase “Rosa tiene 1 año de edad”, posee un valor real, que podría ser por ejemplo 0.9. Dicha frase se

puede emplear en términos de conjuntos como, por ejemplo: “Rosa pertenece al conjunto de gente pequeña”, en términos de conjuntos difusos.

Es significativo diferenciar entre probabilidad y sistemas difusos, si bien es cierto que ambos operan sobre el mismo rango numérico, los criterios son distintos. Siguiendo con el ejemplo anterior en términos probabilísticos sería: Existe un 90% de probabilidad de que Rosa sea pequeña”, en este caso la frase probabilística supone que “Rosa es o no pequeña”, existe un 90% de probabilidad de saber la categoría en que se encuentra. Por el contrario, la frase difusa supone que “Rosa es más o menos pequeña”. Las probabilidades calculan si algo ocurre o no, en cambio, los niveles difusos miden el grado en el cual algo ocurre o alguna condición existe.

1.5.7. Machine Learning.

1.5.7.1. ¿Por qué utilizar Machine Learning en la presente investigación?

Los algoritmos de Machine Learning (ML) o aprendizaje de maquina vienen generando enormes sorpresas en las empresas de software desde los años 50. Un juego de damas con inyección de aprendizaje automático abrió las puertas de numerosos métodos para predecir el comportamiento de los procesos que enfrentan a diario las organizaciones o empresas. Existe una gran variedad de algoritmos de aprendizaje automático que están constantemente aprendiendo de los datos. El proceso de mejoramiento, descripción y predicción de resultados es una de las características principales por la que estos algoritmos están presentes en muchas empresas u organizaciones. Hoy en día millones de personas tienen la necesidad de buscar un producto por internet, desconociendo que los algoritmos de Machine Learning están tomando su historial de navegación para ofrecer productos parecidos a lo que el usuario está buscando y que puedan ser de su interés. El aprendizaje automático permite el entrenamiento de algoritmos para que puedan ser utilizados en tiempo real. Este proceso contribuye a la mejora de la precisión y el aprendizaje a través de la experiencia.

El aprendizaje automático en línea permite el perfeccionamiento de los modelos con la obtención de datos casi en tiempo real, esto permite la adaptación de nuevos patrones generados por la tendencia de asociaciones cambiantes.

Hoy en día se aprovecha en gran dimensión el uso de estos algoritmos de aprendizaje automático debido a los costos moderados de recursos informáticos y almacenamiento. La inversión provee de grandes avances tecnológicos y soluciones en los procesos predictivos con resultados de beneficio en las empresas [45].



Fig. 6 Inteligencia de Aprendizaje de Maquina

Fuente:[45]

APRENDIZAJE SUPERVISADO

Cuando existe la necesidad de contar con un algoritmo que pueda predecir un comportamiento a través de patrones, se hace uso del Aprendizaje Supervisado. Estos tipos de algoritmos aprenden de ejemplos o respuesta a determinados procesos que pueden estar clasificados como clases o etiquetas. Esto se asemeja a la intervención de un maestro y un alumno en la dinámica de enseñanza, donde el maestro muestra ejemplos sobre cómo obtener algún resultado y el alumno aprende del ejemplo para empezar a crear sus propios ejemplos [46]

APRENDIZAJE NO SUPERVISADO

Existen algoritmos que aprenden a través de simples ejemplos aun sin tener algún tipo de respuestas asociada al proceso que se está tratando de resolver. La reestructuración de datos permite en estos algoritmos crear nuevas características para generar nuevos comportamientos y proporcionar indicadores de estimación. Estos algoritmos muestran información sobre la similitud entre personas, animales u objetos, hoy en día podemos notarlos en los sistemas de los supermercados donde a través de compras de un cliente, el algoritmo estudia su reporte de consumo o búsqueda de algún producto en el historial de navegación propiciándole sugerencias al cliente de posibles productos que también puede ser de su necesidad [46].

APRENDIZAJE POR REFUERZO

Un algoritmo también puede aprender aun si tener alguna etiqueta, el constante ensayo o entrenamiento ayudan a mejorar su objetivo. La toma de decisión es aplicada por el algoritmo y en este caso la participación humana es la de proporcionar una realimentación positiva o negativa que ayude al algoritmo a tomar la mejor decisión en un próximo proceso. Como ejemplo podemos tomar el juego de un laberinto donde el jugador intenta escapar de su enemigo y a la vez está corriendo dentro del laberinto, los intentos de escapar, encontrar la salida y las rutas que elige están siendo tomadas y analizadas por el algoritmo, cuando el jugador logre salir ganador el algoritmo habrá aprendido que decisiones debe tomar para que en un próximo juego, el algoritmo muestre la salida más corta [46].

1.5.7.2. Tipos de algoritmos del Machine Learning

1.5.7.2.1. K vecinos más próximos (K-nearest Neighbors – K-NN)

Los K-NN son métodos que sirven para estimar la función de densidad $F(X/C_j)$ de las predictoras x por cada clase C_j a partir de la información recogida por el conjunto de

prototipos que se procesan. Estos tipos de algoritmos tienen un proceso de entrenamiento, es en este entrenamiento donde se guardan los vectores más característicos y las etiquetas de las clases que los ejemplos que sirvieron como entrenamiento. Luego de esta fase se analiza la distancia de un vector que ya ha sido procesado y almacenado y se eligen los k ejemplos más cercanos. Los atributos más importantes suelen perder su peso si en el proceso no se ha apartado los atributos irrelevantes[47].

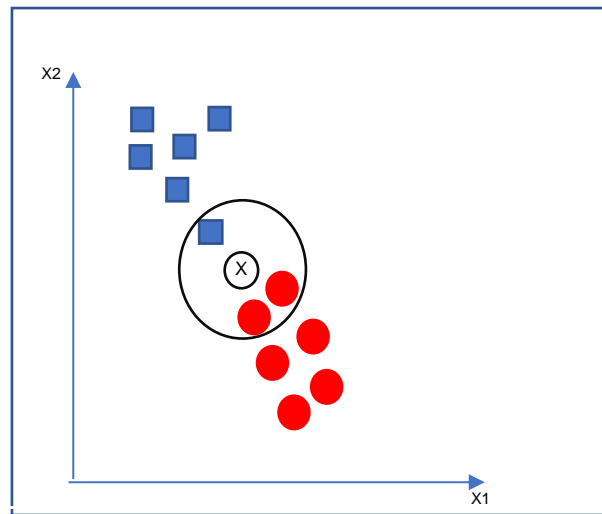


Fig. 7 K-nearest Neighbors – K-NN

Fuente: [47]

En el ejemplo de la figura se puede observar cómo dentro dos clases distintas se han generado tres vecinos más cercanos es decir (K=3).

Un ejemplo de los entrenamientos de X_i es un vector en un espacio multidimensional y su representación vectorial dimensional es de la siguiente forma:

$$X_i = (X_{1i}, X_{2i}, \dots, X_{3i}) \in X$$

Se asigna un punto en el espacio de la clase, si la clase es la más común entre los k ejemplos de ensayo más próximo.

$$d(X_i, X_j) = \sqrt{\sum_{r=1}^p (X_{ri} - X_{rj})^2}$$

1.5.7.2.2. Redes neuronales artificiales (Artificial neural networks)

Una red neuronal artificial aprende mediante entrenamientos consecutivos, es la forma de poder conseguir maximizar su calidad de predicción. En el proceso de entrenamiento se evalúa el error de predicción y se modifica las ponderaciones para mejorar sus resultados.

El Perceptrón La unidad de procesamiento básico en una red neuronal es un nodo, se describe que estos nodos son similares a las neuronas de un cerebro humano: acepta (input) y genera (output). Para crear un valor resumen los nodos suelen procesar los datos de entrada, este valor resumen es la suma de todas las entradas multicapa por sus ponderaciones. La red neuronal también es conocida como la ordenación secuencial que básicamente son 3, los nodos de entrada, de salida e intermedios, estos últimos son conocidos como capa oculta. Las relaciones no lineales han representado situaciones de problemas para las técnicas multivalentes, pero esto es evitado por los nodos que se utilizan con conjunto con la función de activación.

El Modelo neuronal con n entradas contiene:

- Un conjunto de entradas $X_1...X_n$.
- Los pesos sinápticos $W_1...W_n$ correspondientes a cada entrada.
- Una función de agregación, \sum .
- Una función de activación.
- Una salida, Y.

Una neurona tiende a adaptarse al medio que lo rodea y aprender de él, reestructurando el valor de sus pesos sinápticos [47]

El modelo de una neurona Y contiene:

$$Y = f\left(\sum_{i=1}^n w_i x_i\right)$$

1.5.7.2.3. Máquinas de vectores de soporte (Support vector machines)

Una de las singularidades primordiales de este tipo de algoritmos de las SVM es la de separar óptimamente los puntos de un vector etiquetándolos por categorías, estos estarán al lado del hiperplano y los otros puntos de una categoría diferente estarán del otro lado. Intuitivamente estos algoritmos parten de un entrenamiento que permite etiquetar las clases y las representan a través de puntos en el espacio para luego lograr la separación más prudente, esto se realiza para la inserción de nuevas muestras que se coloquen en dicho modelo y que estas también puedan ser clasificadas correctamente.

Como es de saber, en estos métodos de clasificación supervisada los datos(puntos) son vistos como un vector p-dimensional (lista de números p). Un algoritmo de las SVM puede crear un nuevo modelo para predecir si un punto nuevo pertenece a una de las categorías existentes, es por ello que estos métodos se utilizan para solucionar problemas de clasificación y de regresión [47]

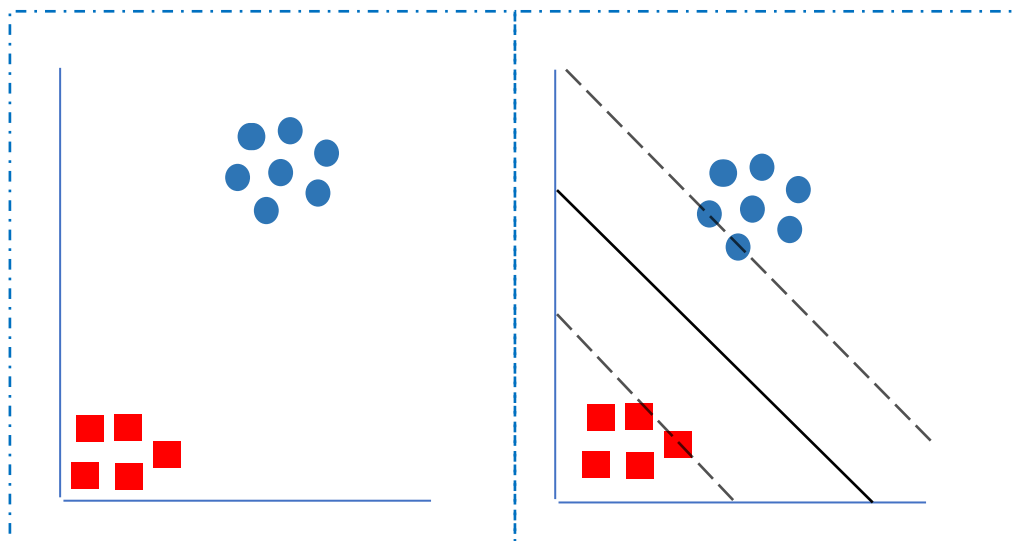


Fig. 8 Máquinas de vectores de soporte

Fuente: [48]

La representación gráfica provee una visualización básica de cómo operan los algoritmos de SVM, pero en realidad el trabajo interno real que se realiza es el de tratar

con más de dos variables predictoras, conjuntos de datos que fácilmente no pueden ser separados, las líneas rectas son el resultado de tratar con curvas de separación y la clasificación se hace en más de 2 categorías.

1.5.7.2.4. Clasificador Bayesiano ingenuo (Naive Bayes Classifier)

El clasificador de Naive Bayes es un constructor de modelos predictivos, en la actualidad se considera uno de los más usados debido su técnica de predicción supervisada que permite obtener resultados favorables en los procesos que requieren evaluar posibles comportamientos [47]

Teorema de Bayes:

$$P\left(\frac{A}{B}\right) = \frac{P(B/A)P(A)}{P(B)}$$

Para poder operar con este tipo de clasificador debemos tener en cuenta que este algoritmo tiene dos partes, la primera que es la construcción del modelo y la siguiente que es la de clasificar nuevos ejemplos con el modelo que anteriormente se ha creado.

Para la creación del método debemos de tener en cuenta 4 pasos importantes:

- ✓ Para cada clase de se debe calcular las probabilidades anteriores.
- ✓ Para mejorar la eficiencia del algoritmo se debe separar las clases.
- ✓ Aplicar los valores de los sucesos equiprobables, para evitar los problemas en caso de que dichos valores sean cero.
- ✓ Obtener valores entre los rangos 0 y 1 previa normalización.

Consideración de un caso equiprobables:

$$P(\text{positivo}) = \frac{1}{2} = 0,5$$

$$P(\text{negativo}) = \frac{1}{2} = 0,5$$

Para la clasificación de un nuevo ejemplo se tiene en cuenta 2 pasos:

- ✓ Cuando surge un nuevo ejemplo se debe se debe crear una clase disponible que determine los valores probabilísticos de los atributos.
- ✓ Aplicación de la fórmula de Naive Bayes.

Para la obtención de n atributos se realiza la siguiente formula:

$$P(A|b_1, b_2, \dots, b_n) = P(A)P(b_1|A)P(b_2|A), \dots, P(b_n|A) = P(A)P(b_j|A)$$

Solución $\text{Armax}_{i=1}^n p(A)P\left(\frac{b}{A}\right)$

Paso 01: De cada atributo se coge los valores 1 que pertenecen a la primera columna de cada tabla.

Paso 02: Aplicamos la formula

$$P(x_5) = 0,5 \cdot (0,5 \cdot 0,25 \cdot 0,5) = 0,03125$$

$$P(\text{negativo}|x_5) = 0,5 \cdot (0,5 \cdot 0,75 \cdot 0,25) = 0,046875$$

Esta segunda fórmula vemos que la clase probablemente es negativa por lo que su clasificación en el nuevo ejemplo es negativa (x_5).

1.5.7.2.5. Árboles de decisión (Decision Trees)

Los árboles de decisión permiten la creación de segmentaciones jerárquicas, estas segmentaciones son formadas a partir de una variable dependiente en la que se agrupan datos comunes o parecidos, siendo estos datos las combinaciones de las variables independientes donde se procesan la totalidad de los casos extraídos de la muestra.

Los elementos de los árboles de decisión son los nodos que representan a las variables de entrada, las ramas que representan los posibles valores de los nodos es decir de las variables de entrada y las hojas son como el resultado o los posibles valores de las variables de salida.

El nodo raíz representa el primer elemento de un árbol de decisión, y es quien tiene la variable de mayor importancia en un proceso clasificatorio. Para el proceso de

clasificación importa evitar que los datos a tratar contengan incoherencias ya que esto perjudica el comportamiento de predicción. Para evitar que esta clasificación contenga sobreajuste existe un método llamado poda que consiste en la modificación de los algoritmos de aprendizaje para evitar que el árbol siga en crecimiento.

Contando con una muestra de entrenamiento y teniendo claro la información de la clase a la que pertenece, podemos realizar la *Maximización de la reducción de impureza*.

Se elige la variable x_1 y se fija un punto de corte que estará representado por la letra c , de tal manera que podamos separar los datos en 2 grupos, por ejemplo, los que tienen $x_1 \leq c$ y los que tienen $x_1 > c$, el nodo inicial ahora tiene 2 nodos, por lo tanto, se establecen dos tipos de observaciones uno que tendrá las observaciones con $x_1 \leq c$ y el otro con $x_1 > c$. Para que el proceso se dé por finalizado se tienen que clasificar todas las observaciones y esto conlleva a que cada nodo tendrá ciclo repetitivo para la selección de la variable y el punto de corte [47].

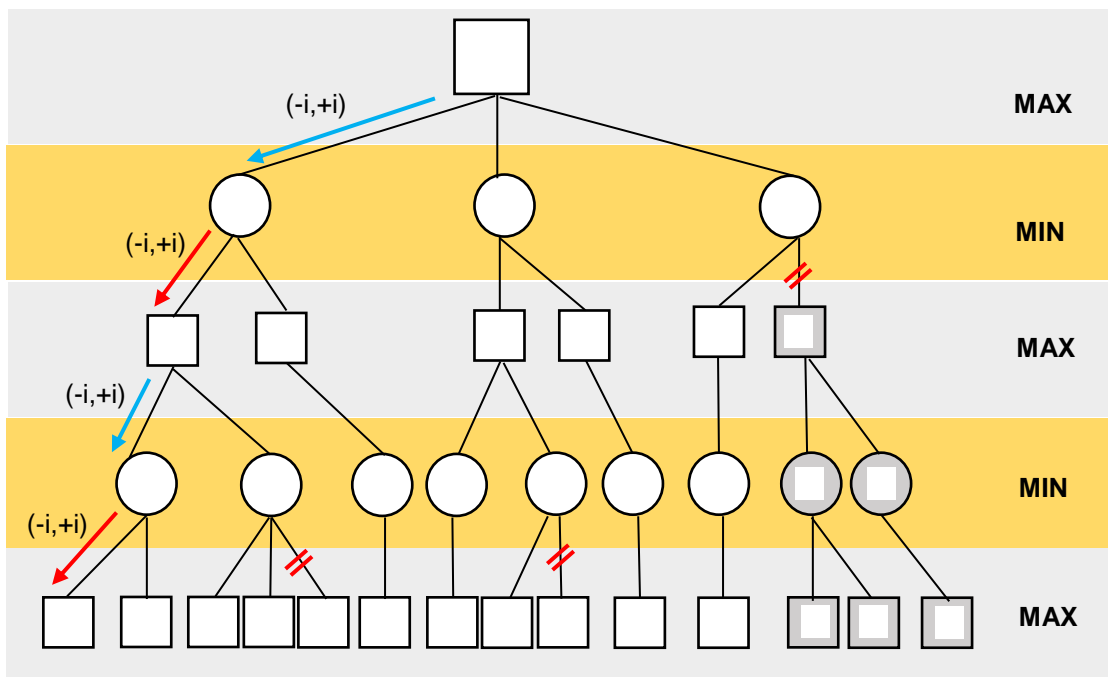


Fig.9 Árboles de decisión
Fuente: [48]

Con las reglas de pre poda se puede obtener:

Pureza del nodo: Para determinar si la construcción o clasificación de un árbol debe finalizar se debe tener en cuenta que todos los registros que se están procesando deben pertenecer a una clase.

Cota de profundidad: esta regla permite tener en cuenta la profundidad del árbol, una vez que esta profundidad se ha ocupado por completo de debe detener el proceso.

Umbral de soporte: No se considera fiable una clasificación que está por debajo del mínimo de ejemplos especificados en los nodos, por la tanto si esto llegara a pasar se debe detener el proceso.

Tenemos dos formas para podar los diferentes algoritmos, la poda por coste-complejidad que permite el equilibrio de clasificación y el crecimiento del árbol, esta forma establece la complejidad por la cantidad de hojas en un árbol y la poda pesimista que coge los casos que son clasificados de forma incorrecta eliminándolos para no perjudicar la precisión del clasificador.

ARBOL CART (Classification and Regression Trees)

Hallar la medida de impureza.

El algoritmo ARBOL CART es conocido como el algoritmo de clasificación y de regresión, permite la creación de árboles de decisión binarios es decir que cada uno de los nodos está dividido exactamente en dos ramas. Este algoritmo acoge variables de entrada y de salida continuas, ordinales y nominales, esto permite la solución de problemas de clasificación y de regresión [48].

Para hallar la medida de la impureza el algoritmo emplea el índice de Gini:

$$G(A_i) = \sum_{j=1}^{M_i} p(A_{ij})G(C/A_{ij})$$

Siendo, $G(C/A_{ij})$ igual a:

$$G(A_{ij}) = \sum_{j=i}^{M_i} p\left(\frac{C_k}{A_{ij}}\right) (1 - p(C_k/A_{ij}))$$

- A_{ij} es el atributo empleado para modificar el árbol.
- J es el número de clases
- M_i es el de valores distintos que tiene el A_i
- $p(A_{ij})$ constituye la probabilidad de que A_i tome su j -ésimo valor y
- $p(C_k/A_{ij})$ representa la probabilidad de que un ejemplo sea de la clase C_k cuando su atributo A_i toma su j -ésimo valor.

El índice de diversidad Gini coge el valor 0 cuando uno de sus grupos es completamente homogéneo y alcanza su mayor valor cuando todas las $p(A_{ij})$ son constantes, entonces el valor para el índice es $\frac{(J-1)}{J}$.

ARBOL C5.0

Cuando se trata de árboles de clasificación el algoritmo C5.0 (en su versión no comercial C4.5), es uno de los más utilizados ya que permite la creación de modelos de árbol de clasificación cogiendo solo variables de salida categórica. Los atributos son analizados por medio de un test estadístico que determina la clasificación permitiendo seleccionar el mejor atributo y colocándolo en la raíz del árbol, luego se crea una rama y su nodo para cada valor posible en su atributo [47].

Si los registros de la base de datos T se agrupan en función de las categorías de las variables de salida S , obteniendo una proporción p_k para cada grupo asociado a un posible resultado, la función de entropía en el caso de dos atributos de salida con probabilidades p y su complementaria, $1-p$, toma la siguiente expresión:

$$INFO(T) = p \log_2(p) (1-p) \log_2(1-p)$$

La variable de entrada obtiene una ganancia teniendo en cuenta la expresión:

$$GANANCIA(X + T) = INFO(T) - INFO(X, T)$$

Donde:

$$INFO(X, T) = \sum_{i=1}^K \frac{T_i}{T} INFO(T_i)$$

- INFO(T,X) es la información proporcionada por la variable de salida S cuando se tiene en cuenta una variable X.
- INFO(X,T_i) es la entropía de la variable de salida S en cada subconjunto T_i determinado por las k categorías de la variable de acceso X.
- Y T_i es el número de registros asociados a una categoría i de la variable X.

El término de ganancia representa la diferencia necesitada para detectar la categoría destino vinculada a un factor T y la información necesitada para detectar esa categoría una vez que se sabe el costo de una variable de acceso para aquel mismo factor. Lo cual esto quiere decir es que dicha variable mostrará menor incertidumbre en el momento de la categorización que lo demás de cambiantes de ingreso [47].

1.5.7.2.6. REGRESIÓN LOGÍSTICA (LOGISTIC REGRESION)

El objetivo principal de la regresión logística es poder predecir el comportamiento de un acontecimiento que surge de una variable dependiente con variables predictoras. Con el uso de esta técnica podemos visualizar el grupo de pertenencia a partir de una variable categórica en función de otro grupo de variables cualitativas o cuantitativas. Existen 02 modalidades de regresión logística.

La Logística binaria permite la explicación de una característica de la variable también conocida como suceso dicotómico, este término se entiende como la división o clasificación en dos partes “todo o nada, perfecto o inútil”.

La Logística Multinomial permite la explicación en un entorno general de variables cualitativas politómica es decir cuando existen más de dos categorías.

La función logística tiene la forma de una curva sigmoïdal, esta forma representa el incremento de una variable desde su expresión lenta hasta su aceleración para terminar con su desaceleración [49].

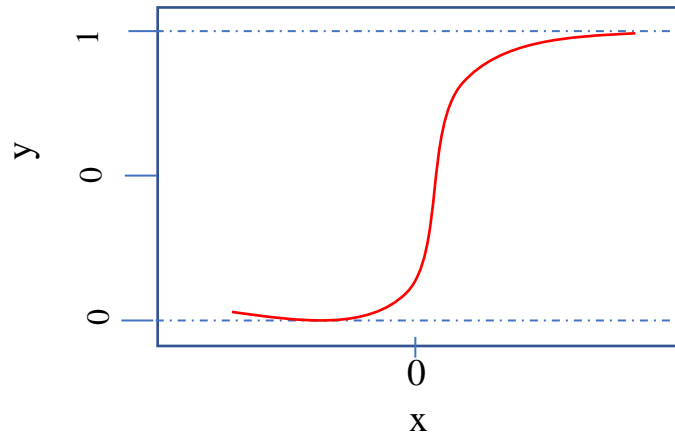


Fig.10 Regresión logistica

Fuente: [50]

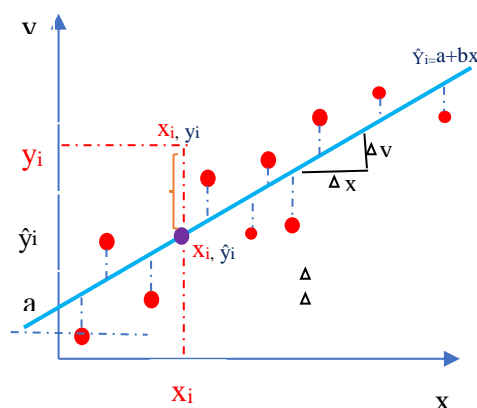
Existen varias etapas por la que una regresión logística realiza su proceso de análisis y estas son la selección de variables del modelo, estimación de los coeficientes de las variables independientes, la clasificación de los casos y análisis de los residuos.

Cuando se trabaja con solo una variable independiente hacemos uso de la regresión logística simple, pero si para el proceso que se está llevando a cabo se requiere del uso de más de una variable independiente estamos usando la regresión logística múltiple.

La expresión de las variables en el modelo de regresión lineal se da de la siguiente forma:

$$Y_i = a + b_1x_{i1} + b_2x_{i2} + \dots + b_px_{ip} + e_i$$

La representación particular del modelo de regresión simple se grafica a través de una recta en el plano que se ajusta por el método de mínimos cuadrados.



$$b = \frac{y}{x}$$

Fig. 11 Regresión Lineal

Fuente: [49]

Se estiman dos sucesos en el modelo de regresión logística binaria, que se codifican con el valor 0 y 1. Si existe la probabilidad de que uno de ellos se cumpla es representado por P , y la probabilidad de que se cumpla lo otro igual a 1 menos la probabilidad de P :

$$Pr Pr (y = 1) = P$$

$$Pr Pr (y = 0) = 1 - P$$

II. MATERIAL Y MÉTODO

II.1. Tipo y Diseño de Investigación.

El tipo de investigación del presente proyecto es de tipo cuantitativa ya que las causas de deserción estudiantil a través de métodos de clasificación es un proceso experimental y con ayuda de los algoritmos de Machine Learning se logrará obtener resultados para la detección temprana de deserción.

II.2. Población y muestra.

II.2.1. Población

Según [49] la población es el grupo de personas u objetos con los cuales se realizará un análisis según la investigación que se está llevando a cabo. En el presente proyecto la población consta de 27 algoritmos de machine learning. Según la literatura científica la lista que se presenta a continuación son algoritmos que permiten realizar el tratamiento de los datos a través de técnicas de aprendizaje, esto permitirá detectar la temprana deserción de los estudiantes y la mejora en la toma de decisiones para beneficio de la organización.

Tabla 1.

Población de algoritmos utilizados para Técnicas de Machine Learning.

Ítem	Algoritmos
1	Bayesian Regression
2	Perceptrón
3	Passive Aggressive Algorithms
4	Dimensionality reduction using Linear Discriminant Analysis
5	Estimation algorithms
6	Support Vector Machines Classification
7	Stochastic Gradient Descent Classification

8	Nearest Neighbors Classification
9	Neighborhood Components Analysis
10	Nearest Centroid Classifier
11	Nearest Centroid Classifier
12	Nearest Neighbors Transformer
13	Gaussian Process Classification (GPC)
14	Gaussian Naive Bayes
15	Multinomial Naive Bayes
16	Bernoulli Naive Bayes
17	Decision Trees Classification
18	Tree algorithms: ID3
19	Tree algorithms: C4.5
20	Tree algorithms: C5.0
21	Tree algorithms: CART
22	Forests of randomized trees
23	AdaBoost
24	Gradient Tree Boosting
25	Multiclass classification
26	Multilabel classification
27	Multi-layer Perceptron

II.2.2. Muestra

La muestra de una investigación permite especificar parte de la población que será estudiada y de la cual se realizarán una serie de procedimientos para cumplir con los objetivos (USMP, 2016). Los algoritmos de clasificación Support Vector Machines Classification, Nearest Neighbors Classification, Decision Trees Classification, Neural

network Classification, Multi-layer Perceptron y Multinomial Naive Bayes son los seleccionados en la presente investigación, porque permiten optimizar el tratamiento de los datos ya que en el campo de la educación estos se encuentran anidados. Esta técnica de Machine Learning promete mejorar la educación implementando modelos de contingencia para prevenir la deserción. De acuerdo a lo redactado, se considera que la muestra es por conveniencia.

II.3. Variables, Operacionalización.

II.3.1. Variable independiente.

Método de Clasificación: Para el caso de la investigación son los algoritmos de muestra los cuales fueron evaluados en su entrenamiento con las métricas de Grado de consumo de CPU, Grado de consumo de memoria, Promedio de tiempo de respuesta.

II.3.2. Variable dependiente.

Deserción de estudiantes: Para el caso de la investigación son los factores influyentes de la deserción, estos fueron evaluados con las métricas de Exactitud, Precisión, Recall.

Tabla 2.

Operacionalización de variables.

Variables	Dimensión	Indicador	Ítem	Técnica e instrumentos de recolección de datos
Variable independiente	Consumo de recursos	Grado de consumo de CPU	$C_c = \sum_j^n \frac{cc_j}{n}$	Instrumentos mecánicos o electrónicos – Registro Electrónico
		Grado de consumo de memoria	$C_m = \sum_j^n \frac{cm_j}{n}$	
		Promedio de tiempo de respuesta	$T_r = \sum_j^n \frac{tf_j - tf_i}{n}$	
Variable dependiente	Rendimiento	Exactitud	$E = \frac{TP + TN}{TP + TN + FP + FN}$	
		Precisión	$P = \frac{TP}{TP + FP}$	
		Recall	$R = \frac{TP}{TP + FN}$	

Los datos serán tratados por los algoritmos que se emplearán en la presente investigación, estos tipos de tratamientos genera un impacto en los resultados según los procesos con los que se les evalúen y los métodos de clasificación que se usen. Es por ello que la presente investigación es de tipo cuasi experimental.

II.4. Técnicas e instrumentos de recolección de datos, validez y confiabilidad.

Instrumentos mecánicos o electrónicos.

Los instrumentos mecánicos electrónicos nos permiten la elaboración de estrategias para poder ordenar los datos en bruto, obtener la información exacta y con los cálculos esperados [51]. La presente investigación requiere del proceso y análisis de grandes volúmenes de datos para la obtención de resultados que predigan la deserción de estudiantes. Estos instrumentos nos ayudarán a evaluar la eficiencia de los algoritmos que se pondrán a prueba para diagnosticar cuál de ellos aporta el mejor rendimiento.

Ficha de Registro Electrónico.

La recolección de datos se hará por medio de fichas de registro electrónico, en ella se visualizarán los resultados que se hayan obtenido del desarrollo de los indicadores. De los ensayos obtendremos los resultados esperados para la presente investigación, es por ello que se requiere utilizar formatos de registro (Anexo 3) para los resultados de precisión, exactitud, Recall, consumo de memoria, consumo de CPU y tiempo de respuesta.

II.5. Procedimiento de análisis de datos.

Los datos recolectados a través de las fichas de registro serán filtrados, validados e interpretados con el objetivo de tener datos limpios, sin valores nulos o anómalos, con la finalidad de obtener información de calidad.

Se utilizará la herramienta de minería de datos de código abierto llamada Weka. Esta herramienta contiene diversos algoritmos de máquinas de conocimiento desarrollados por la universidad de Waikato (Nueva Zelanda).

Con respecto al consumo de recursos, las variables a utilizar son:

Grado de consumo de CPU:

$$C_c = \sum_j^n \frac{cc_j}{n}$$

Donde:

C_c : Es el grado de consumo de CPU del clasificador

cc_j : Es el grado de consumo de CPU en la prueba j

n : Es el total de pruebas

Grado de consumo de memoria

$$C_m = \sum_j^n \frac{cm_j}{n}$$

Donde:

C_m : Es el grado de consumo de memoria del clasificador

cm_j : Es el grado de consumo de memoria en la prueba j

n : Es el total de pruebas

Promedio de tiempo de respuesta

$$T_r = \sum_j^n \frac{tf_j - tf_i}{n}$$

Donde:

T_r : Es el tiempo de respuesta del clasificador

tf_j : Es el tiempo final de respuesta

tf_i : Es el tiempo inicial de respuesta

n : Es el total de pruebas

Exactitud

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

Donde:

TP : Es el total de Verdaderos Positivos.

TN : Es el total de Negativos Verdaderos.

FP : Es el total de Falsos Positivos.

FN : Es el total de Falsos Positivos.

Precisión

$$P = \frac{TP}{TP + FP}$$

Donde:

FP : Es el total de Falsos Positivos.

TP : Es el total de Verdadero Positivos.

Recall

$$R = \frac{TP}{TP + FN}$$

Donde:

FN : Es el total de Falsos Positivos.

VP : Es el total de Verdadero Positivos.

II.6. Criterios éticos.

Consentimiento informado: Las personas o instituciones para poder participar de un tema de investigación debe cumplir un procedimiento, este consiste en aceptar y firmar los acuerdos donde se compromete el consentimiento de información. Este acuerdo permite que los datos recolectados puedan utilizarse por los investigadores del proyecto para su análisis y publicaciones de los resultados obtenidos. Para el presente proyecto de investigación, se informó vía email a la institución Universidad Nacional de Jaén, acerca de la realización del mencionado proyecto, dando como respuesta total consentimiento para el desarrollo del mismo.

Confidencialidad: Es la garantía de que los datos obtenidos en la investigación serán protegidos para que no sean expuestos sin consentimiento de la persona y/o institución. Dicha garantía se realiza a través de normas y/o reglas que limitan el acceso a dichos datos. La presente investigación asegura la protección de la identidad de las personas involucradas en el proyecto.

Criterios de Rigor Científico.

Credibilidad o Valor de Verdad: La información presentada es totalmente verdadera, obtenida por parte de la institución donde se realizó la investigación.

La credibilidad contempla la evaluación de las situaciones en donde una investigación pueda ser reconocida como verosímil, para tal fin, es necesario la indagación de premisas fiables que puedan ser probados en las conclusiones de la investigación, en relación con el desarrollo seguido en la investigación.

Dependencia: Este criterio indica el nivel de estabilidad o consistencia de las conclusiones y descubrimientos de la investigación.

Originalidad: La información recopilada es citada en el apartado de referencias bibliográficas.

III. RESULTADOS.

III.1.Resultados en Tablas y Figuras.

Para el presente proyecto se obtuvo los registros de 358 estudiantes de la carrera de Ingeniería de Industrias Alimentarias de la universidad de Jaén, se conformó un Dataset de 22 atributos y 358 instancias, cada instancia está representada por un estudiante y se utilizaron los algoritmos de Random Forest, J48 y Naive Bayes, Support Vector Machine y RandomTree para el entrenamiento y poder obtener los resultados con respecto a las métricas de grado de consumo de CPU, grado de consumo de Memoria, promedio de tiempo de respuesta, Exactitud, Precisión y Recall.

III.1.1. Métricas de desempeño del consumo de recursos.

Para la obtención del indicador de consumo de CPU se tuvo en cuenta el consumo base antes de ejecutar las iteraciones por cada algoritmo y el consumo total luego de finalizada la iteración, esto con cada algoritmo clasificador, los resultados se muestran en la siguiente figura:

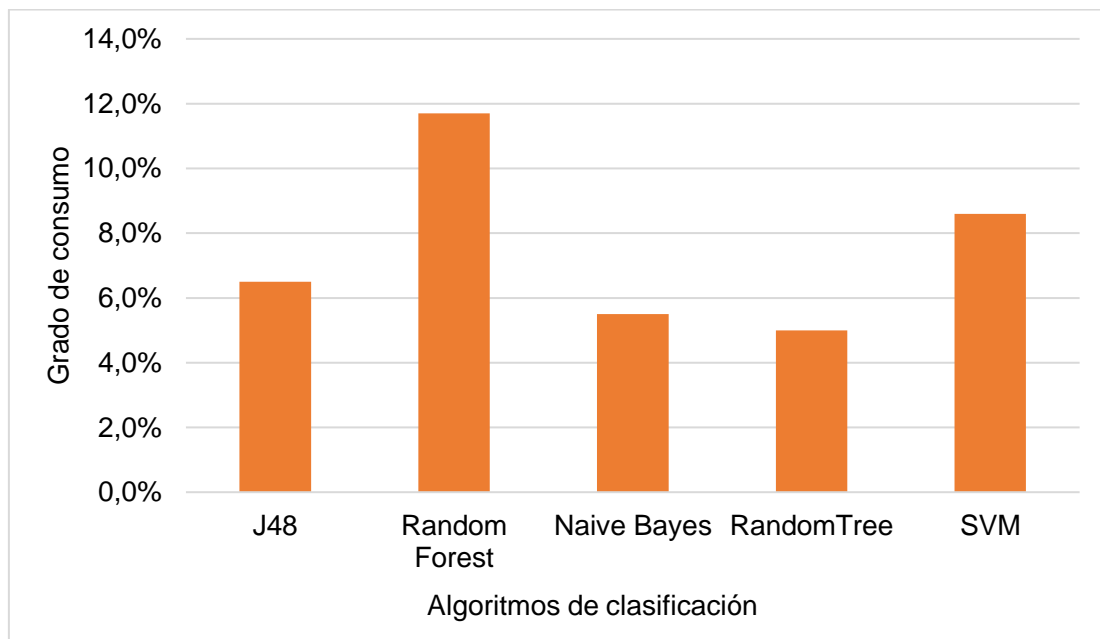


Fig. 12 Grado de consumo de CPU por cada iteración con cada algoritmo clasificador. Elaboración propia.

Fuente: Elaboración propia.

Para obtener el indicador de consumo de memoria RAM en el proceso de entrenamiento se tuvo en cuenta tener como punto base el consumo de memoria actual antes del entrenamiento de cada algoritmo y el consumo de memoria final en la finalización de la iteración del entrenamiento. Se realizaron 5 iteraciones, cada una para un algoritmo clasificador, como lo muestra la siguiente figura:

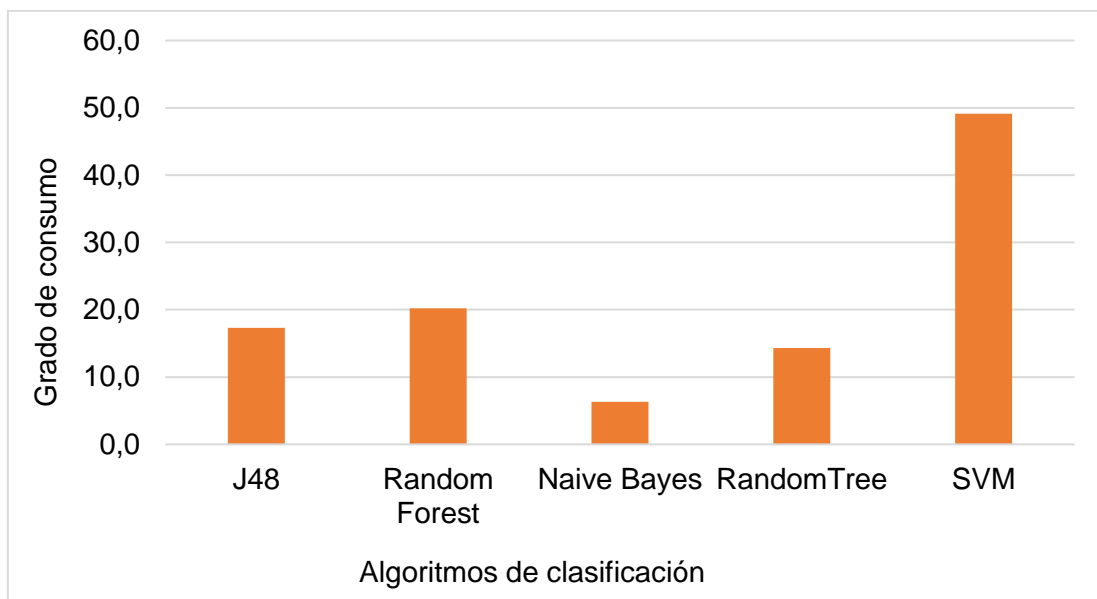


Fig. 13. Grado de consumo de memoria con cada algoritmo predictivo. Elaboración propia.

Fuente:

Elaboración propia.

En el entrenamiento de los algoritmos de clasificación J48, Random Forest, Naive Bayes, Support Vector Machine y RandomTree con los datos ya segmentados, se logró obtener la métrica de tiempo de respuesta en segundos, debido a que los datos obtenidos no son masivos, fue suficiente la utilización de una computadora de gama media con procesador Core I3 y 4 gigabyte de memoria RAM para los procesos de entrenamiento y muestra, siendo también la razón por la que estos procesos se ejecutaron en milésimas de segundos. El algoritmo Naive Bayes obtuvo el mejor desempeño ya que su tiempo de respuesta fue en 0 segundos y obteniendo el mejor desempeño como muestra la siguiente figura:

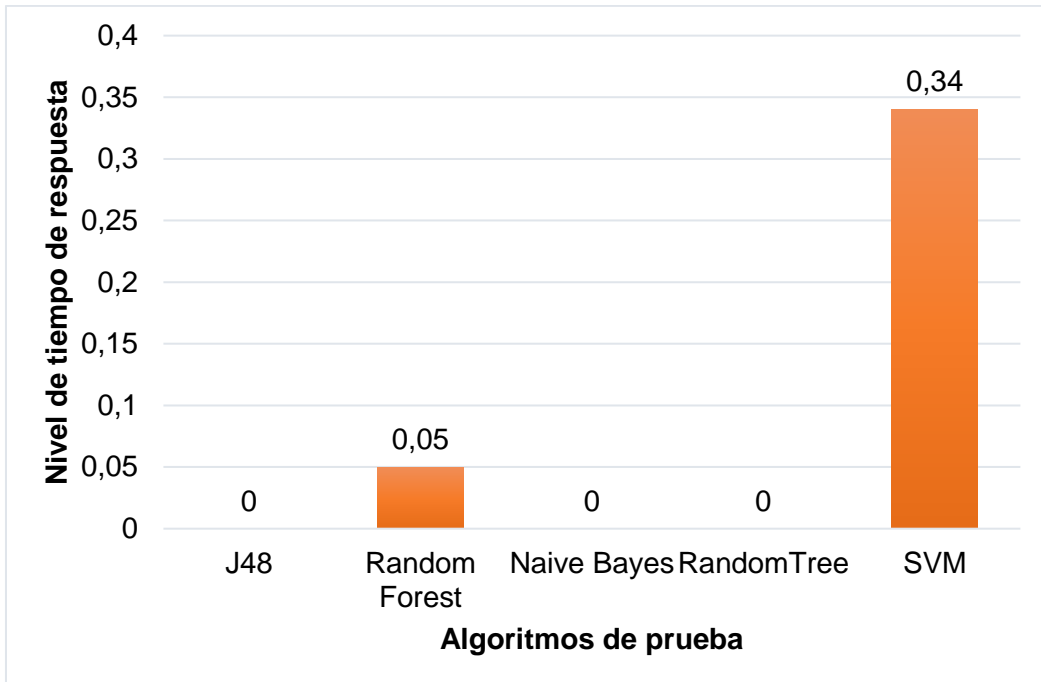


Figura 14. Evaluación del tiempo de respuesta de los clasificadores

de predicción. Elaboración propia.

III.1.2. Métricas de desempeño del método propuesto.

La métrica de exactitud en los 05 clasificadores puestos a prueba en la iteración del Dataset con 358 instancias, dieron un mejor desempeño para el algoritmo Support Vector Machine obteniendo un 98.88% de predicciones correctas sobre las predicciones hechas, estos resultados se obtuvieron de la suma de verdaderos positivos TP con los falsos positivos VN sobre la suma de los verdaderos positivos TP, verdaderos negativos TN, falsos positivos FP y los falsos negativos FN, la fórmula es la siguiente:

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

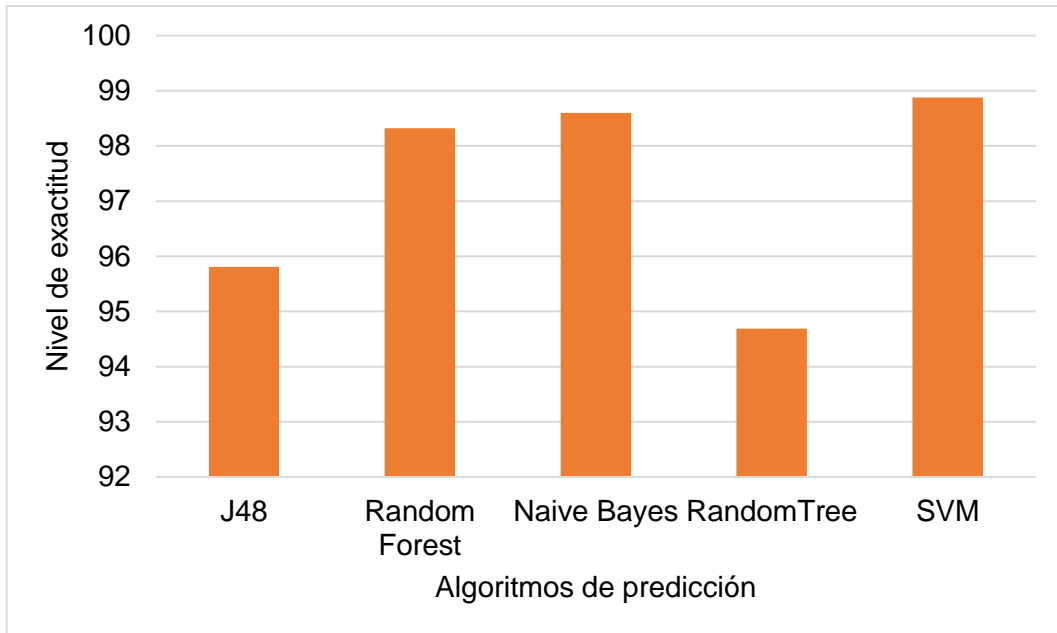


Figura 15. Resultados de Comparación de la métrica de Exactitud, obtenidas del número de predicción correctas. Fuente: Elaboración propia.

Las predicciones obtenidas a través del sistema de clasificación fueron mejorando de acuerdo al tratamiento y entrenamiento de los datos, la siguiente figura muestra la precisión que han tenido los 05 clasificadores, RandomForest y Support Vector Machine obtuvieron un mejor desempeño en cuanto al porcentaje de predicciones correctas, este resultado se obtiene del valor de los verdaderos positivos TP sobre la suma de los verdaderos positivos TP con los falsos positivos FP. Estos valores se obtienen a través de la matriz de confusión. La fórmula es la siguiente:

$$Precisión = \frac{TP}{TP + FP}$$

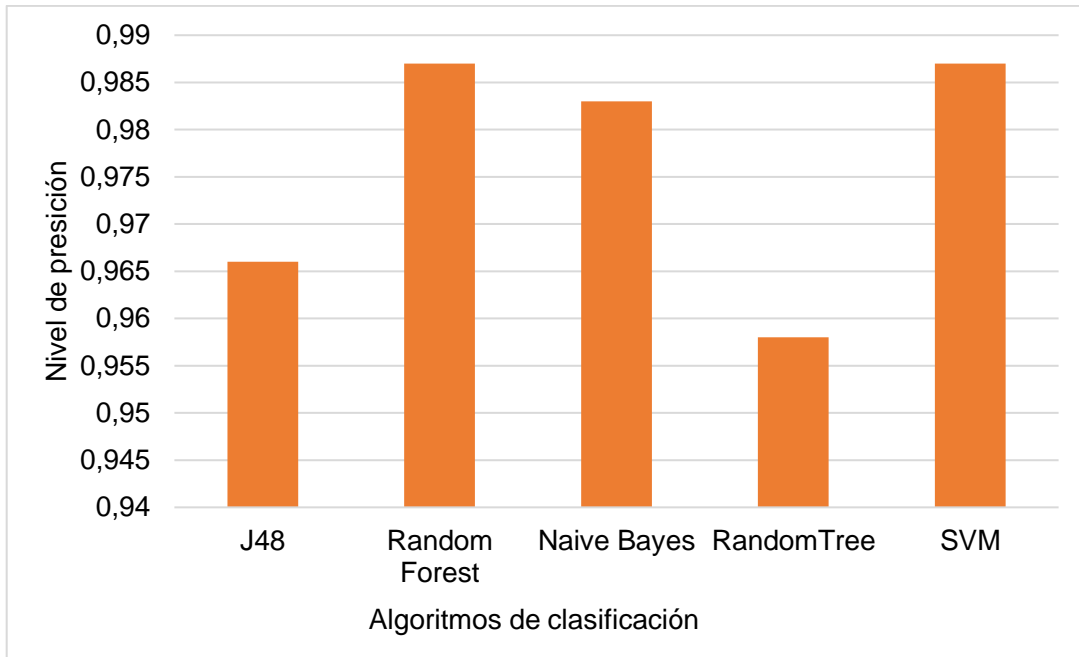


Figura 16. Resultados de Comparación de la métrica de Precisión, obtenidos de la matriz de confusión. Fuente: Elaboración Propia.

Para poder obtener el porcentaje de estudiantes desertores se muestra en la siguiente figura la métrica del Recall o Exhaustividad, de los 05 algoritmos clasificadores se ha podido observar que Naive Bayes y Support Vector Machines tiene un mejor desempeño a diferencias de los algoritmos J48, RandomTree y RandomForest. Para obtener el porcentaje de Recall se divide el total de verdaderos positivos TP sobre la suma de verdaderos positivos TP y falsos negativos, como se muestra en la siguiente formula:

$$Recall = \frac{TP}{TP + FN}$$

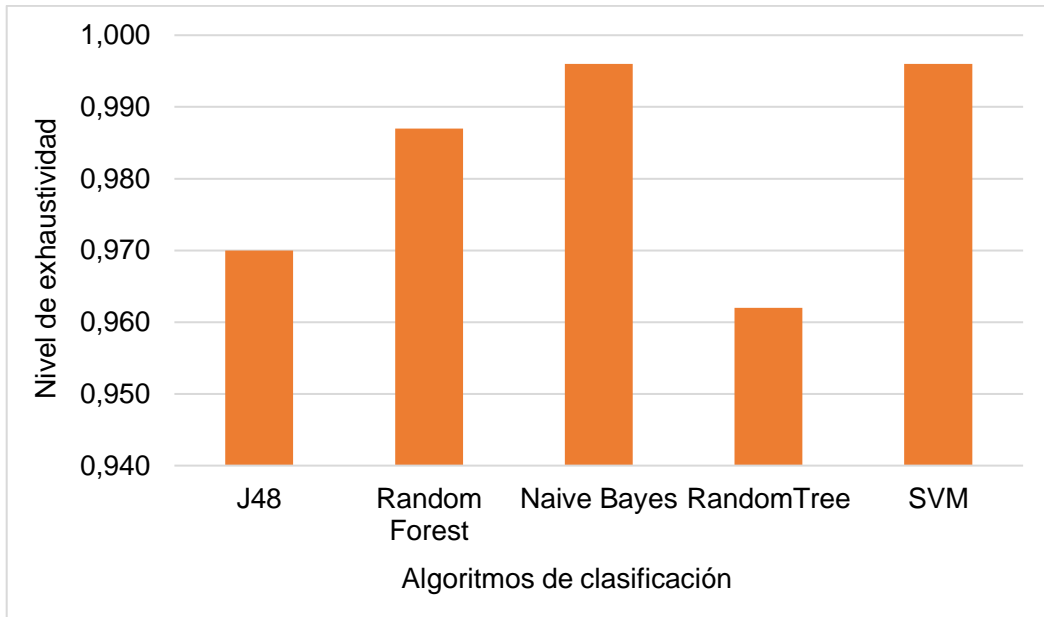


Fig. 17. Resultados de Comparación de la métrica de Recall en los algoritmos clasificadores de muestra.

Fuente: Elaboración propia.

Tabla 3.

Métricas de rendimiento de 5 algoritmos de clasificación para la predicción de la deserción de estudiantes.

Algoritmo Clasificador	Exactitud	Precisión	Recall
Support Vector Machine	98.61%	0.984%	0.995%
Naive Bayes	98.61%	0.980%	1.000%
Random Forest	97.22%	0.961%	1.000%
J48	97.22%	0.961%	1.000%
RandomTree	91.66%	0.922%	0.959%

Fuente: Elaboración propia.

El desempeño del método de clasificación Support Vector Machines procesando una muestra del 20% del Dataset equivalentes a 72 instancias entre estudiantes desertores

y no desertores obtuvo el siguiente resultado teniendo en cuenta los valores de predicción y los valores reales:

Tabla 4.

Resultado del método de clasificación con mejor desempeño en la detección de deserción de estudiantes.

Resultado	Cantidad	Matriz de confusión
Fueron detectados como desertores y lo son	49	VP
Fueron detectados como desertores y lo no son	22	VN
Fueron detectados como no desertores, pero si desertan	0	FN
Fueron detectados como desertores, pero no desertan	1	FP
Cantidad de registros de prueba	72(20%)	

Elaboración propia.

El método de clasificación Support Vector Machine evaluó 22 atributos asignando el peso a cada atributo según el grado de importancia del atributo en la detección de deserción de estudiantes. 08 de los atributos son considerados relevantes para alertar frente a un caso de abandono de estudio.

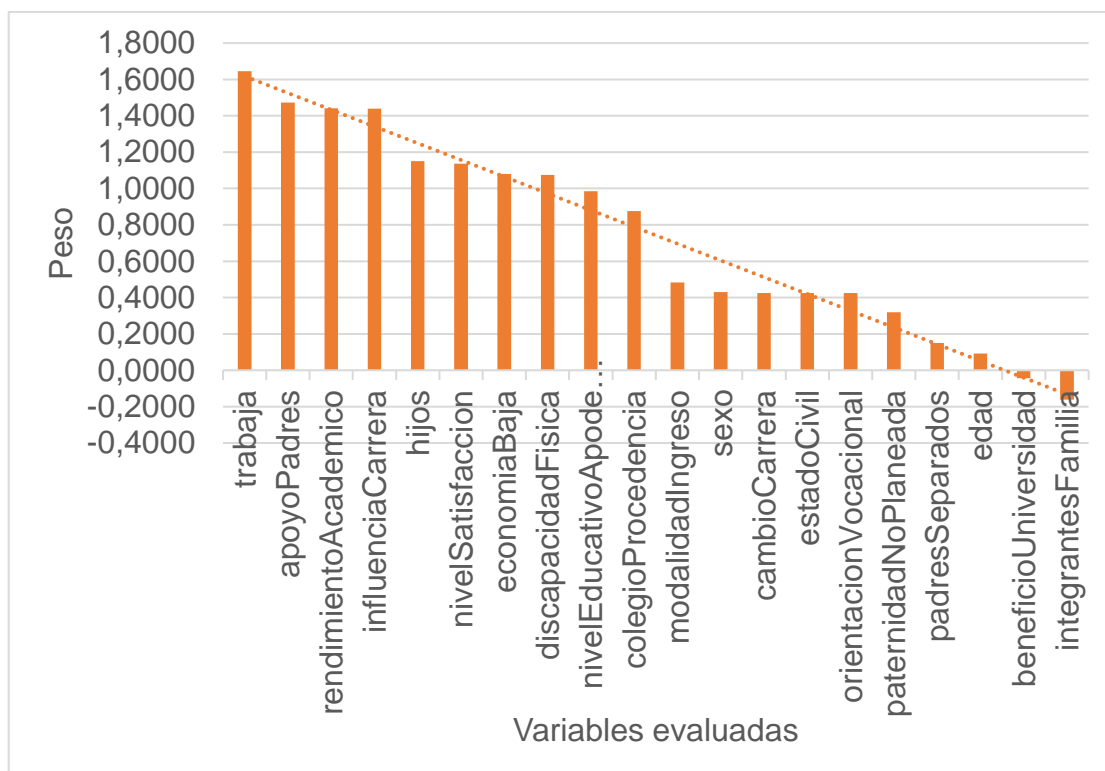


Fig. 18 Asignación de peso según el grado de importancia - Método de clasificación Support Vector Machine.

Fuente: Elaboración propia.

III.2. Discusión de resultados.

Los resultados obtenidos con el algoritmo Support Vector Machine guardan relación con lo mencionado por Cardona, Tatiana & Cudneya Elizabeth, ellos sostienen en su investigación que se debe utilizar el aspecto económico para mejorar la predicción de deserción estudiantil, esta variable ha servido para mejorar los factores de deserción en la presente investigación. De igual forma los autores utilizaron el algoritmo Support Vector Machine y obtuvieron un 78% en cuanto a precisión, mientras que el presente trabajo logro obtener un 98% en esa misma métrica. Cabe resaltar que el logro de este desempeño se debe a que el Dataset en su segunda fase de análisis fue transformado a datos numéricos y el algoritmo SVM tiene un mejor desempeño con este tipo de datos ya que es un algoritmo clasificador que trabaja con funciones.

Los algoritmos clasificadores de tipo árbol como J48, RandomForest y RandomTree mejoran su desempeño cuando son utilizados con filtros que permitan configurar sus hiper parámetros. Ramírez, Patricio & Grandón Elizabeth obtuvieron en su investigación una precisión del 87% con el algoritmos J48, en dicha investigación mencionan que al realizar ajustes en los hiper parámetros de este algoritmo antes del entrenamiento el desempeño suele aumentar, es por ello que para el presente caso de investigación se ha utilizado el filtro Discretize y se ha realizado un ajuste a nivel de instancia para que procese las iteraciones con rangos de 3 bins y mejorar la calidad de los datos antes de ser llevados al entrenamiento. El pretratamiento de los datos y ajuste de los hiper parámetros ayudo a obtener 95% de precisión.

El algoritmo RandomForest es un clasificador de predicción que puede manejar enormes cantidades de datos, tiende a procesar los datos con un método efectivo para valores nulos manteniendo la precisión aun cuando pueden faltar muchos datos. En la investigación de Solis Martin, Tania González Roberto, Fernández Tatiana & Hernández María se utilizó RandomForest para predecir la deserción estudiantil, los autores mencionan que antes de procesar los datos se tenían que eliminar registros incompletos, es así que obtuvieron un 94% de precisión. Para la presente investigación también se contó con un Dataset con muchos datos faltantes, pero con ayuda del filtro ReplaceMissingValues que utiliza la técnica de la media aritmética a nivel de atributos, se pudo replicar datos en los valores nulos, es así que el algoritmo clasificador RandomForest mejoro su desempeño y obtuvo un 98.32% de precisión. Eliminar registros o instancias que carecen de algunos valores puede generar una desventaja ya que en ese proceso también se pueden eliminar datos importantes que pueden aportar a una mejor predicción.

III.3. Aporte práctico.

En esta investigación se requiere contar con una universidad que cuente con la carrera de Ingeniería de Industrias Alimentarias. Para ello se ha tenido en cuenta 03 criterios de evaluación. Como primer criterio se necesita que la universidad sea licenciada, ya que este proceso garantiza que la casa de estudio cuenta con las condiciones básicas de calidad para ofrecer un servicio eficiente a los estudiantes, siendo así se tiene la seguridad que la información que procesan día a día es de calidad. La Superintendencia Nacional de Educación (SUNEDU) es el órgano que vela por este cumplimiento, es por ello que se ha utilizado su portal web para la búsqueda de la universidad. Como segundo criterio se tuvo en cuenta que la universidad tenga como mínimo 05 años formando estudiantes de la carrera de Ingeniería de Industrial Alimentarias. Así mismo el tercer criterio que se ha tomado en cuenta es la ubicación geográfica de la universidad, ésta deberá estar cerca al departamento de Lambayeque en caso de que se requiera de una visita presencial para formalidades que posibiliten la entrega de la información, ya que para manipular la información de una institución o empresa se debe tener en cuenta la integridad, confidencialidad y disponibilidad pertinente que esta amerita.

En el portal de SUNEDU se visualiza una lista de 92 universidades licenciadas y 02 escuelas de posgrado también licenciadas, la actualización de la lista tiene como fecha 04 de enero del año 2021. Las 94 casas de estudios se muestran en el anexo 04.

De las 92 universidades licenciadas, 23 cuentan con la carrera de Ingeniería de Industrias Alimentarias y en la siguiente tabla se ha realizado una evaluación para seleccionar la universidad con las que se va a trabajar aplicando los criterios establecidos.

Tabla 5.

Evaluación y selección de la universidad peruana que ofrece la carrera de ingeniería de industrias alimentarias.

N°	Nombre de la universidad	Criterio	Criterio	Criterio
		01	02	03
1	Universidad Nacional De San Agustín	Si	Si	No
2	Universidad Católica De Santa María	Si	Si	No
3	Universidad Nacional De San Cristóbal De Huamanga	Si	Si	No
4	Universidad Nacional De Cajamarca	Si	Si	No
5	Universidad Nacional De Jaén	Si	Si	Si
6	Universidad Nacional Agraria De La Selva	Si	Si	No
7	Universidad Autónoma De Ica	Si	Si	No
8	Universidad Nacional Del Centro Del Perú	Si	Si	No
9	Universidad Privada Antenor Orrego	Si	Si	Si
10	Universidad Nacional De Barranca	Si	Si	No
11	Universidad Nacional José Faustino Sánchez Carrión	Si	Si	No
12	Universidad Le Cordon Bleu S.A.C.	Si	Si	No
13	Universidad Peruana Unión	Si	Si	No
14	Universidad San Ignacio De Loyola	Si	Si	No
15	Universidad De San Martín De Porres	Si	Si	No
16	Universidad Nacional Agraria La Molina	Si	Si	No
17	Universidad Nacional De La Amazonía Peruana	Si	Si	No
18	Universidad Nacional Daniel Alcides Carrión	Si	Si	No
19	Universidad Nacional De Piura	Si	Si	Si

20	Universidad Nacional De Frontera	Si	Si	No
21	Universidad Nacional De Juliaca	Si	Si	No
22	Universidad Nacional Jorge Basadre Grohmann	Si	Si	No
23	Universidad Nacional Santiago Antúnez De Mayolo	Si	Si	No

Nota: La universidad se seleccionó teniendo en cuenta 03 criterios. Criterio 01: La universidad debe ser licenciada. Criterio 02: La universidad debe tener como mínimo 05 años formando profesionales en la carrera de Ingeniería de Industrias Alimentarias. Criterio 03: La universidad debe estar cerca al departamento de Lambayeque. Elaboración Propia.

De las 23 universidades que ofrecen la carrera de Ingeniería en Industrias Alimentarias se ha elegido la Universidad de Jaén por que cumple con los 03 criterios de evaluación.

Previo a la creación de las variables se hizo una revisión científica de los artículos seleccionados, que permitió determinar las variables más usadas por los autores y realizar un análisis de cuales podrían servir de apoyo para nuestra investigación.

Tabla 6.

Lista de variables creadas por los autores que realizaron estudios en casos similares de deserción estudiantil.

Autores	Características
Ramírez & Grandón (2018)	Edad, género, antecedentes de su ingreso a la universidad (puntaje de la prueba de selección universitaria y puntaje asociado a las notas de enseñanza media), aproximaciones a su situación económica (nivel de ingreso familiar y tipo de colegio de enseñanza media), y datos de su rendimiento

	académico (años de avance, promedio de notas y desviación estándar de notas).
Contreras, Fuentes & Rodríguez (2020)	Factores académicos preuniversitarios, Factores demográficos preuniversitarios, Factores socio-culturales preuniversitarios, Factor socio-económicos preuniversitarios, Factores de Gestión Académica universitaria, Factores Tecnológicos, Factores de Biblioteca, Factores Institucionales, Factores pedagógicos, Factores Intelectuales y Factores afectivos.
Quiñones, Jara, Alvarado, Milla & Gamarra (2020)	Comunidad, Año de nacimiento, género, ingreso al nivel secundario, egreso del nivel secundario, modalidad de ingreso a la universidad de Jaén, Carrera Profesional, cursos desaprobados, ciclo de ingreso.
Miranda & Guzmán (2017)	Estado académico, Año de admisión a la carrera, Año de egreso de enseñanza media del estudiante, Código carrera a la que pertenece el estudiante, Promedio ponderado PSU (Prueba de Selección Universitaria), Preferencia de postulación del estudiante, Nota promedio de los 4 años de Enseñanza Media (NEM), Beneficios estudiantiles, Calificaciones por semestre para cada asignatura cursada por el estudiante.

<p>González & Hernández (2019)</p>	<p>Edad, Código Postal, Grupo Étnico, Capacidad diferentes, número de personas en la familia, nivel de estudio del padre o tutor, nivel de ingreso económico.</p>
<p>Eckert & Suénaga (2015)</p>	<p>Condición de Deserción, Total de Finales Aprobados de 1º año, Proporción de Materias Cursadas del Año 1 (calendario), Cantidad de Fracasos de Cursado del Año 1 (calendario), Número de Finales Aprobados en el Año 1 (calendario), Promedio General de 1º Año, Promedio Materias Aprobadas de 1º Año, Edad de Ingreso, Establecimiento educativo (previo), Localización geográfica (de origen).</p>
<p>Solis, Moreira, González, Fernández, & Hernández (2018)</p>	<p>Alumno que lleva al menos dos semestres sin matricularse, Año de inscripción, Género, Residencia, beca de la institución durante el semestre, Recibió una subvención económica (préstamo) durante el semestre, Tipo de programa en el que está inscrito el estudiante, Lugar donde está matriculado el estudiante, Turno de la escuela, Estudiante admitido en la primera opción de carrera, El estudiante solicitó el cambio de carrera, Semestres no matriculados, Cursos aprobados en el semestre, Año de admisión a la universidad, Cursos necesarios para graduarse.</p>
<p>Cardonaa & Cudneya (2019)</p>	<p>Licenciatura, edad, género, planea trabajar, promedio en curso de matemática, promedio en curso de lenguaje, promedio en curso de inglés.</p>

<p>Viloriaa, Varelac, & Bonerge (2019)</p>	<p>Carrera, Curso, Discapacidad, Coste de la educación, Vive separado de la familia, Tipo de vivienda familiar, de la casa, Servicio de televisión por cable, Servicio de tarjeta de crédito, Servicio de Internet, Servicios básicos, Transporte privado, Servicio de plan telefónico, Servicio de coche propio, Viene en su propio coche, Actualmente trabaja, Aprobado, Abandono de los estudios.</p>
<p>Márquez, Romero, & Ventura, (2015)</p>	<p>Número de horas dedicadas estudiar a diario, métodos de estudio utilizados, lugar normalmente se utiliza para estudiar, tener su propio espacio para estudiar, recursos para el estudio, hábitos de estudio, estudiar en grupo, padres estímulo para el estudio, el estado civil, tener hijos, la religión, tener sanciones administrativas, el tipo de titulación elegida, la influencia en la titulación elegida, el tipo de personalidad, tener una discapacidad física, que sufre una enfermedad crítica, el consumo regular de alcohol, hábitos de fumar, nivel de ingresos de la familia, tener una beca, tener un trabajo, vivir con los padres, el nivel de educación de la madre, el nivel de educación del padre, el número de hermanos, la posición de hijo mayor/medio/joven, el hecho de vivir en un gran ciudad, número de años viviendo en la ciudad, método de transporte utilizado para ir a la escuela, distancia al escuela, nivel de asistencia durante las clases, nivel o aburrimiento durante las clases, interés en el asignaturas, nivel de dificultad de las mismas, nivel de motivación, toma de apuntes en clase, métodos de la enseñanza, la excesiva exigencia de deberes,</p>

	la calidad de la infraestructura escolar, el tener tutor personal, nivel de preocupación del profesor por el bienestar de cada alumno. Edad, sexo, escuela anterior, tipo de escuela, tipo de escuela secundaria, promedio de notas (GPA) en la escuela secundaria, ocupación de la madre, ocupación del padre, número de miembros de la familia, limitaciones para hacer ejercicios, frecuencia de los mismos, tiempo de ejercicios, puntuación obtenida en Lógica, puntuación en Matemáticas, puntuación en Razonamiento Verbal, puntuación en Español, puntuación en Biología, puntuación en Física, puntuación en Química, puntuación en Historia, puntuación en Geografía, puntuación en Cívica, puntuación en Ética, puntuación en Inglés.

El análisis de los factores y características que causan la deserción estudiantil usada por los autores en los artículos científicos seleccionados, ayudo a dimensionar las variables. En las tablas 4,5,6 y 7 se muestras las dimensiones, variables y la frecuencia con la que se han utilizado las variables por los autores.

Tabla 7.

Factores de dimensión personal

Ítem	variables	autor
1	Edad	Ramírez & Grandón (2018) Quiñones, Jara, Alvarado, Milla & Gamarra (2020)

		González & Hernández (2019) Cardona & Cudneya (2019) Márquez, Romero, & Ventura, (2015)
2	Genero	Ramírez & Grandón (2018) Quiñones, Jara, Alvarado, Milla & Gamarra (2020) Solis, Moreira, González, Fernández, & Hernández (2018)
3	Estado civil	Márquez, Romero, & Ventura, (2015)
4	Discapacidad	González & Hernández (2019) Viloriaa, Varelac, & Bonerge (2019) Márquez, Romero, & Ventura, (2015)
5	Nivel de satisfacción	Márquez, Romero, & Ventura, (2015)

Tabla 8.

Factores de dimensión académica.

Ítem	variables	autor
1	Tipo de colegio de procedencia	Ramírez & Grandón (2018) Quiñones, Jara, Alvarado, Milla & Gamarra (2020)

		Márquez, Romero, & Ventura, (2015)
2	Preparación universitaria	Contreras, Fuentes & Rodríguez (2020) Quiñones, Jara, Alvarado, Milla & Gamarra (2020)
3	Modalidad de ingreso	Quiñones, Jara, Alvarado, Milla & Gamarra (2020)
4	Número de veces de postulación	Solis, Moreira, Gonzáles, Fernández, & Hernández (2018)
5	Puntaje de ingreso en el examen de admisión	Ramírez & Grandón (2018)
6	Cambio de carrera	Solis, Moreira, Gonzáles, Fernández, & Hernández (2018)
7	Ciclo de ingreso	Quiñones, Jara, Alvarado, Milla & Gamarra (2020) Eckert & Suénaga (2015) Solis, Moreira, Gonzáles, Fernández, & Hernández (2018)
8	Cursos aprobados	Ramírez & Grandón (2018) Miranda & Guzmán (2017) Eckert & Suénaga (2015)

		Solis, Moreira, Gonzáles, Fernández, & Hernández (2018) Viloriaa, Varelac, & Bonerge (2019) Márquez, Romero, & Ventura, (2015)
9	Cursos desaprobados	Quiñones, Jara, Alvarado, Milla & Gamarra (2020) Solis, Moreira, Gonzáles, Fernández, & Hernández (2018)
10	Promedio ponderado	Ramírez & Grandón (2018)

Tabla 9.

Factores de dimensión económico.

Ítem	variables	autor
1	Beneficiario	Miranda & Guzmán (2017) Solis, Moreira, Gonzáles, Fernández, & Hernández (2018) Márquez, Romero, & Ventura, (2015)
2	Nivel de ingreso	Ramírez & Grandón (2018)

	familiar	Gonzáles & Hernández (2019) Márquez, Romero, & Ventura, (2015)
3	Nivel educativo del apoderado	Gonzáles & Hernández (2019) Márquez, Romero, & Ventura, (2015)
4	Número de integrantes en la familia	Gonzáles & Hernández (2019) Márquez, Romero, & Ventura, (2015)
5	Trabaja	Cardonaa & Cudneya (2019) Viloriaa, Varelac, & Bonerge (2019) Márquez, Romero, & Ventura, (2015)

Tabla 10.

Factores de dimensión social.

Ítem	variables	autor
1	Residencia distrito	Gonzáles & Hernández (2019) Solis, Moreira, Gonzáles, Fernández, & Hernández (2018)
2	Influencia de la	Márquez, Romero, & Ventura, (2015)

	carrera	
--	---------	--

Teniendo en cuenta el análisis de los estudiantes desertores de la carrera de Ingeniería de Industrias Alimentarias se han obtenido 15 variables, estas variables son consideradas según los autores como influyentes en el abandono de los estudios. En la siguiente tabla se muestra las variables con sus valores y referencias que serán usadas en el presente caso de investigación.

Tabla 11.

Creación de 22 variables con valores y referencias.

Ítem	Variable	Valores	Referencia
1	edad	16,17,18,19,20,21,22,23... más de 35	Define los rangos comprendidos de la edad desde los 16 años hasta más de 45 años.
2	género	M y F	Indica el género del estudiante.
3	estadoCivil	Soltero, Comprometido	Indica el estado del estudiante, si está soltero y comprometido si está casado o es conviviente
4	trabaja	Si, No	Indica si el estudiante está

			trabajando.
5	discapacidad	Si, No	Indica si el estudiante tiene alguna discapacidad que le impide avanzar su carrera de forma normal.
6	apoyoPadres	Si, No	Indica si el estudiante tiene el apoyo de sus padres para solventar los gastos de sus estudios superiores.
7	orientacionVocacion al	Si, No	Indica si el estudiante recibió charlas de orientación antes de escoger su carrera profesional.
8	rendimientoAcademico	>=18 & <=20 Excelente >=15 & <=17 Bueno >=11 & <=14 Regular <=10	Indica el puntaje obtenido por el estudiante en el examen de admisión.
9	beneficiario	Si, No	Indica si el estudiante tiene algún beneficio por parte de la universidad.
10	cambioCarrera	Si, No	Indica si el estudiante realizó el cambio de

			carrera.
11	modalidadIngreso	Centro Pre, Examen ordinario y Examen extraordinario - traslado interno	Indica la modalidad con la que ingreso el estudiante a la universidad.
12	nivelEducacionApod	Primario, Secundario, Técnico Superior, Grado Universitario.	Indica hasta que nivel educacional ha cursado el apoderado.
13	nivelSatisfaccion	Insatisfecho, Regular, Buena, Óptima	Indica si el estudiante se siente satisfecho de la enseñanza que recibe.
14	influenciaCarrera	Si, No	Indica si el estudiante fue influenciado por los padres o amigos para escoger la carrera universitaria.
15	Afeccionvicio	Si, No	Indica si el estudiante tiene algún vicio con consumo de drogas, alcohol o juegos de apuestas.
16	padraSeparados	Si, No	Indica la condición familiar de los padres
17	bajaEconomia	Si, No	Indica si el ingreso

			económico familiar es el mínimo vital designado por el estado peruano.
18	paternidadNoPlaneada	Si, No	Indica si el estudiante dejó la carrera por la procreación de un hijo no planeado.
19	colegioProcedencia	Nacional o Privado	Indica la procedencia del estudiante con respecto a su nivel secundario.
20	integrantesFamilia	1,2,3,4,5,6,7 o más	Indica cuantos integrantes tenía la familia cuando estudiaba su carrera profesional.
21	hijos	Si, No	Indica la cantidad de hijos que tenía el estudiante cuando cursaba su carrera profesional.
22	abandonoEstudios	Si, No	Indica si el estudiante dejó los estudios

Elaboración Propia.

Para la obtención de los registros de los estudiantes se realizó reuniones virtuales con la plana jerárquica de la universidad pertenecientes a la carrera de Ingeniería de

Industrias alimentarias y áreas de apoyo como Asuntos académicos, Biblioteca, Admisión y Calidad educativa.

La formación del Dataset en su primera versión tuvo 22 atributos y 358 instancias de estudiantes desertores y no desertores, pertenecientes a la carrera profesional de Ingeniería de Industrias Alimentarias, los estudiantes ingresaron en los años 2012 I, 2012 II, 2013 I, 2013 II, 2014 I, 2014 II, 2015 I, 2015 II, 2016 I, 2016 II. Cabe resaltar que se ha ocultado los nombres de los estudiantes para salvaguardar la identificación de los mismos.

Para la formación del Dataset en su versión final se ha tenido en cuenta la transformación de datos, esta transformación permitió tratar los datos categóricos en datos numéricos como se muestra en la siguiente tabla:

Tabla 12.

Conversión de datos categóricos a datos numéricos.

Variable	Datos categóricos	Datos numéricos
edad	>=17 & <=25	1
	>=26	2
sexo	Femenino	1
	Masculino	2
estadoCivil	Soltero	1
	Comprometido	2
trabaja	No	1
	Si	2

discapacidadFisica	No	1
	Si	2
apoyoPadres	Si	1
	No	2
orientacionVocacional	Si	1
	No	2
rendimientoAcademico	≥ 18 & ≤ 20 excelente	1
	≥ 15 & ≤ 17 bueno	2
	≥ 11 & ≤ 14 regular	3
	≤ 13 bajo	4
beneficioUniversidad	Si	1
	No	2
cambioCarrera	No	1
	Si	2
modalidadIngreso	Centro Pre	1
	Examen ordinario	2
	Examen extraordinario –	3
	Traslado interno	

nivelEducativoApoderado	Grado Universitario	1
	Superior Técnico	2
	Secundario	3
	Primario	4
nivelSatisfaccion	Optimo	1
	Bueno	2
	Regular	3
	Insatisfecho	4
influenciaCarrera	No	1
	Si	2
vicio	No	1
	Si	2
padresSeparados	No	1
	Si	2
economiaBaja	No	1
	Si	2
paternidadNoPlaneada	No	1
	Si	2

colegioProcedencia	Nacional	1
	Privado	2
integrantesFamilia	1,2,3,4,5,6,7 o más	
hijos	No	1
	Si	2
abandonoEstudios	No	1
	Si	2

Elaboración propia.

Una vez hecho la conversión de los datos se formó el Dataset en su versión final, que sería procesado con los filtros de discretización, limpieza, balanceo y luego con los algoritmos predictivos. En la siguiente tabla se muestra las 50 primeras instancias ya con los datos clasificados, las 308 instancias restantes se mostrarán en el anexo 05. Esto para evitar muchas hojas con datos redundantes y priorizar el desarrollo del caso práctico.

Cabe resaltar que la tabla a continuación muestra estudiantes desertores y no desertores y el atributo número 22 que es el atributo clase especifica con la afirmación “SI” si el estudiante es desertor y la afirmación “NO” refiere a que el estudiante culmino su carrera profesional.

Tabla 13.

Dataset en su versión final de estudiantes desertores y no desertores - Conversión de datos categóricos a datos numéricos - estudiantes de la carrera de Ingeniería de Industrias Alimentarias

edad	sexo	estado Civil	trabaja	discapacidad Física	apoyo Padres	orientación Vocacional	rendimiento Académico	beneficio Universidad	cambió Carrera	modalidad Ingreso	nivel Educativo Apoderado	nivel Satisfacción	influencia Carrera	vicio	padres Separados	economía Baja	paternidad No Planeada	colégio Procedeencia	integran familia	hijos	abandonó Estudios
1	0	0	0	0	0	0			0	0	0				0	0	0	0	7	0	NO
1	0	0	1		0	0	2		0	0	1			0	0	1	0	1			SI
1	0	0	0	0	0	0	2		0	2	2			0	0	0	0	0	5	0	NO

1	0	0	1	0	0	0	2		0	0	1	0		0	0	0	0	0	5	0	NO
1	0	0	0	0	0	0	2	1	0	0	1	0	0	0	0	0	0	0	6	0	NO
1	0	0	0	0	0	0	2		0	0			0	0	0	0	0	0	5	0	NO
1	0	0	0	0	0	0	2		0	0			0	0	0	0	0	0	4	0	NO
1	0	0	0	0	0	0	2		0	1			0	0	0	0	0	0	3	0	NO
1	1	0	0	0	0	0			0	0	0	1	0	0	0	0	0	0	5	1	NO
1	0	0	0	0	0	0	1		0	2		1	0	0	0	0	0	0	2	0	NO
1	1	0	1		0	0	1		0	0	1		1	0	0	0	0	1			SI
1	1	0	0	0	0	0	1		0	0		1	0	0	0	0	0	0	6	0	NO
1	1	0	0	0	0	0	1		0	0		1	0	0	0	0	0	0	6	0	NO
1	1	0	1	1	0	0	1	0	0	0	1		1	0	0	0	0	1			SI
1	1	0	0	0	0	0	1		0	2		1	0	0	0	0	0	0	3	0	NO
1	0	0	0	0	0	0	1		0	0		1	0	0	0	0	0	0	1	0	NO
1	1	0	0	0	0	0	1		0	0		1	0	0	0	0	0	0	5	0	NO
1	0	0	0	0	0	0	1		0	1		1	0	0	0	0	0	0	4	0	NO
1	1	0	0	0	0	0	1		0	2		1	0	0	0	0	0	0	3	0	NO
0	0	1	1		1	1	1		1	0	1	1	1	0	0	1	1	0	4		SI

0	1	1	1		1	1	1		1	0	1		1	0	1	1	0	0			SI
1	0	0	1		1	1	1		0	0	1		1	0	0	1	0	0			SI
1	0	0	1		1	1	3		0	0	1		1	0	0	1	0	0			SI
1	0	0	0	0	0		1	1	0	2	0	0	0	0	0	0	0	0	5	0	NO
1	0	0	1		1	1	3		0	0	1		1	0	0	1	0	0			SI
1	1	0	1		1	1	3		0	0	1			0	0	1	0				SI
1	0	0	0	0	0		1		0	0	0		0	0	0	0	0	0	5	0	NO
1	0	0	1	0	0		1		0	0			0	0	0	0	0	0	7	0	NO
1	0	0	1	0	0		1		0	0			0	0	0	0	0	0	2	1	NO
1	1	0	1	0	0		1		0	0			0	0	0	0	0	0	7	0	NO
1	0	0	1	0	0	0	1		0	2			0	0	0	0	0	0	7	0	NO
1	0	0	0	0	0	0	1	0	0	2	1	0	0	0	0	0	0	0	7	0	NO
1	1	0	0	0	0	0	1		0	0	1		0	0	0	0	0	0	3	0	NO
1	0	0	1		1	1	2	1	0	0	1		0	0	0	1	1	1	4		SI
1	1	0	1		1	1	2	1	0	0	1		0	0	0	1	0		1		SI
1	1	0	0	0	0	0	1		0	0	1		0	0	0	0	0	0	6	0	NO
1	0	0	0	0	0	0	1		0	0	1		0	0	0	1	0	0	2	0	NO

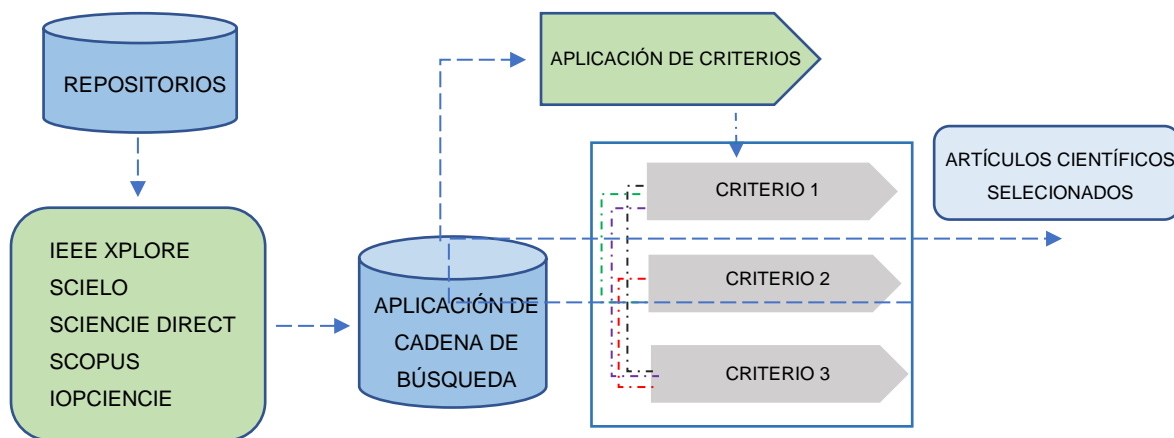
1	0	0	0	0	0	0	1		0	0	1		0	0	0	1	0	0	1	0	NO
1	0	0	0	0	0	0	1		0	0	1		0	0	0	1	0	0	1	0	NO
1	0	0	0	0	0	0	1		0	2	1		0	0	0	1	0	0	7	0	NO
0	0	0	1		1	1	2	1	0	0	1		0	0	0	1	0		6		SI
1	1	0	0	0	0	0	1		0	2	1		0	0	0	1	0	0	3	0	NO
1	1	0	0	0	0	0	1		0	0	1		0	0	0	1	0	0	1	0	NO
1	1	0	0	0	0	0	1		0	0	1		0	0	0	1	0	0	6	0	NO
1	0	0	0	0	0	0	1		0	0	1		0	0	0	1	0	0	7	0	NO
0	1	1	0		1	1	2	1	1	0	1	1	0	0	0	1	0	0	7		SI
1	1	0	0		1	1	2	1	0	0	1	1	0	0	1	1	0	0	2		SI
1	1	0	0	0	0	0	1	0	0	0	1		0	0	0	1	0	0	6	1	NO
1	0	0	0	0	0	0	1		0	2	1		0	0	0	1	0	0	4	0	NO

Nota: Se ha ocultado los nombres de los estudiantes para salvaguardar la identificación de los mismos y se ha priorizado los atributos que ayudaran a detectar la deserción estudiantil. Elaboración propia.

El caso de estudio requiere la realización de estimaciones para detectar la deserción de estudiantes basados en aprendizaje de máquina, estas estimaciones las podemos encontrar en artículos científicos donde se han realizado casos parecidos al de la presente investigación, por esa razón se ha realizado la selección de artículos científicos relevantes en predicción de deserción de estudiantes con algoritmos de aprendizaje de máquina.

Se tuvo en cuenta hacer búsquedas seguras en repositorios de prestigio como los son ScieceDirect, Iopciencie, Scielo y Scopus, estos son fuentes de información verídica que prestan un mayor aporte científico en sus investigaciones.

Se ha tenido en cuenta realizar un proceso de búsqueda como se muestra en la siguiente imagen.



Para la búsqueda de los artículos se han asignado tres criterios. El primer criterio puntualiza que los artículos científicos deber ser publicados a partir del año 2015, así mismo el segundo criterio refiere que los artículos deben de ser similares al caso de estudio y el tercer aplicar la búsqueda mediante cadenas de búsqueda.

Las publicaciones que fueron tomadas en cuenta son las que se encontraron con la siguiente cadena de búsqueda “Students AND desertion” y “Students AND desertion AND machine AND Learning” y se detallan los resultados en las siguientes tablas:

Tabla 14.

Búsqueda de artículos científicos sobre deserción de estudiantes en el repositorio ScieceDirect con cadena de búsqueda

Cadena de Búsqueda	Students AND desertion	Students AND desertion AND machine AND Learning
Cantidad	897	56
Artículos		
Artículos Interés	18	02
Artículos seleccionados	03	02

Tabla15.

Selección de artículos científicos de Predicción de Deserción de Estudiantes, basados en los criterios de evaluación.

Artículos Seleccionados – Repositorio ScieceDirect		
Autor (es)	Nombre de la Investigación	Año

Qian Fu, Zhanghao Gao, Junyi Zhou, Yafeng Zheng	Clsa: A Novel Deep Learning Model For Mooc Dropout Prediction	2021
Amelec Vilorio, Omar Pineda, Noel Varela	Bayesian Classifier Applied To Higher Education Dropout	2019
Jae Youngchung, Sunbokleeb	Dropout Early Warning Systems For High School Students Using Machine Learning	2019
Tatiana Cardona, Elizabeth Cudneya	Predicting Student Retention Using Support Vector Machines	2019
Concepción Burgos, María Campanario, David De La Peña, Juan Lara, David Lizcano, María Martínez	Data Mining For Modeling Students' Performance: A Tutoring Action Plan To Prevent Academic Dropout	2018

Elaboración propia.

Tabla 16.

Búsqueda de artículos científicos sobre deserción de estudiantes en el repositorio Scielo con cadena de búsqueda

Cadena de Búsqueda	Students AND desertion	Students AND desertion AND machine AND Learning
--------------------	---------------------------	--

Cantidad	47	0
Artículos		
Artículos	05	0
Interés		
Artículos	04	0
seleccionados		

Tabla 17.

Selección de artículos científicos de Predicción de Deserción de Estudiantes, basados en los criterios de evaluación.

Artículos Seleccionados – Repositorio Scielo		
Autor (es)	Nombre de la Investigación	Año
Contreras Leonardo, Fuentes Héctor, Rodríguez José	Academic Performance Prediction By Machine Learning As A Success/Failure Indicator For Engineering Students	2020
Ramírez Patricio, Grandón Elizabeth	Prediction Of Student Dropout In A Chilean Public University Through Classification Based On Decision Trees With Optimized Parameters	2018

Miranda Mauricio, Guzmán Jheser	Analysis Of Dropouts Of University Students Using Data Mining Techniques	2017
Eckert Karina, Suénaga Roberto	Analysis Of Attrition-Retention Of College Students Using Classification Technique In Data Mining	2015

Elaboración propia.

Tabla 18.

Búsqueda de artículos científicos sobre deserción de estudiantes en el repositorio IopCiencie con cadena de búsqueda.

Cadena de Búsqueda	Students AND desertion	Students AND desertion AND machine AND Learning
Cantidad	07	00
Artículos		
Artículos Interés	02	00
Artículos seleccionados	02	00

Tabla 19. Selección de artículos científicos de Predicción de Deserción de Estudiantes, basados en los criterios de evaluación.

Artículos Seleccionados – Repositorio IopCiencie		
Autor (es)	Nombre de la Investigación	Año
Márquez Vera, Carlos	Data Mining Applied In School Dropout Prediction	2015
Diaz Pedroza, Chindoy Chasoy, Rosado Gómez	Review Of Techniques, Tools, Algorithms And Attributes For Data Mining Used In Student Desertion	2019

Elaboración propia.

Tabla 20. Búsqueda de artículos científicos sobre deserción de estudiantes en el repositorio Scopus con cadena de búsqueda.

Cadena de Búsqueda	Students AND desertion	Students AND desertion AND machine AND Learning
Cantidad	30	01
Artículos		
Artículos Interés	11	01

Artículos seleccionados	04	01
----------------------------	----	----

Tabla 21.

Selección de artículos científicos de Predicción de Deserción de Estudiantes, basados en los criterios de evaluación.

Artículos Seleccionados – Repositorio IopCiencie		
Autor (es)	Nombre de la Investigación	Año
Márquez Vera, Carlos	Data Mining Applied In School Dropout Prediction	2015
Wan Yaacob, Mohd Sobri, Md Nasir, Wan Yaacob, Norshahidi, Wan Husin	Predicting Student Drop-Out In Higher Institution Using Data Mining Techniques	2020
Diaz Pedroza, Chindoy Chasoy, Rosado Gómez	Review Of Techniques, Tools, Algorithms And Attributes For Data Mining Used In Student Desertion	2019

Elaboración propia.

Se listaron los algoritmos con mejor desempeño en predicción de estudiantes que abandonan sus estudios, teniendo en cuenta para esta selección cumplir con 02 criterios que son los siguientes:

- I. Seleccionar los artículos científicos que tengan un rendimiento mayor al 70% de precisión de predicción.
- II. El escenario de aplicación deberá ser a estudiantes desertores de una universidad en la modalidad presencial.

Para el cumplimiento del primer criterio se ha establecido la comparación cualitativa de los artículos científicos que se han detallado en los trabajos previos, teniendo en cuenta que los artículos que tengan un porcentaje menor o igual al 70% serán considerados como “Bajo rendimiento”, los artículos que tengan mayor al 70% y menor o igual al 85% se consideran como “Rendimiento moderado” y los mayores de 85% se consideran como “Mejor rendimiento”.

Tabla 22.

Modelo de calificación para el desempeño de predicción en los trabajos previos citados.

Rendimiento	Calificación
$\leq 70\%$	Bajo rendimiento
$>70\%$ y ≤ 85	Rendimiento moderado
$>85\%$	Buen rendimiento

Para el cumplimiento del segundo criterio se tendrá en cuenta que el escenario de aplicación sea la modalidad presencial en la casa de estudio donde se ha practicado el caso de investigación, esto debido a que hay casos de investigación del fenómeno de deserción de estudiantes en la modalidad virtual con cursos online que ofrecen plataformas académicas conocidas como Learning Management System (LSM

Tabla 23.

Evaluación y selección de Técnicas de Predicción de algoritmos de aprendizaje automático

N°	Caso de investigación	Algoritmo	Precisión de predicción	Modalidad	Calificación	Seleccionado
1	Prediction of Student Dropout in a Chilean Public University through Classification based on Decision Trees with Optimized Parameters SCIELO	J48	87.27%	Presencial	Rendimiento moderado	Si
2	Academic performance prediction by machine learning as a success/failure indicator for engineering Students	Perceptrón	66.24%	Presencial	Bajo rendimiento	No
3	Awajún and Wampis Student Dropout Estimation Model Using Data Mining	J48	45.00%	Presencial	Bajo rendimiento	No
4	Development of software to predict academic performance	Randomizable	92.63%	Presencial	Buen rendimiento	Si

	using data mining techniques and tools IOP CIENCIE					
5	Analysis of Dropouts of University Students using Data Mining Techniques SCIELO	Red neuronal	73.00%	Presencial	Rendimiento moderado	Si
6	Analysis of Attrition-Retention of College Students Using Classification Technique in Data Mining SCIELO	Naive Bayes	81.10%	Presencial	Bajo rendimiento	Si
7	Perspectives to Predict Dropout in University Students with Machine Learning IEEE	Random Forest	93.00%	Presencial	Buen rendimiento	Si
8	Investigación A predictive model for identifying students with dropout profiles in online courses RESEARCH	Máquina de Vectores de Soporte	92.30%	Virtual	Buen rendimiento	No
9	Bayesian Classifier Applied to Higher	Naive Bayes	91.36%	Presencial	Buen rendimiento	Si

	Education Dropout SICIENDIRECT					
10	Predicting of school failure using data mining. SICIENDIRECT	RandomTree	97.20%	Presencial	Buen rendimiento	Si
11	Predicting Student Retention Using Support Vector Machines.	SVM	78.00%	Presencial	Bajo rendimiento	SI

Nota: Las Técnicas de predicción que serán seleccionadas son las que se consideran con “Rendimiento moderado” o “Buen rendimiento” y con la modalidad del caso de estudio “Presencial”. Elaboración propia.

Tabla 24.

Selección de algoritmos con mejor desempeño en deserción de estudiantes para procesos de prueba.

Investigación	Método de Clasificación	Precisión de Predicción
Predicting Student Retention Using Support Vector Machines.	SVM	78.00%
Prediction of Student Dropout in a Chilean Public University through Classification based on Decision Trees with Optimized Parameters	J48	87.27%
Perspectives to Predict Dropout in University Students with Machine Learning	Random Forest	93.00%
Bayesian Classifier Applied to Higher Education Dropout	Naive Bayes	91.36%
Predicting of school failure using data mining	RandomTree	97.20%

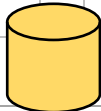
Los algoritmos J48, Random Forest, RandomTree, Naive Bayes y Support Vector Machine han sido seleccionados y se utilizarán para los procesos de prueba que se

realizarán con la data obtenida de la universidad de Jaén para detectar el fenómeno de deserción.

El método de clasificación implementado ayudo en tener en cuenta el proceso que se llevará a cabo para las tareas que se requieren en la predicción de deserción estudiantil, desde la selección de los datos hasta la extracción del conocimiento, interpretación y evaluación.

X	Y	X	Y
Y	Y	X	Y
X	X	X	X
Y			Y

DATA



FILTRADO

EXTRACCIÓN

TRANSFORMACIÓN

CLASIFICACIÓN

NUEVO CONOCIMIENTO

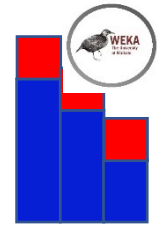
DISCRETIZACIÓN DE LOS DATOS

EVALUACIÓN DE LOS DATOS CON LOS ALGORITMOS DE CLASIFICACIÓN

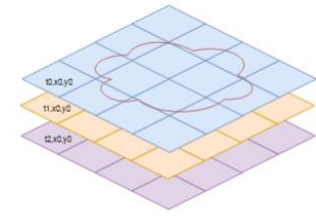
ARCHIVOS EN EXCEL



IDENTIFICACIÓN DE VARIABLES



PRE PROCESAMIENTO DE LOS DATOS



OBTENCIÓN DE PATRONES DE COMPORTAMIENTO

Los algoritmos de Machine Learning son una ciencia empírica, y los filtros utilizados para evaluar el desempeño de predicción suele favorecer a unos y perjudicar a otros, es por ello que para la presente investigación se ha realizado diferentes experimentos con los algoritmos de muestra hasta obtener un método que pueda realizar predicciones eficientes. Los experimentos se detallan en las diferentes tablas durante del desarrollo del aporte practico.

Para la evaluación de los algoritmos de aprendizaje de maquina se utilizó el programa de Weka versión 3.8, este programa permite discretizar la información a través de filtros supervisados y no supervisados para mejorar el desempeño de resultados en un Dataset, a comparación de otros programas de análisis de datos este permite evaluar grandes volúmenes de datos e interactuar con los filtros a nivel de instancias y atributos asi como también el entranamiento y prueba para verificar el rendimiento de predicción. Debido a que la información a procesar no era de gran volumen se utilizó una computadora de gama media con las características de CPU Core I3 con 3.2 MGZ, 04 Gigabytes de memoria RAM y con el sistema operativo Windows 10 home con arquitectura de 64 bits.

Como primer punto se tuvo en cuenta el análisis de los archivos de hoja de cálculo donde se contenía la data, se hizo la extracción de atributos ya que al tener varias fuentes de información (archivos de hoja de cálculo y formularios) es normal tener datos redundantes, cabe resaltar que estos datos no aportan credibilidad al método predictivo que se requiere implementar. Los atributos redundantes fueron la edad, estado civil, género, abandono de estudios, trabaja y discapacidad.

Los archivos fueron convertidos a un formato CSV (Valores separados por comas) para que pueda ser procesado por lo filtros de predicción. Se tuvo en cuenta cambiar el separador de instancias en el Dataset ya que un archivo CSV tiene el signo punto y coma (;) como separador de instancias y la estructura de archivos planos tiene la coma (,) como separados de instancias.

Antes de aplicar los filtros para la limpieza de los datos se hizo la conversión del atributo abandono a clase ya que es el atributo con mayor peso, también se hizo una prueba

de resultado de línea base con el algoritmo ZeroR para tener un resultado base de referencia hacia los algoritmos predictivos que se medirán en adelante. En esta prueba base se obtuvo una precisión de 0.65% es decir 235 instancias correctas y 123 instancias incorrectas, como muestra la siguiente tabla:

Tabla 25:

Proceso de prueba con el Algoritmos ZeroR para obtener el resultado base como punto de referencia antes de medir los algoritmos predictivos de muestra.

Correctly Classified Instances	235	60.6425 %
Incorrectly Classified Instances	123	34.3575 %
Kappa statistic		0
Total, Number of Instances	358	

Tabla 26.

Matriz de confusión - Algoritmos ZeroR

a	b	<- classified as	
235	0		a = No

123	0		b = Si

Luego de obtener el resultado base con el algoritmo ZeroR, se hizo un primer procesamiento de cada algoritmo de clasificación, J48, Random Forest, Random Tree, Naive Bayes y Support Vector Machines.

Primero proceso de entrenamiento

Para el algoritmo J48 en su proceso con 10 pliegues, se obtuvo una precisión del 96.6% y un error del 3.55% con una coherencia de datos de 0.9257 y con una matriz de confusión que se acerca a la clasificación correcta del total de las instancias, como se muestra en las siguientes tablas:

Tabla 27.

Primer proceso - Matriz de confusión del algoritmo J48 con Validación cruzada estratificada de 10 pliegues.

a	b	<- classified as	
229	6		a = No
6	117		b = Si

Elaboración propia.

Tabla 28.

Coefficiente de Sensibilidad – Primer proceso del algoritmo J48.

Tasa de verdaderos positivos	Tasa de falsos positivos
0,974	0,049
0,951	0,026

Elaboración propia.

Tabla 29.

Precisión de predicción del Algoritmo J48 – Primer proceso.

Correctly Classified Instances	346	96.64%
Incorrectly Classified Instances	12	3.35%
Kappa statistic		0.9257
Total, Number of Instances	358	

Elaboración propia.

Para el algoritmo Naive Bayes se obtuvo una precisión del 96.2% y un error del 3.39% con una coherencia de datos de 0.9489 con una matriz de confusión que se acerca a la clasificación correcta de instancias procesadas.

Tabla 30.

Primer proceso - Matriz de confusión del algoritmo Naive Bayes con Validación cruzada estratificada de 10 pliegues.

a	b	<- classified as	
233	2		a = NO
5	118		b = Si

Elaboración propia.

Tabla 31.

Coefficiente de Sensibilidad – Primer proceso del algoritmo Naive Bayes.

Tasa de verdaderos positivos	Tasa de falsos positivos
0,994	0.035
0,965	0.006

Elaboración propia.

Tabla 32.

Precisión de predicción del Algoritmo Naive Bayes – Primer proceso.

Correctly Classified Instances	351	0.9489 %
Incorrectly Classified Instances	7	0.3511 %
Kappa statistic		0.9489
Total Number of Instances	358	

Elaboración propia.

Para el algoritmo Random Forest se obtuvo una precisión del 97.76% y un error del 2,23% con una coherencia de datos de 0.9503 con una matriz de confusión que se acerca a la clasificación correcta de instancias procesadas.

Tabla 33.

Primer proceso - Matriz de confusión del algoritmo Random Forest con Validación cruzada estratificada de 10 pliegues.

a	b	<- classified as	
232	3		a = NO

5	118		b = Si

Tabla 34.

Coefficiente de Sensibilidad – Primer proceso del algoritmo Random Forest.

Tasa de verdaderos positivos	Tasa de falsos positivos
0,987	0.041
0,959	0.013

Tabla 35.

Precisión de predicción del Algoritmo Random Forest – Primer proceso.

Correctly Classified Instances	350	97.7654 %
Incorrectly Classified Instances	8	2.2346 %
Kappa statistic		0.9529
Total Number of Instances	63	

Elaboración propia.

Para el algoritmo RandomTree se obtuvo una precisión del 95.53% y un error del 4,46% con una coherencia de datos de 0.9001 con una matriz de confusión que se acerca a la clasificación correcta de instancias procesadas.

Tabla 36.

Primer proceso - Matriz de confusión del algoritmo RandomTree con Validación cruzada estratificada de 10 pliegues.

a	b	<- classified as	
229	6		a = NO
10	113		b = Si

Tabla 37.

Coefficiente de Sensibilidad – Primer proceso del algoritmo RandomTree.

Tasa de verdaderos positivos	Tasa de falsos positivos
0,974	0.081
0,919	0.026

--	--

Tabla 38

Precisión de predicción del Algoritmo Random Tree – Primer proceso.

Correctly Classified Instances	342	95.5307 %
Incorrectly Classified Instances	16	4.4693 %
Kappa statistic		0.9001
Total Number of Instances	358	

Elaboración propia.

Para el algoritmo Support Vector Machine se obtuvo una precisión del 98.88% y un error del 1,11% con una coherencia de datos de 0.9751 con una matriz de confusión que se acerca a la clasificación correcta de instancias procesadas.

Tabla 39.

Primer proceso - Matriz de confusión del algoritmo Support Vector Machine con Validación cruzada estratificada de 10 pliegues.

a	b	<- classified as	
234	1		a = NO
3	120		b = Si

Tabla 40.

Coefficiente de Sensibilidad – Primer proceso del algoritmo Support Vector Machine.

Tasa de verdaderos positivos	Tasa de falsos positivos
0,996	0.024
0,976	0.004

Tabla 41.

Precisión de predicción del Algoritmo Support Vector Machine – Primer proceso.

Correctly Classified Instances	354	98.8827 %
Incorrectly Classified Instances	4	1.1173 %
Kappa statistic		0.9751
Total Number of Instances	358	

Elaboración propia.

Primer experimento con el filtro Discretize.

Luego de los resultados obtenidos se hizo un análisis de los histogramas de cada atributo con respecto a sus instancias para detectar si alguno de ellos necesitaba de una discretización y mejorar el rendimiento de los algoritmos que se han puesto a prueba.

La discretización realiza un proceso de creación de conjuntos de intervalos finitos para los datos ayudando a mejorar la calidad de los resultados y eliminar el ruido, este proceso se le conoce como Binning, para el presente caso de investigación se usó la discretización no supervisada de igual anchura con el filtro Discretize, convirtiendo los atributos en intervalos con un rango de precisión de 0 a 6.

Al realizar este primer experimento se pudo observar que el desempeño favoreció al algoritmo Naive Bayes y Random Forest y perjudico a J48, Random Tree y Support Vector Machine, como se puede apreciar en la siguiente tabla:

Tabla 42.

Comparación de resultados de los algoritmos clasificadores con el filtro Discretize

Algoritmo Clasificador	Exactitud	Precisión	Recall
J48	94.97%	0.966%	0.957%
Random Forest	98.04%	0.979%	0.991%
Naive Bayes	97.48%	0.967%	0.996 %
RandomTree	95.53%	0.958%	0.974%
Support Vector Machine	98.16%	0.982%	0.994

Segundo experimento- Uso de la media aritmética con el filtro ReplaceMissingValues para llenado de datos faltantes.

Para mejorar el desempeño de los algoritmos también se tuvo en cuenta tratar los datos faltantes, aquellos que no registran valores disponibles y que pueden ser muy importante al momento de hacer una nueva iteración con los algoritmos de entrenamiento, para ello se utilizó el filtro no supervisado ReplaceMissingValues a nivel de atributos, pudiendo observar que el origen de los histogramas pasó de sesgos incoherentes a sesgos de igual proporción y que los datos faltantes fueron remplazados por datos de la media aritmética de cada atributo, la técnica de la media aritmética permite sumar los valores que se tienen como muestra en el atributo, dividirlos por el mismo número de valores y añadir valores a las celdas vacías. Para los siguientes resultados se puede observar que la técnica de ReplaceMissingValues ayudo a mejorar el desempeño de J48, Random Forest, Random Tree, mientras que Naive Bayes y Support Vector Machine mantuvieron su desempeño.

Tabla 43.

Comparación de resultados de los algoritmos clasificadores con el filtro ReplaceMissingValues.

Algoritmo Clasificador	Exactitud	Precisión	Recall
J48	98.88%	0.996%	0.987%
Random Forest	99.72%	1.000%	0.996%
Naive Bayes	96.08%	0.974%	0.966 %
RandomTree	97.76%	0.987%	0.979%

Support Vector Machine	98.88%	0.987%	0.996
------------------------	--------	--------	-------

Tercer experimento - Uso del filtro Normalize

Al observar los conjuntos de datos y los histogramas en el Dataset se pudo detectar que muchos de los atributos no tenían una distribución equitativa, respecto a ello se utilizó el filtro Normalize, esta técnica permite obtener una distribución mejorada de los datos conocida como distribución gaussiana. Se realizó un nuevo experimento para ver el rendimiento de los algoritmos con el filtro mencionada, teniendo como resultado que el clasificador Naive Bayes aumento en un 2% su exactitud en la predicción, mientras que J48, Random Forest, Random Tree y Support Vector Machine mantuvieron su desempeño.

Tabla 44.

Comparación de resultados de los algoritmos clasificadores con el filtro Normalize.

Algoritmo Clasificador	Exactitud	Precisión	Recall
J48	96.64%	0.974%	0.974%
Random Forest	98.04%	0.983%	0.987%
Naive Bayes	98.60%	0.983%	0.996 %
RandomTree	96.08%	0.966%	0.974%
Support Vector Machine	98.88%	0.987%	0.996

Cuarto experimento – Uso de filtros de balanceo a nivel de instancias.

Luego de haber procesado el Dataset con filtros supervisado y no supervisados a nivel de atributo se procedió a utilizar filtros a nivel de instancias para seguir mejorando el desempeño de los algoritmos predictivos.

Se detecto que los atributos como estado civil, discapacidad física, beneficio universitario y cambio de carrera tenían problema de desequilibrio de datos, por lo que se utilizó los filtros de balanceo para tratar de aumentar el buen desempeño de los algoritmos de muestra.

Los filtros Resample, Class Balancer y SpreadSubSample permiten crear un equilibrio de los datos es por ello que se realizó una comparación de los algoritmos J48, Naive Bayes, RandomTree, Random Forest y Support Vector Machines y visualizar cuales de los algoritmos mejoran su desempeño con relación a los resultados anteriores.

Durante el proceso de pruebas se puso observar que los filtro tienen diferentes comportamientos para cada algoritmo, mejorando el desempeño de uno y bajando el desempeño para otros, así como se muestran en las siguientes tablas:

Tabla 45. *Comparación de filtros de balanceo para mejorar el desempeño de resultados del algoritmo J48.*

J48	Exactitud	Precisión	Recall
Sin filtros de balanceo	96.64%	0,974 %	0.974 %
Resample	97.76%	0.989%	0,966%
Class Balancer	97.27%	0,983%	0,972 %
SpreadSubSample	94.30%	0.966%	0.919%

Nota. Para el algoritmo J48 el filtro Resample es el que mejora su desempeño de predicción.

Elaboración Propia

Tabla 46.

Comparación de filtros de balanceo para mejorar el desempeño de resultados del algoritmo Naive Bayes

Naive Bayes	Exactitud	Precisión	Recall
Sin filtros de balanceo	98.60%	0.983 %	0.996%
Class Balancer	98.56%	0.976%	0.996%
SpreadSubSample	98.37%	0.976%	0.972%
Resample	98.04%	0.973%	0.989%

Para el algoritmo Naive Bayes el filtro Class Balancer mejora su desempeño de predicción

Elaboración Propia

Tabla 47.

Comparación de filtros de balanceo para mejorar el desempeño de resultados del algoritmo RandomForest.

RandomForest	Exactitud	Precisión	Recall
Sin filtros de balanceo	97.76%	0.979 %	0.987 %

Resample	98.60%	0.978%	0.994%
Class Balancer	98.54%	0.984%	0.987%
SpreadSubSample	97.15%	0.968%	0.976%

Elaboración Propia.

Nota. Para el algoritmo RandomForest el filtro Resample mejora su desempeño de predicción.

Tabla 48.

Comparación de filtros de balanceo para mejorar el desempeño de resultados del algoritmo RandomTree.

RandomTree	Exactitud	Precisión	Recall
Sin filtros de balanceo	95.53%	0,958%	974%
Resample	98.32%	0.973%	0.994%
SpreadSubSample	93.49%	0.928%	0.943%
Class Balancer	93.40%	0.915%	0.957%

Nota. Para el algoritmo RandomTree el filtro Resample mejora su desempeño de predicción. Elaboración propia.

Tabla 49.

Comparación de filtros de balanceo para mejorar el desempeño de resultados del algoritmo

Support Vector Machine.

Support Vector Machine	Exactitud	Precisión	Recall
Sin filtros de balanceo	98.88%	0.987%	0.996%
Resample	97.76%	0.978%	0.978%
SpreadSubSample	96.74%	0.960%	0.976%
Class Balancer	98.56%	0.976%	0.996%

Elaboración propia.

Entrenamiento y prueba del método de predicción.

Para el entrenamiento del método de predicción se dividió la data en dos segmentos, el primer segmento que contó con el 80% de los datos que es representado por 286 instancias entre estudiantes desertores y no desertores y el segundo segmento de muestra que es del 20%, representado por 72 instancias de igual forma entre desertores y no desertores.

Tabla 50.

Evaluación del algoritmo J48 con el 80% de datos para entrenamiento y 20% de datos como muestra.

J48	Exactitud	Precisión	Recall
Datos entrenamiento	96.85%	0.989%	0.962%
Datos muestra	97.22%	0.961%	1.000 %

Tabla 51.

Evaluación del algoritmo Naive Bayes con el 80% de datos para entrenamiento y 20% de datos como muestra.

Naive Bayes	Exactitud	Precisión	Recall
Datos entrenamiento	98.95%	0.989%	0.995%
Datos muestra	98.61%	0.980%	1.000 %

Tabla 52.

*Evaluación de métricas para el rendimiento de clasificación del algoritmo Random Forest
Segmento de entrenamiento y prueba.*

RandomForest	Exactitud	Precisión	Recall
Datos entrenamiento	97.55%	0.974%	0.989%
Datos muestra	97.22%	0.961%	1.000%

Tabla 53.

*Evaluación de métricas para el rendimiento de clasificación del algoritmo RandomTree
Segmento de entrenamiento y prueba.*

RandomTree	Exactitud	Precisión	Recall
Datos entrenamiento	95.80%	0.948%	0.989%
Datos muestra	91.66%	0.922%	0.959%

Tabla 54.

*Evaluación de métricas para el rendimiento de clasificación del algoritmo Support Vector
Machine Segmento de entrenamiento y prueba.*

Support	Vector	Exactitud	Precisión	Recall
---------	--------	-----------	-----------	--------

Machine			
Datos entrenamiento	98.60%	0.984%	0.995%
Datos muestra	98.61%	0.980%	1.000%

Para la evaluación de la exactitud se hizo en primera instancia un entrenamiento con el 80% de los datos obtenidos, este segmento estuvo compuesto por 286 instancias, que para el presente caso de investigación cada instancia representa un estudiante. El clasificador J48 obtuvo el mejor rendimiento con una precisión del 98.9% de predicación correcta, esto quiere decir que, de las 286 instancias usadas para el entrenamiento, 282 de ellas fueron acertadas.

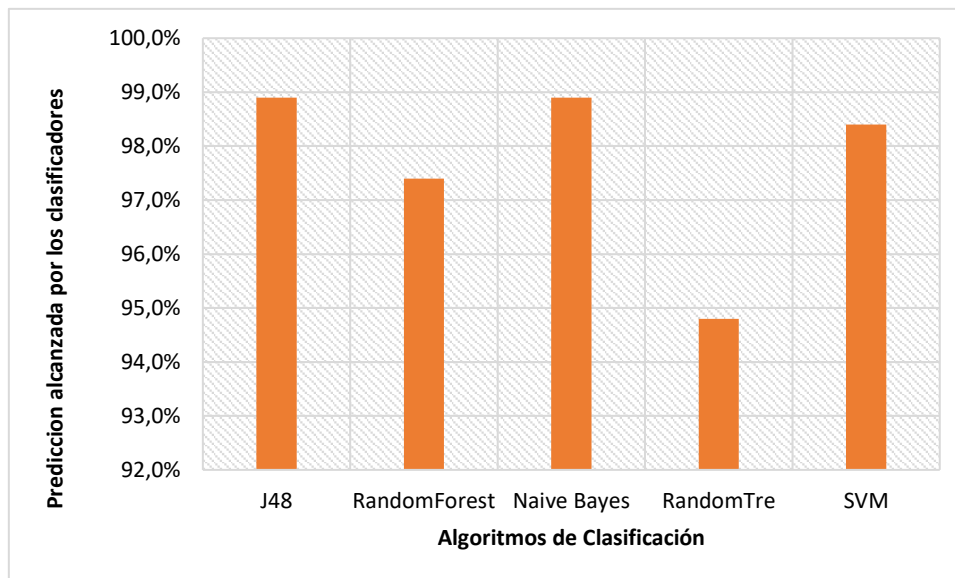


Figura 21. Evaluación de la exactitud con 80% de datos de entrenamiento

Fuente: Elaboración propia.

En segunda instancia se hizo la evaluación de muestra con el 20% de los datos obtenidos, este segmento estuvo compuesto por 72 instancias. El algoritmo Naive Bayes y Support Vector Machine obtuvieron una precisión de 98.0%, es decir que, de 72 instancias, 70 fueron clasificadas correctamente, mientras que J48 y RandomForest obtuvieron un 96.1% y RandomTree 92.2%.

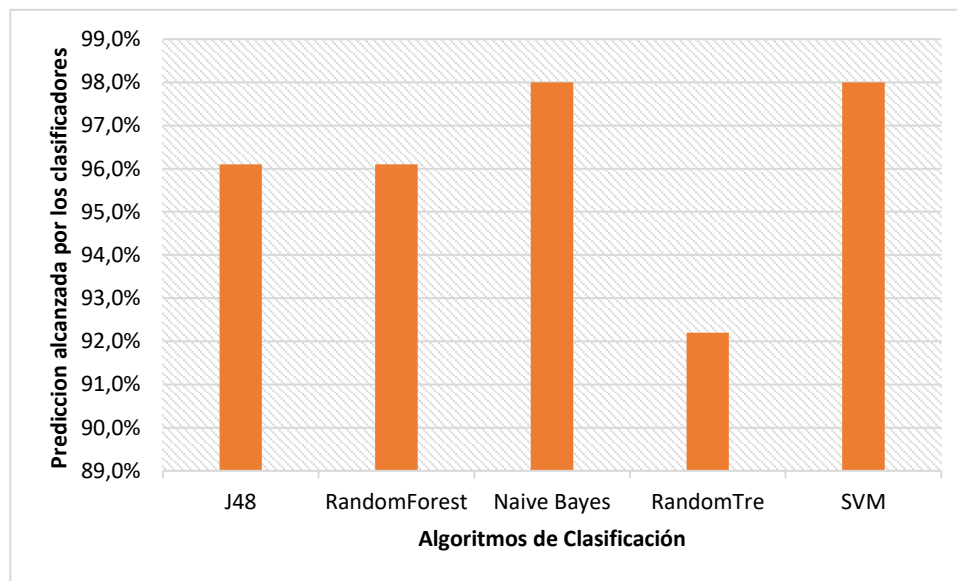


Figura 22. Evaluación de *la exactitud* con el 20% de datos de entrenamiento.

Elaboración propia.

Tabla 55:

Resultados de la matriz de confusión del método propuesto con los algoritmos de clasificación J48, Naive Bayes y Random Forest, RandomTree y Support Vector Machine con el 80% de los datos de entrenamiento y 20% de muestra.

Entrenamiento (80%)	Muestra (20%)
J48 (0.989% de precisión)	J48 (0.961% de precisión)

a	b	<input type="checkbox"/>	Classified as	a	b	<input type="checkbox"/>	Classified as
179	7	a	= no	49	0	a	= no
2	98	b	= si	2	21	b	= si
Naive Bayes (0.989% de precisión)				Naive Bayes (0.980% de precisión)			
a	b	<input type="checkbox"/>	Classified as	a	b	<input type="checkbox"/>	Classified as
185	1	a	= no	49	0	a	= no
2	98	b	= si	1	22	b	= si
RandomForest (0.974% de precisión)				RandomForest (0.961% de precisión)			
a	b	<input type="checkbox"/>	Classified as	a	b	<input type="checkbox"/>	Classified as
184	2	a	= no	49	0	a	= no
5	95	b	= si	2	21	b	= si
RandomTree (0.984% de precisión)				RandomTree (0.922% de precisión)			
a	b	<input type="checkbox"/>	Classified as	a	b	<input type="checkbox"/>	Classified as

184	2	a	= no	47	2	a	= no
5	95	b	= si	4	19	b	= si
Support Vector Machine (0.984% de precisión)				Support Vector Machine (0.980% de precisión)			
a	b	<input type="checkbox"/>	Classified as	a	b	<input type="checkbox"/>	Classified as
185	1	a	= no	49	0	a	= no
3	97	b	= si	1	22	b	= si

Elaboración propia.

IV. CONCLUSIONES Y RECOMENDACIONES

IV.1. Conclusiones.

Para el presente trabajo de investigación, en la fase de levantamiento de información, se creó una base de datos de los estudiantes universitarios de la carrera de ingeniería de industrias alimentarias de la Universidad de Jaén, se conformó un Dataset de 22 atributos y 358 instancias, cada instancia está representada por un estudiante, los cuales fueron almacenados en una hoja de cálculo con extensión CSV. Para la selección de la Universidad Peruana, se consideró como punto central que fuese licenciada por La Superintendencia Nacional de Educación (SUNEDU), con el fin de que garantice una educación de calidad con estándares internacionales para con sus estudiantes.

Para la selección de los algoritmos de clasificación de Machine Learning, se tuvo en cuenta la búsqueda de investigaciones similares que hallan utilizados estos tipos de algoritmos y que sean a partir del año 2015 en adelante, las búsquedas fueron en repositorios de prestigio como: ScienceDirect, Iopciencie, Scielo y Scopus. Los algoritmos seleccionados fueron J48, Random Forest, RandomTree, Naive Bayes y Support Vector Machine.

Para el método propuesto, se procedió a realizar el análisis sobre la data obtenida por la universidad, los archivos entregados en formato Excel y la encuesta realizada ayudaron en el procesamiento de un dataset con 358 instancias y 22 atributos. Con el uso del programa de Weka se procesaron los datos, haciendo uso de los algoritmos de clasificación tales como J48, Random Tree, Random Forest, Naive Bayes y Maquinas de Vectores de Soporte, Los algoritmos fueron tratados con filtros supervisados y no supervisados. Se obtuvieron resultados realizando una división del dataset, el primer segmento constaba del 80% de instancias que se utilizó para entrenamiento y el segundo segmento del 20% de instancias que sería para la prueba de los resultados.

Se pusieron a prueba los 05 algoritmos tales como J48, Random Tree, Random Forest, Naive Bayes y Maquinas de Vectores de Soporte, se utilizó filtros supervisados y no supervisados a nivel de instancias y atributos. Los filtros tienden a dar ventajas o

desventajas en el proceso de entrenamiento de cada algoritmo, para este caso se utilizó el filtro `ReplaceMissingValues` para completar datos faltantes, `SpreadSubSample` y `Class Balancer` para equilibrar clases y `resample` para el entrenamiento y prueba del algoritmo. Luego del proceso de limpieza de datos se hizo una segmentación de dos grupos, para el entrenamiento el primer segmento fue del 80% y para el segundo segmento de prueba el 20% de los datos, logrando tener resultados favorables con los 05 algoritmos clasificadores que se pusieron a prueba.

IV.2. Recomendaciones.

Los datos faltantes o valores nulos no deben ser eliminados ya que esto conlleva a eliminar todo un registro o instancia, esto puede provocar la eliminación de otros datos importantes que pueden ayudar a mejorar el desempeño de predicción. Se recomienda utilizar el filtro `ReplaceMissingValues`, este filtro permite generar valores en las celdas vacías a nivel de atributos con el uso de la media aritmética.

La predicción de un evento con el algoritmo de `Support Vector Machine` se mejora teniendo en cuenta utilizar datos numéricos y realizar una normalización en los datos que no tienen una distribución gaussiana, esto permitirá mejorar la calidad del desempeño del clasificador.

La selección de técnicas de predicción debe ser obtenidas de artículos científicos, esto ayuda a analizar, evaluar, comparar y mejorar los resultados que se tienen como modelo. Para la selección de los artículos científicos, la búsqueda se debe realizar en `IEEEXPLORE`, `CIENCIEDIRECT` y `SCOPUS`.

Los algoritmos de `Machine Learning` tienen un mejor desempeño en el conjunto de datos considerando los factores económicos, sociales, académicos y personales de los estudiantes, con esto se logra aumentar correctamente el porcentaje de predicción en la deserción de los estudiantes.

El fenómeno de deserción estudiantil es un tema difícil de tratar en la actualidad y requiere mantener un seguimiento o comunicación continua con los estudiantes desertores, para determinar nuevas variables y enriquecer los resultados de los algoritmos de clasificación.

REFERENCIAS.

- [1] R. J. Martelo, ; Acevedo, ; Martelo, and P. M, “Análisis Multivariado aplicado a determinar factores clave de la deserción universitaria Multivariate analysis applied to determine key factors of university dropout,” vol. 39, no. 10. 2018.
- [2] S. C. Cáceres, P. Alvarez, M. L. Ortiz, and L. C. Collado, “CI Deserción universitaria: La epidemia que aqueja a los sistemas de educación superior,” *REVISTA PERSPECTIVA*, vol. 20, no. 1, pp. 13–25, Oct. 2019, doi: 10.33198/rp.v20i1.00017.
- [3] D. G. Díaz, J. I. P. Alcaraz, M. D. G. Martínez, C. H. H. Jacome, and E. O. Ruiz, “Nodes: Plataforma para la predicción de deserción escolar utilizando técnicas de inteligencia artificial.”
- [4] L. E. G. Fiegehen, “PAGINA 8 Deserción en educación superior en América Latina y el Caribe,” pp. 1–19, 2008.
- [5] I. Q. VELASCO, “ANÁLISIS DE LAS CAUSAS DE DESERCIÓN UNIVERSITARIA.” pp. 1–48, 2016.
- [6] E. C. Medina, C. B. Chunga, J. A. Aguirre, and E. E. Grandón, 2020 15th Iberian Conference on Information Systems and Technologies (CISTI) : proceedings of CISTI'2020 - 15th Iberian Conference on Information Systems and Technologies : 24 to 27 of June 2020, Seville, Spain. 2020.
- [7] S. V. Orea, A. S. Vargas, and M. G. Alonso, “Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos.”
- [8] I. De Investigacion, P. Elizabeth, O. Farro, I. M. Fernando, and R. Moscol, “FACULTAD DE INGENIERÍA, ARQUITECTURA Y URBANISMO-ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS AUTOR(A): ASESOR.”
- [9] V. M. Coronel Flores, S. Gil Huilca, A. León Figueroa, N. León Román, and J. A. Vilchez Atalaya, “UNIVERSIDAD INCA GARCILASO DE LA VEGA,” 2019.

- [10] P. F. Alania Ricaldi, "UNIVERSIDAD NACIONAL 'DANIEL ALCIDES CARRIÓN' 'APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA PREDECIR LA DESERCIÓN ESTUDIANTIL DE LA FACULTAD DE INGENIERÍA DE LA UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN,'" 2018.
- [11] S. Valero Orea, A. Salvador Vargas, and M. García Alonso, "Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos."
- [12] T. P. S. Ricardo, J. A. J. Toledo, and A. C. Romero, *Detección de Patrones de Deserción Estudiantil con Minería de Datos*. 2017.
- [13] P. E. Ramírez and E. E. Grandón, "Prediction of student dropout in a Chilean public university through classification based on decision trees with optimized parameters," *Formacion Universitaria*, vol. 11, no. 3, pp. 3–10, Oct. 2018, doi: 10.4067/S0718-50062018000300003.
- [14] P. E. Ramírez and E. E. Grandón, "Prediction of student dropout in a Chilean public university through classification based on decision trees with optimized parameters," *Formacion Universitaria*, vol. 11, no. 3, pp. 3–10, Oct. 2018, doi: 10.4067/S0718-50062018000300003.
- [15] L. Jazmín et al., "Análisis de métodos de clasificación para el diagnóstico de fertilidad," no. 114, 2015.
- [16] L. E. Contreras, H. J. Fuentes, and J. I. Rodríguez, "Academic performance prediction by machine learning as a success/failure indicator for engineering students," *Formacion Universitaria*, vol. 13, no. 5, pp. 233–246, 2020, doi: 10.4067/S0718-50062020000500233.
- [17] L. Q. Huatangari, D. M. Jara, N. Alvarado, M. E. Milla, and O. A. Gamarra, "Lenin Quiñones Huatangari et al.: Modelo para la estimación de la deserción estudiantil Awajún y Wampis empleando minería de datos Awajún and Wampis Student Dropout Estimation Model Using Data Mining."
- [18] L. Daniela Forero Zea, Y. Fernanda Piñeros Reina, and J. Ignacio Rodríguez Molano, "Machine Learning for the identification of students at risk of academic desertion."

- [19] M. A. Miranda and J. Guzmán, “Análisis de la deserción de estudiantes universitarios usando técnicas de minería de datos,” *Formacion Universitaria*, vol. 10, no. 3, pp. 61–68, 2017, doi: 10.4067/S0718-50062017000300007.
- [20] D. González Díaz, J. I. Picie Alcaraz, M. D. González Martínez, C. H. Hernández Jacome, and E. Onofre Ruiz, “Nodes: Plataforma para la predicción de deserción escolar utilizando técnicas de inteligencia artificial.”
- [21] K. B. Eckert and R. Suénaga, “Análisis de deserción-permanencia de estudiantes universitarios utilizando técnica de clasificación en minería de datos,” *Formacion Universitaria*, vol. 8, no. 5, pp. 3–12, 2015, doi: 10.4067/S0718-50062015000500002.
- [22] J. A. Martínez Navarro and I. Despujol Zabala, “Machine Learning para la mejora de la experiencia con MOOC: el caso de la Universitat Politècnica de València,” *Revista Interuniversitaria de Investigación en Tecnología Educativa*, pp. 91–104, Jun. 2021, doi: 10.6018/riite.466251.
- [23] M. Solís, T. Moreira, R. González, T. Fernández, and M. Hernández, “Perspectives to Predict Dropout in University Students with Machine Learning.”
- [24] M. A. Santana, E. B. Costa, B. F. S. Neto, I. C. L. Silva, and J. B. A. Rego, “A predictive model for identifying students with dropout profiles in online courses.”
- [25] T. A. Cardona and E. A. Cudney, “Predicting student retention using support vector machines,” in *Procedia Manufacturing*, Elsevier B.V., 2019, pp. 1827–1833. doi: 10.1016/j.promfg.2020.01.256.
- [26] A. Vilorio, O. B. P. Lezama, and N. Varela, “Bayesian classifier applied to higher education dropout,” in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 573–577. doi: 10.1016/j.procs.2019.11.045.
- [27] C. Márquez-Vera, C. R. Morales, and S. V. Soto, “Predicting school failure and dropout by using data mining techniques,” *Revista Iberoamericana de Tecnologías del Aprendizaje*, vol. 8, no. 1, pp. 7–14, 2013, doi: 10.1109/RITA.2013.2244695.
- [28] K. Y. D. Pedroza, B. Y. C. Chasoy, and A. A. R. Gómez, “Review of techniques, tools, algorithms and attributes for data mining used in student desertion,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Oct. 2019. doi: 10.1088/1742-6596/1409/1/012003.

- [29] BORISPEREZ, C. CAMILO, and C. DARIOS, 2018 IEEE 1st Colombian Conference on Applications in Computational Intelligence (ColCACI). IEEE, 2018.
- [30] C. Burgos, M. L. Campanario, D. de la Peña, J. A. Lara, D. Lizcano, and M. A. Martínez, "Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout," *Computers and Electrical Engineering*, vol. 66, pp. 541–556, Oct. 2018, doi: 10.1016/j.compeleceng.2017.03.005.
- [31] R. J. Martelo, ; Acevedo, ; Martelo, and P. M, "Análisis Multivariado aplicado a determinar factores clave de la deserción universitaria Multivariate analysis applied to determine key factors of university dropout," vol. 39, no. 10. 2018.
- [32] M. D. R. Fernández-Hileman, Á. Corengia, and J. Durand, "Deserción y retención universitaria: una discusión bibliográfica," *Pensando Psicología*, vol. 10, no. 17, pp. 85–96, Oct. 2014, doi: 10.16925/pe.v10i17.787.
- [33] P. P. Cruz, "Inteligencia Artificial. Con Aplicaciones a la Ingeniería." [Online]. Available: www.FreeLibros.me
- [34] H. Banda, "Inteligencia Artificial: Principios y Aplicaciones." pp. 1–248, 214AD. [Online]. Available: <https://www.researchgate.net/publication/262487459>
- [35] C. D. Valero Quispe, "Derecho e Inteligencia Artificial en el mundo de hoy: escenarios internacionales y los desafíos que representan para el Perú," *THEMIS Revista de Derecho*, no. 79, pp. 311–322, Nov. 2021, doi: 10.18800/themis.202101.017.
- [36] MAFRE, "LINEA DE TIEMPO EN LA IA."
- [37] MARICIELO MALDONADO, "ARQUITECTURA DE UN SISTEMA EXPERTO."
- [38] P. Carracedo and M. Terrádez, "Minería de datos. Introducción y guía de estudio."
- [39] DIEGO CALVO, "APRENDIZAJE SUPERVISADO Y NO SUPERVISADO," 2017.
- [40] P. Ponce Cruz, "Inteligencia Artificial. Con Aplicaciones a la Ingeniería." [Online]. Available: www.FreeLibros.me
- [41] C. L. Mumford, *Computational intelligence: collaboration, fusion and emergence*. Springer, 2009.
- [42] C. C. Aggarwal, *Data Mining*. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-14142-8.

- [43] D. L. Olson and Delen Dursun, *Advanced Data Mining Techniques*. 2008.
- [44] R. Guallart and P. Maria, “Minería de Datos aplicada al análisis del tratamiento informativo de la drogadicción,” 2010.
- [45] J. Hurwitz and D. Kirsch, *Machine Learning IBM Limited Edition*. 2018. [Online]. Available: <http://www.wiley.com/go/permissions>.
- [46] J. P. Mueller and L. Massaron, *Machine Learning For Dummies*. 2021.
- [47] Y. Xie, “Authoring Books and Technical Documents with R Markdown,” 2017.
- [48] Y. Xie, *bookdown*. Chapman and Hall/CRC, 2016. doi: 10.1201/9781315204963.
- [49] P. López-Roldán and S. Fachelli, “METODOLOGÍA DE LA INVESTIGACIÓN SOCIAL CUANTITATIVA.”
- [50] P. Lopez Roldán and S. Fachelli, *METODOLOGÍA DE LA INVESTIGACIÓN SOCIAL CUANTITATIVA, PRIMERA.*, vol. 1. 2015.
- [51] E. M. Mejía, “PAGINA 51 TÉCNICAS E INSTRUMENTOS DE INVESTIGACIÓN,” pp. 1–239, 2005.

- ANEXOS.

Anexo 01.



FACULTAD DE INGENIERÍA, ARQUITECTURA Y URBANISMO
RESOLUCIÓN N°0445-2021/FIAU-USS

Pimentel, 27 de mayo de 2021

VISTO:

El Acta de reunión N°1305-2021 del Comité de investigación de la Escuela profesional de INGENIERÍA DE SISTEMAS remitida mediante oficio N°0227-2021/FIAU-IS-USS de fecha 19 de mayo de 2021, y;

CONSIDERANDO:

Que, de conformidad con la Ley Universitaria N° 30220 en su artículo 48° que a letra dice: "La investigación constituye una función esencial y obligatoria de la universidad, que la fomenta y realiza, respondiendo a través de la producción de conocimiento y desarrollo de tecnologías a las necesidades de la sociedad, con especial énfasis en la realidad nacional. Los docentes, estudiantes y graduados participan en la actividad investigadora en su propia institución o en redes de investigación nacional o internacional, creadas por las instituciones universitarias públicas o privadas.";

Que, de conformidad con el Reglamento de grados y títulos en su artículo 21° señala: "Los temas de trabajo de investigación, trabajo académico y tesis son aprobados por el Comité de Investigación y derivados a la Facultad o Escuela de Posgrado, según corresponda, para la emisión de la resolución respectiva. El periodo de vigencia de los mismos será de dos años, a partir de su aprobación. En caso un tema perdiera vigencia, el Comité de Investigación evaluará la ampliación de la misma.

Que, de conformidad con el Reglamento de grados y títulos en su artículo 24° señala: La tesis es un estudio que debe denotar rigurosidad metodológica, originalidad, relevancia social, utilidad teórica y/o práctica en el ámbito de la escuela profesional. Para el grado de doctor se requiere una tesis de máxima rigurosidad académica y de carácter original. Es individual para la obtención de un grado; es individual o en pares para obtener un título profesional. Asimismo, en su artículo 25° señala: "El tema debe responder a alguna de las líneas de investigación institucionales de la USS S.A.C.".

Que, según documentos de Vistos el Comité de investigación de la Escuela profesional de INGENIERÍA DE SISTEMAS acuerdan aprobar los temas de las Tesis a cargo de los estudiantes del curso de Investigación I que se detallan en el anexo de la presente Resolución.

Estando a lo expuesto, y en uso de las atribuciones conferidas y de conformidad con las normas y reglamentos vigentes;

SE RESUELVE:

ARTÍCULO 1°: APROBAR, el tema de la Tesis perteneciente a la línea de investigación de INFRAESTRUCTURA, TECNOLOGÍA Y MEDIO AMBIENTE, a cargo de los estudiantes del Programa de estudios de INGENIERÍA DE SISTEMAS según se detalla en el anexo de la presente Resolución.

ARTÍCULO 2°: ESTABLECER, que la inscripción del Tema de la Tesis se realice a partir de emitida la presente resolución y tendrá una vigencia de dos (02) años.

ARTÍCULO 3°: DEJAR SIN EFECTO, toda Resolución emitida por la Facultad que se oponga a la presente Resolución.

REGÍSTRESE, COMUNÍQUESE Y ARCHÍVESE



 Dr. Mario Fernando Ramos Moscol
Decano - Facultad de Ingeniería,
Arquitectura y Urbanismo
UNIVERSIDAD SEÑOR DE SIPÁN S.A.C.



 MBA. María Noelia Siles Rivera
Secretaría Académica / Facultad de Ingeniería,
Arquitectura y Urbanismo
UNIVERSIDAD SEÑOR DE SIPÁN S.A.C.

Cc: Interesado, Archivo

FACULTAD DE INGENIERÍA, ARQUITECTURA Y URBANISMO
RESOLUCIÓN N°0445-2021/FIAU-USS

Pimentel, 27 de mayo de 2021

N°	AUTOR (ES)	TEMA DE TESIS
21	PISFIL CORONADO JOSE LUIS FELIPE	IMPLEMENTACIÓN DE ARQUITECTURA EMPRESARIAL BASADA EN METODOLOGÍA ÁGIL PARA ALINEAR TI CON LOS PROCESOS DE NEGOCIO EN UNA EMPRESA CONSTRUCTORA PERUANA DE OBRAS CIVILES
22	ABAD HERRERA JOHNNY RENSO TEPE ESPINOZA LUIS RAMON	IMPLEMENTACIÓN DE ITIL V4 PARA MEJORAR LOS SERVICIOS DE TI EN EL CENTRO DE SISTEMAS DE INFORMACIÓN DE UNA UNIDAD DE GESTIÓN EDUCATIVA LOCAL PERUANO
23	URRUTIA VASQUEZ MIGUEL JULCA ROJAS ALEX ROGELIO	DESARROLLO DE UN MÉTODO DE IDENTIFICACIÓN AUTOMÁTICA DE ATAQUES SPOOFING DE ENVENENAMIENTO ARP EN LA SUPLANTACIÓN DE IDENTIDAD EN REDES LAN
24	SANCHEZ CELADA ERLIN FERNANDEZ ROMAN ISMAEL	COMPARACIÓN DE ARQUITECTURAS DE IDS HÍBRIDO PARA LA IDENTIFICACIÓN DE ATAQUES DE DOS EN LOS SERVIDORES WEB DE UNA MUNICIPALIDAD PROVINCIAL PERUANA
25	PERALES CHAVEZ JEFFERSON ADRIAN	IMPLEMENTACIÓN DE UN MODELO DE ARQUITECTURA DE INDUSTRIA 4.0 PARA MEJORAR LA INTEROPERABILIDAD ENTRE SISTEMAS DE UNA EMPRESA PERUANA
26	MAGALLANES CARBAJAL KENSER	EVALUACIÓN DE LA EFICIENCIA DE LOS ALGORITMOS DE CRIPTOGRAFÍA PARA CUMPLIR CON LOS NIVELES DE SEGURIDAD DE DATOS DE UNA EMPRESA FINANCIERA PERUANA
27	RACCHUMI LECCA JESÚS MANUEL	DESARROLLO DE UN MIDDLEWARE PARA MEJORAR LA COMUNICACIÓN ENTRE DOS INTERFACES DE LMS Y CRM EN EL PROCESO DE REGISTRO Y EMISIÓN DE CREDENCIALES DE USUARIOS
28	CASTRO QUESQUEN JAIME ELTON	COMPARACIÓN DE ALGORITMOS DE CIFRADO DE DATOS EN EL ASEGURAMIENTO DE VIDEO LLAMADA SOBRE REDES IP
29	PEREZ DIAZ NEILER WILTER CHINCHAY MALDONADO JORGE OBED	IMPLEMENTACIÓN DE TECNOLOGÍA SANDBOX PARA PROTEGER DE ATAQUES RANSOMWARE EN UNA RED INFORMÁTICA LOCAL DE UNA ENTIDAD FINANCIERA
30	MOSCOSO PAREDES ANIBAL	DISEÑO DE UN MODELO DE ARQUITECTURA DE SEGURIDAD DE BAJO COSTO PARA REFORZAR LA SEGURIDAD DE LA RED DEL HOGAR ANTE ATAQUES INFORMÁTICOS
31	MARTINEZ CUMPA JORGE JOSE	EVALUACIÓN DE FACTIBILIDAD DE USO DE TECNOLOGÍA WIRELESS 5GHZ PARA PROPORCIONAR SERVICIOS DE COMUNICACIÓN INALÁMBRICA EN LOS CENTROS POBLADOS RURALES DE LA REGIÓN LAMBAYEQUE
32	CAMPOS BARRERA SANDRO PAUL PASTOR OLIVA CESAR AUGUSTO	IMPLEMENTACIÓN DE UN MÉTODO DE CLASIFICACIÓN PARA DETECTAR LA DESERCIÓN DE ESTUDIANTES DE LA CARRERA DE INGENIERÍA DE INDUSTRIAS ALIMENTARIAS DE UNA UNIVERSIDAD NACIONAL PERUANA BASADO EN APRENDIZAJE DE MAQUINA
33	PICON VASQUEZ ANGEL GABRIEL CESPEDES SALAZAR JUAN CARLOS	DESARROLLO DE UN MÉTODO DE CLASIFICACIÓN AUTOMÁTICA BASADA EN TÉCNICAS ESTADÍSTICAS Y DE MACHINE LEARNING PARA CLASIFICAR A LOS POSTULANTES DE ACUERDO AL PERFIL DE TRABAJO DE UN CALL CENTER
34	MIÑANO SANCHEZ CARLOS JOHNY	COMPARACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA DESCUBRIR INFORMACIÓN RELEVANTE DE VENTAS DE UNA MYPE COMERCIAL
35	MARTOS PAREDES JOEL HAROLD VILLAZON SOSA JAIR AUGUSTO	IMPLEMENTACIÓN DE UN MODELO DE PROCESOS DE SEGURIDAD DE LA INFORMACIÓN PARA UNA PYME PERUANA BASADO EN LA NORMA ISO/IEC 27005 Y LA METODOLOGÍA OCTAVE-S
36	QUISPE PUEMAPE LUIS ALONSO	IMPLEMENTACIÓN DE UN SISTEMA DE GESTIÓN DE SEGURIDAD DE LA INFORMACIÓN APLICANDO LA NORMA ISO/IEC 27001:2014 EN UNA EMPRESA PERUANA DE TELECOMUNICACIONES
37	CHUCO AGUILAR GERSON RAUL	IMPLEMENTACIÓN DE UN SISTEMA DE GESTIÓN DE SEGURIDAD DE LA INFORMACIÓN BASADA EN ISO/IEC 27001 PARA MEJORAR EL NIVEL DE SEGURIDAD DE LOS ACTIVOS DE INFORMACIÓN EN UNA EMPRESA CONSTRUCTORA DE OBRAS CIVILES
38	CAJUSOL ROJAS JOSE DEL CARMEN	IMPLEMENTACIÓN DE UNA PLATAFORMA WEB PARA LA PLANIFICACIÓN Y MONITOREO DE RUTAS DE RECOJO DE RESIDUOS SÓLIDOS DE UN MUNICIPIO DE LA REGIÓN LAMBAYEQUE
39	VALLEJOS RAMOS FERNANDO RAFAEL	DESARROLLO DE UN MÉTODO DE OPTIMIZACIÓN DE USO DE TELA EN EL PROCESO DE ELABORACIÓN DE PRENDAS TEXTILES DE MICROEMPRESAS PERUANAS
40	REQUEJO NAVARRO JERSONS EXFRANSHER	EVALUACIÓN DE ALGORITMOS CRIPTOGRÁFICOS PARA MEJORAR SEGURIDAD EN UNA RED PRIVADA VIRTUAL

FACULTAD DE INGENIERÍA, ARQUITECTURA Y URBANISMO
RESOLUCIÓN N°0445-2021/FIAU-USS

Pimentel, 27 de mayo de 2021

N°	AUTOR (ES)	TEMA DE TESIS
21	PISFIL CORONADO JOSE LUIS FELIPE	IMPLEMENTACIÓN DE ARQUITECTURA EMPRESARIAL BASADA EN METODOLOGÍA ÁGIL PARA ALINEAR TI CON LOS PROCESOS DE NEGOCIO EN UNA EMPRESA CONSTRUCTORA PERUANA DE OBRAS CIVILES
22	ABAD HERRERA JOHNNY RENSO TEPE ESPINOZA LUIS RAMON	IMPLEMENTACIÓN DE ITIL V4 PARA MEJORAR LOS SERVICIOS DE TI EN EL CENTRO DE SISTEMAS DE INFORMACIÓN DE UNA UNIDAD DE GESTIÓN EDUCATIVA LOCAL PERUANO
23	URRUTIA VASQUEZ MIGUEL JULCA ROJAS ALEX ROGELIO	DESARROLLO DE UN MÉTODO DE IDENTIFICACIÓN AUTOMÁTICA DE ATAQUES SPOOFING DE ENVENENAMIENTO ARP EN LA SUPLANTACIÓN DE IDENTIDAD EN REDES LAN
24	SANCHEZ CELADA ERLIN FERNANDEZ ROMAN ISMAEL	COMPARACIÓN DE ARQUITECTURAS DE IDS HÍBRIDO PARA LA IDENTIFICACIÓN DE ATAQUES DE DOS EN LOS SERVIDORES WEB DE UNA MUNICIPALIDAD PROVINCIAL PERUANA
25	PERALES CHAVEZ JEFFERSON ADRIAN	IMPLEMENTACIÓN DE UN MODELO DE ARQUITECTURA DE INDUSTRIA 4.0 PARA MEJORAR LA INTEROPERABILIDAD ENTRE SISTEMAS DE UNA EMPRESA PERUANA
26	MAGALLANES CARBAJAL KENSER	EVALUACIÓN DE LA EFICIENCIA DE LOS ALGORITMOS DE CRIPTOGRAFÍA PARA CUMPLIR CON LOS NIVELES DE SEGURIDAD DE DATOS DE UNA EMPRESA FINANCIERA PERUANA
27	RACCHUMI LECCA JESÚS MANUEL	DESARROLLO DE UN MIDDLEWARE PARA MEJORAR LA COMUNICACIÓN ENTRE DOS INTERFACES DE LMS Y CRM EN EL PROCESO DE REGISTRO Y EMISIÓN DE CREDENCIALES DE USUARIOS
28	CASTRO QUESQUEN JAIME ELTON	COMPARACIÓN DE ALGORITMOS DE CIFRADO DE DATOS EN EL ASEGURAMIENTO DE VIDEO LLAMADA SOBRE REDES IP
29	PEREZ DIAZ NEILER WILTER CHINCHAY MALDONADO JORGE OBED	IMPLEMENTACIÓN DE TECNOLOGÍA SANDBOX PARA PROTEGER DE ATAQUES RANSOMWARE EN UNA RED INFORMÁTICA LOCAL DE UNA ENTIDAD FINANCIERA
30	MOSCOSO PAREDES ANIBAL	DISEÑO DE UN MODELO DE ARQUITECTURA DE SEGURIDAD DE BAJO COSTO PARA REFORZAR LA SEGURIDAD DE LA RED DEL HOGAR ANTE ATAQUES INFORMÁTICOS
31	MARTINEZ CUMPA JORGE JOSE	EVALUACIÓN DE FACTIBILIDAD DE USO DE TECNOLOGÍA WIRELESS 5GHZ PARA PROPORCIONAR SERVICIOS DE COMUNICACIÓN INALÁMBRICA EN LOS CENTROS POBLADOS RURALES DE LA REGIÓN LAMBAYEQUE
32	CAMPOS BARRERA SANDRO PAUL PASTOR OLIVA CESAR AUGUSTO	IMPLEMENTACIÓN DE UN MÉTODO DE CLASIFICACIÓN PARA DETECTAR LA DESERCIÓN DE ESTUDIANTES DE LA CARRERA DE INGENIERÍA DE INDUSTRIAS ALIMENTARIAS DE UNA UNIVERSIDAD NACIONAL PERUANA BASADO EN APRENDIZAJE DE MAQUINA
33	PICON VASQUEZ ANGEL GABRIEL CESPEDES SALAZAR JUAN CARLOS	DESARROLLO DE UN MÉTODO DE CLASIFICACIÓN AUTOMÁTICA BASADA EN TÉCNICAS ESTADÍSTICAS Y DE MACHINE LEARNING PARA CLASIFICAR A LOS POSTULANTES DE ACUERDO AL PERFIL DE TRABAJO DE UN CALL CENTER
34	MIÑANO SANCHEZ CARLOS JOHNY	COMPARACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA DESCUBRIR INFORMACIÓN RELEVANTE DE VENTAS DE UNA MYPE COMERCIAL
35	MARTOS PAREDES JOEL HAROLD VILLAZON SOSA JAIR AUGUSTO	IMPLEMENTACIÓN DE UN MODELO DE PROCESOS DE SEGURIDAD DE LA INFORMACIÓN PARA UNA PYME PERUANA BASADO EN LA NORMA ISO/IEC 27005 Y LA METODOLOGÍA OCTAVE-S
36	QUISPE PUEMAPE LUIS ALONSO	IMPLEMENTACIÓN DE UN SISTEMA DE GESTIÓN DE SEGURIDAD DE LA INFORMACIÓN APLICANDO LA NORMA ISO/IEC 27001:2014 EN UNA EMPRESA PERUANA DE TELECOMUNICACIONES
37	CHUCO AGUILAR GERSON RAUL	IMPLEMENTACIÓN DE UN SISTEMA DE GESTIÓN DE SEGURIDAD DE LA INFORMACIÓN BASADA EN ISO/IEC 27001 PARA MEJORAR EL NIVEL DE SEGURIDAD DE LOS ACTIVOS DE INFORMACIÓN EN UNA EMPRESA CONSTRUCTORA DE OBRAS CIVILES
38	CAJUSOL ROJAS JOSE DEL CARMEN	IMPLEMENTACIÓN DE UNA PLATAFORMA WEB PARA LA PLANIFICACIÓN Y MONITOREO DE RUTAS DE RECOJO DE RESIDUOS SÓLIDOS DE UN MUNICIPIO DE LA REGIÓN LAMBAYEQUE
39	VALLEJOS RAMOS FERNANDO RAFAEL	DESARROLLO DE UN MÉTODO DE OPTIMIZACIÓN DE USO DE TELA EN EL PROCESO DE ELABORACIÓN DE PRENDAS TEXTILES DE MICROEMPRESAS PERUANAS
40	REQUEJO NAVARRO JERSONS EXFRANSHER	EVALUACIÓN DE ALGORITMOS CRIPTOGRÁFICOS PARA MEJORAR SEGURIDAD EN UNA RED PRIVADA VIRTUAL

Anexo 02.



UNIVERSIDAD NACIONAL DE JAÉN
Resolución del Consejo Directivo N° 002-2018-Sunedu/Cd
VICEPRESIDENCIA DE INVESTIGACIÓN
"Año del Bicentenario del Perú: 200 años de Independencia"



Jaén, 15 de julio del 2021.

OFICIO N° 522-2021-OVPI-CO-UNJ.

MAD N° 309568

SEÑOR
MG. ING. VICTOR ALEXCI TUESTA MONTEZA.
DIRECTOR (E) DE LA ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS
UNIVERSIDAD SEÑOR DE SIPAN
PRESENTE.-

ASUNTO: AUTORIZACIÓN PARA EJECUCIÓN DE PROYECTO DE INVESTIGACIÓN.
REFERENCIA: CARTA S/N° 28/05/2021.

Es grato dirigirme a usted para saludarlo cordialmente y en respuesta a su petición mediante el documento de la referencia, manifestarle que esta Vicepresidencia de Investigación autoriza para que los señores **SANDRO PAUL CAMPOS BARRERA y CÉSAR AUGUSTO PASTOR OLIVA** procedan a desarrollar la aplicación de su Proyecto de Investigación titulado "Implementación de un método de clasificación para detectar la deserción de estudiantes de la Carrera de Ingeniería de Industrias Alimentarias de una Universidad Nacional Peruana basado en aprendizaje de maquina", aprobado con Resolución N° 0445-2021/FIAU-USS, de fecha 27 de mayo del 2021.

Así mismo para la aplicación del referido proyecto deberá comunicarse adecuadamente con el Coordinador de la Carrera Profesional de Ingeniería de Industrias Alimentarias **Mg. Frank Fernández Rosillo**, Cel. N° 967108407, para los fines que se estime por conveniente.

Con las muestras de mi especial deferencia, me reitero de usted.

Muy Atentamente,

UNIVERSIDAD NACIONAL DE JAÉN
COMISIÓN ORGANIZADORA



Dr. Víctor Benjamín Carril Fernández
VICEPRESIDENTE DE INVESTIGACIÓN

Cc
Archivo
VBCE/VPPI
MCAO/sec.

Anexo 3. Instrumentos de recolección de datos, con su respectiva validación de los instrumentos.

Consumo de CPU	
Ítem	Valor
Uso	
Velocidad	
Tiempo	

. Formato de informe de consumo de memoria.

Consumo de Memoria	
Ítem	Valor
Uso	
Disponibilidad	
En caché	
Tiempo	

Formato para informe de promedio de tiempo de respuesta.

Promedio de Tiempo de respuesta	
Ítem	Valor
Velocidad	
Tiempo	
CPU	

Memoria	
---------	--

Ítem	Valor
Verdadero positivo (VP)	
Falso Positivo (FP)	
Verdadero Negativo (VN)	
Falso Negativo (FN)	

Ítem	Valor
Tiempo de respuesta	
Grado Consumo CPU	
Grado Consumo de memoria	
Precisión	

Verdadero Positivo (VP)		Falso Positivo (FP)	
Ítem	Valor	Ítem	Valor
Realidad		Realidad	
Predicción del método		Predicción del método	
Numero de resultados		Numero de resultados	
Falso Negativo (FN)		Verdadero Negativo (FP)	
Ítem	Valor	Ítem	Valor
Realidad		Realidad	
Predicción del método		Predicción del método	
Numero de resultados		Numero de resultados	

Tabla 56.

Matriz de confusión de entrenamiento y prueba de los clasificadores J48, RandomForest, RandomTree, Naive Bayes y SVM

Entrenamiento (80%)				Muestra (20%)			
J48 (0.989% de precisión)				J48 (0.961% de precisión)			
a	b	<input type="checkbox"/>	Classified as	a	b	<input type="checkbox"/>	Classified as
179	7	a	= no	49	0	a	= no
2	98	b	= si	2	21	b	= si
Naive Bayes (0.989% de precisión)				Naive Bayes (0.980% de precisión)			
a	b	<input type="checkbox"/>	Classified as	a	b	<input type="checkbox"/>	Classified as
185	1	a	= no	49	0	a	= no
2	98	b	= si	1	22	b	= si
RandomForest (0.974% de precisión)				RandomForest (0.961% de precisión)			
a	b	<input type="checkbox"/>	Classified as	a	b	<input type="checkbox"/>	Classified as
184	2	a	= no	49	0	a	= no

5	95	b	= si	2	21	b	= si
RandomTree (0.984% de precisión)				RandomTree (0.922% de precisión)			
a	b	<input type="checkbox"/>	Classified as	a	b	<input type="checkbox"/>	Classified as
184	2	a	= no	47	2	a	= no
5	95	b	= si	4	19	b	= si
Support Vector Machine (0.984% de precisión)				Support Vector Machine (0.980% de precisión)			
a	b	<input type="checkbox"/>	Classified as	a	b	<input type="checkbox"/>	Classified as
185	1	a	= no	49	0	a	= no
3	97	b	= si	1	22	b	= si

Tabla 57.

Lista de universidades a nivel nacional públicas y privadas licenciadas por SUNEDU.

Nº	NOMBRE DE LA UNIVERSIDAD	TIPO DE GESTIÓN	DEPARTAMENTO/PROVINCIA
1	UNIVERSIDAD NACIONAL INTERCULTURAL FABIOLA SALAZAR LEGUÍA DE BAGUA	PUBLICA	AMAZONAS / BAGUA
2	UNIVERSIDAD NACIONAL TORIBIO RODRÍGUEZ DE MENDOZA DE AMAZONAS	PUBLICA	AMAZONAS / CHACHAPOYAS
3	UNIVERSIDAD TECNOLÓGICA DE LOS ANDES	PRIVADA	APURIMAC / ABANCAY
4	UNIVERSIDAD NACIONAL MICAELA BASTIDAS DE APURÍMAC	PUBLICA	ABANCAY / APURIMAC
5	UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS	PUBLICA	APURIMAC / ANDAHUAYLAS
6	UNIVERSIDAD NACIONAL DE SAN AGUSTÍN	PUBLICA	AREQUIPA / AREQUIPA
7	UNIVERSIDAD LA SALLE	PRIVADA	AREQUIPA / AREQUIPA
8	UNIVERSIDAD CATÓLICA DE SANTA MARÍA	PRIVADA	AREQUIPA / AREQUIPA
9	UNIVERSIDAD CATÓLICA SAN PABLO	PRIVADA	AREQUIPA / AREQUIPA

10	UNIVERSIDAD NACIONAL DE SAN CRISTÓBAL DE HUAMANGA	PUBLICA	AYACUCHO / HUAMANGA
11	UNIVERSIDAD NACIONAL AUTÓNOMA DE HUANTA	PUBLICA	AYACUCHO / HUANTA
12	UNIVERSIDAD NACIONAL DE CAJAMARCA	PUBLICA	CAJAMARCA / CAJAMARCA
13	UNIVERSIDAD NACIONAL AUTÓNOMA DE CHOTA	PUBLICA	CAJAMARCA / CHOTA
14	UNIVERSIDAD NACIONAL DE JAÉN	PUBLICA	CAJAMARCA / JAEN
15	UNIVERSIDAD NACIONAL DEL CALLAO	PUBLICA	CALLAO / PROVIN CONST CALLAO
16	UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO	PUBLICA	CUSCO / CUSCO
17	UNIVERSIDAD ANDINA DEL CUSCO	PRIVADA	CUSCO / CUSCO
18	UNIVERSIDAD NACIONAL INTERCULTURAL DE QUILLABAMBA	PUBLICA	CUSCO / LA CONVENCION
19	UNIVERSIDAD PARA EL DESARROLLO ANDINO	PRIVADA	HUANCAVELICA / ANGARAES
20	UNIVERSIDAD NACIONAL DE HUANCAVELICA	PUBLICA	HUACAVELICA / HUANCAVELICA
21	UNIVERSIDAD NACIONAL AUTÓNOMA DE TAYACAJA "DANIEL HERNÁNDEZ MORILLO"	PUBLICA	HUANCAVELICA / TAYACAJA
22	UNIVERSIDAD NACIONAL HERMILIO VALDIZÁN	PUBLICA	HUANUCO /HUANUCO
23	UNIVERSIDAD DE HUÁNUCO	PRIVADA	HUANUCO /HUANUCO
24	UNIVERSIDAD NACIONAL AGRARIA DE LA SELVA	PUBLICA	HUANUCO / LEONCIO PRADO

25	UNIVERSIDAD AUTÓNOMA DE ICA	PRIVADA	ICA
26	UNIVERSIDAD NACIONAL INTERCULTURAL DE LA SELVA CENTRAL JUAN SANTOS ATAHUALPA	PUBLICA	JUNIN / CHANCHAMAYO
27	UNIVERSIDAD PERUANA LOS ANDES	PRIVADA	JUNIN / HUANCAYO
28	UNIVERSIDAD PRIVADA DE HUANCAYO FRANKLIN ROOSEVELT	PRIVADA	JUNIN / HUANCAYO
29	UNIVERSIDAD NACIONAL DEL CENTRO DEL PERÚ	PUBLICA	JUNIN / HUANCAYO
30	UNIVERSIDAD CONTINENTAL	PRIVADA	JUNIN / HUANCAYO
31	UNIVERSIDAD NACIONAL AUTÓNOMA ALTOANDINA DE TARMA	PUBLICA	JUNIN / TARMA
32	UNIVERSIDAD CÉSAR VALLEJO	PRIVADA	LA LIBERTAD / TRUJILLO
33	UNIVERSIDAD CATÓLICA DE TRUJILLO BENEDICTO XVI	PRIVADA	LA LIBERTAD / TRUJILLO
34	UNIVERSIDAD NACIONAL DE TRUJILLO	PUBLICA	LA LIBERTAD / TRUJILLO
35	UNIVERSIDAD PRIVADA ANTENOR ORREGO	PRIVADA	LA LIBERTAD / TRUJILLO
36	UNIVERSIDAD PRIVADA DEL NORTE	PRIVADA	LA LIBERTAD / TRUJILLO
37	UNIVERSIDAD SEÑOR DE SIPÁN	PRIVADA	LAMBAYEQUE / CHICLAYO
38	UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO	PRIVADA	LAMBAYEQUE / CHICLAYO
39	UNIVERSIDAD NACIONAL DE BARRANCA	PUBLICA	LIMA / BARRANCA
40	UNIVERSIDAD NACIONAL DE CAÑETE	PUBLICA	LIMA / CAÑETE

41	UNIVERSIDAD NACIONAL JOSÉ FAUSTINO SÁNCHEZ CARRIÓN	PUBLICA	LIMA / HUAURA
42	FACULTAD DE TEOLOGÍA PONTIFICIA Y CIVIL DE LIMA	PRIVADA	LIMA
43	UNIVERSIDAD NACIONAL FEDERICO VILLARREAL	PUBLICA	LIMA / LIMA
44	UNIVERSIDAD NACIONAL DE EDUCACIÓN ENRIQUE GUZMÁN Y VALLE	PUBLICA	LIMA / LIMA
45	UNIVERSIDAD PRIVADA NORBERT WIENER	PRIVADA	LIMA / LIMA
46	UNIVERSIDAD PRIVADA SAN JUAN BAUTISTA	PRIVADA	LIMA / LIMA
47	UNIVERSIDAD PRIVADA PERUANO ALEMANA	PRIVADA	LIMA / LIMA
48	UNIVERSIDAD TECNOLÓGICA DEL PERÚ	PRIVADA	LIMA/LIMA
49	UNIVERSIDAD LE CORDON BLEU S.A.C.	PRIVADA	LIMA/LIMA
50	UNIVERSIDAD MARÍA AUXILIADORA	PRIVADA	LIMA/LIMA
51	UNIVERSIDAD CATÓLICA SEDES SAPIENTIAE	PRIVADA	LIMA/LIMA
52	UNIVERSIDAD AUTÓNOMA DEL PERÚ	PRIVADA	LIMA/LIMA
53	UNIVERSIDAD NACIONAL TECNOLÓGICA DE LIMA SUR	PÚBLICA	LIMA/LIMA
54	UNIVERSIDAD JAIME BAUSATE Y MEZA	PRIVADA	LIMA/LIMA
55	UNIVERSIDAD PERUANA UNIÓN	PRIVADA	LIMA/LIMA
56	ESCUELA DE POSGRADO-GERENS S.A.C.	PRIVADA	LIMA/LIMA

57	UNIVERSIDAD CIENTÍFICA DEL SUR	PRIVADA	LIMA/LIMA
58	UNIVERSIDAD ESAN	PRIVADA	LIMA/LIMA
59	UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS	PÚBLICA	LIMA/LIMA
60	UNIVERSIDAD NACIONAL DE INGENIERÍA	PÚBLICA	LIMA/LIMA
61	UNIVERSIDAD DE CIENCIAS Y HUMANIDADES	PRIVADA	LIMA/LIMA
62	UNIVERSIDAD MARCELINO CHAMPAGNAT	PRIVADA	LIMA/LIMA
63	UNIVERSIDAD SAN IGNACIO DE LOYOLA	PRIVADA	LIMA/LIMA
64	UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS	PRIVADA	LIMA/LIMA
65	UNIVERSIDAD DE SAN MARTÍN DE PORRES	PRIVADA	LIMA/LIMA
66	UNIVERSIDAD ANTONIO RUIZ DE MONTOYA	PRIVADA	LIMA/LIMA
67	UNIVERSIDAD NACIONAL AGRARIA LA MOLINA	PÚBLICA	LIMA/LIMA
68	UNIVERSIDAD RICARDO PALMA	PRIVADA	LIMA/LIMA
69	UNIVERSIDAD FEMENINA DEL SAGRADO CORAZÓN	PRIVADA	LIMA/LIMA
70	UNIVERSIDAD DE CIENCIAS Y ARTES DE AMÉRICA LATINA	PRIVADA	LIMA/LIMA
71	UNIVERSIDAD PERUANA CAYETANO HEREDIA	PRIVADA	LIMA/LIMA
72	UNIVERSIDAD DEL PACÍFICO	PRIVADA	LIMA/LIMA
73	UNIVERSIDAD DE LIMA	PRIVADA	LIMA/LIMA

74	PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ	PRIVADA	LIMA/LIMA
75	UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA	PRIVADA	LIMA/LIMA
76	UNIVERSIDAD NACIONAL AUTÓNOMA DE ALTO AMAZONAS	PÚBLICA	LORETO/ALTO AMAZONAS
77	UNIVERSIDAD NACIONAL DE LA AMAZONÍA PERUANA	PÚBLICA	LORETO/MAYNAS
78	UNIVERSIDAD NACIONAL AMAZÓNICA DE MADRE DE DIOS	PÚBLICA	MADRE DE DIOS/TAMBOPATA
79	UNIVERSIDAD NACIONAL DE MOQUEGUA	PÚBLICA	MOQUEGUA/MARISCAL NIETO
80	UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN	PÚBLICA	PASCO/PASCO
81	UNIVERSIDAD NACIONAL DE PIURA	PÚBLICA	PIURA/PIURA
82	UNIVERSIDAD DE PIURA	PRIVADA	PIURA/PIURA
83	UNIVERSIDAD NACIONAL DE FRONTERA	PÚBLICA	PIURA/SULLANA
84	UNIVERSIDAD NACIONAL DEL ALTIPLANO	PÚBLICA	PUNO/PUNO
85	UNIVERSIDAD NACIONAL DE JULIACA	PÚBLICA	PUNO/SAN ROMÁN
86	UNIVERSIDAD NACIONAL DE SAN MARTÍN	PÚBLICA	SAN MARTÍN/SAN MARTÍN
87	ESCUELA DE POSTGRADO NEUMANN BUSINESS SCHOOL S.A.C.	PRIVADA	TACNA/TACNA
88	UNIVERSIDAD NACIONAL JORGE BASADRE GROHMANN	PÚBLICA	TACNA/TACNA
89	UNIVERSIDAD PRIVADA DE TACNA	PRIVADA	TACNA/TACNA

90	UNIVERSIDAD NACIONAL DE TUMBES	PÚBLICA	TUMBES/TUMBES
91	UNIVERSIDAD NACIONAL DE UCAYALI	PÚBLICA	UCAYALI/CORONEL PORTILLO
92	UNIVERSIDAD NACIONAL INTERCULTURAL DE LA AMAZONIA	PÚBLICA	UCAYALI/CORONEL PORTILLO
93	UNIVERSIDAD NACIONAL SANTIAGO ANTÚNEZ DE MAYOLO	PÚBLICA	ÁNCASH/HUARAZ
94	UNIVERSIDAD NACIONAL DEL SANTA	PÚBLICA	

Tabla 58.

Dataset de los estudiantes de Ingeniería de Industrias Alimentarias de la Univeridad de Jaén - Tranformado a datos numéricos

edad	sex	est	tra	dis	apo	ori	ren	ben	ca	mo	niv	niv	infl	vici	pad	eco	pat	col	inte	hijo	aba
	o	ado	baj	cap	yo	ent	dim	efic	mbi	dali	eIE	eIS	uen	o	res	no	ern	egi	gra	s	nd
		Civ	a	aci	Pa	aci	ient	ioU	oC	dad	duc	atis	cia		Se	mia	ida	oPr	nte		on
		il		dad	dre	on	oA	niv	arr	Ing	ativ	fac	Car		par	Baj	dN	oce	sFa		o
				Fisi	s	Vo	cad	ersi	era	res	oA	cio	rer		ado	a	oPI	den	mili		
				ca		cac	emi	dad			po	n	a		s		ane	cia	a		
						ion	co				der						ada				
						al					ado										
2	1	1	1	1	1	1	2	2	2	1	1				1	1	1	1	7	1	NO
2	1	1	2		1	1	3	2	1	1	2			1	1	2	1	2		2	SI
2	1	1	1	1	1	1	3	2	1	3	3			1	1	1	1	1	5	1	NO
2	1	1	2	1	1	1	3	2	1	1	2	1		1	1	1	1	1	5	1	NO
2	1	1	1	1	1	1	3	2	1	1	2	1	1	1	1	1	1	1	6	1	NO
2	1	1	1	1	1	1	3	2	1	1			1	1	1	1	1	1	5	1	NO

2	1	1	1	1	1	1	3	2	1	1			1	1	1	1	1	1	4	1	NO
2	1	1	1	1	1	1	3	2	1	2			1	1	1	1	1	1	3	1	NO
2	2	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	1	5	2	NO
2	1	1	1	1	1	1	2	2	1	3		2	1	1	1	1	1	1	2	1	NO
2	2	1	2		1	1	2	2	1	1	2		2	1	1	1	1	2		2	SI
2	2	1	1	1	1	1	2	2	1	1		2	1	1	1	1	1	1	6	1	NO
2	2	1	1	1	1	1	2	2	1	1		2	1	1	1	1	1	1	6	1	NO
2	2	1	2	2	1	1	2	1	1	1	2		2	1	1	1	1	2		1	SI
2	2	1	1	1	1	1	2		1	3		2	1	1	1	1	1	1	3	1	NO
2	1	1	1	1	1	1	2		1	1		2	1	1	1	1	1	1	1	1	NO
2	2	1	1	1	1	1	2		1	1		2	1	1	1	1	1	1	5	1	NO
2	1	1	1	1	1	1	2		1	2		2	1	1	1	1	1	1	4	1	NO
2	2	1	1	1	1	1	2		1	3		2	1	1	1	1	1	1	3	1	NO
1	1	2	2		2	2	2		2	1	2	2	2	1	1	2	2	1	4	2	SI
1	2	2	2		2	2	2		2	1	2		2	1	2	2	1	1		2	SI
2	1	1	2		2	2	2		1	1	2		2	1	1	2	1	1		2	SI
2	1	1	2		2	2	4		1	1	2		2	1	1	2	1	1		2	SI

2	1	1	1	1	1		2	2	1	3	1	1	1	1	1	1	1	1	5	1	NO
2	1	1	2		2	2	4		1	1	2		2	1	1	2	1	1		2	SI
2	2	1	2		2	2	4		1	1	2			1	1	2	1	1		2	SI
2	1	1	1	1	1		2	2	1	1	1		1	1	1	1	1	1	5	1	NO
2	1	1	2	1	1		2	2	1	1			1	1	1	1	1	1	7	1	NO
2	1	1	2	1	1		2	2	1	1			1	1	1	1	1	1	2	2	NO
2	2	1	2	1	1		2	2	1	1			1	1	1	1	1	1	7	1	NO
2	1	1	2	1	1	1	2	2	1	3			1	1	1	1	1	1	7		NO
2	1	1	1	1	1	1	2	2	1	3	2	1	1	1	1	1	1	1	7		NO
2	2	1	1	1	1	1	2		1	1	2		1	1	1	1	1	1	3		NO
2	1	1	2		2	2	3	2	1	1	2		1	1	1	2	2	2	4	2	SI
2	2	1	2		2	2	3	2	1	1	2		1	1	1	2	1	1	1	2	SI
2	2	1	1	1	1	1	2		1	1	2		1	1	1	1	1	1	6	1	NO
2	1	1	1	1	1	1	2		1	1	2		1	1	1	2	1	1	2	1	NO
2	1	1	1	1	1	1	2		1	1	2		1	1	1	2	1	1	1	1	NO
2	1	1	1	1	1	1	2		1	1	2		1	1	1	2	1	1	1	1	NO
2	1	1	1	1	1	1	2		1	3	2		1	1	1	2	1	1	7	1	NO

1	1	1	2		2	2	3	2	1	1	2		1	1	1	2	1	1	6	2	SI
2	2	1	1	1	1	1	2		1	3	2		1	1	1	2	1	1	3	1	NO
2	2	1	1	1	1	1	2		1	1	2		1	1	1	2	1	1	1	1	NO
2	2	1	1	1	1	1	2		1	1	2		1	1	1	2	1	1	6	1	NO
2	1	1	1	1	1	1	2		1	1	2		1	1	1	2	1	1	7	1	NO
1	2	2	1		2	2	3	2	2	1	2	2	1	1	1	2	1	1	7	1	SI
2	2	1	1		2	2	3	2	1	1	2	2	1	1	2	2	1	1	2	1	SI
2	2	1	1	1	1	1	2	1	1	1	2		1	1	1	2	1	1	6	2	NO
2	1	1	1	1	1	1	2	2	1	3	2		1	1	1	2	1	1	4	1	NO
2	1	1	1	1	1	1	2	2	1	1	2		1	1	1	2	1	1	5	1	NO
2	1	1	1	1	1		2	2	1	1	1		1	1	1	2	1	1	5	1	NO
2	1	1	1	1	1		2	2	1	3		2	1	1	1	2	1	1	3	1	NO
2	2	1	1	1	1		2	2	1	1		2	1	1	1	2	1	1	3	1	NO
2	1	1	1	1	1		2	2	1	1	2	2	1	1	1	2	1	1	5	1	NO
2	2	1	1	1	1		2	2	1	3	2	2	1	1	1	2	1	1	6	1	NO
2	1	1	1	1	1		2	2	1	3	2	2	1	1	1	2	1	1	4	1	NO
2	2	1	1	1	1		2	2	1	3	2	2	1	1	1	2	1	1	2	1	NO

2	2	1	1	1	1		2		1	1	2	2	1	1	1	2	1	1	5	1	NO
1	2	1	1		2	2	3	2	1	1	2	2		1	1	2	1	1	3		SI
2	1	1	1	1	1		2	2	1	1	2	2	1	1	1	2	1	1	4		NO
1	2	1	1	1	1		2	2	1	1	2	2	1	1	1	2	1	1	2		NO
2	2	1	1	1	1		1	2	1	3	2	2	1	1	1	2	1	1	6		NO
2	2	1	1	1	1	1	1	2	1	3	2	2	1	1	1	2	1	1	2		NO
2	1	1	1	1	1	1	1	2	1	1	2	2	1	1	1	2	1	1	7		NO
2	1	1	1	1	1	1	1	2	1	1	2	2	1	1	1	2	1	1	2		NO
2	2	1	1	1	1	1	1		1	3	2	2	1	1	1	2	1	1	4	1	NO
1	1	1	1		2	2	3	2	1	1	2	2	2	1	1	2	1	1	6	1	SI
2	1	1	1	1	1	1	1		1	1	2	2	1	1	1	2	1	1	5	1	NO
2	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	2	1	1	3	1	NO
2	1	1	1	1	1	1	1	2	1	1	2	2	1	1	1	2	1	1	7	1	NO
2	2	1	1	1	1	1	1	2	1	3	1	2	1	1	1	2	1	1	7	1	NO
2	1	1	1	1	1	1	1	2	1	1	3	2	1	1	1	2	1	1	6	1	NO
2	2	1	1	1	1	1	1	2	1	1	3	2	1	1	1	2	1	1	7	1	NO
1	2	2	1		2	1	3	2	2	1	2	2	2	1	1	2	1	1	4	1	SI

2	2	1	1	1	1	1	1	2	1	1	3	2	1	1	1	2	1	1	1	1	NO
2	2	1	1	1	1	1	1	2	1	1	3	2	1	1	1	2	1	1	4	1	NO
2	1	1	1	1	1	1	1	2	1	1	3	2	1	1	1	2	1	1	7	1	NO
1	2	2	1		2	1	3	2	2	1	3	2	2	1	1	2	1	1	5	2	SI
2	1	1	1	1	1		2	2	1	1	3	2	1	1	1	2	1	1	3	1	NO
2	1	1	1	1	1		2	2	1	1	3	2	1	1	1	2	1	1	1	1	NO
2	2	1	1	1	1		2		1	1	3	2	1	1	1	2	1	1	7	1	NO
2	1	1	1	1	1		2		1	1	3	2	1	1	1	2	1	1	6	1	NO
2	2	1	2		2	2	3	2	1	1	3	2	2	1	1	2	1	1	2	1	SI
1	1	1	1	1	1		2		1	1	3	2	1	1	1	2	1	1	6	2	NO
2	1	1	1	1	1		2		1	1	3	2	1	1	1	2	1	1	2	1	NO
1	2	2	2		2	2	3	2	2	1	3	2	2	1	1	2	1	1	1	2	SI
2	1	1	1	1	1		2		1	1	3	2	1	1	1	2	1	1	3	1	NO
2	2	1	1	1	1		2	1	1	1	3	2	1	1	1	2	2	1	2	1	NO
2	1	1	1	1	1		2		1	1	3	2	1	1	1	2	2	1	5	1	NO
2	2	1	1	1	1		2		1	1	3	2	1	1	1	2	2	1	3	1	NO
2	1	1	1	1	1		2		1	3	3	2	1	1	1	2	2	1	5	1	NO

1	1	1	1	1	1		2		1	3	3	2	1	1	1	2	2	1	6		NO
2	1	1	1	1	1		2		1	1	3	2	1	1	1	2	2	1	7		NO
2	1	1	2		2	2	3	2	1	1	3	2	2	1	1	2	2	1	6	2	SI
1	1	1	1	1	1		2		1	1	3	2	1	1	1	2	2	1	1		NO
2	2	1	2	2	2	2	3	2	1	1	3	2	2	1	1	2	1	1	6	2	SI
2	2	1	2		2	2	3	2	1	1	3	2	2	1	1	1	1	1	6	2	SI
2	2	1	2		2	2	3	2	1	1	3	2	2	1	1	1	1	1	3	1	SI
2	1	1	1	1	1		2	2	1	1	3	2	1	1	1	2	2	1	4	1	NO
2	2	1	1	1	1		2	2	1	1	3	2	1	1	1	2	2	1	4	1	NO
2	2	1	1	1	1		2	2	1	1	3	2	1	1	1	2	2	1	4	1	NO
2	2	1	2		1	2	3	2	1	1	4	2	2	1	1	1	1	1	5	1	SI
2	1	1	1	1	1	1	2	2	1	1	2	2	1	1	1	2	2	1	4	1	NO
2	1	1	1	1	1	1	2	2	1	1	2	2	1	1	1	1	2	1	1	1	NO
1	1	1	1	1	1	1	2	2	1	1	2	2	1	1	1	1	2	1	5	1	NO
2	1	1	1	1	1	1	2	2	1	1	2	2	1	1	1	1	2	1	3	1	NO
1	1	1	1	1	1	1	3	2	1	2	2	2	1	1	1	1	2	1	7	1	NO
2	1	1	1	1	1	1	3	2	1	1	1	2	1	1	1	1	1	1	5		NO

2	1	1	1	1	1	1	3	2	1	1	2	2	1	1	1	1	1	1	2		NO
2	2	1	1	1	1	1	3	2	1	1	1	2	1	1	1	1		1	5		NO
2	2	1	1	1	1	1	3		1	1	1	2	1	1	1	1		1	3		NO
2	1	1	1	1	1	1	3		1	1	1	2	1	1	1	1		1	4		NO
2	1	1	1	1	1	1	3		1	1	1	2	1	1	1	1		1	6		NO
1	1	1	1	1	1	1	3	1	1	1	1	2	1	1	1	1		1	7		NO
1	1	1	2		1	2	4	2	1	1	4	2	2	1	1	1		2	5	2	SI
2	1	1	1	1	1	1	3	2	1	1	1	2	1	1	1	1		1	5	1	NO
2	2	1	1	1	1	1	3	2	1	1	1	2	1	1	1	1		1	7	1	NO
2	1	1	1	1	1		3	2	1	1	1	2	1	1	1	1	1	1	1	1	NO
2	2	1	1	1	1		3	2	1	1	1	2	1	1	1	1	1	2	4	1	NO
2	1	1	1	1	1		1	2	1	1	1	2	1	1	1	1	1	2	2	1	NO
2	2	1	2		2	2	4	2	1	1	4	2		1	1	1	1	2	1	1	SI
2	1	1	2		2	2	4	2	1	1	4	2	1	1	2	1	2	2	7	1	SI
2	1	1	1	1	1		1	2	1	3	1	2	1	1	1	1	1	2	4	1	NO
2	2	1	1	1	1		1	2	1	1	1	2	1	1	1	1	1	2	6	1	NO
2	2	1	1	1	1		1	2	1	1	1	2	1	1	1	1	1	2	4	1	NO

2	1	1	1	1	1		1	2	1	1	1	2	1	1	1	1	1	2	7	1	NO
2	1	1	1	1	1		1	2	1	1	1	2	1	1	1	1	1	2	6	1	NO
2	1	1	1	1	1		1	2	1	3	1	2	1	1	1	1	1	2	4	1	NO
1	2	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	2	1	1	NO
2	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	2	5	1	NO
2	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	2	1	1	NO
2	2	1	2		2	1	4	2	1	1	4		1	1	1	1	1	2	2	2	SI
2	1	1	2		2	1	2	2	1	1	2		1	1	1	1	1	2	6	2	SI
2	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	2	2	1	NO
2	1	1	2	1	1	1	1	2	1	1	1	2	1	1	1	1	1	2	1	1	NO
1	1	1	2	1	1	1	1		1	1	1	2	1	1	1	1	1	2	3		NO
2	1	1	2	1	1	1	1		1	1	1	2	1	1	1	1	1	2	4		NO
2	2	1	2	1	1	1	1		1	1	1	2	1	1	1	1	1	2	2		NO
2	2	1	2		2	1	2	1	1	1	2		1	1	1	2	1	1	4	2	SI
1	1	1	2	1	1	1	1		1	1	1	2	1	1	1	1	1	2	7		NO
1	1	1	2	1	1	1	1	2	1	1	1	2	1	1	1	1		2	5	1	NO
2	1	1	2	1	1	1	2	2	1	1	1	2	1	1	1	1		2	1	1	NO

1	1	1	2	1	1	1	2	2	1	1	1	2	1	1	1	1		2	3	1	NO
2	1	1	2	1	1	1	2	2	1	1	1	2	1	1	1	1		2	2	1	NO
2	2	1	2	2	2	1	3	2	1	1	2		1	1	1	1		1	7	1	SI
1	1	1	2	1	1	1	2	2	1	1	1	2	1	1	1	1		2	4	1	NO
2	1	1	2	1	1	1	2	2	1	1	1	2	1	1	1	1		2	2	1	NO
1	1	1	2	1	1	1	2	2	1	1	1	2	1	1	1	1	1	2	4	1	NO
2	2	1	2		2	2	3	2	1	3	2		1	1	1	1	1	1	7	1	SI
2	1	1	2		2	1	3	2	1	1	2		1	1	1	1	1	1	3	1	SI
2	1	1	2	1	1	1	2	2	1	1	1	2	1	1	1	1	1	2	5	1	NO
2	1	1	1		2	1	3		1	3	2	3	1	1	1	2	1	1	1	1	SI
2	2	1	1		2	1	3		1	1	4	3	1	1	2	2	1	1	7	1	SI
1	1	1	2	1	1	1	2	1	1	1	1	2	1	1	1	1	1	2	1	1	NO
2	1	1	1		2	2	3		1	1	4	3	2	1	1	2	2	1	7	1	SI
2	2	1	1		2	2	3		1	1	4	3	2	1	1	2	1	1	2	1	SI
1	1	1	2	1	1	1	2		1	1	1	2	1	1	1	1	1	2	5	1	NO
2	1	1	1		2	2	3		1	1	4	3	2	1	1	2	1	2	5	1	SI
1	2	1	1		2	2	3	2	1	1	4	3	2		1	2	1	2	4	1	SI

2	1	1	1		2	2	3	2	1	1	4	3	2		1	2	1	2	5	1	SI
2	2	1	1	1	1	1	2	2	1	1	1	2	1	1	1	1		2	2	1	NO
2	1	1	1	1	1	1	2	2	1	3	1		1	1	1	1		2	2		NO
2	1	1	1	1	1	1	2	2	1	3	1		1	1	1	1		2	6		NO
2	1	1	1	1	1	1	2	2	1	3	1		1	1	1	1		2	6		NO
1	1	1	1	1	1	1	2	2	1	3	1		1	1	1	1		2	6		NO
1	1	1	1	1	1	1	2	2	1	3	1		1	1	1	1		2	6		NO
1	1	1	1	1	1	1	2	2	1	3	1		1	1	1	1		2	1	1	NO
1	2	1	1	1	1	1	2	2	1	3	1	1	1	1	1	1	1	2	5	2	NO
1	1	1	2		2	2	3	2	1	3	4	3	2		1	2	1	2	7	2	SI
1	2	1	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1	2	3	1	NO
1	1	1	1	1	1	1	2	2	1	3		1	1	1	1	1	1	2	5	1	NO
1	1	1	2		2	2	3	2	1	1	4	3			1	2	2	2	4	2	SI
2	2	1	1	1	1	1	2		1	1		1	1	1	1	1	1	2	5	1	NO
1	2	1	2		2	2	3	2	1	3	3	3			1	1	1	1	6	2	SI
2	1	1	1	1	1		2	2	1	3		1	1	1	1	1	1	2	2	1	NO

1	1	1	2		2	2	2	2	1	3	3	3			1	1	1	1	1	2	SI
2	2	1	1	1	1		2	2	1	1		1	1	1	1	1	1	2	1	1	NO
2	2	1	2		2	2	2	2	1	3	3				1	1	1	1	6	2	SI
1	2	1	1	1	1		2	2	1	3		1	1	1	1	1	1	2	2	1	NO
2	1	1	1	1	1		2	2	1	3		1	1	1	1	1	1	2	2	1	NO
2	1	1	1	1	1		2	2	1	3		1	1	1	1	1	1	2	7	1	NO
1	1	1	2		2	2	2	2	1	3	3				1	1	1	2	6	1	SI
1	1	1	1	1	1		3	2	1	3		1	1	1	1	1		2	1	1	NO
1	1	1	1	1	2		3	2	1	3		1	1	1	1	1			2	1	NO
2	2	1	1	1	2		3	2	1	3		1	1	1	1	1			3	1	NO
1	1	1	1	1	1		3	2	1	3		1	1	1	1	1			2	1	NO
1	1	1	1	1	1		3		1	2		1	1	1	1	1			1	1	NO
2	2	1	1	1	1		3		1	1		1	1	1	1	1			5	1	NO
2	2	1	2	1	1		3		1	1		1	1	1	1	1			1	1	NO
1	1	1	2	1	1		3		1	3		1	1	1	1	1			3	1	NO
2	2	1	2	1	1		3		1	3		1	1	1	1	1		1	2	1	NO
1	2	1	2		2	2	2		1	3	3				1	1	1	2	7	2	SI

1	2	1	1	1	1		3		1	3		1	1	1	1	1	1	1	3	1	NO
2	2	1	2		2	2	2		1	1	3	3			1	1	2	2	6	2	SI
2	1	1	1	1	1		3		1	1		1	1	1	1	1	1	1	2	1	NO
1	2	1	2		2	1	2		1	2	3	3		1	1	2	1	2	6	2	SI
1	1	1	1	1	1		3		1	3		1	1	1	1	1		1	2	1	NO
2	2	1	1	1	1		3		1	3		1	1	1	1	1		1	6	1	NO
1	1	1	1	1	1		3		1	1		1	1	1	1	1		1	6	1	NO
2	2	1	2	2	2	1	2		1	3	3	3			1		1	2	6	2	SI
2	2	1	1	1	1		3	2	1	3		1	1	1	1	1	1	1	7	1	NO
1	1	1	2		1	1	2	1	1	3	3	3	2		1	1	1	2	6	1	SI
2	1	1	1	1	1		1		1	3		1	1	1	1	1	1	1	5	1	NO
2	2	1	2		1	1	2		1	3	3	3	2		1	1	1	2	2	1	SI
2	2	1	1	1	1		1	2	1	3		1	1	1	1	1	1	1	3	1	NO
2	1	1	1	1	2		1	2	1	3		1	1	1	1	1	1	1	5	1	NO
2	1	1	1	1	2		1	2	1	1		1	1	1	1	1		1	1	1	NO
2	1	1	1	1	2		1	2	1	2		1	1	1	1	1	1	1	2	1	NO
1	2	1	2		1	2	2	2	1	3	3	3	2		1	1	1	2	7	2	SI

1	2	1	2		1	2	2	2	1	3	3	3	2		1	1	2	2	1	2	SI
1	1	1	1	1	1		1	2	1	1		1	1	1	1	1	1	1	7	1	NO
1	1	1	2		1	2	2	2	1	3	3	3	2		1	1	1	1	7	1	SI
2	2	1	1		2	2	3	2	2	3	3	3	2		2	1	1	1	5	2	SI
1	1	1	1	1	1		1	2	1	1		1	1	1	1	1	1	1	7	1	NO
1	1	1	1	1	1		1	2	1	1		1	1	1	1	1	1	1	3	1	NO
1	2	2	1		2	2	3	2	2	3	4	3	2		1	1	1	1	6	1	SI
1	1	1	1		2	2	3	2	1	1	4		2		1	2	1	1	4	1	SI
1	2	1	1	1	1		1	1	1	3		1	1	1	1	1	1	1	6	1	NO
2	2	2	1		2	2	3		2	1	4	2	2		1		1	1	6	2	SI
1	1	1	1		2	2	3	2	1	3	4	2	2		1	1	2	1	5	2	SI
1	2	1	1	1	1		1	2	1	3		1	1	1	1	1	1	1	3	1	NO
2	1	1	1	1	1		1	2	1	3		1	1	1	1	1		1	5	1	NO
1	2	1	1	1	1		1	2	1	3		1	1	1	1	1	1	1	1	1	NO
1	1	1	2		2	2	3	2	1	3	4	2	2		1	1	1	1	7	1	SI
1	2	1	1	1	1		1	2	1	3		1	1	1	1	1	1	1	2	1	NO
1	2	1	2		2	2	3	2	1	3	4	2			1	1	1		5	2	SI

1	1	1	2		2	2	3	2	1	3	4	2			1	1	1		7	2	SI
1	2	1	2	2	2	2	1	2	1	3		2			1	1	1		5	2	SI
1	2	1	1	1	1		2		1	3		1	1	1	1	1	1		2	1	NO
2	2	1	2		2	2	3	2	2	3		2			1	1	1		5	2	SI
2	2	1	1	1	1		2		1	3		1	1	1	1	1	1	1	4	1	NO
1	1	1	2		1	1	1		1	3		2			1	2	2	2	1	1	SI
2	2	1	1	1	1		2	2	1	3		1	1	1	1	1	1	1	6	1	NO
1	1	1	1	1	1		2	2	1	1		1	1	1	1	1	1	1	1	1	NO
1	2	1	2		2	1	1	2	1	3		1			1	1	1	2	5	2	SI
1	1	1	2		2	1	1	2	1	3		1			1	1	1	2	4	2	SI
1	1	1	2		2	2	1	2	1	3					1	1	1	2	6	2	SI
1	1	1	2		2	2	1	2	1	3		1	2		2	1	2	2	3	1	SI
1	1	1	1	1	1		2	2	1	3		1	1	1	1	1	1	1	1	1	NO
1	2	1	1	1	1		2	2	1	1			1	1	1	1	1	1	3	2	NO
1	1	1	1	1	1		2		1	3			1		1	1	1	1	4	1	NO
2	1	1	2		2	2	1		1	3	4		2		1	1	1	1	6	2	SI
1	2	1	1	1	1		2	1	1	3					1	1	1	1	1	1	NO

1	1	1	2		2	2	3		1	3	4		2		1	1	1	1	3	2	SI
1	2	1	1		2	2	3	2	1	3	4		2		1	2	1	1	5	2	SI
2	2	1	1		2	2	3	2	1	3	4		2		1	2	1	1	1	2	SI
2	1	1	1	1	1		2	2	1	1					1	1		1	3	1	NO
2	1	1	1	1	2		2	2	1	3		2			1	1		1	7	1	NO
1	1	1	1	1	1		2	2	1	3		2			1	1			5	1	NO
1	1	1	1	1	1		2	2	1	1		2	2		1	1			3	1	NO
1	2	1	1		2	2	3	2	1	2	4		2		1	2	2		5	2	SI
1	2	1	1	1	1		2	2	1	3		2	2		1	1	1		4	1	NO
1	2	1	2		2	2	3	2	1	3	4		2		1	2	1		6	2	SI
1	2	1	1	1	1		2	2	1	1		2	2		1	1			5	1	NO
1	2	1	1	1	1		1	2	1	3		2	2		1	1			6	1	NO
1	1	1	1	1	1		1	2	1	3		2	2		1	1			2	1	NO
1	1	1	2		2	2	3	2	1	3	4		2		1	2			3	2	SI
1	1	1	1	1	1		1	2	1	3		2	2		1	1			2	1	NO
2	1	1	1	1	1		1	2	1	3		2	2		1	1			5	1	NO
1	1	1	1	1	1		1	2	1	3		2	2		1	1		1	1	1	NO

1	1	1	2		2	2	3	2	1	3	4		2		1	2		1	3	2	SI
2	2	1	1		2	2	3	2	1	3	4		2		2	2		1	5	2	SI
1	2	1	1	1	1		1	2	1	2		2	2		1	1		1	7	1	NO
2	2	1	2	1	1		1	2	1	3	1	2	2		1	1	1	1	3	1	NO
1	1	1	1	1	2		1	2	1	3		2	2		1	1	1	1	4	1	NO
1	2	1	1		2	2	3	2	1	1			2		1	2	1	1	5	2	SI
1	2	1	2		2	2	3	2	2	3		2	2		1	1	2	1	5	2	SI
1	2	1	1	1	1		1	2	1	3		2	2		1	1	1	1	2	1	NO
1	1	1	1		2	2	1	2	1	3		2	2		1	1	1	1	1	2	SI
1	1	1	1	1	1		1	2	1	2		2	2		1	1	1	1	5	1	NO
1	1	1	2		2	1	1	2	1	2	3	2	2		1	1	1	1	4	1	SI
1	2	1	2		2	1	1	2	1	3	3	3	2		1	2	1	1	3	2	SI
1	1	1	2		2	2	1	2	1	3	3	3	2		1	1	1	1	4	2	SI
2	2	1	2		2	2	1	2	1	3	3	3	2		1	2	1	1	2	2	SI
1	1	1	1	1	1		1	2	1	2		2	2		1	1	1	1	6	1	NO
1	1	1	1	1	1		1	2	1	3		2			1	1	1	1	5	1	NO
2	1	1	1	1	1		1	2	1	3		2			1	1	1	1	4	1	NO

1	2	1	2		2	2	3	2	1	3	3	3	2		1	2	1	1	7	2	SI
1	2	1	1	1	1		1	2	1	1		2			1	1	1	1	7	1	NO
1	2	1	2		1	2	3	2	1	3	3	3	2		1	2	2	1	5	1	SI
1	1	1	2		1	2	3	2	1	3	3	3	2		1	2	1	1	2	2	SI
1	1	1	2		1	2	3	2	1	3	3	3	2		1	2	1	2	4	2	SI
1	2	2	2		1	2	3	2	2	3	3	3	2		1	2	1	2	4	2	SI
1	1	1	2		1	2	3	2	1	1	3	3	2		1	2	1	2	7	2	SI
1	1	1	1	1	1		1	2	1	3		2			1	1	1	2	3	1	NO
1	1	1	1	1	1		1	2	1	3		2			1	1	1	2	3	1	NO
1	2	1	2		1	2	3	2	1	3	1		2		1	2	1	2	5	2	SI
2	2	1	1	1	1		1	2	1	1		2			1	1		2	5	1	NO
2	1	1	1	1	1		1	2	1	3					1	1		2	2	1	NO
1	1	1	1	1	1		1	2	1	3					1	1		2	3	1	NO
1	2	1	2		1	2	3	2	1	3	1		2		1	2		2	6	2	SI
1	1	1	1		1	2	3	2	1	3	1				1	2		2	1	2	SI
1	1	1	1	1	1		1	2	1	3					1	1		2	4	1	NO
1	1	1	1		1	2	3	2	1	3	1				1	2		2	5	2	SI

1	2	1	1		2	2	3		1	1	1				2	2		2	6	2	SI
1	1	1	1	1	1		2		1	3					1	1	1	2	3	1	NO
1	2	1	2		2	2	3		1	3	1				1	2	1	1	6	2	SI
1	1	1	2	2	2	2	3	2	1	3	1				1	2	2	1	7	2	SI
1	2	1	2		2	2	2	2	1	3	1				1	2	1	1	3	2	SI
1	2	1	2		2	1	2	2	1	3	1	2			1	2		1	5	2	SI
1	2	1	1		2	1	2	2	1	3	1	2			1	2		1	1	2	SI
1	1	1	1		2	1	2	2	1	3	3	2			1	2		1	3	2	SI
1	1	1	1	1	1		2	2	1	1					1	1		2	1	1	NO
1	1	1	1	1	1		2	2	1	3	2				1	1		2	4	1	NO
1	1	1	1	2	1		2	2	1	3	2	1			1	1		2	4	1	NO
1	1	1	1	1	1		2	2	1	1	2	1			1	1		2	6	2	NO
1	1	1	2		2	2	2	2	1	1	3	2			1	2		1	5	2	SI
1	2	1	2		2	2	2	2	1	3	3	2			1	1		1	7	2	SI
1	2	1	2		2	2	2	2	1	3	3	2			1	1		1	4	1	SI
1	1	1	2		1		2	2	1	3	2	1			1	1	1	2	4	1	NO
1	1	1	2	1	1		2	2	1	3	2	1			1	1	1	2	6	1	NO

1	2	1	2		2	2	2	2	1	3	3	2	2		1	1	1	1	7	2	SI
1	1	1	2		2	2	2	2	1	1	3	2	2		1	1	1	2	7	2	SI
1	2	1	2		1	2	4	2	1	1	3	2	2		1	1	1	2	2	2	SI
1	2	1	2	1	1		2	2	1	1	2	1	1		1		1	2	3	1	NO
1	2	1	2	1	1		2	2	1	1	2	1	1				1	2	7	1	NO
1	2	1	2		2	2	4	2	1	1	3	2	2		1	1	1	2	4	2	SI
1	1	1	1		2	2	4	2	1	1	4	2	2		1	1	1	1	3	1	SI
1	1	1	1	1	2		2	2	1	3	2	1	1				1	2	7	1	NO
1	2	1	1	1	1		2	2	1	3	2	1	1				1	2	4	1	NO
1	1	1	1	2	1		2	2	1	3	2	1	1				1	2	5	1	NO
1	1	1	1	1	1		2	2	1	3	2	1	1				1	2	7	1	NO
2	2	1	1	1	1		2	2	1	1	2	1	1		2		1	2	7	1	NO
1	2	1	1		2	2	4	2	1	3	4	2	2		2	2	1	1	2	2	SI
1	2	1	1	1	1		2	2	1	3	2	1	1		2		1	2	3	1	NO
1	1	1	1	1	1		2	2	1	3	2	1	1		2		1	2	2	1	NO
1	1	1	1	1	1		2	2	1	3	2	1	1		2		1	2	2	1	NO
1	1	1	1	1	1		2	2	1	1		1	1		2		1	1	7	1	NO

1	2	1	1	2	1		2	2	1	3	1	1	1		2		1	1	3	1	NO
1	1	1	1	1	1		2	2	1	3	1	1	1		2		1		2	1	NO
1	1	1	1	1	1		2	2	1	3	1	1	1		2		1		6	1	NO
1	2	1	1		2	2	4	2	1	3	4	2	2		1	1	1		3	2	SI
1	2	1	1		1	2	4	2	1	3	4	2	2		1	1	2		4	2	SI
1	1	1	1	1	1		2	2	1	1	1	1	1		2		1		2	1	NO
1	1	1	1	1	2		2	2	1	3	1	1	1		2		1		4	1	NO
1	1	1	1	1	1		2	2	1	1	1	1	1		2		1		7	1	NO
1	1	1	1	1	1		2	2	1	3	1	1	1		2		1		3	1	NO
1	1	1	2		2	2	4	2	1	2	4		1		1	1	1	1	6	1	SI
2	2	1	1	1	1		2	2	1	2	1		1		2		1	1	1	1	NO
1	2	1	2		2	2	4	2	1	3	4				1	1	1	1	5	2	SI
1	1	1	2		2	2	4	2	1	3	4				1	2	1	1	3	2	SI
1	2	1	1	1	1		2	2	1	1	1				2		1	1	2	1	NO
1	1	1	1	1	1		2	2	1	1	1				2		1	1	3	1	NO
1	2	1	2		2	1	4	2	1	1	4				1	1	1	1	3	2	SI
1	2	1	2	2	2	1	3	2	1	1	4				2	1	2	1	7	2	SI

1	2	1	2		1	1	3	2	1	3	4				1	1	1	1	7	2	SI
1	1	1	1	1	1		2	2	1	1	1	3			2		1	1	5	1	NO
2	2	1	2		2	1	3	2	1	3	4				1	1	1	1	6	2	SI
2	1	2	2		2	2	3	2	2	2	4	4			1	1	1	1	2	2	SI
1	1	1	2		2		2	2	1	1		3					1	1		1	NO
1	1	1	2	1	1		2	2	1	3		3					1	1		1	NO
1	1	1	2	1	1		2	2	1	3	1	3					1	1		1	NO
1	1	1	2	1	1		2	2	1	1	1	3					1	1		1	NO
1	1	1	2	1	1		2		1	3	1	3					1	1		1	NO
1	1	1	2	1	1		2		1	2	1	3					1	1		1	NO
1	1	1	2	1	1		2		1	3	1	3					1	1		1	NO

Diseño de formulario para la participación de los estudiantes de la universidad de Jaén de la carrera de Ingeniería de Industrias Alimentarias.

10/12/21 0:30

DESERCIÓN ESTUDIANTIL - INGENIERIA DE INDUSTRIAS ALIMENTARIAS

DESERCIÓN ESTUDIANTIL - INGENIERIA DE INDUSTRIAS ALIMENTARIAS

El objetivo de este formulario es recopilar los datos y factores que influyen en la decisión universitaria de la carrera de Ingeniería de Industrias Alimentarias. Agradecemos de antemano a todos los participantes que nos ayuden proponer nuevos métodos para detener el fenómeno de deserción que ha afectado a la carrera de Ingeniería de Industrias Alimentarias. La información que respondan en este formulario, será estrictamente privada.

***Obligatorio**

1. Apellidos y Nombres

Ejm: Aldana Córdova, José Matthias

2. Edad

Ejm: 18

3. Genero *

Marca solo un óvalo.

Masculino

Femenino

4. Estado Civil

Marca solo un óvalo.

Soltero(a)

Casado(a)

Conviviente

Separado(a)

5. Tipo de colegio secundario de procedencia

Marca solo un óvalo.

- Nacional
 Particular

6. Nombre y lugar del colegio secundario de procedencia, entiéndase por "lugar"
Distrito y Provincia

7. Mencione el Departamento, Provincia y distrito al que pertenecía cuando cursaba la carrera profesional.

8. ¿Ha tenido algún beneficio por parte de la universidad? De ser si la respuesta detalle el beneficio que obtuvo

Ejm: Beca, Comedor Universitario u otros.

9. ¿Cual fue su nivel de satisfacción con respecto a la enseñanza cuando cursaba la carrera profesional?

Marca solo un óvalo.

- Óptima
 Buena
 Regular
 Mala

10. ¿Cuándo cursaba la carrera profesional, usted trabajaba?

Marca solo un óvalo.

- Si
 No

11. ¿Cuál es el nivel educativo máximo alcanzado de su apoderado?

Marca solo un óvalo.

- Primaria
 Secundaria
 Técnica
 Universitaria
 Otros

12. ¿Cuál era el número de integrantes en su familia cuando cursaba la carrera profesional?

13. ¿Al momento de iniciar su carrera profesional, usted ya tenía hijos?

Marca solo un óvalo.

- Si
 No

14. Si usted no concluyo sus estudios Seleccione cuál fue la razón. (Puede seleccionar varias alternativas).

Selecciona todos los que correspondan.

- Discapacidad Física
- Falta de apoyo de mis padres.
- Falta de orientación vocacional
- Cambio de Carrera Profesional
- Desaprobación de cursos
- Inasistencia a clases
- Sanción disciplinaria
- Separación de mis padres
- Falta de recursos económicos
- Escogí la carrera profesional por influencia de mis padres o amistades
- Malas influencias (vicios)
- La universidad esta muy lejos de mi domicilio.
- Embarazo no planeado
- Paternidad no planeada

15. Si conoce algún compañero que no concluyo sus estudios y sabe cual fue la razón por favor Seleccione cuál fue la razón. (Puede seleccionar varias alternativas).

Selecciona todos los que correspondan.

- Discapacidad Física
- Falta de apoyo de mis padres.
- Falta de orientación vocacional
- Cambio de Carrera Profesional
- Desaprobación de cursos
- Inasistencia a clases
- Sanción disciplinaria
- Separación de mis padres
- Falta de recursos económicos
- Escogí la carrera profesional por influencia de mis padres o amistades
- Malas influencias (vicios)
- La universidad esta muy lejos de mi domicilio.
- Embarazo no planeado
- Paternidad no planeada

Respuestas de estudiantes de la universidad de Jaén de la carrera de Ingeniería de Industrias Alimentarias.

10/12/21 0:46

DESERCIÓN ESTUDIANTIL - INGENIERIA DE INDUSTRIAS ALIMENTARIAS

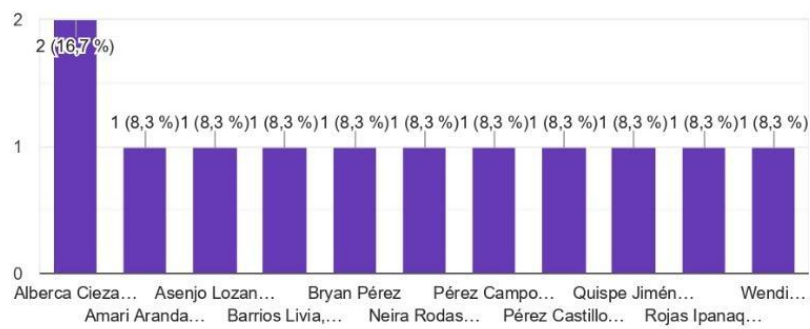
DESERCIÓN ESTUDIANTIL - INGENIERIA DE INDUSTRIAS ALIMENTARIAS

14 respuestas

[Publicar datos de análisis](#)

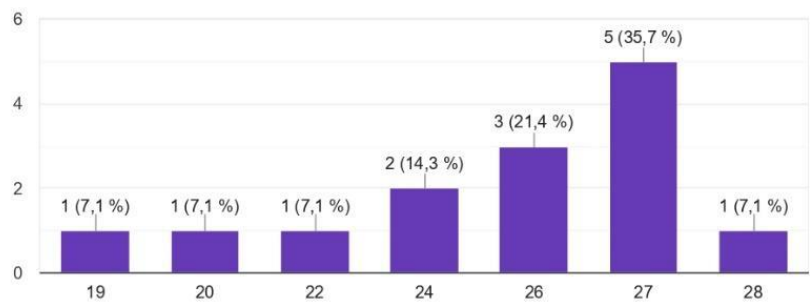
Apellidos y Nombres

12 respuestas



Edad

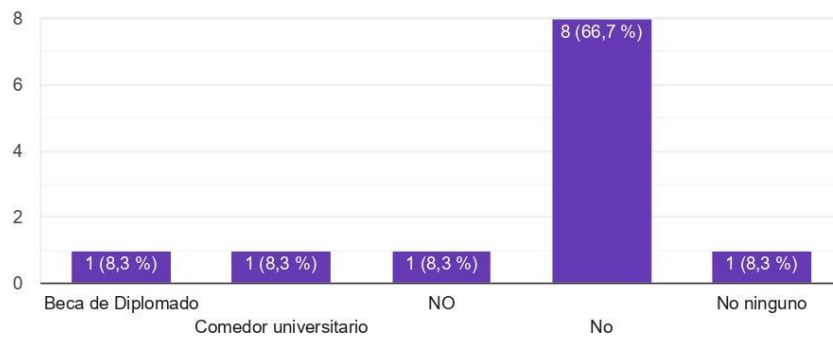
14 respuestas



Nombre y lugar del colegio secundario de procedencia, entiéndase por "lugar" Distrito y Provincia

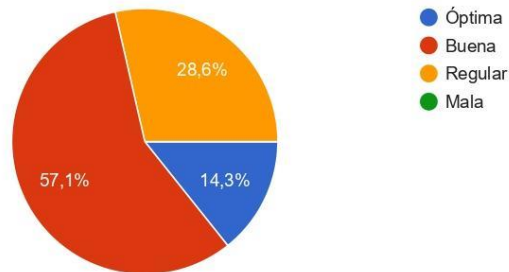
¿Ha tenido algún beneficio por parte de la universidad? De ser si la respuesta detalle el beneficio que obtuvo

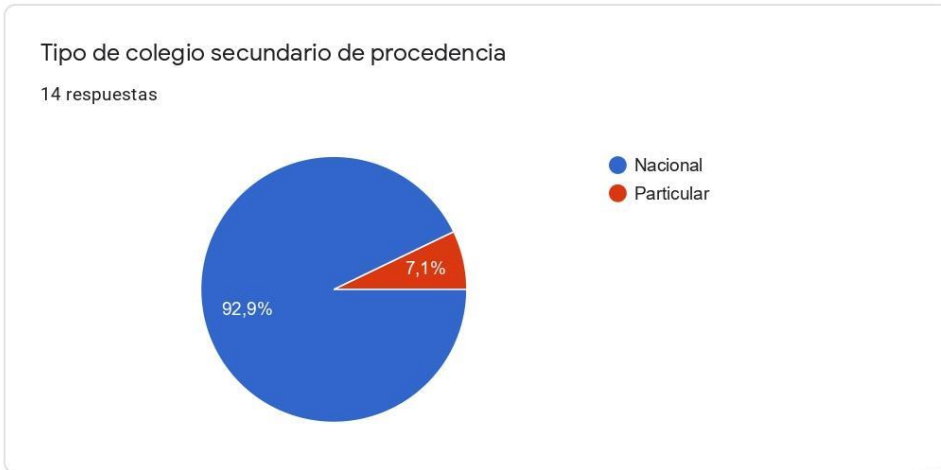
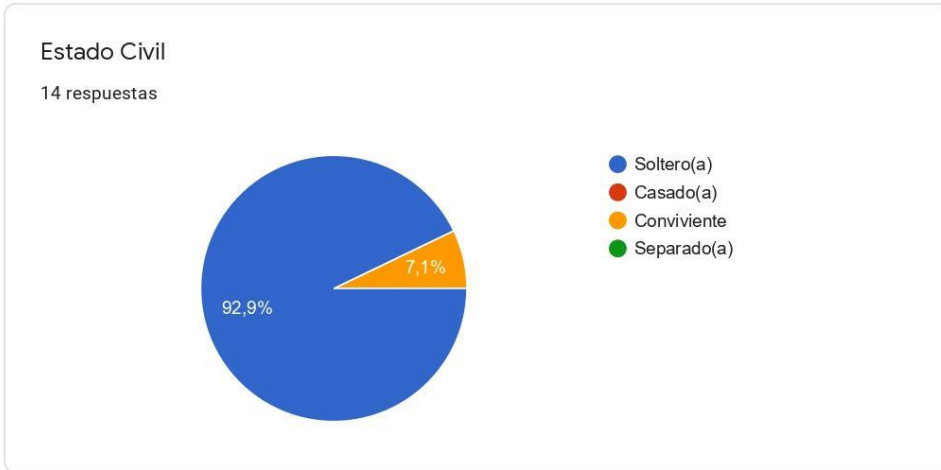
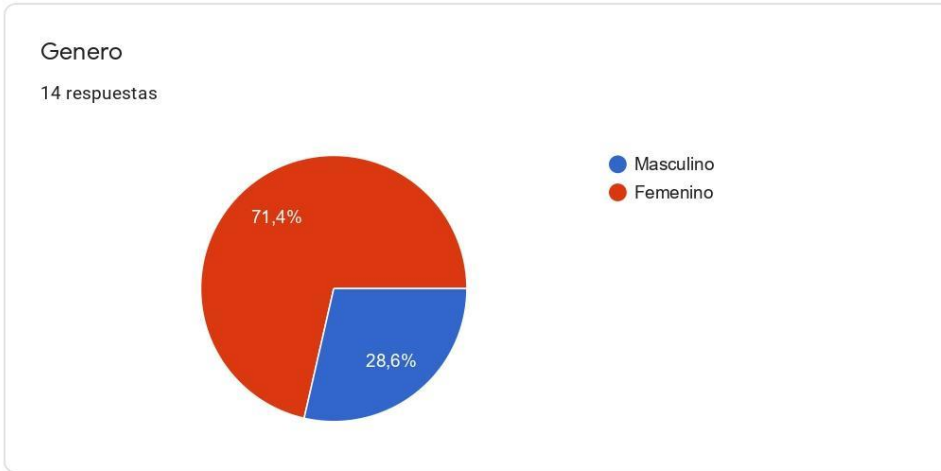
12 respuestas



¿Cual fue su nivel de satisfacción con respecto a la enseñanza cuando cursaba la carrera profesional?

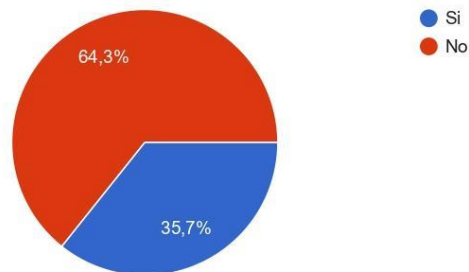
14 respuestas





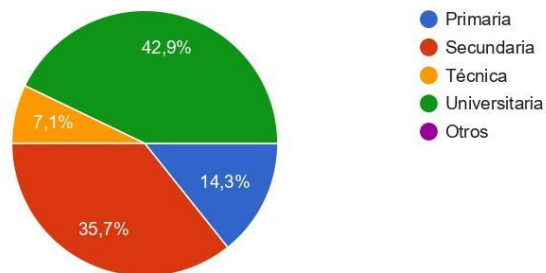
¿Cuándo cursaba la carrera profesional, usted trabajaba?

14 respuestas



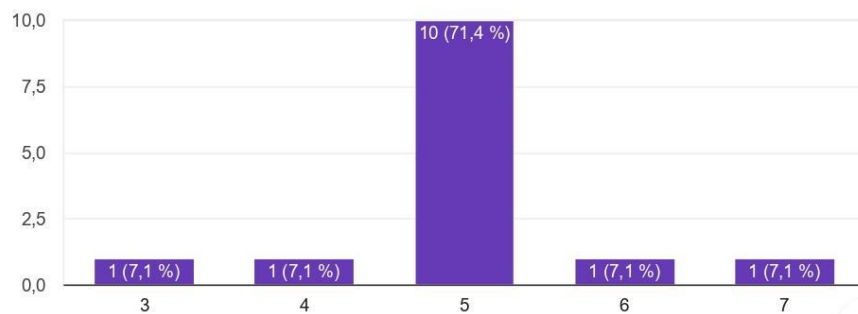
¿Cuál es el nivel educativo máximo alcanzado de su apoderado?

14 respuestas



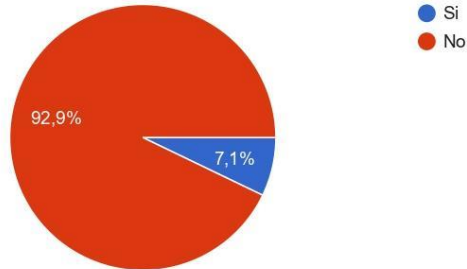
¿Cuál era el número de integrantes en su familia cuando cursaba la carrera profesional?

14 respuestas



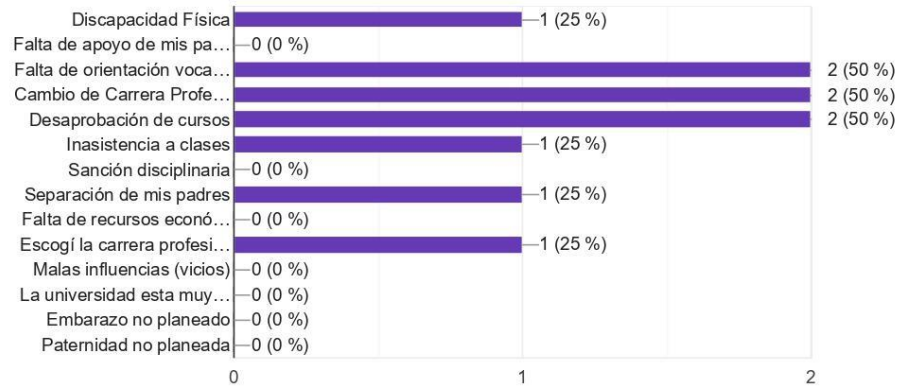
¿Al momento de iniciar su carrera profesional, usted ya tenía hijos?

14 respuestas



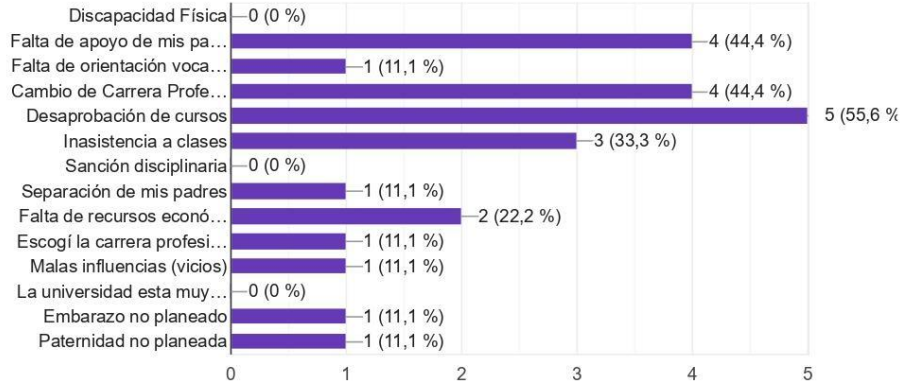
Si usted no concluyo sus estudios Seleccione cuál fue la razón. (Puede seleccionar varias alternativas).

4 respuestas



Si conoce algún compañero que no concluyo sus estudios y sabe cual fue la razón por favor Seleccione cuál fue la razón. (Puede seleccionar varias alternativas).

9 respuestas



Este contenido no ha sido creado ni aprobado por Google. [Notificar uso inadecuado](#) - [Términos del Servicio](#) - [Política de Privacidad](#)

Google Formularios



NOMBRE DEL TRABAJO	AUTOR
CamposBarreraSandroPaul-Turnitin.docx	Sandro Campos Barrera

RECuento DE PALABRAS	RECuento DE CARACTERES
21335 Words	112029 Characters

RECuento DE PÁGINAS	TAMAÑO DEL ARCHIVO
122 Pages	1.0MB

FECHA DE ENTREGA	FECHA DEL INFORME
Dec 5, 2023 9:02 AM GMT-5	Dec 5, 2023 9:04 AM GMT-5

● **17% de similitud general**

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos

- 14% Base de datos de Internet
- 3% Base de datos de publicaciones
- Base de datos de Crossref
- Base de datos de contenido publicado de Crossref
- 10% Base de datos de trabajos entregados

● **Excluir del Reporte de Similitud**

- Material bibliográfico
- Material citado
- Coincidencia baja (menos de 8 palabras)