



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA  
ESCUELA PROFESIONAL DE INGENIERÍA DE  
SISTEMAS**

**Patrones de la deserción estudiantil aplicando un modelo  
predictivo en un IESTP - Lima 2023**

**TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:  
Ingeniero de Sistemas**

**AUTORES:**

Mendoza Chero, Ricardo Ryan ([orcid.org/0000-0002-2129-4701](https://orcid.org/0000-0002-2129-4701))

Silva Arellano, Jose Luis ([orcid.org/0000-0002-2130-0227](https://orcid.org/0000-0002-2130-0227))

**ASESOR:**

Mgtr. Agurto Marchán, Winner ([orcid.org/0000-0002-0396-9349](https://orcid.org/0000-0002-0396-9349))

**LÍNEA DE INVESTIGACIÓN:**

Sistemas de Información y Comunicaciones

**LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:**

Desarrollo económico, empleo y emprendimiento

PIURA — PERÚ

2023

## **DEDICATORIA**

Se dedica esta investigación a nuestros padres que nos han acompañado durante el proceso de este proyecto que, gracias a su confianza y aliento constante, se pudo cumplir con nuestras metas trazadas.

## **AGRADECIMIENTO**

En primer lugar, quisiera agradecer a nuestros padres y docente, quienes nos han brindado su confianza y apoyo durante el proceso de la investigación, al tutor de la asignatura, por sus consejos constantes durante los estudios.

## Índice de contenidos

DEDICATORIA.....	ii
AGRADECIMIENTO .....	iii
ÍNDICE DE CONTENIDOS.....	iv
ÍNDICE DE TABLAS.....	v
ÍNDICE DE FIGURAS .....	vi
RESUMEN.....	vii
ABSTRACT.....	viii
I. INTRODUCCIÓN.....	1
II. MARCO TEÓRICO .....	6
III. METODOLOGÍA.....	20
3.1. Tipo y diseño de investigación.....	20
3.2. Variables y operacionalización .....	20
3.3 Población, Muestra y Muestreo .....	21
3.3.1 Población .....	21
3.3.4. Unidad de análisis.....	21
3.4. Técnicas e Instrumentos de Recolección de Datos .....	22
3.5. Procedimientos .....	25
3.6. Método de análisis de datos .....	27
3.7. Aspectos éticos.....	28
IV. RESULTADOS .....	30
V. DISCUSIÓN .....	36
VI. CONCLUSIONES .....	38
VII. RECOMENDACIONES.....	39
REFERENCIAS .....	40
ANEXOS .....	47

## Índice de tablas

Tabla 1: Unidad de análisis .....	22
Tabla 2: Instrumentos de recolección de datos para la variable: Deserción estudiantil	23
Tabla 3: Instrumentos de recolección de datos para la variable: Modelo Predictivo.....	24
Tabla 4: Evaluación de Clasificación por Variables Predictivas.....	34

## Índice de figuras

Ilustración 1: Diagrama de actividades del proceso de investigación .....	26
Ilustración 2: Matriz de correlación entre variables .....	30
Ilustración 3: Modelo de la variable Promedio ponderado .....	30
Ilustración 4: Modelo de la variable Promedio inasistencias.....	31
Ilustración 5: Validación cruzada del modelo de árbol de decisión de la variable promedio ponderado .....	33
Ilustración 6: Validación cruzada del modelo de árbol de decisión de la variable promedio de inasistencias.....	33

## Resumen

En este estudio de tipo aplicado se utilizó la metodología CRISP-DM para encontrar patrones de abandono de estudiantes en un programa de informática y ciencias de la computación del IESTP. Los datos se recogieron mediante un enfoque cuantitativo y un diseño descriptivo-experimental utilizando cuestionarios y un muestreo de conveniencia no probabilístico. Entre las variables importantes con correlaciones moderadamente positivas se incluyeron "Hábitos alimentarios", "Acceso al ordenador" y "Ejercicio regular", mientras que "Promedio ponderado" y "Promedio de ausencias" mostraron notables correlaciones negativas. Estas variables fueron el foco de los modelos de aprendizaje automático, que se construyeron con fuertes correlaciones y árboles de decisión visuales para apoyar su papel crítico en la predicción del abandono. Los valores de precisión del 77,06% y el 80,73% para "Media ponderada" y "Ausencias medias" obtenidos mediante validación cruzada validaron la capacidad de predicción de los modelos. Según el análisis exhaustivo, el 75,14% de los casos se clasificaron correctamente como "Media ponderada" y el 73,48% como "Ausencias medias". Concluyendo, estos resultados ofrecen valiosos insights para abordar la deserción estudiantil en instituciones similares, destacando la importancia de considerar múltiples factores y la aplicabilidad efectiva de modelos predictivos.

**Palabras clave:** Deserción estudiantil, modelo predictivo, minería de datos, factores de deserción.

## **Abstract**

In this applied-type study used CRISP-DM methodology to find patterns of student dropout in a computer science and informatics program at IESTP. Data were collected using a quantitative approach and a descriptive-experimental design using questionnaires and non-probability convenience sampling. Important variables with moderately positive correlations included "Eating habits," "Computer access," and "Regular exercise," while "Weighted average" and "Average absenteeism" showed notable negative correlations. These variables were the focus of the machine learning models, which were built with strong correlations and visual decision trees to support their critical role in predicting dropout. Accuracy values of 77.06% and 80.73% for "Weighted Mean" and "Mean Absences" obtained through cross-validation validated the predictive ability of the models. According to the comprehensive analysis, 75.14% of the cases were correctly classified as "Weighted mean" and 73.48% as "Mean absences". In conclusion, these results offer valuable insights for addressing student attrition in similar institutions, highlighting the importance of considering multiple factors and the effective applicability of predictive models.

**Keywords:** Student dropout, predictive model, data mining, dropout factors.

## I. INTRODUCCIÓN

El fenómeno de la deserción estudiantil ha sido objeto de gran atención en la comunidad académica. Se han examinado multitud de factores, incluidos factores sociales, culturales, económicos, familiares y personales, así como el propio proceso educativo, con sus variaciones en cuanto a contenido, métodos de impartición y calendario, como posibles predictores de la intención de un estudiante de abandonar los estudios. (OCDE, 2019)

Con respecto a la deserción estudiantil, es notable la escasez de estudios que empleen modelos predictivos para abordar este problema, el cual, sin lugar a dudas, demanda una atención particular. Las consecuencias negativas derivadas de la deserción impactan tanto a los estudiantes como a la sociedad, especialmente en términos económicos. Este fenómeno, que suele manifestarse principalmente entre el primer y último ciclo académico, encuentra sus raíces en diversas razones, como limitaciones de tiempo y recursos financieros, entre otros. Tal como lo subraya el estudio "Prediction model of first-year student desertion at Universidad Bernardo O'Higgins (UBO)" (2018), la creación de modelos predictivos se erige como una herramienta valiosa para comprender y abordar eficazmente la deserción estudiantil, permitiendo anticipar y mitigar las causas subyacentes y, en última instancia, mejorar la retención y el éxito académico de los estudiantes en la Universidad Bernardo O'Higgins.

Diversas investigaciones llevadas a cabo en Perú han corroborado que las limitadas oportunidades, combinadas con similitudes en las características socioeconómicas y educativas, son factores determinantes que explican las elevadas tasas de deserción en la educación superior del país. La deserción académica, entendida como el proceso en el cual un estudiante abandona de manera voluntaria o bajo coacción el campo de estudio elegido, se ve influenciada por diversos factores tanto internos como externos. Este fenómeno puede ser propiciado o desencadenado por circunstancias favorables o desfavorables. Como resalta el estudio "The Numerical Probabilistic Approach to the Processing and Presentation of Remote Monitoring Data" (2019), esta problemática no solo es un desafío local, sino que también requiere una

aproximación numérica y probabilística para procesar y presentar datos de monitoreo remoto, destacando la necesidad de estrategias efectivas y adaptativas para abordar la complejidad de la deserción académica en el contexto peruano.

Al emplear el término "modelo predictivo", se alude de manera inherente a una representación de la realidad, fundamentada en un intento descriptivo de vincular un conjunto de variables. El análisis predictivo no solo constituye un enfoque técnico, sino que también encierra la fusión estratégica de la gestión, la tecnología de la información y la aplicación de modelos estadísticos. Su propósito es abordar la interpretación de datos históricos con miras a proyectar patrones futuros. Este enfoque revela su potencial en el ámbito académico al proporcionar herramientas valiosas para entender y prever el rendimiento estudiantil mediante la extracción de conocimientos a partir de datos educativos. Este paradigma integrador impulsa no solo la eficacia de las estrategias educativas, sino también la toma de decisiones informada en el entorno educativo contemporáneo. Hoy en día contar con un modelo predictivo es de vital importancia para las organizaciones si se tiene como objetivo alinear y volver más eficientes tanto los recursos como estrategias. Los modelos predictivos permiten aproximar resultados futuros con una base proveniente de la ciencia de datos obteniendo múltiples beneficios desde manejar el inventario de una manera eficiente hasta mitigar riesgos. (Academic performance prediction by machine learning as a success/failure indicator for engineering students, 2021).

Los modelos predictivos necesitan un mínimo conocimiento de los datos del Big Data de la empresa para poder funcionar correctamente, por ello es importante la data histórica de la empresa. (Predictive model for the early detection of students with high risk of academic dropout, 2021).

Considerando esto, se formula la problemática general: ¿Cuáles son los patrones de deserción estudiantil aplicando un modelo predictivo en el programa académico de Computación e Informática de un IESTP? y los problemas específicos, a) ¿Cuáles son los factores más influyentes en la deserción estudiantil en el programa de Computación e Informática del IESTP?, b) ¿Cómo

se pueden modelar y representar los patrones de deserción estudiantil a través de diferentes algoritmos de machine learning? c) ¿Cuál es la precisión predictiva de los modelos construidos y cómo se pueden mejorar mediante ajustes y técnicas de optimización?

Los hallazgos de este estudio aportan una base sustancial para el avance de modelos predictivos en el ámbito de institutos de educación superior. Estos resultados, fundamentados en un análisis exhaustivo de patrones de deserción estudiantil, no solo amplían nuestra comprensión de los factores implicados en este fenómeno, sino que también ofrecen una plataforma robusta para la formulación y mejora de futuros modelos predictivos. Al proporcionar una visión detallada de las variables influyentes, correlaciones significativas y la efectividad de modelos de machine learning específicos, esta investigación se posiciona como un referente clave para la comunidad académica y profesional interesada en abordar la problemática de la deserción estudiantil. Además, al resaltar la aplicabilidad y relevancia práctica de los resultados en el contexto educativo, se sugiere un potencial impacto positivo en la toma de decisiones y estrategias institucionales orientadas a mejorar las tasas de retención estudiantil.

La presente investigación tiene como objetivo general Identificar patrones de deserción estudiantil desarrollando un modelo predictivo en el programa académico de Computación e Informática de un IESTP y como objetivos específicos, a) identificar y priorizar los factores más significativos relacionados con la deserción estudiantil en el programa de Computación e Informática, b) Desarrollar y contrastar modelos de aprendizaje automático con el propósito de identificar y representar patrones asociados a la deserción estudiantil. y c) Evaluar la exactitud en la capacidad predictiva de los modelos y realizar modificaciones con el fin de mejorar su rendimiento.

Según Fernández et al. (2018), En su artículo presenta el método KRA-SD, un análisis de relevancia basado en núcleos para la predicción del abandono escolar. La cual consta de dos etapas: una de alineación de kernel que correlaciona medidas académicas y biopsicosociales con etiquetas de salida (abandono/no abandono), y otra de clasificación de características que predice

el abandono utilizando k- nearest más cercanos y máquinas de vectores de soporte. Evaluando la validación y su prueba, usando un marco de entrenamiento, para los datos reales de la Universidad Tecnológica de Pereira. Teniendo como resultado una tasa de precisión de deserción estudiantil de un 85.64%.

Pedroza, Chasoy y Gómez (2019), Nos dice en su artículo que, mediante el uso de la minería de datos, el proyecto pretende determinar los temas recurrentes que diversos autores han abordado en relación con el abandono académico entre los estudiantes. En la investigación se utilizaron diversas fuentes. La exhaustividad de cada publicación en cuanto a técnicas, algoritmos, herramientas y atributos fue un factor clave en el proceso de selección. Los resultados demuestran que el aprendizaje supervisado es el objetivo principal de la mayoría de los estudios, siendo la clasificación la técnica más utilizada. Además, se observa una tendencia a emplear el algoritmo J48, que sitúa el nivel académico como la variable o clase principal que debe predecirse.

Kadiyala y Kumar (2018), En su proyecto nos habla que, para el fin de predecir las concentraciones de dióxido de carbono en los autobuses, se evalúa el rendimiento de una serie de algoritmos de aprendizaje automático, empleándose un enfoque metódico de implementación en Python. Utilizando medidas de rendimiento operacional para evaluar el modelo predictivo, determinando que el algoritmo xgboost rinde mejor que los demás algoritmos de este estudio.

Aqila y Faisal (2023), Recalca en su estudio que se centra en la utilización de un conjunto de datos de Kaggle para el análisis exploratorio de datos (EDA) y árboles de decisión para la clasificación del cáncer de pulmón. Utilizando el conocido algoritmo de árboles de decisión, obteniéndose así un alto recuerdo para la clase positiva y un bajo recuerdo para la clase negativa. En el conjunto de datos se incluyen datos demográficos y clínicos de los pacientes. La investigación clasifica de forma eficaz y bastante precisa el cáncer de pulmón mediante la construcción del modelo de Árboles de Decisión. Los resultados del EDA arrojan luz sobre variables importantes que afectan a la clasificación del

cáncer de pulmón, tanto positiva como negativamente. Este estudio podría contribuir a la creación de modelos de predicción del cáncer de pulmón.

## II. MARCO TEÓRICO

Barroso (2018) en su investigación, Desarrollo de un Modelo Predictivo para la Gestión Proactiva de Órdenes de Trabajo en el Área de Supply de BHP Minerals Americas, tuvo como objetivo el crear un modelo de predicción utilizando métodos de aprendizaje automático que permita a la unidad de Supply anticiparse a posibles caídas en el indicador WOPIFOT, mejorando la gestión de órdenes de trabajo con requerimiento de materiales, haciendo uso de la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), haciendo uso de datos históricos de orden de trabajo y materiales como muestra y el área de Supply de BHP Mineral Americas como población, teniendo como resultados que el GBT mostro resultados prometedores en la métrica de desempeño teóricas tales como el accuary de 82,2%, Cohen's Kappa de 0,625 y AUC de 0,731. También se nos dice que de darse un escenario ideal la herramienta actuaría de forma muy eficaz. Llegando a la conclusión de que se debe tener en cuenta las tecnologías de análisis y ciencia de datos de vanguardia, ya que benefician a la industria orientada a la producción y presentan oportunidades viables para mejorar los procedimientos mediante la automatización de los procedimientos de análisis de datos.

Barrueta Meza y Castillo Villarreal (2018) en su investigación denominada modelo de análisis predictivo para determinar clientes con tendencia a la deserción en bancos peruanos, Utilizando la herramienta SAP Predictive Analytics y la metodología PMBOK como base y guía principal para la creación de documentos de gestión, el objetivo era aplicar un modelo predictivo basado en la minería de datos en empresas minoristas. El número total de clientes de este proyecto previsto es de 20.000; 1.764 se clasifican como "desertores" y 18.236 como "no desertores". Los resultados del proyecto incluyen un documento de investigación, un informe final que presenta el modelo predictivo y un cuadro de mando creado a partir de la aplicación de SAP Predictive Analytics en empresas minoristas, y un plan de proyecto con riesgos, hipótesis y límites identificados. Se llegó a la conclusión que la implementación del modelo predictivo utilizando SAP Predictive Analytics permite reducir pérdidas monetarias al predecir las ventas en el área de Planeamiento Comercial de

empresas retail, proporcionando herramientas valiosas para la toma de decisiones. Además, la evaluación y aprobación de los entregables por el Comité de Proyecto.

Según Yovera (2021) en su investigación Modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C 2021, averiguar cómo afectaba el modelo predictivo al procedimiento de rescisión de contratos de IMP era el principal objetivo del estudio. La metodología utilizada fue CRISP-DM, teniendo una muestra de 4 reportes, con una población de 8 reportes. Se obtuvieron resultados tales como el incremento en el nivel de eficacia de las metas planteadas hasta en un 22%, así como la reducción de reportes hasta del 99.5%. Se llegó a la conclusión que la aplicación del modelo predictivo al procedimiento de resolución de contratos de clientes de Inmobiliaria Miraflores Perú S.A.C. fue exitosa, evidenciada por el aumento significativo en la eficacia de las metas planificadas y la notable reducción en el tiempo de generación de reportes. Esto respalda la utilidad del diseño del modelo predictivo al mejorar el proceso de resolución de los contratos.

En su investigación de análisis de modelos predictivos que estiman la deserción estudiantil de pregrado, Rivas (2019) en un esfuerzo por caracterizar modelos predictivos universitarios que identifiquen a los estudiantes que tienen una alta probabilidad de deserción. Hace uso de la técnica CRISP-DM, estudiantes presenciales como tipo de muestra, y evaluaciones de riesgo con un modelo de Cox en cuatro momentos claves durante los primeros meses de vida universitaria para predecir las conductas de permanencia/deserción en el tiempo. con una población de estudiantes matriculados en primer año de carreras de ingeniería de la Universidad de Atacama. Destacándose diversas herramientas y enfoques, como el Data Mining, KDD, muestreo estratificado, árboles de decisión y análisis de eventos históricos, que son empleados en la construcción de modelos predictivos para extraer información valiosa y tomar decisiones informadas, teniendo como conclusión que a pesar de las limitaciones del estudio, como la escasez de modelos predictores de deserción, se destaca la importancia de los modelos detectados y la necesidad de comprender los datos y variables clave para desarrollar herramientas efectivas de retención estudiantil.

Según Nina y Vilca (2018), en su investigación llamada búsqueda de patrones de comportamiento usando machine learning, para la toma de decisiones gerenciales en la empresa chuchuhuasi, con el objetivo de evaluar el impacto del machine learning a través de patrones de comportamiento a lo largo de los procesos de negocio y generar recomendaciones para la toma de decisiones de gestión de ChuchuHuasi SA., la metodología utilizada fue BPM: RAD, con una población 4000 y una muestra de 2196, como resultado se determinó el proceso de toma de decisiones de gestión utilizando la información recopilada de los clientes que visitaron el establecimiento, haciendo hincapié en los rasgos y preferencias individuales del cliente. La empresa utiliza software genérico para realizar sus tareas; no utiliza ningún software especializado. Se utilizó la herramienta bizagi BPMS ya que tiene varios beneficios relevantes para este proyecto. Como conclusión se tiene que, a través de la generalización de patrones de comportamiento en los procesos de negocio, se comprobó la influencia del machine learning y se formularon recomendaciones al proceso de toma de decisiones de la dirección de ChuchuHuasi S.A.

Alarcón (2021) en su investigación Sistema analítico basado en un modelo predictivo de procesamiento de datos en la big data en la educación superior, cuyo objetivo era procesar datos académicos mediante un sistema analítico basado en un modelo predictivo que considera la relación entre grandes volúmenes de datos y técnicas predictivas integradas. La metodología MAMBO se centró en la mediación empresarial, la adquisición de datos, la gestión, la búsqueda y la clasificación. La población está determinada por los datos académicos de todos los estudiantes universitarios a partir de 2005, y la muestra a tratar se refiere a los ciclos académicos de 2016-I a 2020-I, que corresponde a los últimos 5 años, totalizando 684 039 registros académicos, los resultados destacaron en las áreas de mejora en la suficiencia y claridad de datos, accesibilidad, seguridad, y tiempo de respuesta. Además, resalta la importancia de herramientas tecnológicas y estándares de calidad. Estos resultados proporcionan una base para la identificación de áreas críticas y la implementación de mejoras en el procesamiento de datos académicos. En

conclusión, al analizar los indicadores, los resultados se corroboraron estadísticamente. El rendimiento global del sistema mejoró un 8,03%, la precisión de la clasificación del rendimiento académico aumentó un 1,3% y la precisión de la clasificación del rendimiento académico aumentó un 5,12%.

Sifuentes (2018), cuya investigación fue Modelo predictivos de la deserción estudiantil en una universidad privada del Perú, cuyo objetivo fue determinar cómo el uso de modelos predictivos en asignaturas críticas contribuye a identificar, al inicio del curso, a los estudiantes en riesgo de deserción. Cuya metodología utilizada fue CRISP-DM, En la muestra se consideró la información histórica de 4,478 estudiantes, contando con una población constituida por 14,467 estudiantes universitarios, teniendo como conclusión que, si bien es cierto que el abandono estudiantil es un tema latente en las universidades públicas y privadas de todo el Perú, nuestras investigaciones han demostrado que existen algunos mecanismos o medidas que pueden ayudar a mitigarlo.

Quiroz (2021) señala en su investigación factores asociados al riesgo en la deserción estudiantil en un Instituto de Educación Superior Tecnológico Público Señala que la metodología CRISM-DM utiliza técnicas de minería de datos para estimar la tasa de abandono escolar de los alumnos de enseñanza básica normal. Para esta investigación se utilizaron registros de la base de datos de estudiantes de la Unidad de Gestión Educativa Local (UGEL) de Chiclayo que abarcan los años 2018 al 2020, lo que requirió un procesamiento previo. Se obtuvo un mayor grado de confianza al comparar las técnicas de Redes Neuronales y ETS. El modelo predictivo utilizado en el estudio arroja un resultado en menos de un minuto, demostrando que sus indicadores lograron identificar sus objetivos. Anexando los conceptos anteriores, se define como datos a la información que ingresa al modelo predictivo de manera verídica y puntual. La información de datos que se logró obtener conlleva a que se consoliden y ordenen atravesando por recursos que permitan ser útiles dentro de los procesamientos como algoritmos para aprendizaje.

García Del Río y López Contreras (2018) en su investigación implementación de herramientas de extracción, transformación y carga de datos estructurados en

big data teniendo como objetivo que en un contexto de big data, Utilización de las herramientas Sqoop y Hive para la carga, transformación y extracción de datos estructurados, haciendo uso de la metodología cascada porque el proyecto se basa en métodos de ensayo y error. completó en 19 segundos importando un total de 1,380 registros, así como también el procedimiento tardó 21 segundos en completarse, durante los cuales se importaron 15,957 registros en total. Al final, la importación de 22.170 registros tardó 13,9875 segundos. Una vez finalizado el proyecto, fue posible demostrar que las dos fuentes de datos (SQL y CSV) analizadas con las herramientas, arrojaron resultados idénticos en cuanto a estructura e integridad de los datos.

(Morales Hernández et al. 2022) En su investigación algoritmos de aprendizaje automático para la predicción del logro académico, tuvo como objetivo predecir el rendimiento académico, se utilizaron dos clasificadores supervisados de aprendizaje automático: una red neuronal multicapa y un algoritmo de refuerzo de gradiente. Se entregaron cuestionarios a una muestra de alumnos, padres, profesores y directores de los centros que cumplían los requisitos de las pruebas, teniendo una población de miles de habitantes, teniendo como resultado de la investigación que se encontró que los puntajes en español y matemáticas fueron las variables más importantes para predecir el logro académico. Otras variables como sexo, tipo de escuela y ubicación geográfica también mostraron importancia en la clasificación, la conclusión fue que, a la hora de predecir los niveles de rendimiento académico, los clasificadores de aprendizaje automático multicapa perceptrón (MLP) y gradient boosting (GB) obtuvieron resultados comparables en cuanto a la precisión global de la clasificación (PG).

Para Quiroga y Lanzarini (2019), en su artículo aprendizaje automático. Aplicaciones en reconocimiento de gestos, acciones y señas, cuyo El objetivo principal de este enfoque es crear una transferencia tecnológica tangible al lenguaje de signos ampliando el clasificador dinámico de gestos creado en la tesis doctoral. Así como la muestra consistió en imágenes y videos que contienen señas, acciones y gestos realizados por personas. La población abarcó una variedad de ejemplos para cada subproblema, incluyendo lenguaje de señas, acciones cotidianas y gestos específicos para un dominio particular. Los

resultados esperados incluyen avances en el reconocimiento de señas, acciones y gestos mediante el uso de Aprendizaje Profundo, con énfasis en Redes Convolucionales Profundas. Teniendo en conclusión esta radica en la aplicación efectiva de técnicas de Aprendizaje Profundo, como Redes Convolucionales Profundas, para abordar problemas específicos de reconocimiento en imágenes y videos.

Para Valenzuela (2022) en su investigación Aprendizaje Supervisado: Métodos, Propiedades y Aplicaciones, teniendo como objetivo principal de este trabajo ha sido extender los fundamentos teóricos de la teoría del aprendizaje estadístico - en la que se basan las Máquinas de Vectores Soporte- a los modelos SVR y SVDD, analizando sus fundamentos matemáticos, enfocando en el desarrollo de modelos matemáticos que pueden predecir valores equivalentes de nuevos datos a partir de conjuntos de datos conocidos. En este trabajo, se considera construir un modelo de este tipo.

Según Nakano et al. (2021), en su investigación modelo de aprendizaje no supervisado para la priorización de inventarios cíclicos, cuyo objetivo es crear y evaluar un modelo de asignación rápida de productos para llevar a cabo un enfoque de recuento de ciclos, la metodología utilizada fue CRISP-DM, Se eligió una muestra de 17.007 transacciones de inventario para que sirviera de recopilación inicial de datos del proyecto. Se identificó la prevalencia de discrepancias en los sistemas de información de inventarios y se destacó la importancia de abordar estas discrepancias de manera temprana, se concluye que la detección temprana de discrepancias en inventarios es crucial para evitar consecuencias negativas en las operaciones de bodega y en la relación con los clientes.

Norambuena, et al. (2022) Nos dice en su investigación modelos predictivos basados en uso de analíticas de aprendizaje en educación superior: una revisión sistemática, teniendo como objetivo examinar las características de los modelos de predicción basados en el análisis del aprendizaje en la enseñanza superior. Para estudiar y seleccionar las publicaciones, se siguieron los criterios mundiales PRISMA para la técnica de revisión sistemática. Hasta 500 participantes de los

campos de la ingeniería y la tecnología conformaron las muestras que fueron seleccionados principalmente en Europa, Los resultados mostraron que todas las investigaciones, utilizando como predictores características analíticas, sociodemográficas y sociocognitivas, intentaron pronosticar el éxito académico en el 100% de los casos. Como conclusión se tiene que la mayoría produjeron grandes modelos de predicción ( $R^2 > .30$ ). Se determina que existen importantes deficiencias en la forma en que se construyen actualmente los modelos de predicción del abandono universitario.

Jhong Guillen (2019) en análisis comparativo de las técnicas de minería de datos para la estimación de consumos de energía eléctrica en la empresa Electronorte S.A., teniendo que el objetivo es realizar un examen comparativo de los métodos de minería de datos para la estimación del consumo eléctrico de ElectroNorte S.A., usando como metodologías XP, SCRUM, CRISP-DM, SEMMA, teniendo como muestra las jerarquías de la población de 326.383 suministros, que se desglosan por empresa, unidad de negocio y ciclo de facturación, dado que los esquemas tradicionales HOLTWINTERS y ARIMA tienen limitaciones basadas en el tipo de series temporales y que es un reto maximizar los cálculos realizados en función de sus coeficientes, el resultado es intrigante. Teniendo como conclusión que la creación de una aplicación web que permite realizar simulaciones y evaluar el comportamiento de distintas técnicas, proporcionando una herramienta práctica basada en los hallazgos de la investigación.

Según Contreras y Sánchez (2020), nos dice en su investigación llamada analítica predictiva para conocer el patrón de consumo de los clientes en la Empresa Cienpharma S.A.C. Los registros de ventas de 2014-2015 de la empresa sirvieron de muestra, y el modelo de minería de datos mediante algoritmos de asociación, clústeres y redes neuronales se aplicó a la población de registros de ventas de la base de datos transaccional de la empresa. El modelo de minería de datos se ejecutó utilizando IBM SPSS Modeler y la metodología CRISP-DM, muestran patrones de consumo identificados en la base de datos transaccional de la empresa. Teniendo como conclusión que la implementación de la analítica predictiva a través de técnicas de minería de

datos permitió identificar patrones de consumo en Cienpharma S.A.C., brindando una visión más precisa y detallada de las tendencias de los clientes.

Sotomayor y Rodriguez (2018) en su artículo Factores explicativos de la deserción académica en la Educación Superior Técnico Profesional: el caso de un centro de formación técnica, el objetivo era examinar los factores que contribuyen al abandono académico en el sector de la enseñanza técnica. En el marco de un método cualitativo, se propone un estudio de caso acorde con el objetivo especificado, teniendo como muestra a Directivos, académicos y estudiantes del centro de formación técnica chileno seleccionado como caso de estudio. Con una población de estudiantes de la Educación Superior Técnico Profesional en Chile, con énfasis en aquellos asociados al centro de formación técnica estudiado. Teniendo como resultado la identificación de aspectos contextuales, académicos, organizativos, socioeconómicos, familiares y personales que explican la deserción académica. Llegando a la conclusión que el fenómeno de la deserción académica en la Educación Superior Técnico Profesional es complejo y multidimensional, con factores que van más allá de lo académico. Las conclusiones del estudio ofrecen directrices para abordar este problema, proponiendo líneas de acción para reducir las tasas de deserción.

Yeboah, Otoo y Abaye (2020) en su artículo principios básicos de los algoritmos de clasificación K-Nearest-Neighbor, Support Vector Machine, Random Forest y Neural Network: Una revisión, con el objetivo de examinar y comparar estos métodos de clasificación no paramétricos en los siguientes atributos: robustez de los datos de entrenamiento, sensibilidad a los cambios, ajuste de datos, estabilidad, capacidad para manejar datos de gran tamaño, sensibilidad al ruido, tiempo invertido en el ajuste de parámetros y precisión. En el embolsado, cada árbol de decisión se crea individualmente utilizando una muestra bootstrap del conjunto de datos, y los árboles posteriores no dependen de los anteriores. De una población de artículos de investigación y libros que abordaron el uso de algoritmos de clasificación no paramétricos en el período de 2000 a 2017. En los resultados se observó que RF es sensible a cambios pequeños en el conjunto de datos de entrenamiento y a veces es inestable, mostrando tendencia al sobreajuste. KNN es fácil de implementar y entender, pero se vuelve

significativamente lento a medida que crece el tamaño de los datos, y la elección del valor ideal de K es difícil. SVM y RF son insensibles al ruido o sobreentrenamiento, mostrando habilidad para manejar datos desequilibrados. Se concluye que, a pesar del potencial de NN como algoritmo de clasificación, debido a su procedimiento de ajuste de parámetros consumidor de tiempo, alta complejidad en el procesamiento computacional, la variedad de arquitecturas de NN disponibles y los distintos algoritmos utilizados para el entrenamiento, la mayoría de los investigadores recomiendan SVM y RF como métodos más fáciles y manejables que logran resultados con altas precisiones y son a menudo más rápidos de implementar.

Robles-Velasco et al. (2020) en su investigación aplicación de la regresión logística para la predicción de roturas de tuberías en redes de abastecimiento de agua, teniendo como objetivo reducir el número de roturas inesperadas en redes de abastecimiento de agua mediante el desarrollo de un método predictivo. La muestra, compuesta por un histórico de siete años, se divide aleatoriamente en dos partes, cinco años para el entrenamiento del modelo, y dos para su evaluación. La red está formada por más de 3,800 kilómetros de tuberías y abastece a una población de más de 1 millón de personas. Se logró predecir el 85.9% de las roturas de tuberías utilizando la metodología propuesta, con una precisión total del modelo del 76.6%. El método predictivo desarrollado, basado en la regresión logística y técnicas avanzadas de procesamiento de datos, demostró ser efectivo al predecir un alto porcentaje de roturas de tuberías.

Wang et al. (2020) Se habla en esta investigación Big Data Service Architecture: Visión general, marco de procesamiento e integración de la computación en nube, teniendo como objetivo explorar la aplicación de la tecnología de big data en minería de datos, análisis de datos y compartición de datos, La población que se aborda es la infraestructura y los servicios basados en big data, aplicables a diversas áreas. Teniendo también que en los resultados se presentan las características y capacidades de la arquitectura de servicios de big data, destacando su papel en la extracción de valor potencial de los datos y su contribución al beneficio económico. Lo que llega a la conclusión de que el artículo subraya la importancia y versatilidad de la arquitectura de servicios de

big data, enfatizando su capacidad para proporcionar soluciones personalizadas de procesamiento y análisis de datos para consumidores de servicios.

Zhu et al. (2019) en su investigación Big data y su impacto en los Sistemas de Transporte Inteligente (STI), con el objetivo Examinar la relación entre big data y los Sistemas de Transporte Inteligente (STI), destacando su influencia en el diseño y la aplicación de estos sistemas para mejorar la seguridad, eficiencia y rentabilidad. Se lleva a cabo una revisión exhaustiva que abarca la historia y las características tanto de big data como de los STI. Se exploran diversos estudios de caso que ilustran aplicaciones de análisis de big data en STI, abordando temas como el análisis de accidentes de tráfico, la predicción del flujo de tráfico, la planificación de servicios de transporte público, la planificación de rutas de viaje personal, la gestión y control del transporte ferroviario, y el mantenimiento de activos. Los resultados destacan la variedad de aplicaciones de big data en STI, demostrando su capacidad para mejorar diversos aspectos del sistema. Llegando a la conclusión de que el estudio subraya la importancia de big data en la evolución de los STI y plantea desafíos abiertos para el uso continuo de análisis de big data en este contexto, proporcionando una visión integral de su impacto actual y potencial.

Li, Higgins y Deeks (2019) en su artículo Enfoques y Métodos de Recopilación de Datos en Revisiones Sistemáticas: Directrices y Consideraciones, teniendo como objetivo el describir las metodologías y enfoques apropiados para la recopilación de datos en revisiones sistemáticas, abordándose enfoques específicos para la recopilación de datos en revisiones sistemáticas. La población de interés son los estudios incluidos en revisiones sistemáticas. La muestra se compone de datos extraídos de artículos de revistas y otros informes relevantes. Se resalta la importancia de recopilar datos que describan adecuadamente los estudios incluidos, respalden la creación de tablas y figuras, faciliten la evaluación del riesgo de sesgo y permitan la síntesis y metaanálisis. La conclusión principal destaca la necesidad de utilizar enfoques estructurados y seguir pautas de informes específicas para garantizar la calidad y relevancia de los datos recopilados en revisiones sistemáticas.

Tran et al. (2022), En este trabajo llamado evaluación de Modelos de Aprendizaje Automático para la Predicción del Comportamiento Térmico y Eléctrico de Celdas de Batería de Iones de Litio, teniendo como objetivo Investigar el rendimiento de cuatro modelos de aprendizaje automático utilizados para predecir los comportamientos térmicos (temperatura) y eléctricos (voltaje) de celdas de batería de iones de litio (Li-ion). Desarrollándose cuatro modelos de regresión de aprendizaje automático (regresión lineal, k-vecinos más cercanos, bosque aleatorio y árbol de decisiones) utilizando la biblioteca scikit-learn en Python. Los modelos se validaron con datos experimentales y se compararon mediante la métrica R<sup>2</sup>. La muestra se compone de datos experimentales recopilados de una celda de batería Li-ion prismática con una capacidad de 25 Ah, ciclada bajo condiciones específicas. Los resultados del rendimiento de los cuatro modelos de aprendizaje automático se informaron y compararon utilizando la métrica R<sup>2</sup>. El modelo basado en árbol de decisiones, con un puntaje R<sup>2</sup> de 0.99, se determinó como el mejor modelo en este estudio de caso. Llegando a la conclusión que se destaca que el modelo basado en árbol de decisiones demostró ser altamente preciso en la predicción del comportamiento térmico y eléctrico de las celdas de batería Li-ion, sugiriendo su idoneidad para mejorar la eficiencia y confiabilidad de los sistemas de gestión de baterías en el contexto de vehículos eléctricos.

Patil y Kulkarni (2019) en su investigación predicción de la precisión del árbol de decisión distribuido mediante aprendizaje automático, Teniendo como objetivo el mejorar el rendimiento del tiempo de construcción del modelo en un entorno distribuido para el algoritmo de Árbol de Decisiones. Se propone una versión mejorada del algoritmo de Árbol de Decisiones distribuido para lograr un tiempo de construcción de modelo más eficiente. La muestra consiste en conjuntos de datos grandes en los que se aplicará el algoritmo de Árbol de Decisiones distribuido mejorado para evaluar su rendimiento en términos de tiempo de construcción del modelo. Los resultados se centran en la eficiencia del tiempo de construcción del modelo logrado por la versión mejorada del algoritmo de Árbol de Decisiones distribuido. La conclusión destaca que la versión mejorada del algoritmo de Árbol de Decisiones distribuido logra un rendimiento más

eficiente en el tiempo de construcción del modelo para conjuntos de datos grandes, sin comprometer la precisión del Árbol de Decisiones.

Schröer, Kruse y Gómez (2021) en su investigación revisión sistemática de la literatura sobre la aplicación del modelo de procesos CRISP-DM, el objetivo de esta revisión bibliográfica sistemática es identificar el enfoque de la investigación, las mejores prácticas y los nuevos métodos para aplicar las fases CRISP-DM. Se emplea una revisión sistemática para analizar estudios publicados en IEEE, ScienceDirect y ACM. La muestra consiste en estudios recientes de minería de datos que aplican el modelo CRISP-DM, seleccionados de fuentes como IEEE, ScienceDirect y ACM. Los resultados muestran que varios proyectos de minería de datos desde 2017 se basan en CRISP-DM

Martínez-Plumed et al. (2021), en su investigación CRISP-DM veinte años después: De los procesos de minería de datos a las trayectorias de la ciencia de datos, tuvo un objetivo de investigar si CRISP-DM (CRoss-Industry Standard Process for Data Mining), Se utiliza un enfoque de investigación que analiza la vigencia de CRISP-DM en proyectos contemporáneos de ciencia de datos. La muestra consiste en siete ejemplos del mundo real donde las actividades exploratorias desempeñan un papel crucial, además de 51 casos de uso extraídos del NIST Big Data Public Working Group. Se analizan proyectos que representan tanto actividades dirigidas por objetivos como exploratorias. Los resultados revelan que, en proyectos de ciencia de datos dirigidos por objetivos y procesos, CRISP-DM sigue siendo relevante. La conclusión más significativa es que CRISP-DM sigue siendo efectivo en proyectos de ciencia de datos dirigidos por objetivos y procesos.

Cazacu y Titan (2021), en su investigación adaptación de CRISP-DM a las Ciencias Sociales, teniendo como objetivo introducir la metodología CRISP-DM como un estándar para analizar grandes volúmenes de datos no estructurados. Se emplea un enfoque que aplica la metodología CRISP-DM en proyectos de descubrimiento de conocimiento para ciencias sociales. La muestra consiste en proyectos de descubrimiento de conocimiento en ciencias sociales que enfrentan desafíos específicos relacionados con la complejidad inherente y las

características únicas de este campo. La población abarca el conjunto de proyectos en ciencias sociales que buscan extraer conocimientos útiles e implícitos de grandes volúmenes de datos no estructurados. Los resultados destacan que la aplicación de la metodología CRISP-DM puede estandarizar y mejorar la eficiencia del proceso de preparación de datos en proyectos de ciencias sociales. La conclusión más relevante es que la implementación de la metodología CRISP-DM en proyectos de ciencias sociales puede contribuir significativamente a estandarizar el proceso de preparación de datos, reducir el tiempo dedicado a esta fase y mejorar la calidad general de la investigación en este campo.

Mackieson, Shlonsky y Connolly (2019), en su investigación aumento del rigor y reducción del sesgo en la investigación cualitativa: un análisis de documentos de debates parlamentarios utilizando análisis temático aplicado, teniendo como objetivo del artículo el describir cómo se operacionalizó el análisis temático aplicado en una investigación cualitativa que empleó un análisis de documentos oficiales. La metodología utilizada es el análisis temático aplicado, que proporciona estructura e integra la reflexividad en la investigación cualitativa basada en datos textuales. La muestra consiste en los registros oficiales de debates parlamentarios en el Estado de Victoria, Australia, específicamente relacionados con la introducción de Órdenes de Cuidado Permanente. La población abarca los documentos oficiales relevantes a esta temática. Los resultados destacan la aplicabilidad exitosa del análisis temático aplicado en el contexto de un análisis de documentos oficiales. La conclusión más importante es que el análisis temático aplicado demuestra ser una herramienta valiosa para fortalecer la investigación cualitativa en trabajo social al utilizar el análisis de documentos oficiales.

Asimismo, Pierrakeas et al. (2020) en su investigación Variabilidad de los motivos de abandono en la enseñanza a distancia y predicción de alumnos propensos al abandono, teniendo como objetivo el identificar un algoritmo de aprendizaje automático efectivo y las características más informativas para predecir la deserción de estudiantes en programas de educación de adultos a distancia. habla acerca de limitar el abandono es crucial, ya que teniendo la

capacidad para predecir el abandono de los estudiantes sería muy útil, así como el uso de algoritmos de aprendizaje automático para predecir la deserción estudiantil. Teniendo una muestra que consiste en datos recopilados de estudiantes matriculados en la Hellenic Open University. La población abarca estudiantes de programas de educación de adultos a distancia. Teniendo resultados revelan que los atributos más informativos para predecir la deserción incluyen género del estudiante, participación en la primera reunión presencial y calificaciones en las primeras dos asignaciones escritas. Llegando a la conclusión que la identificación de atributos clave y un algoritmo efectivo para predecir la deserción estudiantil en programas de educación de adultos a distancia.

Según Berens et al. (2018) en su investigación denominada Detección precoz de estudiantes en situación de riesgo: predicción del abandono escolar mediante datos administrativos de los estudiantes y métodos de aprendizaje automático, la cual tiene como objetivo el desarrollar un Sistema de Detección Temprana (EDS) utilizando datos administrativos de estudiantes de una universidad estatal y una universidad privada. Esta investigación hace uso de análisis predictivo, específicamente regresión, redes neuronales, árboles de decisión y el algoritmo AdaBoost. Teniendo una muestra de alrededor de 6,700 estudiantes y 26 programas de licenciatura. Con una población aproximada de 23,000 estudiantes y 90 programas de licenciatura, y una universidad privada de ciencias aplicadas (PUAS) utilizándose un conjunto de datos de entrenamiento de exalumnos que se matricularon entre 2007 y 2017, teniendo como resultado que la precisión de la predicción al final del primer semestre es del 79% para la universidad estatal y del 85% para la universidad privada de ciencias aplicadas. Después del cuarto semestre, la precisión mejora al 90% para la universidad estatal y al 95% para la universidad privada de ciencias aplicadas. La conclusión que tiene es el desarrollo exitoso de un Sistema de Detección Temprana que utiliza análisis predictivo para predecir el éxito estudiantil y distinguir a los estudiantes en riesgo de deserción.

### **III. METODOLOGÍA**

#### **3.1. Tipo y diseño de investigación**

##### **3.1.1. Tipo de investigación**

Esta investigación se clasifica como de tipo aplicada, ya que tiene como objetivo principal identificar y comprender los patrones que contribuyen a la deserción estudiantil desarrollando un modelo predictivo en un Instituto de Educación Superior Tecnológica Pública (IESTP) de Lima, con el fin de desarrollar un modelo predictivo efectivo. Este tipo de investigación utiliza conocimientos aplicados con el fin de detectar un tema en específico. (CONCYTEC, 2018).

Dado que se utilizaron datos numéricos, se adoptó un diseño de investigación cuantitativo, con el fin de estructurar y organizar la investigación (Hernández Sampieri, 2018). Por ello, la investigación se midió mediante una guía de observación, un formulario de registro y entrevistas.

##### **3.1.2. Diseño o Método de Investigación**

El diseño de investigación seleccionado para este estudio es descriptivo - experimental. Se utilizaron como técnicas el cuestionario que tiene como instrumento la encuesta, la revisión de documentos y técnicas como el análisis documental y la observación de campo, que tienen como instrumento la ficha de registro y guía de observación para identificar de manera detallada los patrones que influyeron en la deserción estudiantil desarrollando un modelo predictivo en un IESTP. Este diseño permitirá conocer a fondo las opiniones y experiencias de los actores involucrados en el proceso académico. (Luis Codina 2023).

#### **3.2. Variables y operacionalización**

La variable independiente del estudio fue la deserción estudiantil la cual tiene como dimensión Factores influyentes en la deserción, que consta de 4 indicadores, personales, académicos, sociales y económicos.

Por otra parte, como variable dependiente se tuvo modelo predictivo el cual tuvo como primera y única dimensión la eficiencia, que consta del indicador métricas haciendo referencia a precisión, f1, score y recall.

### **3.3 Población, Muestra y Muestreo**

#### **3.3.1 Población**

Dar a los recursos que necesitan los estudiantes para seguir carreras de investigación y ser competentes es el principal objetivo del curso Metodología de la Investigación I. Se hace hincapié en la importancia de los temas tratados que se trataron como fundamentos esenciales. Al ayudar a los estudiantes a través de los procesos metodológicos en las etapas preliminares de la técnica, el objetivo principal es capacitarlos para diseñar el objeto de estudio de su investigación. Según Cordoba et al. (2023), los participantes deben ser capaces de identificar los rasgos distintivos de cada enfoque metodológico, formular conclusiones defendibles y justificar de forma teórica sus decisiones.

En la población fueron incluidos los registros de los estudiantes inscritos en la carrera de Computación e Informática en 2023, así como cualquier información respaldada y subida en el sistema MIDENU. Silva et al. (2018) afirman que los criterios legítimos de los participantes que desean formar parte del estudio rigen la inclusión.

Por el contrario, la exclusión ocurre cuando los datos recolectados no guardan relación o no se relacionan con el tema de investigación que ha sido sugerido, tal como lo afirman Silva et al. (2018). De esta manera, se omitieron números telefónicos, casos de estudiantes exonerados y datos de años anteriores. El proceso de inclusión y exclusión garantiza que la muestra recolectada sea consistente y lo suficientemente relevante para cumplir con los objetivos de la investigación.

#### **3.3.4. Unidad de análisis**

Según Hernández Sampieri (2018), la forma de medir a los sujetos de la investigación y su contribución al estudio determinan la unidad de

análisis. Para esta investigación comprenden: Los reportes de la carrera de computación e informática del año 2023.

### 3.4. Técnicas e Instrumentos de Recolección de Datos

Esta herramienta, la técnica de recolección de datos, funciones de recogida de datos relacionados con el tema que se estudiará en el estudio; consiste en extraer información que hagan fácil el correcto desarrollo del proyecto. (Cisneros et al., 2022)

En este estudio, la unidad de análisis comprendió los registros individuales de los estudiantes matriculados en la unidad académica de Computación e Informática durante los períodos académicos 2022-1 y 2022-2. Estos datos se extrajeron y examinaron en detalle, proporcionando la base para el análisis y la identificación de patrones de deserción estudiantil.

**Tabla 1: Unidad de análisis**

Año	Unidad de análisis	Total
2022-1 / 2022-2	363 registros	363

Elaboración Propia de los autores.

Encuestas: Aquí se generan un conjunto de preguntas cerradas respecto a una o más variables a medir, se pueden dar por teléfono, correo, online y personalmente (Martínez 2022).

Revisión documental: Esta metodología se utiliza para elegir y recopilar información acerca de la variable desde diferentes enfoques, lo que posibilita una exploración más completa de los conocimientos relacionados con el tema y la variable, abordándolos desde la perspectiva de integración, confirmación y evaluación crítica. (Useche *et al.* 2023).

Ficha de registro: Establece que el fichero sirve para reunir y conservar datos, lo que, además de darle una extensión, le proporciona un valor y una unidad. (Maturrano 2021)

Guía de observación: Se define la observación como la descripción organizada de eventos, comportamientos y elementos en el entorno social seleccionado para el estudio. La observación capacita al observador para detallar situaciones mediante el uso de los cinco sentidos, generando una representación escrita que refleja la situación en análisis. (Martínez 2022). Esta guía permite obtener datos del modelo como: el tiempo que se tarda en entrenar y el porcentaje de rendimiento de precisión.

situación en análisis. (Martínez 2022). Esta guía permite obtener datos del modelo como: el tiempo que se tarda en entrenar y el porcentaje de rendimiento de precisión.

**Tabla 2: Instrumentos de recolección de datos para la variable: Deserción estudiantil**

Nº	INDICADOR		INSTRUMENTO
1	Factores académicos	Turno	Ficha de registro
2		Sección	
3		Curso	
4		Semestre	
5		Nota final del curso	
6		Porcentaje de inasistencias por curso	
7		Promedio inasistencias	
8		Promedio ponderado	
9	Factores	Sexo	Encuesta

10	Distrito de Residencia
11	Acceso a computadora
12	Acceso a internet
13	Apoyo familiar
14	Ingresos familiares
15	Trabajo parcial
16	Becas ayuda
17	Tipo familia
18	Carga familiar
19	Discapacidad enfermedad
20	Ejercicio regular
21	Hábitos alimenticios
22	Actividades sociales
23	Consumo de alcohol
24	Consumo de cigarrillos
25	Desertaron

Elaboración propia de los autores

**Tabla 3: Instrumentos de recolección de datos para la variable: Modelo**

## Predictivo

Nº	INDICADOR	INSTRUMENTO	
1	Precisión del modelo	Presición	Guía de observación
		Rcall	
		F1 Score	
		Instancias correctas	
		Instancias Incorrectas	

Elaboración propia de los autores

### 3.5. Procedimientos

Para el desarrollo de esta investigación se inició por identificar el problema general, así como también identificar los factores más influyentes y disponibles con los que se contaban en la deserción estudiantil en un IESTP, mediante una entrevista con los participantes adecuados y las preguntas específicas, donde se abarcaron conceptos clave relacionados a la deserción, es importante destacar que se considera deserción a los elementos que se producen tanto dentro del sistema educativo como en los contextos social, familiar, personal y ambiental conducen al abandono escolar. Minedu (2020)

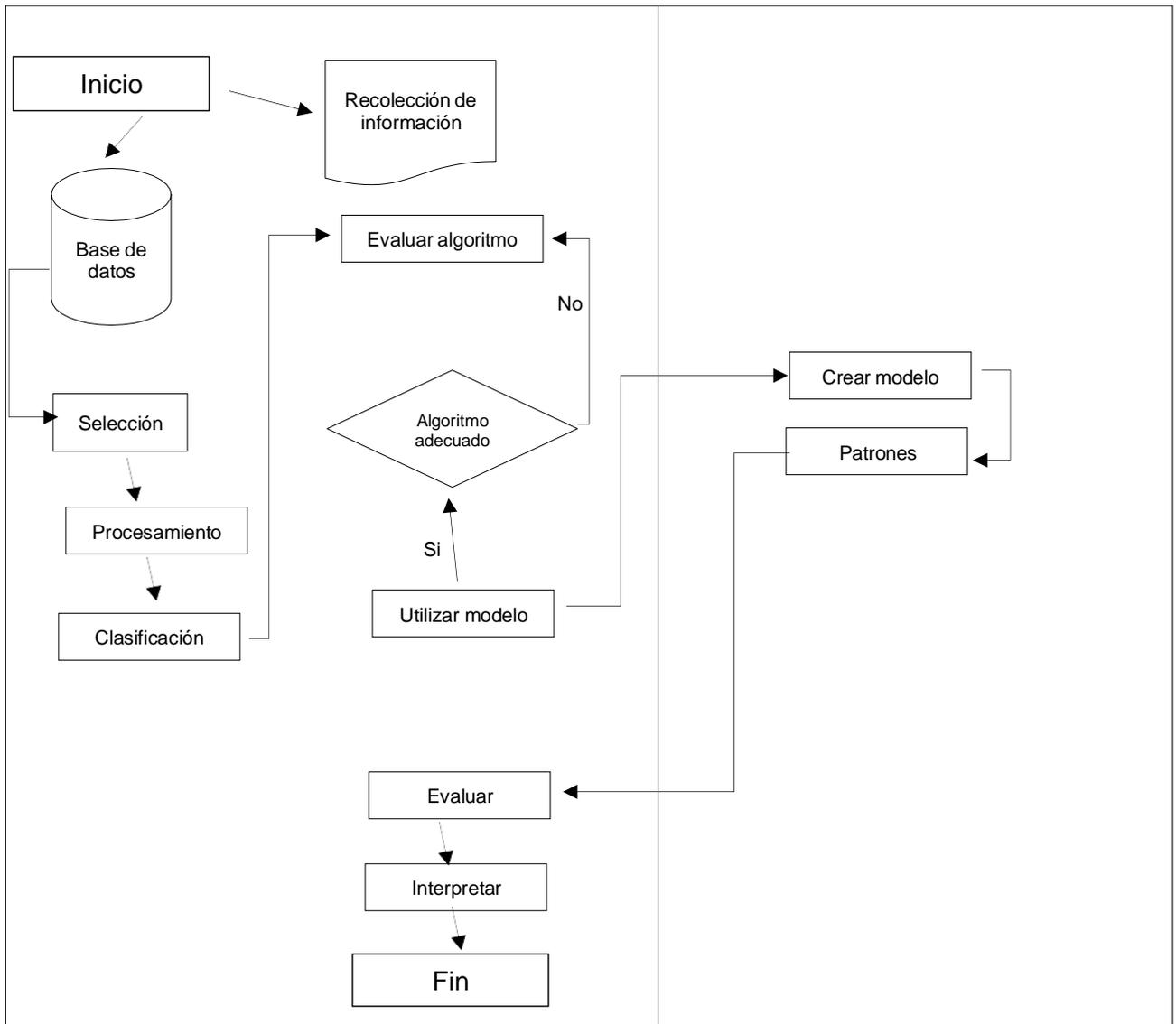
En segunda instancia se definieron las técnicas e instrumentos a utilizar considerando los datos académicos, también para medir la eficiencia del modelo se utilizaron los datos de las métricas del modelo, para ello se hizo uso de los instrumentos como la encuesta, la revisión, la revisión documental, la ficha de registro y guías de observación.

Una vez diseñados y validados los instrumentos de recolección de datos, se procedió a la recopilación de información de los estudiantes de la unidad académica de Computación e Informática durante los periodos académicos

2022-1 y 2022-2. Posteriormente, se llevó a cabo un análisis exhaustivo de los resultados mediante técnicas avanzadas de minería de datos. La metodología aplicada, basada en CRISP-DM (Cross Industry Standard Process for Data Mining), guió cada fase del proceso, desde la comprensión del negocio hasta la implementación del modelo predictivo. La regresión logística (RLO) se empleó como técnica central en la elaboración del modelo predictivo, permitiendo explorar las relaciones entre variables y evaluar su impacto en la deserción estudiantil. El Gráfico N° 1 ilustra visualmente el diagrama de actividades del proceso de investigación, destacando las etapas clave desde la construcción de instrumentos hasta la generación del modelo predictivo. Este método integrado garantizó un análisis minucioso y una interpretación detallada de los patrones identificados. De esta manera, se logró obtener una comprensión más completa de los diversos factores que ejercen influencia en la deserción estudiantil en la unidad académica mencionada.

**Ilustración 1: Diagrama de actividades del proceso de investigación**

Experto	Técnica
---------	---------



Elaboración propia de los autores

### 3.6. Método de análisis de datos

Se realizan operaciones en la que los objetivos del estudio son alcanzados por los investigadores, proporcionando función a las hipótesis que se desarrollaron. En este punto, los datos se presentan de forma oral, auditiva o escrita como textos en las notas, documentos y entrevistas. Esta información debe ser detectable, rastreable y accesible. (Borjas, J 2020)

Se recolectaron datos a través de los instrumentos clave para ubicar las fuentes de la fase de extracción como los registros de los alumnos, se examinaron los datos en Microsoft Excel de distintos registros brindados por la institución.

### **3.7. Aspectos éticos**

La validez y la fiabilidad son componentes fundamentales que definen la exactitud de las puntuaciones y los datos recogidos mediante instrumentos aplicados en el estricto contexto de la investigación científica (Medina y Carrión, 2020).

En este contexto se abordan varias cuestiones importantes, todas ellas sustentadas en ideas fundamentales. El primer componente, la responsabilidad, es una dedicación sin concesiones al avance y uso de la investigación. Según el Código Ético, la validez es un elemento crucial que garantiza la exactitud y coherencia de los resultados.

La utilización de herramientas de resultados de similitud -en este ejemplo, Turnitin- se propugna como medio de preservar la originalidad y evitar el plagio, defendiendo así los derechos de propiedad intelectual de estudios anteriores. Del mismo modo, se incluyen las normas ISO como requisito adicional para mantener los estándares de precisión y calidad. Uno de los conceptos rectores es la autonomía, que da a cada estudiante la libertad de elegir si participa o no en la investigación

La aplicación de la recolección de los datos de las personas que satisfacen las necesidades de la investigación pone de relieve la competencia profesional y científica. Como principio rector, se plantea la transparencia en la obtención de resultados, garantizando su autenticidad y verificabilidad.

El objetivo desinteresado de la investigación de beneficiar tanto a los participantes como a la comunidad en general. Un compromiso ético que integre es el poder investigar sin poner en peligro el medio ambiente ni la biodiversidad si se tiene cuidado con ambos.

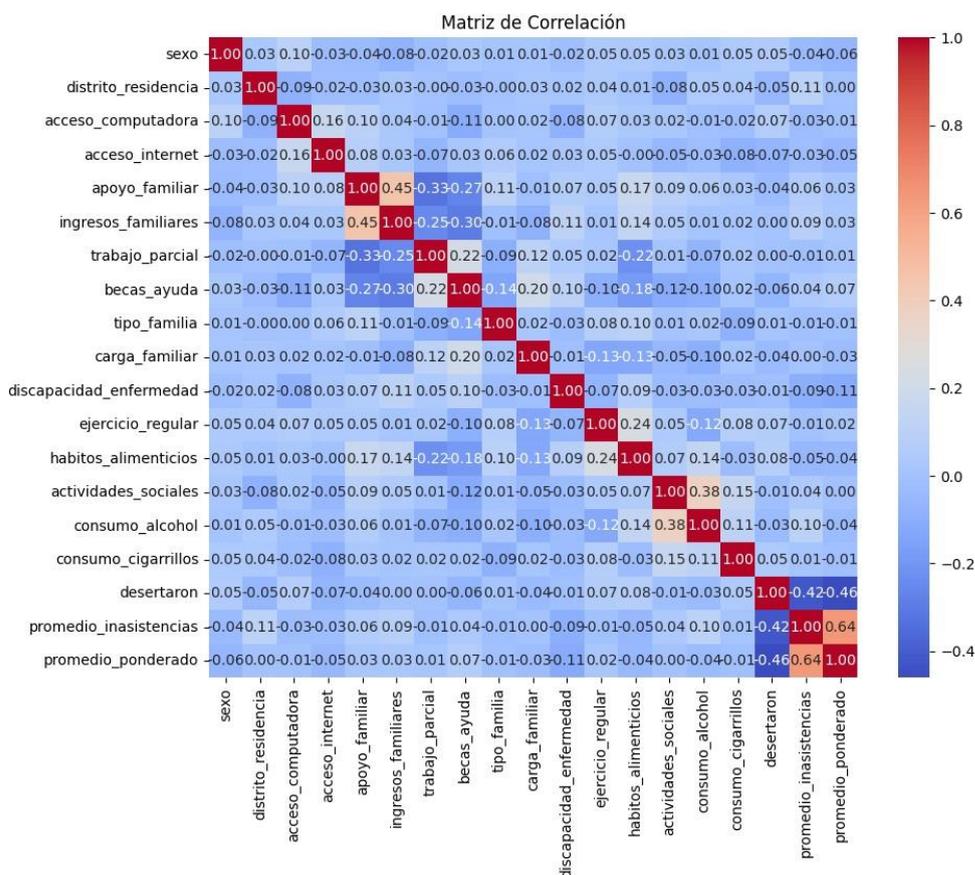
En última instancia, la dignidad humana y la justicia se defienden en todas las etapas, garantizando así que todas las personas implicadas en la investigación recibieron un trato justo y respetuoso. Cuando se combinan, estos principios y facetas proporcionan una sólida base ética que respalda el calibre y la fiabilidad del estudio científico realizado.

#### IV. RESULTADOS

La investigación tuvo como objetivo general Identificar patrones de la deserción estudiantil desarrollando un modelo predictivo en un IESTP - Lima 2023

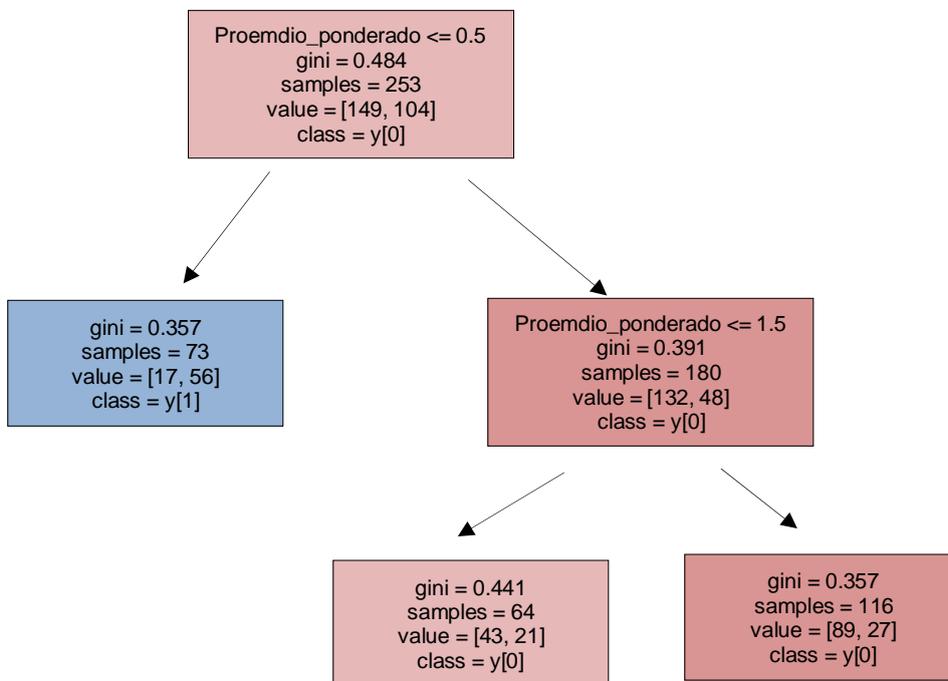
**Objetivo específico 1:** Identificar los factores más influyentes en la deserción estudiantil del programa de Computación e Informática.

#### Ilustración 2: Matriz de correlación entre variables



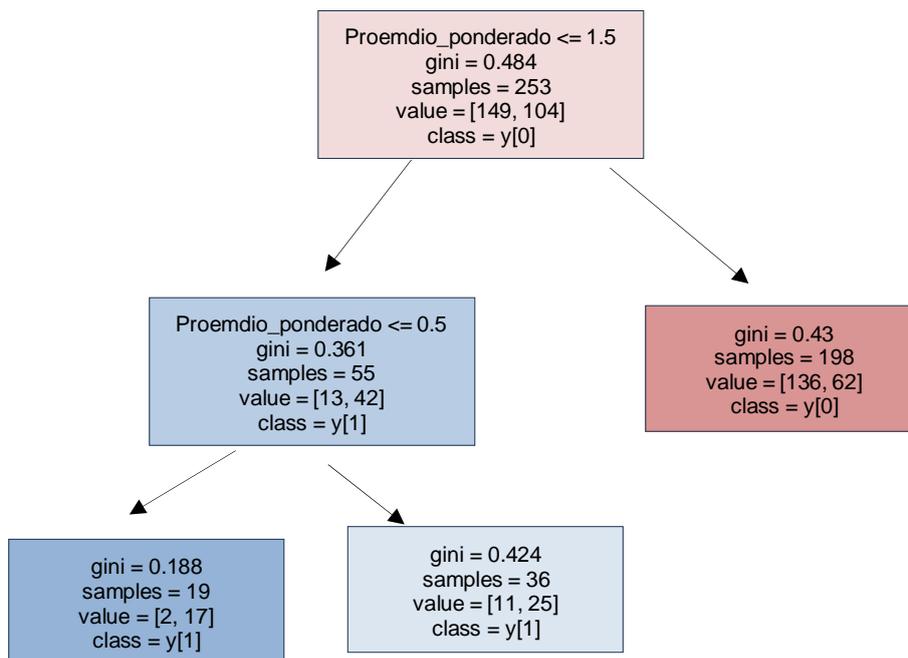
Elaboración Propia de los autores.

#### Ilustración 3: Modelo de la variable Promedio ponderado



Elaboración Propia de los autores.

#### Ilustración 4: Modelo de la variable Promedio inasistencias



Elaboración Propia de los autores.

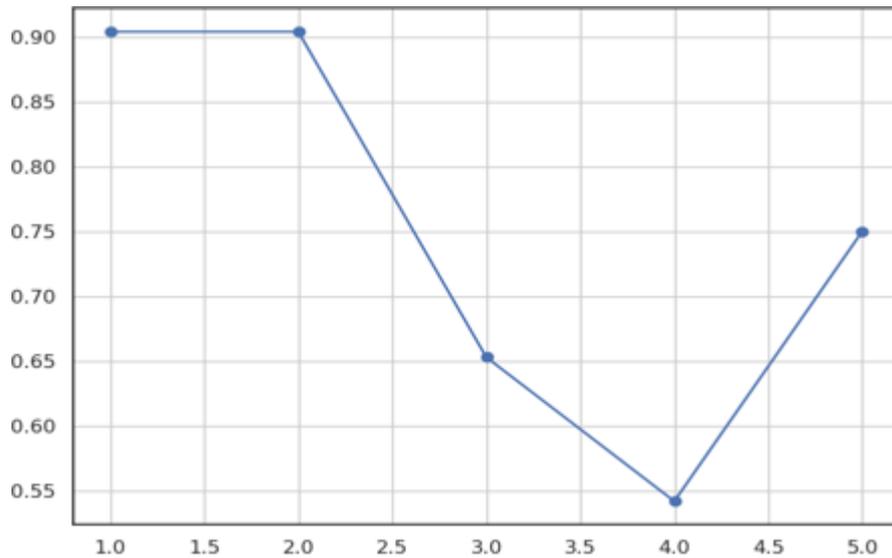
La muestra una matriz de correlación entre las variables revela relaciones significativas como "Hábitos Alimenticios," "Acceso a Computadora," y "Ejercicio Regular" exhibieron correlaciones positivas moderadas con el resultado general, mientras que "Promedio Ponderado" y "Promedio de Inasistencias" mostraron

correlaciones negativas notables. La "Acceso a Internet" tuvo una correlación negativa, aunque más débil. Objetivo Específico 2: Construir y comparar modelos de machine learning para representar patrones de deserción estudiantil.

En el análisis de los modelos, se identificaron "Promedio Ponderado" y "Promedio de Inasistencias" como las variables más influyentes en la predicción del resultado general. La correlación negativa sustancial de estas variables (-0.460446 y -0.421968, respectivamente) indicaron una relación significativa con el abandono. Un bajo promedio ponderado y un alto número de inasistencias están asociados con una mayor probabilidad de abandono. Este hallazgo se refleja en los árboles de decisión generados para estas variables, donde se observan ramas que se dividen en condiciones específicas de promedio ponderado e inasistencias, destacando su importancia en la toma de decisiones del modelo. En contraste, variables como "Hábitos Alimenticios", "Acceso a Computadora", y "Ejercicio Regular" tuvieron correlaciones positivas, pero menos marcadas, sugiriendo una influencia más tenue en comparación con las dos variables mencionadas. La variable "Acceso a Internet" también muestra una correlación negativa, pero su impacto es menos pronunciado. Estos resultados respaldan la elección de "Promedio Ponderado" y "Promedio de Inasistencias" como variables clave para la predicción del abandono, ya que su asociación más fuerte y la representación visual en los árboles de decisión respaldan su papel crucial en la toma de decisiones del modelo.

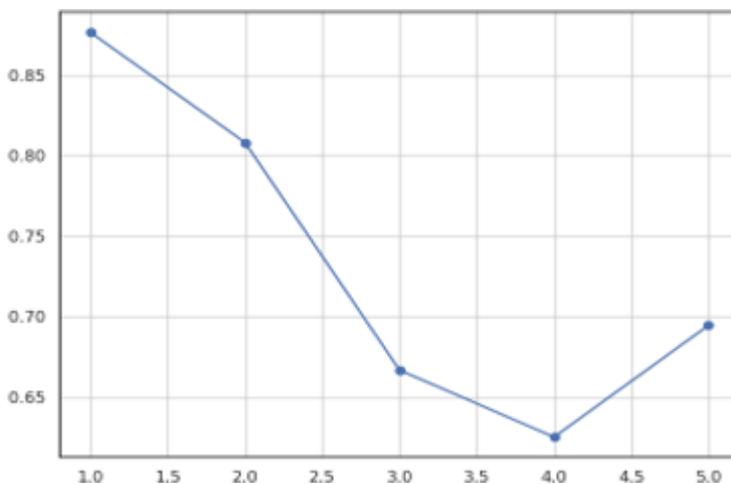
**Objetivo específico 2:** Construir y comparar modelos de machine learning para representar patrones de deserción estudiantil.

**Ilustración 5:** Validación cruzada del modelo de árbol de decisión de la variable promedio ponderado



Elaboración Propia de los autores.

**Ilustración 6:** Validación cruzada del modelo de árbol de decisión de la variable promedio de inasistencias



Elaboración Propia de los autores.

Considerando los resultados de la validación cruzada utilizando árboles de decisión para las variables de promedio ponderado y promedio de inasistencias, se obtuvieron precisiones del 77.06% y 80.73%, respectivamente. En el caso del promedio ponderado, la matriz de confusión revela una correcta predicción de 57 casos de no deserción y 27 casos de deserción, con 14 falsos positivos y 11 falsos

negativos. Por otro lado, para el promedio de inasistencias, se lograron 68 predicciones correctas de no deserción y 20 predicciones correctas de deserción, con 3 falsos positivos y 18 falsos negativos. Estos resultados indican un desempeño positivo de los modelos en la identificación de patrones de deserción estudiantil.

**Objetivo específico 3:** Evaluar la precisión predictiva de los modelos y ajustarlos para mejorar su rendimiento.

Para la evaluación de los modelos predictivos de deserción estudiantil construidos mediante árboles de decisión, se llevó a cabo un análisis detallado de su rendimiento utilizando la técnica de validación cruzada. Este proceso se aplicó para garantizar la robustez y generalización de los modelos, evitando sesgos asociados a particiones específicas de los datos. Los resultados se detallan en la tabla a continuación, donde se presenta el rendimiento porcentual de instancias correctas e incorrectas para dos variables clave: "Promedio Ponderado" y "Promedio de Inasistencias".

**Tabla 4: Evaluación de Clasificación por Variables Predictivas**

Elaboración Propia de los autores.

En la tabla 1 se refleja la capacidad de los modelos para clasificar adecuadamente

Variable	Correctas	Incorrectas	Porcentaje Correctas	Porcentaje Incorrectas	Total (%)
Promedio Ponderado	272	90	75.14	24.86	100
Promedio Inasistencias	266	96	73.48	26.52	100

las instancias en función de las variables consideradas. La tabla muestra que, en promedio, el 75.14% de las instancias relacionadas con "Promedio Ponderado" fueron clasificadas correctamente, mientras que para "Promedio de Inasistencias", el porcentaje de instancias correctamente clasificadas fue del 73.48%. Estos

hallazgos proporcionan una visión cuantitativa del desempeño de los modelos y respaldan la validez de su aplicación en la predicción de patrones de deserción estudiantil.

## V. DISCUSIÓN

Objetivo Específico 1: Identificar los Factores Influyentes en la Deserción Estudiantil del Programa de Computación e Informática.

La necesidad de comprender los elementos que influyen directamente en la retención estudiantil llevó a descubrir y priorizar los factores vinculados a la deserción estudiantil en el programa de Ciencias de la Computación e Informática. En línea con las investigaciones que fueron realizadas por Berens et al. (2018) y Pierrakeas et al. (2020), que enfatizaron en su investigación la importancia de identificar las características críticas, este objetivo es situar estos hallazgos en el contexto particular de la computación y la informática.

Una base sólida para lograr este objetivo es proporcionada por la teoría de la identificación temprana de estudiantes en riesgo por parte de Berens et al. (2018) y el énfasis en la diversidad de razones de abandono de Pierrakeas et al. (2020). Al adaptar estas ideas al programa académico concreto, será posible tanto identificar los elementos relevantes en la causa de abandono como clasificarlos en función al impacto que tienen cada uno.

Objetivo Específico 2: Desarrollar y contrastar modelos de aprendizaje automático con el propósito de identificar y representar patrones asociados a la deserción estudiantil.

La utilización de modelos de aprendizaje automático, como se sugiere en este objetivo, se alinea con los hallazgos de Pedroza, Chasoy y Gómez (2019) y Fernández et al. (2018), que utilizaron técnicas de vanguardia para predecir el abandono escolar. Combinando estos puntos de vista, este objetivo lo que busca es utilizar los beneficios de los algoritmos de aprendizaje automático en el escenario particular del abandono de los estudiantes de informática y ciencias de la computación.

El uso de modelos basados en análisis de relevancia y categorización de rasgos, de acuerdo con la técnica de Fernández et al. (2018), la cual permitió identificar patrones distintivos dentro del programa. Además, la comparación de estos modelos con la gama de algoritmos empleados por Pedroza, Chasoy y Gómez

(2019) ofrece la posibilidad de determinar cuál es la mejor estrategia para manejar el problema en esta situación particular.

Objetivo Específico 3: Evaluar la Precisión Predictiva de los Modelos y Ajustarlos para Mejorar su Rendimiento.

Según los enfoques recomendados por Aqila y Faisal (2023), evaluar el poder predictivo de los modelos sugeridos es un paso crucial. Haciendo también hincapié en el uso de árboles de decisión para la categorización destacando la importancia de evaluar la capacidad de los modelos en la predicción.

Hablaremos de cómo las métricas de evaluación Se abordaron las métricas de evaluación tales como la precisión y eficiencia utilizándose en la discusión de este objetivo para medir el éxito de los modelos predictivos en el entorno de la computación y la informática. La sugerencia de Kadiyala y Kumar (2018) nos dice que es necesario un enfoque iterativo para que resuene con la discusión sobre las mejoras de rendimiento y sugiere una estrategia adaptable basada en los resultados de la evaluación.

## VI. CONCLUSIONES

Se logró identificar patrones de deserción estudiantil en el programa de Computación e Informática del IESTP - Lima 2023, mediante la aplicación de un modelo predictivo.

Se identificaron factores determinantes en la deserción estudiantil, la matriz de correlación evidenció relaciones significativas, resaltando la importancia de elementos como "Hábitos Alimenticios", "Acceso a Computadora" y "Ejercicio Regular". Además, variables clave como "Promedio Ponderado" y "Promedio de Inasistencias" mostraron correlaciones negativas notables, indicando su relevancia en el abordaje de la problemática.

La aplicación de modelos de machine learning, centrados en "Promedio Ponderado" y "Promedio de Inasistencias", reveló su destacado papel como variables predictoras clave en la predicción de la deserción estudiantil. Estos modelos demostraron correlaciones negativas sustanciales, consolidando su relevancia en la toma de decisiones del modelo. En contraste, otras variables como "Hábitos Alimenticios" o "Acceso a Computadora" mostraron correlaciones positivas, aunque menos pronunciadas, indicando una influencia más tenue.

Finalmente, se confirmó la robustez de los modelos mediante la validación cruzada con un 77.06% para "Promedio Ponderado" y un 80.73% para "Promedio de Inasistencias", además la matriz de confusión arrojó detalles sobre falsos positivos y falsos negativos, proporcionando una visión más profunda de la eficacia de los modelos en la anticipación de patrones de deserción estudiantil.

## **VII. RECOMENDACIONES**

Se propone avanzar en el desarrollo de modelos de predicción cada vez más complejos, como algoritmos de aprendizaje profundo o redes neuronales, para aumentar la precisión en la identificación de patrones relacionados con la deserción estudiantil. Estos modelos tienen la capacidad de capturar relaciones complejas y no lineales entre las variables, lo que proporcionaría una comprensión más profunda de los factores que afectan la permanencia de los estudiantes.

Se sugiere ampliar el modelo actual incorporando variables contextuales específicas de la IESTP. Esto podría abarcar aspectos como la calidad de la interacción entre estudiantes y docentes, la disponibilidad de recursos académicos y el acceso a programas de apoyo estudiantil. La inclusión de estas variables permitiría una evaluación más completa de los determinantes de la deserción.

Se recomienda realizar investigaciones adicionales que validen el modelo desarrollado en diversos entornos educativos. Esto implica aplicar el modelo en instituciones similares o con características diferentes para evaluar su robustez y aplicabilidad general. La validación en múltiples contextos contribuirá a fortalecer la utilidad y confiabilidad del modelo.

Dada la relevancia de El papel de los factores socioeconómicos en el abandono escolar, se recomienda profundizar en la investigación sobre cómo estos factores interactúan con las variables académicas. Esto podría incluir un análisis de la influencia En cuanto a las circunstancias socioeconómicas de las familias y los estudiantes en el proceso de toma de decisiones académicas.

## REFERENCIAS

ALARCÓN GARCÍA, R.E., 2021. Sistema analítico basado en un modelo predictivo de procesamiento de datos en la big data en la educación superior. En: Accepted: 2021-12-30T02:25:46Z, Repositorio Institucional - USS [en línea], [Consulta: 13 diciembre 2022].

<http://repositorio.uss.edu.pe//handle/20.500.12802/9040>.

AQILA, A. y FAISAL, M., 2023. Lung Cancer EDA Classification Using the Decision Trees Method in Python. *Informatics and Software Engineering*, vol. 1, no. 1, ISSN 2988-2222.

DOI 10.58777/ise.v1i1.56.

FERNÁNDEZ, J., ROJAS, A., DAZA, G., GÓMEZ, D., ÁLVAREZ, A. y OROZCO, Á., 2018. Student Desertion Prediction Using Kernel Relevance Analysis. En: Y. HERNÁNDEZ HEREDIA, V. MILIÁN NÚÑEZ y J. RUIZ SHULCLOPER (eds.), *Progress in Artificial Intelligence and Pattern Recognition*. Cham: Springer International Publishing, pp. 263-270. *Lecture Notes in Computer Science*, ISBN 978-3-030-01132-1.

DOI 10.1007/978-3-030-01132-1\_30.

KADIYALA, A. y KUMAR, A., 2018. Applications of python to evaluate the performance of decision tree-based boosting algorithms. *Environmental Progress & Sustainable Energy*, vol. 37, no. 2, ISSN 1944-7450.

DOI 10.1002/ep.12888.

PEDROZA, K.Y.D., CHASOY, B.Y.C. y GÓMEZ, A.A.R., 2019. Review of techniques, tools, algorithms and attributes for data mining used in student desertion. *Journal of Physics: Conference Series*, vol. 1409, no. 1, ISSN 1742-6596.

DOI 10.1088/1742-6596/1409/1/012003.

BARROSO, Salgado Javier Antonio. *Modelo predictivo basado en machine learning de ordenes de trabajo riesgosas para mantenimiento de equipos mineros* [en línea]. Tesis, Universidad de Chile, 2018 [consultado el 11 de octubre de 2023].

<http://repositorio.uchile.cl/handle/2250/170038>

RENZO ANDRÉ BARRUETA MEZA y EDGAR JEAN PAUL CASTILLO VILLARREAL, 2018. Modelo de análisis predictivo para determinar clientes con tendencia a la deserción en bancos peruanos.

<http://hdl.handle.net/10757/626023>

BOATENG, E.Y., OTOO, J. y ABAYE, D.A., 2020. Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review. *Journal of Data Analysis and Information Processing*, vol. 8, no. 4,

DOI 10.4236/jdaip.2020.84020.

BORJAS GARCÍA, Jorge Edgardo. Validez y confiabilidad en la recolección y análisis de datos bajo un enfoque cualitativo. *TRASCENDER, CONTABILIDAD Y GESTIÓN* [en línea]. 2020, (15), 79–97 [consultado el 11 de octubre de 2023]. ISSN 2448-6388.  
doi:10.36791/tcg.v0i15.90

BRAVO PEÑA, L.C., 2022. Degradación y deforestación en la cuenca del río Conchos (México): modelado predictivo mediante regresión logística (1985-2016). En: Accepted: 2022-06-14T20:29:02ZjournalAbbreviation: Degradation and deforestation in the Conchos river basin (Mexico): predictive modeling through logistic regression (1985-2016)DOI: 10.30827/cuadgeo.v6i1i1.21629 [en línea]. [consulta: 27 octubre 2023].  
<http://cathi.uacj.mx/handle/20.500.11961/22073;jsessionid=5054716A17AD1357D1F10D596D9C57E5>.

LI, T., HIGGINS, J.P. y DEEKS, J.J., 2019. Collecting data. *Cochrane Handbook for Systematic Reviews of Interventions* [en línea]. S.l.: John Wiley & Sons, Ltd, pp. 109-141. [consulta: 25 noviembre 2023]. ISBN 978-1-119-53660-4.  
<https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119536604.ch5>.

PATIL, S. y KULKARNI, U., 2019. Accuracy Prediction for Distributed Decision Tree using Machine Learning approach. 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI) [en línea]. S.l.: s.n., pp. 1365-1371. [consulta: 25 noviembre 2023]. DOI 10.1109/ICOEI.2019.8862580.  
<https://ieeexplore.ieee.org/abstract/document/8862580>.

SCHRÖER, C., KRUSE, F. y GÓMEZ, J.M., 2021. A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, vol. 181, ISSN 1877-0509.  
DOI 10.1016/j.procs.2021.01.199.

TRAN, M.-K., PANCHAL, S., CHAUHAN, V., BRAHMBHATT, N., MEVAWALLA, A., FRASER, R. y FOWLER, M., 2022. Python-based scikit-learn machine learning models for thermal and electrical performance prediction of high-capacity lithium-ion battery. *International Journal of Energy Research*, vol. 46, no. 2, ISSN 1099-114X.  
DOI 10.1002/er.7202.

WANG, J., YANG, Y., WANG, T., SHERRATT, R.S. y ZHANG, J., 2020. Big Data Service Architecture: A Survey. *Journal of Internet Technology*, vol. 21, no. 2, ISSN 2079-4029.  
DOI 10.3966/160792642020032102008.

ZHU, L., YU, F.R., WANG, Y., NING, B. y TANG, T., 2019. Big Data Analytics in Intelligent Transportation Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, ISSN 1558-0016.  
DOI 10.1109/TITS.2018.2815678.

CAZACU, M. y TITAN, E., 2021. Adapting CRISP-DM for Social Sciences. BRAIN. Broad Research in Artificial Intelligence and Neuroscience, vol. 11, no. 2Sup1, ISSN 2067-3957.  
DOI 10.18662/brain/11.2Sup1/97.

MARTÍNEZ-PLUMED, F., CONTRERAS-OCHANDO, L., FERRI, C., HERNÁNDEZ-ORALLO, J., KULL, M., LACHICHE, N., RAMÍREZ-QUINTANA, M.J. y FLACH, P., 2021. CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 8, ISSN 1558-2191.  
DOI 10.1109/TKDE.2019.2962680.

MACKIESON, P., SHLONSKY, A. y CONNOLLY, M., 2019. Increasing rigor and reducing bias in qualitative research: A document analysis of parliamentary debates using applied thematic analysis. Qualitative Social Work, vol. 18, no. 6, ISSN 1473-3250.  
DOI 10.1177/1473325018786996.

Concytec (2018) [repositorio.concytec.gob.pe](https://repositorio.concytec.gob.pe)

Contreras Arteaga, A. I., & Sánchez Cotrina, F. W. (2020). *Analítica predictiva para conocer el patrón de consumo de los clientes en la Empresa Cienpharma S.A.C. utilizando IBM SPSS Modeler y la metodología CRISP-DM.* <https://hdl.handle.net/20.500.12759/6629>

Contreras-Bravo, L. E., Tarazona-Bermúdez, G. M. y Rodríguez-Molano, J. I. (2021). Tecnología y analítica del aprendizaje: una revisión a la literatura. Revista Científica, 41(2), 150-168.  
<https://doi.org/10.14483/23448350.17547>

Education at a Glance 2019: OECD INDICATORS. *Observatorio de la Infancia y Adolescencia de Andalucía (OIA-A)* [en línea]. [sin fecha] [consultado el 11 de octubre de 2023].  
[https://www.observatoriodelainfancia.es/oia/esp/documentos\\_ficha.aspx?id=5861](https://www.observatoriodelainfancia.es/oia/esp/documentos_ficha.aspx?id=5861)

HENDERI, H., WAHYUNINGSIH, T. y RAHWANTO, E., 2021. Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. *International Journal of Informatics and Information Systems*, vol. 4, no. 1, ISSN 2579-7069.  
DOI 10.47738/ijjis.v4i1.73.

HERNANDEZ SIAMPIERI, Roberto y Cristhian Paulina MENDOZA TORRES. *Metodología de la investigación: las rutas cuantitativa, cualitativa y mixta.* Editado por MCGRAW-HILL INTERAMERICANA, 2018. [sin fecha].  
<https://books.google.com.pe/books?id=GH1dwAEACAAJ>

HOYOS OSORIO, Jhoan Keider y Genaro DAZA SANTACOLOMA. Predictive model to identify college students with high dropout rates. *Revista Electrónica de*

*Investigación Educativa* [en línea]. 2023, **25**, 1–10 [consultado el 11 de octubre de 2023]. ISSN 1607-4041.  
doi:10.24320/redie.2023.25.e13.5398

JHONG GUILLEN, S.J., 2019. Análisis comparativo de las técnicas de minería de datos para la estimación de consumos de energía eléctrica en la empresa Electronorte S.A. En: Accepted: 2020-02-25T23:09:57Z, Repositorio Institucional - USS [en línea], [Consulta: 13 diciembre 2022].  
<http://repositorio.uss.edu.pe/handle/20.500.12802/6768>.

MATURRANO, E.F.L., 2021. El fichaje de investigación como estrategia para la formación de competencias investigativas. *EDUCARE ET COMUNICARE Revista de investigación de la Facultad de Humanidades*, vol. 9, no. 1, ISSN 2312-9018.  
DOI 10.35383/educare.v9i1.594. Perú

MARTÍNEZ, D.V.S., 2022. Técnicas e instrumentos de recolección de datos en investigación. *TEPEXI Boletín Científico de la Escuela Superior Tepeji del Río*, vol. 9, no. 17, ISSN 2007-7629.  
DOI 10.29057/estr.v9i17.7928.

Medina-Díaz, María del R., Verdejo-Carrión Ada L. . Validez y confiabilidad en la evaluación del aprendizaje mediante las metodologías activas. *Alteridad. Revista de Educación* [en línea]. 2020, 15(2), 270-284[fecha de Consulta 11 de Octubre de 2023]. ISSN: 1390-325X.  
<https://www.redalyc.org/articulo.oa?id=467763400011>

MORALES HERNÁNDEZ, M.Á., GONZÁLEZ CAMACHO, J.M., ROBLES VÁSQUEZ, H., VALLE PANIAGUA, D.H. del, DURÁN MORENO, J.R., MORALES HERNÁNDEZ, M.Á., GONZÁLEZ CAMACHO, J.M., ROBLES VÁSQUEZ, H., VALLE PANIAGUA, D.H. del y DURÁN MORENO, J.R., 2022. Algoritmos de aprendizaje automático para la predicción del logro académico. *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo* [en línea], vol. 12, no. 24. [Consulta: 13 diciembre 2022]. ISSN 2007-7467. DOI 10.23913/ride.v12i24.1180.  
[http://www.scielo.org.mx/scielo.php?script=sci\\_abstract&pid=S2007-74672022000100035&lng=es&nrm=iso&tlng=es](http://www.scielo.org.mx/scielo.php?script=sci_abstract&pid=S2007-74672022000100035&lng=es&nrm=iso&tlng=es).

NAKANO EDGAR, R.J. y TORRES, E.F., 2021. Modelo de aprendizaje no supervisado para la priorización de inventarios cíclicos. En: Accepted: 2021-10-16T07:36:14Z Artwork Medium: Digital Interview Medium: Digital publisher: Universidad Icesi [en línea], [Consulta: 13 diciembre 2022].  
[http://repository.icesi.edu.co/biblioteca\\_digital/handle/10906/89010](http://repository.icesi.edu.co/biblioteca_digital/handle/10906/89010).

NINA ASTO, C.M. y VILCA MALPARTIDA, J.A., 2018. Búsqueda de patrones de comportamiento usando Machine Learning, para la toma decisiones gerenciales en la empresa Chuchuhuasi. En: Accepted: 2019-09-11T19:55:11Z, Universidad Andina del Cusco [en línea], [Consulta: 13 diciembre 2022].  
<http://repositorio.uandina.edu.pe/handle/20.500.12557/2795>.

NORAMBUENA, J.M., BADILLA-QUINTANA, M.G. y ANGULO, Y.L., 2022. Modelos predictivos basados en uso de analíticas de aprendizaje en educación superior: una revisión sistemática. Texto Livre [en línea], vol. 15. [Consulta: 13 diciembre 2022]. ISSN 1983-3652. DOI 10.35699/1983-3652.2022.36310. <http://www.scielo.br/j/tl/a/Gs8gvXkzMkks5sVZjgBdhPP/abstract/?lang=es>.

QUIROGA, F. y LANZARINI, L.C., 2019. APRENDIZAJE AUTOMÁTICO. APLICACIONES EN RECONOCIMIENTO DE GESTOS, ACCIONES Y SEÑAS. *Investigación Joven*, vol. 6, no. Especial, pp. 174-174. ISSN 2314-3991. <https://revistas.unlp.edu.ar/InvJov/article/view/7096>

Revista Universidad y Empresa. *Revista Universidad y Empresa* [en línea], [sin fecha]. [Consulta: 13 diciembre 2022]. Disponible en: <https://revistas.urosario.edu.co/index.php/empresa>.

QUIROZ, J.M., 2021. Factores asociados al riesgo en la deserción estudiantil en un Instituto de Educación Superior Tecnológico Público. *Revista de Investigación de la Universidad Norbert Wiener*, vol. 10, no. 2, ISSN 2663-7677. DOI 10.37768/unw.rinv.10.02.006.

PÉREZ, Alexis Matheu et al. Modelo de predicción de la deserción estudiantil de primer año en la Universidad Bernardo O'Higgins. *Educação e Pesquisa* [en línea]. 2018, **44** [consultado el 11 de octubre de 2023]. ISSN 1678-4634. doi:10.1590/s1678-4634201844172094

RIVAS GUERRA, C.E., 2019. Análisis de modelos predictivos que estimen la deserción de estudiantes de pregrado: una revisión de la literatura científica. <https://repositorio.upn.edu.pe/handle/11537/22982>

RODRÍGUEZ-CRUZ, Yunier y María PINTO. Modelo de uso de información para la toma de decisiones estratégicas en organizaciones de información. *Transinformação* [en línea]. 2018, **30**(1), 51–64 [consultado el 11 de octubre de 2023]. ISSN 2318-0889. doi:10.1590/2318-08892018000100005

ROBLES-VELASCO, A., CORTÉS, P., MUÑUZURI, J. y BARBADILLA-MARTÍN, E., 2020. Aplicación de la regresión logística para la predicción de roturas de tuberías en redes de abastecimiento de agua. *Dirección y Organización*, no. 70, ISSN 2171-6323. DOI 10.37610/dyo.v0i70.570.

SIFUENTES BITOCCHI, O., 2018. Modelos predictivos de la deserción estudiantil en una universidad privada peruana. En: Accepted: 2020-12-03T16:51:44Z, *Universidad Continental* [en línea], [consulta: 20 octubre 2023]. DOI 10.15381/idata.v21i2.15602. [https://www.researchgate.net/publication/329847510\\_Modelos\\_predictivos\\_de\\_la\\_desercion\\_estudiantil\\_en\\_una\\_universidad\\_privada\\_peruana](https://www.researchgate.net/publication/329847510_Modelos_predictivos_de_la_desercion_estudiantil_en_una_universidad_privada_peruana)

SILVA, S. S. M., Madawa W. JAYAWARDANA y Denny MEYER. Statistical methods to model and evaluate physical activity programs, using step counts: a systematic review. *PLOS ONE* [en línea]. 2018, **13**(11), e0206763 [consultado el 11 de octubre de 2023]. ISSN 1932-6203. doi:10.1371/journal.pone.0206763

SOTOMAYOR SOLOAGA, Pedro y David RODRÍGUEZ-GÓMEZ. Factores explicativos de la deserción académica en la Educación Superior Técnico Profesional: el caso de un centro de formación técnica. *Revista de Estudios y Experiencias en Educación* [en línea]. 2020, **19**(41), 199–223 [consultado el 20 de octubre de 2023]. ISSN 0718-5162. doi:10.21703/rexe.20201941sotomayor11

TIMARÁN-PEREIRA, R., CAICEDO-ZAMBRANO, J. y HIDALGO-TROYA, A., 2019. Árboles de decisión para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas Saber 11°. *Revista de Investigación, Desarrollo e Innovación*, vol. 9, no. 2, pp. 363-378. ISSN 2389-9417. DOI 10.19053/20278306.v9.n2.2019.9184.

GARCÍA DEL RÍO, A. y LÓPEZ CONTRERAS, I.R., 2018. IMPLEMENTACIÓN DE HERRAMIENTAS DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE DATOS ESTRUCTURADOS EN BIG DATA. <http://erecursos.uacj.mx/handle/20.500.11961/4665>.

USECHE, María, ARTIGAS, Wileidys, QUEIPO, Beatriz, y otros . Técnicas e instrumentos de recolección de datos cuali-cuantitativos. [en línea] Colombia : Universidad de la Guajira, 201986 páginas ISBN 9789566037040. [Fecha consulta : 11 de octubre 2023]. <https://repositoryinst.uniguajira.edu.co/handle/uniguajira/467>

VALENZUELA GONZÁLEZ, G., 2022. Aprendizaje Supervisado: Métodos, Propiedades y Aplicaciones. En: Accepted: 2022-09-29T08:43:37Z [en línea], [Consulta: 13 diciembre 2022]. <https://riuma.uma.es/xmlui/handle/10630/25147>.

YAKUBU, Mohammed Nasiru y A. Mohammed ABUBAKAR. Applying machine learning approach to predict students' performance in higher educational institutions. *Kybernetes* [en línea]. 2021, **ahead-of-print**(ahead-of-print) [consultado el 11 de octubre de 2023]. ISSN 0368-492X. doi:10.1108/k-12-2020-0865

YOVERA JIBAJA, A.E., 2021. Modelo de análisis predictivo para el proceso de resolución de contratos de clientes en la Inmobiliaria Miraflores Perú S.A.C 2021. En: Accepted: 2022-06-24T19:34:22Z, Repositorio Institucional - UCV [en línea], [Consulta: 13 diciembre 2022]. <https://repositorio.ucv.edu.pe/handle/20.500.12692/91064>.

Contreras, Leonardo E., et al. "Academic performance prediction by machine learning as a success/failure indicator for engineering students." vol. 13, no. 5,

2020.

<https://doi.org/10.4067/S0718-50062020000500233>.

Hooda, Monica. "Learning Analytics Lens: Improving Quality of Higher Education." *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 5, 2020, p. 21.

<https://doi.org/10.30534/ijeter/2020/24852020>.

BERENS, J., SCHNEIDER, K., GÖRTZ, S., OSTER, S. y BURGHOFF, J., 2018. Early Detection of Students at Risk – Predicting Student Dropouts Using Administrative Student Data and Machine Learning Methods [en línea]. SSRN Scholarly Paper. 2018. Rochester, NY: s.n. [consulta: 23 noviembre 2023]. 3275433.

<https://papers.ssrn.com/abstract=3275433>.

PIERRAKEAS, C., KOUTSONIKOS, G., LIPITAKIS, A.-D., KOTSIANTIS, S., XENOS, M. y GRAVVANIS, G.A., 2020. The Variability of the Reasons for Student Dropout in Distance Learning and the Prediction of Dropout-Prone Students. En: M. VIRVOU, E. ALEPIS, G.A. TSIHRINTZIS y L.C. JAIN (eds.), *Machine Learning Paradigms: Advances in Learning Analytics* [en línea]. Cham: Springer International Publishing, pp. 91-111. [consulta: 23 noviembre 2023]. Intelligent Systems Reference Library, ISBN 978-3-030-13743-4.

[https://doi.org/10.1007/978-3-030-13743-4\\_6](https://doi.org/10.1007/978-3-030-13743-4_6).

## ANEXOS

### Anexo: Matriz de operacionalización de variables

Variable de Estudio	Definición Conceptual	Definición Operacional	Dimensiones	Indicadores	Escala de Medición	Instrumento
Deserción Estudiantil	La deserción estudiantil se refiere al abandono voluntario de la actividad académica por parte de un estudiante antes de completar el período de estudios establecido en una institución educativa,	La deserción estudiantil se medirá a través de la variable "desertaron" en la tabla de registros. Si un estudiante tiene el valor "Sí" en la columna "desertaron", se considerará que desertó; de lo contrario, se	-Factores	- Factores socioeconómicos	- Nominal - Nominal (M/F) - Nominal - Razón(discreta)	Ficha de Registro
	generalmente dentro de los tres primeros semestres académicos. (Prediction model of first-year student desertion at Universidad Bernardo O'Higgins (UBO)*, 2018)	considerará que no desertó. La deserción se considerará cuando un estudiante no continúa sus estudios durante el mismo semestre en el que se matriculó inicialmente.		-Factores Académicos	-De razón discreta (0 a 20) -De razón (0% a 100%)	Ficha de Registro
Modelo Predictivo	Son estrategias que predicen resultados, utilizando aprendizaje automático, la minería de datos históricos, el big data y el reconocimiento de patrones para la mejora de la toma de decisiones mediante técnicas de análisis de datos" (Shenzhen, 2022).	El modelo predictivo se medirá por su eficiencia mediante las siguientes métricas: precisión, f1 score y recall	Eficiencia	-Nivel de precisión del modelo		Guía de observación

Elaboración: Fuente propia de los autores

## Anexo: Instrumentos de recolección de datos

### Ficha de registro N°1

#### Indicador: Factores Personales

El objetivo de esta ficha de análisis documental es obtener los datos académicos de los estudiantes del programa académico de computación e informática, datos los cuales son provenientes de los registros de los docentes de las asignaturas del programa académico, se tuvo en cuenta la siguiente estructura:

Registro	CAMPOS							
	Turno	Sección	Curso	Semestre	Nota final de curso	Porcentaje de inasistencias	Promedio ponderado	Promedio de asistencia
R1								
R2								
R3								
Rn								

Elaboración Propia de los autores.

Nº	Campo	Tipo	Descripción
1	Turno	CHAR	Corresponde a los turnos de los cursos, donde D = Diurno y N = Nocturno. No debe tener valores blancos ni nulos.
2	Sección	CHAR	Corresponde a la sección del curso, donde existen solo A y B. No debe tener valores blancos ni nulos
3	Curso	String	Corresponde a los cursos del programa académico, donde el nombre del curso debe llevar la nomenclatura correspondiente. No debe tener valores blancos ni nulos
4	Semestre	String	Pueden ser valores como: I, II, III, IV, V, VI. No debe tener valores blancos ni nulos
5	Nota final de curso	Float	La nota final debe ser un número decimal en un rango específico, como de 0 a 20. No debe tener valores blancos ni nulos
6	Porcentaje de inasistencias	Float	El número de inasistencias debe ser un número entero no negativo. No debe tener valores blancos ni nulos
7	Promedio ponderado	Float	Es el promedio del semestre
8	Promedio de asistencia	Float	Es el promedio de asistencia del semestre

Elaboración Propia de los autores.

## Guía de observación

Indicador: Nivel de precisión del modelo

Objetivo: El objetivo de esta guía de observación es obtener los datos de las métricas del modelo como precisión, F1 Score y Rcall. Estas métricas evaluarán la capacidad del modelo para predecir con precisión si un estudiante desertará o no.

<b>Nº</b>	<b>Instancias correctas</b>	<b>Instancias incorrectas</b>	<b>Precisión</b>	<b>F1 Score</b>	<b>Rcall</b>

Elaboración propia de los autores

## Anexo: base de datos utilizada para la investigación en formato .csv

formu\_numerico13 - Excel

Archivo Inicio Insertar Disposición de página Fórmulas Datos Revisar Vista Ayuda ¿Qué desea hacer?

Calibri 11 Fuente Alineación Número

Formato condicional Dar formato como tabla Estilos de celda Insertar Eliminar Ordenar y filtrar Buscar y seleccionar Complementos

POSIBLE PÉRDIDA DE DATOS Algunas características del libro se pueden perder si lo guarda como CSV (delimitado por comas). Para conservar estas características, guárdelo como archivo de Excel. No mostrar de nuevo Guardar como...

sexo

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
sexo	distrito_res	acceso_con	acceso_inti	apoyo_fam	ingresos	trabajo_pa	becas_ayuc	tipo_famili	carga_fam	discapacida	ejercicio_re	habitos_al	actividades	consumo_s	consumo_c	desertaron	promedio	promedio_ponderado								
1	1	2	1	1	2	1	0	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0
2	1	3	1	1	2	1	0	0	1	0	1	1	0	0	1	1	0	0	0	0	0	0	1	1	0	0
3	0	1	1	1	2	1	1	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	2	2	0
4	1	0	1	1	2	2	1	0	0	0	1	1	1	1	1	1	1	0	2	2	2	2	2	2	2	0
5	1	2	1	1	2	1	1	0	0	0	1	1	1	1	1	1	1	0	2	2	1	1	1	2	0	0
6	1	2	1	1	2	1	1	0	0	0	1	1	1	1	1	1	0	1	2	0	1	1	2	0	0	0
7	1	0	1	1	2	1	0	0	1	0	0	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0
8	1	0	1	1	2	1	0	1	1	1	0	0	1	0	1	0	0	2	1	0	0	2	1	1	0	0
9	0	2	1	1	1	1	1	0	1	0	0	1	1	0	0	0	0	2	2	2	2	2	2	2	0	0
10	1	0	1	1	2	1	1	0	0	0	0	0	1	1	1	1	0	0	2	2	2	2	2	2	0	0
11	1	3	1	1	3	1	0	0	1	0	0	1	1	1	1	1	1	0	2	2	2	2	2	2	0	0
12	1	0	1	1	2	1	1	0	1	0	0	1	1	1	1	0	0	2	2	2	2	2	2	2	0	0
13	1	0	1	1	2	1	1	0	1	0	0	1	1	1	1	0	0	1	0	0	1	0	0	0	0	0
14	0	1	1	1	2	1	1	0	1	0	0	1	1	0	1	0	0	2	1	0	0	2	1	1	0	0
15	1	0	1	1	2	1	1	0	0	0	1	1	1	1	1	0	0	2	2	2	2	2	2	2	0	0
16	1	1	1	1	2	1	0	0	1	0	0	1	1	1	1	0	0	2	2	2	2	2	2	2	0	0
17	0	0	1	1	2	1	0	0	1	0	0	0	1	0	0	0	0	2	1	0	0	2	1	1	0	0
18	1	1	1	1	2	1	0	0	1	0	0	0	1	1	1	1	0	1	1	0	1	1	0	0	0	0
19	1	0	1	1	2	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	0	0	0	0	0	0
20	1	2	1	1	2	1	1	0	1	0	0	0	1	1	1	1	0	0	2	1	1	1	2	0	0	0
21	1	0	1	1	2	1	1	0	0	0	0	1	1	1	1	0	1	1	0	1	1	0	1	1	0	0
22	1	0	1	1	2	1	0	0	1	0	0	1	1	0	1	0	1	1	0	1	1	1	0	1	0	0
23	1	0	1	1	2	1	1	0	1	0	0	0	1	0	0	0	0	1	1	0	0	1	1	0	0	0
24	1	0	1	1	2	1	1	0	1	0	0	1	1	1	1	1	1	0	2	2	2	2	2	2	0	0
25	0	2	1	1	2	1	0	0	0	0	1	0	1	1	1	1	0	0	2	2	2	2	2	2	0	0
26	1	1	1	1	2	1	1	1	0	0	0	1	1	1	1	0	0	2	2	2	2	2	2	2	0	0

datosmodi - Excel

Insertar Disposición de página Fórmulas Datos Revisar Vista Ayuda ¿Qué desea hacer?

Calibri 11 Fuente Alineación Número

Formato condicional Dar formato como tabla Estilos de celda Insertar Eliminar Ordenar y filtrar Buscar y seleccionar Complementos

DIDA DE DATOS Algunas características del libro se pueden perder si lo guarda como CSV (delimitado por comas). Para conservar estas características, guárdelo como archivo de Excel. No mostrar de nuevo Guardar como...

dni

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
sexo	edad	curso	semestre	turno	seccion	inasistencias	nota_final	Desertaron																		
F	20	CFD	1	D	A	0	0	No																		
M	22	CFD	1	D	A	0	0	No																		
M	23	CFD	1	D	A	0	13	No																		
M	22	CFD	1	D	A	0	13	No																		
M	18	CFD	1	D	A	0	13	No																		
M	19	CFD	1	D	A	0	13	No																		
M	22	CFD	1	D	A	0	14	No																		
M	25	CFD	1	D	A	0	14	No																		
M	17	CFD	1	D	A	0	14	No																		
M	23	CFD	1	D	A	0	14	No																		
F	22	CFD	1	D	A	0	15	No																		
M	24	CFD	1	D	A	0	15	No																		
M	20	CFD	1	D	A	0	15	No																		
M	21	CFD	1	D	A	0	15	No																		
M	19	CFD	1	D	A	0	15	Yes																		
M	23	CFD	1	D	A	0	15	Yes																		
M	22	CFD	1	D	A	0	15	Yes																		
M	25	CFD	1	D	A	0	15	No																		
M	22	CFD	1	D	A	0	15	No																		
M	23	CFD	1	D	A	0	15	No																		



# UNIVERSIDAD CÉSAR VALLEJO

## CONSTANCIA DE VALIDACIÓN

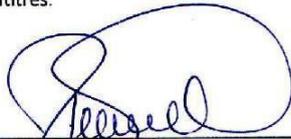
Yo, Rosita Eliabeth Yovera Morales, con DNI N° 44052819  
de profesión Ingeniera de sistemas y el grado de Maestría.

Por medio de la presente hago constar que he revisado con fines de Validación para los **Instrumentos de recolección de datos de la guía de observación bajo el proceso observado de porcentaje de inasistencias**, para la investigación titulada, **Patrones de la deserción estudiantil aplicando un modelo predictivo en un IESTP - Lima 2023**, elaborada por los estudiantes Mendoza Chero Ricardo Ryan y Silva Arellano José Luis.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

Instrumentos de recolección de datos guía de observación N° 1	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO	EXCELENTE
1. Claridad			X		
2. Objetividad		X			
3. Actualidad			X		
4. Organización		X			
5. Suficiencia			X		
6. Intencionalidad			X		
7. Consistencia			X		
8. Coherencia		X			
9. Metodología			X		

En señal de conformidad firmo la presente en la ciudad de Piura a los 17 días del mes de Noviembre del Dos mil veintitrés.

  
Firma del Experto



# UNIVERSIDAD CÉSAR VALLEJO

## CONSTANCIA DE VALIDACIÓN

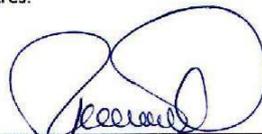
Yo, Rosita Elizabel y Jovana Morales, con DNI N° 44052829  
de profesión Ingeniera de sistemas y el grado de Maestría.

Por medio de la presente hago constar que he revisado con fines de Validación para los **Instrumentos de recolección de datos de la guía de observación bajo el proceso observado de la capacidad del modelo para predecir si un estudiante desertará o no en factor de su nivel socioeconómico**, para la investigación titulada, **Patrones de la deserción estudiantil aplicando un modelo predictivo en un IESTP - Lima 2023**, elaborada por los estudiantes Mendoza Chero Ricardo Ryan y Silva Arellano José Luis.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

Instrumentos de recolección de datos guía de observación N° 2	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO	EXCELENTE
1. Claridad			X		
2. Objetividad			X		
3. Actualidad			X		
4. Organización		X			
5. Suficiencia			X		
6. Intencionalidad		X			
7. Consistencia			X		
8. Coherencia		X			
9. Metodología			X		

En señal de conformidad firmo la presente en la ciudad de Piura a los 17 días del mes de Noviembre del Dos mil veintitrés.

  
Firma del Experto



# UNIVERSIDAD CÉSAR VALLEJO

## CONSTANCIA DE VALIDACIÓN

Yo, Rosita Elizabeth Yovera Morales, con DNI N° 44052829  
de profesión Ingeniera de Sistemas y el grado de Maestría.

Por medio de la presente hago constar que he revisado con fines de Validación para los **Instrumentos de recolección de datos de la encuesta bajo el proceso observado de encontrar factores que puedan influir en la deserción**, para la investigación titulada, **Patrones de la deserción estudiantil aplicando un modelo predictivo en un IESTP - Lima 2023**, elaborada por los estudiantes Mendoza Chero Ricardo Ryan y Silva Arellano José Luis.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

Instrumentos de recolección de datos encuesta	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO	EXCELENTE
1. Claridad			X		
2. Objetividad			X		
3. Actualidad			X		
4. Organización				X	
5. Suficiencia				X	
6. Intencionalidad			X		
7. Consistencia		X			
8. Coherencia			X		
9. Metodología			X		

En señal de conformidad firmo la presente en la ciudad de Piura a los 17 días del mes de Noviembre del Dos mil veintitrés.

  
Firma del Experto



# UNIVERSIDAD CÉSAR VALLEJO

## CONSTANCIA DE VALIDACIÓN

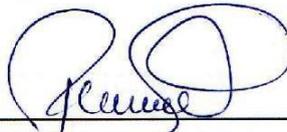
Yo, Rosita Elizabeth Yovera Novales, con DNI N° 44051829  
de profesión Ingeniera de sistemas y el grado de Maestría.

Por medio de la presente hago constar que he revisado con fines de Validación para los **Instrumentos de recolección de datos de la ficha de registro bajo el proceso observado de recolectar la información de los factores académicos**, para la investigación titulada, **Patrones de la deserción estudiantil aplicando un modelo predictivo en un IESTP - Lima 2023**, elaborada por los estudiantes Mendoza Chero Ricardo Ryan y Silva Arellano José Luis.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

Instrumentos de recolección de datos ficha de registro	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO	EXCELENTE
1. Claridad			X		
2. Objetividad			X		
3. Actualidad			X		
4. Organización				X	
5. Suficiencia			X		
6. Intencionalidad		X			
7. Consistencia		X			
8. Coherencia		X			
9. Metodología			X		

En señal de conformidad firmo la presente en la ciudad de Piura a los 17 días del mes de Noviembre del Dos mil veintitrés.

  
Firma del Experto



# UNIVERSIDAD CÉSAR VALLEJO

## CONSTANCIA DE VALIDACIÓN

Yo, ELMER ALFREDO CHUNGA ZAPATA, con DNI N° 02871710  
de profesión INGENIERO INFORMÁTICO y el grado de MAESTRO

Por medio de la presente hago constar que he revisado con fines de Validación para los **Instrumentos de recolección de datos de la guía de observación bajo el proceso observado de la capacidad del modelo para predecir si un estudiante desertará o no en factor de su nivel socioeconómico**, para la investigación titulada, **Patrones de la deserción estudiantil aplicando un modelo predictivo en un IESTP - Lima 2023**, elaborada por los estudiantes Mendoza Chero Ricardo Ryan y Silva Arellano José Luis.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

Instrumentos de recolección de datos guía de observación N° 2	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO	EXCELENTE
1. Claridad			X		
2. Objetividad				X	
3. Actualidad			X		
4. Organización			X		
5. Suficiencia				X	
6. Intencionalidad			X		
7. Consistencia			X		
8. Coherencia			X		
9. Metodología			X		

En señal de conformidad firmo la presente en la ciudad de Piura a los 17 días del mes de Noviembre del Dos mil veintitrés.

Firma del Experto



# UNIVERSIDAD CÉSAR VALLEJO

## CONSTANCIA DE VALIDACIÓN

Yo, ELMER ALFREDO CHUNCA ZAPATA, con DNI N° 02871710  
de profesión INGENIERO INFORMÁTICA y el grado de MAESTRO.

Por medio de la presente hago constar que he revisado con fines de Validación para los **Instrumentos de recolección de datos de la encuesta bajo el proceso observado de encontrar factores que puedan influir en la deserción**, para la investigación titulada, **Patrones de la deserción estudiantil aplicando un modelo predictivo en un IESTP - Lima 2023**, elaborada por los estudiantes Mendoza Chero Ricardo Ryan y Silva Arellano José Luis.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

Instrumentos de recolección de datos encuesta	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO	EXCELENTE
1. Claridad			X		
2. Objetividad			X		
3. Actualidad			X		
4. Organización			X		
5. Suficiencia			X		
6. Intencionalidad			X		
7. Consistencia			X		
8. Coherencia			X		
9. Metodología			X		

En señal de conformidad firmo la presente en la ciudad de Piura a los 17 días del mes de Noviembre del Dos mil veintitrés.

Firma del Experto



# UNIVERSIDAD CÉSAR VALLEJO

## CONSTANCIA DE VALIDACIÓN

Yo, ELMER ALFREDO CHUNGA ZAPATA, con DNI N° 02871710  
de profesión INGENIERO INFORMÁTICO y el grado de MAESTRO.

Por medio de la presente hago constar que he revisado con fines de Validación para los **Instrumentos de recolección de datos de la ficha de registro bajo el proceso observado de recolectar la información de los factores académicos**, para la investigación titulada, **Patrones de la deserción estudiantil aplicando un modelo predictivo en un IESTP - Lima 2023**, elaborada por los estudiantes Mendoza Chero Ricardo Ryan y Silva Arellano José Luis.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

Instrumentos de recolección de datos ficha de registro	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO	EXCELENTE
1. Claridad		X			
2. Objetividad			X		
3. Actualidad				X	
4. Organización			X		
5. Suficiencia			X		
6. Intencionalidad			X		
7. Consistencia				X	
8. Coherencia			X		
9. Metodología				X	

En señal de conformidad firmo la presente en la ciudad de Piura a los 17 días del mes de Noviembre del Dos mil veintitrés.

Firma del Experto



# UNIVERSIDAD CÉSAR VALLEJO

## CONSTANCIA DE VALIDACIÓN

Yo, ELMER ALFREDO CHUNGA ZAPATA, con DNI N° 02871710  
de profesión INGENIERO INFORMÁTICO y el grado de MAESTRO.

Por medio de la presente hago constar que he revisado con fines de Validación para los **Instrumentos de recolección de datos de la guía de observación bajo el proceso observado de porcentaje de inasistencias**, para la investigación titulada, **Patrones de la deserción estudiantil aplicando un modelo predictivo en un IESTP - Lima 2023**, elaborada por los estudiantes **Mendoza Chero Ricardo Ryan** y **Silva Arellano José Luis**.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

Instrumentos de recolección de datos guía de observación N° 1	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO	EXCELENTE
1. Claridad			X		
2. Objetividad				X	
3. Actualidad				X	
4. Organización				X	
5. Suficiencia			X		
6. Intencionalidad				X	
7. Consistencia				X	
8. Coherencia			X		
9. Metodología				X	

En señal de conformidad firmo la presente en la ciudad de Piura a los 17 días del mes de Noviembre del Dos mil veintitrés.

Firma del Experto



# UNIVERSIDAD CÉSAR VALLEJO

## CONSTANCIA DE VALIDACIÓN

Yo, Jván Michell Castillo Jarameng, con DNI N° 02883813  
de profesión Inge. Informática y el grado de Doctor.

Por medio de la presente hago constar que he revisado con fines de Validación para los **Instrumentos de recolección de datos de la guía de observación bajo el proceso observado de la capacidad del modelo para predecir si un estudiante desertará o no en factor de su nivel socioeconómico**, para la investigación titulada, **Patrones de la deserción estudiantil aplicando un modelo predictivo en un IESTP - Lima 2023**, elaborada por los estudiantes Mendoza Chero Ricardo Ryan y Silva Arellano José Luis.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

Instrumentos de recolección de datos guía de observación N° 2	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO	EXCELENTE
1. Claridad		X			
2. Objetividad		X			
3. Actualidad			X		
4. Organización		X			
5. Suficiencia		X			
6. Intencionalidad				X	
7. Consistencia			X		
8. Coherencia			X		
9. Metodología		X			

En señal de conformidad firmo la presente en la ciudad de Piura a los 16 días del mes de Noviembre del Dos mil veintitrés.

Firma del Experto



# UNIVERSIDAD CÉSAR VALLEJO

## CONSTANCIA DE VALIDACIÓN

Yo, Juan Mitchell Castillo Jiménez, con DNI N° 02883813  
de profesión Ing. Informático y el grado de Doctor.

Por medio de la presente hago constar que he revisado con fines de Validación para los **Instrumentos de recolección de datos de la encuesta bajo el proceso observado de encontrar factores que puedan influir en la deserción**, para la investigación titulada, **Patrones de la deserción estudiantil aplicando un modelo predictivo en un IESTP - Lima 2023**, elaborada por los estudiantes Mendoza Chero Ricardo Ryan y Silva Arellano José Luis.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

Instrumentos de recolección de datos encuesta	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO	EXCELENTE
1. Claridad			X		
2. Objetividad		X			
3. Actualidad		X			
4. Organización		X			
5. Suficiencia		X			
6. Intencionalidad			X		
7. Consistencia		X			
8. Coherencia		X			
9. Metodología		X			

En señal de conformidad firmo la presente en la ciudad de Piura a los 16 días del mes de Noviembre del Dos mil veintitrés.

Firma del Experto



# UNIVERSIDAD CÉSAR VALLEJO

## CONSTANCIA DE VALIDACIÓN

Yo, Juan Mitchell Castillo Jimenez, con DNI N° 02883813  
de profesión Ing. Informático y el grado de Doctor.

Por medio de la presente hago constar que he revisado con fines de Validación para los **Instrumentos de recolección de datos de la ficha de registro bajo el proceso observado de recolectar la información de los factores académicos**, para la investigación titulada, **Patrones de la deserción estudiantil aplicando un modelo predictivo en un IESTP - Lima 2023**, elaborada por los estudiantes Mendoza Chero Ricardo Ryan y Silva Arellano José Luis.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

Instrumentos de recolección de datos ficha de registro	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO	EXCELENTE
1. Claridad			X		
2. Objetividad		X			
3. Actualidad		X			
4. Organización			X		
5. Suficiencia		X			
6. Intencionalidad			X		
7. Consistencia		X			
8. Coherencia		X			
9. Metodología			X		

En señal de conformidad firmo la presente en la ciudad de Piura a los 16 días del mes de Noviembre del Dos mil veintitrés.

Firma del Experto



# UNIVERSIDAD CÉSAR VALLEJO

## CONSTANCIA DE VALIDACIÓN

Yo, Juan Michell Castillo Jiménez, con DNI N° 02883813  
de profesión Inf. Informático y el grado de Doctor.

Por medio de la presente hago constar que he revisado con fines de Validación para los Instrumentos de recolección de datos de la guía de observación bajo el proceso observado de porcentaje de inasistencias, para la investigación titulada, **Patrones de la deserción estudiantil aplicando un modelo predictivo en un IESTP - Lima 2023**, elaborada por los estudiantes Mendoza Chero Ricardo Ryan y Silva Arellano José Luis.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

Instrumentos de recolección de datos guía de observación N° 1	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO	EXCELENTE
1. Claridad		X			
2. Objetividad			X		
3. Actualidad				X	
4. Organización				X	
5. Suficiencia			X		
6. Intencionalidad			X		
7. Consistencia			X		
8. Coherencia		X			
9. Metodología		X			

En señal de conformidad firmo la presente en la ciudad de Piura a los 16 días del mes de Noviembre del Dos mil veintitrés.

Firma del Experto

## Anexo 6. Autorización de aplicación del instrumento firmado por la entidad



### AUTORIZACIÓN DE LA ORGANIZACIÓN PARA PUBLICAR SU IDENTIDAD EN LOS RESULTADOS DE LAS INVESTIGACIONES

#### Datos Generales

<b>Nombre de la Organización:</b>	<b>RUC:</b> 20179930201
IESTP José Pardo	
<b>Nombre del titular o representante legal:</b>	
Mg. Régulo Montoya Vargas	<b>DNI:</b> 06786388

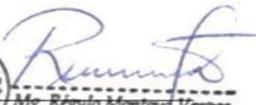
#### Consentimiento:

De conformidad con lo establecido en el artículo 7º, literal "f" del colegio de ética en investigación de la universidad Cesar Vallejo (\*), autorizo , no autorizo  publicar LA IDENTIDAD DE LA ORGANIZACIÓN, en la cual se lleva a cambio la investigación:

<b>Nombre del trabajo de la investigación:</b>	
Modelo de datos predictivos como soporte para la toma de decisiones del proceso académico en el IESTP José Pardo, Lima 2022.	
<b>Nombre del programa académico:</b>	
Escuela Profesional de Ingeniería de Sistemas	
<b>Autor: Nombres y Apellidos:</b>	<b>DNI:</b>
Ricardo Ryan Mendoza Chero	74492120

En caso de autorizarse, soy consiente que la investigación será alojada en el repositorio institucional de la UCV, la misma que será de acceso abierto para los usuarios y podrá ser referenciada en futuras investigaciones, dejando en claro que los derechos de propiedad intelectual corresponden exclusivamente al autor (a) del estudio.

Lugar y fecha:

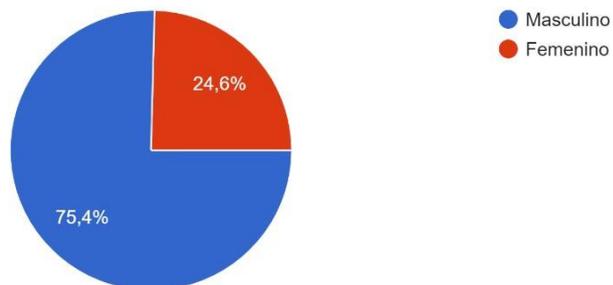
Firma:   
 Mg. Régulo Montoya Vargas  
Director General (e)

(\*)Código de ética en investigación de la universidad cesar vallejo-artículo 7º, literal "f" para difundir o publicar los resultados de un trabajo de investigación es necesario mantener bajo anonimato el nombre de la institución dónde se llevó a cabo el estudio, salvo el caso en que haya un acuerdo formal con el gerente o director de la organización, para que se difunda la identidad de la institución, por ello, tanto en los proyectos de investigación como en las tesis, no se deberá incluir la denominación de la organización, ni en el cuerpo de la tesis ni en los anexos, pero si será necesario describir sus características.

## Anexo 7. Tabulación de cuestionario

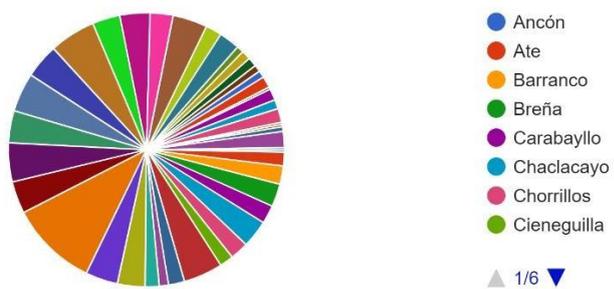
### 2. Sexo

370 respuestas



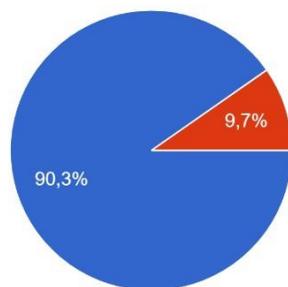
### 3. Distrito donde vives

370 respuestas



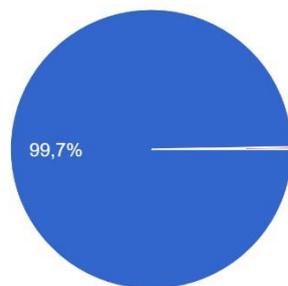
4. ¿Tienes acceso a una computadora?

370 respuestas



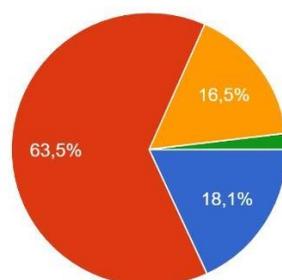
5. ¿Tienes acceso a conexión a Internet?

370 respuestas



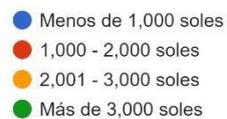
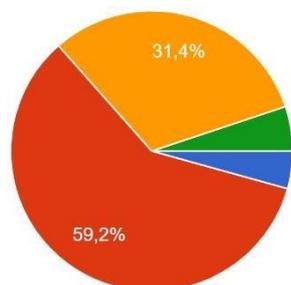
6. ¿Recibes apoyo de tu familia para tus estudios?

370 respuestas



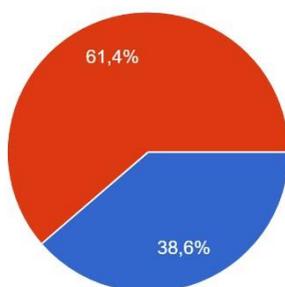
### 7. Ingresos familiares mensuales

370 respuestas



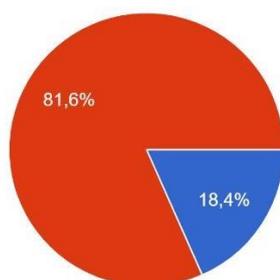
### 8. ¿Trabajas o tienes un empleo a tiempo parcial?

370 respuestas



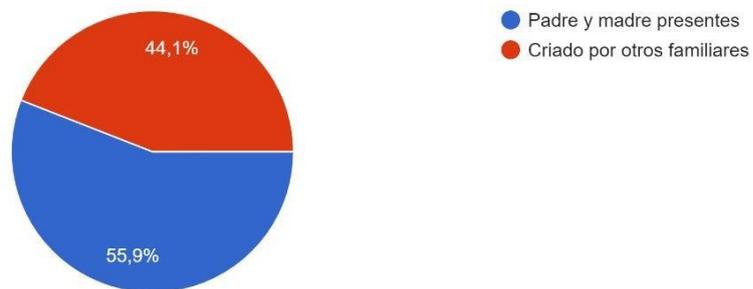
### 9. ¿Tienes acceso a becas o ayuda financiera?

370 respuestas



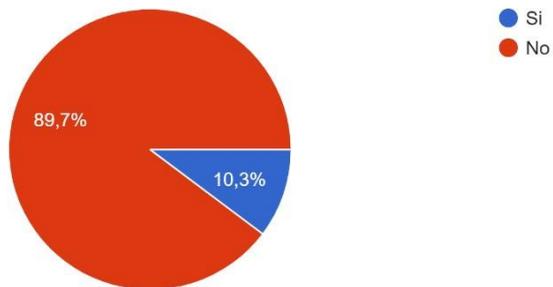
### 10. Tipo de familia

370 respuestas



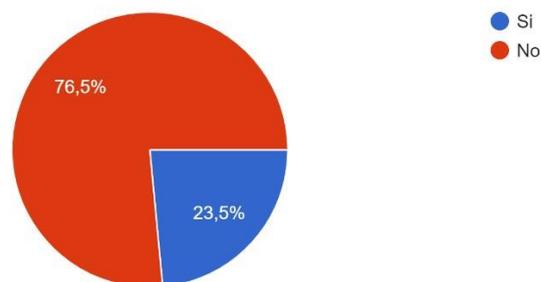
### 11. ¿Tienes a cargo a un miembro de la familia, como un hijo, hermano, tío abuelo u otro familiar enfermo o con discapacidad?

370 respuestas



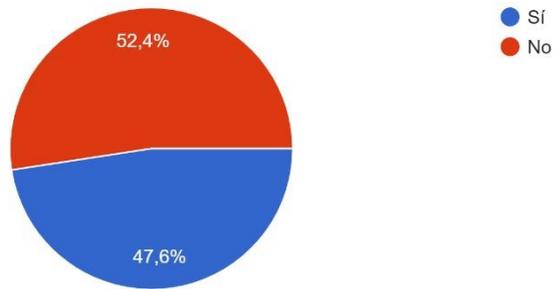
### 12. ¿Tienes una discapacidad o enfermedad que requiera medicación?

370 respuestas



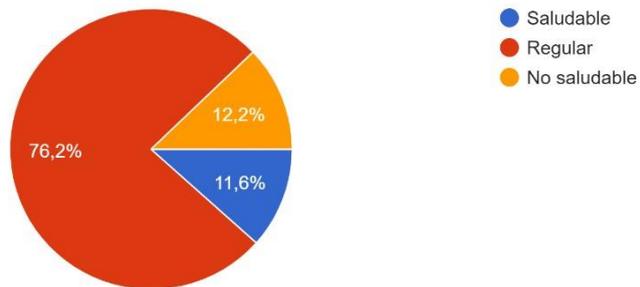
13. ¿Realizas ejercicio regularmente?

370 respuestas



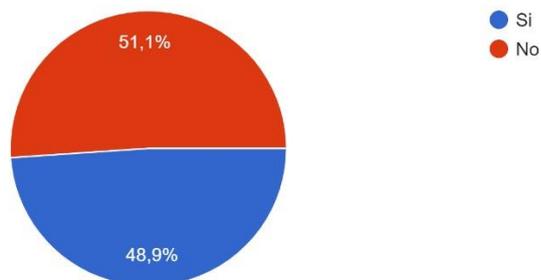
14. Hábitos alimenticios

370 respuestas



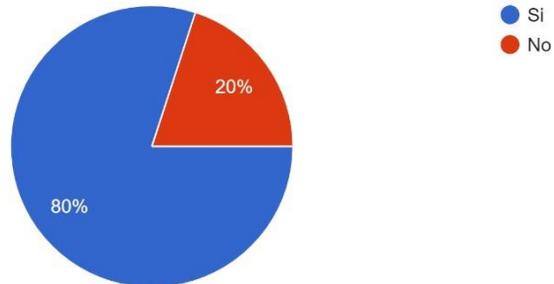
15. ¿Participas en actividades sociales o salidas frecuentes?

370 respuestas



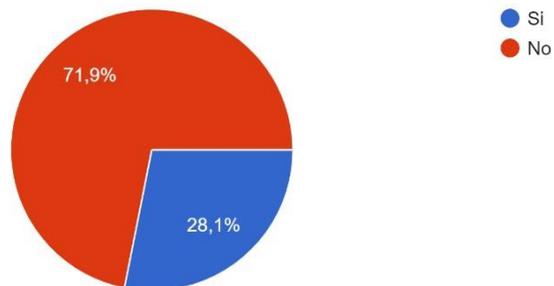
16. ¿Consumes alcohol?

370 respuestas



17. ¿Consumes cigarrillos?

370 respuestas





**UNIVERSIDAD CÉSAR VALLEJO**

**FACULTAD DE INGENIERÍA Y ARQUITECTURA  
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

### **Declaratoria de Autenticidad del Asesor**

Yo, AGURTO MARCHAN WINNER, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - PIURA, asesor de Tesis titulada: "Patrones de la deserción estudiantil aplicando un modelo predictivo en un IESTP - Lima 2023", cuyos autores son MENDOZA CHERO RICARDO RYAN, SILVA ARELLANO JOSE LUIS, constato que la investigación tiene un índice de similitud de 17.00%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

PIURA, 23 de Noviembre del 2023

<b>Apellidos y Nombres del Asesor:</b>	<b>Firma</b>
AGURTO MARCHAN WINNER <b>DNI:</b> 40673760 <b>ORCID:</b> 0000-0002-0396-9349	Firmado electrónicamente por: WAGURTOM el 23- 11-2023 12:04:01

Código documento Trilce: TRI - 0662364