



UNIVERSIDAD CÉSAR VALLEJO

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

**Machine Learning para la toma de decisiones en la gestión de
ventas, Lima 2024**

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:

Ingeniero de Sistemas

AUTORES:

Anchante Aybar, Percy Andersson (orcid.org/0000-0002-4419-7189)

Fernandez Chavez, Davio Dovise (orcid.org/0000-0001-6038-4294)

ASESOR:

Mgtr. Roman Nano, Franklin Rodolfo (orcid.org/0000-0001-7397-6993)

LÍNEA DE INVESTIGACIÓN:

Sistemas de Información y Comunicaciones

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo económico, empleo y emprendimiento

LIMA – PERÚ

2024

Declaratoria de autenticidad del asesor



UNIVERSIDAD CÉSAR VALLEJO

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

Declaratoria de Autenticidad del Asesor

Yo, ROMAN NANO FRANKLIN RODOLFO, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA NORTE, asesor de Tesis titulada: "Machine Learning para la toma de decisiones en la Gestión de Ventas, Lima 2024", cuyos autores son FERNANDEZ CHAVEZ DAVIO DOVISE, ANCHANTE AYBAR PERCY ANDERSSON, constato que la investigación tiene un índice de similitud de 13.00%, 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.

LIMA, 04 de Julio del 2024

Apellidos y Nombres del Asesor:	Firma
ROMAN NANO FRANKLIN RODOLFO DNI: 06158550 ORCID: 0000-0001-7397-6993	Firmado electrónicamente por: FROMANN el 08-07- 2024 21:31:32

Código documento Trilce: TRI - 0792731



Declaratoria de originalidad del autor(es)



ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Declaratoria de Originalidad de los Autores

Nosotros, ANCHANTE AYBAR PERCY ANDERSSON, FERNANDEZ CHAVEZ DAVIO DOVISE estudiantes de la de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA NORTE, declaramos bajo juramento que todos los datos e información que acompañan la Tesis titulada: "Machine Learning para la toma de decisiones en la Gestión de Ventas, Lima 2024", es de nuestra autoría, por lo tanto, declaramos que la Tesis:

1. No ha sido plagiada ni total, ni parcialmente.
2. Hemos mencionado todas las fuentes empleadas, identificando correctamente toda cita textual o de paráfrasis proveniente de otras fuentes.
3. No ha sido publicada, ni presentada anteriormente para la obtención de otro grado académico o título profesional.
4. Los datos presentados en los resultados no han sido falseados, ni duplicados, ni copiados.

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

Nombres y Apellidos	Firma
FERNANDEZ CHAVEZ DAVIO DOVISE DNI: 73663227 ORCID: 0000-0001-6038-4294	Firmado electrónicamente por: DFERNANDEZC1 el 16-07-2024 12:48:54
ANCHANTE AYBAR PERCY ANDERSSON DNI: 72469020 ORCID: 0000-0002-4419-7189	Firmado electrónicamente por: PAANCHANTEA el 16-07-2024 12:48:25

Código documento Trilce: INV - 1751338

Dedicatoria

De: Fernandez Chavez Davio

A mi mamá Patty, por ser mi pilar y fuente inagotable de amor y apoyo. Tu fuerza y dedicación han sido mi inspiración constante. A Jean, por tu paciencia y comprensión. Gracias por estar a mi lado en cada paso, brindándome siempre tu apoyo incondicional. A mi abuelo Davio, por enseñarme el valor del esfuerzo y la perseverancia. Tu sabiduría y ejemplo han sido fundamentales en mi formación y a Gabriela, por tu aliento y compañía. Tu apoyo ha sido invaluable en este camino.

De: Anchante Aybar Percy

A mi padre por su apoyo y dedicación han sido la base de todos mis logros. Tu dedicación y sacrificio han sido una fuente constante de inspiración y motivación en mi vida. Gracias por enseñarme a perseverar y a creer en mí mismo. A mi familia, por su apoyo incondicional y por confiar en mí. Gracias por su paciencia, comprensión y por brindarme un lugar seguro. A mis amigos, por su compañía, alegría y apoyo durante todo este viaje. Gracias por estar en cada paso del camino, celebrando mis éxitos y levantándome en mis fracasos. Su amistad ha sido muy fundamental para llegar hasta aquí.

Agradecimiento

A nuestras familias, por su apoyo incondicional, paciencia y aliento constante en este arduo camino. Su motivación fue fundamental para alcanzar esta meta.

Agradecer a mis amigos leales, refugio de risas y descanso en medio de tantas horas de esfuerzo. Gracias por su comprensión y palabras de ánimo en los momentos de mayor estrés.

Aquellas personas que, con sus consejos, gestos de aliento y oportunidades brindadas contribuyeron a hacer posible este logro.

Y a nosotros mismos, por nuestra perseverancia, dedicación y compromiso inquebrantable para materializar este sueño compartido.

Índice de contenidos

Carátula	i
Declaratoria de autenticidad del asesor.....	ii
Declaratoria de originalidad del autor(es)	iii
Dedicatoria.....	iv
Agradecimiento.....	v
Índice de contenidos.....	vi
Índice de tablas.....	vii
Índice de figuras	viii
Resumen	ix
Abstract.....	x
I. INTRODUCCIÓN	1
II. METODOLOGÍA	11
III. RESULTADOS	19
IV. DISCUSIÓN.....	31
V. CONCLUSIONES	35
VI. RECOMENDACIONES	36
REFERENCIAS.....	38
ANEXOS	44

Índice de tablas

Tabla 1: Metodología Seleccionada	10
Tabla 2: Indicador Productividad - Confiabilidad	15
Tabla 3: Indicador Eficiencia - Confiabilidad	15
Tabla 4: Descripción de Estadísticas de Eficiencia Pre y Post ML	19
Tabla 5: Descripción de Estadísticas de Productividad Pre y Post ML	20
Tabla 6: Eficiencia – Evaluación de Normalidad	22
Tabla 7: Productividad – Evaluación de Normalidad	23
Tabla 8: T - Student Eficiencia	26
Tabla 9: T - Student Productividad	29

Índice de figuras

Figura 1: Productividad Mensual	2
Figura 2: Eficiencia Mensual	2
Figura 3: Crisp DM	10
Figura 4: Distribución T - Student.....	17
Figura 5: Índice de Eficiencia Pre y Post ML	20
Figura 6: Índice de Productividad Pre y Post ML	21
Figura 7: Eficiencia Pretest	22
Figura 8: Eficiencia Post Test	23
Figura 9: Productividad Pretest	24
Figura 10: Productividad Post Test	24
Figura 11: Nivel de Eficiencia.....	25
Figura 12: T - Student Eficiencia.....	27
Figura 13: Distribución T - Student.....	27
Figura 14: Nivel de Productividad.....	28
Figura 15: T - Student Productividad.....	30
Figura 16: Distribución T - Student.....	30

Resumen

Esta investigación examina el impacto del Machine Learning (ML) en la gestión de ventas. El objetivo principal fue determinar la influencia del ML en la eficiencia y productividad de la gestión de ventas en una empresa manufacturera en Lima, Perú. Se realizó un estudio cuantitativo preexperimental con una muestra por conveniencia de 12 meses de datos de ventas de una población de 60 meses. Se implementó un algoritmo de árbol de decisión para la predicción de ventas. Los resultados mostraron un aumento en la eficiencia del 70% al 93%, y en la productividad del 69% al 86%. El análisis mediante la prueba t de Student confirmó la significancia estadística de estas mejoras ($p < 0.05$). Se concluye que el ML influye positivamente en la toma de decisiones para la gestión de ventas, mejorando la eficiencia operativa y la productividad del equipo comercial en el contexto empresarial peruano. Este estudio proporciona evidencia empírica sobre el potencial del ML para transformar las prácticas de gestión de ventas, ofreciendo una base sólida para futuras implementaciones de tecnologías avanzadas en procesos comerciales

Palabras clave: Machine learning, Gestión de ventas, Toma de decisiones, Eficiencia, Productividad

Abstract

This research examines the impact of Machine Learning (ML) on sales management. The main objective was to determine the influence of ML on the efficiency and productivity of sales management in a manufacturing company in Lima, Peru. A quantitative pre-experimental study was conducted with a convenience sample of 12 months of sales data from a population of 60 months. A decision tree algorithm was implemented for sales prediction. The results showed an increase in efficiency from 70% to 93%, and in productivity from 69% to 86%. Analysis using the student's t-test confirmed the statistical significance of these improvements ($p < 0.05$). It is concluded that ML positively influences decision-making for sales management, improving operational efficiency and productivity of the commercial team in the Peruvian business context. This study provides empirical evidence on the potential of ML to transform sales management practices, offering a solid foundation for future implementations of advanced technologies in commercial processes

Keywords: Machine learning, Sales management, Decision making, Efficiency, Productivity

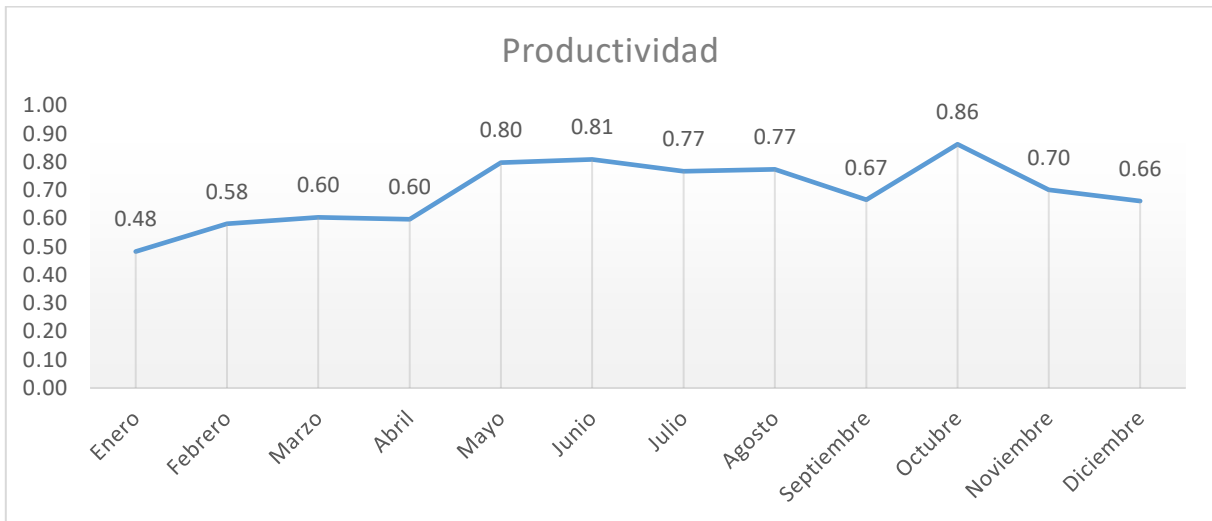
I. INTRODUCCIÓN

Los entornos organizacionales y empresariales se enfrentaron a un problema en común que trascendió los sectores: una gestión poco eficiente de los pronósticos de ventas y su falta de implementación. Según Ferro et al. (2023), la ejecución de la predicción de ventas en las organizaciones se realizó de forma manual y mediante métodos obsoletos, lo cual consumió recursos valiosos de la empresa, tanto en tiempo como en dinero. La importancia de abordar este tema radicó en la mejora tanto la eficiencia como la productividad en la gestión de venta.

Según Carhuavilca (2023), en su informe estadístico de la INEI publicado en febrero de 2024, la producción nacional en el año 2023 se vio afectada por una reducción del -0,55%, precisando en el sector manufactura una baja del -6,65% en el año 2023 y dando el cierre en el mes de diciembre con -10,92%, conservando un trayecto menguante por 11 meses continuos. Esto se debió principalmente a menores niveles de producción en la actividad fabril primaria (-2,77%) y del sector fabril no primario (-8,05%), así como a la disminución en otros sectores como la construcción, financiero, telecomunicaciones, agropecuario y pesca.

Por ello se realizó el análisis a una empresa manufacturera, donde el área de gestión de marketing se encargaba de realizar la medición de sus ventas mensualmente mediante reportes, en esta plataforma el personal, según el canal de venta, se encargaba de ingresar la cantidad de productos vendidos realizados, para que así el gerente comercial pueda descargar el formato en Excel, graficar y realizar sus comparativas con sus metas establecidas por el mismo en base a años anteriores. Este proceso provocaba tardía al momento de querer realizar un análisis del total de ventas realizadas en soles ya que quería evaluar la eficiencia de sus canales de ventas. Dichos gráficos eran elaborados de forma manual en Excel (hoja de cálculo) por el Gerente Comercial. Una vez realizados estos reportes el Gerente enviaba la información al área de operación para así realizar la toma de decisiones en cómo mejorar en sus ventas la productividad

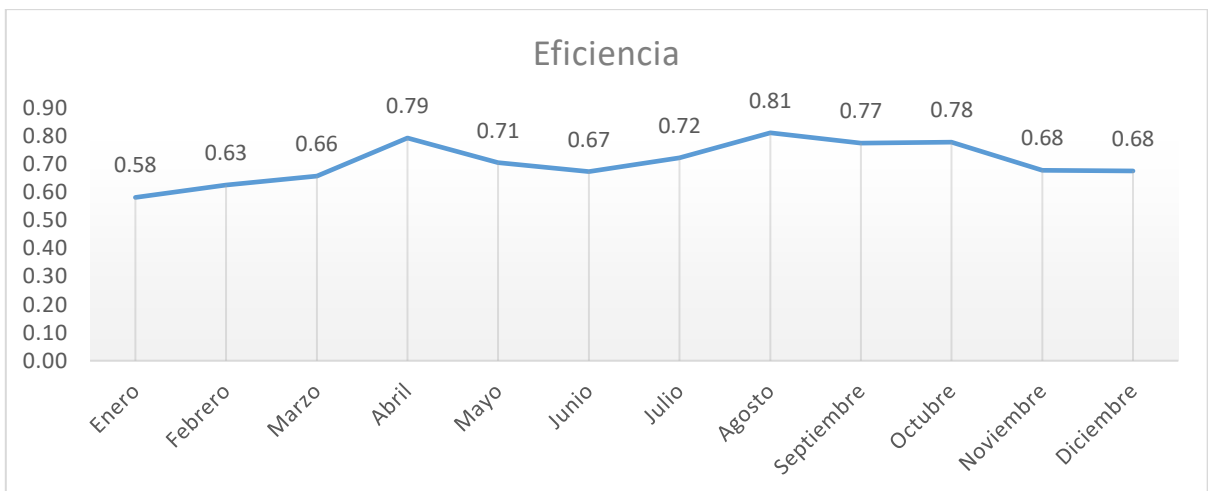
Figura 1: Productividad Mensual



Fuente de Elaboración Propia

El desarrollo de la Figura 1, se realizó a los datos de ventas establecidos por los canales para visualizar como el gerente comercial evaluaba la productividad de sus ventas, en el rango de tiempo de 12 meses, se observa que línea traza un número de ventas eficientes realizadas mensuales. Para el indicador de productividad, el pretest dio un resultado del 69% en el periodo de tiempo de enero del 2023 a diciembre del 2023.

Figura 2: Eficiencia Mensual



Fuente de Elaboración Propia

La Figura 2, se realizó con la data del gerente comercial para visualizar la eficiencia de venta en soles realizada por canal, en un periodo de tiempo de enero del 2023 a diciembre del 2023, en la cual se muestran picos altos y bajos en la eficiencia del canal. Para el indicador de eficiencia, el pretest dio un resultado del 70% en el periodo de tiempo de enero 2023 a diciembre del 2023

Sin embargo, a estas empresas les gustaría conocer el desarrollo productivo de sus ventas a futuro. Por ello, surgió la necesidad del uso de Machine Learning (ML) para optimizar la producción, mejorar la distribución y facilitar toma de decisiones. Los modelos predictivos basados en técnicas de Machine Learning (ML) tuvieron el potencial de proporcionar predicciones más precisas y robustas que los métodos tradicionales. Sin embargo, la aplicación de estas técnicas en la industria manufacturera peruana aún representó un reto.

Frente a este problema se realizó la siguiente pregunta: ¿De qué forma influyó el ML en la toma de decisiones para la gestión de ventas de una empresa manufacturera en Perú? Además, se plantearon las siguientes preguntas específicas: ¿De qué forma influyó el ML en la eficiencia para la toma de decisiones en la gestión de ventas de una empresa manufacturera en Perú? y ¿De qué manera influyó el ML en la productividad para la toma de decisiones en la gestión de ventas de una empresa manufacturera en Perú?

Por lo que la justificación metodológica del siguiente proyecto de investigación es de enfoque aplicado, por lo que se utilizó una metodología cuantitativa de tipo experimental. Para la recolección de la información, se utilizaron datos históricos de ventas de productos, los cuales se organizaron y utilizaron para realizar un pronóstico de las ventas futuras.

De igual forma, la incorporación de modelos de ML para pronosticar las ventas futuras demostró tener un impacto sólido en la justificación económica. Debido a que estos modelos brindaron la capacidad de pronosticar con mayor precisión las ventas futuras,

lo que permitió tomar decisiones fundamentadas en aspectos como la producción, la distribución y las estrategias de marketing, potenciando así la eficiencia operativa impactando positivamente en la rentabilidad y el rendimiento económico.

Por último, como justificación tecnológica, la presente investigación tuvo como finalidad desarrollar modelos de predicción, como árboles de decisión y regresión lineal, utilizando machine learning. Esto permitió un mejor análisis para tomar decisiones en los procesos de ventas. Es importante destacar que esta técnica permite relaciones distintas variables, utilizando una métrica de error que se reduce durante el proceso iterativo para efectuar las predicciones de manera precisa. Esto, a su vez, proporciona una mejor toma de decisiones.

Como objetivo general de investigación se establece, de qué forma influyó el Machine Learning en la toma de decisiones para la gestión de ventas de una empresa manufacturera en Perú. Los objetivos específicos fueron: establecer de qué forma influyó el Machine Learning en la eficiencia para la toma de decisiones de la gestión de ventas en una empresa manufacturera en Perú, y establecer de qué forma influyó el ML en la productividad de una empresa manufacturera en Perú en la gestión de ventas.

Así mismo Almeida et. al. (2023) en investigación tuvo como objetivo verificar si los modelos de ML pueden actuar como predictores del volumen de madera a transportar y apoyar en la toma de decisiones estratégicas. Los algoritmos de ML para la elaboración de los modelos utilizados en la investigación fueron un regresor de CatBoost, árbol de decisión, y K Neighbors para poder proporcionar a las métricas valores mínimos y máximos al coeficiente de determinación. Los resultados mostraron que los patrones de ML pueden actuar en la predicción del volumen de madera a transportar y apoyar la toma de decisiones estratégicas. El modelo de aprendizaje por pila, que combina los resultados de tres modelos de aprendizaje automático diferentes, obtuvo el mejor resultado, con un coeficiente de determinación de 0,70 y un error porcentual absoluto medio del 6%.

Según Hayder, I., Ghazwan , A. y Younis, H. (2023) nos dice que su objetivo de investigación fue ver cómo reaccionan los clientes al ofrecer ofertas a plazo fijo con la ayuda de su historial. Durante el proceso tuvo un problema de clasificación binaria donde debía predecir si el cliente aceptaría o rechazaría ofertas. Se probaron cuatro modelos de ML: k-vecinos más cercanos (K-NN), árbol de decisión, Naive Bayes y support vector machine (SVM). Los resultados demostraron que el árbol de decisión fue el algoritmo que obtuvo el mejor rendimiento, con una precisión del 91% frente a la precisión del 89% del algoritmo SVM. Respecto a los dos algoritmos restantes tuvieron un desempeño inferior en esta clasificación binaria.

Así mismo Saltos, D. y Villacis, Oscar. (2022) en su investigación tuvo como objetivo poder evaluar la factibilidad del pronóstico de ventas de la empresa. La metodología el cual se dio uso fue del aprendizaje no supervisado y usando el modelo de regresión lineal, procedieron a dar uso para ayudar e impulsar en el desempeño de las ventas realizadas. La data recopilada y utilizada fue netamente información brindada por la empresa (base de datos), la cual contiene 3113 registros que abarcan desde el año 2018 hasta mediados del 2021. Como resultados de la investigación se demuestra que la regresión lineal es un modelo estable debido a su tendencia hacia la normalidad entres sus datos con una salida del 73% de aciertos por lo cual demuestra que es un modelo confiable para poder realizar los respectivos pronósticos.

Por otro lado, Vargas (2023) buscó desarrollar un sistema web que emplea algoritmos de aprendizaje automático para pronosticar el rendimiento de los profesores universitarios. El estudio, de carácter cuantitativo y con un diseño cuasiexperimental, utilizó un conjunto de datos de 2449 registros para evaluar el desempeño docente. Se aplicó el algoritmo K-Means para agrupar los datos en tres clústeres óptimos, identificados mediante el método del codo de Jambú. En cuanto a los algoritmos de aprendizaje supervisado, se evaluaron varios modelos, incluyendo Regresión Logística (58.31% de precisión), Máquinas de Vectores de Soporte (SVM) (59.80%), Decision Tree Classifier (87.34%), Random Forest Classifier (90.57%), K-Neighbors Classifier (87.59%) y Gaussian NB (53.85%). El algoritmo Random Forest Classifier

fue seleccionado para el sistema web debido a su precisión superior del 90.57% en la etapa de prueba. La validación del modelo predictivo mediante Validación Cruzada (Cross Validation) resultó en un 91% de precisión, lo que confirma la solidez del modelo y su capacidad para minimizar el sesgo y la variabilidad en la predicción del rendimiento.

En el trabajo de investigación de LI, Ping et. al. (2024) nos menciona que la creciente competitividad en el mercado está haciendo que los principales fabricantes centren su atención en lograr una gestión de operaciones altamente eficiente y optimizada. Basado en esta situación, este documento toma máquinas expendedoras como objeto de investigación. Por otro lado, se construye un modelo de previsión de ventas de máquinas expendedoras basado en redes neuronales (NN). La investigación muestra que: (1) La precisión del algoritmo C4.5 puede alcanzar el 87 % en el máximo y el 68 % en el mínimo. Cuando se utiliza la red neuronal de retro propagación (BPNN) para predecir las ventas de máquinas expendedoras, la curva de los datos predichos coincide básicamente con la curva de los datos reales, lo que demuestra que su precisión es alta. El objetivo de este trabajo es construir un DSS cómodo y seguro tomando como ejemplo las máquinas expendedoras. Los resultados muestran que la precisión del algoritmo mejorado alcanza casi el 90 %, y el tiempo de ejecución mínimo se aproxima al de C4.5.

Según Di Mao, Rui et. al. (2024) en su trabajo de investigación menciona que los minoristas no cuentan con una orientación adecuada de cómo utilizar la analítica para interpretar la relación entre los atributos del producto y calidad del mismo. Es por ello que presentan un estudio basado en el enfoque de machine learning donde se valida la eficacia del marco con un conjunto de datos constituido por 1.599 muestras de vino tinto de la región portuguesa de Minho. El enfoque de la investigación mejora la precisión y solidez de los algoritmos tradicionales de aprendizaje automático mediante el desarrollo de un algoritmo híbrido. Este algoritmo se basa en el aprendizaje conjunto, un método que combina múltiples algoritmos de aprendizaje para obtener resultados más efectivos. Específicamente, el método de aprendizaje conjunto integra

diversos algoritmos, como SVM, la red neuronal recurrente (RNN), la máquina de aprendizaje extremo (ELM) y el bosque aleatorio (RF). Los resultados que se obtuvieron para el algoritmo de SVM tuvo una precisión del 67.16% para predecir la calidad de grado 5. Entre los algoritmos evaluados, el modelo de Random Forest destacó con una impresionante precisión general del 80,71%, demostrando su eficacia para manejar los desafíos predictivos del conjunto de datos.

Sin embargo, Yaranga, Italo (2022) en su investigación nos dice que el estudio realizado es preexperimental con un enfoque aplicado. La muestra consistió en 500 registros de la empresa Futech Perú SAC. Los resultados se fundamentan en indicadores de sensibilidad, precisión y especificidad, conseguida mediante la evaluación de cinco algoritmos de ML para una predicción óptima. Se observó que, usando árboles de decisión, el modelo logró una precisión del 82,6%, para K Nearest Neighbor un 86,8%, Naipes Bayes 94,6 %, Random Forest 96,2% y para el algoritmo SVM un 99,8% de precisión destacando que este modelo es el que logra una mejor predicción para detectar riesgos crediticios de clientes.

Gonzales, A y Rivera, S (2020) en su investigación nos dice que la gestión de ventas es una serie de procesos que están relacionados y que intervienen de manera directa para establecer una correcta relación entre la empresa y los clientes potenciales. En la gestión de ventas, crucial tener en cuenta varios factores clave: la planeación de ventas, que esta implica la preparación anticipada de las actividades futuras basadas en decisiones estratégicas y en la capacidad de adaptarse a cambios constantes; el proceso de venta, que abarca las actividades específicas de cada canal, desde el primer hasta transacción final; las capacitaciones son fundamentales para perfeccionar las habilidades, conocimientos y aptitudes de los consumidores, mejorando su productividad y la relación con los clientes; las estrategias de ventas.

Jiang, Qinhan (2022) en su estudio para grado de master titulado Métodos de Clasificación con Python: Aplicaciones Empresariales que se centró en la aplicación de métodos de clasificación mediante Python en contextos empresariales, se evidencia que un enfoque práctico demuestra que un árbol de decisiones tiene la capacidad de

segmentar los datos en diversos conjuntos, que en su vez se subdividen en subconjuntos adicionales, lo que aumenta el nivel de interrogantes hasta alcanzar una solución. Así mismo la ventaja más notable que ofrecen los árboles de decisión es su capacidad para explicar la clasificación y las decisiones a través de pruebas y ejemplos concretos. La construcción de un árbol de decisiones implica dividir los datos en conjuntos más pequeños, seleccionando las características más pertinentes para cada conjunto. Cada nodo interno corresponde a una característica del conjunto de datos, mientras que cada rama refleja una decisión. Por último, los nodos hoja representan los resultados potenciales de la clasificación.

Según El Naqa, I. y Murphy, M.J. (2015) en su libro nos dice que el ML es una rama fundamental de la inteligencia artificial que se basa en la capacidad de las máquinas para aprender de los datos sin requerir una programación explícita. Ha demostrado ser exitoso en diversas aplicaciones, incluyendo tecnologías como P.L.N (Procesamiento de Lenguaje Natural) y el reconocimiento de voz.

Por otro lado, Alan West (1991) en su libro nos habla sobre la gestión de ventas y el papel fundamental del equipo. Debido a que, en gran parte de las empresas, el equipo de ventas es parte fundamental ya que es la parte más visible, incluso más que el sistema de procesamiento del pedido. Por ello, todos los sectores de los mercados tienen exigencias muy concretas en la manera que las empresas organicen su esfuerzo en las ventas. Por lo cual definen técnicas de ventas, debido a que a medida que la compañía crece los retos aumentan, los conocimientos exigidos para que puedan actuar como vendedores de la misma.

Castañeda, Miguel (2020) en su investigación define la eficiencia como la capacidad de alcanzar un resultado eficiente sin malgastar recursos, esfuerzo, dinero o tiempo innecesariamente. Por ello se obtiene una óptima eficiencia cuando se genera una cantidad similar de servicios o productos a menor costo.

Sin embargo, Rodríguez, Diana (2022), nos comenta que la eficiencia y productividad son frecuentemente muy usados. Por ello los define como el menor nivel de elementos

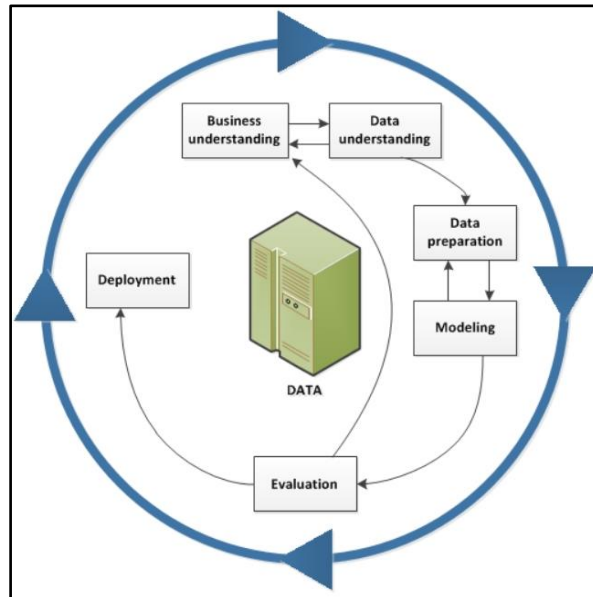
que se debe usar para lograr ejecutar las operaciones deseadas en un sistema productivo. La eficiencia es importante debido a que calcula la cantidad que se produce en base a costos por unidad, semanal o mensual producida y tiene que ver con la meta de mayor producción o menos que se imponga la empresa. Por lo que se trabajará en base a venta en soles mensuales / meta de venta en soles establecida.

Así mismo Castañeda, Miguel (2020) en base a la productividad nos dice que se basa en aprovechar la eficiencia de los recursos, se medirá por la productividad que se ha obtenido y los insumos que han utilizado. Esto va garantizar competitividad como la rentabilidad.

Por otro lado, Rodríguez, Diana (2022), nos habla sobre la productividad, la cual consiste en la relación que hay entre la elaboración de un bien o servicio y el conjunto de los insumos para producirlos. Por ello la productividad es calculada entre los productos de entradas utilizados para producir, y estas salidas o productos que describen el resultado. Por ello se establece como relación de Entrada - Salida que se basan en una conexión de proceso. Por ello nuestra productividad se mide entre la cantidad de productos vendidos mensuales / la meta de productos que se establece la empresa.

Respecto a la metodología según Pineda, C. (2022) en su libro menciona que CRISP-DM es una metodología es la más usada para proyectos de ML y Data Mining, siendo un modelo flexible y cíclico especialmente diseñado para ser adaptado por empresas y cubrir necesidad del proceso, es decir, dar prioridad en descubrir patrones o información útil el cual se encuentran en los datos

Figura 3: Crisp DM



Fuente: IBM

Tabla 1: Metodología Seleccionada

Expertos	EVALUACIÓN DE METODOLOGÍA			METODOLOGÍA SELECCIONADA
GALVEZ TAPIA, ORLEANS	21	35	28	CRISP DM
RIVERA CRISOSTOMO, RENNEE	16	32	19	CRISP DM
BERMEJO TERRORES, HENRRY	23	32	19	CRISP DM
TOTAL	60	99	66	

Fuente: Castañeda, Miguel (2020)

Cómo se logra observar en la tabla 1, la metodología en selección para el presente proyecto es CRISP-DM ya que presenta alto grado de similitud con los requerimientos planteados y el trabajo realizado con los datos, como se refleja en la encuesta realizada por Castañeda, Migue (2020), dicha metodología tuvo una aceptación del 99%, respeto a las otras. Así mismo, como hipótesis general se planteó la aplicación de ML para la mejora de la toma de decisiones en la gestión de ventas. Y como hipótesis específicas se detalló cómo el uso de ML aumentó tanto la eficiencia como la productividad en la gestión de ventas.

II. METODOLOGÍA

Tipo, enfoque y diseño de investigación

Según Castro, J., Gómez., Leidy, K. y Camargo, E. (2022) Una investigación aplicada es aquella cuyo objetivo es resolver problemas prácticos o específicos. Para ello, se basa en los conocimientos ya existentes en la investigación previa, que es el tipo de investigación que busca generar nuevos conocimientos, sin un objetivo práctico específico. A través de este estudio se identificó que el concepto de investigación aplicada ya que les brindó diversas características y enfoques de análisis: aquellos que pueden ir desde un punto de vista empresarial hasta por el académico y hasta el humanista. Por ello este estudio es de enfoque aplicado, por lo que nos basamos en conocimientos ya existentes en base a la data, por lo que se aplican conocimientos de aprendizaje automático a través de un algoritmo de aprendizaje de árbol de decisión para pronosticar las ventas.

Por ello se consideró para la investigación un enfoque cuantitativo, el cual Medina et al., (2023) menciona que para este tipo de investigación se recolectaron datos numéricos para analizar el fenómeno estudiado. Este método permite medir y cuantificar variables, facilitando así la realización de análisis estadísticos y la formulación de generalizaciones basadas en evidencia empírica.

De igual manera, el diseño adoptado en este estudio es de tipo experimental - preexperimental. Según Serrano et al. (2020), nos dice que este diseño permite a los investigadores analizar las variables para examinar las conductas resultantes de la investigación. En otras palabras, el propósito de este diseño es manipular los parámetros de las variables, tanto dependiente como independiente, para observar y medir los efectos de dicha manipulación.

$$G = O1 \rightarrow X \rightarrow O2$$

Pretest - Tratamiento – Post Test

En el cual:

G: Grupo experimental de estudio

X: Tratamiento o procesamiento

O1-O2: Mediciones de pretest / post Test de la solución de ML

Variables y operacionalización

Variable Independiente: Machine Learning (ML)

De acuerdo con Chopra y Khurana (2023) menciona que el ML es parte fundamental de la Inteligencia Artificial ya que se enfoca en el análisis e interpretación de estructuras y patrones presentes en los datos y la máquina es capaz de aprender, tomar decisiones y razonar con estos patrones y poder extraer soluciones inteligentes.”

Variable Dependiente: Gestión de Ventas

En el libro de John O’shaughnessy titulado Marketing Competitivo (1991), nos dice que en la gestión de ventas se basa en la estrategia, cuyo objetivo es determinar correctamente las cosas por realizar, determinar las guías por niveles más alto en base a su planificación en la empresa; mientras las decisiones de su eficiencia van relacionadas con aquellas operaciones que día a día son necesarias para los obtener objetivos propuestos en su estrategia. Por ello lo clasifican en estrategias de ventas, estrategias por clientes individuales y problemas operacionales que van relacionadas con la gestión de una fuerza de ventas.

Definición Operacional

Variable Independiente: Machine Learning

Según Lantz, Brett (2023) “Debido a que el aprendizaje automático, en esencia, se trata de convertir los datos en acción, las herramientas de explicabilidad ayudan a generar confianza en un modelo, lo que conduce a una mayor adopción y un mayor

impacto.”

Variable Dependiente: Gestión de ventas

Castañeda, Miguel (2020), nos habla sobre la gestión de ventas en la cual nos dice que aplicación de machine learning para su investigación dieron resultados positivos para su mejora en los procesos de ventas. En la cual establecieron la eficiencia y productividad como indicadores para su investigación. En la cual su eficiencia aumentó a un 82% y su productividad en un 81%.

Por ello teniendo en cuenta la gestión de ventas se establecen 2 dimensiones, primero la eficiencia la cual busca identificar el nivel de eficiencia de la predicción de ventas con las ventas reales realizadas. Y por último la productividad la cual busca identificar la gestión de ventas en base a la producción de avenida mensual con la cantidad vendida y potenciar la toma de decisiones.

Población, muestra y muestreo

Población

Sergio Jiménez (2022), en su libro titulado Investigación recogida de información de mercados nos habla de que la población viene hacer el conjunto total de elementos o individuos ya sea finitos o infinitos, de lo que se desea obtener alguna información. No obstante, puede tratarse de empresas, productos, personas o directamente un sector en concreto.

En base a lo establecido se considera que la población es de 60 meses el cual se contabiliza desde enero 2019 hasta diciembre del 2023.

Como criterios de inclusión se estima que la presente investigación incorpora datos provistos por la empresa en relación al sector comercial. Así mismo como criterios de exclusión no se está excluyendo ningún dato debido a que la información brindada en por la empresa es de ventas confirmadas por el área comercial. Por último, como delimitación geográfica se conoce que distribuyen productos a nivel nacional en distintos departamentos del Perú.

Muestra

Según Hernández et al. (2014, p 196) indica que “La muestra es un pequeño grupo que fue seleccionado de la población. Así mismo es un grupo que forma parte de la población identificados por atributos específicos.

Fórmula para hallar la muestra

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

Se establece que N (población), Z (Confianza 95%), P (Proporción 5%), Precisión (0.05), q (1-p (0.95)).

$$n = \frac{60 * 1.96^2 * 0.05 * 0.95}{0.05^2 * (60 - 1) + 1.96^2 * 0.05 * 0.95}$$

Muestreo

Según Hernández et al. (2014) El proceso de muestreo consiste en elegir una porción de un grupo más grande con el objetivo de reunir la suficiente información para abordar el problema planteado. La elección de las muestras depende de la cantidad de casos, ya que esto afecta directamente a las inferencias y al resultado final.

Se ha optado en la presente investigación en un muestreo por conveniencia, no probabilístico.

Técnica e instrumento de recolección de datos

La presente investigación titulada “Machine learning para la toma decisiones en la gestión de ventas, lima 2024” se emplea la técnica de cuestionario en la investigación (ver anexos). Además, se utilizan dos fichas de registro como instrumento de investigación.

Tabla 2: Indicador Productividad - Confiabilidad

		Productividad Pretest	Productividad Post Test
Productividad Pretest	Correlación de Pearson Sig. (bilateral) N	1 12	,810" ,001 12
Productividad Post test	Correlación de Pearson Sig. (bilateral) N	,810" ,001 12	1 12

Correlación entre las variables muestra una significancia estadística con un nivel de confianza del 0,001%.

Fuente: Elaboración Propia

La tabla desarrollada en SPSS, se observa un grado de confiabilidad muy aceptable, el cual quiere decir que el indicador es confiable para Productividad.

Tabla 3: Indicador Eficiencia - Confiabilidad

		Eficiencia Pretest	Eficiencia Post Test
Eficiencia Pretest	Correlación de Pearson Sig. (bilateral) N	1 12	,892" ,000 12
Eficiencia Post test	Correlación de Pearson Sig. (bilateral) N	,892" ,000 12	1 12

Correlación entre variables muestra una significación estadística con un nivel de confianza del 0,000%.

Fuente: Elaboración Propia

Procedimientos

Para recolectar los datos primero nos reunimos con el Gerente Comercial en la cual se le presentó el proyecto de investigación con los objetivos propuestos. se procedió a realizar la recolección de los datos. Por lo siguiente se realizó el fichaje con la información necesaria para la medición de los indicadores propuestos, así mismo poder realizar la implementación del proyecto de investigación propuesto la cual busca poder ayudar en la toma decisiones en base a la variable dependiente que es gestión de ventas.

Método de análisis de datos

Respecto a la validación de la hipótesis se empleó la prueba paramétrica de los rangos de signo de T-Student.

Según Valderrama, Santiago (2015) nos dice que la prueba T-Student tiene como finalidad verificar diferenciar de forma estadística que hay en los grupos con respecto a las variables consideradas. La presente prueba es importante ya que permite detectar si existe alguna diferencia en las entre las medias de los grupos, teniendo en cuenta la variabilidad de los datos.

Fórmula T - Student

$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{N}}}$	Donde: <ul style="list-style-type: none">• X es la media muestral• S es la desviación estándar muestral• N es el tamaño de la muestra
--	---

Región de Rechazo

La región de rechazo es $t = t_x$ donde t_x es tal que:

P [t > t_x] = 0.05 donde **t_x = Valor Tabular**

Luego Región de rechazo: **t > t_x**

Promedio

$$\bar{X} = \frac{\sum_{i=1}^n Xi}{n}$$

Desviación Estándar

$$s^2 = \frac{\sum_{i=1}^n (Xi - \bar{X})^2}{n - 1}$$

El gráfico que se muestra ilustra de manera clara la forma de la distribución T-Student. En él se pueden observar dos áreas principales: la zona donde no procedería a rechazarse la hipótesis nula (comúnmente llamada de aceptación) y la zona de rechazo. Estas áreas representan los diferentes resultados posibles al analizar una muestra estadística.

Figura 4: Distribución T - Student

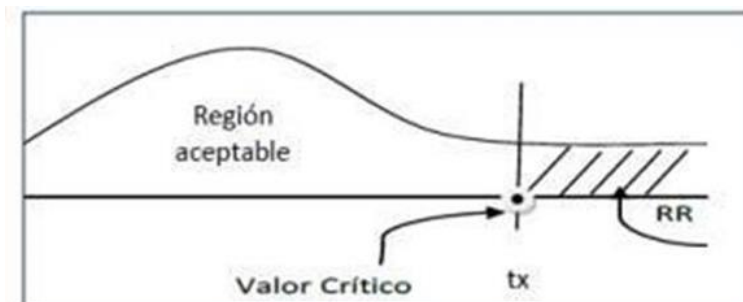


Gráfico de distribución

En la figura siguiente, la determinación en base si aceptar o rechazar la hipótesis nula dependerá de la ubicación del valor T en la figura siguiente: si T está ubicada en la región de rechazo, se aceptará la hipótesis alternativa; si T se encuentra en la región de no rechazo, se mantendrá la nula hipótesis.

Para el análisis de los datos obtenidos se empleará el software IBM SPSS v25, que proporcionará un enfoque tanto descriptivo como inferencial. En cuanto al análisis

descriptivo, se presentarán tablas y gráficos para cada uno de los indicadores. El presente análisis se llevará a cabo con la prueba de Shapiro-Wilk ya que, cuando se examinan muestras de 50 elementos o menos, se aplica una prueba paramétrica que mide la correlación de los datos correspondientes. (Juárez Manayay, 2021).

Aspectos Éticos

El presente informe de investigación se ha elaborado con medidas e instrucciones cruciales, en línea con aspectos legales y éticos. Se emplea información clave sobre la gestión de ventas de la empresa para integrar en el Sistema de Pronóstico. Los objetivos establecidos pretenden añadir un valor considerable al proceso de toma de decisiones, buscando potenciar tanto la productividad como la eficiencia en la gestión de ventas.

III. RESULTADOS

Esta investigación se estructuró en dos fases siguiendo un diseño preexperimental. La etapa inicial consistió en establecer una línea base mediante la medición de los indicadores clave antes de implementar los modelos de ML. Posteriormente, tras la aplicación del ML se procedió a registrar nuevamente los indicadores de productividad y eficiencia para evaluar el impacto.

En la siguiente tabla se visualizará los resultados descriptivos del indicador de Eficiencia:

Tabla 4: Descripción de Estadísticas de Eficiencia Pre y Post ML

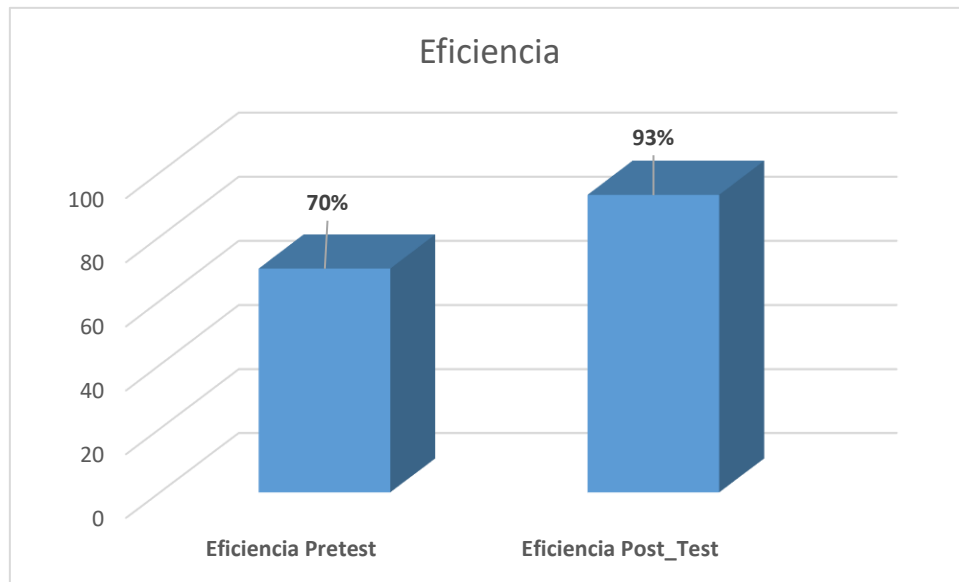
Descripción de Estadísticas						
	N	Mínimo	Máximo	Media	Desviación	Varianza
Pretest Eficiencia	12	0,58	0,81	0,70	0,070	0,005
Post Test Eficiencia	12	0,75	1,07	0,93	0,086	0,007
N Validos (Lista)	12					

Fuente: Elaboración Propia

La implementación del ML generó una mejora positiva en la gestión de ventas para la eficiencia, elevándose de 0.70 en la etapa previa a la implementación (pre-test) a 0.93 en la etapa posterior (post-test), tal como se muestra en la tabla. Este incremento se refleja también en el valor mínimo de eficiencia, que pasó de 0.58 a 0.75 tras la incorporación de esta tecnología.

En cuanto a la dispersión de los datos, se observó un ligero aumento en la variabilidad tras la implementación del ML, pasando de 0.005 en el pre-test a 0.007 en el post-test.

Figura 5: Índice de Eficiencia Pre y Post ML



Fuente: Elaboración Propia

Por lo siguiente, los resultados descripción de nuestro indicador de Productividad se observará en la presente tabla:

Tabla 5: Descripción de Estadísticas de Productividad Pre y Post ML

Descripción de Estadísticas						
	N	Mínimo	Máximo	Media	Desviación	Varianza
Pretest Productividad	12	0,48	0,86	0,69	0,114	0,013
Post Test Productividad	12	0,61	0,97	0,86	0,104	0,011
N Validos (Lista)	12					

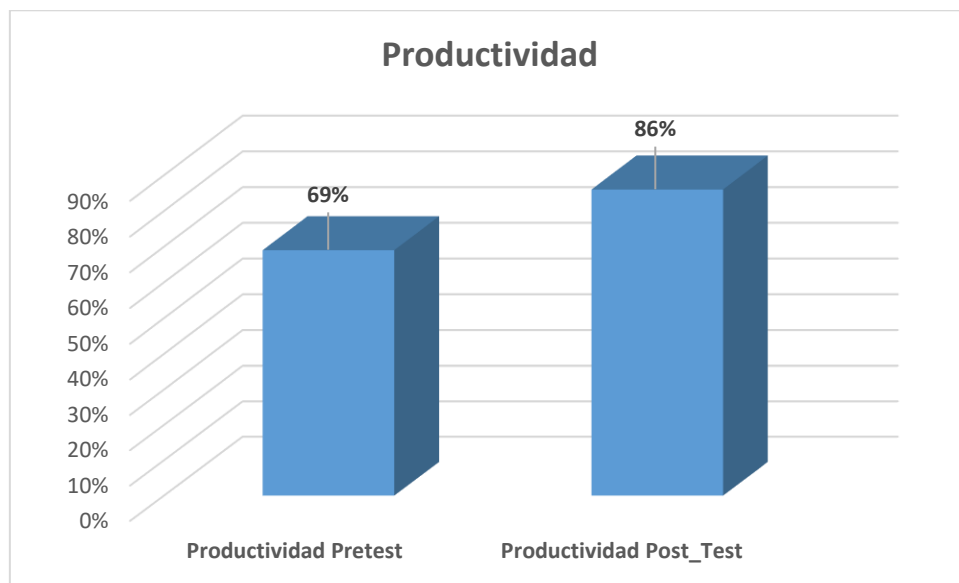
Fuente: Elaboración Propia

En cuanto a la productividad en la gestión de ventas, se observó un aumento significativo tras la implementación del ML. La media de productividad pasó de 0.69 en la etapa previa (pre-test) a 0.86 en la etapa posterior (post-test), como se aprecia en

la tabla. Este incremento también se refleja en la productividad mínima, que aumentó de 0.48 a 0.61 después de la implementación.

La variabilidad, medida como varianza, fue de 0.011 en el post-test, mostrando una ligera disminución respecto al valor de 0.013 obtenido en el pre-test.

Figura 6: Índice de Productividad Pre y Post ML



Fuente: Elaboración Propia

Se examinaron nuestros indicadores de productividad y eficiencia para evaluar su normalidad utilizando Shapiro-Wilk, dado que nuestra muestra (12 meses) eran menores a 50. Los datos se analizaron en SPSS 25.0 con un nivel de confianza del 95%. Se consideró que un valor de significancia (Sig.) menor a 0.05 indicaría una distribución no normal, mientras que un valor igual o superior a 0.05 sugeriría una distribución normal.

En caso la Significancia (SIG.) sea inferior a 0.05, se deduce que los datos no tienen una distribución normal

En caso la Significancia (SIG.) sea mayor a 0.05, se deduce que los datos llevan una distribución normal.

En el cual:

Significancia: P- es valor de contraste crítico.

Por lo que los resultados que se lograron obtener son los siguientes:

Con el fin de optar por la prueba estadística más apropiada, se analizó cuál era la distribución de los datos de eficiencia para definir si estos se ajustaban a una distribución normal.

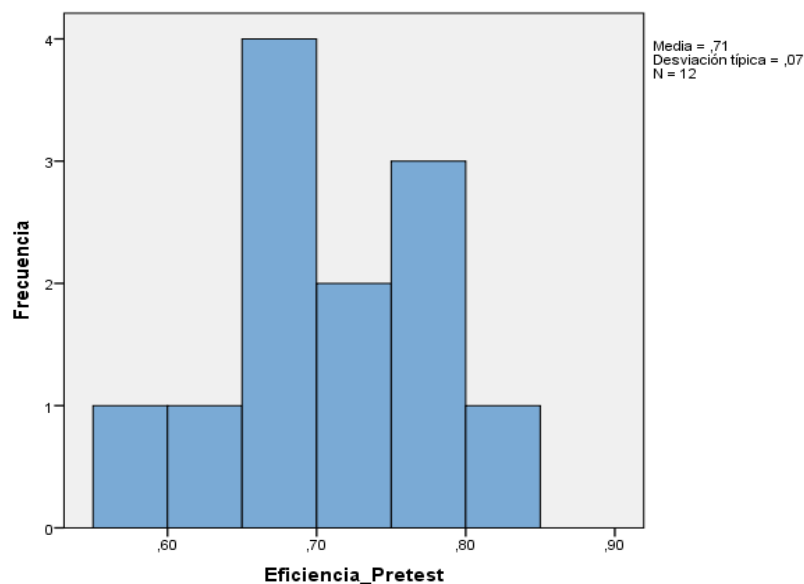
Tabla 6: Eficiencia – Evaluación de Normalidad

Shapiro - Wilk			
	Estadístico	gl	Sig.
Eficiencia Pretest	0,959	12	0,776
Eficiencia Post Test	0,961	12	0,793

Fuente: Elaboración Propia

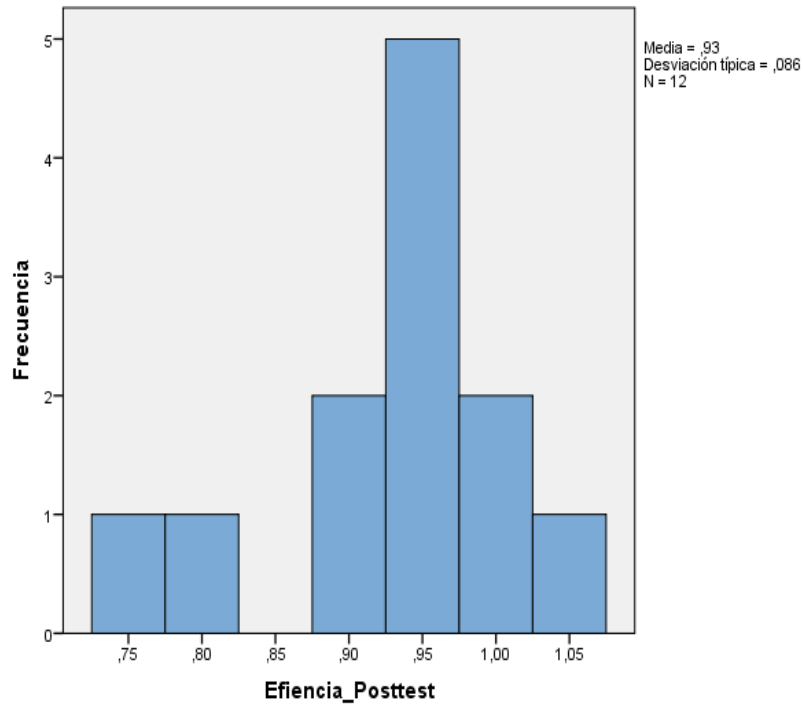
Una vez aplicado Saphiro Wilk, como se observa en la Tabla 6, muestran que la data de eficiencia presenta normalidad en su distribución tanto en la etapa previa (Sig. = 0.776) como posterior (Sig. = 0.793) a la implementación, dado que ambos valores de significancia superan el umbral de 0.05.

Figura 7: Eficiencia Pretest



Fuente: Elaboración Propia

Figura 8: Eficiencia Post Test



Fuente: Elaboración Propia

Con el objetivo de elegir la prueba estadística más adecuada, se examinó cómo estaban distribuidos los datos de productividad para así poder definir si se ajustaban a una normalidad en su distribución.

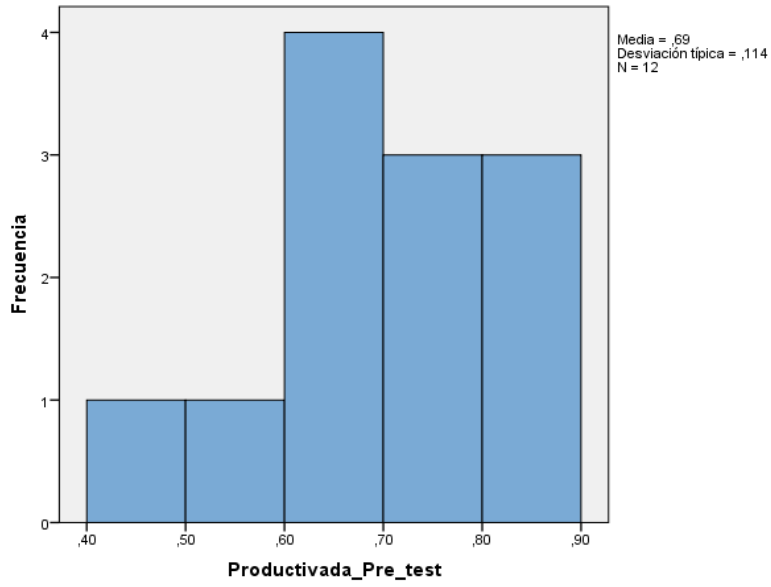
Tabla 7: Productividad – Evaluación de Normalidad

Shapiro - Wilk			
	Estadístico	gl	Sig.
Productividad Pretest	0,170	12	0,644
Productividad Post Test	0,154	12	0,098

Fuente: Elaboración Propia

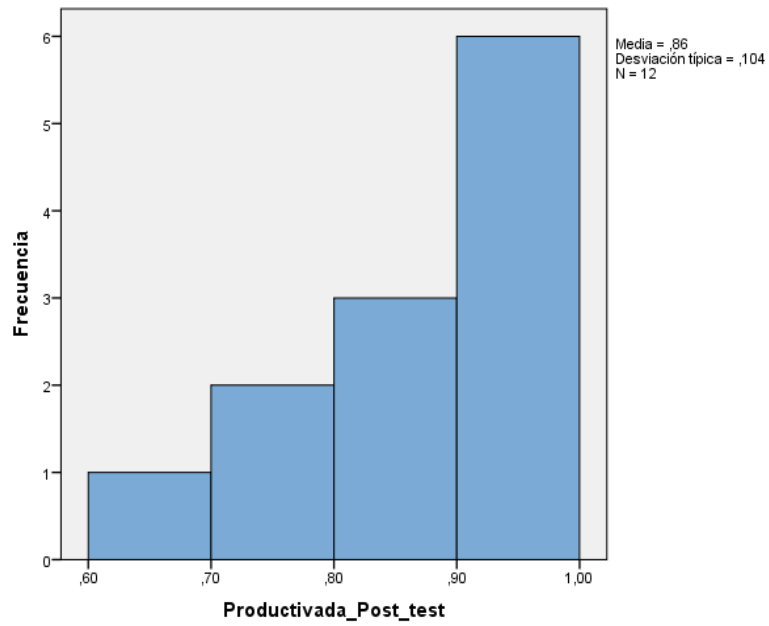
Los resultados de la Tabla N°7 revelan que una Sig. = 0.644) como en el post-test Sig. = 0.098, los valores de significancia son mayores a 0.05. Esto indica la normalidad de los datos en ambos casos en base a la de productividad.

Figura 9: Productividad Pretest



Fuente: Elaboración Propia

Figura 10: Productividad Post Test



Fuente: Elaboración Propia

H1: Indica que ML mejora la gestión de ventas en el indicador de Eficiencia.

Indicador: Eficiencia

Hipótesis Estadística

Variables

- **EFa:** Eficiencia en la gestión de ventas antes de utilizar ML.
- **EFd:** Eficiencia en la gestión de ventas después de utilizar ML.

H0: Establece que no existe un aumento en la eficiencia tras la implementación del ML.

$$H0: EFa \geq EFd$$

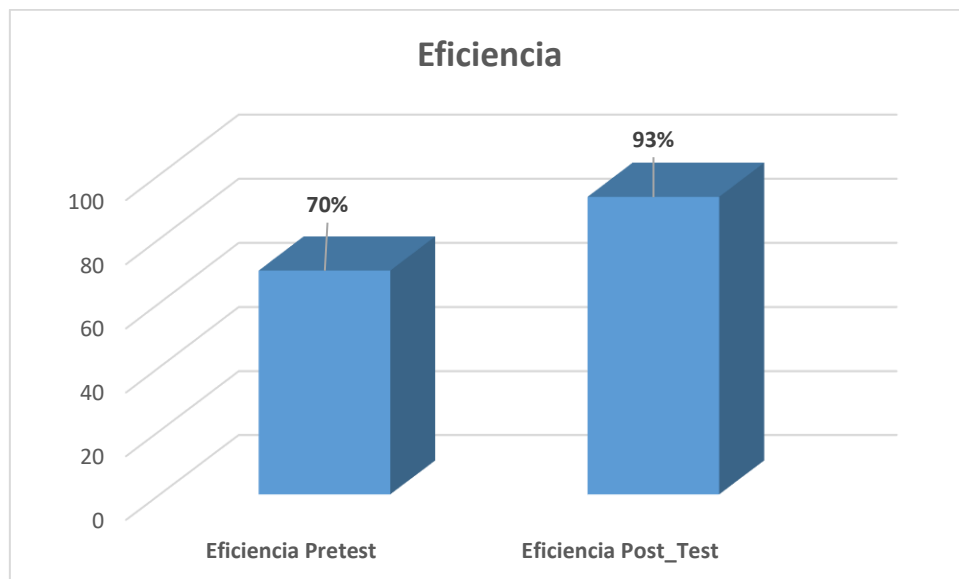
Sin ML con el indicador de Eficiencia no es mejor que ML con el indicador de Eficiencia.

HA: Sostiene que la eficiencia sí aumenta con el uso de estas tecnologías

$$HA: EFa < EFd$$

El indicador de Eficiencia con el ML es superior que el indicador de eficiencia sin ML.

Figura 11: Nivel de Eficiencia



Fuente: Elaboración Propia

Se observa en la figura 13 un incremento notable en la eficiencia, pasando del 70% al

93%.

Los datos tanto en el Pretest y Post Test muestran una distribución normal, se aplicó T-Student para analizar la hipótesis. El valor T calculado de -8.651, siendo inferior al valor crítico de 1.7959, respalda el rechazo de la hipótesis nula. (visualizar tabla N°3).

Tabla 8: T - Student Eficiencia

Media	Evaluación T Student		
Eficiencia Pretest	T	gl	Sig. (bi)
Eficiencia Post Test	-8,651	11	,000

Fuente: Elaboración Propia

Formula T Student Aplicada

$$\frac{d}{S / \sqrt{n}}$$

$$tc = \frac{0.7067 - 0.9317}{0.0901/\sqrt{12}}$$

$$tc = \frac{-0.225}{0.0901/3.464}$$

$$tc = \frac{-0.225}{0.026}$$

$$tc = -8.65$$

El cálculo realizado mediante la fórmula T de Student permite rechazar con un 95% de confianza la hipótesis Nula, por lo cual implica aceptar la hipótesis alternativa. El valor T obtenido, según se observa en la Figura N°14, está en la región de rechazo, lo cual confirma que el aprendizaje automático aumenta la eficiencia en la gestión de ventas.

Figura 12: T - Student Eficiencia

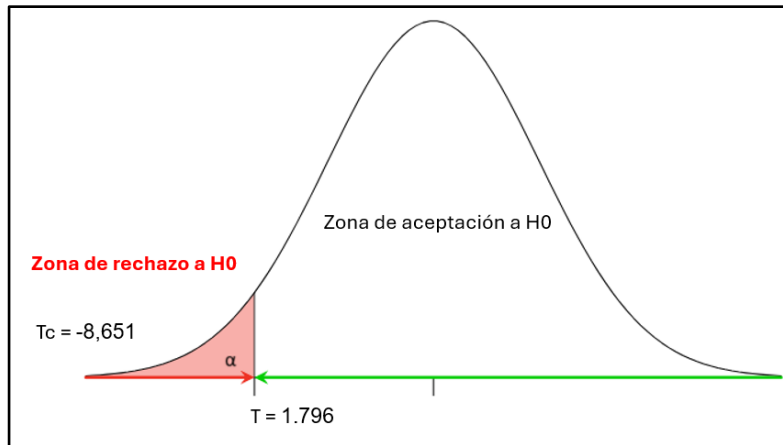


Figura 13: Distribución T - Student

<i>r</i>	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576

Fuente: Elaboración Propia

H2: El ML mejora la productividad en la gestión de ventas.

Indicador: Productividad

Hipótesis Estadística

Variables

PRa: La productividad antes de usar ML

PRd: La Productividad luego de usar ML

H0: ML no aumenta la productividad en la gestión de ventas.

HA: $PRa < PRd$

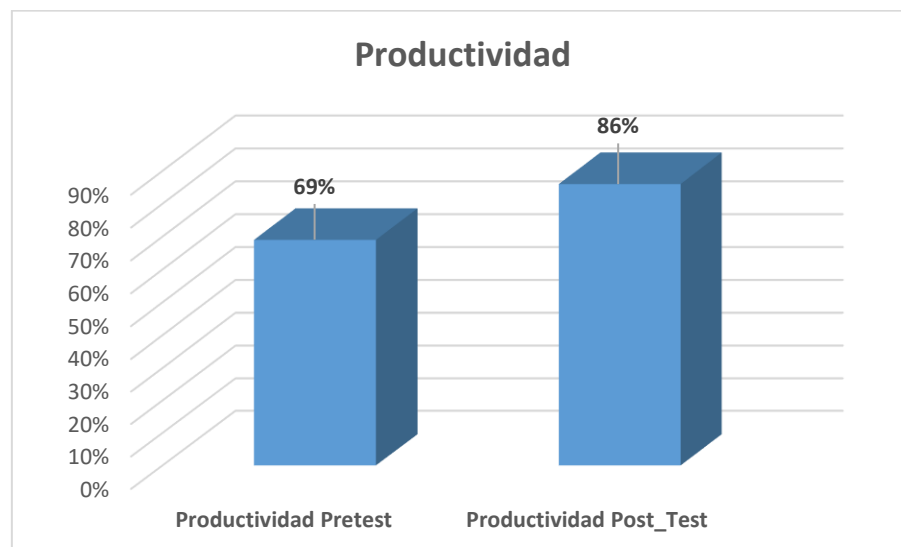
El indicador de Productividad sin el ML es mejor que el indicador de eficiencia con ML.

HA: Como hipótesis Alterna tenemos que ML mejor la productividad para la gestión de ventas.

HA: $PRa > PRd$

El indicador de Productividad con el ML es mejor que el indicador de eficiencia sin ML.

Figura 14: Nivel de Productividad



Fuente: Elaboración Propia

La figura ilustra un incremento notable en la productividad, que se eleva del 69% al 86%.

Dado que los datos de productividad muestran que en el pretest como en el post-test exhibieron una distribución normal, se empleó la prueba T de Student para muestras relacionadas con el objetivo de definir si la implementación del ML generó una diferencia estadísticamente significativa en la productividad. El valor T calculado de -3.634, al ser inferior al valor crítico de 1.7959, confirma la existencia de dicha diferencia significativa. (visualizar tabla N°4).

Tabla 9: T - Student Productividad

Media		Evaluación T Student	
Productividad Pretest	T	gl	Sig. (bi)
Productividad Post Test	-3,62	11	,004

Fuente: Elaboración Propia

Formula T Student Aplicada

$$tc = \frac{d}{S / \sqrt{n}}$$

$$tc = \frac{0.6917 - 0.8637}{0.1642 / \sqrt{12}}$$

$$tc = \frac{-0.172}{0.1642 / 3.464}$$

$$tc = \frac{-0.172}{0.0474}$$

$$tc = -3.62$$

El cálculo realizado mediante la fórmula T de Student permite rechazar la hipótesis nula con un 95% de confianza, lo que implica aceptar la hipótesis alternativa. El valor

de T obtenido, ubicado en la zona de rechazo de la figura N°18, lo cual confirma que la aplicación del ML aumenta significativamente la productividad en la gestión de ventas.

Figura 15: T - Student Productividad

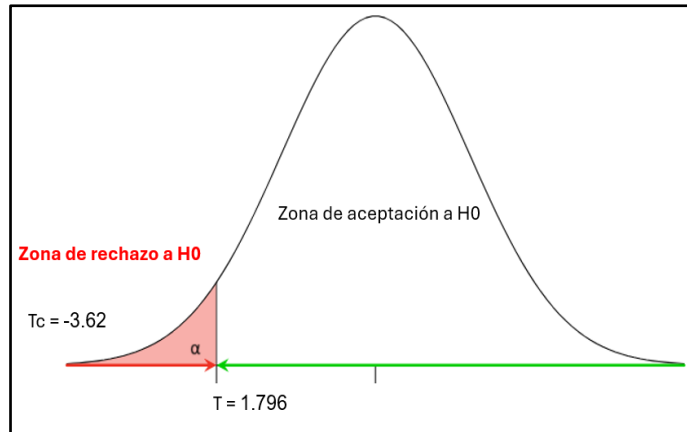


Figura 16: Distribución T - Student

r	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576

Fuente: Elaboración Propia

IV. DISCUSIÓN

La presente investigación los resultados obtenidos muestran que el ML incrementa la eficiencia de 70% a 93% teniendo así un aumento de 23% y respecto a la productividad de 69% a 86% aumentando así un 17% conforme a los plazos marcados por la gestión de ventas.

Para evaluar los datos se utilizó Shapiro-Wilk, así determinamos su normalidad. Dado que los datos eran normales, se procedió a aplicar la prueba t de Student para evaluar si había una diferencia estadísticamente significativa entre las medias de los grupos.

Se detalla que la significancia resulta un 0,001. De tal manera se rechazan las hipótesis nulas (H0) y se aceptan las alternas (H1) en los indicadores planteados.

De la misma manera Almeida et. al. en su investigación llegó a concluir que el ML influye de manera *significativa* en la predicción del volumen de madera a transportar donde los resultados de tres modelos de aprendizaje automático diferentes destacan el aprendizaje por pila el cual obtuvo el mejor resultado, con un coeficiente de determinación de 0,70 y un error porcentual absoluto medio del 6%.

En el Primer indicador Eficiencia se obtuvo una mejora en la eficiencia en un 23% en comparación con el pretest. Los autores Hayder, I., Ghazwan , A. y Younis, H. (2023) usan un modelo de predicción mediante algoritmos de aprendizaje automático el cual utilizarían cuatro modelos de machine learning los cuales son K-NN, árbol de decisión, Naive Bayes y SVM. Usaron estos modelos para ver cómo reaccionan los clientes al suscribir depósitos u ofertas a plazo fijo con la ayuda de su historial dando así resultados favorables el cual muestra que el algoritmo que tiene mejor rendimiento es el de árbol de decisión con una precisión del 91% destacando por encima del algoritmo SVM con un 89%. En cambio, en nuestra investigación aplicamos el algoritmo de árbol de decisión para ver/medir la cantidad de las ventas por canal dándonos así un resultado de 93% de eficiencia el cual muestra una mejora moderada aplicando este modelo de machine learning frente a un 70% de eficiencia medida en el pre-test.

Los autores Saltos, D. y Villacis, Oscar. En su trabajo de investigación aplican la metodología del aprendizaje no supervisado utilizando el algoritmo de regresión lineal donde se busca analizar la viabilidad del pronóstico de ventas de la empresa. Respecto a los datos se tomaron de la base de datos los cuales constan de 3113 desde 2018 al 2021. La aplicación de este modelo les dio como resultado que el algoritmo de regresión lineal es estable, esto se debe a su tendencia a la normalidad de sus datos teniendo una salida del 73% el cual se afirma que es un modelo confiable para poder realizar los respectivos pronósticos. En contraste a nuestro trabajo de investigación se aplicó el método del aprendizaje supervisado usando el algoritmo de Decisión Tree (DT) donde se busca la predicción de nuestros indicadores si influye en la toma de decisiones para gestión de ventas. Con relación a los datos se tomaron de 60 meses el cual se contabiliza desde enero 2019 hasta diciembre del 2023. El resultado obtenido fue que respecto hay un aumento en la eficiencia aplicando el algoritmo de árboles de decisión el cual de un 70% asciende a un al valor de 93%. Así mismo respecto al resultado de productividad de un 69% asciende a un valor de 86 %. De tal manera se puede notar que la aplicación de este algoritmo de machine learning influye considerablemente en ambos indicadores planteando así que apoya en la toma de decisiones de la empresa en base a su eficiencia y productividad.

En el estudio de LI, Ping (2024) aplica el modelo de clasificación de árbol de decisión, además utiliza el algoritmo de Redes Neuronales, de esta manera se busca predecir las ventas de máquinas expendedoras. Los resultados muestran que la precisión del algoritmo mejorado alcanza casi el 90 %, y el tiempo de ejecución mínimo se aproxima al de C4.5. El tiempo de ejecución del algoritmo C4.5 mejorado en distintos conjuntos de datos es menor que el del algoritmo sin modificar. Con respecto a nuestra investigación se aplicó el algoritmo de árbol de decisión siendo efectiva en la toma de decisiones para la proactividad y eficiencia que se manejen en empresas u organizaciones.

En su trabajo de investigación, Di Mao, Rui, et al. (2022) implementaron un enfoque

de aprendizaje conjunto que integra diversos algoritmos, incluyendo SVM, RNN, ELM y RF, para mejorar la precisión y robustez de las predicciones de calidad del producto. Utilizando un conjunto de datos de 1.599 muestras de vino tinto, demostraron que el modelo de Random Forest alcanzó una precisión del 80,71%, superando significativamente a otros algoritmos como SVM, que obtuvo una precisión del 67,16%. En nuestro caso respecto al enfoque de Di Mao, Rui, et al. utiliza un método de aprendizaje conjunto más complejo que incluye múltiples algoritmos, el Random Forest destacó con una precisión general del 80,71%. En comparación, los árboles de decisión, el cual se hace uso en el presente trabajo, aunque tal vez no alcancen la misma precisión en todos los casos, ofrecen una ventaja significativa en términos de interpretabilidad y facilidad de uso. Esto los convierte en una herramienta preferida para la toma de decisiones en entornos empresariales donde la comprensión clara y la implementación rápida son cruciales.

Para Yaranga, Italo (2022) el cual evaluó cinco algoritmos de machine learning utilizando una muestra de 500 registros de clientes. Los resultados demostraron diferentes niveles de precisión para cada algoritmo: árboles de decisión alcanzaron una precisión del 82,6%, K Nearest Neighbor (KNN) un 86,8%, Naive Bayes un 94,6%, Random Forest un 96,2%, y SVM un 99,8%. Al comparar estos estudios con el nuestro trabajo de investigación se resalta que el del algoritmo de árboles de decisión en la gestión de ventas, se destaca que lo importante es escoger el algoritmo adecuado para cada contexto específico. Los árboles de decisión ofrecen ventajas significativas en términos de interpretabilidad y simplicidad, permitiendo a los gestores de ventas identificar factores clave y tomar decisiones informadas rápidamente.

Según Gonzales, A y Rivera, S donde nos dice que la gestión de ventas son una serie de procesos que están relacionados y que intervienen de manera directa para establecer una correcta relación entre la empresa y los clientes potenciales. En la gestión de ventas, es crucial tener en cuenta varios factores clave: la planeación de ventas, el proceso de venta, las capacitaciones, la relación con los clientes. y por último la disponibilidad. Para nuestro trabajo de investigación la gestión de ventas al igual

que Gonzales y Rivera nos enfocamos en la correcta intervención entre la empresa y los clientes potenciales con la adición del machine learning para lograr encontrar una correcta gestión de ventas.

El autor Castañeda en su trabajo de investigación la cual busca determinar la influencia del aprendizaje supervisado para la gestión de ventas utilizando los algoritmos de aprendizaje supervisado los cuales son Regresión Lineal, Regresión Lineal Simple y Redes Neuronales donde la metodología la cual trabajó fue el de CRISP DM. Los datos se recopilan a través de registros y encuestas. Los resultados alcanzados fueron que el ML en el desarrollo de la gestión de ventas aumentó la eficiencia en un 27% del 55% al 82% y la productividad en un 29% del 52% al 81%. A diferencia de nuestro trabajo de investigación, también utilizamos la eficiencia y la productividad como variables en nuestro caso el modelo de Decision Tree de para la aplicación y realizar los pronósticos de las ventas y así ayudar a tener un control en la toma d decisiones de la producción y/o eficiencia del canal de ventas.

V. CONCLUSIONES

La presente investigación concluye con que su objetivo general es establecer la influencia del Machine Learning (ML) en la toma de decisiones para la gestión de ventas. Los resultados obtenidos demuestran un impacto positivo y significativo en dos dimensiones clave: eficiencia y productividad.

Objetivo Específico 1: Influencia del Machine Learning en la eficiencia de la toma de decisiones para la gestión de ventas

El análisis estadístico revela un aumento del 32.86% en la eficiencia de la gestión de ventas tras la implementación del ML, pasando de una media de 0.70 a 0.93. Este incremento, respaldado por la prueba t de Student ($t(11) = -8.651, p < 0.001$), confirma la eficacia del ML para optimizar los procesos de toma de decisiones y alcanzar las metas de ventas de manera más efectiva.

Objetivo Específico 2: Influencia del Machine Learning en la productividad de la toma de decisiones para la gestión de ventas

La productividad también experimentó un aumento considerable del 24.64% tras la implementación del ML, con una media que pasó de 0.69 a 0.86. Este resultado, validado por la prueba t de Student ($t(11) = -3.62, p = 0.004$), demuestra la capacidad del ML para potenciar el rendimiento del equipo de ventas y alcanzar mayores volúmenes de ventas en relación con las metas establecidas.

Objetivo General: Influencia del ML en la toma de decisiones para la gestión de ventas

En conjunto, los incrementos del 32.86% en eficiencia y del 24.64% en productividad evidencian la influencia positiva del ML en la toma de decisiones para la gestión de ventas. El ML ha demostrado ser una herramienta eficaz para optimizar los procesos comerciales, mejorar el desempeño del equipo de ventas y alcanzar los objetivos planteados.

VI. RECOMENDACIONES

1. Profundización del objeto de estudio:

Para abordar la problemática de una gestión de ventas poco eficiente y mejorar tanto la productividad como la eficiencia, se recomienda:

- Realizar un estudio longitudinal que abarque un período más extenso, idealmente de 3 a 5 años. Esto permitiría evaluar la consistencia a largo plazo de las mejoras en productividad y eficiencia obtenidas mediante el uso de Machine Learning (ML) en la gestión de ventas (Carhuavilca, 2023).
- Expandir el estudio para incluir diferentes sectores además del manufacturero, permitiendo comparar cómo el ML impacta la productividad y eficiencia en diversos contextos empresariales (Ferro et al., 2023).

2. Recomendaciones metodológicas para futuros investigadores:

- Implementar un diseño experimental completo con grupo de control, comparando directamente la productividad y eficiencia de equipos de ventas que utilizan ML frente a los que no lo hacen. Esto proporcionaría evidencia más robusta sobre el impacto del ML en estos indicadores clave (Vargas, 2023).
- Explorar y comparar diferentes algoritmos de ML además del árbol de decisión utilizado. Por ejemplo, investigar el rendimiento de Random Forest o redes neuronales en la predicción de ventas y su impacto en la productividad y eficiencia (Yaranga, 2022).
- Incorporar análisis de series temporales junto con ML para mejorar la precisión de las predicciones de ventas, lo cual podría aumentar aún más la eficiencia en la gestión de inventarios y recursos (Saltos & Villacis, 2022).

3. Recomendaciones técnicas para empresas:

- Implementar un sistema de recolección de datos en tiempo real que integre información de ventas, inventario y productividad de los vendedores. Esto proporcionará una base más sólida para el entrenamiento de modelos de ML, potencialmente mejorando aún más la eficiencia y productividad (Jiang, 2022).
- Desarrollar un dashboard que muestre en tiempo real los indicadores de productividad y eficiencia, junto con las predicciones del modelo de ML. Esto permitirá a los gerentes de ventas tomar decisiones más informadas y ágiles (Gonzales & Rivera, 2020).
- Establecer un programa de capacitación continua para el personal de ventas

sobre cómo interpretar y actuar sobre las predicciones del ML para maximizar su productividad y eficiencia (Alan West, 1991).

4. Recomendaciones normativas:

- Establecer políticas claras sobre cómo se medirán y utilizarán los indicadores de productividad y eficiencia en conjunción con las predicciones del ML, asegurando un uso ético y transparente de estos datos (El Naqa & Murphy, 2015).
- Implementar un protocolo de revisión regular del modelo de ML, evaluando su impacto continuo en la productividad y eficiencia. Esto asegurará que el modelo siga siendo relevante y efectivo a lo largo del tiempo (Pineda, 2022).

5. Recomendaciones para la mejora continua del conocimiento científico:

- Fomentar la colaboración entre academia e industria para desarrollar métricas más sofisticadas de productividad y eficiencia en ventas que puedan ser mejoradas mediante ML. Esto podría llevar a insights más profundos sobre cómo la tecnología puede optimizar estos aspectos clave del rendimiento en ventas (Chopra & Khurana, 2023).
- Realizar estudios comparativos entre diferentes industrias para identificar factores contextuales que influyen en la efectividad del ML para mejorar la productividad y eficiencia en ventas. Esto podría llevar al desarrollo de mejores prácticas adaptadas a diferentes escenarios empresariales (LI et al., 2024).

REFERENCIAS

- Almeida RO, Munis RA, Camargo DA, da Silva T, Sasso Júnior VA, Simões D. Prediction of Road Transport of Wood in Uruguay: Approach with Machine Learning. *Forests*. 2022; 13(10):1737. Disponible en: <https://doi.org/10.3390/f13101737>
- BARRY, Jay. Principios de Administración de Operaciones [en línea] México: Pearson Educación. 2004 [Fecha de consulta: 13 de noviembre de 2023] Capítulo 4: Siete pasos en el sistema de pronóstico. Disponible en: https://www.google.com.pe/books/edition/Principles_of_Operations_Management/jVlwSsVHUfAC?hl=es&gbpv=1&dq=sistema+de+pronosticos&pg=PA106&printsec=frontcover ISBN: 9702605253
- BCRP . ACTIVIDAD ECONÓMICA: SETIEMBRE 2023 [en línea]. Noviembre 2023, n° 84. [Fecha de consulta: 20 de abril de 2024]. Disponible en : <https://www.bcrp.gob.pe/docs/Publicaciones/Notas-Estudios/2023/nota-de-estudios-84-2023.pdf>
- Carhuavilca , Dante, et al. Producción Nacional [en línea]. Febrero 2024, n° 2. [Fecha de consulta: 20 de abril de 2024]. Disponible en : <https://m.inei.gob.pe/media/MenuRecursivo/boletines/02-informe-tecnico-produccion-nacional-diciembre-2023.pdf>
- Carhuavilca , Dante, et al. Producción Nacional [en línea]. Junio 2023, n° 6. [Fecha de consulta: 20 de abril de 2024]. Disponible en : <https://m.inei.gob.pe/media/MenuRecursivo/boletines/06-informe-tecnico-produccion-nacional-abr-2023.pdf>
- CARLOS, Pineda. Aprendizaje automático y profundo en python [en línea]. Bogotá - Colombia 2022. Pag 45, Disponible en: https://www.google.com.pe/books/edition/Aprendizaje_autom%C3%A1tico_y_profundo_en_py/mgNcEAAAQBAJ?hl=es-419&gbpv=0&kptab=overview ISBN: 9789587923155
- CASTAÑEDA, Miguel. Machine Learning para la gestión de ventas en la Empresa Vertice Empresarial S.A.C, Tesis (Licenciado en Ingeniería de Sistemas). Lima: Universidad César Vallejo, 2020. Disponible en: <https://hdl.handle.net/20.500.12692/56129>

CASTRO, John., GÓMEZ, Leidy y CAMARGO, Esperanza, La investigación aplicada y el desarrollo experimental en el fortalecimiento de las competencias de la sociedad del siglo XXI. Revista colombiana de ingeniería [en línea]. Julio-septiembre 2022 [Fecha de consulta: 5 de octubre de 2023] Disponible en: <https://doi.org/10.14483/22487638.19171> ISSN 2248-7638

CHOPRA, Deepti, ROOPAL K. Introduction to Machine Learning with Python [en línea]. India: Bentham Science Publishers, 2023 [fecha de consulta: 18 de abril de 2024]. Capitulo 2. Introduction To Machine Learning. Disponible en : <https://ebookcentral.proquest.com/lib/biblioucv/detail.action?docID=304107021> SBN: 9789815124422

EL NAQA, Issam y MURPHY, Martin. Machine Learning in Radiation Oncology Revista USA de Medicina e Ingeniería [en línea]. Enero-diciembre 2015 [Fecha de consulta: 21 de octubre de 2023] Disponible en: https://doi.org/10.1007/978-3-319-18305-3_1

Estadística, población, muestra y variables. [Blog]. Lima: Durand, J., (19 de abril de 2014). [Fecha de consulta: 15 de noviembre de 2023]. Recuperado de: <https://es.slideshare.net/jcarlos2509/estadistica-poblacion-muestra-y-variables>

FERRO, Rubén, COSSIO, Gonzalo y FERNANDÉZ, Nicolas. Pronóstico de ventas a través de una aplicación web aplicando machine learning y variables exógenas. [en lí-nea] Bogotá : Escuela Colombiana de Ingeniería Julio Garavito, 202393 páginas . Disponible en : <https://repositorio.escuelaing.edu.co/handle/001/2815>

GONZALES, Asshly e RIVERA, Susan. Sales management in the pharmaceutical company GR Farmanova S.A.C. from the district of Villa El Salvador, period 2019. Tesis (Bachiller en Contabilidad). Lima: Universidad Autónoma del Perú, 2021. Disponible en: <https://hdl.handle.net/20.500.13067/1228>

GONZALEZ, Asshly y RIVERA, Susan. Gestión de ventas en la empresa farmacéutica GR Farmanova S.A.C. del distrito de Villa El Salvador, periodo 2019. Tesis (Bachiller en Contabilidad). Lima: Universidad Autónoma del Perú, 2021 Disponible en: <https://hdl.handle.net/20.500.13067/1228>

Haenlein, M., & Kaplan, A. M. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, 61(4), 5-14. Disponible en: https://www.researchgate.net/publication/334539401_A_Brief_History_of_Artificial_Intelligence_On_the_Past_Present_and_Future_of_Artificial_Intelligence

- Hayder, Israa., NABI, Ghazwan y YOUNIS, Hussain, Predicting Reaction Based on Customer's Transaction using Machine Learning Approaches. Revista Indonesa de ingeniería eléctrica e informática [en línea]. Junio-septiembre 2022 [Fecha de consulta: 1 de noviembre de 2023] Disponible en: https://www.researchgate.net/publication/367762092_Predicting_reaction_based_on_customer's_transaction_using_machine_learning_approaches
- HERNÁNDEZ, Roberto., FERNÁNDEZ, Carlos y Baptista, María del Pilar. Metodología de la investigación (6° ed.). México: McGraw Hill Interamericana Editores S.A. de C.V [Fecha de consulta: 28 de noviembre de 2023] Capítulo 8: Selección de la muestra. Disponible en: https://apiperiodico.jalisco.gob.mx/api/sites/periodicooficial.jalisco.gob.mx/files/metodologia_de_la_investigacion_-_roberto_hernandez_sampieri.pdf
- HUERA, Y. SINCHIGUANO, M. (2023) Aplicación wb con machine learning para la gestión de ventas en el negocio popular Tecno Maedera del Cantón Santo Domingo. [Tesis de grado] Pontificia Universidad Católica del Ecuador, Sede Santo Domingo. Disponible en : <https://repositorio.puce.edu.ec/handle/123456789/42244>
- LANTZ, Brett. Machine learning with R : learn techniques for building and improving machine learning models, from data preparation to model tuning, evaluation, and working with big data.[en línea]. Inglaterra: Packt Publishing, 2023 [fecha de consulta: 05 de marzo de 2024] Capítulo 1. Introducing Machine Learning. Disponible en : <https://app.knovel.com/kn/resources/kpMLRL0001/toc?cid=to> ISBN: 978-1-80107-605-0
- LANZ Lukas, BRIKER Roman, GERPOTT Fabiola. Employees Adhere More to Unethical Instructions from Human Than AI Supervisors: Complementing Experimental Evidence with Machine Learning, 2024. Pag 3, Disponible en: <https://doi.org/10.1007/s10551-023-05393-1>
- LI, Ping et. al. Construction and optimization of vending machine decision support system based on improved C4.5 decision tree. Heliyon [en línea]. Febrero-2024, Volume 10, Issue 3 [Fecha de consulta: 10/03/2024]. Disponible en <https://doi.org/10.1016/j.heliyon.2024.e25024>ISSN: 2405-8440

- LUZURIAGA, Jaramillo., et. al. Histogram and normal distribution: Shapiro-Wilk and Kolmogorov Smirnov applied in SPSS [en línea]. Octubre-2023, Volume 4, LATAM VII [Fecha de consulta: 20/05/2024]. Disponible en <https://doi.org/10.56712/latam.v4i4.1242>ISSN: 2405-8440
- MEDINA, Miguel., et. al. Método mixto de investigación: Cuantitativo y cualitativo. [en línea]. Julio - 2023. Instituto Universitario de Innovación Ciencia y Tecnología Inudi Perú. [Fecha de consulta: 12/03/2024]. Disponible en: <https://doi.org/10.35622/inudi.b.105>ISBN: 978-612-5069-95-5
- MENDOZA, Santiago. Pasos para elaborar proyectos de investigación científica: cuantitativa, cualitativa y mixta [en línea] Lima: Editorial San Marcos. 2013 [Fecha de consulta: 9 de octubre de 2023] Capítulo 2: El método Experimental. Disponible en: <https://books.google.com.pe/books?id=LC4MxQEACAAJ>ISBN: 9786123028787
- O'SHAUGHNESSY, Jhon. Marketing Competitivo [en línea] Madrid: Diaz de Santos, S. A. 1991 [Fecha de consulta: 30 de noviembre de 2023] Capítulo 12: Gestión de ventas. Disponible en: https://www.google.com.pe/books/edition/Marketing_competitivo/Zua9_PZcmGcC?hl=es-419&gbpv=0ISBN: 9788487189777
- OLIVEIRA, Rodrigo, et al. Prediction of Road Transport of Wood in Uruguay: Approach with Machine Learning. Revista brasileña de Agricultura [en línea]. Septiembre-octubre 2022 [Fecha de consulta: 15 de noviembre de 2023] Disponible en: <https://doi.org/10.3390/f13101737>EISSN: 1999-4907
- QINHAN, Jiang. Métodos de Clasificación con Python: Aplicaciones Empresariales, Tesis (Master en Dirección de Empresas). Valencia: Universitat Politècnica de València, 2023. Disponible en: <http://hdl.handle.net/10251/195236>
- RODRIGUEZ, Cristina. La productividad en el servicio [en línea]. Bogotá - Colombia 2022. Disponible en: https://www.google.com.pe/books/edition/La_productividad_en_el_servicio/7XSXEAAQBAJ?hl=es-419&gbpv=0ISBN: 9789586606790
- RUI MA, Di Mao et. al. From vineyard to table: Uncovering wine quality for sales management through machine learning. Journal of Business Research [en línea]. Abril-2024, Volume 176 [Fecha de consulta: 02/04/2024]. Disponible en <https://doi.org/10.1016/j.jbusres.2024.114576>ISSN: 0148-2963

- SALTOS, Douglas y VILLACIS, Oscar. Implementación de Machine Learning en el área de ventas de la empresa Zapec S.A. Tesis (Licenciado en Negocios Internacionales). Guayaquil: Universidad Católica de Santiago de Guayaquil, 2022. Disponible en: <http://repositorio.ucsg.edu.ec/handle/3317/18337>
- Saura, J. R., Reyes-Menendez, A., & Martinez-Navarro, J. (2023). Artificial intelligence and machine learning in business: A review and research agenda. *Technological Forecasting and Social Change*, 188, 122235. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S0040162522007855>
- SERGIO, Jimenez, Martinez. Investigación y recogida de información de mercados.[en línea]. Antequera - Málaga 2022. Pag 28, Disponible en: https://www.google.com.pe/books/edition/Investigaci%C3%B3n_y_recogida_de_informaci%C3%B3n/-7WpEAAAQBAJ?hl=es-419&gbpv=0ISBN: 9788411031912
- SERRANO, A. A., SANZ, L. G., RODRIGO, I. L., GORDO, E. G., ÁLVARO, B. G., & BREA, L. R. (2020). MÉTODOS DE INVESTIGACIÓN DE ENFOQUE EXPERIMENTAL.
- Shipunov Alexey. (2019, enero 5). Book: Visual Statistics Use R! (Shipunov). Statistics LibreTexts. Disponible en: <https://www.academia.edu/download/55568285/Experimental.pdf>
- VARGAS, Habana. Sistema web basado en técnicas de machine learning para predecir el desempeño de los docentes en el nivel de educación superior universitarias, Tesis (Licenciado en Ingeniería de sistemas y computación). Chiclayo: Universidad Católica Santo Toribio De Mogrovejo, 2023. Disponible en: <http://hdl.handle.net/20.500.12423/6128>
- Wedel, M., & Kannan, P. K. (2023). *Marketing Analytics: Data-Driven Techniques with Microsoft Excel*. Wiley. Disponible en: <https://www.wiley.com/en-be/Marketing+Analytics%3A+Data-Driven+Techniques+with+Microsoft+Excel-p-9781118373439>ISBN: 978-1119886529
- WEST, Alan. Gestión de la Distribución Comercial [en línea]. Madrid: Diaz de Santos, S. A. 1991 [Fecha de consulta: 19 de noviembre de 2023]. Capítulo 14. Gestión de ventas: Su papel en la distribución. Disponible en: <https://books.google.com.pe/books?id=10pWjstZnWwC>ISBN: 9788487189814

YARANGA, Italo. Machine Learning para predecir el riesgo crediticio de un cliente en la empresa FUTECH PERU S.A.C. Tesis (Licenciado en Ingeniería de Sistemas). Lima: Universidad César Vallejo, 2022. Disponible en: <https://hdl.handle.net/20.500.12692/104048>

ANEXOS

APLICACIÓN DE CRISP-DM A LA PROBLEMÁTICA

Machine Learning en la Gestión Fases

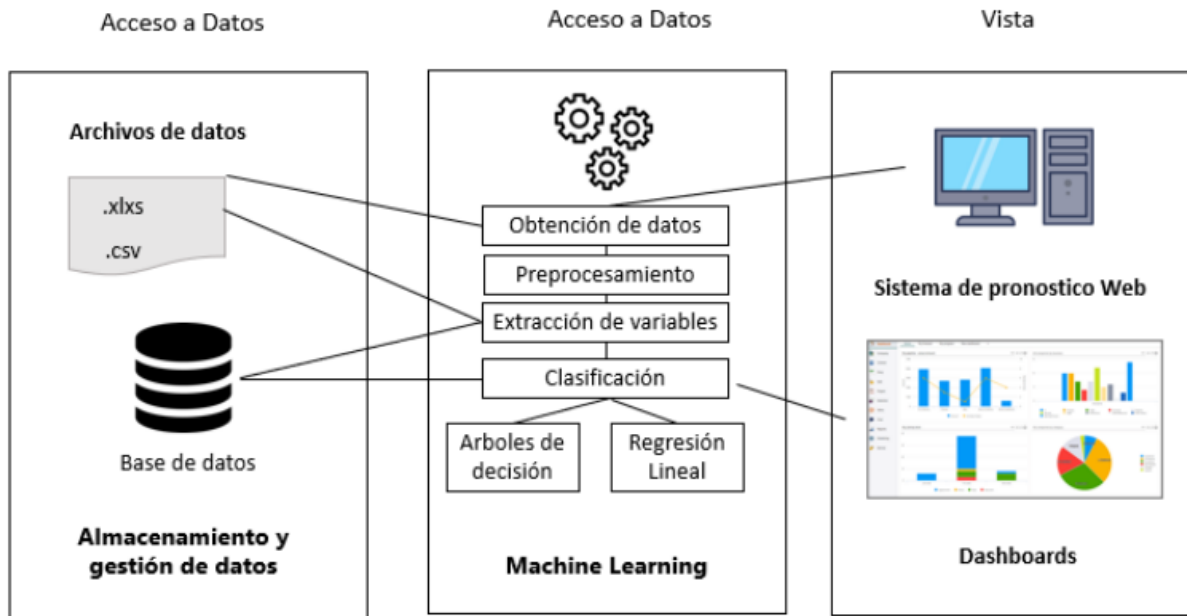


Figura 20

Gráfico de las fases de implementación el machine learning elaboración propia

FASES DE IMPLEMENTACIÓN

1. Comprensión del Negocio

En este sentido, al detallar cada tarea que sustenta esta fase inicial de comprensión del negocio, se pretende definir los requisitos y metas de este proyecto actual. Este enfoque permitirá transformarlos en objetivos técnicos y organizarlos en un plan concreto para el proyecto.

1.1. Determinar los Objetivos del Negocio

Los objetivos que se aplicarán en este proyecto de machine learning son determinar cómo el uso del Sistema de Pronóstico mediante machine learning impacta la gestión

de ventas. Además, se analizará cómo el Machine Learning influye en la eficiencia y productividad de la gestión de ventas en esta empresa.

Contexto

Las empresas se encuentran en una posición donde dispone de una base de datos actualizada que registra las ventas realizadas. No obstante, carece de procesos sistematizados en la gestión de ventas que permitan una planificación adecuada y un pronóstico preciso para una toma de decisiones más eficiente y ágil.

Objetivos del Negocio

El objetivo general planteado para el negocio es ver de qué manera influye el uso de las técnicas del aprendizaje automático para la toma de decisiones en la empresa, de tal qué manera, que gracias a las predicciones del programa puedan tener un mejor manejo de la productividad y eficiencia de sus procesos. Así mismo se definieron los siguientes objetivos:

Establecer de qué forma influyó el Machine Learning en la eficiencia para la toma de decisiones de la gestión de ventas en una empresa manufactura en el Perú

Establecer de qué forma influyó el Machine Learning en la productividad para la gestión de ventas en una empresa manufactura en el Perú

Los informes pueden ser muy útiles para las coordinaciones al momento de validaciones de producción y eficiencia. Esto permitirá a la empresa lograr mejorar la calidad de sus procesos aplicados en su producción y eficiencia de ventas por canal.

Criterios de éxito del negocio

En pos de fortalecer la gestión de ventas, se propone implementar un proyecto de machine learning con el objetivo de mejorar la precisión de los pronósticos de ventas y optimizar la eficiencia operativa. El proyecto busca integrar de manera fluida el Sistema de Pronóstico con la base de datos de ventas existente. Se aspira a obtener una aceptación positiva por parte del equipo de ventas, respaldada por una mejora tangible en la productividad y la capacidad del sistema para facilitar decisiones estratégicas basadas en análisis profundos. La adaptabilidad a cambios del mercado y el cumplimiento normativo son también aspectos clave para garantizar el éxito a largo plazo de esta iniciativa.

1.2. Valuación de la Situación

En este punto se valida que contamos con información de gestiones de ventas, la cuales fueron brindadas en excel con datos de ventas mensuales y anuales. La cual es realmente útil para realizar el sistema de pronóstico utilizando machine learning.

Inventario de recursos

Considerando los recursos de software disponibles, contamos con Sql Server, el cual proporciona herramientas para gestionar la información de manera más efectiva. Por lo tanto, se establecerá una base de datos para almacenar los datos y llevar a cabo un análisis más exhaustivo.

Enfocándonos en los recursos del Hardware con lo que trabajaremos detallamos a continuación:

- Modelo: PC Escritorio
- Procesador: Intel © Core i5 9na Gn 2,90 GHz
- Memoria RAM: 16,00 GB
- Almacenamiento: SSD 2TB NVme - DISCO DURO 2TB 7200 RPM
- Tarjeta gráfica: Geforce RTX 3060
- Monitor TFT: MSI 27 pulgadas 144hz

1.3. Determinar los Objetivos del proyecto

Para determinar los objetivos del proyecto se tuvieron en cuenta los siguientes puntos:

- Predecir la eficiencia de las ventas en base a una cantidad de meses establecida
- Verificar la productividad de las ventas en base a la producción con las ventas realizadas mensuales.
- Automatización de los reportes en dashboard para una mejor lectura y toma de decisiones.

1.4. Realizar el Plan del Proyecto

El presente proyecto se realiza en fases para facilitar el tiempo de entrega:

- Fase 1: Análisis de los datos del negocio (Excel)

- Fase 2: Preparación de los datos (selección, limpieza y conversión)
- Fase 3: Elección del algoritmo de machine learning a utilizar y ejecución de las muestras sig.
- Fase 4: Análisis de los resultados obtenidos, de requerirse repetir fase 3
- Fase 5: Generación de gráficos de los datos del modelo predictivo, en base a los objetivos del negocio con los resultados obtenidos.
- Fase 6: Muestra de resultado final.

2. Comprensión de los Datos

Se llevó a cabo la recolección de datos con el objetivo de iniciar la exploración del problema, comprender la naturaleza de los datos y evaluar su calidad inicial.

2.1 Recolección de los Datos

Procedemos a listar los datos que usaremos a lo largo del desarrollo del proyecto.

- **Año:**

2019

2020

2021

2022

2023

- **Meses**

Enero

Febrero

Marzo

Abril

Mayo

Junio

Julio

Agosto

Septiembre

Octubre

Noviembre

Diciembre

- **Tipo de avena**

Avena Granel 10k

Avena Envasada 2.4k

Avena valor Agregado 1.8k

- **Canal de Venta:**

Distribuidor

Brooker

- **Ventas Distribuidor**

- **Ventas Brooker**

- **Total cantidad de avena**

- **Total de venta en soles**

2.2 Descripción de los datos

TABLA_VENTAS_PRODUCTO: En esta tabla se realizó el registro y almacenado de datos, partiendo desde el 2019 hasta el 2023 con la información brindada por el área responsable. La cual está dividida por los 5 años, 12 meses que se evalúan, el tipo de avena según corresponda al análisis del objetivo de la presente empresa y por último el total de avena vendida en cantidad de productos.

- **AÑO:** 5 AÑOS
- **MESES:** 12 meses
- **TIPO DE AVENA**

AVENA_GRANEL_10K_CANT

AVENA_ENVASADA2_4_CANT

AVENA_VALOR_AGREGADO1_8_CANT

- **TOTAL_VENTAS_CANT**

TABLA_VENTAS_CANAL_VENTA: En esta tabla se realizó el registro y almacenado de datos, partiendo desde el 2019 hasta el 2023 con la información brindada por el área responsable. La cual está dividida por los 5 años, 12 meses que se evalúan, la cantidad de soles vendidos por cada uno de los meses según el canal de venta que corresponda y por último los totales de ventas por cada mes.

AÑO: 5 AÑOS

- **MESES:** 12 meses
- **CANAL DE VENTA**

VENTAS_DISTRIBUIDOR

VENTAS_BROOKER

- **TOTAL_VENTAS_EN_SOLES**

TABLA_VENTAS_PRODUCTO

ANIO	MESES	AV_GRANEL_10K_CANT	AV_ENVASADA2_4_CANT	AV_VALOR_AGRDO1_8_CANT	TOTAL_AVENA	TOTAL_VENTAS_EN_SOLES	
5	2019	Junio	2753	3473	2601	8827	44915.71
6	2019	Julio	2365	2143	825	5333	24681.71
7	2019	Agosto	6420	3415	2705	12540	58943.05
8	2019	Setiembre	5986	3196	794	9976	42831.02
9	2019	Octubre	7246	3211	792	11249	47549.72
10	2019	Noviembre	2882	1492	805	5179	23380.89
11	2019	Diciembre	4572	1280	294	6146	25030.70
12	2020	Enero	3550	3595	886	8031	36349.70
13	2020	Febrero	2915	2029	1604	6548	31670.58
14	2020	Marzo	5883	3230	1080	10193	44626.70
15	2020	Abril	2764	6407	1810	10981	53228.34
16	2020	Mayo	4115	12976	2034	19125	90811.92
17	2020	Junio	3826	22464	4859	31149	154442.23
18	2020	Julio	3277	13742	2758	19777	96431.04
19	2020	Agosto	6441	15263	3200	24904	118433.06
20	2020	Setiembre	7155	11241	1774	20170	92037.72
21	2020	Octubre	9150	12023	1198	22371	99049.46

TABLA_VENTAS_CANAL_VENTA

ANIO	MESES	VENTAS_DISTRIBUIDOR	VENTAS_BROOKER	TOTAL_AVENA	TOTAL_VENTAS_EN_SOLES
2019	Junio	8826	7021	15847	181765.09
2019	Julio	5332	12063	17395	199520.65
2019	Agosto	12540	8191	20731	237784.57
2019	Setiembre	9977	10139	20116	230730.52
2019	Octubre	11250	11522	22772	261194.84
2019	Noviembre	5179	9473	14652	168058.44
2019	Diciembre	6146	6850	12996	149064.12
2020	Enero	4561	7544	12105	138844.35
2020	Febrero	3363	8062	11425	131044.75
2020	Marzo	7653	10855	18508	212286.76
2020	Abril	4628	6748	11376	130482.72
2020	Mayo	9034	5356	14390	165053.30
2020	Junio	13593	8057	21650	248325.50
2020	Julio	8919	6883	15802	181248.94
2020	Agosto	10233	7885	18118	207813.46
2020	Setiembre	10130	9136	19266	220981.02
2020	Octubre	12259	9724	21983	252145.01

3. Preparación de los Datos

Durante esta etapa, se enfoca en la preparación de los datos para adaptarlos a las técnicas de machine learning que se emplearán. Este proceso implica seleccionar el subconjunto de datos a utilizar, depurarlos para mejorar su calidad, incorporar nueva información a partir de los datos existentes y formatearlos de acuerdo con los requisitos de la herramienta de modelado que se utilizará.

4. Modelado

En la etapa siguiente, se seleccionará la técnica más idónea para los objetivos establecidos. Una vez que se haya elaborado el plan de prueba a los algoritmos elegidos, se aplicarán estas técnicas a los datos con el fin de crear el modelo correspondiente. Seguido de ello, se procederá a evaluar el modelo para lograr identificar si cumple con los criterios preestablecidos. Las técnicas seleccionadas para este propósito son:

- Árboles de decisión (Modelo 1)
- Regresión Lineal Simple (Modelo 2)

4.1. Árbol de Decisión

```
1 # Importar librerías
2 import seaborn as sns
3 import pandas as pd
4 from sklearn.tree import DecisionTreeRegressor
5 from sklearn.model_selection import train_test_split
6 from sklearn.metrics import mean_squared_error
7 from sklearn.preprocessing import MinMaxScaler
8 import numpy as np
9 from sklearn.metrics import r2_score
10 import matplotlib.pyplot as plt
11 import pickle
12
13 # Cargar datos
14 df = pd.read_csv("/workspaces/codespaces-blank/TESIS_ML_DJANGO/DATA_IUP_2025_arbol_decision.csv")
15 #df.head(12)
16 df[5:]
17
18 # Diccionario para mapear los meses a valores numéricos
19 meses_a_numeros = {
20     'Enero': 1, 'Febrero': 2, 'Marzo': 3, 'Abril': 4, 'Mayo': 5, 'Junio': 6,
21     'Julio': 7, 'Agosto': 8, 'Setiembre': 9, 'Octubre': 10, 'Noviembre': 11, 'Diciembre': 12
22 }
23
24 # Crear una nueva columna con los valores numéricos de los meses
25
26 df['MESES_NUM'] = df['MESES'].map(meses_a_numeros)
27
28 X = df[['ANIO', 'MESES_NUM', 'TOTAL_AVENA']]
29 y = df['TOTAL_VENTAS_EN_SOLES']
30
31 # Dividir los datos en conjunto de entrenamiento y prueba
32 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)
```

```

33
34 # Escalar las características
35 scaler = MinMaxScaler(feature_range=(0,1))
36 X_train_scaled = scaler.fit_transform(X_train)
37 X_test_scaled = scaler.transform(X_test)
38
39 # Crear el modelo de árbol de regresión
40 regressor = DecisionTreeRegressor(max_depth=6)
41
42 # Entrenar el modelo
43 regressor.fit(X_train_scaled, y_train)
44
45 # Hacer predicciones en el conjunto de prueba
46 y_pred = regressor.predict(X_test_scaled)
47
48 # Evaluar el modelo
49 r2 = r2_score(y_test, y_pred)
50 print(f"R-squared: {r2}")
51
52 # Guardar el modelo y el escalador como archivos .sav
53 pickle.dump(regressor, open("avena_sales_prediction_model.sav", "wb"))
54 pickle.dump(scaler, open("avena_sales_scaler.sav", "wb"))
55
56 # Función para hacer predicciones
57 def predict_total_ventas(new_data, model, scaler):
58     new_data_scaled = scaler.transform(new_data)
59     return model.predict(new_data_scaled)
60
61 # Ejemplo de cómo usar el modelo guardado
62 # Cargar el modelo y el escalador
63 loaded_model = pickle.load(open("avena_sales_prediction_model.sav", "rb"))
64 loaded_scaler = pickle.load(open("avena_sales_scaler.sav", "rb"))
65

```

```

66 def predict_total_avena(new_data, model):
67     # Convertir los nuevos datos en un DataFrame
68     new_df = pd.DataFrame(new_data)
69     # Hacer la predicción
70     prediction = model.predict(new_df)
71     return prediction[0]
72
73 # Nuevos datos de ejemplo para hacer predicciones
74 new_data_1 = {'ANIO': [2024], 'MESES_NUM': [1], 'TOTAL_AVENA': [20284]}
75 new_data_2 = {'ANIO': [2024], 'MESES_NUM': [2], 'TOTAL_AVENA': [16378]}
76 new_data_3 = {'ANIO': [2024], 'MESES_NUM': [3], 'TOTAL_AVENA': [20069]}
77 new_data_4 = {'ANIO': [2024], 'MESES_NUM': [4], 'TOTAL_AVENA': [20744]}
78 new_data_5 = {'ANIO': [2024], 'MESES_NUM': [5], 'TOTAL_AVENA': [21515]}
79 new_data_6 = {'ANIO': [2024], 'MESES_NUM': [6], 'TOTAL_AVENA': [20540]}
80 new_data_7 = {'ANIO': [2024], 'MESES_NUM': [7], 'TOTAL_AVENA': [18906]}
81 new_data_8 = {'ANIO': [2024], 'MESES_NUM': [8], 'TOTAL_AVENA': [21231]}
82 new_data_9 = {'ANIO': [2024], 'MESES_NUM': [9], 'TOTAL_AVENA': [20263]}
83 new_data_10 = {'ANIO': [2024], 'MESES_NUM': [10], 'TOTAL_AVENA': [16968]}
84 new_data_11 = {'ANIO': [2024], 'MESES_NUM': [11], 'TOTAL_AVENA': [17731]}
85 new_data_12 = {'ANIO': [2024], 'MESES_NUM': [12], 'TOTAL_AVENA': [23560]}
86
87 # Lista para almacenar los nuevos datos
88 new_data_list = [new_data_1, new_data_2, new_data_3, new_data_4, new_data_5, new_data_6,
89                 new_data_7, new_data_8, new_data_9, new_data_10, new_data_11, new_data_12]
90 # Lista para almacenar las predicciones
91 predictions = []
92
93 # Hacer predicciones con los nuevos datos y almacenarlas en la lista
94 for data in new_data_list:
95     prediction = predict_total_avena(data, regressor)
96     predictions.append(prediction)
97 # Crear un DataFrame con las predicciones
98 predictions_df = pd.DataFrame(predictions, columns=['Predicción VENTA_TOTAL_EN_SOLES_X_CANAL'])
99 # Mostrar el DataFrame
100 print(predictions_df)

```

4.2. Regresión Lineal

```
1 # Importar librerías
2 import seaborn as sns
3 import pandas as pd
4 from sklearn.linear_model import LinearRegression
5 from sklearn.model_selection import train_test_split
6 from sklearn.metrics import mean_squared_error
7 from sklearn.preprocessing import MinMaxScaler
8 import numpy as np
9 from sklearn.metrics import r2_score
10 import matplotlib.pyplot as plt
11 import pickle
12
13 # Cargar datos
14 df = pd.read_csv("/workspaces/codespaces-blank/TESES_ML_DJANGO/DATA_IUP_2025_regresion_lineal.csv")
15
16 # Diccionario para mapear los meses a valores numéricos
17 meses_a_numeros = {
18     'Enero': 1, 'Febrero': 2, 'Marzo': 3, 'Abril': 4, 'Mayo': 5, 'Junio': 6,
19     'Julio': 7, 'Agosto': 8, 'Setiembre': 9, 'Octubre': 10, 'Noviembre': 11, 'Diciembre': 12
20 }
21
22 # Crear una nueva columna con los valores numéricos de los meses
23 df['MESES_NUM'] = df['MESES'].map(meses_a_numeros)
24
25 X = df[['AÑO', 'MESES_NUM', 'TOTAL_AVENA']]
26 y = df['TOTAL_VENTAS_EN_SOLES']
27
28 # Dividir los datos en conjunto de entrenamiento y prueba
29 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)
30
31 # Escalar las características
32 scaler = MinMaxScaler(feature_range=(0, 1))
33 X_train_scaled = scaler.fit_transform(X_train)
34 X_test_scaled = scaler.transform(X_test)
35
36 # Crear el modelo de regresión lineal
37 regressor = LinearRegression()
```

```
38 # Entrenar el modelo
39 regressor.fit(X_train_scaled, y_train)
40
41 # Hacer predicciones en el conjunto de prueba
42 y_pred = regressor.predict(X_test_scaled)
43
44 # Evaluar el modelo
45 r2 = r2_score(y_test, y_pred)
46 print(f"R-squared: {r2}")
47
48 # Guardar el modelo y el escalador como archivos .sav
49 pickle.dump(regressor, open("modelo_regresion_lineal.sav", "wb"))
50 pickle.dump(scaler, open("modelo_regresion_scaler.sav", "wb"))
51
52 # Función para hacer predicciones
53 def predict_total_ventas(new_data, model, scaler):
54     new_data_scaled = scaler.transform(new_data)
55     return model.predict(new_data_scaled)
56
57 # Cargar el modelo y el escalador
58 loaded_model = pickle.load(open("modelo_regresion_lineal.sav", "rb"))
59 loaded_scaler = pickle.load(open("modelo_regresion_scaler.sav", "rb"))
60
61 # Función para hacer predicciones con nuevos datos
62 def predict_total_avena(new_data, model, scaler):
63     # Convertir los nuevos datos en un DataFrame
64     new_df = pd.DataFrame(new_data)
65     # Escalar los datos nuevos
66     new_df_scaled = scaler.transform(new_df)
67     # Hacer la predicción
68     prediction = model.predict(new_df_scaled)
69     return prediction[0]
70
```

```

71 # Nuevos datos de ejemplo para hacer predicciones
72 new_data_1 = {'ANIO': [2024], 'MESES_NUM': [1], 'TOTAL_AVENA': [20284]}
73 new_data_2 = {'ANIO': [2024], 'MESES_NUM': [2], 'TOTAL_AVENA': [16378]}
74 new_data_3 = {'ANIO': [2024], 'MESES_NUM': [3], 'TOTAL_AVENA': [20069]}
75 new_data_4 = {'ANIO': [2024], 'MESES_NUM': [4], 'TOTAL_AVENA': [20744]}
76 new_data_5 = {'ANIO': [2024], 'MESES_NUM': [5], 'TOTAL_AVENA': [21515]}
77 new_data_6 = {'ANIO': [2024], 'MESES_NUM': [6], 'TOTAL_AVENA': [20540]}
78 new_data_7 = {'ANIO': [2024], 'MESES_NUM': [7], 'TOTAL_AVENA': [18906]}
79 new_data_8 = {'ANIO': [2024], 'MESES_NUM': [8], 'TOTAL_AVENA': [21231]}
80 new_data_9 = {'ANIO': [2024], 'MESES_NUM': [9], 'TOTAL_AVENA': [20263]}
81 new_data_10 = {'ANIO': [2024], 'MESES_NUM': [10], 'TOTAL_AVENA': [16968]}
82 new_data_11 = {'ANIO': [2024], 'MESES_NUM': [11], 'TOTAL_AVENA': [17731]}
83 new_data_12 = {'ANIO': [2024], 'MESES_NUM': [12], 'TOTAL_AVENA': [23560]}
84
85 # Lista para almacenar los nuevos datos
86 new_data_list = [new_data_1, new_data_2, new_data_3, new_data_4, new_data_5, new_data_6,
87                 new_data_7, new_data_8, new_data_9, new_data_10, new_data_11, new_data_12]
88 # Lista para almacenar las predicciones
89 predictions = []
90
91 # Hacer predicciones con los nuevos datos y almacenarlas en la lista
92 for data in new_data_list:
93     prediction = predict_total_avena(data, loaded_model, loaded_scaler)
94     predictions.append(prediction)
95 # Crear un DataFrame con las predicciones
96 predictions_df = pd.DataFrame(predictions, columns=['Predicción VENTA_TOTAL_EN_SOLES_X_CANAL'])
97 # Mostrar el DataFrame
98 print(predictions_df)
99

```

5. Evaluación

En la siguiente etapa metodológica, se enfoca en evaluar o valorar los modelos, poniendo énfasis en los objetivos propuestos. Una vez concluida la presente evaluación, es crucial determinar si se han alcanzado los objetivos planteados. Si se confirma su logro, se procede con la implementación, se hace imperativo identificar cualquier aspecto omitido y llevar a cabo una revisión exhaustiva de todo el proceso.

5.1 Evaluar los Resultados

En el ámbito de la empresa, se había definido como principal indicador de logro la capacidad de generar pronósticos con un nivel de confiabilidad satisfactorio. Sin embargo, sin embargo, este criterio puede resultar ambiguo, lo que hace imprescindible apoyarse en parámetros de éxito más concretos y cuantificables desde una perspectiva técnica. Para evaluar objetivamente la calidad de las predicciones, es fundamental contar con una base sólida, como los indicadores estadísticos derivados de la ejecución de los modelos. En cualquier caso, fundamentándose en los datos obtenidos mediante el análisis de datos, podemos proceder a examinar cada modelo para así eliminar aquel que no alcance un umbral mínimo de rendimiento.

- **Modelo 1 para objetivo**

Este modelo resultó más factible, dado que permite realizar una predicción más confiable en base a la gestión de ventas, con un 88% de fiabilidad, lo cual consideramos aceptable desde el punto de vista de la empresa.

- **Modelo 2 para objetivo**

Este modelo no resultó viable, ya que no proporcionó garantías suficientes, con tan solo un 49% de confianza al momento de la predicción, ya que se busca lograr un modelo predictivo más confiable sobre la gestión de ventas que se llevarán a cabo.

Modelo Aprobado

En base a los resultados en esta sección, el modelo aprobado es el 1, ya que cumple con los criterios de éxito propuestos por la empresa, mientras el modelo 2 es descartado por no cumplir con los requisitos de la empresa ni con los del ML.

6. Distribución

Esta es la etapa final de la metodología Crisp DM y su objetivo es detallar al cliente cómo implementar el proyecto desarrollado en las fases anteriores, además de presentar los resultados obtenidos de manera comprensible para el cliente. Otro propósito de esta fase es establecer una estrategia para el mantenimiento del proyecto que sea escalable, para en un futuro producir informes que incluya posibles propuestas en la toma de decisiones del negocio.

6.1 Planear la Implementación

Para la implantación del presente proyecto de investigación en la empresa es sumamente necesario poder tener acceso real a la base de datos, en la que contiene información completa y relativa de las ventas de productos y el total de ingresos que presenta mensualmente la empresa, a mayor información, permite al modelo predecir mejor. Comenzando desde ahí, los siguientes pasos son los mismos que se propusieron en este documento iniciando desde la comprensión del negocio hasta la implantación. Teniendo en cuenta que habrán etapas, como la preparación de datos y comprensión de la empresa, que al obtener los datos reales de la empresa, podrían

ser más complejas de entender y tomar más tiempo para ejecutarse al momento de la limpieza y ordenamiento de datos.

6.2 Planear Mantenimiento y Monitoreo

El mantenimiento y monitoreo del presente proyecto luego de la implementación es una fase crucial, ya que los datos procesados frecuentemente pueden ser alterados o por los canales de venta. Estos cambios pueden deberse a diversas razones, como una codificación incorrecta o un error en la seguridad. Dado que un gran volumen de datos en constante movimiento, debe realizarse una copia de seguridad de datos recomendablemente cada tres meses, ya que en ese lapso de tiempo se pueden entregar informes al gerente comercial sobre el proceso de sus ventas ya sea en productos o en ingresos mensuales para así pueda apoyarse en la toma de decisiones para un mejor análisis y evaluar a sus canales de ventas. Sin embargo, este plan podría cambiar según lo requerido por la empresa.

Para un buen monitoreo y mantenimiento, se establece los siguientes procesos:

- Copia de seguridad de manera trimestral de la información o datos.
- Los resultados que se obtengan en cada extracción de datos deben transformarse en hojas de cálculo y generar gráficos diversos para facilitar la visualización e interpretación de los resultados en caso se desee evaluar.
- Clasificación de los datos en función de los modelos para volver a ingresarlos de requerirse.

6.3. Producir el Informe Final

En esta parte se presentan los gráficos estadísticos respecto a los puntos clave del proyecto y la experiencia que se obtuvo durante su desarrollo. El público dirigido sería el área de operaciones de tal manera que puedan tener la comprensión de estos gráficos para que realicen sus respectivas comparativas y así tomar medidas correspondientes en base a su toma de decisiones para el área.

La aplicación de la metodología CRISP-DM en el presente proyecto ha permitido

identificar un comportamiento predictivo en la eficiencia de la cantidad de ventas en soles por canal y en la productividad en términos de cantidad de ventas. Se ha desarrollado un plan para la extracción, normalización y codificación de los datos, lo cual ha permitido la realización del proceso de datos anuales utilizando el modelo de árbol de decisión.

6.4 Revisar el Proyecto

La última fase de la metodología de Crisp DM, es sumamente importante evaluar tanto los aciertos como los errores, para así identificar posibles mejoras para mejorar los modelos, la ejecución de la minería de los datos y obtener resultados más precisos.

Primero, como se mencionó anteriormente en otros apartados, la mayor limitación de este proyecto ha sido la falta de una base de datos real sobre la cual trabajar, lo cual ha condicionado en gran medida los resultados obtenidos.

Plataforma web completa con Python y librería Chart.js para visualizar datos en gráficos.

Eficiencia:



Productividad:



Descargar Gráficos:

ADMINISTRACION

GESTIÓN DE VENTAS

EFICIENCIA

VENTAS DE PRODUCTOS

Año

2019

2020

2021

2022

PRODUCTIVIDAD

VENTAS EN SOLES

Año

2019

2020

2021

2022

TRANSACTIONS

[Download Graphics](#)

[Cerrar sesión](#)

Ventas mensuales en Cantidad de Productos 2019

Historial de descargas recientes

- chart_3.png
54.8 KB • Hace 3 minutos
- chart_2.png
32.9 KB • Hace 3 minutos
- chart_1.png
33.8 KB • Hace 32 minutos

Historial de descargas completo

Gráfico de Ventas Anual

Registro de Datos:

Home Redes

ADMINISTRACION

GESTIÓN DE VENTAS

EFICIENCIA

VENTAS DE PRODUCTOS

Año

2019

2020

2021

2022

PRODUCTIVIDAD

VENTAS EN SOLES

Año

2019

2020

2021

2022

TRANSACTIONS

[Download Graphics](#)

[Cerrar sesión](#)

Ventas mensuales en Cantidad de Productos 2020

Ventas 2020

Predicción Ventas

Año:

Total Avena para el mes 1:

Total Avena para el mes 2:

Total Avena para el mes 3:

Total Avena para el mes 4:

Total Avena para el mes 5:

Total Avena para el mes 6:

Total Avena para el mes 7:

Total Avena para el mes 8:

Total Avena para el mes 9:

Total Avena para el mes 10:

Total Avena para el mes 11:

Total Avena para el mes 12:

Obtener Predicciones

Anexo 1. Tabla de operacionalización de variables.

Variable Independiente	Definición Conceptual	Definición Operacional	Dimensión	Indicador
Machine Learning	De acuerdo con Chopra y Khurana (2023) menciona que el Machine Learning “Es un campo de la inteligencia computacional que implica el análisis e interpretación de estructuras y patrones presentes en los datos y la máquina es capaz de aprender, tomar decisiones y razonar con estos patrones y poder extraer soluciones inteligentes.”	Según Lantz, Brett (2023) “Debido a que el aprendizaje automático, en esencia, se trata de convertir los datos en acción, las herramientas de explicabilidad ayudan a generar confianza en un modelo, lo que conduce a una mayor adopción y un mayor impacto.”		
Variable Dependiente	Definición Conceptual	Definición Operacional	Dimensión	Indicador
Gestión de Ventas	En el libro de John O'shaughnessy titulado Marketing Competitivo (1991), nos dice que en la gestión de ventas se basa en la estrategia, cuyo objetivo es determinar correctamente las cosas por realizar, determinar las guías por niveles más alto en base a su planificación en la empresa; mientras las decisiones de su eficiencia van relacionadas con aquellas operaciones que día a día son necesarias para los obtener objetivos propuestos en su estrategia.	Castañeda, Miguel (2020), nos habla sobre la gestión de ventas en la cual nos dice que aplicación de machine learning para su investigación dieron resultados positivos para su mejora en los procesos de ventas. En la cual establecieron la eficiencia y productividad como indicadores para su investigación. En la cual su eficiencia aumentó a un 82% y su productividad en un 81%.	Eficiencia	Nivel de Eficiencia
			Productividad	Nivel de Productividad

Anexo 2. Matriz de Consistencia.

Problema	Objetivos	Hipótesis	Operacionalización de Variables				
			Variable	Dimensión	Indicador	Fórmula	Metodología
General	General	General	Independiente				Tipo de Investigación: Aplicada Diseño: Pre - Experimental Muestreo: muestreo no probabilístico Por conveniencia Técnica de Recolección de datos: Ficha de registro
PG: ¿De qué forma influye el Machine learning en la toma de decisiones para la gestión de ventas?	OG. Establecer de qué forma influye el Machine learning en la toma de decisiones para la gestión de ventas.	HG La aplicación de Machine Learning mejora en la toma de decisiones de la gestión de ventas.	Machine Learning				
Específico	Específico	Específico	Dependiente				
PE1 ¿De qué forma influye Machine learning en la eficiencia para la toma de decisiones en la gestión de ventas?	OE1 Establecer de qué forma influye Machine Learning en la eficiencia para la toma de decisiones de la gestión de ventas.	HE1 El uso de machine learning aumenta la eficiencia para la gestión de ventas.	Gestión de Ventas	Eficiencia	Nivel de Eficiencia	$NE = VSCM / MVS$ NE: Nivel de eficiencia VSCM: Venta en Soles por Canal Mensual MVS: Meta de Ventas en Soles	
PE2 ¿De qué manera influye Machine Learning en la productividad para la toma de decisiones en la gestión de ventas?	OE2 Establecer de qué forma influye Machine Learning en la productividad para la gestión toma decisiones de la gestión ventas.	HE2 El uso de machine learning refuerza la productividad para la gestión de ventas.		Productividad	Nivel de Productividad	$NP = VCM / MVC$ NP: Nivel de Productividad VCM: Ventas en Cantidad Mensual MVC: Meta de venta en cantidad	

Anexo 3. Cuestionario



UNIVERSIDAD CÉSAR VALLEJO

Cuestionario

Estimado Jorry Baca Cuba Gerente Comercial de Industrias Unidas del Perú, se agradece su participación de este cuestionario, el cual tiene como objetivo netamente académico. Este cuestionario es anónimo, por favor sírvase a indicar la frecuencia de acción de su organización marcando con una equis "X", considerando la siguiente escala para cada enunciado:

MUY ALTA	ALTA	MODERADO	BAJA	MUY BAJA
5	4	3	2	1

Enunciado	MA	A	M	B	MB
Dimensión 1: Productividad	5	4	3	2	1
1. ¿Cómo evaluaría la capacidad actual de IUP para cumplir con sus metas de ventas establecidas?	X				
2. En su opinión, ¿qué tan efectivos son los métodos actuales de pronóstico de ventas de IUP para maximizar su producción y alcanzar las metas de ventas?		X			
3. ¿Qué nivel de precisión percibe en las proyecciones de ventas actuales de IUP en relación con sus metas de ventas?		X			
4. ¿Cómo considera la coordinación entre los departamentos de IUP para optimizar la producción y alcanzar las metas de ventas?		X			
5. ¿Qué tan ágil percibe la respuesta de IUP ante cambios inesperados en la demanda del mercado para ajustar las estrategias de ventas y cumplir con las metas establecidas?		X			
6. En su opinión, ¿qué tan rápido se adapta IUP a nuevas tecnologías para mejorar su productividad y alcanzar las metas de ventas?			X		
7. ¿Cómo calificaría la eficacia del uso actual de recursos humanos en las actividades de ventas de IUP para lograr las metas establecidas?		X			
8. ¿Qué nivel de productividad percibe en los procesos logísticos de IUP para asegurar una distribución oportuna y contribuir al logro de las metas de ventas?			X		
9. ¿Qué opinión tiene sobre el nivel de automatización en la cadena de ventas de IUP y su contribución al logro de las metas de ventas establecidas?		X			
10. ¿Cómo valora la capacidad de IUP para innovar en sus estrategias de ventas y adaptarse a las tendencias del mercado para lograr las metas establecidas?		X			
Dimensión 2: Eficiencia					
1. ¿Cómo calificaría la eficiencia actual de los diferentes canales de venta de IUP en términos de contribuir al logro de las metas de ventas establecidas?			X		
2. En su opinión, ¿qué tan efectivos son los canales de venta actuales de IUP para generar ingresos y contribuir al logro de las metas de ventas establecidas?			X		
3. ¿Qué nivel de satisfacción tiene con la rentabilidad de los diferentes canales de venta utilizados por IUP en relación con las metas de ventas establecidas?		X			
4. ¿Cómo considera la eficacia de la estrategia de precios de IUP en relación con la competencia en el mercado y su impacto en el logro de las metas de ventas?		X			
5. ¿Qué tan ágil percibe que es la respuesta de IUP ante cambios en las tendencias de consumo en términos de ajustar los canales de venta para alcanzar las metas establecidas de ventas?			X		
6. En su opinión, ¿qué tan efectivo es el aprovechamiento de oportunidades de nuevos mercados mediante los canales de venta de IUP para lograr las metas de ventas establecidas?		X			
7. ¿Cómo evaluaría la efectividad del servicio al cliente ofrecido a través de los diferentes canales de venta de IUP en términos de contribuir al logro de las metas de ventas establecidas?			X		
8. ¿Qué nivel de efectividad percibe en las estrategias de marketing utilizadas por IUP para promover sus productos a través de los diferentes canales de venta y contribuir al logro de las metas de ventas establecidas?			X		
9. ¿Cómo considera el nivel de inversión en tecnología de información para optimizar los canales de venta de IUP y contribuir al logro de las metas de ventas establecidas?			X		
10. ¿Cómo percibe la capacidad de IUP para adaptarse a las preferencias cambiantes de los clientes a través de sus canales de venta y contribuir al logro de las metas de ventas establecidas?			X		

Jorry Baca Cuba
DNI: 09542221
Gerente Comercial de Industrias
Unidas del Perú

Anexo 4. Anexo Entrevista



Entrevista al Gerente de Marketing de Industrias Unidas del Perú

Entrevista

La presente entrevista muestra las preguntas que se realizaron al gerente de Marketing de Industrias Unidas del Perú, para así poder recolectar información y nos pueda brindar un mejor panorama del estado de la empresa.

¿Cuál es el negocio de la empresa?

Industrias Unidas del Perú es una empresa que se encarga de la producción y venta de productos hechos a base de maca, quinua, kiwicha y soya.

¿El área de sistemas esta cumpliendo con los objetivos planteado? ¿Por qué?

Actualmente el área de sistema no está cumpliendo con los objetivos planteados, ya que deberían enviar indicadores diarios, semanal y mensualmente, los cuales son de veras importancia para la toma de decisiones para el área de operaciones.

¿Cree usted que tiene los recursos necesarios para tener un buen control de la empresa?

Actualmente llevamos un control y manejo de datos manual para la toma de decisiones, ya que solo se usa herramientas como Excel para comparar los datos.

¿Cree usted que el personal de sistemas es el adecuado? ¿Por qué?

Si, ya que cumplen con las funciones designadas, así mismo llevan capacitaciones constantes para lograr los objetivos planteados en base a seguridad y funcionalidad en la empresa.

¿Se estan llegando a los objetivos propuestos con el uso de indicadores? ¿Por qué?

En base a los indicadores no se esta llegando a los objetivos propuestos para lograr tener una mejor toma de decisiones en la productividad, ya que hay una demora en el proceso de datos. Esto hace que haya un margen de error.

¿Cómo mide la productividad de sus ventas?

La productividad la medimos mediante las ventas realizadas de carácter mensual, se evalúa el número de productos vendidos, la cantidad de kilos vendidos de avena, comparándolo con las ventas anuales y la productividad de la fábrica.

¿Cómo realiza ese análisis para evaluar la productividad?

Todos los datos son planificados y evaluados del Excel de forma manual con fórmulas en la hoja de cálculo, en base a las cantidades vendidas. Al finalizar el mes se realiza un reporte y vemos si la producción fue optima.

¿Usted cree que se cumplen las funciones correctamente?

En base a la productividad sentimos que en ocasiones existe una mala toma de decisiones ya que es muy incierto la cantidad de ventas que se pueden realizar a futuro, por ende, se suele realizar a veces una sobreproducción.





¿Qué herramientas tecnológías utiliza para elaborar dichos informes?

El área de sistemas utiliza herramientas de Excel y PowerBi para una mejor visualización de los datos presentado en dashboard.

¿Tiene limitaciones o dificultades para la generación de dichos informes?

Si, debido a que se necesita una correcta limpieza de los datos y manipulación estos para un correcto informe y este mismo se presente correctamente en los dashboard sobre las ventas al fin de cada semana, quincena y mes para el área de operaciones.

Jorry Baca Cuba
DNI: 09542221
Gerente Comercial de Industrias
Unidas del Perú

Percy Anchante Aybar
DNI: 72469020
Investigador
Universidad Cesar Vallejo

Davio Fernandez Chavez
DNI: 73663227
Investigador
Universidad Cesar Vallejo



Anexo 5. Anexo Ficha de Registro Eficiencia Pretest

PRE - Ficha de Registro				
Investigadores	Anchante Aybar, Percy Andersson			
	Fernandez Chavez, Davio			
Empresa	Industrias Unidas del Perú			
Investigación	Machine Learning para la toma de decisiones en la gestión de ventas, Lima 2024			
Variables	Gestión de Ventas			
Dimensión	Eficiencia			
Indicador	Nivel de eficiencia			
Formula	Ventas en soles por canal mensual / Meta de ventas en soles			
N°	Fecha	Ventas en soles por canal mensual	Meta de ventas en soles	Eficiencia
1	ene-23	232657	400000	0.58
2	feb-23	187856	300000	0.63
3	mar-23	230191	350000	0.66
4	abr-23	237934	300000	0.79
5	may-23	246777	350000	0.71
6	jun-23	235594	350000	0.67
7	jul-23	216852	300000	0.72
8	ago-23	243520	300000	0.81
9	Set-23	232417	300000	0.77
10	oct-23	194623	250000	0.78
11	nov-23	203375	300000	0.68
12	dic-23	270233	400000	0.68



Jorry Baca Cuba
 DNI: 09542221
 Gerente Comercial de Industrias
 Unidas del Perú

Anexo Ficha de Registro Eficiencia Post Test

POST - Ficha de Registro				
Investigadores		Anchante Aybar, Percy Andersson		
		Fernandez Chavez, Davio		
Empresa		Industrias Unidas del Perú		
Investigación		Machine Learning para la toma de decisiones en la gestión de ventas, Lima 2024		
Variables		Gestión de Ventas		
Dimensión		Eficiencia		
Indicador		Nivel de eficiencia		
Formula		Ventas en soles por canal mensual / Meta de ventas en soles		
N°	Fecha	Ventas en soles por canal mensual	Meta de ventas en soles	Eficiencia
1	ene-24	224319	300000	0.75
2	feb-24	187856	200000	0.94
3	mar-23	224319	250000	0.90
4	abr-23	237934	250000	0.95
5	may-23	244976	300000	0.82
6	jun-23	235594	220000	1.07
7	jul-23	220981	250000	0.88
8	ago-23	243520	250000	0.97
9	Set-23	232417	250000	0.93
10	oct-23	195483	200000	0.98
11	nov-23	203375	200000	1.02
12	dic-23	270233	280000	0.97



Jorry Baca Cuba
DNI: 09542221
Gerente Comercial de Industrias
Unidas del Perú

Anexo Ficha de Registro Productividad Pretest

PRE - Ficha de Registro				
Investigadores		Anchante Aybar, Percy Andersson		
		Fernandez Chavez, Davio		
Empresa		Industrias Unidas del Perú		
Investigación		Machine Learning para la toma de decisiones en la gestión de ventas, Lima 2024		
Variables		Gestión de Ventas		
Dimensión		Productividad		
Indicador		Nivel de productividad		
Formula		Ventas en Cantidad Mensual/ Meta de ventas en Cantidad		
N°	Fecha	Ventas en Cantidad Mensual	Meta de ventas en Cantidad	Productividad
1	ene-23	19313	40000	0.48
2	feb-23	17428	30000	0.58
3	mar-23	25358	42000	0.60
4	abr-23	23888	40000	0.60
5	may-23	25526	32000	0.80
6	jun-23	23454	29000	0.81
7	jul-23	22228	29000	0.77
8	ago-23	23226	30000	0.77
9	Set-23	14646	22000	0.67
10	oct-23	21564	25000	0.86
11	nov-23	10514	15000	0.70
12	dic-23	119144	180000	0.66



Jorry Baca Cuba
 DNI: 09542221
 Gerente Comercial de Industrias
 Unidas del Perú

Anexo Ficha de Registro Productividad Post Test

POST - Ficha de Registro				
Investigadores		Anchante Aybar, Percy Andersson		
		Fernandez Chavez, Davio		
Empresa		Industrias Unidas del Perú		
Investigación		Machine Learning para la toma de decisiones en la gestión de ventas, Lima 2024		
VARIABLES		Gestión de Ventas		
Dimensión		Productividad		
Indicador		Nivel de productividad		
Formula		Ventas en Cantidad Mensual/ Meta de ventas en Cantidad		
N°	Fecha	Ventas en Cantidad Mensual	Meta de ventas en Cantidad	Productividad
1	ene-24	17428	18000	0.97
2	feb-24	17428	18000	0.97
3	mar-24	25102	30000	0.84
4	abr-24	23995	26000	0.92
5	may-24	25102	26000	0.97
6	jun-24	23335	25000	0.93
7	jul-24	22478	26000	0.86
8	ago-24	23335	30000	0.78
9	sep-24	14728	18000	0.82
10	oct-24	22478	25000	0.90
11	nov-24	10915	18000	0.61
12	dic-24	119144	150000	0.79

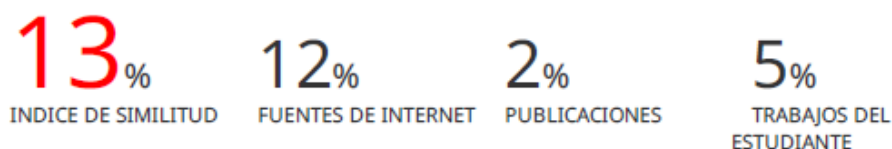


Jorry Baca Cuba
 DNI: 09542221
 Gerente Comercial de Industrias
 Unidas del Perú

Anexo 6. Reporte de similitud en software Turnitin

Turnitin Machine Learning para la toma de decisiones en la Gestión de Ventas, Lima 2024.docx

INFORME DE ORIGINALIDAD



FUENTES PRIMARIAS

1	hdl.handle.net Fuente de Internet	5%
2	Submitted to Universidad Cesar Vallejo Trabajo del estudiante	3%
3	repositorio.ucv.edu.pe Fuente de Internet	2%
4	es.slideshare.net Fuente de Internet	<1%
5	prezi.com Fuente de Internet	<1%
6	wiki2.org Fuente de Internet	<1%
7	www.tdx.cat Fuente de Internet	<1%
8	(Carlinda Leite and Miguel Zabalza). "Ensino superior: inovação e qualidade na docência", Repositório Aberto da Universidade do Porto, 2012.	<1%

Anexo 7. Anexo Carta de Autorización



CARTA DE AUTORIZACIÓN DE LA EMPRESA INDUSTRIAS UNIDAS DEL PERU S.A

Señor:

Jorry Baca Cuba
Gerente Comercial

De manera atenta manifiesto nuestro interés y conociendo de la propuesta del proyecto de investigación titulada:

"Machine Learning para la toma de decisiones en la Gestión de Ventas en Industrias Unidas del Perú, Lima 2024"

Elaborado por los estudiantes:

Anchante Aybar Percy Andersson con Código de estudiante N°7002440517

Fernandez Chavez Davio Dovise con Código de estudiante N° 6700203627.

En este sentido, la presente carta tiene como finalidad autorizar las acciones correspondientes al proyecto de investigación. Así mismo, se reitera que el material brindado y la información recolectada será netamente con fines de estudio. En caso contrario, al ser usado con otros fines se incurrirá en una falta grave.

Sin otro particular, se agradece el tiempo y dedicación que demande el presente estudio, cuyos resultados nos beneficiaran como institución.

Jorry Baca Cuba
DNI: 09542221
Gerente Comercial de Industrias
Unidas del Perú

