



**UNIVERSIDAD CÉSAR VALLEJO**

**ESCUELA DE POSGRADO**

**PROGRAMA ACADÉMICO DE MAESTRÍA EN INGENIERÍA  
DE SISTEMAS CON MENCIÓN EN TECNOLOGÍAS DE LA  
INFORMACIÓN**

Machine Learning para predecir la adquisición de plataformas  
educativas de la empresa Difucien Ecuador, 2022

**TESIS PARA OBTENER EL GRADO ACADÉMICO DE:**

Maestra en Ingeniería de Sistemas con Mención en Tecnologías de la  
Información

**AUTORA:**

Guzman Velez, Dayana Mariuxi (orcid.org/0000-0001-5651-392X)

**ASESOR:**

Dr. Pacheco Torres, Juan Francisco (orcid.org/0000-0002-8674-3782)

**LÍNEA DE INVESTIGACIÓN:**

Sistema de Información y Comunicaciones

**LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:**

Desarrollo económico, empleo y emprendimiento

**TRUJILLO – PERÚ**

**2022**

## DEDICATORIA

A mis padres, por ser un pilar fundamental e inculcarme la importancia del estudio en mi vida estudiantil.

A mis hermanos, por siempre creer en mí y brindarme su apoyo.

A mi esposo por darme su apoyo incondicional, por creer en mí y darme confianza en cada proyecto de mi vida.

## **AGRADECIMIENTO**

A Dios por brindarme bienestar, salud y permitirme tener la oportunidad de estudiar para estar preparada académicamente.

A mi familia y esposo por estar a mi lado en cada paso importante de mi vida.

## ÍNDICE DE CONTENIDOS

CARÁTULA .....	i
DEDICATORIA.....	ii
AGRADECIMIENTO.....	iii
ÍNDICE DE CONTENIDOS .....	iv
ÍNDICE DE TABLAS .....	v
ÍNDICE DE GRÁFICOS Y FIGURAS .....	vii
RESUMEN .....	viii
ABSTRACT .....	ix
I. INTRODUCCIÓN .....	1
II. MARCO TEÓRICO .....	5
III. METODOLOGÍA.....	15
3.1. Tipo y diseño de investigación .....	15
3.2. Variables y operacionalización.....	16
3.3. Población, muestra y muestreo.....	18
3.4. Técnicas e instrumentos de recolección de datos.....	19
3.5. Procedimientos .....	19
3.6. Método de análisis de datos.....	20
3.7. Aspectos éticos .....	20
IV. RESULTADOS .....	21
V. DISCUSIÓN.....	33
VI. CONCLUSIONES.....	39
VII. RECOMENDACIONES .....	40
REFERENCIAS.....	41
ANEXOS .....	49

## ÍNDICE DE TABLAS

Tabla 1 Indicadores e instrumentos de medición. ....	18
Tabla 2 Medidas estadísticas del Indicador Tiempo para elegir plataformas de adquisición de clientes. ....	21
Tabla 3 Medidas estadísticas del Indicador Tiempo para la elección de clientes potenciales para la empresa. ....	22
Tabla 4 Medidas estadísticas del Indicador Índice de adquisiciones de plataformas educativas que se realizan en un trimestre. ....	23
Tabla 5 Medidas estadísticas del Indicador Tiempo para el proceso de cierre de ventas de productos académicos. ....	24
Tabla 6 Prueba de normalidad de los indicadores pretest y postest. ....	25
Tabla 7 Prueba de rangos de Wilcoxon en las medias relacionadas al indicador Tiempo para elegir plataformas de adquisición de clientes. ....	26
Tabla 8 Prueba de Wilcoxon aplicado al indicador Tiempo para elegir plataformas de adquisición de clientes. ....	27
Tabla 9 Prueba de rangos de Wilcoxon en las medias relacionadas al indicador Tiempo para la elección de clientes potenciales para la empresa. ....	28
Tabla 10 Prueba de Wilcoxon aplicado al indicador Tiempo para la elección de clientes potenciales para la empresa. ....	28
Tabla 11 Prueba de rangos de Wilcoxon en las medias relacionadas al indicador Índice de adquisiciones de plataformas educativas que se realizan en un trimestre. ....	29
Tabla 12 Prueba de Wilcoxon aplicado al indicador Índice de adquisiciones de plataformas educativas que se realizan en un trimestre. ....	30
Tabla 13 Prueba de rangos de Wilcoxon en las medias relacionadas al indicador Tiempo para el proceso de cierre de ventas de productos académicos. ....	31
Tabla 14 Prueba de Wilcoxon aplicado al indicador Tiempo para el proceso de cierre de ventas de productos académicos. ....	31
Tabla 15 Guía de observación N°1. ....	49
Tabla 16 Guía de observación N°2. ....	49
Tabla 17 Guía de observación N°3. ....	50
Tabla 18 Guía de observación N°4. ....	50
Tabla 19 TABLA DE OPERACIONALIZACIÓN DE VARIABLES. ....	51

Tabla 20 MATRIZ DE CONSISTENCIA .....	53
Tabla 21 Tabla de indicadores de la investigación. Elaboración propia. ....	56

## ÍNDICE DE GRÁFICOS Y FIGURAS

Figura 1 Medias del Indicador Tiempo para elegir plataformas de adquisición de clientes. ....	21
Figura 2 Medias del Indicador Tiempo para la elección de clientes potenciales para la empresa.....	23
Figura 3 Medias del Indicador Índice de adquisiciones de plataformas educativas que se realizan en un trimestre. ....	24
Figura 4 Medias del Indicador Tiempo para el proceso de cierre de ventas de productos académicos. ....	25
Figura 5 Resumen de los datos de la base de datos. ....	63
Figura 6 Gráfica de Presupuesto en la adquisición de plataformas educativas. ...	63
Figura 7 Resultado de modelo 1 de regresión.....	64
Figura 8 Resultado de la predicción del modelo 1.....	64
Figura 9 Resumen del segundo modelo de predicción. ....	64
Figura 10 Resultados de las pruebas del segundo modelo.....	64
Figura 11 Resumen del tercer modelo de predicción. ....	64
Figura 12 Resultado de la predicción del tercer modelo. ....	64
Figura 13 Formulario con resultados de la predicción parte 1.....	64
Figura 14 Formulario con resultados de la predicción parte 2.....	64
Figura 15 Formulario con resultados de la predicción parte 3.....	64
Figura 16 Formulario con resultados de la predicción parte 4.....	64
Figura 17 Formulario con resultados de la predicción parte 5.....	64

## RESUMEN

El actual trabajo de investigación tuvo como objetivo predecir la adquisición de plataformas educativas. Se utilizó Machine Learning para predecir, buscando teoría relacionada se encontró que el algoritmo de regresión lineal es usado con frecuencia para predecir ventas, por ello, fue el aplicado en este estudio. Para realizar el análisis de los datos, se usó el software estadístico R Studio, en el cual se realizaron modelos predictivos para obtener una solución que permitiera mejorar el cierre de ventas de productos. En el apartado de recolección de datos, se utilizaron como instrumentos fichas de observación en el PreTest y PostTest. Luego de aplicar la solución de Machine Learning, los resultados obtenidos fueron que se redujo en un 18,01% (14 minutos) el tiempo para elegir plataformas; también, el tiempo para elección de clientes potenciales se redujo en un 32,29% (65, 50 minutos). Otro resultado a considerar, fue que el índice de adquisiciones de plataformas aumentó en un 16,50%, del mismo modo, el tiempo para cerrar ventas se redujo en un 24,35%, gracias a la implementación de Machine Learning. Como conclusión, se indica que la aplicación de un algoritmo de aprendizaje automático mejora significativamente los procesos de cierre de ventas de productos académicos.

**Palabras clave:** Machine Learning, Regresión lineal, R Studio, Predicción de ventas, Algoritmos, Plataformas educativas.



## ABSTRACT

The current research work aimed to predict the acquisition of educational platforms. Machine Learning was used for the prediction, looking for related theory, it was found that the line regression algorithm is frequently used to predict sales, therefore, it was applied in this study. To perform the data analysis, the R Studio statistical software was used, in which predictive models were made to obtain a solution that would improve the closing of product sales. In the data collection section, observation sheets were used as instruments in the PreTest and PostTest. After applying the Machine Learning solution, the results obtained were that the time to choose platforms was reduced by 18.01% (14 minutes); also, the time for choosing potential clients was reduced by 32.29% (65, 50 minutes). Another result to consider was that the platform acquisition index increased by 16.50%, in the same way, the time to close sales was reduced by 24.35%, thanks to the implementation of Machine Learning. As a conclusion, it is indicated that the application of an automatic learning algorithm significantly improves the sales closing processes of academic products.

**Keywords:** Machine Learning, Linear Regression, R Studio, Sales Prediction, Algorithms, Educational Platforms.

## I. INTRODUCCIÓN

Las plataformas virtuales educativas han marcado una gran diferencia en la educación. Esto crea una nueva forma de adquirir conocimientos, en una sociedad que usa los medios digitales en gran medida en su diario vivir. Las tecnologías actuales en gran medida ayudan a desarrollar técnicas de instrucción aprendizaje, en las cuales los estudiantes participan activamente por el uso de estas tecnologías, esto es algo que hace unos años se veía como algo lejano (Barrera Rea & Guapi Mullo, 2018).

Difucien Ecuador, es una empresa que desarrolla y comercializa plataformas digitales educativas destinadas a estudiantes de educación media y superior. Se enfoca en buscar que las instituciones de educación superior utilicen herramientas que impulsen a los estudiantes a interesarse por la investigación científica y desarrollo de habilidades de aprendizaje.

La empresa cuenta con un amplio repertorio de productos; por lo cual, en la mayoría de ocasiones se hace complicado identificar cuál es el producto indicado para diferentes tipos de clientes. La atención al cliente es un aspecto clave en sus procesos, y es por eso, que mostrar todo el catálogo es una actividad que se considera al momento ineficiente; debido a la cantidad de tiempo que demora la presentación de productos. Esto genera apuros por parte de los clientes, quienes no pueden atender de manera extendida las presentaciones de los asesores por la cantidad de tiempo que conlleva tomar en cuenta la información de todos los productos.

Según lo expuesto, se tiene un conflicto a la hora de elegir las plataformas educativas por ello el número de adquisiciones se ve perjudicado. La toma de decisión actual de los clientes es prolongada porque no se deciden a qué producto dirigirse, esto causa que el tiempo de cierre de venta se vuelva superior al que en promedio se espera. Implementando Machine Learning se espera mejorar el tiempo del cierre de ventas y la cantidad de plataformas vendidas.

Otro problema que se evidencia es el tiempo de elección de clientes potenciales, porque se deben tomar en cuenta muchos factores, entre ellos la solvencia

económica para adquirir los productos, a qué áreas de estudio se enfocan, la cantidad de sujetos que usarán los productos, entre otros lo cual conlleva análisis y tiempo.

Es por ello, que se plantea utilizar Machine Learning como una herramienta que ayude a predecir el comportamiento de los clientes y sus intereses de compra, antes de ofrecer los productos. Con esto se quiere lograr mejorar los tiempos de venta, ofrecer una buena atención al cliente y que todas las partes involucradas tengan una buena experiencia comercial.

Machine Learning es una herramienta que se utiliza con frecuencia para detectar patrones de comportamiento de un conjunto de datos. Los resultados de esta detección son determinados por la máquina y ayuda a la toma de decisiones de las organizaciones (Asto & Malpartida, 2020)

Para este trabajo de investigación se plantea la siguiente formulación del problema general: ¿De qué manera influye Machine Learning en la adquisición de plataformas educativas de la empresa Difucien Ecuador, 2022? Como variable dependiente está la *Adquisición de plataformas educativas*, variable independiente *Machine Learning* y variable interviniente *Proceso de venta de los asesores*.

La investigación tiene varios problemas específicos que se encuentran a continuación: ¿De qué manera influirá Machine Learning en el tiempo para elegir plataformas educativas? ¿De qué manera influirá Machine Learning en el número promedio de adquisiciones de plataformas educativas en un trimestre? ¿De qué manera influirá Machine Learning en el tiempo para el proceso de cierre de ventas de productos académicos? ¿De qué manera influirá Machine Learning en el tiempo para la elección de clientes potenciales para la empresa?

Considerando lo descrito, podemos argumentar las siguientes razones para justificar el trabajo de investigación: de forma operativa, se justifica porque es importante que las empresas tengan estrategias de captación de clientes, en donde el objetivo sea mejorar o incrementar las ventas, ofreciendo productos que convengan a las necesidades de los clientes y brindando una experiencia de compra eficaz y efectiva.

En el ámbito tecnológico, actualmente existen sistemas que permiten administrar procesos de venta, como por ejemplo CRM (Customer Relationship Management) o buscadores de clientes, pero estos no predicen ventas. Se justifica el proyecto de investigación porque implementando un modelo predictivo basado en Machine Learning podremos aprovechar al máximo esta tecnología para lograr llegar a las metas de venta antes del tiempo previsto.

Como justificación económica, se espera con esta implementación mejorar los ingresos brutos de la empresa porque al predecir las adquisiciones mejorarán las ventas a corto plazo.

En el ámbito social, un sistema de Machine Learning para predecir la adquisición de plataformas educativas, no solo beneficiará a la empresa, sino que ayudará a los clientes a decidir por la compra de recursos de suma importancia para los alumnos de establecimientos de nivel superior.

El objetivo general que se plantea es, predecir la adquisición de plataformas educativas de la empresa Difucien Ecuador mediante la implementación de un sistema de Machine Learning. Como objetivos específicos de la investigación tenemos los siguientes: Disminuir el tiempo para elegir plataformas educativas que van a adquirir los clientes mediante la predicción con Machine Learning. Aumentar el número promedio de adquisiciones de plataformas educativas que se realizan en un trimestre a través de Machine Learning. Disminuir el tiempo para el proceso de cierre de ventas de productos académicos mediante la implementación de Machine Learning. Disminuir el tiempo para la elección de clientes potenciales para la empresa.

En base a lo anteriormente tratado, se plantea la siguiente hipótesis: la implementación de Machine Learning permitirá predecir significativamente la adquisición de plataformas educativas de la empresa Difucien Ecuador. Entre las hipótesis específicas tenemos:

- El uso de Machine Learning ayudará a disminuir el tiempo para elegir plataformas educativas que van a adquirir los clientes.
- Machine Learning permitirá aumentar el número promedio de adquisiciones de plataformas educativas en un trimestre.

- La implementación de Machine Learning permitirá disminuir el tiempo del proceso de cierre de ventas de productos académicos.
- Machine Learning permitirá disminuir el tiempo para elección de clientes potenciales.

## II. MARCO TEÓRICO

Se han tomado en cuenta trabajos de investigación internacionales y nacionales que sustentan la presente investigación. En las revisiones internacionales encontramos a Cheriyan et al. (2018), con su trabajo realizado en Reino Unido, en dónde nos indican que las predicciones de venta generan gran impacto en los negocios y que las metodologías de minería de datos son efectivas para extraer datos para la toma de decisiones. En su trabajo, realizan un detallado análisis de modelos predictivos, estudian diversas técnicas y medidas de predicciones para mejorar las ventas futuras. Los resultados se basaron en la confiabilidad y precisión de las técnicas, en dónde encontraron que el modelo con mejores características es algoritmo Gradient Boost, que demostró en las pruebas realizadas gran exactitud al momento de realizar las predicciones de ventas.

El autor Odegua (2020), en su trabajo realizado en Nigeria nos brinda una perspectiva en la predicción de ventas de un comisariato, en donde indica que los métodos estadísticos tradicionales han dejado muchos aspectos sin abordar y que en su mayoría dan como resultados modelos predictivos que funcionan mal. En este estudio utilizan tres algoritmos de aprendizaje automático: K-Nearest Neighbor, Gradient Boosting and Random forest. Los resultados muestran que el algoritmo Random Forest tuvo un mejor rendimiento, Gradient Boosted cumple su función, pero, con dificultad y K-Nearest Neighbor, aunque rápido, tiene un rendimiento más pobre entre los tres. El algoritmo Random Forest se desempeñó mejor, ya que tenía una media absoluta más baja de error que los otros dos modelos y esto aumenta el poder predictivo.

Siguiendo la misma línea, el autor Boada (2016), en Venezuela realizó un artículo de estudio de variables causales y desarrolló un modelo estadístico exponencial multivariable como instrumento computarizado de evaluación de las variables cuantitativas y cualitativas sobre la futura demanda de productos. Se realizaron predicciones con base a índices de inflación y la cantidad de vendedoras que se estimaba, dando como resultado la mejora de procesos en los departamentos de Logística, Mercadeo, Ventas y Finanzas.

Najar (2018), en su trabajo de investigación que fue realizado en España, en la empresa Café Candelas que comercializa café e infusiones a hoteles, restaurantes, etc. Utilizó información de las ventas de los años 2015, 2016 y 2017 para analizar comportamientos en el consumo de productos. El objetivo del proyecto fue realizar un procesamiento intensivo de información empleando técnicas de modelado estadístico y machine learning. Se utilizaron diferentes librerías del software RStudio para dar soporte a las tareas de aprendizaje no supervisado y de visualización de datos para la presentación de los resultados. Como resultado se pudo determinar cuáles son las provincias con preferencias similares de consumo y poder orientar estrategias para que la organización llegue a cumplir sus metas más fácilmente. Por lo que Café Candelas debería de emplear similares técnicas de marketing para la segmentación del mercado de estas provincias.

Por otro lado, Asto & Malpartida (2020), en su trabajo de investigación elaborado en Perú, se centraron en la problemática de la empresa Vértice que utilizaba un proceso no normalizado para administrar ventas por los encargados del área de software; también, se retrasaban al realizar procesos de gestión de reportes. Concluyeron que Machine Learning incrementó el porcentaje de validez a un 82,04% habiendo anteriormente 54,64%. Consiguientemente, se asevera que Machine Learning desarrolla una manera eficiente de gestionar ventas. La ejecución de Machine Learning contribuyó a mejorar los cuadros de producción y la eficiencia en las personas encargadas del área comercial, de igual manera en los procesos relacionados a la automatización de reportes en la menor cantidad de tiempo.

En la investigación en Colombia de Cabrera (2016), se prepararon distintas metodologías para predecir el flujo diario en un hotel. El entrenamiento y ratificación se efectuó manejando cuatro metodologías de Machine Learning. La información se dividió en tres conjuntos: entrenamiento, pruebas y validación. Lo resultante demuestra efectividad del método ridge regresión en la predicción, se manejaron procedimientos para realizar datasets que se basaron en reservaciones con 90, 60, 30, 20, 15,10 y 7 días de antelación, incluyendo datos sobre días de feriados, temporadas altas, semanas y meses en los cuales la demanda es mayor.

El autor Wisesa et al. (2020), aplica en esta investigación datos de ventas B2B para el respectivo análisis. Los datos B2B pudieron proporcionar información sobre cómo la empresa de telecomunicaciones debe administrar su equipo de ventas, productos y flujos de presupuesto. Estudió y analizó modelos predictivos comprensibles utilizando una técnica de aprendizaje automático para mejorar la predicción de la venta futura. En este estudio, También utilizó la técnica de aprendizaje automático para analizar la confiabilidad de las ventas B2B. Además, al final de esta investigación, introdujo otras medidas y técnicas utilizadas para predecir las ventas.

El estudio de Davidson (2019), se basa en la predicción de las mejores ofertas para encontrar soluciones para estacionamientos. Aplicó métodos estadísticos que se usan tradicionalmente y Machine Learning para la predicción, dando como resultado que los algoritmos de Machine Learning Random Forest y Extreme Gradient Boosting tuvieron mejores rendimientos.

Otro trabajo relacionado es de Poojary et al. (2021), en el que utiliza Machine Learning para la predicción de costos aproximados de metales y joyas, esta problemática surge porque diariamente se alteran los valores de estos materiales, lo cual hace difícil conocer su precio en el futuro. Dando como consecuencia que se redujo el valor de las joyas en un 25%, lo cual resultó que se mejoren los precios para los clientes.

El trabajo de Raizada & Saini (2021), se basa en comparar métodos de Machine Learning para predecir ventas. Utilizan técnicas como regresión lineal, Random Forest, K-NN Algorithm, Support Vector Machine (SVM), para la predicción de ventas de 45 sucursales de la tienda Walmart. Esta predicción ayudará a los empresarios a tomar decisiones para realizar estrategias de marketing, descuentos y promociones.

Los autores Narayana et al. (2021), Meiyazhagan & Rajkumar (2021), Huo (2021), Ramachandra et al. (2021), Sakib (2021), Li & Luo (2019) y Fotache et al. (2021), concuerdan en que el aprendizaje automático es una técnica excelente para predicción de ventas en cualquier ámbito como por ejemplo para el uso en centros comerciales, comercio electrónico, días de promociones en Black Friday, consumos



en restaurantes, eventos deportivos e incluso superando a los tradicionales métodos estadísticos.

En cuanto a los antecedentes a Nivel Nacional, se encuentra el trabajo de investigación de Utreras (2020), cuyo objetivo fue analizar el rendimiento de modelos basados en datos de comerciales y bienes raíces que se encontraron en bases experimentales. El modelo CART fue el que obtuvo mejores resultados, logrando alcanzar desde un 80 a 94% de precisión. Esto brinda una perspectiva del buen rendimiento que se tiene al aplicar modelos de predicción con aprendizaje autónomo en bases de datos, lo cual, es una alternativa de bajo costo que se puede implementar en empresas en crecimiento.

En cuanto al enfoque conceptual se considera la siguiente fundamentación teórica. Brink et al. (2017), nos brindan un concepto de Machine Learning, en dónde se expone que consiste en técnicas que se usan para predecir patrones para la posterior toma de decisiones. El Machine Learning se puede aplicar en gran medida en el campo comercial, sus usos van desde la atención al cliente hasta la predicción de ventas, uso en procesos industriales y análisis de comportamientos. Puede resolver problemas que se dificultan realizarlos manualmente debido a la cantidad de datos por procesar, en estos casos, podría encontrar similitudes que utilizando el proceso manual no se lograría.

Machine Learning tiene una clasificación basada en outputs, en la cual se encuentran tres tipos: Aprendizaje reforzado, Aprendizaje no supervisado y Aprendizaje supervisado. El aprendizaje supervisado es cuando un profesor agrega datos para que se realice entrenamientos para que el algoritmo los aprenda para obtener los resultados deseados. El aprendizaje no supervisado no cuenta con la opción de una variable que supervise el proceso (Ospina, 2021).

El Aprendizaje supervisado es una de las clases de contrariedades de machine learning que son comunes de encontrar, en este tipo de aprendizaje la máquina aprende mediante patrones o modelos. Este algoritmo es conocido por entrenarse con preguntas y respuestas, que se conocen como características (features) y etiquetas (labels). El algoritmo almacena información que encuentra de respuestas

a ciertas preguntas, para futuras interacciones o predicciones que se realicen (Joakin, 2021).

El aprendizaje supervisado es equivalente a tener un concepto de supervisor, es decir, el objetivo principal es facilitar a un agente la medición exacta del error, esta característica se logra manteniendo un conjunto de información integrado por entradas que suministran salidas deseadas (Joakin, 2021).

El agente podrá remediar los parámetros para lograr disminuir toda la operación y funcionalidad con respecto a una pérdida global. En cada iteración, el algoritmo verifica si los datos proporcionados son coherentes y si son necesariamente maleables, la precisión aumenta considerablemente y la diferencia entre los valores esperados y pronosticados es casi nula. Se necesita que el escenario sea supervisado, para lograr entrenar el algoritmo y que este funcione sin problema con nuevas muestras proporcionadas (Joakin, 2021).

Es imprescindible que el sistema obtenga la capacidad de generalización y de evitar un inconveniente denominado sobreajuste, esto genera que ocurra un sobre aprendizaje que se debe a una capacidad en exceso, uno de los efectos más importantes de este inconveniente es predecir de manera correcta solo las muestras usadas en la etapa de entrenamiento, esto ocurre sobre todo si el índice de error para las muestras restantes es sumamente alta. Un ejemplo de esto es si se realiza un proceso para reconocer fotografías, de distintos animales y reconocer cuales son perros y gatos (Joakin, 2021).

Lo que se debe realizar en primera instancia es alimentar el sistema con una serie de fotografías para decir si ciertamente se trata de un perro o gato, a este proceso es a lo que conocemos como entrenamiento, lo importante de esto, es que luego proporcionaremos fotografías nuevas y el sistema tendrá evaluar distintos parámetros para devolvernos una respuesta y conocer si es perro o gato (Joakin, 2021).

Los inconvenientes del aprendizaje supervisado pueden dividirse en complicaciones de regresión y clasificación. Un inconveniente de clasificación se genera la variable de salida es una clase, por ejemplo “amarillo”, “verde” o “salud”

y “sin salud”. El aprendizaje supervisado por regresión se presenta cuando la variable de salida es un parámetro real, como “dinero” o “gravedad” (Serra, 2021).

Los problemas más comunes de la clasificación y la regresión son la recomendación y la predicción de sucesiones temporales.

Los ejemplos de algoritmos supervisados de aprendizaje automático más reconocidos son los siguientes: Soporte de máquinas de vectores (Support Vector Machine) para inconvenientes de clasificación. Regresión lineal para inconvenientes de regresión. Bosque aleatorio (Random Forest) para dificultades de categorización y regresión (Serra, 2021).

El aprendizaje no supervisado es el método en el que solo hay datos de entrada (X) y no contiene variables de salida. La principal función del aprendizaje no supervisado es componer la forma o distribución de los datos de entrada para reconocerlos y generar aprendizaje de ellos (Serra, 2021).

Un ejemplo de este tipo de aprendizaje en la vida cotidiana es catalogar monedas de diferente denominación en pilas separadas. Nadie le enseñó al algoritmo como clasificarlas, pero solo observando sus cualidades como la denominación, puede distinguir las monedas y a apilarlas en los conjuntos correctos (Serra, 2021).

Es conocido como aprendizaje no supervisado porque, si lo diferenciamos con el aprendizaje supervisado que ya se estudió anteriormente, aquí no existen respuestas correctas y no hay un agente, Los algoritmos se alojan en los dispositivos para conocer y mostrar la distribución interesante en los datos (Serra, 2021).

Los inconvenientes del aprendizaje no supervisado pueden dividirse en asociación y agrupamiento. Asociación: este tipo de aprendizaje se basa en tener normas de asociación para descubrir aquellas que se relacionen con grandes proporciones de datos, por ejemplo, alguien compra X pero también puede comprar Y. Agrupación: en este tipo se expresan con clasificaciones inherentes de información, como clasificación o agrupación de clientes por procedimiento de compra (Serra, 2021).

Los tipos más habituales de algoritmos de aprendizaje no supervisado son los siguientes: Algoritmo de Apriori para inconvenientes de aprendizaje de normas de agrupación. k-means para inconvenientes de agrupamiento (Serra, 2021).

El aprendizaje reforzado es un procedimiento más complicado y es el modelo menos común de los algoritmos, pero sus resultados son sorprendentes. Usa un tipo de recompensas para generar aprendizaje y no utiliza etiquetas. Es comparado con el aprendizaje por refuerzo en el ámbito de la psicología. En este proceso se usan comentarios tanto negativos como positivos de vez en cuando para robustecer los comportamientos. Es como entrenar una mascota, si tiene excelentes comportamientos se les van a otorgar obsequios que se vuelven recurrentes. Los malos actos se castigan y son por lo general los menos comunes (Serra, 2021).

Este método basado en recompensas es pieza clave en el aprendizaje reforzado. De vez en cuando este tipo de aprendizaje por refuerzo se conoce como el que aprende errores. En primera instancia, si practicamos colocando el algoritmo de refuerzo en un ambiente, habrá muchos errores. Si adicionamos en el algoritmo una buena señal para los comportamientos positivos y una mala para los negativos, se puede reforzar el algoritmo para optar por elegir los buenos comportamientos sobre los malos (Serra, 2021).

Predecir ventas se refiere al alcance o estimación que se tiene en comercializaciones de productos de una empresa en cierto tiempo establecido y dedicado hacia un nicho de mercado en específico. Es una metodología que se basa en indicar a las organizaciones cuáles serán sus ingresos en base a los productos ofertados, cuál mejorará la manera de toma de decisiones e indicará hacia dónde deben apuntar (Carrillo, 2020).

Las predicciones mediante el aprendizaje automático, que es una clasificación de inteligencia artificial, permiten tener resultados precisos, fundado por tendencias de mercado y esperando tener una mejora en procesos de venta en su utilización. Este proceso no es lineal, en comparación a los tradicionales, se adapta a contingencias mediante procesos matemáticos. Según Forbes se puede originar un valor agregado de \$2.6T Ventas y Marketing y \$2T en producción y proyección de cadena de insumos en el año 2020. En el presente, varios países utilizan la inteligencia

artificial para mejorar los métodos, entre ellos están Estados Unidos, Alemania y Dinamarca (Gonzalez & Robayo, 2020).

Las proyecciones de tendencia, promedio móvil simple, análisis de regresión, técnica Box Jenkins (ARIMA), suavización exponencial promedio móvil ponderado, son modelos tradicionales, precisos, combinados y probabilísticos. Éstos son manejados para obtener predicciones y, sobre todo, ventajas y desventajas entre ellos. De todas formas, estos algoritmos no son capaces de obtener resultados satisfactorios en un ambiente cambiante a menudo y con altas indecisiones. Ahora en los actuales momentos se busca algoritmos como regresión de soporte vectorial y redes neuronales artificiales que se basan en metodologías que se basan en sistemas no lineales (Peña, 2018).

Clasificación de los modelos tradicionales de pronóstico. Los pronósticos son unos métodos que permiten la organización de negocios en un tiempo lejano. Se debe elegir un modelo que permita predecir y aumentar los resultados antiguos, mediante procesos que involucren tecnología para procesar inmensas cantidades de datos y que brinde como resultado algo relacionado con la situación actual. Según el autor, los siguientes son modelos de predicción basados en series de tiempo: Técnica Box Jenkins. Análisis de regresión. Promedio móvil simple. Suavización exponencial. Proyecciones de tendencia. Promedio móvil ponderado (Peña, 2018).

Las técnicas de pronósticos se basan en una comparación de los diferentes errores resultantes, para calcular la efectividad de un pronóstico. Cuando nos referimos a errores hablamos sobre las particularidades que ocurrieron entre lo que sucedió en realidad y lo pronosticado. En las ciencias estadísticas, los fallos o errores se denominan residuales. Siempre que el resultado de la predicción no supere el rango de confianza, no se lo cataloga como un error. Por ello, su utilización constante hace tender a una divergencia como un error (Peña, 2018).

R es un lenguaje de programación enfocado al ámbito estadístico, pero en realidad tiene más funcionalidades que permiten tener mayor detalle en la interacción con los datos. Brinda un espacio para utilizar funciones de procesamiento de datos sin tener que realizar un código extenso (Navarro Danielle, 2019).

Las principales funcionalidades son las matrices y los vectores que se utilizan para procesar grandes cantidades de datos de manera sencilla. Una ventaja de R Studio es que se pueden crear un sin número de gráficas dependiendo lo que estemos buscando (Shipunov Alexey, 2019).

La regresión lineal es una de las metodologías del campo de la estadística que se usan con mayor frecuencia. Una de las razones por las que se le atribuye su frecuente uso es la excelencia de sus modelos, los cuales permiten que se reflejen los procesos del objeto de estudio de manera exacta, concisa y elemental (Hernández-Lalinde et al., 2019).

La regresión lineal se basa en el estudio o entendimiento de cómo los resultados numéricos están enlazados con los valores de la variable de entrada. La técnica se utiliza con frecuencia de dos formas: para generar respuesta hacia una predicción o para asociar un conjunto de datos (Jon, Nicholas, 2017).

También se conoce como un conjunto de técnicas con mucho poder en el ámbito estadístico, por lo general se la involucra con gráficos o diagramas de dispersión trazados por una línea recta. Se utilizan para realizar predicciones o para analizar la relación de variables cuantitativas (Diez David et al., 2022).

Según el autor se puede utilizar un modelo de regresión lineal para comprender la relación entre dos variables cuantitativas y conocer si el resultado es estadísticamente significativo. Además, permite realizar modelos que integren múltiples variables y la predicción dependerá de los valores de las mismas (Poldrack, Russell A., 2021).

Un modelo de regresión lineal simple tiene la funcionalidad de definir la relación que existe entre una variable que se conoce como "Predictora" que es la de entrada, y la salida que sería el resultado del sistema. Este modelo también es conocido con sus siglas (SLR) simple linear regression (Lilja & Linse, 2022).

Lo interesante de la regresión lineal es que se pueden integrar variables explicativas para mejorar la variación de nuestro resultado, a esto se lo conoce como regresión lineal múltiple. Con este tipo se pueden obtener predicciones claras y precisas (Roback Paul Legler, 2021).

La regresión lineal múltiple y otros modelos que se basan en lo experimental son casos que se utilizan con frecuencia en la regresión lineal. Pueden presentarse al contemplar una cantidad de población y un modelo referente a los términos de la variable predictora (Olive David J., 2017).

### **III. METODOLOGÍA**

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

##### **Tipo de investigación**

Para este trabajo de investigación, según su finalidad se consideró un tipo de investigación de tipo aplicada. Según el autor Sánchez (2020), la define como un modelo para una investigación práctica o funcional que saca provecho de lo aprendido en la aplicación de una investigación teórica que se enfoca en solucionar problemáticas.

Según su enfoque se consideró una investigación de tipo Cuantitativa, el autor Flores (2019), la describe como una investigación que estudia acontecimientos que están sujetos a medición mediante el uso de métodos estadísticos para el procesamiento de la información recolectada.

Tomando en cuenta el método de obtención de datos la actual investigación fue de tipo experimental, que según Guevara, Verdesoto, Alban et al. (2020), se considera un procedimiento en el cual se somete a un conjunto de sujetos a ciertas circunstancias, procesos o estímulos (variable independiente), para conocer la respuesta que se provocan (variable dependiente). En este tipo de investigación se manipulan más de una variable de estudio, con el objetivo de inspeccionar si existe algún incremento o disminución de las mismas y su resultado en comportamientos examinados.

##### **Diseño de investigación**

Para este estudio de investigación se planteó un diseño experimental de tipo pre-experimental. En este tipo de diseño los investigadores se enfocan en estudiar variables para analizar conductas que resultan de la investigación. Dicho de otro modo, lo que busca este diseño experimental es restaurar los parámetros de la variable independiente y verificar los resultados o efectos de la variable dependiente. Cuando se quiere efectuar esto, se deben tomar en cuenta situaciones drásticas o rigurosas para poder comprender las razones por las cuales se origina un acontecimiento (Serrano et al., 2020).



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

Pre-test – Tratamiento – Post-test

Dónde:

G: Grupo experimental

X: Tratamiento

O1-O2: Mediciones pre-test/post-test de la solución de Machine Learning

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

En este apartado se presentan las definiciones operacionales y conceptuales de las variables sujetas a investigación.

Definición conceptual de la variable independiente Machine Learning, el aprendizaje automático, es una metodología científica que se basa en utilizar computadores y aparatos inteligentes para que estudien y aprendan a sacar por sí mismos las muestras y semejanzas que existen en los datos. Estas muestras o patrones sirven para tomar decisiones y pronosticar conductas (Valdez Alvarado, 2018).

El Aprendizaje Automático es una clasificación de la Inteligencia Artificial, en donde los dispositivos pueden tener un aprendizaje de ellos mismos, sin estar exactamente codificados por las personas. La metodología de aprendizaje automático puede formar muestras o patrones estudiando información de tiempos pasados que se los conoce como Datos de entrenamiento. Estos patrones son utilizados para predecir eventos en el futuro (Valdez Alvarado, 2018).

Definición operacional de Machine Learning, este aprendizaje es el que se usa para comprender métodos y obtener un modelamiento de conocimiento mediante el aprendizaje. Machine Learning está inmerso en muchas áreas, las que mayormente destacan son las siguientes, transporte autónomo, seguridad informática, genética y su finalidad es reconocer o pronosticar patrones complejos fundados en data.

La escala de medición para esta variable de estudio es de Razón porque es de tipo cuantitativa.

Definición conceptual de la variable dependiente Predecir adquisición/ventas, la predicción de ventas básicamente está basada en la comprensión de ventas de un servicio o producto en cada una de las etapas de su proceso y que tienden a asimilar que los nuevos productos van a tener un comportamiento como los que se usan como guías. Es un procedimiento de costo no tan elevado y que tiene como desventaja entender que las variables de procedimientos pasados se conservan y que los efectos que tienen los nuevos productos son dotados por los comportamientos de los objetos que se toman para estudio (Moreno,2019).

Definición operacional de Predecir adquisición/ventas, es que se define como un estimado de una exigencia futura de productos. Una predicción es desarrollada mediante técnicas cualitativas y cuantitativas, también puede ser por una combinación de las mismas, y principalmente se enfocan en componentes intrínsecos/internos o extrínsecos/externos. Por lo general los métodos de predicción pretenden pronosticar alguna de las unidades de la demanda: estacionalidad, aleatoriedad, ciclo y tendencia.

Los indicadores de la variable dependiente de investigación son los siguientes:

1. Tiempo para elegir plataformas de adquisición de clientes.
2. Tiempo para la elección de clientes potenciales para la empresa.
3. Índice de adquisiciones de plataformas educativas que se realizan en un trimestre.
4. Tiempo para el proceso de cierre de ventas de productos académicos.

En la siguiente tabla se encuentran los indicadores con su instrumento de medición.

**Tabla 1 Indicadores e instrumentos de medición.**

Indicador	Instrumento	Cantidad	Fórmula
Tiempo para elegir plataformas de adquisición de clientes	Ficha de observación	53	Tiempo para elegir plataformas = (Tiempo final para elección de plataformas educativas - Tiempo de inicio para elección de plataformas educativas)
Tiempo para la elección de clientes potenciales para la empresa.	Ficha de observación	53	Tiempo para la elección de clientes potenciales = (Tiempo final elección de clientes potenciales - Tiempo de inicio para elección de clientes potenciales)
Índice de adquisiciones de plataformas educativas que se realizan en un trimestre.	Ficha de observación	53	Índice de adquisiciones de plataformas educativas = (Plataformas educativas adquiridas / Plataformas educativas pendientes de adquirirse en un trimestre) *100
Tiempo para el proceso de cierre de ventas de productos académicos.	Ficha de observación	53	Tiempo para la elección de clientes potenciales = (Tiempo final para el proceso de cierre de venta - Tiempo de inicio para el proceso de cierre de venta)

**Fuente: Elaboración propia.**

La escala de medición es de Razón porque permitirá medir de forma cuantitativa la variable.

### 3.3. Población, muestra y muestreo

#### Población

La población se fundamenta en tener un conjunto de componentes que contienen particularidades que se toman como objeto de estudio. La población que se utilizará para esta investigación son los 60 clientes de la empresa.

#### Muestra

Se va a determinar mediante una fórmula la muestra de tipo aleatorio simple, considerando un nivel de confianza del 95% y un margen de error del 5%. El resultado es una muestra de 53 clientes.

$$\text{Tamaño de la muestra} = \frac{z^2 * p * q * N}{[e^2 * (N - 1)] + z^2 * p * q}$$

Dónde:

N = 60

Z = Nivel de confianza (95%) = 1.96

$$p = 0.5$$

$$q = 0.5$$

$$e = 5\%$$

$$\text{Tamaño de la muestra} = \frac{(1.96^2) * (0.5) * (0.5) * (60)}{(0.05^2)(60-1) + (1.96^2) * (0.5) * (0.5)} = 53 \text{ clientes.}$$

## **Muestreo**

El muestreo seleccionado es probabilístico Aleatorio Simple porque cada elemento se selecciona de manera independiente y toma de manera ecuánime la elección de las muestras que surgen de una población.

### **3.4. Técnicas e instrumentos de recolección de datos**

#### **Técnica de recolección de datos**

La técnica que se utilizará en la actual investigación, es la observación mediante estas técnicas podremos cumplir lo estipulado en los indicadores.

#### **Instrumentos**

Para el presente estudio se utilizaron fichas para registrar los resultados obtenidos de la técnica de observación. Estas fichas ayudaron a lograr los objetivos planteados en la introducción.

#### **Validez**

Los instrumentos utilizados en la investigación fueron evaluados por tres expertos que están especializados en el área de estudio. Ellos validaron la precisión de los recursos utilizados en las pruebas que se aplicaron a la muestra.

### **3.5. Procedimientos**

Para esta investigación se realizó una recolección de datos mediante la técnica de observación. Además, se utilizaron algoritmos y técnicas de Machine Learning para lograr los objetivos de la investigación. Para realizar el procedimiento correspondiente se tomaron datos históricos de las ventas y clientes de la empresa.

Se utilizó el método inductivo en el apartado de población y muestra que es estudiar de lo particular a lo general, esperando que la parte estudiada (muestra) sea distintiva o lo que representa el entorno (población); para que mediante esto se puedan asegurar las conclusiones obtenidas de la investigación.

### **3.6. Método de análisis de datos**

El análisis de los datos se realizó por medio de la herramienta el uso del software IBM SPSS Statistics 25, la cual nos brindó un enfoque estadístico descriptivo e inferencial acerca de los datos procesados. El análisis descriptivo se mostró mediante figuras y tablas con medidas estadísticas de cada uno de los indicadores.

Con respecto al análisis inferencial, se usó el Test Kolmogorov-Smirnov, como prueba de normalidad de datos, debido a que la muestra es superior a 50, el resultado permitió conocer si se utilizan pruebas paramétricas o no paramétricas. Con respecto a la comprobación de hipótesis se aplicó el método estadístico Wilcoxon porque los datos no son normales en dónde se evaluaron medias del grupo de estudio para comprobar si la hipótesis es nula o válida.

### **3.7. Aspectos éticos**

Se tomó en cuenta información de la empresa en este estudio con la debida autorización y firmando un acuerdo de confidencialidad. En cuanto a la estructura de la investigación se toman en cuenta las normativas dictaminadas por la Universidad César Vallejo para elaborar trabajos de posgrado. Adicionalmente, para comprobar la originalidad de la información se utilizará el sistema antiplagio Turnitin.

## IV. RESULTADOS

### Estadística descriptiva

#### Medidas estadísticas del Indicador Tiempo para elegir plataformas de adquisición de clientes.

*Tabla 2 Medidas estadísticas del Indicador Tiempo para elegir plataformas de adquisición de clientes.*

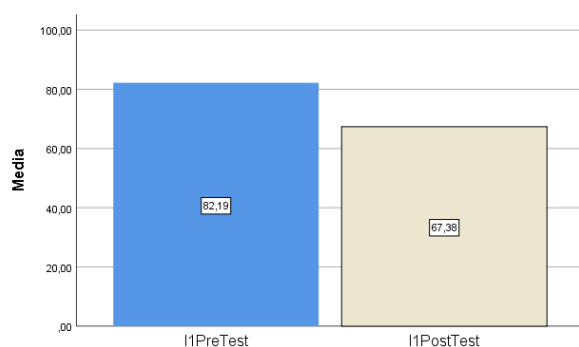
Estadísticos descriptivos					
	N	Mínimo	Máximo	Media	Desv. Desviación
Tiempo para elegir plataformas – PreTest	53	48,00	123,00	82,1864	23,43403
Tiempo para elegir plataformas – PostTest	53	28,00	106,20	67,3792	23,50918
N válido (por lista)	53				

*Fuente: Elaboración propia.*

En la tabla 2 se muestran las medidas estadísticas del indicador Tiempo para elegir plataformas de adquisición de clientes. Las medidas dieron como resultado que la media va de 82,1864 minutos (Pretest) a 67,3792 minutos (PostTest). Se evidencia que existe una mejora en la reducción de tiempo de elección de plataformas en el PostTest teniendo una diferencia de 14,8072 minutos (18,01%).

En los rangos de tiempo se presentó un incremento del pretest (48,00 a 123,00) al postTest (28,00 a 106,20), pasando de 75 al 78.2. Con respecto a la desviación estándar en el Pretest es  $\pm 23,43403$  (28,51% de la media) y en el postTest  $\pm 23,50918$  (34,89% de la media).

*Figura 1 Medias del Indicador Tiempo para elegir plataformas de adquisición de clientes.*



*Fuente: Elaboración propia.*

En la figura 1 correspondiente a las medias del indicador Tiempo para elegir plataformas de adquisición de clientes, se observa una reducción del 18,01% del postTest en relación al pretest.

### Medidas estadísticas del Indicador Tiempo para la elección de clientes potenciales para la empresa.

**Tabla 3 Medidas estadísticas del Indicador Tiempo para la elección de clientes potenciales para la empresa.**

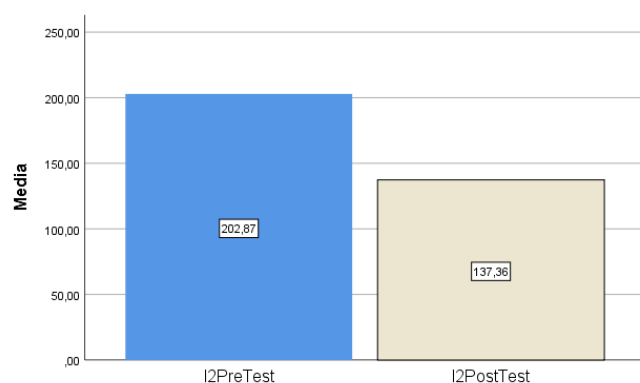
Estadísticos descriptivos					
	N	Mínimo	Máximo	Media	Desv. Desviación
Tiempo para elección de clientes – PreTest	53	96,00	336,00	202,8679	94,43721
Tiempo para elección de clientes – PostTest	53	48,00	240,00	137,3585	76,34313
N válido (por lista)	53				

**Fuente: Elaboración propia.**

En la tabla 3 se observan las medidas estadísticas del indicador Tiempo para la elección de clientes potenciales para la empresa. Las medidas dieron como resultado que la media va de 202,8679 minutos (Pretest) a 137,3585 minutos (PostTest). Existe una mejora en la reducción de tiempo de elección de clientes potenciales en el PostTest teniendo una diferencia de 65,5094 minutos (32,29%).

En los rangos de tiempo se presentó un incremento del pretest (96,00 a 336,00) al postTest (48,00 a 240,00), pasando de 240 al 192. Con respecto a la desviación estándar en el Pretest es  $\pm 94,43721$  (46,55% de la media) y en el postTest  $\pm 23,50918$  (55,57% de la media).

**Figura 2 Medias del Indicador Tiempo para la elección de clientes potenciales para la empresa.**



**Fuente: Elaboración propia.**

En la figura 2 que muestra las medias del indicador Tiempo para la elección de clientes potenciales para la empresa, se observa una reducción del 32,29% del postTest en relación al pretest.

### **Medidas estadísticas del Indicador Índice de adquisiciones de plataformas educativas que se realizan en un trimestre.**

**Tabla 4 Medidas estadísticas del Indicador Índice de adquisiciones de plataformas educativas que se realizan en un trimestre.**

<b>Estadísticos descriptivos</b>					
	N	Mínimo	Máximo	Media	Desv. Desviación
Índice de adquisiciones de plataformas – PreTest	53	,00	100,00	34,9057	22,64888
Índice de adquisiciones de plataformas – PostTest	53	,00	100,00	51,4151	23,72473
N válido (por lista)	53				

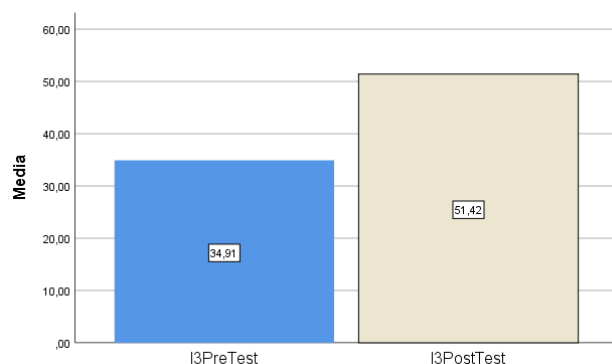
**Fuente: Elaboración propia.**

En la tabla 4 se muestran las medidas estadísticas del indicador Índice de adquisiciones de plataformas educativas que se realizan en un trimestre. Las medidas dieron como resultado que la media porcentual va de un 34,9057% (Pretest) a un 51,4151% (PostTest). Se evidencia que existe una mejora en las adquisiciones de plataformas en el PostTest teniendo una diferencia de 16,5094%.

En los rangos de tiempo se presentaron valores iguales en el pretest y postTest (0,00 a 100,00). Con respecto a la desviación estándar en el Pretest es  $\pm 22,64888$  (64,89% de la media) y en el postTest  $\pm 23,72473$  (46,14% de la media).



**Figura 3 Medias del Indicador Índice de adquisiciones de plataformas educativas que se realizan en un trimestre.**



**Fuente:** *Elaboración propia.*

En el gráfico 3 perteneciente a las medias del indicador Índice de adquisiciones de plataformas educativas que se realizan en un trimestre, se observa un aumento del 16,5094%. del postTest en relación al pretest.

### **Medidas estadísticas del Indicador Tiempo para el proceso de cierre de ventas de productos académicos**

**Tabla 5 Medidas estadísticas del Indicador Tiempo para el proceso de cierre de ventas de productos académicos**

<b>Estadísticos descriptivos</b>					
	N	Mínimo	Máximo	Media	Desv. Desviación
Tiempo para el proceso de cierre de ventas – PreTest	53	168,00	5840,01	2246,0762	1694,11442
Tiempo para el proceso de cierre de ventas - PostTest	53	120,00	4380,00	1698,9434	1379,27359
N válido (por lista)	53				

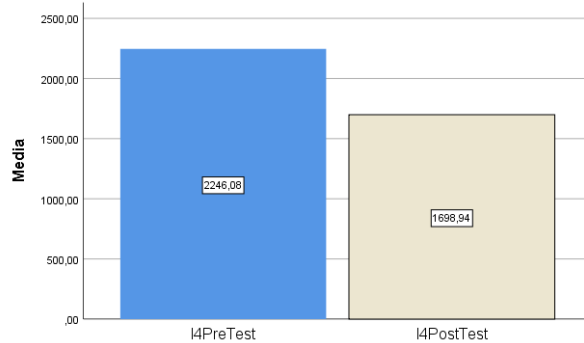
**Fuente:** *Elaboración propia.*

En la tabla 5 se observan las medidas estadísticas del indicador Tiempo para el proceso de cierre de ventas de productos académicos. Las medidas dieron como resultado que la media va de 2246,0762 minutos (Pretest) a 1698,9434 minutos (PostTest). Existe una mejora en la reducción de tiempo de proceso de cierre de ventas de productos académicos en el PostTest teniendo una diferencia de 547,1328 minutos (24,35%).

En los rangos de tiempo se presentó un incremento del pretest (168,00 a 5840,01) al postTest (120,00 a 4380,00), pasando de 5672,01 al 4260,00. Con respecto a la

desviación estándar en el Pretest es  $\pm 1694,11442$  (75,42% de la media) y en el postTest  $\pm 1379,27359$  (81,18% de la media).

**Figura 4 Medias del Indicador Tiempo para el proceso de cierre de ventas de productos académicos.**



**Fuente: Elaboración propia.**

En la figura 4 se muestran las medias del indicador Tiempo para el proceso de cierre de ventas de productos académicos, se observa una reducción del 24,35% del postTest en relación al pretest.

## Estadística inferencial

**Tabla 6 Prueba de normalidad de los indicadores pretest y posttest**

	Kolmogorov-Smirnov <sup>a</sup>			
	Estadístic	o	gl	Sig.
Tiempo para elegir plataformas – PreTest	,143	53	,008	
Tiempo para elegir plataformas – PostTest	,144	53	,008	
Tiempo para elección de clientes – PreTest	,263	53	,000	
Tiempo para elección de clientes – PostTest	,257	53	,000	
Índice de adquisiciones de plataformas – PreTest	,329	53	,000	
Índice de adquisiciones de plataformas – PostTest	,222	53	,000	
Tiempo para el proceso de cierre de ventas - PreTest	,249	53	,000	
Tiempo para el proceso de cierre de ventas - PostTest	,305	53	,000	

**Fuente: Elaboración propia.**

Se utilizó la prueba de Normalidad Kolmogorov-Smirnova porque la muestra es mayor a 50. En la tabla 6 se muestran los resultados del nivel de significancia que es Sig = 0,008 en el pretest y postTest del primer indicador. Los valores de significancia de los indicadores 2, 3 y 4 dio como resultado Sig = 0,000. Con estos

valores se evidencia que no existe normalidad porque los valores de significancia son menores a 0,050

Según los resultados de los valores de significancia se utilizaron pruebas no paramétricas. Para comparar el valor de las medias se tomó en cuenta la prueba Wilcoxon.

### Prueba de hipótesis

#### Hipótesis de investigación 1:

- **H1:** El uso de Machine Learning ayudará a disminuir el tiempo para elegir plataformas educativas que van a adquirir los clientes.
- **Indicador:** Tiempo para elegir plataformas de adquisición de clientes.

#### Hipótesis estadísticas

##### Definición de variables:

- **TPEPACa:** Tiempo para elegir plataformas de adquisición de clientes antes de la implementación
- **TPEPACd:** Tiempo para elegir plataformas de adquisición de clientes después de la implementación
- **Hipótesis Nula Ho:** El uso de Machine Learning ayudará a aumentar el tiempo para elegir plataformas educativas que van a adquirir los clientes.

$$H_0: TPEPACd - TPEPACa \geq 0$$

- **Hipótesis Alterna Ha:** El uso de Machine Learning ayudará a disminuir el tiempo para elegir plataformas educativas que van a adquirir los clientes.

$$H_a: TPEPACd - TPEPACa < 0$$

### Prueba Wilcoxon Indicador Tiempo para elegir plataformas de adquisición de clientes.

*Tabla 7 Prueba de rangos de Wilcoxon en las medias relacionadas al indicador Tiempo para elegir plataformas de adquisición de clientes.*

		Rangos		
		N	Rango promedio	Suma de rangos
PostTest – PreTest	Rangos negativos	53 <sup>a</sup>	27,00	1431,00
	Rangos positivos	0 <sup>b</sup>	,00	,00
	Empates	0 <sup>c</sup>		

Total	53		
-------	----	--	--

- a. PostTest < PreTest
- b. PostTest > PreTest
- c. PostTest = PreTest

**Fuente:** *Elaboración propia.*

En la tabla 7 se muestran los valores de la prueba de rangos, en dónde 53 pares se encuentran en rangos negativos (reducción de tiempo) y no existen pares en el rango positivo (aumento de tiempo), lo cual quiere decir que Machine Learning redujo el tiempo para elegir plataformas.

**Tabla 8 Prueba de Wilconxon aplicado al indicador Tiempo para elegir plataformas de adquisición de clientes.**

Estadísticos de prueba <sup>a</sup>	
	PostTest - PreTest
Z	-6,334 <sup>b</sup>
Sig. asintótica(bilateral)	,000

- a. Prueba de rangos con signo de Wilcoxon
- b. Se basa en rangos positivos.

**Fuente:** *Elaboración propia.*

La tabla 8 nos muestra un valor de significancia de 0,000 que es menor a 0,050, entonces, se rechaza la hipótesis nula y se acepta la hipótesis de la investigación que indica que Machine Learning ayudó a disminuir significativamente el tiempo para elegir plataformas educativas que van a adquirir los clientes.

### Hipótesis de investigación 2:

- **H1:** Machine Learning permitirá disminuir el tiempo para elección de clientes potenciales.
- **Indicador:** Tiempo para la elección de clientes potenciales para la empresa.

### Hipótesis estadísticas

#### Definición de variables:

- **TPECPEa:** Tiempo para la elección de clientes potenciales para la empresa antes de la implementación
- **TPECPEd:** Tiempo para la elección de clientes potenciales para la empresa después de la implementación

- **Hipótesis Nula  $H_0$ :** Machine Learning permitirá aumentar el tiempo para elección de clientes potenciales.

$$H_0: TPECPEd - TPECPEa \geq 0$$

- **Hipótesis Alterna  $H_a$ :** Machine Learning permitirá disminuir el tiempo para elección de clientes potenciales.

$$H_a: TPECPEd - TPECPEa < 0$$

**Prueba Wilcoxon Indicador Tiempo para la elección de clientes potenciales para la empresa.**

*Tabla 9 Prueba de rangos de Wilcoxon en las medias relacionadas al indicador Tiempo para la elección de clientes potenciales para la empresa.*

		Rangos		
		N	Rango promedio	Suma de rangos
PostTest – PreTest	Rangos negativos	53 <sup>a</sup>	27,00	1431,00
	Rangos positivos	0 <sup>b</sup>	,00	,00
	Empates	0 <sup>c</sup>		
	Total	53		

a. PostTest < PreTest

b. PostTest > PreTest

c. PostTest = PreTest

**Fuente: Elaboración propia.**

En la tabla 9 se muestran los valores de la prueba de rangos, en dónde 53 pares se encuentran en rangos negativos (reducción de tiempo) y no existen pares en el rango positivo (aumento de tiempo), lo cual quiere decir que Machine Learning redujo el tiempo para la elección de clientes potenciales para la empresa.

*Tabla 10 Prueba de Wilcoxon aplicado al indicador Tiempo para la elección de clientes potenciales para la empresa.*

Estadísticos de prueba <sup>a</sup>	
	PostTest - PreTest
Z	-6,445 <sup>b</sup>
Sig. asintótica(bilateral)	,000

a. Prueba de rangos con signo de Wilcoxon

b. Se basa en rangos positivos.

**Fuente: Elaboración propia.**

La tabla 10, nos muestra un valor de significancia de 0,000 que es menor a 0,050, entonces, se rechaza la hipótesis nula y se acepta la hipótesis de la investigación que indica que Machine Learning permitió disminuir significativamente el tiempo para elección de clientes potenciales.

### Hipótesis de investigación 3:

- **H1:** Machine Learning permitirá aumentar el número promedio de adquisiciones de plataformas educativas en un trimestre.
- **Indicador:** Índice de adquisiciones de plataformas educativas que se realizan en un trimestre

### Hipótesis estadísticas

#### Definición de variables:

- **IAPETa:** Índice de adquisiciones de plataformas educativas que se realizan en un trimestre antes de la implementación
- **IAPETd:** Índice de adquisiciones de plataformas educativas que se realizan en un trimestre después de la implementación
- **Hipótesis Nula Ho:** Machine Learning permitirá disminuir el número promedio de adquisiciones de plataformas educativas en un trimestre.

$$H_0: IAPETd - IAPETa \geq 0$$

- **Hipótesis Alternativa Ha:** Machine Learning permitirá aumentar el número promedio de adquisiciones de plataformas educativas en un trimestre.

$$H_a: IAPETd - IAPETa < 0$$

### Prueba Wilcoxon del indicador Índice de adquisiciones de plataformas educativas que se realizan en un trimestre.

*Tabla 11 Prueba de rangos de Wilcoxon en las medias relacionadas al indicador Índice de adquisiciones de plataformas educativas que se realizan en un trimestre.*

		Rangos		
		N	Rango promedio	Suma de rangos
PostTest – PreTest	Rangos negativos	1 <sup>a</sup>	17,00	17,00
	Rangos positivos	34 <sup>b</sup>	18,03	613,00
	Empates	18 <sup>c</sup>		
	Total	53		

a. PostTest < PreTest

b. PostTest > PreTest

c. PostTest = PreTest

**Fuente: Elaboración propia.**

En la tabla 11 se muestran los valores de la prueba de rangos, en donde 1 par se encuentran en rangos negativos (reducción), 34 pares en el rango positivo (aumento), y 18 en empate lo cual quiere decir que Machine Learning aumentó el número promedio de adquisiciones de plataformas educativas en un trimestre.

**Tabla 12 Prueba de Wilconxon aplicado al indicador Índice de adquisiciones de plataformas educativas que se realizan en un trimestre.**

Estadísticos de prueba <sup>a</sup>	
	PostTest - PreTest
Z	-5,460 <sup>b</sup>
Sig. asintótica(bilateral)	,000

a. Prueba de rangos con signo de Wilcoxon

b. Se basa en rangos negativos.

**Fuente: Elaboración propia.**

La tabla 12 nos muestra un valor de significancia de 0,000 que es menor a 0,050, entonces, se rechaza la hipótesis nula y se acepta la hipótesis de la investigación que indica que Machine Learning permitió aumentar el número promedio de adquisiciones de plataformas educativas en un trimestre.

#### **Hipótesis de investigación 4:**

- **H1:** La implementación de Machine Learning permitirá disminuir el tiempo del proceso de cierre de ventas de productos académicos.
- **Indicador:** Tiempo para el proceso de cierre de ventas de productos académicos

#### **Hipótesis estadísticas**

##### **Definición de variables:**

- **TPCVPAa:** Tiempo para el proceso de cierre de ventas de productos académicos antes de la implementación
- **TPCVPA d:** Tiempo para el proceso de cierre de ventas de productos académicos después de la implementación

- **Hipótesis Nula Ho:** La implementación de Machine Learning permitirá aumentar el tiempo del proceso de cierre de ventas de productos académicos.

$$H_0: \text{TPCVPA}_d - \text{TPCVPA}_a \geq 0$$

- **Hipótesis Alternativa Ha:** La implementación de Machine Learning permitirá disminuir el tiempo del proceso de cierre de ventas de productos académicos.

$$H_a: \text{TPCVPA}_d - \text{TPCVPA}_a < 0$$

**Prueba Wilcoxon Indicador Tiempo para el proceso de cierre de ventas de productos académicos.**

**Tabla 13 Prueba de rangos de Wilcoxon en las medias relacionadas al indicador Tiempo para el proceso de cierre de ventas de productos académicos.**

		Rangos		
		N	Rango promedio	Suma de rangos
PostTest – PreTest	Rangos negativos	53 <sup>a</sup>	27,00	1431,00
	Rangos positivos	0 <sup>b</sup>	,00	,00
	Empates	0 <sup>c</sup>		
	Total	53		

a. PostTest < PreTest

b. PostTest > PreTest

c. PostTest = PreTest

**Fuente: Elaboración propia.**

En la tabla 13 se muestran los valores de la prueba de rangos, en donde 53 pares se encuentran en rangos negativos (reducción), 0 pares en el rango positivo (aumento), es decir que Machine Learning redujo el tiempo del proceso de cierre de ventas de productos académicos.

**Tabla 14 Prueba de Wilcoxon aplicado al indicador Tiempo para el proceso de cierre de ventas de productos académicos.**

Estadísticos de prueba <sup>a</sup>	
PostTest - PreTest	
Z	-6,404 <sup>b</sup>
Sig. asintótica(bilateral)	,000

a. Prueba de rangos con signo de Wilcoxon

b. Se basa en rangos positivos.

**Fuente: Elaboración propia.**



La tabla 14 nos muestra un valor de significancia de 0,000 que es menor a 0,050, entonces, se rechaza la hipótesis nula y se acepta la hipótesis de la investigación que indica que Machine Learning permitió disminuir el tiempo del proceso de cierre de ventas de productos académicos.

## V. DISCUSIÓN

Los resultados obtenidos en esta tesis de investigación reflejan que existe una mejora considerable en los indicadores de la variable dependiente una vez aplicado el experimento y recolectados los datos post test.

Se utilizó la prueba no paramétrica Kolmogorov-Smirnov para conocer la normalidad de los datos, esto resultó en que no existe normalidad de datos en los indicadores de esta investigación. Por ello, se usó la prueba Wilcoxon para comparar las medias, dando como efecto que el valor de significancia sea 0,000. Por lo tanto, se rechazaron las hipótesis nulas  $H_0$  y se aceptaron las hipótesis alternas  $H_1$  en los cuatro indicadores.

El autor Carlos García Mateo (2020), describe en su trabajo cómo realizar predicción de ventas de acciones de una entidad bancaria utilizando Rstudio y el algoritmo regresión lineal al igual que esta investigación. Estas herramientas permitieron predecir la tendencia, alzas y bajas de precio de las acciones dando resultados favorables a los objetivos de la investigación, esto demuestra que utilizar estas tecnologías es idóneo para predecir.

Wu (2022), en su investigación, nos indica que Machine Learning es una técnica que se utiliza con mucha frecuencia en campos de la ingeniería, salud, negocios, etc. En su auge, ha tenido éxito en el comercio y tecnología. Según el autor, Machine Learning se basa en el estudio de muestras de datos, y es una de las metodologías más precisas para realizar predicciones. Se utiliza como punto de referencia para la toma de decisiones en una compañía. En este trabajo se estudian varios algoritmos de aprendizaje automático, dando como resultado que la regresión lineal es uno de los más eficientes.

El autor, utiliza este algoritmo para analizar la relación entre los años de experiencia de los empleados y su salario, para conocer cuál es un estimado de incremento de sueldo con el pasar de los años y el porcentaje de inversión. Para nuestro estudio, se utilizó la regresión lineal porque según otros autores que realizan comparaciones de algoritmos concuerdan en que es preciso, por este motivo se lo aplicó para realizar la predicción de plataformas educativas.

En el primer indicador Tiempo para elegir plataformas de adquisición de clientes se evidenció una mejora en la reducción del 18,01% en el proceso para elección, en cambio en el estudio del autor Silva (2021), se utilizó Árboles de Decisión para predecir la elección y compra de libros teniendo un resultado óptimo en recomendaciones, e interés de parte de los usuarios, mejorando en un 16,00% sus ventas.

Los autores Dudhia et al. (2020), nos indican en su investigación que la elección de productos en portales web populares de la India depende en gran medida de las recomendaciones de otros compradores. Utilizó la técnica de filtrado colaborativo con Machine Learning y Deep Learning, tomando en cuenta factores como, reseñas, puntuación del producto, y el comportamiento de los compradores, teniendo como resultado un 30% de mejora en la aceptación de productos. En cambio, en nuestra investigación, se utilizó solamente Machine Learning para predecir la elección de productos y se obtuvo una mejora del 18,01%, este porcentaje es muy bueno considerando que se usó de una sola técnica.

Un estudio similar es el de Zhang & Qiu (2021), ellos utilizan un modelo de red neuronal para predecir qué producto será elegido por los consumidores. Recopilaron datos de reseñas en línea de los productos, comentarios en redes sociales y otras plataformas y según estas características, aplicaron el modelo predictivo que permitió analizar el impacto de sus ventas a futuro. Como resultado se redujo el tiempo en un 15% para saber cuáles productos son los preferidos, lo cual permitió conocer su popularidad y compararlos con las ventas de los competidores. Concluyeron que el modelo era sensato y práctico comparando lo resultante con datos reales de años anteriores.

En contraste con nuestro trabajo, se utilizó una técnica diferente para realizar el modelo predictivo, sin embargo, el porcentaje de mejora es similar demostrando que ambas técnicas brindan resultados favorables en cuanto a la adquisición de productos y estudio de sus características.

Singh et al. (2021) aplicaron modelos de Machine Learning para facilitar la toma de decisiones comerciales, tener procesos más claros y precisos. Utilizaron los algoritmos Random Forest y XGBoost Machine para predecir diferentes pedidos de

productos, con ello consiguieron una mejora del 20% en la métrica de ventas y tiempos de entregas, porque lograron anticipar los pedidos de los clientes. En nuestro caso sucedió algo parecido, nos anticipamos en conocer cuál plataforma el cliente podría elegir para agilizar el proceso de ventas.

En el segundo indicador Tiempo para la elección de clientes potenciales se obtuvo una mejora en la reducción de tiempo del 32,29% en comparación con el pretest. Los autores Asencio, Luz de los Ángeles et al. (2021), y Bernachea, Carla et al. (2021), utilizan un modelo de clasificación conveniente y con mejores resultados en comparación de otros algoritmos probados que fue el de regresión logística. Los primeros autores utilizan Machine Learning para clasificar automáticamente a los clientes con el algoritmo K-NN, y el segundo para calcular la probabilidad de deserción de clientes mediante programación en Python dando resultados favorables y cumpliendo sus objetivos. En cambio, en nuestra investigación la regresión lineal fue aplicada porque su uso va enfocado a predicción de ventas.

Centeno, José, Vera, Adriana, et al. (2020), en su trabajo utilizaron cuatro modelos de predicción, siendo xtreme gradient boosting el que mejores resultados brindó, con una exactitud del 70,77%. Con ello, pudieron segmentar clientes y aplicar métodos de venta para lograr detener su deserción y mejorar la elección de clientes, a diferencia de nuestra investigación que está enfocada en conocer clientes potenciales.

Los autores Rincon & Gonzalez (2021), en su trabajo utilizan machine learning para identificar clientes potenciales usando una metodología probabilística. La identificación de clientes mejoró en un 30% logrando detectar mayores oportunidades. En comparación, con nuestro indicador tuvimos una mejora del 32,29% debido a que se utilizó otro método de predicción.

Los autores Centeno, José et al. (2020), utilizan Machine Learning para estudiar el comportamiento y predecir cuáles clientes renovarán pólizas y cuáles no. Evaluaron cuatro algoritmos, pero fue el xtreme gradient boosting el que tuvo mejores resultados, brindando una precisión del 70,77%, con lo cual se pudo dividir clientes y aplicar estrategias focalizadas de venta para lograr mantener la liquidez. A

diferencia de nuestro indicador, en este trabajo se observa un resultado diferente porque utilizan otra metodología para predecir.

Con respecto a tercer indicador Índice de adquisiciones de plataformas educativas que se realizan en un trimestre, los autores Afanador Jiménez et al. (2022), realizan una investigación similar para predecir la demanda de productos en una panificadora, al igual que en nuestra investigación, utilizan la regresión lineal pero también redes neuronales para establecer la producción diaria. Esto resultó en que la utilidad aumentó en un 17,23% y hubo una reducción de costos de 6,36%. En cambio, aplicando solamente aprendizaje automático en nuestra investigación, se logró que haya un 16,51% de aumento en la adquisición de plataformas que se realizan en un periodo. Los autores de la investigación utilizaron dos algoritmos, por lo tanto, obtuvieron un porcentaje mayor de mejora.

El autor Bursztyn, (2018) indica que un modelo de pronóstico preciso puede ser un factor decisivo en la supervivencia de una empresa de venta directa. Siendo así, la importancia de la comparación de dos modelos estadísticos polares para sus predicciones, entre ellos Regresión lineal y Random Forest, teniendo como resultado diferencias como que, aunque Random Forest es mucho más computacional y complejo, no llega a ser una diferencia estadísticamente significativa a comparación de la regresión lineal. Esto se debe a que es un método que proporciona información sobre la descripción de variables que no se obtienen con el modelo Random Forest, siendo una clara ventaja del modelo lineal.

De tal manera, este código del modelo lineal ayudó a obtener precisión en el índice de adquisiciones de la empresa. En nuestro trabajo de investigación también se utilizó regresión lineal para describir variables y tener modelo con exactitud para realizar predicciones.

Según Najjar (2018), el aplicó técnicas de modelado estadístico y machine learning utilizando históricos de los últimos 3 años para el análisis del comportamiento y evolución de los patrones de consumo por producto, canal y ámbitos geográficos. El cual obtuvo resultados entre los modelos aplicados como el análisis de clustering, con el que determinó las agrupaciones de provincias con preferencias similares de consumo, pero el análisis de regresión lineal permitió la proyección de

ventas para los próximos años en cuanto a la mejora de sus servicios, siendo favorable para la toma de decisiones y aumento del índice de adquisiciones. Al igual que en nuestra tesis se basa en un modelo de regresión para tener una proyección de ventas a futuro.

La investigación de Choudhary et al. (2018), tuvo como objetivo predecir ventas en línea para aumentar ganancias, utilizaron el algoritmo de regresión lineal para analizar datos de dos años para realizar la predicción de las ventas del siguiente año, mejorando en un 14 %. Compararon datos reales con el resultado del pronóstico para conocer la precisión del algoritmo. Indicaron que los resultados ayudarán a la toma de decisiones para el aumento posterior de ventas. En relación con nuestro estudio, se utiliza el algoritmo para el mismo objetivo que es aumentar ganancias. Existe una diferencia en el resultado de ambos trabajos de 2,51%, siendo nuestro modelo el que tuvo mejores resultados.

En el cuarto indicador Tiempo para el proceso de cierre de ventas de productos académicos se obtuvo una reducción del 24,35% del tiempo. Se analizó una literatura del autor Castañeda Miguel (2020), que utiliza un modelado de regresión lineal y redes neuronales para mejorar la gestión de ventas aumentando su eficiencia de un 54,64% a un 82,04%. En comparación con nuestra investigación, podemos observar que los porcentajes varían porque el autor además de usar regresión también incluye en su estudio un análisis con redes neuronales, por ello la mejora en eficiencia.

El autor Joakin (2021), en su trabajo de investigación utiliza Machine Learning con redes neuronales para predecir las ventas y mejorar la toma de decisiones de los directivos de la empresa. Utilizó un modelo que predijo las actividades comerciales de los siguientes 7 días, luego usaron esa misma metodología para realizar la predicción de todo un año. Con esta metodología pudo realizar la comparación de ventas entre los años 2018 y 2020, esto dio como resultado que según la predicción las ventas aumentarán en un 33,33% con respecto a los años anteriores.

Los resultados entre las investigaciones pueden variar dependiendo el algoritmo utilizado y el contexto estudiado. En nuestro caso, se realizó el estudio en un

periodo corto de tiempo en comparación del autor de este trabajo que lo realizó durante un año y es por ello que se da la diferencia en los resultados.

Silva Marcus Vinícius et al., (2021), utilizaron el aprendizaje automático para garantizar las ventas de una compañía brasileña que está en el sector alimentario. Para el análisis se utilizó información recopilada de cuatro años que se encontraba en un ERP. Se dividieron los datos en grupos para realizar pruebas y predecir las ventas anuales y mensuales. Como resultado se evidenció una mejora en el tiempo para procesar los alimentos entre 4,2 y 4.8 segundos, con ello aumenta la producción y se reduce el tiempo de venta en un 25%. Si comparamos este resultado con el obtenido en esta investigación (24,35%), casi no existe diferencia en el porcentaje de reducción de tiempos de venta, aunque la predicción haya sido aplicada en otro sector productivo.

Los autores Indriyanti et al. (2021), realizaron una investigación utilizando el algoritmo de regresión lineal para pronosticar ventas de motocicletas tomando en cuenta grandes cantidades de datos, correspondientes a 5 años atrás. Su finalidad fue predecir la tendencia comercial de los próximos meses. Según el argumento de los autores, los resultados fueron muy exactos, se logró un porcentaje de 63,77% de precisión y el porcentaje de error fue mínimo en comparación a lo esperado. Con estos resultados, el tiempo para el cierre de ventas disminuyó logrando tener mayor rentabilidad. En nuestro estudio, se realizó la predicción con la misma técnica, pero con menor cantidad de datos, mientras que los autores utilizan mayor cantidad de datos por la trayectoria de la compañía en la que aplican el experimento, logrando un muy buen porcentaje de precisión.

El trabajo de Acosta-Velásquez et al. (2021), estuvo enfocado en la predicción de ventas netas postpandemia en empresas colombianas. Utilizaron datos históricos para el análisis con Machine Learning. Utilizaron alrededor de dos millones de datos para realizar pruebas predictivas. Evaluaron algoritmos como Random Forest, Extreme Learning Machine, XGBoost y Regresión lineal, pero el que mejor desempeño tuvo fue Random Forest. La diferencia en el desempeño puede estar ligada al gran tamaño de datos que se analizaron siendo el algoritmo más robusto Random Forest, sin embargo, como en nuestra investigación se analizó una cantidad menor de datos se utilizó regresión lineal.

## **VI. CONCLUSIONES**

1. Como principal conclusión se logró determinar el algoritmo adecuado para la implementación de Machine Learning, el cual permitió predecir sin inconvenientes la adquisición de plataformas educativas por parte de los clientes. Para el desarrollo del objetivo general de esta investigación se cumplieron los objetivos específicos, llegando a las siguientes conclusiones.
2. Se concluye que Machine Learning disminuyó el tiempo en un 18,01% con respecto a las plataformas elegidas en el pretest, lo cual ayudó a que se conozca por cuál plataforma el cliente estará interesado para agilizar el proceso de ventas.
3. En el segundo indicador la reducción de tiempo de elección de clientes tuvo un porcentaje más considerable, un 32,29% en comparación a las técnicas que se venía aplicando por parte del personal de la empresa. Esto hará que los clientes potenciales conozcan los productos y exista mayor probabilidad de un cierre de ventas.
4. El porcentaje de adquisición de plataformas educativas en un trimestre aumentó en un 17,23%, por lo tanto, se puede concluir que el algoritmo utilizado mejoró la venta de productos académicos y el cliente adquirió un producto que se ajuste a sus necesidades.
5. Aplicando la solución de Machine Learning, el proceso de cierre de ventas mejoró porque se redujo el tiempo en un 24,35%. Por ello podemos decir que la metodología aplicada optimizó el proceso de venta en todas sus fases.



## **VII. RECOMENDACIONES**

1. Se debe realizar un análisis previo de los datos a utilizar para lograr determinar el modelo idóneo y evitar que los datos tengan inconsistencias al momento de realizar las predicciones.
2. Como recomendación para los asesores de la empresa, deben tomar en cuenta las características del cliente, número de estudiantes, carreras de la institución y los resultados de la herramienta para ofrecer al cliente plataformas acorde a sus necesidades.
3. Es recomendable que el personal de la empresa tome en cuenta el análisis de la base de datos de clientes para encontrar características similares entre los nuevos prospectos. Con ello, habrá posibilidad de tener apertura porque estarán interesados por productos similares que se ofertan.
4. La adquisición de plataformas, si bien es cierto depende del presupuesto de la institución, pero también hay factores que ayudan a que adquieran productos, entre ellos, la calidad y características que mejoren la problemática del cliente. Es por ello, que se recomienda utilizar soluciones tecnológicas para facilitar la elección de productos idóneos para los clientes y lograr un aumento de ventas.
5. Se recomienda utilizar metodologías adecuadas en todas las fases de venta para lograr que el porcentaje de cierre aumente en un periodo determinado de tiempo.
6. Para los investigadores que se quieren adentrar al mundo del aprendizaje automático, se recomienda que apliquen varios algoritmos a su muestra de estudio para lograr una comparación y verificar el que tiene mayor exactitud en la predicción. Por motivos de tiempo esto no se realizó en esta investigación, pero se tomará en cuenta para un trabajo futuro.

## REFERENCIAS

- Acosta-Velásquez, R. D., Fajardo-Moreno, W. S., & Espinosa-Leal, L. (2021). Machine Learning in the Prediction of Net Sales for Colombian Companies in a Post-pandemic Scenario. *PASEW-21, CABES-21, MESSH-21, ICISSET-21, FBES-21, L3S2E-21 & BEMHSS-21 2021 European International Conferences*. <https://doi.org/10.17758/eirai11.f1221104>
- Afanador Jiménez, M., Casadiegos Chaparro, S. J., Campo Maichel, I., & Casallas Estrella, J. S. (2022). *Diseño de un modelo de pronóstico de demanda basado en machine learning y un modelo multi-objetivo para planeación de la producción en una industria panificadora*. <http://repository.javeriana.edu.co/handle/10554/59123>
- Alban, G. P. G., Arguello, A. E. V., & Molina, N. E. C. (2020). Metodologías de investigación educativa (descriptivas, experimentales, participativas, y de investigación-acción). *RECIMUNDO*, 4(3), 163-173. [https://doi.org/10.26820/recimundo/4.\(3\).julio.2020.163-173](https://doi.org/10.26820/recimundo/4.(3).julio.2020.163-173)
- Asencio, Luz de los Ángeles, Chiang, Ricardo, Crisóstomo, Fernanda, Hernández, Gisela, & Lajo, Almendra. (2021). *Técnicas de Machine Learning para la clasificación automática de clientes en una empresa de seguros*. <https://webcache.googleusercontent.com/search?q=cache:5G3ZFCvreqYJ:https://repositorio.esan.edu.pe/handle/20.500.12640/2933+&cd=1&hl=es&ct=clnk&gl=ec&client=firefox-b-d>
- Asto, C. M. N., & Malpartida, J. A. V. (2020). “*BÚSQUEDA DE PATRONES DE COMPORTAMIENTO USANDO MACHINE LEARNING, PARA LA TOMA DE DECISIONES GERENCIALES EN LA EMPRESA CHUCHUHUASI*”. 105.
- Barrera Rea, V. F., & Guapi Mullo, A. (2018). La importancia del uso de las plataformas virtuales en la educación superior. *Atlante Cuadernos de Educación y Desarrollo*,

julio. <https://www.eumed.net/rev/atlante/2018/07/plataformas-virtuales-educacion.html>

Bernachea, Carla, Chilet, Edward, Guzmán, Paola, Inche, Victor, & Leon, Johana. (2021).

*Técnica de Machine Learning para el cálculo de la probabilidad de fuga de los clientes de la empresa Bitel.*

<https://webcache.googleusercontent.com/search?q=cache:UE0yJBf7PzsJ:https://repositorio.esan.edu.pe/handle/20.500.12640/2929+&cd=1&hl=es&ct=clnk&gl=ec&client=firefox-b-d>

Boada, A. (2016). “Sistema Forecast”. Predicción automatizada en empresas de venta directa. *Opción*, 32(11), 121-142.

Brink, H., Richards, J. W., & Fetherolf, M. (2017). *Real-world machine learning*. Manning.

Bursztyn, V. (2018). *Direct Sales Forecasting: Statistical Methods vs. Artificial Intelligence*. 43.

Cabrera, I. F. P. (2016). *Desarrollo de un modelo basado en Machine Learning para la predicción de la demanda de*. 174.

*Cap.4-Investigación cuantitativa y cualitativa.pdf*. (s. f.). Recuperado 28 de abril de 2022, de <http://repositorio.utmachala.edu.ec/bitstream/48000/14232/1/Cap.4-Investigaci%C3%B3n%20cuantitativa%20y%20cualitativa.pdf>

Carlos García Mateo. (2020). *Estudio Estadístico de Valores Bursátiles Utilizando R: Una Aproximación Didáctica Opensource*.

<https://uvadoc.uva.es/bitstream/handle/10324/37286/TFG-B.1360.pdf?sequence=1&isAllowed=y>

CARRILLO, H. (2020). Pronóstico de ventas y tendencias de mercado para la organización cooperativa mercaorinoquia. Modalidad de grado. *UNIVERSIDAD COOPERATIVA DE COLOMBIA*.

Castañeda, Miguel. (2020). *Machine learning para la gestión de ventas en la Empresa*  
*Vértice* *Empresarial* *S.A.C.*

<https://webcache.googleusercontent.com/search?q=cache:->

[hur8Fl53kIJ:https://repositorio.ucv.edu.pe/handle/20.500.12692/56129+&cd=1&hl=es&ct=clnk&gl=ec&client=firefox-b-d](https://repositorio.ucv.edu.pe/handle/20.500.12692/56129+&cd=1&hl=es&ct=clnk&gl=ec&client=firefox-b-d)

*Castañeda\_RMAD-SD.pdf*. (s. f.).

Centeno, José, Vera, Adriana, & Campuzano, Juan. (2020). *DSPACE en ESPOL: Deserción de clientes en el sector asegurador. Evidencia mediante el análisis de modelos de Machine Learning.*

<https://webcache.googleusercontent.com/search?q=cache:iqCsG0R->

[dj8J:https://www.dspace.espol.edu.ec/handle/123456789/53316+&cd=1&hl=es&ct=clnk&gl=ec&client=firefox-b-d](https://www.dspace.espol.edu.ec/handle/123456789/53316+&cd=1&hl=es&ct=clnk&gl=ec&client=firefox-b-d)

Cheriyana, S., Ibrahim, S., Mohanan, S., & Treasa, S. (2018). Intelligent Sales Prediction Using Machine Learning Techniques. *2018 International Conference on Computing, Electronics Communications Engineering (iCCECE)*, 53-58.

<https://doi.org/10.1109/iCCECOME.2018.8659115>

Choudhary, R., Prasad, S., & Gopalakrishnan. (2018). Prediction of Sales Value in Online shopping using Linear Regression. *2018 4th International Conference on Computing Communication and Automation (ICCCA)*.

<https://doi.org/10.1109/CCAA.2018.8777620>

Davidson, B. E. K., Jason. (2019). Using Machine Learning to Predict Sales Conditional on Bid Acceptance. *Using Machine Learning to Predict Sales Conditional on Bid Acceptance*. <https://www.internationaljournalssrg.org/IJEMS/paper-details?Id=456>

Diez David, Barr Christopher, & Çetinkaya-Rundel Mine. (2022, noviembre 21). *Book: OpenIntro Statistics*. *Statistics* *LibreTexts*.

[https://stats.libretexts.org/Bookshelves/Introductory\\_Statistics/Book%3A\\_OpenIntro\\_Statistics\\_\(Diez\\_et\\_al\)](https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book%3A_OpenIntro_Statistics_(Diez_et_al)).

Dudhia, D. J., Dave, S. R., & Yagnik, S. (2020). Self Attentive Product Recommender – A Hybrid Approach with Machine Learning and Neural Network. *2020 International Conference for Emerging Technology (INCET)*, 1-4. <https://doi.org/10.1109/INCET49848.2020.9154034>

Fotache, M., Cojocariu, I.-C., & Berteau, A. (2021). High-Level Machine Learning Framework for Sports Events Ticket Sales Prediction. *CompSysTech*. <https://doi.org/10.1145/3472410.3472426>

Gonzalez, S. D., & Robayo, V. V. (2020). *PRONÓSTICO DE DEMANDA UTILIZANDO INTELIGENCIA ARTIFICIAL*. 52.

Hernández-Lalinde, J., Espinosa-Castro, J.-F., García Álvarez, D., & Bermúdez-Pirela, V. (2019). Sobre el uso adecuado de la regresión lineal: Conceptualización básica mediante un ejemplo aplicado a las ciencias de la salud. *Revista AVFT - Archivos Venezolanos de Farmacología y Terapéutica*. <https://bonga.unisimon.edu.co/handle/20.500.12442/5045>

Hugo Sánchez. (2020). *Manual de términos en investigación científica, tecnológica y humanística*. 79.

Huo, Z. (2021). Sales Prediction based on Machine Learning. *2021 2nd International Conference on E-Commerce and Internet Technology (ECIT)*. <https://doi.org/10.1109/ECIT52743.2021.00093>

Indriyanti, A. D., Prehanto, D., Prisma, I., & Nuryana, I. (2021). *The web-based estimation of motorcycles sales using linear regression method*. <https://doi.org/10.1088/1757-899X/1098/5/052115>

- Joakin, I. (2021). *Aplicación de tecnologías de aprendizaje automático para predecir negocios y tomar decisiones empresariales*. 125.
- Jon, Nicholas. (2017). (PDF) *Modern Data Science with R - Multiple Regression*.  
[https://www.researchgate.net/publication/321758086\\_Modern\\_Data\\_Science\\_with\\_R\\_-\\_Multiple\\_Regression\\_mdsr-bookgithubio](https://www.researchgate.net/publication/321758086_Modern_Data_Science_with_R_-_Multiple_Regression_mdsr-bookgithubio)
- Li, J., & Luo, T. (2019). *Machine Learning Algorithm Generated Sales Prediction for Inventory Optimization in Cross-border E-Commerce*.  
<https://doi.org/10.25236/IJFET.2019.010107>
- Lilja, D. J., & Linse, G. M. (2022). *Linear Regression Using R: An Introduction to Data Modeling, 2nd Edition*. University of Minnesota Libraries Publishing.  
<https://doi.org/10.24926/8668/1301>
- Meiyazhagan, K., & Rajkumar, S. (2021). Business Automation And Sales Prediction In Shopping Malls With Machine Learning Approach. *Undefined*.  
<https://www.semanticscholar.org/paper/Business-Automation-And-Sales-Prediction-In-Malls-Meiyazhagan-Rajkumar/3d09fbe76002be39381fd5decb2cbe54b8844536>
- Najar, J. D. V. (2018). *Análisis de Comportamiento de Consumos de Clientes*. 29.
- Narayana, C. V., Likhitha, C. L., Bademiya, S., & Kusumanjali, K. (2021). Machine Learning Techniques To Predict The Price Of Used Cars: Predictive Analytics in Retail Business. *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*.  
<https://doi.org/10.1109/ICESC51422.2021.9532845>
- Navarro Danielle. (2019, enero 11). *Book: Learning Statistics with R - A tutorial for Psychology Students and other Beginners (Navarro)*. Statistics LibreTexts.  
[https://stats.libretexts.org/Bookshelves/Applied\\_Statistics/Book%3A\\_Learning\\_Sta](https://stats.libretexts.org/Bookshelves/Applied_Statistics/Book%3A_Learning_Sta)

tistics\_with\_R\_-

\_A\_tutorial\_for\_Psychology\_Students\_and\_other\_Beginners\_(Navarro)

Odegua, R. (2020). *Applied Machine Learning for Supermarket Sales Prediction*.

Olive David J. (2017). *Linear Regression*. <https://link.springer.com/book/10.1007/978-3-319-55252-1>

Ospina, J. E. C. (2021). *Desarrollo de solución analítica para la predicción de la demanda de línea*. 25.

Poldrack, Russell A. (2021). *Book: Statistical Thinking for the 21st Century*. [https://stats.libretexts.org/Bookshelves/Introductory\\_Statistics/Book%3A\\_Statistical\\_Thinking\\_for\\_the\\_21st\\_Century\\_\(Poldrack\)](https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book%3A_Statistical_Thinking_for_the_21st_Century_(Poldrack))

Poojary, T., Ramchandani, D., Pawaskar, I., & Lemos, N. (2021). *Intelligent Sales Prediction using Machine Learning*. <https://www.semanticscholar.org/paper/Intelligent-Sales-Prediction-using-Machine-Learning-Poojary-Ramchandani/36ab51197995dabe2329ef4baebb3da9e1659205>

Raizada, S., & Saini, J. R. (2021). Comparative Analysis of Supervised Machine Learning Techniques for Sales Forecasting. *International Journal of Advanced Computer Science and Applications*. <https://doi.org/10.14569/ijacsa.2021.0121112>

Ramachandra, H., Balaraju, G., Rajashekar, A., & Patil, H. (2021). Machine Learning Application for Black Friday Sales Prediction Framework. *Undefined*. <https://www.semanticscholar.org/paper/Machine-Learning-Application-for-Black-Friday-Sales-Ramachandra-Balaraju/eefa0620ccad0c2e14f6c5c3a746d4ef2bdb2298>

- Rincon, N. F., & Gonzalez, D. (2021). *Ejecutar un modelo de machine learning para identificar los clientes potenciales basados en un proceso probabilístico para la empresa dell technologies*. <http://repository.unimilitar.edu.co/handle/10654/38562>
- Roback Paul Legler, L. J. (2021). *Beyond Multiple Linear Regression*. <https://bookdown.org/roback/bookdown-BeyondMLR/>
- Sakib, S. M. N. (2021). *Restaurant Sales Prediction Using Machine Learning*. <https://doi.org/10.31224/osf.io/wa927>
- Sánchez Flores, F. A. (2019). Fundamentos Epistémicos de la Investigación Cualitativa y Cuantitativa: Consensos y Disensos. *Revista Digital de Investigación en Docencia Universitaria*, 101-122. <https://doi.org/10.19083/ridu.2019.644>
- 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*. 33.
- Shipunov Alexey. (2019, enero 5). *Book: Visual Statistics Use R! (Shipunov)*. Statistics LibreTexts. [https://stats.libretexts.org/Bookshelves/Introductory\\_Statistics/Book%3A\\_Visual\\_Statistics\\_Use\\_R\\_\(Shipunov\)](https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book%3A_Visual_Statistics_Use_R_(Shipunov))
- Silva Marcus Vinícius, Melk de Carvalho Luiz, & Batista de Souza Flávio Henrique. (2021, marzo 16). *Strategic Management of Sales Assisted by Machine Learning: A Case Study in a Large Food Business - Mokslinès Leidybos Deimantas - Diamond Scientific Publishing*. <https://www.dpublication.com/abstract-of-11th-restconf/12-112/>
- Silva Morante Christian Raymond. (2021). *PROTOTIPO DE APLICACIÓN WEB MEDIANTE MACHINE LEARNING PARA PROPORCIONAR LA PREDICCIÓN DE ARTÍCULOS DE LIBROS EN PREFERENCIA AL USUARIO*.



- Singh, A., Tharanum, L., Qureshi, M. M., & Rachana. (2021). *Risk Management with Backorder in Supply Chain using Machine Learning Techniques*.  
<https://www.semanticscholar.org/paper/Risk-Management-with-Backorder-in-Supply-Chain-Singh-Tharanum/4a949aab80654e16f555966b7d1e447da2ceeff1>
- Utreras, C. A. P. (2020). *Análisis y predicción de las tendencias de venta en el mercado usando árboles de regresión*. 58.
- Valdez Alvarado, A. (2018, mayo 29). *INTRODUCCIÓN AL MACHINE LEARNING*.  
<https://doi.org/10.13140/RG.2.2.28886.19527>
- Wisesa, O., Andriansyah, A., & Khalaf, O. I. (2020). *Prediction Analysis for Business To Business (B2B) Sales of Telecommunication Services using Machine Learning Techniques*. <https://doi.org/10.29252/MJEE.14.4.145>
- Wu, Y. (2022). Linear regression in machine learning. *Other Conferences*.  
<https://doi.org/10.1117/12.2628053>
- Yolcu, U., Egrioglu, E., & Aladag, C. H. (2013). A new linear & nonlinear artificial neural network model for time series forecasting. *Decision Support Systems*, 54(3), 1340-1347. <https://doi.org/10.1016/j.dss.2012.12.006>
- Zhang, G., & Qiu, H. (2021). Competitive Product Identification and Sales Forecast Based on Consumer Reviews. *Mathematical Problems in Engineering*.  
<https://doi.org/10.1155/2021/2370692>

## ANEXOS

### Anexo 1: Instrumento de Recolección de datos

#### 1. Guía de observación N°1: Tiempo para elegir plataformas de adquisición para clientes

*Tabla 15 Guía de observación N°1.*

Guía de observación de medición del indicador Tiempo para elegir plataformas de adquisición de clientes				
Investigador:		Dayana Guzmán Vélez		
Proceso observado:		Adquisición de plataformas		
Pre-Test				
N° de Obs.	Fecha	Tiempo de inicio para elección de plataformas educativas	Tiempo final para elección de plataformas educativas	Tiempo para elegir plataformas = (Tiempo final para elección de plataformas educativas - Tiempo de inicio para elección de plataformas educativas)
1				
2				
3				
4				
5				
6				
N				

*Fuente: Elaboración propia.*

#### 2. Guía de observación N°2: Tiempo para la elección de clientes potenciales para la empresa.

*Tabla 16 Guía de observación N°2*

Guía de observación de medición del indicador Tiempo para la elección de clientes potenciales para la empresa.				
Investigador:		Dayana Guzmán Vélez		
Proceso observado:		Elección de clientes potenciales		
Pre-Test				
N° de Obs.	Fecha	Tiempo de inicio para elección de clientes potenciales	Tiempo final para elección de clientes potenciales	Tiempo para la elección de clientes potenciales = (Tiempo final elección de clientes potenciales - Tiempo de inicio para elección de clientes potenciales)
1				
2				
3				
4				
5				
6				
N				

*Fuente: Elaboración propia.*

### 3. Guía de observación N°3: Índice de adquisiciones de plataformas educativas que se realizan en un trimestre.

Tabla 17 Guía de observación N°3

Guía de observación de medición del indicador índice de adquisiciones de plataformas educativas que se realizan en un trimestre				
Investigador:		Dayana Guzmán Vélez		
Proceso observado:		Adquisición de plataformas educativas en un trimestre		
Pre-Test				
N° de Obs.	Fecha	Plataformas educativas adquiridas	Plataformas educativas pendientes de adquirirse en un trimestre	Índice de adquisiciones de plataformas educativas = (Plataformas educativas adquiridas / Plataformas educativas pendientes de adquirirse en un trimestre)*100
1				
2				
3				
4				
5				
6				
N				

Fuente: Elaboración propia.

### 4. Guía de observación N°4: Tiempo para el proceso de cierre de ventas de productos académicos.

Tabla 18 Guía de observación N°4

Guía de observación de medición del indicador Tiempo para el proceso de cierre de ventas de productos académicos				
Investigador:		Dayana Guzmán Vélez		
Proceso observado:		Proceso de cierre de ventas de productos académicos		
Pre-Test				
N° de Obs.	Fecha	Tiempo de inicio para el proceso de cierre de venta	Tiempo final para el proceso de cierre de venta	Tiempo para el proceso de cierre de ventas de productos académicos = (Tiempo final para el proceso de cierre de venta - Tiempo de inicio para el proceso de cierre de venta)
1				
2				
3				
4				
5				
6				
N				

Fuente: Elaboración propia.

## TABLA DE OPERACIONALIZACIÓN DE VARIABLES

Tabla 19 TABLA DE OPERACIONALIZACIÓN DE VARIABLES.

VARIABLES DE ESTUDIO	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIÓN	INDICADORES	ESCALA DE MEDICIÓN
<b>INDEPENDIENTE:</b> Machine Learning	Es una metodología científica que se basa en utilizar computadores y aparatos inteligentes para que estudien y aprendan a sacar por sí mismos las muestras y semejanzas que existen en los datos. Estas muestras o patrones sirven para tomar decisiones y pronosticar conductas (Valdez Alvarado, 2018).	Es el que se usa para comprender métodos y obtener un modelamiento de conocimiento mediante el aprendizaje. Machine Learning está inmerso en muchas áreas, las que mayormente destacan son las siguientes, transporte autónomo, seguridad informática, genética y su finalidad es reconocer o pronosticar patrones complejos fundados en data.	- Algoritmos - Técnicas - Grandes cantidades de información - Entrenamiento de algoritmo		De razón.
<b>DEPENDIENTE:</b> Predecir adquisición/ventas	La predicción de ventas básicamente está basada en la comprensión de ventas de un servicio o producto en cada una de las etapas de su proceso y que tienden a asimilar que los nuevos productos van a tener un comportamiento como los que se usan como guías. Es un	Se define como un estimado de una exigencia futura de productos. Una predicción es desarrollada mediante técnicas cualitativas y cuantitativas, también puede ser por una combinación de las mismas, y principalmente se enfocan en componentes intrínsecos/internos o extrínsecos/externos. Por lo general los métodos de	- Clientes - Presupuesto - Tipo de plataforma - Necesidad - Aprobación de autoridades	1. Tiempo para elegir plataformas de adquisición de clientes. 2. Tiempo para la elección de clientes potenciales para la empresa. 3. Índice de adquisiciones de plataformas educativas que se	De razón.

	<p>procedimiento de costo no tan elevado y que tiene como desventaja entender que las variables de procedimientos pasados se conservan y que los efectos que tienen los nuevos productos son dotados por los comportamientos de los objetos que se toman para estudio (Moreno,2019).</p>	<p>predicción pretenden pronosticar alguna de las unidades de la demanda: estacionalidad, aleatoriedad, ciclo y tendencia.</p>		<p>realizan en un trimestre. 4. Tiempo para el proceso de cierre de ventas de productos académicos.</p>	
--	--	--	--	---	--

**Fuente: Elaboración propia.**

## MATRIZ DE CONSISTENCIA

Tabla 20 MATRIZ DE CONSISTENCIA

MATRIZ DE CONSISTENCIA					
TÍTULO DE LA TESIS: Machine Learning para predecir la adquisición de plataformas educativas de la empresa Difucien Ecuador, 2022					
FORMULACIÓN DEL PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES Y DIMENSIONES	METODOLOGÍA	JUSTIFICACIÓN
<p><b>GENERAL:</b> ¿De qué manera influirá Machine Learning en la adquisición de plataformas educativas de la empresa Difucien, Ecuador 2022?</p>	<p><b>GENERAL:</b> Predecir la adquisición de plataformas educativas de la empresa Difucien Ecuador a través de la implementación de un sistema de Machine Learning</p>	<p><b>GENERAL:</b> La implementación de Machine Learning permitirá predecir la adquisición de plataformas educativas de la empresa Difucien Ecuador</p>	<p><b>INDEPENDIENTE:</b> Machine Learning</p> <p><b>Dimensiones:</b></p> <ul style="list-style-type: none"> <li>- Algoritmos</li> <li>- Técnicas</li> <li>- Grandes cantidades de información</li> <li>- Entrenamiento de algoritmo</li> </ul>	<p><b>Tipo y Diseño de Investigación:</b></p> <p><b>Tipo de investigación:</b> Experimental</p> <p><b>Diseño de Investigación:</b> Preexperimental</p> <p><b>Población:</b> La población total de clientes que adquieren las plataformas es de 80</p> <p><b>Muestra:</b> Tomando en cuenta un nivel de confianza del 95% y un margen de error del 5% la muestra es: 67 clientes</p> <p><b>Muestreo:</b></p>	<p><b>Operativa:</b> Se justifica porque es importante que las empresas tengan estrategias de captación de clientes, en donde el objetivo sea mejorar o incrementar las ventas, ofreciendo productos que convengan a las necesidades de los clientes y brindando una experiencia de compra eficaz y efectiva.</p> <p><b>Tecnológica:</b> Implementando un modelo predictivo basado en Machine Learning podremos aprovechar al máximo esta tecnología para lograr llegar a las</p>

				El muestreo seleccionado es probabilístico Aleatorio Simple porque cada elemento se selecciona de manera independiente y toma de forma equitativa la selección de las muestras a partir de una población	metas de venta antes del tiempo previsto.  <b>Económica:</b> Se espera con esta implementación mejorar los ingresos brutos de la empresa porque al predecir las adquisiciones mejorarán las ventas a corto plazo.
<p><b>ESPECÍFICOS:</b></p> <p>– ¿De qué manera influirá Machine Learning en el tiempo promedio para elegir plataformas educativas?</p> <p>– ¿De qué manera influirá Machine Learning en el número promedio de adquisiciones de plataformas educativas en un trimestre?</p> <p>– ¿De qué manera influirá Machine Learning en el tiempo promedio para el proceso de</p>	<p><b>ESPECÍFICOS:</b></p> <p>– Disminuir el tiempo promedio para elegir plataformas educativas que van a adquirir los clientes mediante la predicción con Machine Learning.</p> <p>– Aumentar el número promedio de adquisiciones de plataformas educativas que se realizan en un trimestre a través de Machine Learning.</p> <p>– Disminuir el tiempo promedio para el proceso de cierre de ventas de productos</p>	<p><b>ESPECÍFICAS:</b></p> <p>– La predicción con Machine Learning ayudará a disminuir el tiempo promedio para elegir plataformas educativas que van a adquirir los clientes</p> <p>– Machine Learning permitirá aumentar el número promedio de adquisiciones de plataformas educativas en un trimestre</p> <p>– La implementación de Machine Learning permitirá disminuir el tiempo</p>	<p><b>DEPENDIENTE:</b></p> <p>Adquisición de plataformas educativas</p> <p><b>Dimensiones:</b></p> <ul style="list-style-type: none"> <li>- Clientes</li> <li>- Presupuesto</li> <li>- Tipo de plataforma</li> <li>- Necesidad</li> <li>- Aprobación de autoridades</li> </ul>	<p><b>Técnicas e Instrumentos de recolección de Datos:</b></p> <p>Encuestas de Fichas de observación</p>	<p><b>Social:</b></p> <p>La implementación de un sistema de Machine Learning para predecir la adquisición de plataformas educativas, no solo beneficiará a la empresa, sino que ayudará a los clientes a decidir por la compra de recursos de suma importancia para los estudiantes de instituciones de educación superior.</p>

<p>cierre de ventas de productos académicos?          – ¿De qué manera influirá Machine Learning en el tiempo para la elección de clientes potenciales para la empresa?</p>	<p>académicos mediante la implementación de Machine Learning.          – Disminuir el tiempo promedio para la elección de clientes potenciales para la empresa</p>	<p>del proceso de cierre de ventas de productos académicos.          – Machine Learning permitirá disminuir el tiempo para elección de clientes potenciales.</p>			
---	--	--	--	--	--

***Fuente: Elaboración propia.***



## TABLA DE INDICADORES DE LA INVESTIGACIÓN

Tabla 21 Tabla de indicadores de la investigación. Elaboración propia.

Indicador	Instrumento	Cantidad	Fórmula
Tiempo para elegir plataformas de adquisición de clientes	Ficha de observación	53	Tiempo para elegir plataformas = (Tiempo final para elección de plataformas educativas - Tiempo de inicio para elección de plataformas educativas)
Tiempo para la elección de clientes potenciales para la empresa.	Ficha de observación	53	Tiempo para la elección de clientes potenciales = (Tiempo final elección de clientes potenciales - Tiempo de inicio para elección de clientes potenciales)
Índice de adquisiciones de plataformas educativas que se realizan en un trimestre.	Ficha de observación	53	Índice de adquisiciones de plataformas educativas = (Plataformas educativas adquiridas / Plataformas educativas pendientes de adquirirse en un trimestre)*100
Tiempo para el proceso de cierre de ventas de productos académicos.	Ficha de observación	53	Tiempo para la elección de clientes potenciales = (Tiempo final para el proceso de cierre de venta - Tiempo de inicio para el proceso de cierre de venta)

*Fuente: Elaboración propia.*

## Solicitud de aprobación para el desarrollo de la tesis



### **“AÑO DEL FORTALECIMIENTO DE LA SOBERANÍA NACIONAL”**

Trujillo, 02 de junio de 2022

#### **CARTA N° 092-2022-UCV-VA-EPG-F01/J**

Sra. Nelly Victoria Montesdeoca León

**Representante Legal**

**DIFUCIEN ECUADOR CIA LTDA**

**Presente.** -

#### **ASUNTO: AUTORIZACIÓN PARA EL DESARROLLO DE TESIS**

Es grato dirigirme a usted para saludarle cordialmente y así mismo presentar a la estudiante **DAYANA MARIUXI GUZMÁN VÉLEZ**, del programa de **MAESTRÍA EN INGENIERÍA DE SISTEMAS CON MENCIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN**, de la Escuela de Posgrado de la Universidad César Vallejo.

La estudiante en mención solicita autorización para aplicar los instrumentos necesarios para el desarrollo de su tesis denominada: **“MACHINE LEARNING PARA PREDECIR LA ADQUISICIÓN DE PLATAFORMAS EDUCATIVAS DE LA EMPRESA DIFUCIEN ECUADOR, 2022”**, en la institución que Ud. Dirige.

El objetivo principal de este trabajo de investigación es predecir la adquisición de plataformas educativas de la empresa Difucien Ecuador, mediante la implementación de un sistema de Machine Learning.

Agradeciendo la atención que brinde a la presente, aprovecho la oportunidad para expresarle mi consideración y respeto.

**Atentamente.** -



Mg. Ricardo Benites Aliaga  
Jefe de la Escuela de Posgrado-Trujillo  
Universidad César Vallejo

#### **ADJUNTO:**

- Instrumentos de recolección de datos.

## Carta de aprobación para el desarrollo de la tesis



03/06/2022

**Ing. Dayana Guzmán Vélez**  
Presente

En atención de su solicitud de autorización para el desarrollo de su Tesis denominada **"Machine Learning para predecir la adquisición de plataformas educativas de la empresa Difucien Ecuador, 2022"** y de la aplicación de instrumentos para recolectar datos. Como representante legal de la empresa Difucien Ecuador CIA LTDA, se aprueba su requerimiento por el tiempo que estime necesario.

Atentamente.



Escaneado y autenticado por:  
**NELLY VICTORIA  
MONTEDEOCA LEÓN**

**Nelly Montedeoca León**  
Representante Legal de  
Difucien Ecuador CIA LTDA

## CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO - EXPERTO 1

### CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO

VARIABLE: ADQUISICIÓN DE PLATAFORMAS EDUCATIVAS DE LA EMPRESA DIFUCIEN ECUADOR, 2022

N°	INDICADORES	Claridad <sup>1</sup>		Pertinencia <sup>2</sup>		Relevancia <sup>3</sup>		Sugerencias
		Si	No	Si	No	Si	No	
1	INDICADOR 1: TIEMPO PARA ELEGIR PLATAFORMAS PARA ADQUISICIÓN DE CLIENTES Fórmula: Tiempo final – Tiempo de inicio	X		X		X		
2	INDICADOR 2: TIEMPO PARA LA ELECCIÓN DE CLIENTES POTENCIALES PARA LA EMPRESA Fórmula: Tiempo final – Tiempo de inicio	X		X		X		
3	INDICADOR 3: INDICE DE ADQUISICIONES DE PLATAFORMAS EDUCATIVAS QUE SE REALIZAN EN UN TRIMESTRE Fórmula: (Plataformas educativas adquiridas / Plataformas educativas pendientes de adquirirse en un trimestre) * 100	X		X		X		
4	INDICADOR 4: TIEMPO PARA EL PROCESO DE CIERRE DE VENTAS DE PRODUCTOS ACADÉMICOS Fórmula: Tiempo final – Tiempo de inicio	X		X		X		

Observaciones (precisar si hay suficiencia): SUFICIENTE

Opinión de aplicabilidad:   Aplicable [ X ]           Aplicable después de corregir [ ]           No aplicable [ ]

Apellidos y nombres del juez evaluador: Ing. Msc. JOSÉ LUIS TUBAY VERGARA           C.C: 1205414509

07/07/2022

Especialista: Metodólogo [ ]           Temático [ X ]

Grado: Maestro [ X ]           Doctor [ ]

<sup>1</sup>Claridad: Se entiende con dificultad algún enunciado del ítem, es conciso, exacto y directo.

<sup>2</sup>Pertinencia: Si el ítem pertenece a la dimensión.

<sup>3</sup>Relevancia: El ítem es apropiado para representar al componente o dimensión específica del constructo.

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión

JOSE LUIS  
TUBAY  
VERGARA

Firmado digitalmente  
por JOSE LUIS TUBAY  
VERGARA  
Fecha: 2022.07.10  
08:36:12 -05'00'

Firma del experto Informante

## CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO - EXPERTO 2

### CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO

VARIABLE: ADQUISICIÓN DE PLATAFORMAS EDUCATIVAS DE LA EMPRESA DIFUCIEN ECUADOR, 2022

N°	INDICADORES	Claridad <sup>1</sup>		Pertinencia <sup>2</sup>		Relevancia <sup>3</sup>		Sugerencias
		Si	No	Si	No	Si	No	
1	INDICADOR 1: TIEMPO PARA ELEGIR PLATAFORMAS PARA ADQUISICIÓN DE CLIENTES Fórmula: Tiempo final – Tiempo de inicio	X		X		X		
2	INDICADOR 2: TIEMPO PARA LA ELECCIÓN DE CLIENTES POTENCIALES PARA LA EMPRESA Fórmula: Tiempo final – Tiempo de inicio	X		X		X		
3	INDICADOR 3: INDICE DE ADQUISICIONES DE PLATAFORMAS EDUCATIVAS QUE SE REALIZAN EN UN TRIMESTRE Fórmula: (Plataformas educativas adquiridas / Plataformas educativas pendientes de adquirirse en un trimestre) * 100	X		X		X		
4	INDICADOR 4: TIEMPO PARA EL PROCESO DE CIERRE DE VENTAS DE PRODUCTOS ACADÉMICOS Fórmula: Tiempo final – Tiempo de inicio	X		X		X		

Observaciones (precisar si hay suficiencia): SUFICIENTE

Opinión de aplicabilidad:   Aplicable [ X ]           Aplicable después de corregir [ ]           No aplicable [ ]

Apellidos y nombres del juez evaluador: PhD. OVIEDO BAYAS BYRON WLADIMIR           C.C: 0914200373

11/07/2022

Especialista: Metodólogo [X]           Temático [ ]

Grado: Maestro [ ]           Doctor [ X ]

<sup>1</sup>Claridad: Se entiende con dificultad algún enunciado del ítem, es conciso, exacto y directo.

<sup>2</sup>Pertinencia: Si el ítem pertenece a la dimensión.

<sup>3</sup>Relevancia: El ítem es apropiado para representar al componente o dimensión específica del constructo.

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión



Firma del experto Informante

## CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO - EXPERTO 3

### CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO

VARIABLE: ADQUISICIÓN DE PLATAFORMAS EDUCATIVAS DE LA EMPRESA DIFUCIEN ECUADOR, 2022

N°	INDICADORES	Claridad <sup>1</sup>		Pertinencia <sup>2</sup>		Relevancia <sup>3</sup>		Sugerencias
		Si	No	Si	No	Si	No	
1	INDICADOR 1: TIEMPO PARA ELEGIR PLATAFORMAS PARA ADQUISICIÓN DE CLIENTES Fórmula: Tiempo final – Tiempo de inicio	X		X		X		
2	INDICADOR 2: TIEMPO PARA LA ELECCIÓN DE CLIENTES POTENCIALES PARA LA EMPRESA Fórmula: Tiempo final – Tiempo de inicio	X		X		X		
3	INDICADOR 3: INDICE DE ADQUISICIONES DE PLATAFORMAS EDUCATIVAS QUE SE REALIZAN EN UN TRIMESTRE Fórmula: (Plataformas educativas adquiridas / Plataformas educativas pendientes de adquirirse en un trimestre) * 100	X		X		X		
4	INDICADOR 4: TIEMPO PARA EL PROCESO DE CIERRE DE VENTAS DE PRODUCTOS ACADÉMICOS Fórmula: Tiempo final – Tiempo de inicio	X		X		X		

Observaciones (precisar si hay suficiencia): **SUFICIENTE**

Opinión de aplicabilidad:   Aplicable [ X ]           Aplicable después de corregir [ ]           No aplicable [ ]

Apellidos y nombres del juez evaluador: Ing. Msc. ZHUMA MERA EMILIO RODRIGO           C.C.: 1715393193

14/07/2022

Especialista: Metodólogo [ ]           Temático [ X ]

Grado: Maestro [ X ]           Doctor [ ]

EMILIO  
RODRIGO  
ZHUMA  
MERA

Firmado digitalmente por  
EMILIO RODRIGO  
ZHUMA MERA  
Fecha: 2022.07.15  
21:55:57 -05'00'

<sup>1</sup>Claridad: Se entiende con dificultad algún enunciado del ítem, es conciso, exacto y directo.

<sup>2</sup>Pertinencia: Si el ítem pertenece a la dimensión.

<sup>3</sup>Relevancia: El ítem es apropiado para representar al componente o dimensión específica del constructo.

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión

Firma del experto Informante

## Carta de aceptación de los resultados de la investigación



21/07/2022

**Ing. Dayana Guzmán Vélez**  
Presente:

Mediante la presente, se comunica que la empresa Difucien Ecuador CIA. LTDA, está de acuerdo con los resultados del trabajo de investigación "Machine Learning para predecir la adquisición de plataformas educativas de la empresa Difucien Ecuador, 2022".

Reiteramos el apoyo brindado a usted, para el proceso de desarrollo de esta tesis.

Atentamente,



Escanea el código QR para  
NELY VICTORIA  
MONTESDEOCA LEÓN

**Nelly Montesdeoca León**  
Representante Legal de  
Difucien Ecuador CIA. LTDA

Empresa: DIFUCIEN ECUADOR CIA. LTDA.  
Dirección: Av. de los Shyris N34-40 y República del Salvador / Edificio Tapia / Piso 9 / Oficina 901  
Teléfonos: 02 2430 993 / 02 2241 182 / 02 2439 189 / 02 2433 016 / Ext: 204 RUC: 1792483262001

## Desarrollo del algoritmo de Machine Learning

Se utilizó el software RStudio para desarrollar el modelo de regresión lineal para Machine Learning. Cargamos la base de datos que será objeto de estudio y realizamos un resumen de los datos (Cantidad mínima, Cantidad máxima, mediana).

**Figura 5 Resumen de los datos de la base de datos.**

```
base_de_datos <- read.csv("D:\\Dayana\\METODOLOGÍA DE INVESTIGACIÓN\\Presentar\\CSVPrediccionenR.csv")
resumen <- summary(base_de_datos)
resumen

> resumen <- summary(base_de_datos)
> resumen
Cantidad_de_estudiantes Carreras Modalidad_virtual Presupuesto_para_adquisición_de_plataformas
Min. : 226 Min. : 2.00 Min. :0.0000 Min. : 20000
1st Qu.: 2750 1st Qu.:11.75 1st Qu.:0.0000 1st Qu.: 80000
Median : 7000 Median :22.50 Median :0.0000 Median :100000
Mean :11188 Mean :24.02 Mean :0.2667 Mean :129833
3rd Qu.:13000 3rd Qu.:31.25 3rd Qu.:1.0000 3rd Qu.:162500
Max. :52000 Max. :63.00 Max. :1.0000 Max. :500000
Compró_enalgúnmomento_plataformas_de_Difucien Plataformas_presentadas
Min. :0.00 Min. : 3.000
1st Qu.:1.00 1st Qu.: 3.000
Median :1.00 Median : 4.000
Mean :0.95 Mean : 6.317
3rd Qu.:1.00 3rd Qu.: 9.000
Max. :1.00 Max. :11.000
Cuántas_reuniones_de_presentación_de_productos_tuvieron_en_este_año Cuánto_tiempo_duraron_las_reuniones
Min. :1.00 Min. :0.000
1st Qu.:1.75 1st Qu.:1.000
Median :2.00 Median :1.000
Mean :2.05 Mean :1.483
3rd Qu.:3.00 3rd Qu.:2.000
Max. :4.00 Max. :3.000
Elegióalguna_plataforma.SI.NO Plataformas_elegidas
Min. :0.00 Min. :0.000
1st Qu.:1.00 1st Qu.:1.000
Median :1.00 Median :1.500
Mean :0.95 Mean :1.817
3rd Qu.:1.00 3rd Qu.:2.250
Max. :1.00 Max. :5.000
```

**Fuente: Elaboración propia.**

Se realizó una gráfica tomando en cuenta la influencia del Presupuesto en la adquisición de plataformas educativas, para ver la tendencia de compra.

**Figura 6 Gráfica de Presupuesto en la adquisición de plataformas educativas.**

```
plot(base_de_datos$Plataformas_elegidas,base_de_datos$Presupuesto_para_adquisición_de_plataformas,main="Influencia del presupuesto en la adquisición de plataformas educativas")
```

**Fuente: Elaboración propia.**

Se realizó el algoritmo de regresión lineal para predecir la adquisición de plataformas educativas según el presupuesto. Mostramos un resumen que nos indica que el modelo tiene una precisión del 32%.



**Figura 7 Resultado de modelo 1 de regresión**

```
summary(regresion)

> regresion <- lm(Plataformas.elegidas ~ Presupuesto_para_adquisición_de_plataformas, data=base_de_datos)
> summary(regresion)

Call:
lm(formula = Plataformas.elegidas ~ Presupuesto_para_adquisición_de_plataformas,
    data = base_de_datos)

Residuals:
    Min       1Q   Median       3Q      Max
-2.1726 -0.5789 -0.1805  0.5914  2.6242

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      7.821e-01  2.304e-01   3.395  0.00125 **
Presupuesto_para_adquisición_de_plataformas  7.968e-06  1.505e-06   5.295  1.91e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9459 on 58 degrees of freedom
Multiple R-squared:  0.3258,    Adjusted R-squared:  0.3142
F-statistic: 28.03 on 1 and 58 DF,  p-value: 1.914e-06
```

**Fuente: Elaboración propia.**

Colocamos los valores que queremos que el algoritmo tome en cuenta para realizar la predicción de plataformas. Se realizaron múltiples pruebas, en este caso con valores del Presupuesto para adquisición, la cantidad de estudiantes y carreras, dando como resultado la predicción de la cantidad de plataformas que podría adquirir el cliente.

**Figura 8 Resultado de la predicción del modelo 1.**

```
#Modelo de regresión
data_nueva <- data.frame (Presupuesto_para_adquisición_de_plataformas=90000,Cantidad_de_estudiantes=12354,Carreras=33)
predict(regresion,data_nueva)

> data_nueva <- data.frame (Presupuesto_para_adquisición_de_plataformas=90000,Cantidad_de_estudiantes=12354,Carreras=33)
> predict(regresion,data_nueva)
      1
1. 51561

data_nueva <- data.frame (Presupuesto_para_adquisición_de_plataformas=200000,Cantidad_de_estudiantes=20000,Carreras=45)
predict(regresion,data_nueva)

> data_nueva <- data.frame (Presupuesto_para_adquisición_de_plataformas=200000,Cantidad_de_estudiantes=20000,Carreras=45)
> predict(regresion,data_nueva)
      1
2. 381828

data_nueva <- data.frame (Presupuesto_para_adquisición_de_plataformas=400000,Cantidad_de_estudiantes=16000,Carreras=60)
predict(regresion,data_nueva)
```

**Fuente: Elaboración propia.**

El segundo modelo realizado fue el de predicción de las plataformas que van a elegir los clientes, se cargaron los datos de la base y las variables que se ayudarán en la predicción. El modelo tiene un 73% de confiabilidad, lo que nos quiere decir que es un buen modelo.

**Figura 9 Resumen del segundo modelo de predicción.**

```
base_de_datos <- read.csv("D:\\dayana\\METODOLOGÍA DE INVESTIGACIÓN\\Presentar\\CSVPrediccionR-v1.2.csv")
resumen <- summary(base_de_datos)
resumen
plot(base_de_datos$Plataformas_elegidas,base_de_datos$Presupuesto_para_adquisición_de_plataformas,main="Influencia del
regresion <- lm(Plataformas_elegidas ~ Presupuesto_para_adquisición_de_plataformas +Cantidad_de_estudiantes + Carreras,
summary(regresion)
#Modelo de regresión
data_nueva <- data.frame (Presupuesto_para_adquisición_de_plataformas=50000,Cantidad_de_estudiantes=20000,Carreras=35)
predict(regresion,data_nueva)

> summary(regresion)

Call:
lm(formula = Plataformas_elegidas ~ Presupuesto_para_adquisición_de_plataformas +
  Cantidad_de_estudiantes + Carreras, data = base_de_datos)

Residuals:
    Min       1Q   Median       3Q      Max
-8.184 -2.908 -1.779  1.655 16.532

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      2.035e+00  1.084e+00  1.877  0.0632 .
Presupuesto_para_adquisición_de_plataformas  1.370e-05  5.285e-06  2.592  0.0109 *
Cantidad_de_estudiantes  1.017e-07  3.765e-06  0.027  0.9785
Carreras          9.019e-03  3.537e-02  0.255  0.7992
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.887 on 108 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared:  0.07322,    Adjusted R-squared:  0.04747
F-statistic: 2.844 on 3 and 108 DF,  p-value: 0.04113
```

**Fuente: Elaboración propia.**

A continuación, se muestran algunas de las pruebas realizadas para la predicción.

**Figura 10 Resultados de las pruebas del segundo modelo.**

```
> data_nueva <- data.frame (Presupuesto_para_adquisición_de_plataformas=50000,Cantidad_de_estudiantes=20000,Carreras=3
5)
> predict(regresion,data_nueva)
      1
3.037519

> data_nueva <- data.frame (Presupuesto_para_adquisición_de_plataformas=20000,Cantidad_de_estudiantes=50000,Carreras=6
0)
> predict(regresion,data_nueva)
      1
5.321049
>

> data_nueva <- data.frame (Presupuesto_para_adquisición_de_plataformas=100000,Cantidad_de_estudiantes=30000,Carreras=45)
> predict(regresion,data_nueva)
      1
3.813728
```

**Fuente: Elaboración propia.**

El tercer modelo fue para predecir clientes potenciales, se colocaron las variables, entre ellas la plataforma que elegiría el cliente. En el resumen del modelo nos indica que tiene una confiabilidad del 100%, es decir, es un buen modelo.

Figura 11 Resumen del tercer modelo de predicción.

```
regression <- lm(Cliente ~ Cantidad_de_estudiantes + Carreras + Modalidad_virtual + Compró_en_algún_momento_plataformas_de_Difucien + Plataformas_presentadas,
summary(regression))
```

```
> regression <- lm(Cliente ~ Cantidad_de_estudiantes + Carreras + Modalidad_virtual + Compró_en_
a=base_de_datos)
> summary(regression)
```

Call:  
lm(formula = Cliente ~ Cantidad\_de\_estudiantes + Carreras + Modalidad\_virtual +  
Compró\_en\_algún\_momento\_plataformas\_de\_Difucien + Plataformas\_presentadas,  
data = base\_de\_datos)

Residuals:  
Min 1Q Median 3Q Max  
-33.093 -12.569 -1.369 13.076 40.277

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.833e+01	1.051e+01	1.744	0.0840 .
Cantidad_de_estudiantes	6.157e-06	1.275e-05	0.483	0.6301 .
Carreras	2.134e-01	1.286e-01	1.660	0.0999 .
Modalidad_virtual	1.174e+01	1.974e-03	5945.542	<2e-16 ***
Compró_en_algún_momento_plataformas_de_Difucien	1.095e+01	9.734e+00	1.125	0.2631 .
Plataformas_presentadas	-1.288e+00	5.543e-01	-2.323	0.0221 *

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.49 on 106 degrees of freedom  
(1 observation deleted due to missingness)  
Multiple R-squared: 1, Adjusted R-squared: 1  
F-statistic: 2.513e+08 on 5 and 106 DF, p-value: < 2.2e-16

Fuente: Elaboración propia.

En la siguiente imagen se muestran los resultados de las pruebas predictivas.

Figura 12 Resultado de la predicción del tercer modelo.

```
data_nueva <- data.frame (Cantidad_de_estudiantes=60000,Carreras=55,Modalidad_virtual=1,Compró_en_algún_momento_plataformas_de_Difucien=1,Plataformas_presentadas=3)
predict(regression,data_nueva)
```

```
> data_nueva <- data.frame (Cantidad_de_estudiantes=60000,Carreras=55,Modalidad_virtual=1,Compró_en_algún_momento_plataformas_de_Difucien=1,Plataformas_presentadas=3)
> predict(regression,data_nueva)
1
49.26605
```

```
> data_nueva <- data.frame (Cantidad_de_estudiantes=20000,Carreras=25,Modalidad_virtual=1,Compró_en_algún_momento_plataformas_de_Difucien=0,Plataformas_presentadas=5)
> predict(regression,data_nueva)
1
29.09063
```

```
> data_nueva <- data.frame (Cantidad_de_estudiantes=45000,Carreras=40,Modalidad_virtual=0,Compró_en_algún_momento_plataformas_de_Difucien=1,Plataformas_presentadas=3)
> predict(regression,data_nueva)
1
34.23543
```

```
> data_nueva <- data.frame (Cantidad_de_estudiantes=15000,Carreras=12,Modalidad_virtual=0,Compró_en_algún_momento_plataformas_de_Difucien=0,Plataformas_presentadas=4)
> predict(regression,data_nueva)
1
15.83624
```

Fuente: Elaboración propia.

A continuación, se encuentra el formulario elaborado en R que muestra el proceso y resultados de los modelos de predicción.

**Figura 13 Formulario con resultados de la predicción parte 1.**

## Reporte de predicción de adquisición de plataformas de la empresa Difucien Ecuador

Dayana Guzmán

### Regresión lineal Predicción de la Cantidad de plataformas que puede adquirir

```
base_de_datos <- read.csv("D:\\Dayana\\METODOLOGIA DE INVESTIGACION\\Presentar\\CSVPrediccionenR.csv")
resumen <- summary(base_de_datos)
resumen
```

```
## Cantidad_de_estudiantes Carreras Modalidad_virtual
## Min. : 226 Min. : 2.00 Min. :0.0000
## 1st Qu.: 2750 1st Qu.:11.75 1st Qu.:0.0000
## Median : 7000 Median :22.50 Median :0.0000
## Mean :11188 Mean :24.02 Mean :0.2667
## 3rd Qu.:13000 3rd Qu.:31.25 3rd Qu.:1.0000
## Max. :52000 Max. :63.00 Max. :1.0000
## Presupuesto_para_adquisicion_de_plataformas
## Min. : 20000
## 1st Qu.: 80000
## Median :100000
## Mean :129833
## 3rd Qu.:162500
## Max. :500000
## Compr_en_algun_momento_plataformas_de_Difucien Plataformas_presentadas
## Min. :0.00 Min. : 3.000
## 1st Qu.:1.00 1st Qu.: 3.000
## Median :1.00 Median : 4.000
## Mean :0.95 Mean : 6.317
```

```
plot(base_de_datos$Plataformas_elegidas,base_de_datos$Presupuesto_para_adquisicion_de_plataformas,type="b",main="Influencia del presupuesto en la adquisición de plataformas", xlab="Plataformas elegidas", ylab="Presupuesto")
```

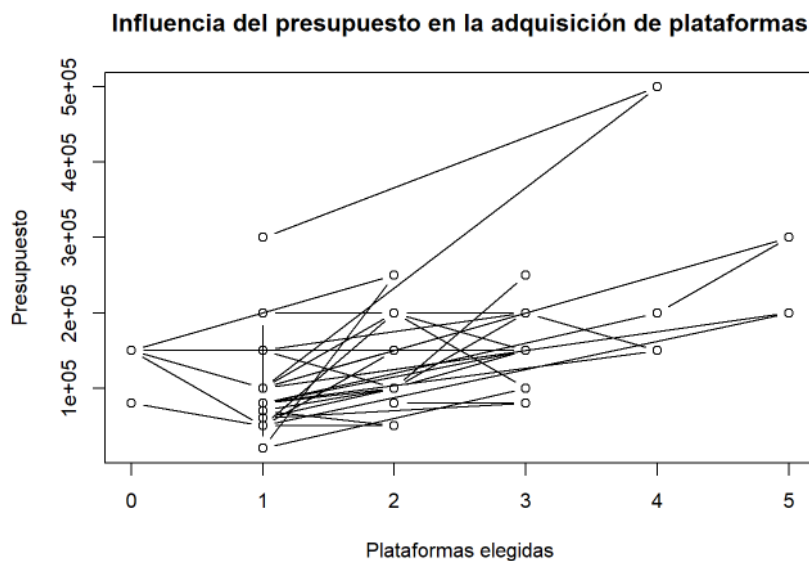


Figura 14 Formulario con resultados de la predicción parte 2.

## Modelo de regresión lineal

```
regresion <- lm(Plataformas.elegidas ~ Presupuesto_para_adquisicion_de_plataformas +Cantidad_de_estudiantes + Carreras, data=base_de_datos)
summary(regresion)
```

```
##
## Call:
## lm(formula = Plataformas.elegidas ~ Presupuesto_para_adquisicion_de_plataformas +
##   Cantidad_de_estudiantes + Carreras, data = base_de_datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0608 -0.5087 -0.1869  0.5384  2.6393
##
## Coefficients:
##              Estimate Std. Error t value
## (Intercept)    7.795e-01  2.738e-01  2.847
## Presupuesto_para_adquisicion_de_plataformas  8.042e-06  1.659e-06  4.848
## Cantidad_de_estudiantes -7.258e-06  1.408e-05 -0.515
## Carreras        3.090e-03  1.190e-02  0.260
##              Pr(>|t|)
## (Intercept)    0.00615 **
## Presupuesto_para_adquisicion_de_plataformas 1.03e-05 ***
## Cantidad_de_estudiantes    0.60825
## Carreras        0.79617
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9604 on 56 degrees of freedom
## Multiple R-squared:  0.3291, Adjusted R-squared:  0.2932
## F-statistic: 9.157 on 3 and 56 DF,  p-value: 5.027e-05
```

Figura 15 Formulario con resultados de la predicción parte 3.

```
data_nueva <- data.frame (Presupuesto_para_adquisicion_de_plataformas=400000,Cantidad_de_estudiantes=16000,Carreras=60)
predict(regresion,data_nueva)
```

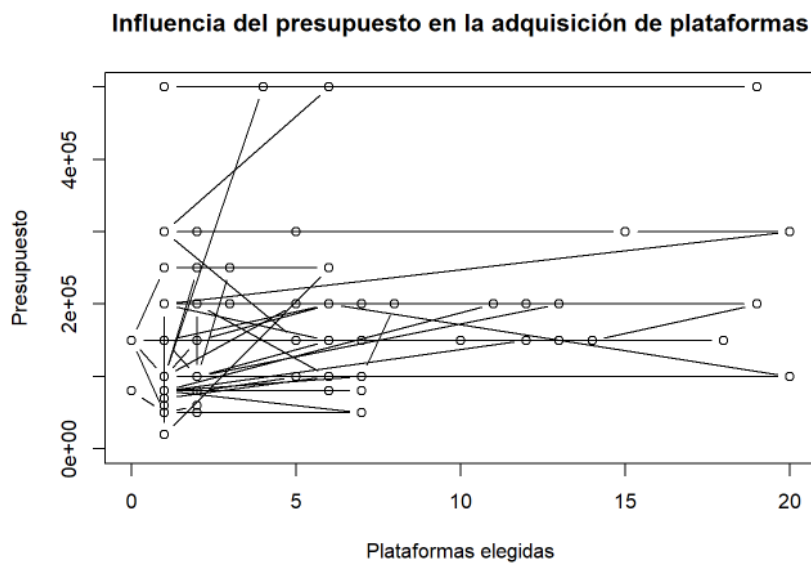
```
##      1
## 4.065631
```

## Regresión lineal Predicción de elección de plataformas

```
base_de_datos <- read.csv("D:\\Dayana\\METODOLOGIA DE INVESTIGACION\\Presentar\\CSVPrediccionenR-v1.2.csv")
resumen <- summary(base_de_datos)
resumen
```

```
##      Cliente      Cantidad_de_estudiantes      Carreras      Modalidad_virtual
## Min.   : 1      Min.   : 12      Min.   : 0.00      Min.   : 0.0
## 1st Qu.: 15     1st Qu.: 3017     1st Qu.:14.00     1st Qu.: 0.0
## Median : 26     Median : 9541     Median :27.00     Median : 0.0
## Mean   : 5269    Mean   : 23669    Mean  :25.93     Mean  : 446.6
## 3rd Qu.: 42     3rd Qu.: 13000    3rd Qu.:31.25     3rd Qu.: 0.0
## Max.   :587000   Max.   :1312701   Max.   :63.00     Max.   :50000.0
## NA's   :1       NA's   :1         NA's   :1         NA's   :1
## Presupuesto_para_adquisicion_de_plataformas
## Min.   : 1
## 1st Qu.:100000
## Median :150000
## Mean   :156964
## 3rd Qu.:200000
## Max.   :500000
## NA's   :1
## Compro_en_algun_momento_plataformas_de_Difucien Plataformas_presentadas
## Min.   : 0.000      Min.   : 3.000
## 1st Qu.: 1.000      1st Qu.: 3.000
```

```
plot(base_de_datos$Plataformas_elegidas,base_de_datos$Presupuesto_para_adquisicion_de_plataformas,type="b",main="Influencia del presupuesto en la adquisición de plataformas", xlab= "Plataformas elegidas", ylab="Presupuesto")
```



**Figura 16 Formulario con resultados de la predicción parte 4.**

```
data_nueva <- data.frame (Presupuesto_para_adquisicion_de_plataformas=350000,Cantidad_de_estudiantes=60000,Carreras=55)
predict(regresion,data_nueva)
```

```
##      1
## 7.295607
```

## Regresión lineal Predicción de clientes potenciales

```
regresion <- lm(Cliente ~ Cantidad_de_estudiantes + Carreras + Modalidad_virtual + Compro_en_algun_momento_plataformas_de_Difucien + Plataformas_presentadas, data=base_de_datos)
summary(regresion)
```

```
##
## Call:
## lm(formula = Cliente ~ Cantidad_de_estudiantes + Carreras + Modalidad_virtual +
##   Compro_en_algun_momento_plataformas_de_Difucien + Plataformas_presentadas,
##   data = base_de_datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -33.093 -12.569  -1.369  13.076  40.277
##
## Coefficients:
##              Estimate Std. Error t value
## (Intercept)    1.833e+01  1.051e+01  1.744
## Cantidad_de_estudiantes  6.157e-06  1.275e-05  0.483
## Carreras        2.134e-01  1.286e-01  1.660
## Modalidad_virtual  1.174e+01  1.974e-03 5945.542
## Compro_en_algun_momento_plataformas_de_Difucien  1.095e+01  9.734e+00  1.125
## Plataformas_presentadas -1.288e+00  5.543e-01  -2.323
##
##>>>
```

```
regresion <- lm(Plataformas_elegidas ~ Presupuesto_para_adquisicion_de_plataformas +Cantidad_de_estudiantes + Carreras, data=base_de_datos)
summary(regresion)
```

```
##
## Call:
## lm(formula = Plataformas_elegidas ~ Presupuesto_para_adquisicion_de_plataformas +
##   Cantidad_de_estudiantes + Carreras, data = base_de_datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
##  -8.141  -2.913  -1.815   1.639  16.491
##
## Coefficients:
##              Estimate Std. Error t value
## (Intercept)    2.097e+00  1.100e+00  1.905
## Presupuesto_para_adquisicion_de_plataformas  1.354e-05  5.322e-06  2.545
## Cantidad_de_estudiantes  9.289e-08  3.780e-06  0.025
## Carreras        8.237e-03  3.557e-02  0.232
##
## Pr(>|t|)
## (Intercept)    0.0594 .
## Presupuesto_para_adquisicion_de_plataformas  0.0124 *
## Cantidad_de_estudiantes  0.9804
## Carreras        0.8173
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.907 on 107 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.07049, Adjusted R-squared:  0.04442
## F-statistic: 2.705 on 3 and 107 DF, p-value: 0.04906
```

**Figura 17** Formulario con resultados de la predicción parte 5.

```
##                                Pr(>|t|)
## (Intercept)                    0.0840 .
## Cantidad_de_estudiantes        0.6301
## Carreras                       0.0999 .
## Modalidad_virtual              <2e-16 ***
## Compro_en_algun_momento_plataformas_de_Difucien 0.2631
## Plataformas_presentadas       0.0221 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.49 on 106 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  1, Adjusted R-squared:  1
## F-statistic: 2.513e+08 on 5 and 106 DF,  p-value: < 2.2e-16
```

```
data_nueva <- data.frame (Cantidad_de_estudiantes=15000,Carreras=12,Modalidad_virtual=0,Compro_en_algun_momento_p
lataformas_de_Difucien=0,Plataformas_presentadas=3, Plataformas_elegidas=19)
predict(regresion,data_nueva)
```

```
##          1
## 17.12411
```





**ESCUELA DE POSGRADO**

**MAESTRÍA EN INGENIERÍA DE SISTEMAS CON MENCIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN**

**Declaratoria de Autenticidad del Asesor**

Yo, PACHECO TORRES JUAN FRANCISCO, docente de la ESCUELA DE POSGRADO MAESTRÍA EN INGENIERÍA DE SISTEMAS CON MENCIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN de la UNIVERSIDAD CÉSAR VALLEJO SAC - TRUJILLO, asesor de Tesis titulada: "Machine Learning para predecir la adquisición de plataformas educativas de la empresa Difucien Ecuador, 2022", cuyo autor es GUZMAN VELEZ DAYANA MARIUXI, constato que la investigación cumple con el índice de similitud establecido, y 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.

TRUJILLO, 07 de Agosto del 2022

<b>Apellidos y Nombres del Asesor:</b>	<b>Firma</b>
PACHECO TORRES JUAN FRANCISCO <b>DNI:</b> 18167212 <b>ORCID</b> 0000-0002-8674-3782	Firmado digitalmente por: JPACHECO el 07-08- 2022 10:04:19

Código documento Trilce: TRI - 0398483