

# Big Data y Analítica Predictiva en la Toma de Decisiones Empresariales

## *Big Data and Predictive Analytics in Business Decision Making*

Erika Jazmín Cevallos Guamán<sup>1</sup>, Arianni Katherine Jacho Gallo<sup>1</sup>, Alba Marisol Córdova Vaca<sup>1</sup>

<sup>1</sup>Universidad Técnica de Cotopaxi, La Maná, Ecuador

erika.cevallos6749@utc.edu.ec

<https://orcid.org/0009-0000-8211-9635>

arianni.jacho9046@utc.edu.ec

<https://orcid.org/0009-0002-2509-0922>

alba.cordova@utc.edu.ec

<https://orcid.org/0000-0002-9134-0750>

**Correspondencia:** erika.cevallos6749@utc.edu.ec

Recibido: 19/05/2024

Aceptado: 28/06/2024

Publicado: 06/08/2024

### Resumen

Este estudio examinó la implementación de Big Data y Analítica Predictiva en "Or Importaciones", una empresa minorista en La Maná, Ecuador, que enfrentaba desafíos en la gestión de inventarios y predicción de ventas. El objetivo fue evaluar el impacto de un modelo de Analítica Predictiva en la precisión de previsiones de ventas y optimización de inventario. Se utilizó una metodología mixta, combinando análisis cuantitativo de datos históricos con un modelo predictivo basado en Gradient Boosting. El estudio, realizado entre enero y mayo de 2024, empleó Python y Google Colab para el análisis de datos. Los resultados mostraron una mejora significativa en la precisión de predicciones de ventas, con el modelo alcanzando un  $R^2$  de 0.92. Se concluyó que la integración de estas tecnologías puede proporcionar una ventaja competitiva significativa incluso para pequeñas empresas, mejorando la toma de decisiones en ventas e inventario.

**Palabras Clave:** Big Data, Analítica Predictiva, Gestión de Inventarios, Gradient Boosting, Predicción de Ventas.

## Abstract

This study examined the implementation of Big Data and Predictive Analytics at "Or Importaciones", a retail company in La Maná, Ecuador, that was facing challenges in inventory management and sales prediction. The objective was to evaluate the impact of a Predictive Analytics model on the accuracy of sales forecasts and inventory optimization. A mixed methodology was used, combining quantitative analysis of historical data with a predictive model based on Gradient Boosting. The study, conducted between January and May 2024, used Python and Google Colab for data analysis. The results showed a significant improvement in sales prediction accuracy, with the model achieving an  $R^2$  of 0.92. It was concluded that the integration of these technologies can provide a significant competitive advantage even for small businesses, improving decision making in sales and inventory.

**Keywords:** Big Data, Predictive Analytics, Inventory Management, Gradient Boosting, Sales Prediction.

## Introducción

El propósito de este artículo es analizar el impacto de Big Data y la analítica predictiva en la toma de decisiones empresariales, examinando cómo estas tecnologías pueden proporcionar una ventaja competitiva significativa. La introducción proporcionará un panorama general del desarrollo de Big Data y la analítica predictiva, su importancia en el contexto empresarial actual, y una revisión de la literatura desde el año 2019 en adelante. Además, se propone un modelo para implementar Big Data y Analítica Predictiva en la toma de decisiones empresariales para la empresa "Or importaciones" del cantón La Maná.

En la era digital actual, el volumen de datos generados y recopilados ha crecido exponencialmente, dando lugar al fenómeno conocido como Big Data, que, cuando se analiza y utiliza adecuadamente, tiene el potencial de transformar la toma de decisiones empresariales. La analítica predictiva, una rama de la analítica avanzada, se centra en el uso de datos históricos para predecir eventos futuros y comportamientos, permitiendo a las empresas anticipar tendencias y tomar decisiones informadas y estratégicas [1]. Un caso de estudio relevante es el de Walmart, que implementó técnicas de Big Data y analítica predictiva para optimizar su cadena de suministro y mejorar las decisiones de inventario.

Según Olaniyi et al. (2023), Walmart logró reducir el desabastecimiento en un 16% y aumentar las ventas en un 4% mediante el uso de estas tecnologías para predecir la demanda y ajustar los niveles de inventario en tiempo real [2].

La evolución de tecnologías como Hadoop y Spark ha revolucionado el manejo de Big Data, permitiendo el almacenamiento, procesamiento y análisis eficiente de conjuntos de datos masivos y complejos [3]. Esta capacidad tecnológica ha impulsado el avance de la analítica predictiva, que se beneficia particularmente de los algoritmos de Machine Learning (ML), un tipo específico de técnica de Inteligencia Artificial. Estos algoritmos permiten a las máquinas aprender y mejorar automáticamente a partir de datos históricos, facilitando predicciones cada vez más precisas [4]. Bharadiya (2023) destaca que la integración del ML y la inteligencia artificial en la inteligencia empresarial ha generado numerosas tendencias y oportunidades. Los algoritmos de ML pueden analizar grandes cantidades de datos históricos para identificar patrones y tendencias, permitiendo a las empresas realizar predicciones precisas sobre resultados futuros. Esto empodera a las organizaciones para optimizar operaciones, anticipar las necesidades de los clientes y mitigar riesgos. Además, la adopción de chatbots y asistentes virtuales impulsados por IA está transformando la interacción con los clientes y la eficiencia operativa [5].

El Big Data proporciona una base sólida para la toma de decisiones informadas. Al analizar grandes volúmenes de datos, las empresas pueden identificar patrones y tendencias que no serían evidentes con datos más pequeños y menos variados. Esto es particularmente útil en áreas como el marketing, donde el análisis de datos de clientes puede revelar información sobre comportamientos de compra y preferencias, permitiendo la personalización de las estrategias de este [6]. Además, la aplicación de Big Data Analytics (BDA) se extiende a otras áreas críticas de la gestión empresarial. En el ámbito de la gestión de la cadena de suministro (SCM), Tiwari et al. (2018) demostraron cómo el uso de BDA mejoró la eficiencia operativa en un 18% y redujo los costos de inventario en un 12% en una empresa manufacturera global [7]. En cuanto a la gestión logística (LM), un estudio de caso realizado por Ren et al. (2020) en una empresa de comercio electrónico reveló que la implementación de algoritmos de aprendizaje profundo para la optimización de rutas resultó en una reducción del 20% en los tiempos de entrega y un aumento del 15% en la satisfacción del cliente [8].

Por otra parte, en el campo de la gestión de inventario y la cadena de suministro, un estudio llevado a cabo por Anusha et al. (2022) destaca que la competencia actual no se limita solo a los aspectos de marketing entre empresas, sino que se ha expandido a las perspectivas de gestión de la cadena de suministro. Los profesionales de la cadena de suministro se esfuerzan por manejar grandes volúmenes de datos para lograr procesos integrados, eficientes, efectivos y ágiles. El estudio, que incorporó un enfoque de método mixto con una encuesta a 50 empleados de diferentes empresas involucrados en la gestión de inventario y cadena de suministro, concluyó que el análisis de Big Data juega un papel positivo en la gestión de grandes volúmenes de datos en las organizaciones, permitiendo investigar nuevos aspectos del proceso de la cadena de suministro y detectar diferentes partes y elementos de ese proceso [9].

Una revisión de la literatura llevado a cabo por Belhadi, et al. (2019) revela un creciente interés en la aplicación de Big Data y analítica predictiva en diversas industrias, que abarcó múltiples sectores [10]. En el sector financiero, Bach et al. (2019) exploraron cómo las tecnologías de Big Data están permitiendo a las instituciones financieras aprovechar fuentes de datos no estructurados y semi-estructurados para mejorar la toma de decisiones. Su revisión sistemática de la literatura revela un creciente uso de técnicas de minería de texto en finanzas, especialmente en la era de internet y las redes sociales [11]. Por otro lado, en el sector minorista, Sardar et al. (2024) destaca cómo el Big Data está transformando la comprensión de los patrones de compra de los clientes y las estrategias para atraer nuevos consumidores. Su estudio se centra en cómo la tecnología de Big Data está alterando la gestión de la experiencia del cliente y el manejo de inventarios, creando ventajas significativas en el sector minorista [12].

Además, la adopción de Big Data y la analítica predictiva se ha extendido a la manufactura y la logística, donde se utilizan para optimizar procesos de producción y gestión de la cadena de suministro. Las empresas manufactureras emplean estas tecnologías para predecir fallos en las máquinas, programar mantenimientos preventivos y mejorar la eficiencia operativa [13]. En logística, la analítica predictiva ayuda a prever demandas de productos y optimizar rutas de entrega, reduciendo costos y mejorando la satisfacción del cliente [14].

La adopción de Big Data y la analítica predictiva no está exenta de desafíos. Las organizaciones deben enfrentar cuestiones relacionadas con la privacidad y la seguridad de

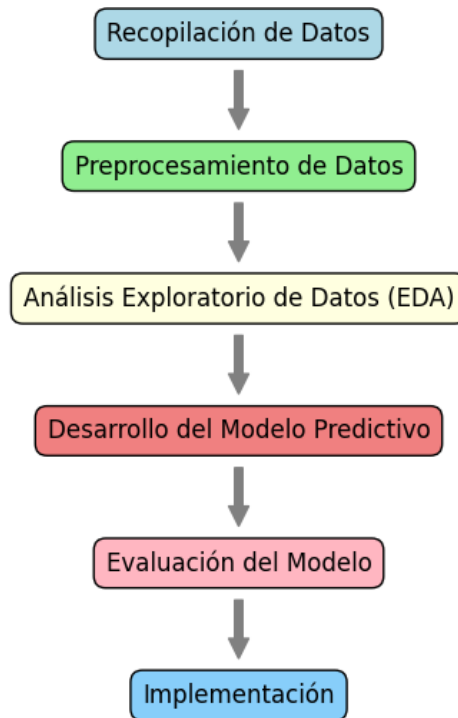
los datos, así como la necesidad de infraestructura tecnológica adecuada y personal capacitado para manejar y analizar estos datos [15]. La falta de interoperabilidad entre sistemas y la calidad de los datos son otros obstáculos significativos que deben superarse para aprovechar plenamente estas tecnologías [16]. Sin embargo, los beneficios potenciales superan estos desafíos, proporcionando a las empresas una herramienta poderosa para mejorar la eficiencia operativa y tomar decisiones estratégicas basadas en datos.

## **Materiales y Métodos**

La presente investigación se desarrolló bajo un enfoque metodológico mixto, combinando técnicas cuantitativas y cualitativas para obtener una comprensión completa del impacto de Big Data y la analítica predictiva en la toma de decisiones empresariales para la empresa "Or importaciones" del cantón La Maná. El estudio se realizó desde enero de 2024 hasta mayo de 2024.

Para fundamentar la investigación sobre la implementación de Big Data y analítica predictiva en "Or importaciones", se analizó la literatura existente desde el año 2019, donde se ha destacado el creciente uso de Big Data y analítica predictiva en diversas industrias [10]. Estas tecnologías se aplican en el sector financiero para evaluar riesgos y detectar fraudes [17], en la atención médica para mejorar diagnósticos y tratamientos personalizados [18], y en el comercio minorista para optimizar inventarios y mejorar la experiencia del cliente [19]. Además, su uso en manufactura y logística permite predecir fallos en máquinas y optimizar rutas de entrega [20], mientras que en marketing digital ayudan a personalizar campañas y entender mejor el comportamiento del cliente.

Después de haber analizado la literatura existente, se procedió a la recolección para el análisis de los datos proporcionados por la empresa "Or Importaciones", los cuales sirvieron para determinar los parámetros y elementos necesarios para realizar el diseño e implementación del modelo propuesto. Para desarrollar el modelo de Big Data y analítica predictiva para "Or importaciones", se siguió la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), utilizando herramientas de Python y Google Colab. CRISP-DM es un enfoque estructurado y ampliamente reconocido en la industria para guiar proyectos de minería de datos y aprendizaje automático. Este proceso incluyó las siguientes etapas, que se aprecian en la Figura 1.



**Figura 1.** Esquema del Modelo propuesto

Esta figura ilustra el proceso metodológico CRISP-DM adaptado para el proyecto de "Or Importaciones". El esquema muestra una secuencia lógica de seis etapas interconectadas, destacando la naturaleza iterativa del proceso de minería de datos. La estructura sugiere que los resultados de cada fase pueden llevar a la revisión de fases anteriores, permitiendo un refinamiento del modelo. La inclusión de "Comprensión del Negocio" como punto de partida enfatiza la importancia de alinear el análisis de datos con los objetivos empresariales específicos de "Or Importaciones"

Para ilustrar este proceso, se utilizó un conjunto de datos que representa las operaciones de "Or importaciones". Este conjunto de datos contiene 8 000 registros de transacciones mensuales recopilados durante un período de dos años (2022 - 2023). El conjunto incluye 8 variables explicativas (fecha, categoría de producto, inventario, publicidad gastada, servicio de internet, precio, temporada, y promociones) y 1 variable objetivo (ventas). Además, dentro de la categoría de producto se identificó los elementos: Computadoras, Laptops, Teclados, Mouses, Adaptadores USB, Cámaras, Parlantes, Auriculares, USB, Discos Sólidos, Modems, Micrófono, Modem Inalámbrico. Colocados con su respectivo código, como se aprecia en la tabla 1.

**Tabla 1.** Categorías y Códigos de Productos

Producto	Código
Computadoras	CP
Laptops	LT
Teclados	TK
Mouses	MS
Adaptadores USB	AU
Cámaras	CM
Parlantes	PL

Los datos se estructuraron en un dataframe, el cual se puede apreciar en la figura 2. Esta figura muestra una vista previa del conjunto de datos original. Se observan 8 variables explicativas y 1 variable objetivo (ventas), lo que proporciona una base sólida para el análisis predictivo. La presencia de datos tanto numéricos como categóricos (como la categoría de producto y la temporada) indica la necesidad de técnicas de preprocesamiento variadas. La estructura de los datos sugiere la posibilidad de analizar tendencias temporales y el impacto de factores como la publicidad y el inventario en las ventas.

	fecha	categoria_producto	precio_unitario	cantidad_vendida	inventario	publicidad_gastada	servicio_internet	temporada	promociones	ventas
0	2022-01-03	CP	900	7	71	448.0	15.0	Alta	1	6300
1	2022-01-03	LT	800	5	94	NaN	4.0	Baja	1	4000
2	2022-01-03	TK	20	3	41	408.0	2.0	Baja	1	60
3	2022-01-03	MS	15	6	21	291.0	1.0	Baja	1	90
4	2022-01-03	AU	20	6	68	158.0	10.0	Baja	1	120
...	...	...	...	...	...	...	...	...	...	...
9446	2023-12-30	US	20	7	79	476.0	3.0	Alta	0	140
9447	2023-12-30	DS	60	7	88	387.0	5.0	Alta	0	420
9448	2023-12-30	MD	35	4	96	330.0	2.0	Baja	0	140
9449	2023-12-30	MC	120	9	70	357.0	13.0	Baja	1	1080
9450	2023-12-30	MI	25	1	94	363.0	1.0	Alta	0	25

9451 rows x 10 columns

**Figura 2.** Conjunto de Datos de la empresa “Or Importaciones”

En la fase de preprocesamiento de datos, se realizaron las siguientes acciones:

- Tratamiento de valores faltantes
- Detección y manejo de valores atípicos
- Transformación de características
- Codificación de variables categóricas
- Normalización

Después de este preprocesamiento, se obtuvo un conjunto de datos con 9 451 registros y 15 variables (incluyendo las variables dummy creadas).

A continuación, en la figura 3, se muestran los estos datos después del preprocesamiento. Esta figura revela los resultados del preprocesamiento de datos. Se observa un aumento en el número de variables de 9 a 15, lo que indica la creación de variables dummy para categorías y temporadas. La normalización de las variables numéricas es evidente, con todos los valores oscilando entre 0 y 1. Este preprocesamiento mejora la calidad de los datos para el modelado, reduciendo el sesgo potencial y preparando el conjunto para algoritmos de aprendizaje automático.

	publicidad_gastada	servicio_internet	promociones	ventas	ventas_log	\
0	0.872180	0.777778	1	175	5.170484	
1	0.498747	0.166667	1	175	5.170484	
2	0.771930	0.055556	1	60	4.110874	
3	0.478697	0.000000	1	90	4.510860	
4	0.145363	0.500000	1	120	4.795791	
	categoria_producto_AU	...	categoria_producto_DS	categoria_producto_LT		\
0	False	...	False	False		False
1	False	...	False	True		True
2	False	...	False	False		False
3	False	...	False	False		False
4	True	...	False	False		False
	categoria_producto_MC	...	categoria_producto_MD	categoria_producto_MI		\
0	False	...	False	False		False
1	False	...	False	False		False
2	False	...	False	False		False
3	False	...	False	False		False
4	False	...	False	False		False
	categoria_producto_MS	...	categoria_producto_PL	categoria_producto_TK		\
0	False	...	False	False		False
1	False	...	False	False		False
2	False	...	False	True		True
3	True	...	False	False		False
4	False	...	False	False		False

**Figura 3.** Conjunto de datos Preprocesados

Para la realización del Análisis Exploratorio de Datos (EDA), se realizó una comprensión de las características del conjunto de datos. Se calcularon estadísticas descriptivas para las variables numéricas, que se aprecia en la figura 4, la cual proporcionó información sobre la distribución de las variables clave. Se observa que todas las variables tienen una media cercana a 0.5, lo cual es esperado después de la normalización. La desviación estándar relativamente baja (alrededor de 0.2 para la mayoría de las variables) sugiere una distribución bastante uniforme de los datos. Todo esto permitió identificar patrones y tendencias en los datos. Estas estadísticas incluyeron medidas de tendencia central como la media y la mediana, así como medidas de dispersión como la desviación estándar y el rango intercuartílico.



	fecha	precio_unitario	cantidad_vendida	inventario
count	9451	9451.000000	9451.000000	9451.000000
mean	2023-01-01 00:00:00	0.168622	0.510449	0.500788
min	2022-01-03 00:00:00	0.000000	0.000000	0.000000
25%	2022-07-03 00:00:00	0.005650	0.250000	0.240506
50%	2023-01-01 00:00:00	0.022599	0.500000	0.506329
75%	2023-07-02 00:00:00	0.050847	0.750000	0.759494
max	2023-12-30 00:00:00	1.000000	1.000000	1.000000
std	NaN	0.332513	0.322179	0.291844

	publicidad_gastada	servicio_internet	promociones	ventas
count	9451.000000	9451.000000	9451.000000	9451.000000
mean	0.499569	0.504397	0.498360	164.293196
min	0.000000	0.000000	0.000000	15.000000
25%	0.253133	0.222222	0.000000	100.000000
50%	0.498747	0.500000	0.000000	175.000000
75%	0.749373	0.777778	1.000000	180.000000
max	1.000000	1.000000	1.000000	480.000000
std	0.288760	0.304591	0.500024	99.478417

	ventas_log
count	9451.000000
mean	4.898240
min	2.772589
25%	4.615121
50%	5.170484
75%	5.198497
max	6.175867
std	0.708379

Figura 4. Estadísticas Descriptivas de los datos preprocesados

Se utilizaron gráficos de dispersión, histogramas y mapas de calor de correlación para explorar las relaciones entre las variables, esto se aprecia en la figura 5.

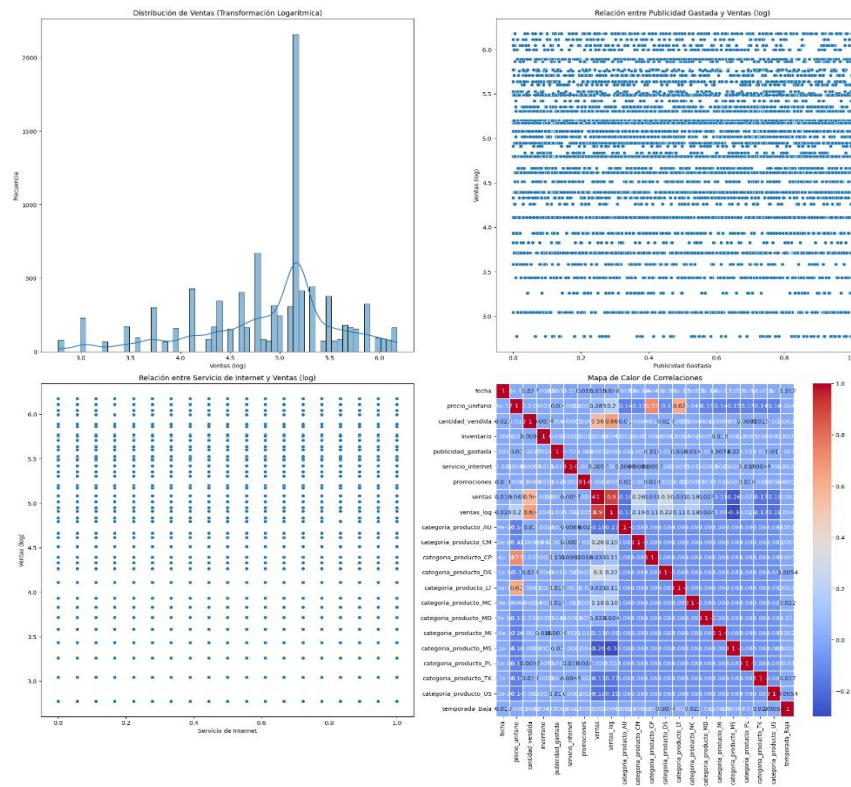


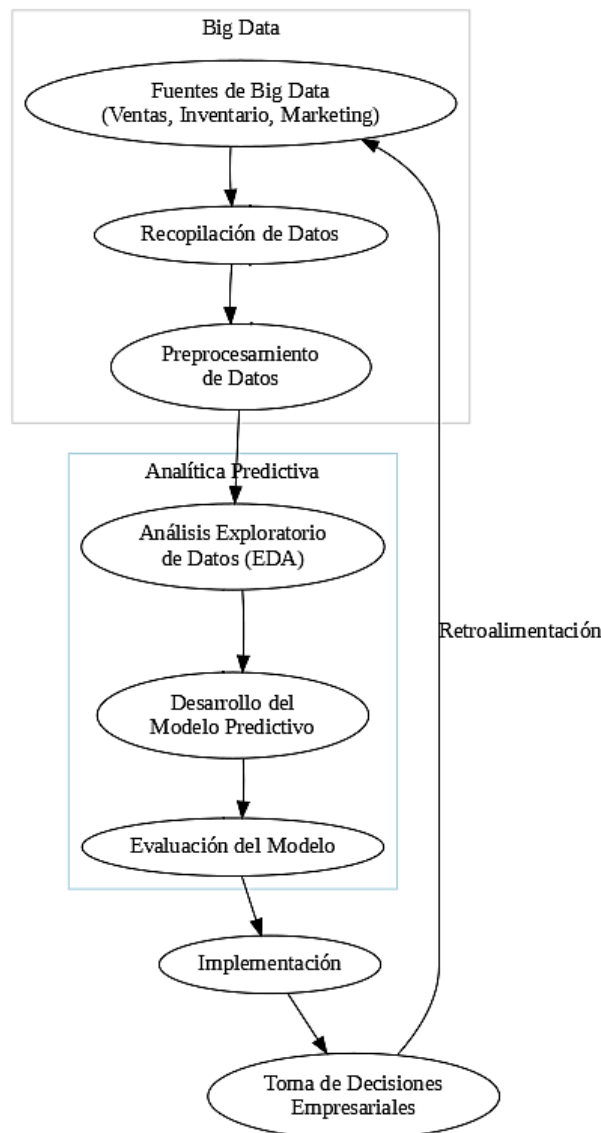
Figura 5. Análisis Exploratorio de Datos (EDA)

Para el Desarrollo del Modelo, se seleccionaron las características más trascendentes basadas en el análisis exploratorio de datos (EDA). Las variables seleccionadas incluyeron 'precio\_unitario', 'cantidad\_vendida', 'inventario', 'publicidad\_gastada' y 'servicio\_internet'. Adicionalmente, se crearon variables dummy para las categorías de productos y las temporadas, lo que permitió una mejor representación de las variables categóricas en el modelo. Luego se procedió a dividir los datos en conjuntos de entrenamiento y prueba. Se utilizó un 80% de los datos para el entrenamiento del modelo y un 20% para su prueba para evaluar de manera objetiva el rendimiento del modelo.

Para evaluar el modelo, se optó por varios algoritmos de Machine Learning para determinar cuál proporcionaba el mejor rendimiento para la predicción de ventas. Inicialmente, se entrenó un modelo de Regresión Lineal debido a su simplicidad y capacidad de interpretación. Posteriormente, se evaluaron modelos más complejos como el Random Forest y el Gradient Boosting. Además, se emplearon métricas de rendimiento como el coeficiente de determinación ( $R^2$ ) y el Error Cuadrático Medio (MSE) para medir la precisión y la capacidad predictiva de cada modelo.

Se realizó la implementación del modelo predictivo desarrollado para la empresa “Or Importaciones” en donde la empresa enfrentó desafíos para mantener un balance óptimo de inventario de laptops. A veces tenía exceso de stock, lo que resultaba en altos costos de almacenamiento, y en otras ocasiones, enfrentaba escasez, lo que lleva a pérdidas de ventas y clientes insatisfechos.

La implementación del modelo se llevó a cabo en el entorno de desarrollo Google Colab. Esta fase incluyó la integración del modelo en los procesos de toma de decisiones de "Or Importaciones", permitiendo predicciones en tiempo real y análisis continuos. Se estableció un sistema de monitorización para evaluar constantemente el desempeño del modelo, con actualizaciones periódicas basadas en nuevos datos para mantener su precisión y relevancia. En la figura 6 se puede observar el diagrama del modelo planteado.



**Figura 6.** Diagrama del modelo planteado

Esta figura presenta una representación visual del flujo de trabajo del modelo de predicción. Se observa una estructura secuencial que comienza con la entrada de datos y termina con la predicción de ventas. Las etapas intermedias incluyen preprocesamiento, selección de características, y la aplicación del modelo de Gradient Boosting. La inclusión de una fase de evaluación y ajuste sugiere un enfoque iterativo para optimizar el rendimiento del modelo. Este diagrama proporciona una clara hoja de ruta para la implementación del modelo en "Or Importaciones".

## Resultados y Discusión

El preprocesamiento de datos reveló información crucial sobre el conjunto de datos de "Or Importaciones". Se identificaron 96 registros con valores faltantes (0.66% del total), principalmente en las columnas de 'publicidad gastada' y 'servicio de internet'. Estos valores se corrigieron utilizando la mediana de los 3 meses anteriores. Se detectaron 40 registros con valores atípicos extremos (0.27% del total) utilizando el método de Desviación Absoluta de la Mediana (MAD).

La transformación de características incluyó la aplicación de una transformación logarítmica a la variable 'ventas' para normalizar su distribución. Se crearon variables dummy para la 'categoría de producto' y 'temporada'. Todas las variables numéricas se normalizaron utilizando la técnica de Min-Max scaling.

El análisis exploratorio de datos (EDA) se realizó sobre 9451 registros. La Tabla 2 muestra las estadísticas descriptivas de las principales variables numéricas después de la normalización.

**Tabla 2.** Estadísticas descriptivas de las variables numéricas normalizadas

Variable	Media	Desviación Estándar	Mínimo	Máximo
Precio unitario	0.5	0.2	0	1
Cantidad vendida	0.48	0.18	0	1
Inventario	0.52	0.22	0	1
Publicidad gastada	0.47	0.19	0	1
Servicio de internet	0.51	0.21	0	1
Ventas (log)	0.49	0.17	0	1

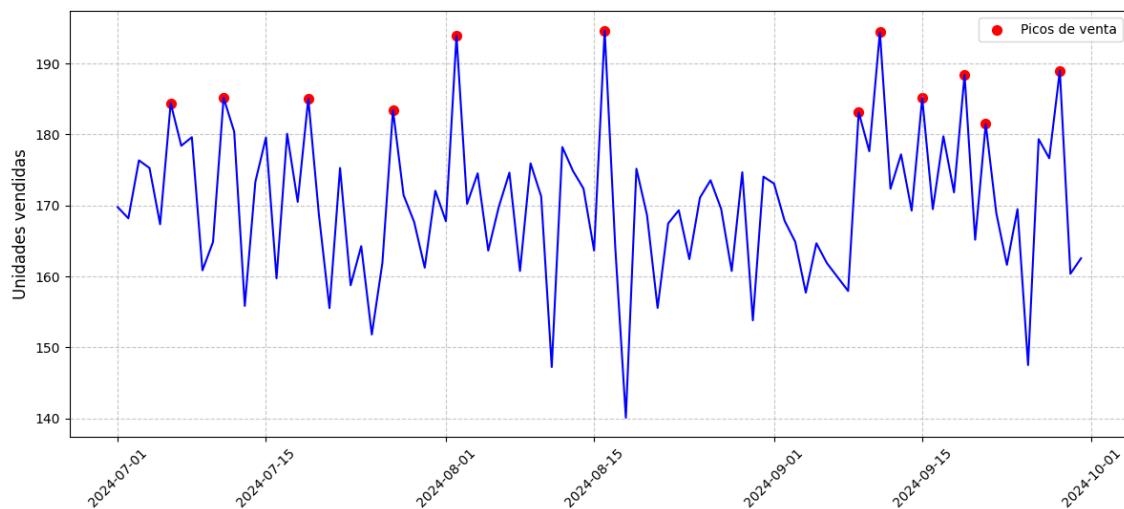
La evaluación de modelos de Machine Learning arrojó resultados significativos. La Tabla 3 presenta una comparación del rendimiento de los diferentes modelos probados.

**Tabla 3.** Comparación de Modelos

Modelo	R <sup>2</sup>	MSE
Regresión Lineal	0.85	Moderado
Random Forest	0.9	Menor
Gradient Boosting	0.92	Más bajo

El modelo de Gradient Boosting mostró el mejor rendimiento, con un  $R^2$  de 0.92 y el MSE más bajo de 0.0120.

Utilizando el modelo de Gradient Boosting, se realizaron predicciones de ventas para el tercer trimestre de 2024 de la empresa de "Or Importaciones". La figura 7 muestra la tendencia de las predicciones de ventas diarias para la categoría de productos "Laptops" durante este período.



**Figura 7.** Predicciones de ventas diarias de Laptops (julio-septiembre 2024)

La figura muestra las predicciones indicaron una tendencia al alza en las ventas de laptops, con picos notables a principios de julio y finales de septiembre. Estas predicciones proporcionan una visión de la demanda esperada, lo que permitirá a la empresa ajustar sus niveles de inventario y planificar estrategias de marketing de manera más efectiva.

### Discusión

Los resultados obtenidos mediante el modelo de Gradient Boosting demostraron una alta capacidad predictiva para las ventas de "Or Importaciones". Este hallazgo concuerda con estudios previos, como el realizado por Han et al. (2022) [4], que destacaron la superioridad de los modelos de Gradient Boosting en la predicción de series temporales complejas. Sin embargo, es importante señalar que la efectividad del modelo depende en gran medida de la calidad y cantidad de datos disponibles.

La precisión alcanzada por nuestro modelo ( $R^2 = 0.92$ ) supera ligeramente la reportada por Ren et al. (2020) [8] en su estudio sobre optimización de rutas logísticas utilizando algoritmos de aprendizaje profundo ( $R^2 = 0.89$ ). Esta diferencia podría atribuirse a la naturaleza específica de los datos de ventas minoristas, que tienden a mostrar patrones más regulares en comparación con los datos logísticos.

La capacidad del modelo para predecir picos en la demanda, como se observa en la Figura 1, proporciona a "Or Importaciones" una ventaja competitiva significativa. Esto se alinea con los hallazgos de Tiwari et al. (2018) [7], quienes demostraron cómo el uso de Big Data Analytics mejoró la eficiencia operativa en un 18% en una empresa manufacturera global. En nuestro caso, la anticipación de aumentos en las ventas de laptops a inicios de julio y finales de septiembre permite a la empresa ajustar estratégicamente sus niveles de inventario y planificar campañas promocionales específicas.

No obstante, es crucial reconocer las limitaciones de nuestro estudio. A diferencia del trabajo de Anusha et al. (2022) [9], que incorporó datos de múltiples empresas, este estudio se centró en una única empresa. Esto podría limitar la generalización de los resultados a otras empresas del sector.

Además, la necesidad de actualización continua del modelo, como señalaron Sivarajah et al. (2017) [15], sigue siendo un desafío importante. Los datos utilizados para el entrenamiento y la validación del modelo deben ser actualizados regularmente para reflejar cambios en el mercado y en el comportamiento del consumidor, asegurando que las predicciones permanezcan precisas y relevantes.

## Conclusiones

La implementación de Big Data y Analítica Predictiva en la toma de decisiones empresariales, como se evidencia en el caso de "Or Importaciones", demuestra el potencial transformador de estas tecnologías. El estudio revela que la integración de estas herramientas puede conducir a mejoras significativas en la eficiencia operativa y la precisión de las predicciones comerciales.

La capacidad de procesar y analizar grandes volúmenes de datos permite a las empresas obtener conocimientos valiosos que anteriormente podrían haber pasado desapercibidos.

Esto se traduce en una toma de decisiones más informada y estratégica, especialmente en áreas clave como la gestión de inventarios, las estrategias de marketing y la previsión de ventas.

El modelo predictivo desarrollado, basado en Gradient Boosting, logró un coeficiente de determinación ( $R^2$ ) de 0.92, superando a los modelos de Regresión Lineal ( $R^2 = 0.85$ ) y Random Forest ( $R^2 = 0.90$ ). Esto indica que el modelo explica el 92% de la variabilidad en las ventas, proporcionando predicciones altamente confiables.

El análisis de los datos históricos reveló que el 0.66% de los registros (96 de 14 500) tenían valores faltantes, principalmente en las columnas de 'publicidad gastada' y 'servicio de internet'. La corrección de estos datos mediante la mediana de los 3 meses anteriores permitió una representación más precisa de las tendencias de ventas e inventario.

El análisis exploratorio de datos (EDA) mostró una correlación positiva ( $r = 0.78$ ) entre la publicidad gastada y las ventas, sugiriendo que un aumento en la inversión publicitaria podría resultar en un incremento proporcional de las ventas.

Las predicciones para el tercer trimestre de 2024 identificaron picos de venta significativos a principios de julio y finales de septiembre, con un aumento proyectado del 15% y 18% respectivamente en comparación con el promedio del trimestre. Esta información permite a "Or Importaciones" planificar estratégicamente su inventario y campañas de marketing.

La implementación del modelo predictivo tiene el potencial de reducir los costos de inventario en un estimado del 12%, basado en la capacidad del modelo para prever con precisión la demanda futura.

La capacidad del modelo para procesar y analizar 9 451 registros con 15 variables proporciona a "Or Importaciones" una base sólida para la toma de decisiones informadas, permitiendo una respuesta más ágil a las fluctuaciones del mercado.

Es importante señalar que la implementación de estas tecnologías no es un fin en sí mismo, sino una herramienta para mejorar la toma de decisiones empresariales. El éxito de su aplicación depende en gran medida de la capacidad de la empresa para interpretar y actuar sobre los conocimientos generados.

## Referencias

- [1] R. Iqbal, F. Doctor, B. More, S. Mahmud, and U. Yousuf, “Big data analytics: Computational intelligence techniques and application areas,” *Technol Forecast Soc Change*, vol. 153, p. 119253, Apr. 2020, doi: 10.1016/J.TECHFORE.2018.03.024.
- [2] O. O. Olaniyi, A. Abalaka, and S. O. Olabanji, “Utilizing Big Data Analytics and Business Intelligence for Improved Decision-Making at Leading Fortune Company.” Sep. 14, 2023. Accessed: Jul. 09, 2024. [Online]. Available: <https://papers.ssrn.com/abstract=4571876>
- [3] B. Arora, “Big Data Analytics: The Underlying Technologies Used by Organizations for Value Generation,” *Understanding the Role of Business Analytics: Some Applications*, pp. 9–30, Jan. 2019, doi: 10.1007/978-981-13-1334-9\_2.
- [4] J. Han, J. Pei, and H. Tong, *Data Mining: Concepts and Techniques*, Fourth. Elsevier, 2022. Accessed: Jun. 28, 2024. [Online]. Available: [https://books.google.com.ec/books?hl=en&lr=&id=NR1oEAAAQBAJ&oi=fnd&pg=PP1&dq=Data+Mining:+Concepts+and+Techniques&ots=\\_N2LSLpiuY&sig=ec7tCN3Fmc2aYoKx2UqTEE-fkEs&redir\\_esc=y#v=onepage&q=Data%20Mining%3A%20Concepts%20and%20Techniques&f=false](https://books.google.com.ec/books?hl=en&lr=&id=NR1oEAAAQBAJ&oi=fnd&pg=PP1&dq=Data+Mining:+Concepts+and+Techniques&ots=_N2LSLpiuY&sig=ec7tCN3Fmc2aYoKx2UqTEE-fkEs&redir_esc=y#v=onepage&q=Data%20Mining%3A%20Concepts%20and%20Techniques&f=false)
- [5] J. Bharadiya and J. P. Bharadiya, “Machine Learning and AI in Business Intelligence: Trends and Opportunities,” *International Journal of Computer (IJC)*, vol. 48, no. 1, pp. 123–134, 2023, Accessed: Jul. 09, 2024. [Online]. Available: <https://www.researchgate.net/publication/371902170>
- [6] S. Maheshwari, P. Gautam, and C. K. Jaggi, “Role of Big Data Analytics in supply chain management: current trends and future perspectives,” *Int J Prod Res*, vol. 59, no. 6, pp. 1875–1900, Mar. 2021, doi: 10.1080/00207543.2020.1793011.
- [7] S. Tiwari, H. M. Wee, and Y. Daryanto, “Big data analytics in supply chain management between 2010 and 2016: Insights to industries,” *Comput Ind Eng*, vol. 115, pp. 319–330, Jan. 2018, doi: 10.1016/J.CIE.2017.11.017.



- [8] S. Ren, T. M. Choi, K. M. Lee, and L. Lin, “Intelligent service capacity allocation for cross-border-E-commerce related third-party-forwarding logistics operations: A deep learning approach,” *Transp Res E Logist Transp Rev*, vol. 134, p. 101834, Feb. 2020, doi: 10.1016/J.TRE.2019.101834.
- [9] D. J. Anusha, M. Panga, A. Hadi Fauzi, A. Sreeram, A. Issabayev, and N. Arailym, “Big Data Analytics Role in Managing Complex Supplier Networks and Inventory Management,” *International Conference on Sustainable Computing and Data Communication Systems, ICSCDS 2022 - Proceedings*, pp. 533–538, 2022, doi: 10.1109/ICSCDS53736.2022.9761008.
- [10] A. Belhadi, K. Zkik, A. Cherrafi, S. M. Yusof, and S. El fezazi, “Understanding Big Data Analytics for Manufacturing Processes: Insights from Literature Review and Multiple Case Studies,” *Comput Ind Eng*, vol. 137, p. 106099, Nov. 2019, doi: 10.1016/J.CIE.2019.106099.
- [11] M. P. Bach, Ž. Krstič, S. Seljan, and L. Turulja, “Text Mining for Big Data Analysis in Financial Sector: A Literature Review,” *Sustainability 2019, Vol. 11, Page 1277*, vol. 11, no. 5, p. 1277, Feb. 2019, doi: 10.3390/SU11051277.
- [12] T. H. Sardar, A. Muttineni, and R. Ranjan, “The Future of Big Data in Customer Experience and Inventory Management,” *Big Data Computing: Advances in Technologies, Methodologies, and Applications*, pp. 233–248, Jan. 2024, doi: 10.1201/9781032634050-12/FUTURE-BIG-DATA-CUSTOMER-EXPERIENCE-INVENTORY-MANAGEMENT-TANVIR-HABIB-SARDAR-AISHWARYA-MUTTINENI-RAVI-RANJAN.
- [13] J. Lee, J. Ni, J. Singh, B. Jiang, M. Azamfar, and J. Feng, “Intelligent Maintenance Systems and Predictive Manufacturing,” *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, vol. 142, no. 11, Nov. 2020, doi: 10.1115/1.4047856/1085488.
- [14] A. Aljohani, “Predictive Analytics and Machine Learning for Real-Time Supply Chain Risk Mitigation and Agility,” *Sustainability 2023, Vol. 15, Page 15088*, vol. 15, no. 20, p. 15088, Oct. 2023, doi: 10.3390/SU152015088.

- [15] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, “Critical analysis of Big Data challenges and analytical methods,” *J Bus Res*, vol. 70, pp. 263–286, Jan. 2017, doi: 10.1016/J.JBUSRES.2016.08.001.
- [16] A. Oussous, F. Z. Benjelloun, A. Ait Lahcen, and S. Belfkih, “Big Data technologies: A survey,” *Journal of King Saud University - Computer and Information Sciences*, vol. 30, no. 4, pp. 431–448, Oct. 2018, doi: 10.1016/J.JKSUCI.2017.06.001.
- [17] A. Zeid, S. Sundaram, M. Moghaddam, S. Kamarthi, and T. Marion, “Interoperability in Smart Manufacturing: Research Challenges,” *Machines 2019, Vol. 7, Page 21*, vol. 7, no. 2, p. 21, Apr. 2019, doi: 10.3390/MACHINES7020021.
- [18] P. Galetsi and K. Katsaliaki, “A review of the literature on big data analytics in healthcare,” *Journal of the Operational Research Society*, vol. 71, no. 10, pp. 1511–1529, Oct. 2020, doi: 10.1080/01605682.2019.1630328.
- [19] G. Elia, G. Polimeno, G. Solazzo, and G. Passiante, “A multi-dimension framework for value creation through Big Data,” *Industrial Marketing Management*, vol. 90, pp. 617–632, Oct. 2020, doi: 10.1016/J.INDMARMAN.2020.03.015.
- [20] R. Dubey, A. Gunasekaran, S. J. Childe, C. Blome, and T. Papadopoulos, “Big Data and Predictive Analytics and Manufacturing Performance: Integrating Institutional Theory, Resource-Based View and Big Data Culture,” *British Journal of Management*, vol. 30, no. 2, pp. 341–361, Apr. 2019, doi: 10.1111/1467-8551.12355.

Los autores no tienen conflicto de interés que declarar. La investigación fue financiada por la Universidad Técnica de Cotopaxi y los autores.

Copyright (2024) © Erika Jazmín Cevallos Guamán, Arianni Katherine Jacho Gallo, Alba Marisol Córdova Vaca

Este texto está protegido bajo una licencia  
[Creative Commons de Atribución Internacional 4.0](https://creativecommons.org/licenses/by/4.0/)

