

# Use of Big Data and Predictive Analytics in Digital Platform Marketing: A Systematic Review

Abigail N. Quispe-Florida<sup>1</sup>  and Cristofher Z. Vargas<sup>2</sup> 

<sup>1,2</sup>Universidad Tecnológica del Perú, Lima, Peru, u18310388@utp.edu.pe, crzuniga@utp.edu.pe

*Resumen—Nowadays, marketing companies face several challenges such as high competition, market complexity and technological innovation. Faced with this, they are increasingly turning to the use of big data and predictive analytics. However, knowledge in this area is limited as technological changes make it difficult to keep pace with the actual performance of these strategies, so the objective of this review is to identify the big data techniques used in digital platform marketing for strategic sales. A total of 764 papers were identified in the Scopus and Web of Science databases, where 28 passed the selection criteria under the PRISMA statement. The main findings indicate that the use of big data and machine learning techniques can significantly improve digital marketing and sales strategies. These techniques enable better customer segmentation, content personalization and prediction of buying behavior. In conclusion, although Big Data and machine learning have great potential to optimize strategic digital marketing and sales strategies, more empirical research is required to validate the findings. The present review provides a comprehensive view of the current state and trends in this emerging area of knowledge.*

*Palabras clave—Marketing, Big Data, Analytics, Algorithm, Digital Platform*

**Digital Object Identifier:** (only for full papers, inserted by LACCEI).

**ISSN, ISBN:** (to be inserted by LACCEI).

**DO NOT REMOVE**

# Uso de Big Data y Analítica Predictiva en el Marketing de Plataformas Digitales: Una Revisión sistemática

Abigail N. Quispe-Florida<sup>1</sup>  and Cristofher Z. Vargas<sup>2</sup> 

<sup>1,2</sup>Universidad Tecnológica del Perú, Lima, Peru, u18310388@utp.edu.pe, crzuniga@utp.edu.pe

**Resumen**—En la actualidad las empresas de marketing se enfrentan a diversos desafíos como alta competencia, complejidad del mercado e innovación tecnológica. Ante esto, se recurre cada vez más al uso de big data y la analítica predictiva. Sin embargo, el conocimiento en esta área es limitado ya que los cambios tecnológicos hacen que sea difícil seguir el ritmo del rendimiento real de estas estrategias, por ello el objetivo de la presente revisión es identificar las técnicas de big data utilizadas en el marketing de plataformas digitales para la obtención de ventas estratégicas. Fueron identificados un total de 764 documentos en las bases de datos Scopus y Web of Science, donde 28 pasaron los criterios de selección bajo la declaración PRISMA. Los principales hallazgos indican que el uso de técnicas de big data y aprendizaje automático puede mejorar significativamente las estrategias de marketing y ventas digitales. Estas técnicas permiten una mejor segmentación de clientes, personalización de contenido y predicción del comportamiento de compra. En conclusión, aunque el Big Data y el aprendizaje automático tienen un gran potencial para optimizar las estrategias de marketing y ventas digitales estratégicas, se requieren más investigaciones empíricas para validar los hallazgos. La presente revisión permite tener una visión integral del estado actual y las tendencias en esta área de conocimiento emergente.

**Palabras clave**—Marketing, Big Data, Analytics, Algorithm, Plataformas Digitales

## I. INTRODUCCIÓN

En el entorno empresarial actual, las empresas de marketing se ven confrontadas con desafíos considerables, tales como una competencia intensa, la creciente complejidad del mercado y los rápidos avances tecnológicos. En respuesta a estas condiciones, las organizaciones buscan constantemente estrategias para obtener una ventaja competitiva, y en este sentido, el uso estratégico de big data y analítica predictiva se contempla como una herramienta esencial para tomar decisiones informadas. Esta situación ha impulsado la adopción creciente de tecnologías de big data en el ámbito del marketing, donde las empresas buscan no solo comprender en profundidad las dinámicas del mercado, sino también anticipar el comportamiento de los clientes para ajustar y optimizar sus estrategias de ventas [1, 2]. A pesar de la creciente popularidad de estas prácticas, se destaca que el conocimiento en este ámbito aún está en una fase inicial [3]. En este contexto, se genera la necesidad de llevar a cabo una Revisión Sistemática de la Literatura (RSL) enfocada en las ventas estratégicas en plataformas digitales. Esta revisión busca identificar las mejores prácticas, tendencias emergentes y áreas sin explorar

del marketing digital. Por ello, el objetivo de esta RSL es examinar y analizar de manera crítica la literatura académica del uso de técnicas de big data en el marketing de plataformas digitales, centrándose específicamente en la obtención de ventas estratégicas, encontrando metodologías, enfoques y herramientas empleadas en este contexto, ofreciendo una visión comprensiva de las prácticas actuales y, posiblemente, identificando áreas donde la investigación futura podría aportar nuevos conocimientos y avances significativos. La búsqueda bibliográfica se realizó en las principales bases de datos de literatura académica, tomando como principal a Scopus se incluirán artículos publicados en inglés desde el año 2020 hasta la fecha ya que para iniciar esta revisión se tomó como referencia la revisión [1]. Se identificarán las principales áreas de consenso y desacuerdo en la literatura, y se ofrecerán recomendaciones para futuras investigaciones.

El presente documento se organiza de la siguiente manera: En la sección II se presenta la metodología empleada para seleccionar la literatura que se ajuste a la problemática del estudio. En la sección III se presentan los resultados, sintetizando ordenadamente los hallazgos encontrados en las diferentes áreas de interés. Esto permite tener una visión integral de los aspectos más relevantes con relación al tema central de la revisión. Posteriormente, en la sección IV se discuten los principales hallazgos, limitaciones del estudio realizado y recomendaciones para futuros estudios. Finalmente, en la sección V se exponen las conclusiones del presente estudio.

## II. METODOLOGÍA

### A. Estrategia de búsqueda

La estrategia de búsqueda que se realizó en esta investigación fue basada en el método PICOC, la cual proviene de su acrónimo en inglés (*Problem, Intervention, Comparison, Output, Context*), donde se delimitaron las componentes respectivas; siendo el Marketing el Problema a abordar, Big Data la intervención a realizar, las Ventas estratégicas fueron los Resultados y finalmente, Plataformas Digitales siendo el contexto en el cual se realizó la búsqueda. Note que, para realizar la búsqueda de literatura no se consideró la componente Comparación, reduciendo así al método PIOC. Posteriormente, siguiendo esta metodología, se identificaron las palabras clave en inglés por cada componente PICOC, las cuales se muestran en la tabla I.

TABLA I  
COMPONENTES PIOC Y PALABRAS CLAVE

<b>Población</b>	Marketing	Sell, Retail, Merchandise, Trade, Promot
<b>Intervención</b>	Big Data	Big data analysis, Statistical Analysis, Metadata, Algorithm, Data management
<b>Resultado</b>	Ventas Estrategicas	Performance, Achievement, Analysis of results, Product reception, Harvest
<b>Contexto</b>	Plataformas Digitales	Website, App site, Site

Teniendo en consideración las componentes PIOC, se formula la siguiente pregunta de revisión: ¿Qué técnicas de big data se usan en el marketing de plataformas digitales para la obtención de ventas estratégicas?. Así mismo, esta pregunta se desglosa en las siguientes sub-preguntas: ¿Cuál es la intención del uso de big data?, ¿Con que canales de ventas se pueden lograr las ventas estratégicas?, ¿Cuáles son las estrategias utilizadas en plataformas digitales? y ¿De qué manera se podría potenciar el marketing?

Para dar respuesta a estas preguntas se buscó literatura científica en dos bases de datos, Scopus y Web of Science. Siendo la ecuación de búsqueda de la forma: (sell OR retail OR merchandise OR trade OR promot) AND (“Big data analysis” OR “statistical analysis” OR metadata OR algorithm OR “data management”) AND (performance OR achievement OR “analysis of results” OR “product reception” OR harvest) AND (website OR “app site” OR site).

### B. Criterios de Selección

Para delimitar la muestra de literatura a revisar se aplicaron los criterios de selección de documentos.

Criterios de inclusión:

- CI1: Artículos que tengan información sobre marketing usando big data.
- CI2: Artículos que tengan información sobre ventas estratégicas en plataformas digitales.
- CI3: Artículos enfocados al marketing digital.

Criterios de exclusión:

- CE1: Artículos redactados en un idioma distinto al inglés.
- CE2: Documentos con fechas anteriores al 2020.
- CE3: Artículos que no se alineen con el tema de investigación.

### C. Descripción del proceso de selección PRISMA

La lógica del sistema de selección PRISMA (*Preferred Reporting Items for Systematic reviews and Meta-Analyses*) comienza con la identificación de la investigación relevante a través de una búsqueda exhaustiva y sistemática en bases de datos electrónicos para encontrar artículos. Se establecen criterios de inclusión y exclusión para seleccionar los estudios que se incorporarán a la revisión sistemática. Estos criterios se aplican de manera consistente y transparente para garantizar la objetividad de la búsqueda [4].

A partir de la búsqueda de literatura en las bases de datos Scopus y WoS, se reportó un total de 764 artículos, luego se realizó la eliminación de 147 documentos duplicados,

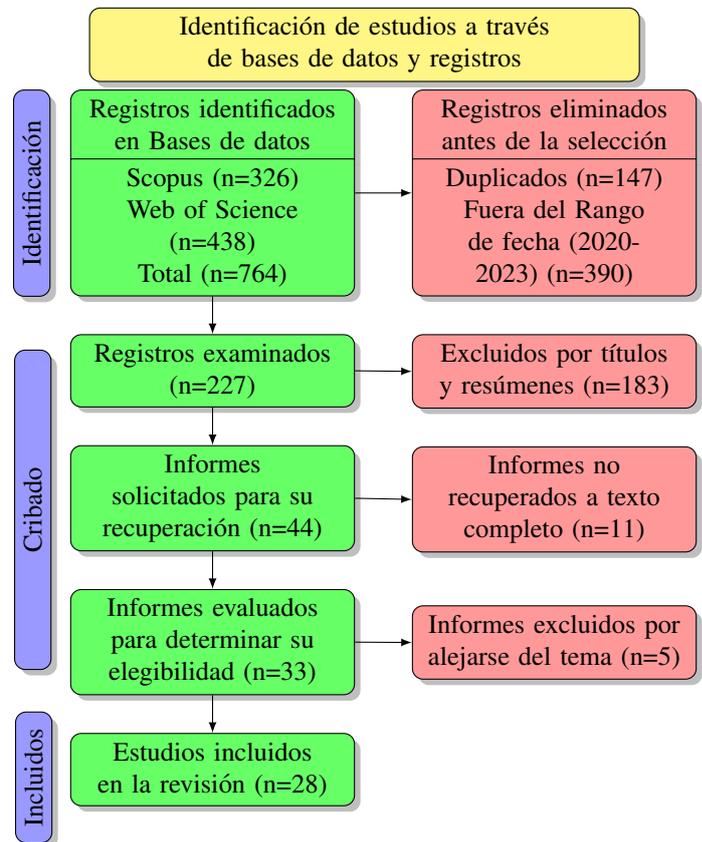


Fig. 1: Diagrama de flujo de PRISMA

quedando un total de 617, luego se realizó un filtro de año de la publicación considerando sólo los artículos desde el 2020 hasta el 2023, dando como resultado 390 documentos excluidos de la revisión por lo que quedaron 227 documentos, entre esos artículos se excluyeron 183 ya que no cumplían con el tema de revisión, quedando así 44 de los cuales 11 no fueron recuperados a texto completo. Se logró recuperar 33 artículos a texto completo, entre ellos 5 fueron excluidos luego de ser revisados a texto completo y finalmente, la muestra total que se revisará incluye a 28 artículos. Un resumen de este proceso de selección se puede observar a partir del diagrama de flujo PRISMA mostrado en la figura 1.

## III. RESULTADOS

### A. Definición del Big Data

El término “Big Data” se refiere a la gran cantidad de volúmenes de datos recopilados, adquiridos de múltiples accesos de información y que requieren de técnicas de análisis especializadas [1, 3], también existen los que se enfocan en datos de búsquedas web y en redes sociales de los usuarios, que pueden analizarse en profundidad para obtener información detallada sobre los usuarios en línea, donde se encuentran nombres, género, ubicación, incluso datos privados como el número de seguro social, datos de salud, imágenes, entre otros [5, 6, 7, 8, 9]. Se enfoca en el uso de técnicas de aprendizaje automático y procesamiento de lenguaje natural

para analizar sentimientos, emociones y opiniones mediante reseñas de productos [10, 11, 12]. Los cuales pueden ser utilizados en sistemas de recomendación en tiempo real que pueden manejar grandes cantidades de datos de usuarios e ítems. Esto se hace para optimizar estrategias de ofertas y así atraer clientes potenciales en entornos de comercio electrónico [13, 14, 15, 16]. Otra de las características que tiene el Big Data es que también sirve para clasificar datos de operaciones de ventas de productos, análisis de casos, predecir valores para recomendaciones en plataformas de e-commerce y distintos pedidos en línea, minimizando errores con ayuda de algoritmos inteligentes [17, 18, 19, 20, 21, 22, 23, 24].

### B. Área de aplicación

Siguiendo con el análisis se encontraron distintos artículos que tenían como principal idea el uso de Big Data, sin embargo, cada uno de ellos se centró en temas diferentes, pero con la misma relevancia en cuanto al análisis del estudio de revisión, como se puede apreciar en la tabla II, donde se hace mención al área de aplicación y sus objetivos. Algunos de estos artículos tienen como principal objetivo mantener la privacidad de los datos de las personas para evitar ataques de privacidad comerciales y no comerciales [5] una de estas propone un nuevo protocolo llamado PaOSLo (*Profile Aware ObScure Logging*) para preservar la privacidad de los usuarios en las búsquedas web, ocultando su perfil de búsqueda real ante los motores de búsqueda [8]. Así mismo, existen otros artículos que están enfocados en clasificar las reseñas de clientes en plataformas de comercio electrónico, redes sociales como Twitter y otra fuentes [3, 9, 12, 19] estudiando los sentimientos que generan distintas noticias o titulares que puedan influir en el comportamiento del usuario [10] se propone utilizar algoritmos metaheurísticos de aprendizaje automático a partir de los datos históricos de transacciones de comercio electrónico para ayudar a tomar decisiones y así atraer más clientes potenciales a través de estrategias de publicidad y de oferta apropiadas y también detectar similitudes entre productos en Marketplace basándose en títulos e imágenes de productos para maximizar los ingresos [6, 10, 13, 15, 25] mejorando las recomendaciones en plataformas de comercio electrónico que tienen múltiples módulos o secciones (ej. ofertas del día, ventas flash, etc.) [23, 26], creando ofertas y promociones personalizadas para sus clientes a través de marcos de Big Data que les permitan manejar grandes volúmenes de ventas con modelos eficientes [14]. También se encontró la aplicación de minería de datos con la intención de explorar la adopción de tecnologías de inteligencia artificial, dando soporte a decisiones de mantenimiento y clasificación eficiente de enormes cantidades de datos, para mejorar la gestión de relaciones con clientes y aumentar los beneficios económicos de las empresas que proveen servicios para negocios en línea y procesos de negocio [7, 18, 21, 22]. Por último, también mejora la eficiencia de los procesos de recolección, producción, distribución y entrega de pedidos considerando la llegada dinámica de pedidos durante el día y fecha de entrega específica e investigar

TABLA II  
ÁREA DE APLICACIÓN

Área de aplicación	Objetivo
Privacidad	Proteger la privacidad de los datos personales
Clasificación de reseñas	Identificar el sentimiento o la opinión de los usuarios
Análisis de sentimientos	Detectar el impacto emocional de las noticias o titulares
Toma de decisiones	Ayudar a tomar decisiones estratégicas
Mejora de recomendaciones	Personalizar las recomendaciones a los usuarios
Minería de datos	Explorar la adopción de tecnologías de inteligencia artificial
Eficiencia de procesos	Mejorar la eficiencia de los procesos empresariales

compensaciones entre capacidad de respuesta, riesgo y costo [24, 27, 28].

### C. Área de enfoque

La preocupación crítica de la privacidad en las OSN (*Online Social Networks*) es una de las áreas de enfoque que debe ser contemplada en el uso del Big Data. El enfoque metodológico, abarca desde el pre-procesamiento de datos hasta técnicas avanzadas de clustering y medidas para garantizar la privacidad de datos [5, 8], existen otros enfoques como el siguiente punto de las metodologías utilizadas que se centra en el análisis de 3 modelos NHPP (*NonHomogeneous Poisson Process*) jerárquicos para la modelación de interacciones de usuarios NHPP [17] incorporando aspectos contextuales y metadatos. También se destaca el uso del algoritmo AdaBoost para la clasificación, la extracción de características de texto e imágenes [25] y la evaluación exhaustiva del rendimiento a través de experimentos numéricos [24]. Este enfoque multidimensional muestra la complejidad de las interacciones de usuarios y puede contribuir al avance en el campo de la modelación del comportamiento en línea para la planificación y programación continua en el tiempo, luego se tiene una metodología integral que emplea algoritmos de aprendizaje por refuerzo, en particular TD3 (*Twin Delayed Deep Deterministic Policy Gradient*) [15], para entrenar políticas óptimas de generación de factores de oferta, segmentadas por franjas horarias. Además, se incorporan técnicas de procesamiento de lenguaje natural, como *Bag of Words* y TF-IDF (*Term frequency – Inverse document frequency*) [10] para la representación de textos. La perspectiva se amplía al utilizar modelos de machine learning, como Naive Bayes y XGBoost, en conjunto con modelos de deep learning, tales como *Convolutional Neural Networks* (CNN) y *Long Short-Term Memory* (LSTM), con el fin de anticipar emociones. Una contribución adicional del estudio es la implementación de modelos DSSM (*Deep Structured Semantic Models*) para calcular la similitud semántica entre las consultas de usuarios y las descripciones de productos, facilitando la identificación de la intención de compra [11]. Además, se destaca el desarrollo de un sistema web con backend en Python Django y frontend en Angular [22]; estas metodologías destacan por su enfoque multidisciplinario y su aplicación práctica en el desarrollo de

TABLA III  
METODOLOGÍA UTILIZADA

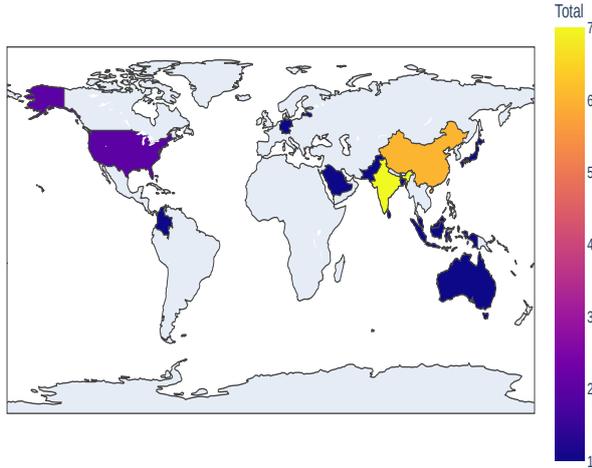


Fig. 2: Distribución de publicaciones por país

un sistema web innovador. Esta parte de las metodologías utilizadas aborda una variedad de enfoques y técnicas en el ámbito del aprendizaje automático supervisado y la aplicación de IoT (*Internet of Things*) en aplicaciones empresariales. Inicialmente, se evalúan algoritmos generales para medir su confiabilidad, haciendo hincapié en el uso de búsquedas de palabras y métricas como la exposición del perfil de usuario [12, 13] en donde se utilizan técnicas de análisis de sentimientos, técnicas de regresión lineal y técnicas de clasificación [9]. Posteriormente, se presenta una propuesta de aplicaciones empresariales de IoT que utiliza técnicas de clustering global y local para categorizar datos eficientemente, con el objetivo de mejorar la retención de clientes y la satisfacción en el sector [1, 21, 23]. Además, se resalta la importancia del tema de investigación que se enfoca en mejorar la personalización del marketing digital mediante el uso de inteligencia artificial (IA) [7, 29]. Se menciona el desarrollo de un modelo de programación lineal entera mixta para la planificación y programación continua en el tiempo [27]. Asimismo, se propone un nuevo estimador, sin sesgo, de la función de pérdida ideal para entrenar modelos de factorización matricial [26]. En términos de revisión sistemática de literatura, este trabajo abarca diversas áreas, desde la evaluación de algoritmos y métricas en el aprendizaje supervisado hasta la aplicación de IOT en el contexto empresarial y la personalización del marketing digital. La inclusión de modelos matemáticos multiobjetivo y el uso de ejemplos numéricos para demostrar la eficiencia de los modelos fortalecen la contribución del trabajo a la literatura científica [28]. En resumen, este estudio ofrece una perspectiva integral que combina la teoría y la aplicación práctica en múltiples campos de investigación [23, 29].

Área de enfoque	Metodología
Privacidad en OSN	Preprocesamiento de datos, clustering avanzado, medidas de privacidad
Modelación de interacciones de usuarios	Análisis de modelos NHPP jerárquicos, AdaBoost, evaluación exhaustiva
Generación de factores de oferta	Aprendizaje por refuerzo, TD3, procesamiento de lenguaje natural, Bag of Words, TF-IDF, Naive Bayes, XGBoost, CNN, LSTM, DSSM
Aprendizaje automático supervisado y aplicación de IOT	Evaluación de algoritmos, análisis de sentimientos, regresión lineal, clasificación, clustering global
Predicción del comportamiento del mercado de valores	Datos de ventas de Kaggle, análisis de características cualitativas y cuantitativas de las redes sociales, regresión lineal, random forest
Minería de datos de marketing	Aprendizaje automático supervisado, máquinas de soporte vectorial (SVM)
Técnicas generales	Web scraping, limpieza y transformación de datos, diseño e implementación de modelos de machine learning

Además se utilizaron datos de ventas conjuntos de datos de Kaggle analizando características cualitativas y cuantitativas de las redes sociales y otras fuentes para predecir el comportamiento del mercado de valores en donde entrenaron dos modelos de aprendizaje automático como Regresión Lineal y Random Forest [14]; además, se utilizan técnicas de aprendizaje automático supervisado, específicamente Máquinas de Soporte Vectorial (SVM, por sus siglas en inglés) para la minería de datos de marketing [6, 18] por último, están la técnicas como Web Scraping para obtener datos, limpieza y transformación de datos, diseño e implementación de modelos de Machine Learning como (árboles de decisión, Random Forest, Regresión Lineal y Redes Neuronales) [19], como se puede apreciar en la tabla III. Estos enfoques de revisión fueron hechos en distintos países, 7 de ellos en India, 6 en China, 2 en Estados Unidos e Irán y 1 en cada 1 de otros países como se muestra en la figura 2.

#### D. Herramientas utilizadas para la medición

Existen diferentes herramientas que se utilizaron para este análisis y cada una de ellas enfocada y aplicada a un área de enfoque distinto como se ve en la tabla IV. Prevención y privacidad: LECC (Privacidad diferencial local mediante mecanismo exponencial), PSO-GA (Algoritmo genético de optimización de enjambre de partículas), PFC (Medios C difusos que preservan la privacidad), DOA (Matriz ajena diferencial) [5]. Procesamiento de Texto y Aprendizaje Automático: Clasificador ingenuo de Bayes (NBC), Máquina de vectores de soporte (SVM), TF-IDF (Frecuencia de texto y Frecuencia de documento inversa), TensorFlow/Keras [12, 30] Optimización Estadística y Computacional: MLE (estimación de máxima verosimilitud), LSE (estimación de mínimos cuadrados), AG (Algoritmo Genético), SA (recocido simulado), PSO (optimización de enjambre de partículas) [17]. Agrupación de

TABLA IV  
ÁREA Y HERRAMIENTAS

Área	Herramienta
Prevención y privacidad	LECC, PSO-GA, PFC, DOA
Procesamiento de texto y aprendizaje automático	NBC, SVM, TF-IDF, Tensor-Flow/Keras
Optimización estadística y computacional	MLE, LSE, AG, SA, PSO
Agrupación de perfiles de usuario	K-means, K-medias multivariante
Comparativa de metaheurísticas	ALO, DA, GOA, HHO, MFO, MVO, SCA, SSA, WOA, GWO
Big data y aprendizaje automático	Apache Spark con MLlib, Regresión lineal, Bosque aleatorio
Procesamiento de datos a gran escala	Análisis de sentimiento, regresión lineal, clasificación, análisis de redes sociales, análisis de big data
Análisis de sentimiento y clasificación	VADER Lexicon, métricas de clasificación
Métricas de rendimiento	Coefficiente de determinación (R <sup>2</sup> ), error cuadrático medio (RMSE), coeficiente de variación (CV)
Aprendizaje automático en Python	Aprendizaje automático en Python
Métricas de evaluación de modelos	Pérdida de registro, DCG (ganancia acumulada descontada), Recordar MAP (precisión media)

Perfiles de Usuario: Algoritmo de clustering K-means implementado en la herramienta Weka para agrupar perfiles de usuario según similitud y Algoritmo de agrupamiento multivariante K-medias para generación de patrones de compra para predecir intereses de los clientes y generar recomendaciones personalizadas [1, 8] Comparativa de Metaheurísticas: ALO (Optimización de León Artificial), DA (Algoritmo de evolución diferencial), GOA (Algoritmo de optimización de saltamontes), HHO (optimización de Harris Hawks), MFO (optimización de búsqueda de alimento para monos), MVO (Optimizador multiverso), SCA (Algoritmo seno coseno), SSA (Algoritmo de enjambre de salpas), WOA (Algoritmo de optimización de ballenas), GWO (Optimizador del lobo gris) [13]. Big Data y Aprendizaje Automático: Apache Spark con MLlib, Regresión Lineal, Bosque aleatorio (precisión) [14] Publicidad Digital y Subastas: Estrategia propuesta evaluada con conjunto de datos de subastas de impresiones publicitarias [15] Procesamiento de Datos a Gran Escala: Análisis de sentimiento, Regresión Lineal, Clasificación, Análisis de Redes Sociales, Análisis de Big Data, estas herramientas ayudan a procesar grandes cantidades de datos de diferentes fuentes y extraer información relevante para tomar decisiones comerciales informadas [9]. Análisis de sentimiento y clasificación: VADER Lexicon para puntajes de sentimiento, métricas de clasificación: precisión, recuperación, puntuación F1, Matrices de confusión. Se utiliza el léxico VADER para evaluar sentimientos en titulares de noticias, midiendo precisión, recuerdo y F1-score, tanto en macro como micro promedios. Se destaca el uso de F1-score para evaluar la identificación precisa de productos relevantes, complementado con matrices de confusión para una comprensión detallada de la clasificación del modelo. Este enfoque integral proporciona una evaluación exhaustiva de la capacidad del modelo en el análisis efectivo de sentimientos en titulares [6, 10, 11] Métricas de Rendimiento: Coeficiente de Determinación (R<sup>2</sup>), Error Cuadrático Medio (RMSE), Coeficiente de variación (CV) [19]. Aprendizaje automático en Python: Aprendizaje-scikit. Se utiliza la biblioteca de aprendizaje automático scikit-learn en Python para el análisis [21]. Competición en Kaggle: Precisión de modelos (resultados de más del 90 % de precisión) [22]. Revisión Teórica de Modelos: AdaBoost Árboles de Decisión Bosque aleatorio [25]. Métricas de Evaluación de Modelos: Pérdida de registro, DCG (ganancia acumulada descontada), Recordar MAP (precisión media) [26] Métricas de Evaluación de Modelos: Precisión, Ganancia acumulada descontada normalizada (nDCG) para medir rendimiento [23].

#### E. Variables de medición para las ventas

Para esta sección se encontraron distintas variables que entran en juego para ser aplicadas en la revisión, primero se tiene el análisis de fallas de sistemas, el cual analiza el modelo con 4 conjuntos de datos de fallas de sistemas reales, 3 de Big Data y 1 de un sistema de tamaño medio [17], luego, se tienen los datos del cliente y estrategias publicitarias en las cuales se recolectaron 4 parámetros/características de los clientes: edad del cliente, grupo de edad del cliente, puntaje de

compra del cliente e ingreso anual del cliente [21]. También se mide el desempeño de Rankings de recolección de datos [23] con el objetivo de maximizar el número de clientes potenciales atraídos por la estrategia de publicidad [13]. Por otro lado, están los algoritmos de rendimiento de aprendizaje automático en general en donde se evalúa el rendimiento en términos de precisión de agrupamiento, generación de recomendaciones y retención de clientes [1, 29], con esto se puede medir la predicción del comportamiento que dan otras variables de medición como las de predicción de compras y ventas con atributos independientes de búsqueda por frecuencia, métricas dependiendo del número de clics o predecir el comportamiento del mercado [10, 12, 14, 15]. Por último, se tiene el análisis datos de ventas pasadas como costos de transporte, costos por retrasos, número de pedidos recolectados y entregados, tiempo de recolecta, entre otros [24], los cuales sirven para generar predicciones identificando la relevancia entre la consulta y el producto [11, 22] y la demanda de productos en los mercados, precios mínimos y máximos de venta [27, 7] con títulos e imágenes de los productos [25] como se muestra en la tabla V.

#### F. Plataformas utilizadas para el análisis

Para el manejo y ejecución de puesta en marcha se abordaron diversas plataformas digitales como Facebook y Yelp [5], así como también, plataformas como Alibaba, Amazon, Yahoo Finance y Kaggle [10, 12, 23]. También ejecutaron la simulación mediante Java con sockets [8], utilizando Apache Spark como marco de Big Data para entrenar modelos [12]. La implementación incluyó el uso de Python para el entrenamiento y evaluación del modelo, con redes neuronales para el actor y el crítico [15]. La recopilación de datos abarcó diversas fuentes sociales, como *Moneycontrol*, *IIFL*, *Economic*

TABLA V  
VARIABLES DE VENTAS

Análisis de fallas de sistema	Evalúa el rendimiento del modelo con conjuntos de datos de fallas de sistemas reales
Datos del cliente y estrategias publicitarias	Recolecta datos de los clientes con el objetivo de maximizar el número de clientes potenciales atraídos por estrategias de publicidad
Algoritmos de rendimiento de aprendizaje automático	Evalúa el rendimiento de los algoritmos de aprendizaje automático en términos de precisión de agrupamiento, generación de recomendaciones y retención de clientes
Predicción de compras y ventas	Utiliza atributos independientes de búsqueda por frecuencia, métricas dependiendo del número de clics o predecir el comportamiento del mercado
Análisis de datos de ventas pasadas	Utiliza datos de ventas pasadas para generar predicciones, identificar la relevancia entre la consulta y el producto

*Times*, *Business Standard*, *Reuters* y *LiveMint* para predecir el mercado de valores [9], así como también, emplearon SVM con el paquete Liblinear como clasificador, y realizaron pruebas del algoritmo con datos de Taobao [16, 23]. El proceso también involucró herramientas como Dexi.io para web scraping, Python para limpieza y análisis de datos [19] y tecnologías web como Django, Angular.js y MySQL [22]. Además, el algoritmo fue probado en Twitter [3], comercio electrónico y Home Depot [11]. Para la implementación en Marketplace [25] se recurrió a MATLAB, resolviendo modelos mediante GAMS y CPLEX. Finalmente, la resolución del modelo matemático se llevó a cabo con el software GAMS [24, 27, 28]. Este enfoque integral abarca diversas plataformas y técnicas, proporcionando una perspectiva amplia y detallada en el campo de la predicción del mercado y el análisis de datos.

#### G. Resultados de las estrategias de ventas

Existen los resultados por interpretación en donde se proponen modelos integrales de predicción y medición como el modelo de SVM para agilizar la predicción de la calidad de una venta calculado en reseñas [12, 18], enfocándose en maximizar los clics bajo diversos presupuestos en el conjunto de datos de prueba [15]. El sistema desarrollado tiene el potencial de personalizar la experiencia de compra en línea y las recomendaciones de productos según el género predicho de los clientes [6]. Aunque no se presentan resultados concretos, se destaca el uso del modelo SVM para mejorar estrategias de marketing y ventas [15]. El sistema demostró eficacia en la predicción de demanda, precios y desempeño de sucursales y empleados [22]. Además, se evidencia que el enfoque propuesto mejora la retención de clientes mediante recomendaciones personalizadas según sus intereses [1]. El algoritmo, basado en algoritmos genéticos, se muestra eficiente en la agrupación de pedidos, superando a métodos heurísticos en términos de costos y productividad [24]. Además, se destaca la determinación de puntos óptimos de venta en los mercados, considerando la caída de precios para maximizar las ganancias en la cadena de suministro [27]. También, están los resultados que sí pudieron ser medibles en cuanto a los resultados que arrojaron en cada uno de los estudios con

cada una de las herramientas con que fueron ejecutadas, se concluye que el optimizador PSO con MAD destaca como el más efectivo para el modelo híbrido con Big Data [17]. El algoritmo MFO se destaca al lograr el mejor rendimiento para atraer clientes potenciales [13] En el ámbito de Spark MLlib, los modelos entrenados exhiben una mayor precisión en la predicción de precios y ventas futuras en comparación con aquellos entrenados sin Spark [14]. El modelo propuesto WL-DSSM supera a otros, como DSSM, en la identificación precisa de productos relevantes (medido con F1-score), mejorando la experiencia del usuario y potencialmente impulsando las ventas esto permite mejorar la experiencia de usuario y potencialmente las ventas al conectar mejor las consultas con los productos buscados [11]. Para la predicción del valor de tasación, el Random Forest se destaca con un R2 promedio superior a 0.9 [19]. La eficiente aplicación de clustering y clasificación en centros comerciales puede aumentar los ingresos generales, permitiendo la categorización de clientes en clusters y el envío de ofertas personalizadas [21]. El algoritmo propuesto supera a otros métodos de aprendizaje por refuerzo multiagente en diversas métricas de evaluación [23]. Finalmente, el modelo AdaBoost logra una precisión destacada del 91.81 % en la detección de similitud de productos. Estas conclusiones refuerzan la eficacia y versatilidad de diversos enfoques y algoritmos en diferentes contextos de análisis y predicción [25].

#### H. Enfoque del marketing

El estudio se enfoca en la capacidad técnica para identificar la intención de compra de los usuarios mediante técnicas de procesamiento de lenguaje natural y aprendizaje automático [11]; esto permite la personalización y recomendación de contenido a los usuarios en tiempo real para plataformas online [30], con el objetivo de mejorar la retención de clientes minoristas a través de estrategias de marketing one-to-one [1]. Los resultados del estudio podrían aplicarse para mejorar las recomendaciones de productos similares a los usuarios y para ayudar a los vendedores a categorizar nuevos productos [23, 25]. En resumen, el estudio se enfoca en la comodidad del comercio electrónico con clasificador automático de reseñas de clientes y en la mejora de las recomendaciones personalizadas para los usuarios [12]. El enfoque es técnico, sobre modelos de confiabilidad de software [17] y está centrado en la privacidad de búsquedas en la web [8]. Está enfocado en la mejora de algoritmos de recomendación [26], aprendizaje automático supervisado [29], desde la perspectiva del anunciante. Busca maximizar clics dentro del presupuesto disponible [15] para predecir valores de tasación inmobiliaria utilizando técnicas de inteligencia artificial [19] optimizando operaciones logísticas y operaciones de la cadena de suministro [24, 27]. En síntesis, en la tabla VI se puede ver el enfoque de marketing del estudio.

TABLA VI  
ENFOQUE DE MARKETING

Enfoque	Objetivo
Personalización de la experiencia de compra	Mejorar la satisfacción del cliente y aumentar las ventas
Optimización de la cadena de suministro	Reducir costos, mejorar la eficiencia y aumentar la rentabilidad
Predicción de ventas	Anticipar la demanda y tomar decisiones estratégicas
Atracción de clientes potenciales	Generar interés en los productos o servicios

#### IV. DISCUSIÓN

Los principales hallazgos del estudio indican que el uso de big data y técnicas de aprendizaje automático [29] puede mejorar significativamente las estrategias de marketing y ventas en plataformas digitales [5]. Específicamente, permiten una mejor segmentación de clientes [10], personalización de contenido y predicción de comportamiento de compra [12], estos hallazgos concuerdan con otros estudios previos que han encontrado beneficios tangibles del Big Data para el marketing [1], como mayor efectividad en campañas y mejor experiencia de usuario. Sin embargo, la investigación en esta área es aún emergente dada la continua evolución tecnológica.

Un resultado relevante es que algoritmos como MFO, Spark MLlib y Random Forest mostraron un alto rendimiento en tareas de predicción y personalización [31]. Esto destaca el potencial de distintas técnicas de aprendizaje automático en este contexto [29]. Sin embargo, existe cierta limitación del estudio ya que se basa principalmente en la revisión de literatura académica, por lo que se requieren más investigaciones empíricas que validen los hallazgos en entornos reales. Para futuros estudios podrían enfocarse en aplicaciones sectoriales específicas del Big Data para marketing [7], evaluando métricas concretas de ventas y satisfacción del cliente en plataformas digitales [21]. Asimismo, se podría analizar la integración de distintas técnicas para potenciar los resultados. La revisión sistemática brinda una visión global del estado del conocimiento y destaca el gran potencial de Big Data [17] y aprendizaje automático [11] para impulsar el marketing y las ventas estratégicas en el comercio digital [31, 14]. Sin embargo, para validar y refinar estos hallazgos se necesitan más investigaciones en entornos reales, así como también, la combinación de estos hallazgos con las implicaciones éticas para comprender mejor el tipo de regulaciones que brinde un equilibrio entre la innovación y la privacidad de los usuarios. En ese sentido, se espera que la automatización y la personalización sean partes claves de estas nuevas tendencias en el marketing para ventas en plataformas digitales.

#### V. CONCLUSIÓN

Según la revisión sistemática realizada, se concluye que el uso de técnicas de big data y aprendizaje automático, como algoritmos de agrupamiento, clasificación y predicción, pueden mejorar significativamente las estrategias de marketing y ventas estratégicas en plataformas digitales. Los principales hallazgos indican que permiten una mejor segmentación

de clientes, personalización de contenido y predicción del comportamiento de compra. Sin embargo, se requieren más investigaciones empíricas para validar estos resultados.

La revisión contribuye a la literatura existente al proveer una visión integral del estado actual del conocimiento en esta área emergente. Las limitaciones incluyen el enfoque teórico y la falta de validación en entornos reales. Futuras investigaciones podrían centrarse en aplicaciones sectoriales específicas, evaluando métricas concretas de ventas y satisfacción del cliente.

Se recomienda analizar la integración de distintas técnicas de Big Data y aprendizaje automático para potenciar los resultados en el marketing digital. Asimismo, se sugiere explorar el impacto ético y social de la aplicación de estas tecnologías en las estrategias comerciales digitales. Es necesario desarrollar guías y políticas que promuevan un uso responsable, transparente y justo de los datos y algoritmos predictivos para proteger los intereses de los consumidores.

#### REFERENCIAS

- [1] A. Miklosik and N. Evans, "Impact of big data and machine learning on digital transformation in marketing: A literature review," 2020.
- [2] A. ROSÁRIO, L. B. MONIZ, and R. CRUZ, "Data science applied to marketing: A literature review," 2021.
- [3] M. Anshari, M. N. Almunawar, S. A. Lim, and A. Al-Mudimigh, "Customer relationship management and big data enabled: Personalization & customization of services," 2019.
- [4] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, D. Moher, J. J. Yepes-Nuñez, G. Urrútia, M. Romero-García, and S. Alonso-Fernández, "Declaración prisma 2020: una guía actualizada para la publicación de revisiones sistemáticas," *Revista Española de Cardiología*, vol. 74, 2021.
- [5] S. A. Alanazi, "Robust sentimental class prediction based on cryptocurrency-related tweets using tetrad of feature selection techniques in combination with filtered classifier," *Applied Sciences (Switzerland)*, vol. 12, 2022.
- [6] M. R. Kumar, J. Venkatesh, and A. M. Z. Rahman, "Data mining and machine learning in retail business: developing efficiencies for better customer retention," *Journal of Ambient Intelligence and Humanized Computing*, 2021.
- [7] R. Gangarde, A. Sharma, and A. Pawar, "Enhanced clustering based on privacy preservation to ensure k-anonymity, t-closeness, l-diversity, and balanced privacy utility," *Computers, Materials and Continua*, vol. 75, 2023.
- [8] M. M. Khan, M. G. Sohrab, and M. A. Yousuf, "Customer gender prediction system on hierarchical e-commerce data," *Beni-Suef University Journal of Basic and Applied Sciences*, vol. 9, 2020.

- [9] A. Nikolajeva, "Use of artificial intelligence technologies in e-commerce and business processes," 2021.
- [10] M. Ullah, R. U. Khan, I. U. Khan, N. Aslam, S. S. Aljameel, M. I. U. Haq, and M. A. Islam, "Profile aware obscure logging (paoslo): A web search privacy-preserving protocol to mitigate digital traces," *Security and Communication Networks*, vol. 2022, 2022.
- [11] P. Mehta, S. Pandya, and K. Kotecha, "Harvesting social media sentiment analysis to enhance stock market prediction using deep learning," *PeerJ Computer Science*, vol. 7, 2021.
- [12] S. Kamal, S. Sharma, V. Kumar, H. Alshazly, H. S. Hussein, and T. Martinetz, "Trading stocks based on financial news using attention mechanism," *Mathematics*, vol. 10, 2022.
- [13] J. Ma, X. Guo, and X. Zhao, "Identifying purchase intention through deep learning: analyzing the q & d text of an e-commerce platform," *Annals of Operations Research*, 2022.
- [14] W. Zhao, "Classification of customer reviews on e-commerce platforms based on naive bayesian algorithm and support vector machine," vol. 1678, 2020.
- [15] R. Yazdani, M. J. Taghipourian, M. M. Pourpasha, and S. S. Hosseini, "Attracting potential customers in e-commerce environments: A comparative study of meta-heuristic algorithms," *Processes*, vol. 10, 2022.
- [16] M. J. Awan, M. S. M. Rahim, H. Nobanee, A. Yasin, O. I. Khalaf, and U. Ishfaq, "A big data approach to black friday sales," *Intelligent Automation and Soft Computing*, vol. 27, 2021.
- [17] M. Liu, L. Jiaying, Z. Hu, J. Liu, and X. Nie, "A dynamic bidding strategy based on model-free reinforcement learning in display advertising," *IEEE Access*, vol. 8, 2020.
- [18] Y. Gao and Y. Liu, "Construction and application of international commercial dispute resolution mechanism model," *Security and Communication Networks*, vol. 2022, 2022.
- [19] A. K. Prasad, M. D. Kumar, V. D. J. Macedo, B. R. Mohan, and N. A. Prasad, "Machine learning approach for prediction of the online user intention for a product purchase," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, 2023.
- [20] Y. Chang, "Research on marketing data mining model based on support vector machine algorithm," 2022.
- [21] N. Martinez, J. Tellez, J. Barrero, and L. Chaves, "Automatic method for the prediction of the commercial appraisal of a property in bogota city," 2021.
- [22] R. Guha, "Improving the performance of an artificial intelligence recommendation engine with deep learning neural nets," 2021.
- [23] Y. Alzahrani, J. Shen, and J. Yan, "Multi-layer efficient data classification methods for enterprise business applications," 2022.
- [24] D. P. Nilmini, D. Nawinna, and B. Attanayaka, "Data-driven business intelligence platform for smart retail stores," 2021.
- [25] X. He, B. An, Y. Li, H. Chen, R. Wang, X. Wang, R. Yu, X. Li, and Z. Wang, "Learning to collaborate in multi-module recommendation via multi-agent reinforcement learning without communication," 2020.
- [26] F. H. N. Shavaki and F. Jolai, "A rule-based heuristic algorithm for joint order batching and delivery planning of online retailers with multiple order pickers," *Applied Intelligence*, vol. 51, 2021.
- [27] M. A. Muhsin, D. R. Wijaya, E. Hernawati, and A. Wi-dayanti, "Adaboost algorithm for marketplace product similarity detection," 2022.
- [28] Y. Saito, S. Yaginuma, Y. Nishino, H. Sakata, and K. Nakata, "Unbiased recommender learning from missing-not-at-random implicit feedback," 2020.
- [29] W. Albrecht and M. Steinrucke, "Continuous-time scheduling of production, distribution and sales in photovoltaic supply chains with declining prices," *Flexible Services and Manufacturing Journal*, vol. 32, 2020.
- [30] M. Shekarian, S. V. R. Nooraie, and M. M. Parast, "An examination of the impact of flexibility and agility on mitigating supply chain disruptions," *International Journal of Production Economics*, vol. 220, 2020.
- [31] Y. Ma, B. M. Narayanaswamy, H. Lin, and H. Ding, "Temporal-contextual recommendation in real-time," 2020.