

INTEGRATION OF BIG DATA WITH MACHINE LEARNING FOR PREDICTIVE SALES ANALYSIS: A SYSTEMATIC REVIEW

Henry Jesús García Cancho¹; Vayrol Miguel Melgarejo Choque²; Nelson Jesús Campos Rosendo³; Jaime Arturo Bravo Ruiz⁴

^{1,2,3 y 4}Universidad Tecnológica del Perú, Lima, Perú.

u20232299@utp.edu.pe, u19212515@utp.edu.pe, c19357@utp.edu.pe, c27763@utp.edu.pe

Abstract– *This study examines the implementation of Machine Learning techniques in sales forecasting, highlighting the impact of Big Data integration in transforming business strategies. Large-scale companies like Amazon, Google, and Microsoft are leading this shift, using Machine Learning to enhance the accuracy and efficiency of trend forecasting, thereby strengthening their competitiveness. However, small and medium-sized enterprises (SMEs) face significant challenges in adopting these advanced technologies due to limitations in infrastructure and expertise, which restrict their ability to leverage predictive sales analysis. Predictive analysis allows for demand forecasting and process improvement, but SMEs face obstacles when trying to implement complex techniques such as Random Forest or neural networks, due to data complexity and the selection of relevant variables. This study evaluates the main Machine Learning techniques used in sales forecasting and the specific challenges SMEs face in this context. The methodology follows PRISMA guidelines and uses the PICO framework to organize the search in the Scopus database. The analysis addresses strategies applied by large corporations and the barriers SMEs face when attempting to implement similar technologies. Additionally, technological solutions that improve sales forecasting efficiency and accuracy are explored, overcoming obstacles such as scalability and data quality. This approach provides a comprehensive overview of the current state of research in sales forecasting, highlighting challenges and proposing strategies for SMEs to adopt Machine Learning and Big Data, helping to reduce the competitive gap with large companies.*

Keywords– *Big Data, Machine Learning, sales prediction, small and medium-sized enterprises (SMEs), predictive analytics*

INTEGRACIÓN DE BIG DATA CON MACHINE LEARNING PARA ANÁLISIS PREDICTIVO EN VENTAS: UNA REVISIÓN SISTEMÁTICA

Henry Jesús García Cancho¹, Vayrol Miguel Melgarejo Choque²; Nelson Jesús Campos Rosendo³; Jaime Arturo Bravo Ruiz⁴

^{1,2,3} y ⁴Universidad Tecnológica del Perú, Lima, Perú.

u20232299@utp.edu.pe, u19212515@utp.edu.pe, c19357@utp.edu.pe, c27763@utp.edu.pe

Resumen– Este estudio examina la implementación de técnicas Machine Learning en la predicción de ventas, destacando el impacto de integración Big Data en la transformación de estrategias comerciales. Empresas de gran escala como Amazon, Google y Microsoft lideran este cambio, utilizando Machine Learning para mejorar la precisión y eficiencia en la predicción de tendencias, lo que refuerza su competitividad. No obstante, las pequeñas y medianas empresas (SMEs) enfrentan dificultades significativas para adoptar estas tecnologías avanzadas, debido a limitaciones en infraestructura y experiencia, lo que limita su capacidad de aprovechar el análisis predictivo de ventas. El análisis predictivo permite anticipar la demanda y mejorar procesos, pero las SMEs encuentran obstáculos al intentar implementar técnicas complejas como Random Forest o redes neuronales, debido a la complejidad de los datos y la elección de variables relevantes. Este estudio evalúa las principales técnicas de Machine Learning utilizadas en la predicción de ventas desafiando particulares que enfrentan las SMEs. La metodología sigue lineamientos de PRISMA, utiliza el enfoque PICO para organizar la búsqueda en la base de datos Scopus. El análisis aborda estrategias aplicadas por grandes corporaciones y las barreras que las SMEs enfrentan al intentar implementar tecnologías similares. Además, se exploran soluciones tecnológicas que mejoran la eficiencia y precisión en la predicción de ventas, superando obstáculos como escalabilidad y calidad de datos. Este enfoque proporciona una visión global actual de la investigación en predicción de ventas, subrayando los desafíos y proponiendo estrategias para que las SMEs adopten Machine Learning y Big Data, ayudando a reducir la brecha competitiva con grandes empresas.

Palabras clave– Big Data, Machine Learning, predicción de ventas, pequeñas y medianas empresas (SMEs), análisis predictivo

I. INTRODUCCIÓN

El análisis predictivo ha surgido como un recurso clave para las organizaciones que buscan maximizar la eficiencia y efectividad en sus procesos para estrategias de ventas, permitiéndoles anticiparse a las demandas del mercado y mejorar sus decisiones comerciales. El crecimiento de tecnologías como el Big Data y las técnicas de Machine Learning (ML) ha permitido que grandes volúmenes de datos sean transformados en información valiosa para predecir tendencias y comportamientos futuros. Empresas de renombre mundial como Amazon, Google y Microsoft han liderado este cambio, demostrando cómo la

incorporación de la inteligencia artificial en sus operaciones ha incrementado la precisión y eficiencia de sus estrategias de ventas, potenciando su competitividad [1]. Sin embargo, a pesar de estos avances, las pequeñas y medianas empresas (SMEs), especialmente en economías emergentes, enfrentan serios obstáculos para adoptar estas tecnologías de vanguardia. La falta de infraestructura adecuada, junto con los retos en el manejo y la evaluación de grandes cantidades de datos, junto con la carencia de experiencia en la implementación de modelos predictivos representan obstáculos que dificultan a estas empresas aprovechar las ventajas del análisis predictivo. [2]. A pesar de que algunos mercados emergentes han comenzado a superar estas barreras, la adopción generalizada de Big Data y ML en las SMEs sigue siendo un desafío pendiente.

La incapacidad de muchas SMEs para implementar eficazmente el Big Data Predictive Analytics (BDPA) crea una brecha significativa en su capacidad para competir con grandes corporaciones. Sectores como el comercio electrónico y las ventas minoristas se ven especialmente afectados, ya que la falta de precisión, escalabilidad y puntualidad en los modelos predictivos limita la capacidad de estas empresas para ajustarse de manera ágil a las condiciones del mercado y optimizar sus procesos comerciales [5].

Además, el uso de algoritmos de ML presenta retos adicionales para las SMEs, como la correcta selección de características y el procesamiento de datos, lo que restringe el potencial de los modelos predictivos avanzados. Estas limitaciones no solo afectan la optimización de estrategias comerciales, sino que también generan una barrera significativa para la innovación y el crecimiento en mercados emergentes [4].

A pesar de que el uso de ML y Big Data en el análisis predictivo ha demostrado ser efectivo en grandes empresas, la investigación que aborda los desafíos específicos de las SMEs en la adopción de estas tecnologías es escasa [4]. Este estudio busca llenar ese vacío al explorar las principales técnicas de ML aplicadas al análisis predictivo de ventas, y evaluar los desafíos y oportunidades que enfrentan las SMEs, en particular en términos de privacidad, escalabilidad y precisión [3].

El propósito de esta revisión sistemática es ofrecer una comprensión más profunda de cómo estas tecnologías pueden ser implementadas con éxito en las SMEs, y proponer estrategias que faciliten su adopción, contribuyendo así a cerrar la brecha competitiva con grandes corporaciones. Esta distribución de publicaciones destaca el interés global en la aplicación de tecnologías avanzadas para optimizar la

precisión y eficiencia en el análisis de datos comerciales.

II. METODOLOGÍA

La metodología de esta revisión sistemática se enfoca en analizar técnicas de ML en el contexto del análisis predictivo de ventas, haciendo hincapié en la integración de Big Data. La investigación se fundamenta en los lineamientos definidos por los Elementos Preferidos para Informes de Revisiones Sistemáticas y Metaanálisis (PRISMA 2020).[6], adoptados por sus claras pautas para garantizar revisiones sistemáticas de alta calidad y transparencia. La pregunta de investigación fue formulada utilizando el método PICO, lo que permitió estructurar una búsqueda precisa y enfocada en la literatura científica relevante, la búsqueda de los artículos se realizó exclusivamente en Scopus el 4 de octubre sin imponer un límite de fecha, para incluir la mayor cantidad de estudios relevantes, la eliminación de duplicados se realizó de forma manual aplicando filtros de búsqueda directamente en Scopus, para el proceso de selección, se revisaron los estudios según criterios de inclusión y exclusión basados en la relevancia para el análisis predictivo de ventas, el uso de Big Data y técnicas de ML.

Una de las ventajas de esta metodología es que proporciona una visión integral de las investigaciones existentes sobre la predicción de ventas, facilitando la identificación de vacíos en el conocimiento y áreas que necesitan mayor exploración. Además, al fundamentarse en estudios empíricos previos, se logra una mayor confianza en los resultados obtenidos [8].

A. Formulación de preguntas de investigación

Para realizar la búsqueda bibliográfica, se utilizó la base de datos Scopus, conocida por su amplio repositorio de artículos científicos revisados por pares, trabajos de investigación y conferencias académicas, Scopus no solo permite acceder a una gran variedad de documentos, sino que también ofrece herramientas avanzadas para evaluar el impacto y relevancia de los estudios mediante análisis de citas y métricas. La estrategia de búsqueda se estructuró siguiendo el método PICO, que facilita la formulación de preguntas claras y específicas, esenciales para llevar a cabo una revisión sistemática detallada. La pregunta de revisión planteada es: “¿Qué técnicas de ML con Big Data se utilizan para medir la eficiencia, precisión y escalabilidad en predicciones de ventas?”. Esta pregunta busca explorar la aplicación de algoritmos de aprendizaje automático en el contexto del análisis predictivo, enfocado en mejorar la capacidad de anticipar el comportamiento de las ventas y evaluar el rendimiento de dichas técnicas.

La formulación de preguntas bajo esta metodología presenta varias ventajas. Primero, ayuda a definir claramente la población objetivo del estudio, permitiendo obtener información más precisa y relevante. Segundo, facilita la identificación de las intervenciones que se están

estudiando, lo que permite comparar diferentes enfoques y determinar cuál es más efectivo. En este estudio, la pregunta general y las subpreguntas fueron formuladas de la siguiente manera:

Pregunta general

Q1: ¿Qué técnicas de ML con Big Data se utilizan para medir la eficiencia, precisión y escalabilidad en predicciones de ventas?

Preguntas específicas

RQ1: ¿Cómo impacta el uso de algoritmos de ML en la predicción de ventas en comparación con los métodos tradicionales?

RQ2: ¿Cuáles son las principales técnicas de ML utilizadas en la predicción de ventas con grandes volúmenes de datos?

RQ3: ¿De qué manera las técnicas de ML como redes neuronales y Random Forest, mejoran la precisión en la predicción de ventas?

RQ4: ¿Cómo influye la precisión y el volumen de los datos en la eficacia de los modelos de ML aplicados a la predicción de ventas?

B. Especificación de palabras clave

Para garantizar una búsqueda exhaustiva y relevante de la literatura científica, se formuló una ecuación de búsqueda basada en los términos clave identificados en la estrategia PICO. Estos términos fueron seleccionados cuidadosamente para abarcar tanto las técnicas de ML como su aplicación en el análisis predictivo de ventas, incluyendo intervenciones con Big Data y la comparación con métodos tradicionales [4]. La ecuación de búsqueda fue formulada y aplicada en la base de datos Scopus, utilizando una combinación de operadores booleanos y términos clave que abarcan diversas áreas relacionadas con la predicción de ventas y la eficiencia de los algoritmos de aprendizaje automático [3]. Con el fin de garantizar que la búsqueda fuese exhaustiva y sistemática, se empleó un conjunto de palabras clave en inglés, dado que estas permiten agrupar resultados de diferentes idiomas. A continuación, se presenta la Matriz PICO, que contiene los términos clave, tanto en español como en inglés, utilizados en el proceso de búsqueda:

TABLA I
MATRIZ PICO

Metodología	Descripción	Palabras clave en español	Palabras clave en inglés
P	Problema/Población	Predicción de ventas	Sales prediction, sale predictions, prediction algorithms, prediction methods
I	Intervención	Técnicas de ML con Big data	Neuronal networks, decision trees, forecasting, Random Forest, Boosting, ML

		KNN, Support Vector Machine	Máquinas de vectores soporte	XGBoost, KNN, SVM
O	Resultados	Eficiencia, precisión y escalabilidad en predicción de ventas	Precisión en predicciones, tendencias de ventas	Sales trends, Market trends, Prediction accuracy, Forecast accuracy

C. *Formulación de ecuaciones y motores de búsqueda*

Para encontrar la información adecuada y específica, se formuló una ecuación de búsqueda utilizando conectores lógicos (OR, AND) y operadores booleanos, el motor de búsqueda utilizado fue Scopus, y la ecuación de búsqueda fue la siguiente:

(TITLE-ABS-KEY("Sales prediction" OR "Prediction algorithms" OR "Enterprise Sales")) AND (TITLE-ABS-KEY("Neural networks" OR "Decision trees" OR "Forecasting" OR "Random Forest" OR "Gradient Boosting" OR "Big data" OR "ML")) AND (TITLE-ABS-KEY("Regression analysis" OR "Deep learning" OR "XGBoost" OR "KNN" OR "Support Vector Machine")) AND (TITLE-ABS-KEY("Sales trends" OR "Prediction accuracy" OR "Scalability" OR "Demand forecasting"))

D. *Criterios de inclusión y exclusión*

Con el fin de realizar la búsqueda de manera eficaz, se definieron los siguientes criterios de inclusión y exclusión: Criterios de inclusión (en concordancia con el enfoque PICO)

TABLA II
CRITERIOS DE INCLUSIÓN

CI1	Uso de Big Data en Predicción de Ventas
CI2	Aplicación de algoritmos de ML
CI3	Artículos de comparación en Técnicas tradicionales y modernas en la predicción de ventas.
CI4	Estudios que midan la precisión, eficiencia y escalabilidad

TABLA III
CRITERIOS DE EXCLUSIÓN

CE1	Estudios que limita o no representen Big Data
CE2	Artículos que no estén escritos en inglés
CE3	Términos y palabras claves no relacionadas con la predicción de ventas con el uso de ML
CE4	Artículos publicados fuera del rango temporal seleccionado

E. *Proceso de selección y PRISMA*

Para asegurar una revisión sistemática rigurosa, se aplicó el enfoque PRISMA. Este enfoque facilitó el proceso de identificación, filtrado y selección de los estudios más relevantes, siguiendo de forma estricta, se aplicaron los criterios de inclusión y exclusión previamente [9]. Inicialmente, se identificaron 499 registros en la base de datos Scopus. Durante la etapa de identificación, se removieron 98 registros duplicados, 57 registros considerados no elegibles mediante herramientas de automatización, y otros 220 registros fueron descartados por diversas razones, lo que resultó en un total de 124 registros para la siguiente fase de cribado, en la fase de cribado, se examinaron los 124 registros resultantes, de los cuales 47 estudios fueron excluidos por no cumplir con los criterios de relevancia, dejando 77 estudios para ser evaluados más a fondo. Todos los 77 informes fueron revisados para su posible recuperación y no hubo ninguno que no pudiera ser recuperado para evaluación. A continuación, se llevó a cabo una evaluación de elegibilidad sobre 77 estudios seleccionados. Durante esta fase, se aplicaron criterios adicionales de exclusión, basados en el año de publicación (solo se consideraron estudios publicados entre 2020 y 2024), el idioma (solo se aceptaron estudios en inglés) y la relevancia del contenido (focalizado en Big Data y técnicas de ML aplicadas a la predicción de ventas, evaluando precisión, eficiencia y escalabilidad), tras esta evaluación, se excluyeron los estudios que no cumplían con estos requisitos, resultando en 49 estudios finales incluidos en la revisión.

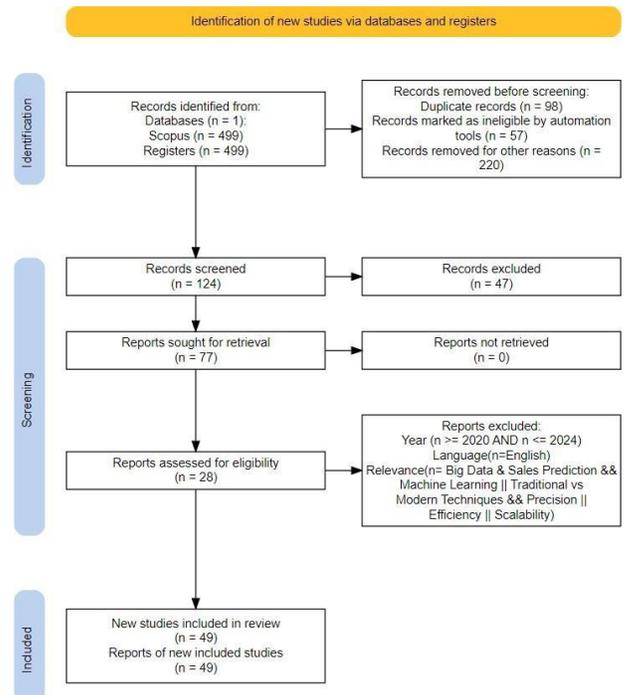


Fig. 1 Diagrama prisma

III. RESULTADOS

Los resultados de la investigación fueron estructurados mediante dos enfoques analíticos fundamentales, con el objetivo de optimizar la estructuración de la información y hacerla más accesible para su comprensión.

- Datos bibliométricos: Incluyen información bibliográfica, generalmente de carácter cuantitativo.
- Datos académicos: Comprenden la información extraída de artículos, que pueden ser tanto como cualitativa como cuantitativa.

A. Resultados de Análisis Bibliométrico

1. Artículos clasificados según el enfoque de la investigación

Como corolario, se identificó que varios estudios exploran las aplicaciones más recientes y desarrollos en la integración de Big Data con técnicas de ML para mejorar el análisis predictivo en ventas. En la Fig. 2 se muestra la estructuración y categorización de los datos recopilados, destacando las áreas temáticas más relevantes y los enfoques predominantes en la literatura actual.

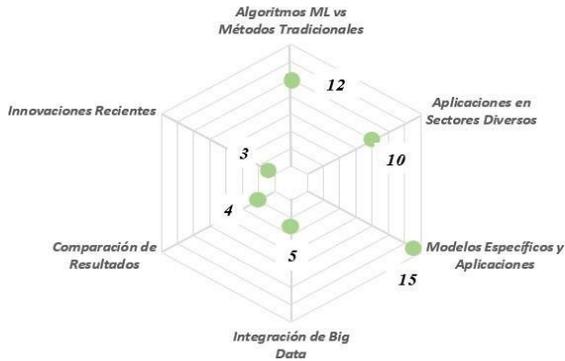


Fig. 2 Cantidad de artículos según tema de investigación.

Los artículos seleccionados se organizaron según temas comunes, identificándose que 12 de ellos abordan comparativas entre algoritmos de Machine Learning y métodos tradicionales, destacando las fortalezas y limitaciones de cada enfoque en el contexto del análisis predictivo de ventas. Asimismo, 10 artículos describen aplicaciones en sectores diversos, mostrando cómo se implementan estas técnicas en industrias como el retail y el transporte. Por otro lado, 15 artículos se enfocan en modelos específicos y sus aplicaciones, detallando el uso de algoritmos avanzados como redes neuronales y técnicas híbridas. 5 artículos abordan la integración de Big Data con herramientas de ML para optimizar la calidad del análisis de datos y su impacto en la toma de decisiones. Asimismo, 4 artículos comparaciones entre los resultados generados mediante técnicas de ML frente a métodos tradicionales, y 3 artículos investigan las innovaciones recientes y tendencias en el área.

2. Artículos según año de publicación

Las investigaciones recopiladas también se organizaron de acuerdo con el año de publicación para facilitar un análisis más claro y comprensible. Este criterio abarca desde el 2020 hasta el 2024, ya que se observó un crecimiento constante en la cantidad de estudios publicados durante este período. El año con la mayor cantidad de publicaciones fue el 2023, con un total de 15 artículos, lo que evidencia un auge en las investigaciones relacionadas con el tema. En comparación, el 2022 y el 2021 registraron 12 y 11 artículos respectivamente, mientras que el 2024, a pesar de ser un año reciente, ya cuenta con 7 artículos. El 2020 presentó un menor número de publicaciones, con 6 artículos, como se muestra en la Fig3.

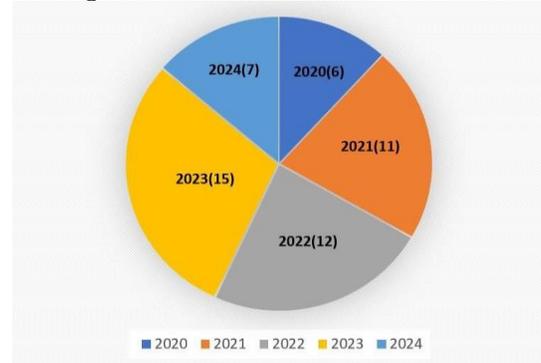


Fig. 3 Distribución de los artículos en función del año de publicación.

3. Distribución geográfica de los artículos según su país de origen

En el transcurso de la elaboración de la revisión sistemática de la literatura, se ilustran los países implicados en la Fig. 4. Se destaca a China como el país con mayor cantidad de artículos seleccionados, con un total de 23 publicaciones, reflejando su crecimiento significativo en investigaciones relacionadas con la Integración de Big Data y técnicas de ML para el análisis predictivo en ventas. Este interés se alinea con las estrategias tecnológicas del país, que buscan posicionarlo como líder en el uso de Big Data para optimizar procesos comerciales y aumentar la competitividad en el mercado global.

Otros países que han contribuido de manera notable incluyen a Corea del Sur con 4 publicaciones, Estados Unidos con 3, y Australia e India con 2 publicaciones cada uno. Además, países como Canadá, Croacia, República Checa, Egipto, Alemania, Grecia, Italia, Japón, Pakistán, España, Suecia, Suiza, y Reino Unido aportaron en menor proporción, con entre 1 y 2 publicaciones cada uno. Esta distribución de publicaciones destaca el interés global en la aplicación de tecnologías avanzadas para optimizar la precisión y eficiencia en el análisis de datos comerciales.



Fig. 4 Cantidad de artículos según página de publicación

B. Resultados Académicos

1. Distribución de artículos por revista

Un enfoque adicional para organizar la información recopilada fue analizar la distribución de los artículos según las revistas en las que fueron publicados. Destacó la revista IEEE Access con un total de 15 artículos, evidenciando su importancia en el área. Otras publicaciones relevantes incluyeron Computational Intelligence and Neuroscience, Computers, Materials and Continua, Environmental Research Letters, IEEE Transactions on Learning Technologies, Intelligent Systems in Accounting, Finance and Management, ISPRS International Journal of Geo-Information, Scientific Programming y Systems, con 2 artículos cada una, lo que refleja un interés sostenido en estos tópicos.

En cuanto a las demás publicaciones, cada una de ellas contribuyó con un solo artículo. Algunas de las revistas que participaron fueron, entre otras, 3 Biotech, Advances in Multimedia, Buildings, Forecasting, Frontiers in Energy Research, y varias más, como IEEE Robotics and Automation Letters, IEEE Transactions on Instrumentation and Measurement, IEEE Transactions on Parallel and Distributed Systems, y otras destacadas en el ámbito académico como International Journal of Advanced Computer Science and Applications, International Journal of Data Warehousing and Mining, Journal of Marine Science and Engineering, y muchas más en áreas como matemáticas, ingeniería y tecnología.

Estos datos, representados gráficamente en la Fig. 5, subrayan la amplitud de las fuentes académicas que contribuyen al estudio del análisis predictivo y el uso de técnicas de Machine Learning.

Se identificó que la implementación de técnicas de ML y Big Data en ventas está aumentando en diversas aplicaciones. En el contexto del análisis predictivo en ventas, se observa que algoritmos avanzados como Random Forest, redes neuronales y XGBoost es comúnmente empleado por su habilidad para procesar grandes cantidades de datos. y reconocer patrones complejos [55].

Estas técnicas permiten capturar interacciones no lineales y mejorar la precisión en comparación con los métodos tradicionales, optimizando así los resultados en sectores comerciales dinámicos.

Por otra parte, los estudios también exploran aplicaciones prácticas en distintas industrias, donde el uso de algoritmos se combina con grandes conjuntos de datos estructurados.

Este enfoque permite una predicción más precisa y adaptable a las fluctuaciones del mercado.

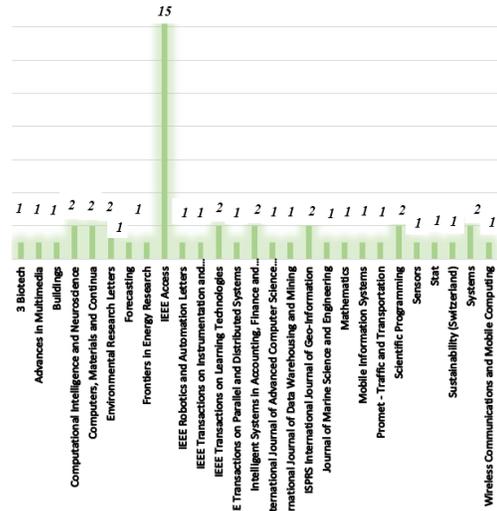


Fig. 5 Cantidad de artículos según página de publicación

2. ¿Qué problemas se han identificado en la predicción de ventas con métodos tradicionales?

En la figura 6 se organizan los problemas principales encontrados en los métodos tradicionales de predicción de ventas, tales como la baja precisión al manejar datos complejos, lo que reduce la capacidad de ajustarse a las variaciones dinámicas en la demanda del mercado [12], [19].

Otra limitación señalada está relacionada con la dependencia de suposiciones lineales [14], [21], esto restringe la capacidad de realizar análisis de datos a gran escala y dificulta la identificación de patrones no lineales que representen de manera más precisa el comportamiento del mercado [23], [26], [40].

Además, los métodos tradicionales no cuentan con la capacidad para actualizarse de manera automática, lo cual reduce su efectividad en la predicción de ventas en entornos de alta volatilidad y grandes volúmenes de datos [17], [19].



Fig. 6 Problemas en la predicción de ventas con métodos tradicionales

3. ¿Qué ventajas ofrece el uso de algoritmos de Machine Learning en la predicción de ventas?

En la figura 7 se destacan los principales objetivos del uso de algoritmos de ML en la predicción de ventas, tales como la capacidad de realizar análisis precisos en tiempo real, lo cual permite una respuesta rápida ante cambios en las tendencias del mercado [11], [15].

Otro de los objetivos es mejorar la precisión al manejar grandes volúmenes de datos y patrones complejos en series temporales [18], [22]. Además, se busca optimizar las estrategias comerciales mediante la identificación de patrones ocultos y el ajuste dinámico de predicciones basadas en datos históricos y actuales [24], [27].



Fig. 7 Objetivos del uso de ML

4. ¿Qué beneficios trae la implementación de algoritmos de ML en la predicción de ventas?

La implementación de algoritmos de ML en la predicción de ventas ofrece beneficios significativos, como la mejora en la exactitud y la habilidad para gestionar grandes volúmenes y complejidad de datos, lo que posibilita una predicción más precisa y flexible [13], [20]. Otros beneficios incluyen la disminución de los costos operativos al optimizar los inventarios y anticiparse a cambios en la demanda, minimizando el desperdicio y el exceso de stock [14], [27]. Además, los algoritmos de ML facilitan la personalización de campañas de marketing y el ajuste dinámico de precios en función de las condiciones del mercado y el

comportamiento del cliente, lo cual incrementa la competitividad de la empresa y mejora la satisfacción del cliente [16], [19], [48], [50].

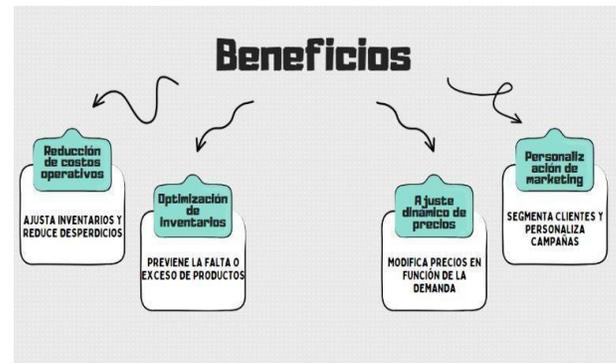


Fig.8 Beneficios de ML en la predicción de ventas.

5. ¿En qué áreas de las estrategias de ventas se pueden aplicar técnicas de ML?

Las técnicas de ML tienen aplicaciones en áreas clave de las estrategias de ventas, tales como la optimización de inventarios, que permite prever con mayor precisión la demanda de productos y ajustar el stock eficientemente para satisfacer las necesidades del mercado, también se aplican en la personalización de campañas de marketing, donde el ML permite segmentar clientes de forma precisa, ofreciendo productos y promociones relevantes para cada grupo objetivo [17], [21]. En el área de fijación de precios, los modelos de ML permiten ajustar los precios en tiempo real, maximizando la rentabilidad de la empresa y adaptándose continuamente a los cambios en la demanda del mercado [19], [23],[53].



Fig.9 Áreas de Aplicación de ML en Estrategias de Ventas

6. ¿Qué aspectos de calidad se mejoran en las predicciones de ventas al usar ML?

La aplicación de ML en la predicción de ventas mejora diversos aspectos de calidad, tales como la reducción en los márgenes de error, lo que aumenta la confiabilidad de los pronósticos y permite una mejor planificación estratégica [18], [24]. Además, la capacidad de actualización automática de los modelos asegura que la precisión se mantenga en entornos con cambios rápidos, favoreciendo una toma de decisiones más informada [21], [27]. Asimismo, los modelos

de ML mejoran la robustez y adaptabilidad al procesar diversas variables y la capacidad de adaptar las predicciones en tiempo real, lo cual permite a las empresas responder eficazmente a las fluctuaciones de la demanda [25], [28].

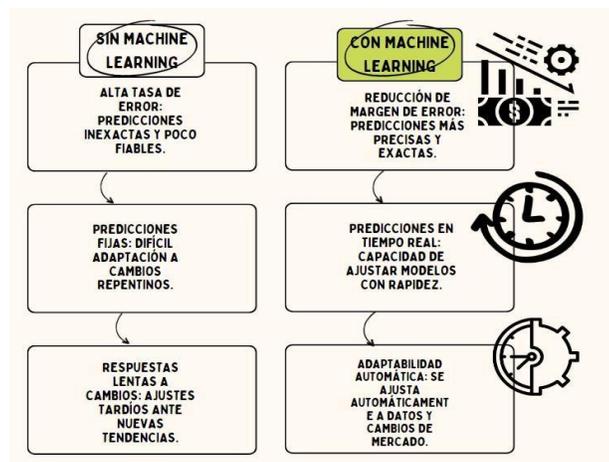


Fig.10 Beneficios de ML en la Precisión y Calidad de Predicciones de Ventas

RQ1: ¿Cómo impacta el uso de algoritmos de ML en la predicción de ventas en comparación con los métodos tradicionales?

En el estudio realizado, se identificaron investigaciones que comparan la eficacia de los algoritmos de ML en comparación con los enfoques tradicionales en la predicción de ventas. [53]. Algoritmos como XGBoost, SVM, Random Forest y redes neuronales han mostrado avances considerables en la precisión de las predicciones, gracias a su habilidad para procesar grandes cantidades de datos [39]. Además, los modelos de ML pueden ajustarse y optimizarse de manera más eficiente mediante la afinación de parámetros [31], lo que mejora su desempeño en comparación con los métodos tradicionales. Los hallazgos obtenidos coinciden con estos resultados, ya que se observó que el manejo de la complejidad de los datos subjetivos y objetivos es más eficaz con ML. Por ejemplo, algoritmos como XGBoost alcanzan un 82% de precisión en comparación con los métodos tradicionales. Del mismo modo, algoritmos como SVM han demostrado un incremento del 10.4% en la precisión. Además, el uso de técnicas como Random Forest y redes neuronales ha mostrado una mejora en la precisión del 5%, utilizando variables básicas [45],[38],[30],[31], capturando patrones complejos que los métodos tradicionales, como ARIMA o la regresión lineal, no logran modelar tan eficazmente en entornos dinámicos [11],[54].

Se destaca que la implementación de algoritmos de ML en la predicción de ventas no solo mejora la precisión, sino que también permite una mejor adaptación a factores externos como la estacionalidad o los cambios en el mercado [50],[16],[57]. Entre los beneficios más importantes se incluyen la optimización de inventarios, la habilidad para llevar a

cabo análisis predictivos más sólidos y una mayor eficacia en la manipulación y análisis de extensos conjuntos de datos procedentes de múltiples fuentes. Por lo tanto, estas técnicas superan a los métodos tradicionales en varias áreas, ya que estos últimos tienden a tener dificultades para capturar patrones complejos debido a su naturaleza lineal y presentan limitaciones en su precisión en comparación con las técnicas modernas [39],[49],[44],[12].

A pesar de estos avances, la precisión de los modelos predictivos puede verse influenciada por la complejidad del entorno y las características particulares de cada empresa. En este sentido, el ML proporciona soluciones adaptativas que permiten ajustes continuos a través del análisis de datos en tiempo real, lo que facilita la toma de decisiones más fundamentadas y orientadas a la estrategia dentro de las organizaciones. Sectores como el financiero se benefician significativamente al obtener predicciones más precisas sobre las fluctuaciones del mercado, lo que mejora la toma de decisiones en mercados minoristas, e-commerce, entre otros [13],[37],[26],[46],[50]. El tiempo de respuesta también es un factor clave en el rendimiento. Los algoritmos de ML se ajustan de manera más ágil a los cambios del mercado, en contraste con los enfoques tradicionales, que son más lentos y, por lo tanto, limitan su efectividad. Las técnicas de ML proporcionan beneficios importantes al mejorar la comprensión dinámica de los datos de manera más precisa, basándose en un análisis integral de la información disponible [18],[19],[17].

RQ2: ¿Cuáles son las principales técnicas de ML utilizadas en la predicción de ventas con grandes volúmenes de datos?

En la información recabada de la investigación, se han identificado diversas técnicas de ML que se utilizan para la predicción de ventas con grandes volúmenes de datos. Estas técnicas incluyen algoritmos con mecanismos de utilidad de compensaciones efectivas en el entorno de ventas, manejo de secuencias de datos de alta dimensionalidad y en entornos que requieren la ejecución de múltiples tareas concurrentes [20],[23],[33],[47]. Entre las más populares se encuentran algoritmos como XGBoost, Random Forest y Long Short-Term Memory (LSTM), que han demostrado ser eficaces para manejar la complejidad y la variabilidad de los datos de ventas [36],[55]. Dichas técnicas permiten detectar patrones ocultos en series temporales y analizar múltiples variables de manera simultánea.

Por ende, algunas técnicas se distinguen por su habilidad para gestionar volúmenes significativos de datos y atributos complejos, como es el caso de Random Forest [32], que ha demostrado ofrecer predicciones más precisas y confiables. Asimismo, otros métodos como la regresión basada en árboles de decisión y las redes neuronales profundas también se aplican en la predicción de ventas, mejorando la precisión y permitiendo su adaptación a entornos dinámicos que demandan grandes cantidades de datos [41],[21].

Los sectores que más se benefician del uso de estas técnicas incluyen la manufactura y la logística, donde la optimización de inventarios,

finanzas y estrategias de ventas es esencial [43],[15]. Además, en el ámbito del retail, la tecnología y el e-commerce, el uso de estas técnicas es altamente beneficioso debido a la necesidad de análisis predictivo y modelado del comportamiento del consumidor, impactando positivamente en la planificación y optimización de recursos [39],[42].

RQ3: ¿De qué manera las técnicas de ML, como redes neuronales y Random Forest, mejoran la precisión en la predicción de ventas?

En la actualidad, las técnicas de ML, como redes neuronales y Random Forest, mejoran la precisión en la predicción de ventas al captar patrones complejos y manejar múltiples variables, superando a los métodos tradicionales.

Estas técnicas capturan e identifican con mayor precisión que otros algoritmos [14],[38],[51]. Las redes neuronales destacan por su capacidad para capturar relaciones complejas y no lineales entre variables, lo que las hace más eficaces en la predicción de ventas en escenarios [15],[17],[52],[56]. En comparación con otras técnicas, Random Forest sobresale por su mayor precisión debido a su estructura de múltiples árboles, lo que permite mitigar el impacto de variables ruidosas o atípicas. Esto mejora su rendimiento respecto a algoritmos de regresión o árboles de decisión individuales, ya que reduce la varianza al combinar múltiples árboles. Además, Random Forest tolera mejor el ruido en los datos, lo que lo hace más efectivo que métodos como la regresión lineal en la predicción de ventas [54],[51],[49].

Por otro lado, factores como la carga de trabajo del agente, la precisión de predicciones anteriores y las decisiones de los clientes entre precio y precisión influyen en la exactitud de los modelos de predicción. Se resalta que factores que involucren una integración de características estructurales y dependencias a largo plazo afectan significativamente la precisión de los modelos predictivos al tener características de dependencia [22],[23],[24].

RQ4: ¿Cómo influye la calidad y cantidad de los datos en la efectividad de los modelos de ML aplicados a la predicción de ventas?

La exactitud y la disponibilidad de datos juegan un papel crucial en el desempeño de los modelos de aprendizaje automático para prever las ventas. Estos modelos requieren conjuntos de datos extensos, precisos y relevantes para identificar patrones significativos y generar predicciones exactas [38],[57]. Datos de baja calidad, como aquellos afectados por la estacionalidad, picos o ruido, pueden distorsionar la estructura esencial de un modelo predictivo de ventas, afectando su efectividad en diferentes sistemas de ML diseñados para mejorar la precisión en las predicciones [10],[25],[27]. Los datos incompletos o ruidosos afectan negativamente la precisión de las predicciones. Sin el uso de algoritmos avanzados y técnicas de optimización, sería difícil mitigar los problemas relacionados con el manejo adecuado de la calidad de los datos y la minimización de los errores son aspectos fundamentales para optimizar el rendimiento de

los modelos que impactan la precisión general del modelo [28],[29].

Un conjunto de datos correctamente estructurado es esencial para que los modelos de ML operen de manera eficiente. El tamaño del dataset afecta la capacidad de los modelos para generalizar y evitar el sobreajuste. Se considera tanto la universalidad como la distinción que permite que el marco UDM realice predicciones más precisas. De este modo, se optimiza la capacidad de las empresas para optimizar sus procesos de ventas y tomar decisiones informadas [34],[35].

DISCUSIÓN

La revisión sistemática de literatura evaluó 49 artículos que abordan la integración de Big Data con técnicas de ML para el análisis predictivo en ventas. Los resultados del análisis indican que la combinación de Big Data y ML puede tener un impacto positivo en las estrategias comerciales, contribuyendo a una mayor precisión en la predicción de ventas y optimización de recursos. La revisión identificó sinergias entre Big Data y ML, enfocándose en la optimización de la eficiencia y toma de decisiones. Estos dos elementos, que forman la base de análisis predictivo en ventas,

se alinean con los objetivos de optimizar la exactitud y la capacidad de expansión de las predicciones de ventas. Además, la revisión presenta evidencia de los beneficios de integrar Big Data y ML en el análisis predictivo de ventas. Por ejemplo, el estudio [5] se discuten los beneficios, debilidades y problemas abiertos de cada enfoque, así como los caminos futuros. Destaca la necesidad de optimizar métricas como puntualidad, precisión y escalabilidad en las aplicaciones de Big Data, lo que permitiría a las organizaciones avanzar del análisis retrospectivo al predictivo prospectivo. También, el estudio [19], muestra como resultado que los algoritmos de ML, como XGBoost, superan a los métodos tradicionales en la predicción de los niveles de llenado de los contenedores de basura al manejar eficazmente la variabilidad de los datos y proporcionar predicciones más precisas, logrando un 82% de precisión en comparación con el 59% del modelo de referencia. Además, en un último estudio nos menciona la integración de Random Forest en los Pymes, donde ofrece mayor precisión en las predicciones en comparación con métodos tradicionales donde alcanza una precisión del 92.01% lo cual es significativamente superior a la precisión obtenida por métodos como la media o KNN impute.

Los algoritmos de ML, como el Random Forest, suelen ofrecer mayor precisión en las predicciones en comparación con métodos tradicionales. En el caso descrito, el Random Forest alcanzó una precisión del 92.01%, lo cual es significativamente superior a la precisión obtenida por métodos como la media o KNN impute [36].

Por otro lado, la revisión también identificó algunas limitaciones como la accesibilidad y calidad de los datos en las investigaciones, así como la complejidad de integrar técnicas de ML en infraestructuras tecnológicas existentes. Sin embargo, los hallazgos sugieren que la integración de Big data con ML es una herramienta prometedora para mejorar las estrategias de ventas y perfeccionar la toma de decisiones dentro de las organizaciones.

V. CONCLUSIONES

Para concluir, esta investigación identificó que la integración de Machine Learning (ML) y Big Data es una estrategia esencial para incrementar la eficiencia, precisión y escalabilidad en la predicción de

ventas, especialmente en sectores como el retail, e-commerce y tecnología. La implementación de algoritmos avanzados, como redes neuronales, Random Forest, XGBoost y LSTM, permite no solo mejorar los procesos de toma de decisiones en las organizaciones, sino también responder con mayor rapidez a las condiciones cambiantes del mercado y gestionar grandes volúmenes de datos.

Además, estas técnicas ayudan a las empresas a optimizar sus inventarios y recursos, impulsando un mejor desempeño en distintos sectores económicos. Sin embargo, se encontraron barreras significativas, como la necesidad de datos de alta calidad, la complejidad de ajuste de los modelos y los altos requerimientos computacionales. Para abordar estos desafíos, *se están desarrollando enfoques más accesibles y escalables* que permitan la adopción efectiva de ML y Big Data en empresas de menor y mediana escala (pymes). En cuanto a sostenibilidad y escalabilidad, las innovaciones en ML han demostrado ser herramientas eficaces para la gestión óptima de recursos. Estas técnicas no solo mejoran los resultados financieros de las empresas, sino que también fomentan prácticas más sostenibles y precisas en comparación con los métodos tradicionales, cerrando la brecha competitiva entre las pymes y las grandes empresas. Como recomendaciones, para profundizar en el entendimiento de la aplicación de ML y Big Data en la predicción de ventas, se recomienda realizar búsquedas en bases de datos científicas como Scopus, Emerald Insight, IEEE Xplore, ScienceDirect y WoS. Estas fuentes ofrecen estudios relevantes sobre la eficiencia, precisión y escalabilidad de los modelos predictivos.

Se recomienda enfocar la investigación en sectores con grandes volúmenes de datos, como retail y tecnología, que emplean técnicas avanzadas como redes neuronales, Random Forest y XGBoost. También se sugiere adaptar estas tecnologías a otros sectores con procesos similares para acelerar la innovación. Finalmente, se propone fomentar la colaboración entre pymes, promoviendo el uso de ML y Big Data para impulsar el crecimiento económico sostenible y mejorar la competitividad, a través de estrategias que incluyan recursos accesibles, capacitación y herramientas tecnológicas adaptadas a sus necesidades.

REFERENCIAS BIBLIOGRAFICAS

- [1] A. Mantri and R. Mishra, "Empowering small businesses with the force of big data analytics and AI: A technological integration for enhanced business management", *Journal of High Technology Management Research*, vol. 221 no. 119795, 2023, doi: 10.1016/j.hitech.2023.100476
- [2] W. Yu, T. Dillon, F. Mostafa, W. Rahayu, Y. Liu, "A Global manufacturing big data ecosystem for fault detection in predictive maintenance", *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 183-192, 2020, doi: 10.1109/TII.2019.2915846.
- [3] M. L. D. Tewu, Suwarno, P. Lisdiono, R. Friska, and A. J. Pramono, "Enterprise risk management and supply chain management: The mediating role of competitive advantage and decision making in improving firms performance," *Uncertain Supply Chain Management*, vol. 12, no. 2, pp. 1131-1140, 2024, doi: 10.5267/j.uscm.2023.11.021.
- [4] M. A. Alaketu et al., "Comparative Analysis of Intrusion Detection Models using Big Data Analytics and ML Techniques," *International*

- Arab Journal of Information Technology*, vol. 21, no. 2, pp. 326-337, 2024, doi: 10.34028/21/2/14.
- [5] A. Jamarani, S. Haddadi, R. Sarvzadeh, M. Haghi Kashani, M. Akbari, and S. Moradi, "Big data and predictive analytics: A systematic review of applications," *Artificial Intelligence Review*, vol. 57, no. 7, 2024, doi: 10.1007/s10462-024-10811-5.
- [6] Akyazi T., Goti A., Oyarbide A., Alberdi E. y Bayon F., "A Guide for the Food Industry to Meet the Future Skills Requirements Emerging with Industry 4.0", *Foods*, vol. 9, n.º 4, p. 492, abril de 2020.. <https://doi.org/10.3390/foods9040492>
- [7] S. Ahmad, S. Wasim, S. Irfan, S. Gogoi, A. Srivastava y Z. Farheen, «Cualitativo v/s. Investigación Cuantitativa: una revisión resumida.» *Poblacion*, 2019.
- [8] Tashakkori A. y Teddlie C., *SAGE Handbook of Mixed Methods in Social & Behavioral Research*. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publ., Inc., 2010 <https://doi.org/10.4135/9781506335193>
- [9] Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis *Campbell Systematic Reviews*, 18, e1230. <https://doi.org/10.1002/cl2.1230>
- [10] Hirata E.; Matsuda T., "Forecasting Shanghai Container Freight Index: A Deep-Learning-Based Model Experiment", *Journal of Marine Science and Engineering*, Volume 10, n.º 5, p. 593, 2022, doi: 10.3390/jmse10050593
- [11] Shamsad H.; Ullah F.; Ullah A.; Kebande V.R.; Ullah S.; Al-Dhaqm A., "Forecasting and Trading of the Stable Cryptocurrencies With Machine Learning and Deep Learning Algorithms for Market Conditions", *IEEE Access*, Volume 11, 2023, doi: 10.1109/ACCESS.2023.3327440
- [12] Mu G.; Dai L.; Ju X.; Chen Y.; Huang X., "MS-IHHO-LSTM: Carbon Price Prediction Model of Multi-Source Data Based on Improved Swarm Intelligence Algorithm and Deep Learning Method", *IEEE Access*, Volume 12, 2024, doi: 10.1109/ACCESS.2024.3409822
- [13] Mu G.; Gao N.; Wang Y.; Dai L., "A Stock Price Prediction Model Based on Investor Sentiment and Optimized Deep Learning", *IEEE Access*, Volume 11, 2023, doi: 10.1109/ACCESS.2023.3278790
- [14] Lu S.; Wei W.; Zhu Z.; Liang Y.; Liu H., "Operation Risk Assessment of Hydroelectric Energy Storage Based on Data Visualization and Convolutional Neural Network", *Frontiers in Energy Research*, Vol. 9, 2022, pag. 827942, doi: 10.3389/fenrg.2021.827942
- [15] Li D.; Li X.; Gu F.; Pan Z.; Chen D.; Madden A., "A Universality-Distinction Mechanism-Based Multi-Step Sales Forecasting for Sales Prediction and Inventory Optimization", *Systems*, Vol. 11, n.º 6, 2023, pag. 311, doi: 10.3390/systems11060311
- [16] El Majzoub A.; Rabhi F.A.; Hussain W., "Evaluating interpretable machine learning predictions for cryptocurrencies", *Intelligent Systems in Accounting, Finance and Management*, Vol. 30, pag. 3, 2023, doi: 10.1002/isaf.1538
- [17] Bousono-Calzon C.; Bustarviejo-Munoz J.; Aceituno-Aceituno P.; Escudero-Garzas J.J., "On the Economic Significance of Stock Market Prediction and the No Free Lunch Theorem", *IEEE Access*, Vol. 7, 2019, doi: 10.1109/ACCESS.2019.2921092
- [18] Kang Y.; Ozdogan M.; Zhu X.; Ye Z.; Hain C.; Anderson M., "Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest", *Environmental Research Letters*, Vol. 15, n.º 6, 2020, doi: 10.1088/1748-9326/ab7df9
- [19] Sanjar K.; Bekhzod O.; Kim J.; Paul A.; Kim J., "Missing data imputation for geolocation-based price prediction using KNN-MCF method", *ISPRS International Journal of Geo-Information*, Vol. 9, n.º 4, pag. 227, 2020, doi: 10.3390/ijgi9040227
- [20] Li X.; Dong H., "A Transaction Trade-Off Utility Function Approach for Predicting the End-Price of Online Auctions in IoT", *Wireless Communications and Mobile Computing*, Vol. 2021, 2021, doi: 10.1155/2021/6656421
- [21] Sun H.; Yao Z.; Miao Q., "Design of Macroeconomic Growth Prediction Algorithm Based on Data Mining", *Mobile Information Systems*, Vol. 2021, 2021, doi: 10.1155/2021/2472373
- [22] Chen Z.; Hu J.; Min G.; Zomaya A.Y.; El-Ghazawi T., "Towards Accurate Prediction for High-Dimensional and Highly-Variable Cloud Workloads with Deep Learning", *IEEE Transactions on Parallel and Distributed Systems*, Vol. 31, n.º 4, 2020, doi: 10.1109/TPDS.2019.2953745
- [23] Volpe G.; Mangini A.M.; Fanti M.P., "A Deep Reinforcement Learning Approach for Competitive Task Assignment in Enterprise

- Blockchain", IEEE Access, Vol. 11, 2023, doi: 10.1109/ACCESS.2023.3276859
- [24] Meher P.K.; Satpathy S., "Improved recognition of splice sites in A. thaliana by incorporating secondary structure information into sequence- derived features: a computational study", 3 Biotech, Vol 11. n°. 11, pag. 484, 2021, doi: 10.1007/s13205-021-03036-8
- [25] S. Lin, X. Nong, J. Luo, and C. Wang, "A novel multi-model stacking ensemble learning method for metro traction energy prediction," *IEEE Access*, vol. 10, pp. 129231-129244, 2022, doi: 10.1109/ACCESS.2022.3228441.
- [26] C. Zhang, Y. Zhao, and H. Zhao, "A novel hybrid price prediction model for multimodal carbon emission trading market based on CEEMDAN algorithm and window-based XGBoost approach," *Mathematics*, vol. 10, no. 21, p. 4072, 2022, doi: 10.3390/math10214072.
- [27] K. P. Fourkiotis and A. Tsadiras, "Applying machine learning and statistical forecasting methods for enhancing pharmaceutical sales predictions," *Forecasting*, vol. 6, no. 1, pp. 170-186, 2024, doi: 10.3390/forecast610010.
- [28] P. Jiang, X. Zhou, T. Liu, X. Guo, D. Ma, C. Zhang, Y. Li, and S. Liu, "Prediction dynamics in cotton aphid using unmanned aerial vehicle multispectral images and vegetation indices," *IEEE Access*, vol. 11, pp. 5908-5918, 2023, doi: 10.1109/ACCESS.2023.3235912.
- [36] L. Hang, D. Liu, and F. Xie, "A hybrid model using PCA and BP neural network for time series prediction in Chinese stock market with TOPSIS analysis," *Sci. Prog.*, vol. 2023, Article ID 9963940, 2023, doi: 10.1155/2023/9963940.
- [37] Y. Zhao, Z. Wei, and J. Wen, "Prediction of soil heavy metal content based on deep reinforcement learning," *Sci. Prog.*, vol. 2022, Article ID 1476565, 2022, doi: 10.1155/2022/1476565.
- [38] A. El Majzoub, F. A. Rabhi, and W. Hussain, "Evaluating interpretable machine learning predictions for cryptocurrencies," *Intell. Syst. Account. Finance Manag.*, vol. 30, no. 3, pp. 137-149, 2023, doi: 10.1002/isaf.1538.
- [39] Q. Jiao, L. Bao, H. Bai, H. Niu, and C. Han, "SVM-BTS based trajectory identification and prediction method for civil rotorcraft UAVs," *IEEE Access*, vol. 11, pp. 137248-137263, 2023, doi: 10.1109/ACCESS.2023.3338727.
- [40] C. Bousono-Calzon, J. Bustarviejo-Munoz, P. Aceituno-Aceituno, and J. J. Escudero-Garzas, "On the economic significance of stock market prediction and the no free lunch theorem," *IEEE Access*, vol. 7, pp. 8731844-75177, 2019, doi: 10.1109/ACCESS.2019.2921092.
- [41] Y. Kang, M. Ozdogan, X. Zhu, Z. Ye, C. Hain, and M. Anderson, "Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest," *Environ. Res. Lett.*, vol. 15, no. 6, p. 064005, 2020, doi: 10.1088/1748-9326/ab7df9.
- [42] G. E. Atteia, H. A. Mengash, and N. A. Samee, "Evaluation of using parametric and non-parametric machine learning algorithms for Covid-19 forecasting," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 10, pp. 647-657, 2021, doi: 10.14569/IJACSA.2021.0121071.
- [43] D. R. Bickel, "Testing prediction algorithms as null hypotheses: Application to assessing the performance of deep neural networks," *Stat.*, vol. 9, no. 1, e270, 2020, doi: 10.1002/sta4.270.
- [44] K. Sanjar, O. Bekhzod, J. Kim, A. Paul, and J. Kim, "Missing data imputation for geolocation-based price prediction using KNN-MCF method," *ISPRS Int. J. Geo-Information*, vol. 9, no. 4, p. 227, 2020, doi: 10.3390/ijgi9040227.
- [45] H. Huang, L. Huang, R. Song, F. Jiao, and T. Ai, "Bus single-trip time prediction based on ensemble learning," *Comput. Intell. Neurosci.*, vol. 2022, Article ID 6831167, 2022, doi: 10.1155/2022/6831167.
- [46] M. Mravik, T. Vetrisevi, K. Venkatachalam, M. Sarac, N. Bacanin, and S. Adamovic, "Diabetes prediction algorithm using recursive ridge regression L2," *Comput. Mater. Continua*, vol. 71, no. 1, pp. 457-471, 2022, doi: 10.32604/cmc.2022.020687.
- [47] X. Ren, B. Yang, N. Luo, J. Li, Y. Li, T. Xue, and A. Yang, "The prediction of sinter drums strength using hybrid machine learning
- [29] J. Dai, "Improving random forest algorithm for university academic affairs management system platform construction," *Advances in Multimedia*, 2022, Article ID 8064844, doi: 10.1155/2022/8064844.
- [30] C. Imhof, I.-S. Comsa, M. Hlosta, B. Parsaeifard, I. Moser, and P. Bergamin, "Prediction of dilatatory behavior in eLearning: A comparison of multiple machine learning models," *IEEE Trans. Learn. Technol.*, vol. 16, no. 5, pp. 648-663, 2023, doi: 10.1109/TLT.2022.3221495.
- [31] B. Qiu and W. Fan, "Machine learning based short-term travel time prediction: Numerical results and comparative analyses," *Sustainability*, vol. 13, no. 13, p. 7454, 2021, doi: 10.3390/su13137454.
- [32] A. Vorkapić, R. Radonja, and S. Martinčić-Ipsić, "Predicting seagoing ship energy efficiency from the operational data," *Sensors*, vol. 21, no. 8, p. 2832, 2021, doi: 10.3390/s21082832.
- [33] D. Li, X. Li, F. Gu, Z. Pan, D. Chen, and A. Madden, "A universality-distinction mechanism-based multi-step sales forecasting for sales prediction and inventory optimization," *Systems*, vol. 11, no. 6, p. 311, 2023, doi: 10.3390/systems11060311.
- [34] Y. Yang, G. Liu, H. Zhang, Y. Zhang, and X. Yang, "Predicting the compressive strength of environmentally friendly concrete using multiple machine learning algorithms," *Buildings*, vol. 14, no. 1, p. 190, 2024, doi: 10.3390/buildings14010190.
- [35] N. Gao, Y. Hong, J. Chen, and C. Pang, "Regional expressway freight volume prediction algorithm based on meteorological information," *Promet - Traffic and Transportation*, vol. 35, no. 2, pp. 195-211, 2023, doi: 10.7307/ptt.v35i2.40. algorithms," *Comput. Intell. Neurosci.*, vol. 2022, Article ID 4790736, 2022, doi: 10.1155/2022/4790736.
- [48] H. Yang and X. Ren, "Design and development of a rural tourism marketing system using deep learning," *IEEE Access*, vol. 12, pp. 64795- 64806, 2024, doi: 10.1109/ACCESS.2024.3396081.
- [49] Q. Zhu, C. Wang, W. Jin, J. Ren, and X. Yu, "Deep transfer learning based on LSTM model for reservoir flood forecasting," *Int. J. Data Warehousing Mining*, vol. 20, no. 1, 2024, doi: 10.4018/IJDWM.338912.
- [50] S. De and P. Prabu, "A sampling-based stack framework for imbalanced learning in churn prediction," *IEEE Access*, vol. 10, pp. 68017-68028, 2022, doi: 10.1109/ACCESS.2022.3185227.
- [51] C. Imhof, I.-S. Comsa, M. Hlosta, B. Parsaeifard, I. Moser, and P. Bergamin, "Prediction of dilatatory behavior in eLearning: A comparison of multiple machine learning models," *IEEE Trans. Learn. Technol.*, vol. 16, no. 5, pp. 648-663, 2023, doi: 10.1109/TLT.2022.3221495.
- [52] A. Muminov, O. Sattarov, and D. Na, "Enhanced Bitcoin price direction forecasting with DQN," *IEEE Access*, vol. 12, pp. 29093-29112, 2024, doi: 10.1109/ACCESS.2024.3367719.
- [53] M. Adnan, M. I. Uddin, E. Khan, F. S. Alharithi, S. Amin, and A. A. Alzahrani, "Earliest possible global and local interpretation of students' performance in virtual learning environment by leveraging explainable AI," *IEEE Access*, vol. 10, pp. 129843-129864, 2022, doi: 10.1109/ACCESS.2022.3227072.
- [54] G. Xue, J. Song, X. Kong, Y. Pan, C. Qi, and H. Li, "Prediction of natural gas consumption for city-level DHS based on attention GRU: A case study for a northern Chinese city," *IEEE Access*, vol. 7, pp. 8830455-130685, 2019, doi: 10.1109/ACCESS.2019.2940210.
- [55] Y. Zhang, J. Ma, L. Hu, K. Yu, L. Song, and H. Chen, "A haze feature extraction and pollution level identification pre-warning algorithm," *Comput. Mater. Continua*, vol. 64, no. 3, pp. 1929-1944, 2020, doi: 10.32604/cmc.2020.010556.
- [56] H.-S. Moon and J. Seo, "Fast user adaptation for human motion prediction in physical human-robot interaction," *IEEE Robotics Autom. Lett.*, vol. 7, no. 1, pp. 120-127, 2022, doi: 10.1109/LRA.2021.3116319.
- [57] Z. Huang, J. Shao, W. Guo, W. Li, J. Zhu, Q. He, and D. Fang, "Tool wear prediction based on multi-information fusion and genetic algorithm- optimized Gaussian process regression in milling," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, doi: 10.1109/TIM.2023.3280531.