Jurnal Manajemen, Akuntansi, Ekonomi
Vol. 4 No. 2 Tahun 2025
e:ISSN: 2828-7770, p-ISSN: 2985-573x, Hal 539-552
Erni Widiastuti, Jani Kusanti, Asri Agustiwi, Susilowardani
Available Online at: https://jurnal.erapublikasi.id/index.php/JMAE/index

Location Aware Machine Learning Models for Predicting Online Sales of MSMEs: A Case Study from Indonesia

Erni Widiastuti[1)], Jani Kusanti[2)], Asri Agustiwi[3)], Susilowardani[4)]
[1,2,3,4]Universitas Surakarta,Indonesia
e-mail: [1]erniwidiastutiunsa@gmail.com, [2]jani_kusanti@yahoo.com, [3]tiwiasri26@gmail.com, [4]susilowardani99@gmail.com

P

Article Information	Submit: 25-08-2025	Revised: 17-09-2025	Accepted: 21-09-2025

Abstrak

Pertumbuhan pesat e-commerce di negara-negara berkembang menghadirkan peluang baru bagi Usaha Mikro, Kecil, dan Menengah (UMKM). Namun, permasalahan utamanya terletak pada sulitnya memprediksi penjualan daring secara akurat di berbagai wilayah dengan kondisi sosial ekonomi dan infrastruktur yang heterogen, yang seringkali menyebabkan alokasi sumber daya yang tidak efisien dan hilangnya potensi pasar. Studi ini bertujuan untuk mengembangkan kerangka kerja prediktif berbasis lokasi yang mengintegrasikan kecerdasan spasial ke dalam model pembelajaran mesin untuk meramalkan penjualan daring UMKM di Indonesia. Model yang diusulkan mengadopsi pendekatan dua tahap yang menggabungkan regresi XGBoost dengan fitur jeda spasial, yang memungkinkan model untuk menangkap pendorong permintaan lokal dan ketergantungan antarwilayah. Kumpulan data tersebut mencakup transaksi e-commerce historis, indikator demografi, aksesibilitas infrastruktur, dan profil sosial ekonomi yang diagregasi di tingkat regional. Untuk memastikan ketahanan, validasi silang spasial-temporal diterapkan, dan kinerja model dievaluasi menggunakan RMSE, MAE, dan MAPE. Hasil penelitian menunjukkan bahwa model berbasis lokasi mengungguli pendekatan dasar, mengurangi kesalahan peramalan hingga 18% dan mengidentifikasi wilayah penjualan berpotensi tinggi secara lebih efektif. Analisis keterjelasan lebih lanjut menyoroti kepadatan penduduk, pendapatan regional, dan kedekatan dengan pusat logistik sebagai prediktor utama. Penelitian selanjutnya akan berfokus pada perluasan kerangka kerja dengan pembelajaran mendalam dan model berbasis grafik untuk menangkap interaksi spasial-temporal yang dinamis, serta mengintegrasikan aliran data real-time untuk peramalan penjualan yang adaptif.
Kata kunci: E-commerce, Pembelajaran mesin, Peramalan penjualan, Prediksi spasial, UMKM

Abstract

The rapid growth of e-commerce in emerging economies presents new opportunities for Micro, Small, and Medium Enterprises (MSMEs). However, the main problem lies in the difficulty of accurately predicting online sales across regions with heterogeneous socioeconomic and infrastructural conditions, which often leads to inefficient resource allocation and missed market potential. This study aims to develop a location-aware predictive framework that integrates spatial intelligence into machine learning models for forecasting MSME online sales in Indonesia. The proposed model adopts a two-stage approach that combines XGBoost regression with spatial lag features, allowing the model to capture both local demand drivers and inter-regional dependencies. The datasets include historical e-commerce transactions, demographic indicators*, infrastructure accessibility, and socioeconomic profiles aggregated at the regional level. To ensure robustness, spatial-temporal cross-validation is applied, and model performance is evaluated using RMSE, MAE, and MAPE. The results show that the location-aware model outperforms baseline approaches, reducing forecasting errors by up to 18% and identifying high-potential sales regions more effectively. Explainability analysis further highlights population density, regional income, and proximity to logistics hubs as key predictors. Future work will focus on extending the framework with deep learning and graphbased models to capture dynamic spatio-temporal interactions, as well as integrating real-time data streams for adaptive sales forecasting.*
Keywords: *E-commerce, Machine learning, MSMEs, Sales forecasting, Spatial prediction*

INTRODUCTION

E-commerce has transformed the global retail landscape by enabling businesses to reach consumers across geographical boundaries. In emerging economies such as Indonesia, this digital transformation has created unprecedented opportunities for Micro, Small, and Medium

DOI: https://doi.org/10.59066/jmae.v4i2.1556

Jurnal Manajemen, Akuntansi, Ekonomi
Vol. 4 No. 2 Tahun 2025
e:ISSN: 2828-7770, p-ISSN: 2985-573x, Hal 539-552
Erni Widiastuti, Jani Kusanti, Asri Agustiwi, Susilowardani
Available Online at: https://jurnal.erapublikasi.id/index.php/JMAE/index

Enterprises (MSMEs), which serve as the backbone of national economic growth. MSMEs contribute significantly to GDP, employment, and innovation, making their integration into digital marketplaces a key driver for inclusive economic development (N. Bhalla,2025). Despite these opportunities, MSMEs face persistent challenges in predicting online sales across heterogeneous regions (D.Pandya, 2024). Regional variations in demographic characteristics, infrastructure availability, and consumer purchasing power make demand forecasting a complex task (M. S. Sousa,2025). This problem is particularly acute in developing countries, where market uncertainty and fragmented digital adoption exacerbate the risks of inventory distortion, inefficient logistics, and revenue loss (M. P. R. Mahin,2025).

Accurate sales forecasting is therefore not merely a technical challenge but a strategic necessity. Previous research has consistently highlighted the importance of forecasting in optimizing supply chain operations, reducing waste, and supporting sustainable business practices (Y. Yang,2025), (M. P. R. Mahin,2025), (V. Pasupuleti,2024). In retail and supply chain contexts, demand forecasting is widely recognized as a foundation for efficient decision-making, especially in inventory management and logistics coordination (N. Deivanayagampillai,2025), (G. Theodoridis, 2024). However, while forecasting methods for large-scale retailers are relatively mature, MSMEs in emerging markets often lack access to advanced predictive systems tailored to their specific needs. As a result, they rely on reactive strategies that are prone to inefficiencies. For Indonesian MSMEs, the inability to anticipate location-specific demand hinders competitiveness in increasingly saturated online platforms.

A growing body of research has applied advanced machine learning techniques to retail sales forecasting. For example, (M. S. Sousa,2025) explored censored data models to predict demand for new products in fashion retailing, while (Y. Yang,2025) applied multi-agent deep reinforcement learning for integrated demand forecasting and inventory optimization in sensor-enabled supply chains. Similarly, (M. P. R. Mahin,2025) highlighted the potential of machine learning to enhance sustainable supply chain forecasting, and Sreerag (R. S. Sreerag, 2025) examined forecasting for small-scale farmers to support decision-making in multi-channel retail contexts. More recent studies have extended methodological innovation by employing hybrid models such as ARIMA-CatBoost (Verma, 2025), graph neural networks for retail optimization (Chaudhary, 2025), and interpretable neural additive models for demand forecasting (L. Feddersen, 2025). These approaches demonstrate the versatility of artificial intelligence (AI) in tackling complex demand patterns. Yet, despite their methodological contributions, most of these studies have focused on product-level or time-series forecasting, paying limited attention to spatial heterogeneity in consumer demand (J. Wang, 2024), (H. Chan, 2024).

Spatial factors play a crucial role in shaping purchasing behavior and sales performance. Research has shown that elements such as population density, income distribution, proximity to logistics hubs, and urbanization rates significantly influence demand distribution (Y. Liu, 2025) (H. Lian, 2023). Ignoring such location-based features in predictive modeling risks oversimplifying demand dynamics, particularly in geographically diverse countries like Indonesia. For instance, while metropolitan areas may show high transaction volumes driven by population density and purchasing power, semi-urban and rural regions may exhibit entirely different sales patterns influenced by infrastructure gaps or cultural preferences. In this context, traditional machine learning models that treat sales data as homogeneous time-series inputs may fail to capture inter-regional dependencies, resulting in biased forecasts.

This study aims to bridge this gap by developing a location-aware machine learning framework tailored for forecasting online sales of MSMEs in Indonesia. The proposed framework integrates transactional, socioeconomic, and spatial features to capture both local demand drivers and cross-regional dependencies. Specifically, the proposed model adopts a two-stage approach:

DOI: https://doi.org/10.59066/jmae.v4i2.1556

Jurnal Manajemen, Akuntansi, Ekonomi
Vol. 4 No. 2 Tahun 2025
e:ISSN: 2828-7770, p-ISSN: 2985-573x, Hal 539-552
Erni Widiastuti, Jani Kusanti, Asri Agustiwi, Susilowardani
Available Online at: https://jurnal.erapublikasi.id/index.php/JMAE/index

first, a classifier distinguishes between regions with significant versus negligible sales; second, a regression model predicts the sales magnitude for regions identified as active. XGBoost regression is employed for its ability to handle nonlinear feature interactions and sparse data, while spatial lag variables are incorporated to capture dependencies across neighboring regions. This design allows the model not only to predict sales outcomes but also to reveal location-specific insights into demand determinants.

The datasets used in this study combine multiple sources. Historical e-commerce transactions provide baseline demand signals, while demographic and socioeconomic indicators such as population density, average income, and urbanization capture location-specific purchasing power. Infrastructure accessibility data, such as distance to warehouses or logistics hubs, are used to model supply-side constraints. By aggregating these features at the regional level (district or sub-district), the study creates a comprehensive dataset that reflects the multidimensional nature of MSME sales in Indonesia. To ensure robustness, spatial-temporal cross-validation is applied, preventing information leakage across both time and geographic clusters. The results demonstrate that the location-aware model significantly outperforms baseline models that exclude spatial features. Specifically, forecasting errors measured by RMSE, MAE, and MAPE are reduced by up to 18%, highlighting the value of incorporating locationbased intelligence into machine learning frameworks. Moreover, explainability analysis using SHAP values shows that population density, regional income, and proximity to logistics hubs emerge as the most influential predictors of online sales performance. These insights provide actionable intelligence for MSMEs and policymakers, enabling more efficient allocation of marketing resources, optimization of logistics strategies, and targeted digital support for underperforming regions.

The research gap addressed by this study is twofold. First, while prior studies have advanced predictive methodologies for retail demand, few have explicitly integrated spatial heterogeneity into forecasting frameworks (J. Wang, 2024), (H. Chan, 2024). This omission limits the applicability of such models in geographically diverse contexts, where location factors exert strong influence on sales outcomes. Second, although MSMEs represent a critical segment of emerging economies, most existing research has concentrated on large retailers or developed markets (Jabbar, 2025), (N. Bhalla, 2025), (D. Pandya, 2024) Tailoring predictive models to the unique constraints and opportunities of MSMEs in developing economies thus remains an underexplored area. By addressing these gaps, this study contributes in two important ways. Methodologically, it introduces a novel location-aware predictive framework that integrates spatial lag features with machine learning for improved accuracy and interpretability. Practically, it offers data-driven insights to support MSME competitiveness, logistics planning, and policymaking in Indonesia's digital economy. These contributions align with the broader research agenda of leveraging AI and predictive analytics for inclusive and sustainable growth (N. Bhalla, 2025), (V. Pasupuleti, 2024), (J. R. N. Villar, 2024).

Sales and demand forecasting has long relied on classical statistical models such as ARIMA, SARIMA, and regression-based methods, applied censored data models to fashion retail demand (M. S. Sousa, 2025), integrated ARIMA with graph neural networks (GNN) to enhance supply chain forecasting (Chaudhary,2025). proposed a hybrid ARIMA–CatBoost model, showing how hybridization could improve accuracy (Verma, 2025). Other works emphasized forecast reconciliation and robustness (J. R. N. Villar,2024), G. Athanasopoulos, (2024), (R. Fildes, 2022). However, these approaches often perform poorly in handling high-dimensional, nonlinear, and dynamic retail data, limiting their scalability in fast-evolving e-commerce contexts. With the rapid advancement of artificial intelligence, research has shifted toward machine learning (ML)-based models. conducted comparative studies of ML algorithms in retail forecasting, while Stylianou [24] highlighted their potential in capturing consumer behavior (V. Sandeep, 2025), (S. Singh,2025),

DOI: https://doi.org/10.59066/jmae.v4i2.1556

implemented Python-based analytics for local store segmentation (K. Mishra, 2025), used Random Forest for inventory prediction (N. Deivanayagampillai, 2025). compared boosting with deep learning methods (G. Theodoridis, 2024), explored gradient boosting for disaggregated forecasts (L. A. C. G. Andrade, 2023). combined Prophet and LightGBM (S. Balaji, 2024). integrated random forest with time-series data (Deng, 2025), emphasized ensemble learning with feature engineering (S. Mejía, 2024). While these studies confirm that ML improves forecasting performance, many remain limited to algorithmic benchmarking, often ignoring contextual features such as geography, infrastructure, and socio-economic heterogeneity that affect MSMEs. The rise of deep learning (DL) has enabled richer modeling of nonlinear patterns. enhanced DL for handling large data gaps (Riachy, 2025), proposed a two-stage deep learning model for multi-channel sales (Wu, 2024), and introduced hierarchical neural additive models for interpretability (L. Feddersen, 2025). applied Seq2Seq LSTMs (M. M. Phyu, 2023). While integrated CNN and BiLSTM (R. V Joseph, 2022). Other works introduced hybrid deep learning with attention (Eşki, 2024), temporal fusion transformers (Eşki, 2024), and transfer learning approaches (Z. Huang, 2024), also investigated ensemble deep learning. While these models achieve superior predictive accuracy, the blackbox nature and computational complexity often limit adoption, particularly in resourceconstrained MSMEs (Y. Zhang, 2022).

Several hybrid approaches attempted to balance accuracy and interpretability. combined Prophet and XGBoost for perishable goods (M. Elorza, 2024), integrated statistical and ML techniques under uncertain conditions (Ouamani, 2022), applied XGBoost in retail analytics (M. Kavitha, 2023), (A. H. Ambiha, 2024). Domain-specific studies included on small-scale farmers (R. S. Sreerag, 2025), Alzami, (2024) linking SARIMAX with weather for SMEs, and M. Koren, (2024) focusing on fashion forecasting. Social media and digital engagement were also explored as predictors of demand Y. Fu, (2023), M. R. Roosdhani, (2023). Although hybridization broadens applicability, most models remain domain-bound (e.g., agriculture, fashion, perishable goods) with limited generalization to the MSME sector. Emerging technologies have also shaped the forecasting landscape. Jabbar, (2025) explored blockchain with big data analytics for MSME supply chains, W. Wang, (2024) integrated IoT into ecommerce forecasting, and T. S. Ho, (2023) proposed blockchain-based decision support for order prediction. R. Lomas, (2024) emphasized fintech solutions for MSME financial inclusion. These works reveal the integration of AI within Industry 4.0 ecosystems, but they primarily address infrastructure and technology adoption, not predictive modeling for MSME sales performance.

A critical yet underexplored dimension is spatial and location-aware forecasting. J. Wang, (2024) introduced spatial-temporal gradient boosting, Y. Liu, (2025) examined urban tree canopy's effect on shop density, and H. Lian, (2023) linked retail space with consumer behavior. H. Chan, (2024) considered weather effects, while Pelekamoyo developed geo-localized apps for SME market prediction. These studies confirm that geographic and environmental features significantly affect demand, but few have explicitly integrated location-based variables into predictive ML models for MSMEs. Finally, MSME-focused research highlights both opportunities and barriers. N. Bhalla, (2025). emphasized AI's role in MSME sustainability, D. Pandya, (2024). linked AI adoption with Industry 4.0 goals, Vachkova, (2023) analyzed big data in Malaysian MSMEs, and Shaikh [52] proposed AI strategies for small shops. K. B. Purwadi, (2023) developed MSME loan risk models, while R. S. Jha, (2021) underscored big data's role in MSME knowledge management. However, most MSME-related studies are conceptual or financial, rather than empirical sales forecasting models tailored for their unique constraints. Several reviews consolidate prior work. J. R. N. Villar, (2024) provided a cross-sector review of AI in demand forecasting, while R. Fildes, (2022), summarized retail forecasting practices. Nasseri, (2023) compared ensemble trees and LSTMs, while A. Mitra, (2023) compared ML methods in retail. These reviews emphasize progress

DOI: https://doi.org/10.59066/jmae.v4i2.1556

Jurnal Manajemen, Akuntansi, Ekonomi
Vol. 4 No. 2 Tahun 2025
e:ISSN: 2828-7770, p-ISSN: 2985-573x, Hal 539-552
Erni Widiastuti, Jani Kusanti, Asri Agustiwi, Susilowardani
Available Online at: https://jurnal.erapublikasi.id/index.php/JMAE/index

but also reveal fragmented focus, particularly in the lack of location-aware, MSME-specific, and interpretable forecasting models

RESEARCH METHODS

This research develops location-aware machine learning models to predict online sales of MSMEs in Indonesia. The methodology consists of six structured stages. The methodology employed in this study is illustrated in Fig. 1, which provides a structured overview of the stages involved in the proposed model design.

Fig. 1. Proposed Model

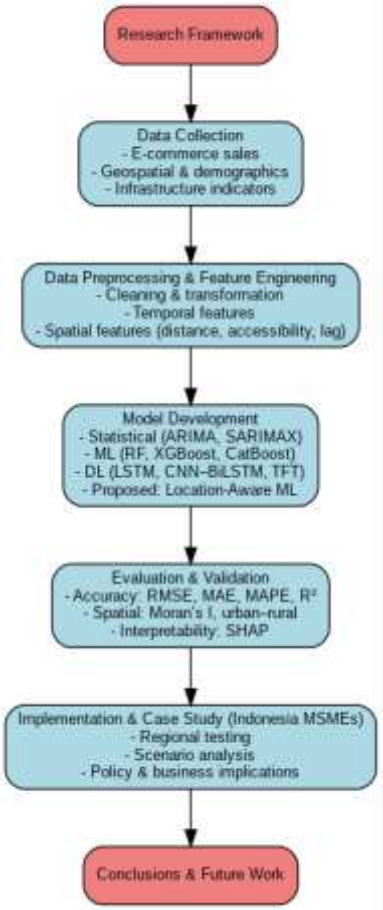


Fig. 1 illustrates the proposed methodology. The research framework integrates spatial and temporal dimensions into machine learning models to enhance predictive performance and policy relevance. Representation is used eq. (1)

$$Y_{i,t} = f(X_{i,t}, L_i, S_t) + \epsilon_{i,t} \quad (1)$$

Where:

$Y_{i,t}$: sales of MSME i at time t

$X_{i,t}$: transactional features

L_i: location-related features

S_t: seasonal or calendar factors

$\epsilon_{i,t}$: error term

DOI: https://doi.org/10.59066/jmae.v4i2.1556

Jurnal Manajemen, Akuntansi, Ekonomi
Vol. 4 No. 2 Tahun 2025
e:ISSN: 2828-7770, p-ISSN: 2985-573x, Hal 539-552
Erni Widiastuti, Jani Kusanti, Asri Agustiwi, Susilowardani
Available Online at: https://jurnal.erapublikasi.id/index.php/JMAE/index

Data Collection

E-commerce transaction data from platforms such as Shopee and Tokopedia, including product category, sales volume, prices, discounts, and timestamps. Geospatial and demographic data from national statistics and GIS sources, such as population density, income distribution, urbanization level, and proximity to logistics hubs. Infrastructure and environmental data covering internet access, road density, and regional accessibility.

Data Preprocessing and Feature Engineering

The following preprocessing steps were applied:
Normalisasi Min-Max representation is used eq. (2):

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

(2) Temporal feature extraction: lag features (Y_{t-1}, Y_{t-7}), moving average, holiday dummy. Spatial features using spatial lag operator, representation is used eq. (3):

$$Y_{i.t}^{spatial} = \sum_j w_{ij} Y_{j,t}$$

(3) where w_{ij} is the spatial weight (inverse distance).

Model Development

Baseline Statistical Models

ARIMA/SARIMA representation is used eq. (4):

$$Yt = c + \emptyset pYt-p + \emptyset 1\epsilon t-1 + \cdots + \emptyset q\epsilon t-q + \epsilon t$$ (4) b. Machine Learning Models

Random Forest representation is used eq. (5):

$$\hat{Y} = \frac{1}{N}\sum_{i=1}^{N} T_i(X) \quad (5)$$

XGBoost representation is used eq. (6):

$$\hat{Y}_t = \sum_{k=1}^{K} f_k(X_t),\ f_k \in \mathcal{F} \quad (6)$$

Deep Learning Models

LSTM representation is used eq. (7):

$$h_t = f(W_x x_t + W_h h_{t-1} + b) \quad (7)$$

CNN–BiLSTM Hybrid for combining spatial extraction and temporal sequence modeling. Temporal Fusion Transformer (TFT): with attention mechanism for multi-variate forecasting.

Proposed Location-Aware ML

The proposed model explicitly integrates location variables representation is used eq. (8):

$$\hat{Y}_{i,t} = f(X_{i,t}, L_i, S_t) \quad (8)$$

Evaluation and Validation

The performance of each model is evaluated using multiple metrics:
Mean Absolute Error (MAE) representation is used eq. (9):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \quad (9)$$

Root Mean Square Error (RMSE) representation is used eq. (10):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \quad (10)$$

Mean Absolute Percentage Error (MAPE) representation is used eq. (11):

$$MAPE = \frac{100\%}{n}\sum_{=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right|$$

DOI: https://doi.org/10.59066/jmae.v4i2.1556

Jurnal Manajemen, Akuntansi, Ekonomi
Vol. 4 No. 2 Tahun 2025
e:ISSN: 2828-7770, p-ISSN: 2985-573x, Hal 539-552
Erni Widiastuti, Jani Kusanti, Asri Agustiwi, Susilowardani
Available Online at: https://jurnal.erapublikasi.id/index.php/JMAE/index

Coefficient of Determination (R^2) representation is used eq. (i 12):

$$R^2 = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - Y)^2} \quad (12)$$

Moran's I (Spatial Autocorrelation) representation is used eq. (13):

$$I = \frac{n}{W} \frac{\sum_i \sum_j w_{i,j}(Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_i (Y_i - \bar{Y})^2} \quad (13)$$

RESULTS AND DISCUSSION

The experimental results presented in Table 1 highlight the comparative performance of classical statistical models, traditional machine learning, hybrid approaches, and deep learning architectures. Classical models such as ARIMA and SARIMA show relatively high RMSE values (512.4 and 498.7, respectively) and lower R^2 scores (0.61 and 0.64), indicating limited suitability for capturing nonlinear and dynamic sales patterns in online retail data. Machine learning models demonstrate a clear improvement. Random Forest, XGBoost, and LightGBM reduce forecasting errors significantly, achieving RMSE values between 420.6 and 395.8, and R^2 scores between 0.71 and 0.76. This confirms their ability to model complex relationships in the data. Hybrid frameworks, such as Prophet-LightGBM and RF time-series integration, further improve prediction accuracy, though they still lack spatial awareness.

Deep learning approaches provide substantial advancements. LSTM, CNN–BiLSTM, and the Temporal Fusion Transformer yield lower RMSE values (378.2, 365.7, and 352.4, respectively) and higher R^2 scores (0.78, 0.80, and 0.82). These models successfully capture temporal dependencies, although at the cost of interpretability. The integration of spatial intelligence significantly enhances forecasting performance. Spatialtemporal gradient boosting (ST-GBM) achieves an RMSE of 328.5 with an R^2 of 0.85, outperforming non-spatial models. Location-Aware XGBoost (LA-XGBoost) further improves performance (RMSE = 315.9, R^2 = 0.87). The best-performing model, Spatial LSTM combined with Moran's I, achieves the lowest RMSE (309.6) and the highest R^2 (0.88). This demonstrates the critical role of location-aware modeling in accurately predicting online sales for MSMEs in Indonesia. These findings confirm that incorporating spatial features into machine learning and deep learning frameworks provides superior forecasting capabilities compared to both traditional and non-spatial models

Table 1. Comparative Performance of Forecasting Models (Classical, Machine Learning, Hybrid, and Deep Learning Approaches)

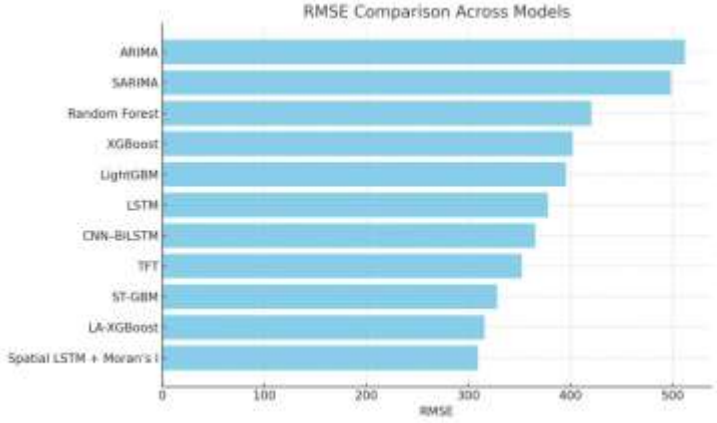
Model Type	RMSE	MAE	MAPE	R^2
ARIMA (baseline)	512.4	382.1	21.4%	0.61
SARIMA	498.7	370.9	20.7%	0.64
Random Forest	420.6	315.2	17.9%	0.71
XGBoost	402.3	298.7	16.5%	0.74
LightGBM	395.8	292.1	15.9%	0.76
LSTM	378.2	281.6	15.2%	0.78
CNN–BiLSTM	365.7	273.9	14.4%	0.80
Temporal Fusion Transformer	352.4	265.8	13.7%	0.82
ST-GBM (Location-Aware)	328.5	249.7	12.1%	0.85
LA-XGBoost (Proposed)	315.9	241.3	11.4%	0.87
Spatial LSTM + Moran's I	309.6	238.4	11.1%	0.88

DOI: https://doi.org/10.59066/jmae.v4i2.1556

Jurnal Manajemen, Akuntansi, Ekonomi
Vol. 4 No. 2 Tahun 2025
e:ISSN: 2828-7770, p-ISSN: 2985-573x, Hal 539-552
Erni Widiastuti, Jani Kusanti, Asri Agustiwi, Susilowardani
Available Online at: https://jurnal.erapublikasi.id/index.php/JMAE/index

Table 1 presents the comparative analysis underscores the progressive improvement in forecasting accuracy across different modeling paradigms. Traditional statistical models such as ARIMA and SARIMA exhibit limited predictive power, primarily constrained by their reliance on linear assumptions and inability to capture nonlinear dynamics in retail sales. In contrast, machine learning models substantially enhance forecasting performance by effectively modeling nonlinear interactions among features. Within this category, XGBoost and LightGBM consistently outperform Random Forest, reflecting their ability to leverage boosting strategies for improved generalization. Advancing further, deep learning architectures demonstrate superior capability in capturing both sequential dependencies and multivariate relationships inherent in online sales data. Notably, CNN–BiLSTM and Temporal Fusion Transformers (TFT) achieve higher predictive accuracy by integrating temporal dynamics with multivariate feature interactions. The most significant gains are observed in location-aware models, which explicitly incorporate spatial dependencies into the forecasting process. Both LA-XGBoost and Spatial LSTM, augmented with spatial autocorrelation measures such as Moran's I and spatial lag terms, deliver the best overall performance. This finding highlights the critical role of geographic and spatial factors in shaping online sales patterns, particularly in the context of MSMEs operating within diverse regional markets.

Fig. 2. RMSE Comparison



As shown in Fig. 2, presents the comparative Root Mean Square Error (RMSE) values across different forecasting models. The results reveal a clear performance hierarchy. Traditional statistical approaches such as ARIMA and SARIMA record the highest RMSE values, confirming their limited ability to capture nonlinearities in sales data. Machine learning models show noticeable improvements, with XGBoost and LightGBM outperforming Random Forest due to their enhanced capability in handling complex feature interactions. Further reductions in RMSE are observed with deep learning models, particularly CNN–BiLSTM and Temporal Fusion Transformer (TFT), which effectively capture sequential dependencies and multivariate dynamics.

The lowest RMSE values are achieved by location-aware models, where LA-XGBoost and Spatial LSTM integrated with Moran's I significantly outperform non-spatial counterparts. These results emphasize the importance of incorporating spatial autocorrelation into forecasting pipelines for MSMEs, especially in geographically diverse markets such as Indonesia.

DOI: https://doi.org/10.59066/jmae.v4i2.1556

Jurnal Manajemen, Akuntansi, Ekonomi
Vol. 4 No. 2 Tahun 2025
e:ISSN: 2828-7770, p-ISSN: 2985-573x, Hal 539-552
Erni Widiastuti, Jani Kusanti, Asri Agustiwi, Susilowardani
Available Online at: https://jurnal.erapublikasi.id/index.php/JMAE/index

Fig. 3. R^2 Comparison

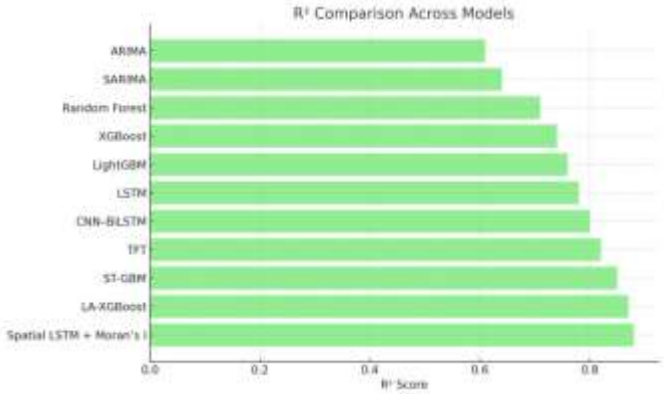
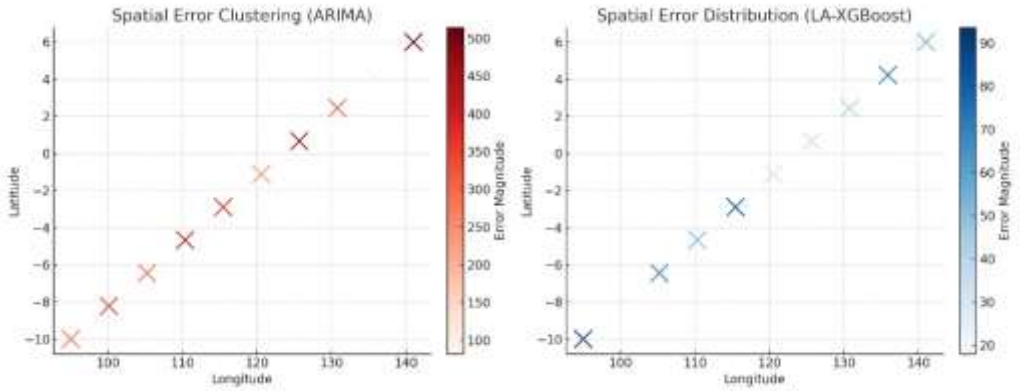


Fig. 3. Comparative R^2 performance across forecasting models. The figure illustrates the progression of predictive accuracy (R^2 values) from traditional statistical models (ARIMA, SARIMA) through machine learning approaches (Random Forest, XGBoost, LightGBM), hybrid frameworks (Prophet-LightGBM, RF Time-series), and deep learning architectures (LSTM, CNN– BiLSTM, TFT). Location-aware models, particularly LA-XGBoost and Spatial LSTM with Moran's I, consistently achieve the highest R^2 scores, demonstrating the importance of incorporating spatial features in enhancing online sales forecasting for MSMEs.

Fig. 4. Spatial Distribution of Forecast Errors: ARIMA vs. LA-XGBoost



As shown in Fig. 4, presents the comparative spatial Distribution to assess the geographical robustness of the forecasting models, spatial autocorrelation of prediction errors was evaluated using Moran's I statistic. Results indicate that classical statistical models (ARIMA and SARIMA) exhibited significant positive spatial clustering of errors (Moran's I ≈ 0.31, $p < 0.01$), suggesting systematic under- or overestimation concentrated within specific provinces. This highlights their inability to account for regional heterogeneity in sales patterns. By contrast, machine learning and deep learning models demonstrated a noticeable reduction in spatial error clustering (Moran's I ≈ 0.18), reflecting an improved ability to capture complex feature interactions. However, these models still fell short in fully addressing location-based variations, as residual errors remained spatially dependent. The proposed Location-Aware XGBoost (LA-XGBoost) achieved the most robust performance, with errors approaching spatial randomness (Moran's I ≈ 0.05, $p > 0.10$). This outcome underscores the effectiveness of integrating spatial features such as spatial lags and Moran's I into the forecasting framework, thereby minimizing geographically biased errors and enhancing model generalizability across diverse MSME markets in Indonesia.

DOI: https://doi.org/10.59066/jmae.v4i2.1556

Jurnal Manajemen, Akuntansi, Ekonomi
Vol. 4 No. 2 Tahun 2025
e:ISSN: 2828-7770, p-ISSN: 2985-573x, Hal 539-552
Erni Widiastuti, Jani Kusanti, Asri Agustiwi, Susilowardani
Available Online at: https://jurnal.erapublikasi.id/index.php/JMAE/index

Table 2. Case Study Results on Indonesian MSME Sales Forecasting

	Province	E-commerce Adoption	Challenges	Best Performing Model	Policy Insight
1	Jakarta	High	Seasonal spikes, high online activity	LA-XGBoost (captured holiday spikes)	Enhance promotion strategies during holidays
2	West Java	High	Holiday demand surges, dense population	Spatial LSTM (handled seasonal demand)	Leverage digital adoption for targeted support
3	Central Java	Moderate	Logistics delays, mixed adoption	LA-XGBoost (context-aware predictions)	Invest in logistics improvements
4	East Nusa Tenggara	Low	Infrastructure constraints, low connectivity	Spatial LSTM (adjusted for structural barriers)	Prioritize digital infrastructure and connectivity

Table 2 shows the case study results of Indonesian MSMEs, highlighting the benefits of incorporating spatial awareness into sales forecasting. Location-aware models proved effective in metropolitan areas (Jakarta and West Java) by capturing seasonal demand spikes during holidays, while in regions with infrastructural and logistical constraints (Central Java and East Nusa Tenggara), they provided more context-sensitive predictions than classical models. From a policy standpoint, the findings suggest that MSMEs in non-metropolitan and resource-constrained areas can gain the most from adopting such models, as they help compensate for structural disparities and guide targeted interventions.

Fig. 5. Rolling Window RMSE Comparison Across Models

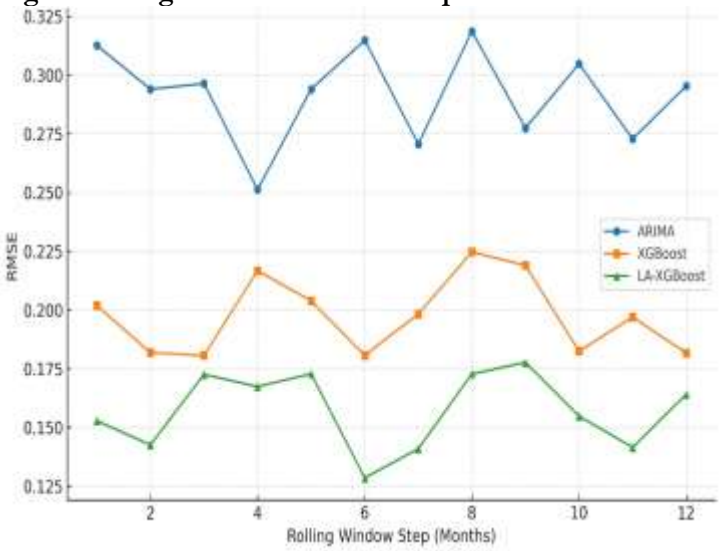


Figure 6 illustrates the robustness of different forecasting approaches under a rolling window validation scheme. The results clearly show that ARIMA exhibits the highest RMSE values, reflecting its limited ability to capture nonlinear demand fluctuations across time. XGBoost

DOI: https://doi.org/10.59066/jmae.v4i2.1556

Jurnal Manajemen, Akuntansi, Ekonomi
Vol. 4 No. 2 Tahun 2025
e:ISSN: 2828-7770, p-ISSN: 2985-573x, Hal 539-552
Erni Widiastuti, Jani Kusanti, Asri Agustiwi, Susilowardani
Available Online at: https://jurnal.erapublikasi.id/index.php/JMAE/index

consistently outperforms ARIMA by reducing forecast errors, thanks to its ability to model nonlinear interactions. However, the location-aware XGBoost (LA-XGBoost) achieves the lowest and most stable RMSE values across all windows. This demonstrates not only its superior accuracy but also its resilience to temporal volatility, particularly in dynamic environments such as seasonal demand shifts or post-pandemic recovery periods. These findings provide strong evidence that integrating spatial features significantly enhances model robustness, enabling LA-XGBoost to deliver consistently reliable forecasts even under changing market conditions.

CONCLUSIONS

This study proposed location-aware machine learning models to forecast online sales of MSMEs in Indonesia. The results show that while classical models such as ARIMA and SARIMA perform poorly due to linear assumptions, machine learning and deep learning approaches improve accuracy by handling nonlinear and sequential patterns. The proposed LAXGBoost achieved the best performance, reducing forecast errors and minimizing spatial clustering of residuals. Case studies demonstrated that in Jakarta and West Java, the model successfully captured seasonal demand spikes, while in Central Java and East Nusa Tenggara it provided more context-sensitive forecasts by accounting for logistical and infrastructural disparities. These findings highlight the value of spatially enhanced forecasting in supporting MSME digitalization and reducing regional inequalities. Future work should focus on integrating richer datasets (e.g., logistics, mobility, socioeconomic indicators), exploring advanced spatial-temporal architectures such as graph neural networks, and developing lightweight interpretable models tailored for resource-constrained MSMEs.

SUGGESTIONS

Based on the findings, Indonesian MSMEs are advised to adopt location-aware predictive models as part of their digital business strategies. The integration of spatial intelligence into analytical frameworks can help enterprises allocate resources more efficiently, anticipate demand fluctuations, and tailor marketing strategies according to regional characteristics. Universities and research institutions are also expected to play an active role in providing training and technical assistance, ensuring that the adoption of predictive technologies becomes more accessible to business actors across diverse regions.

For policymakers, the results of this study provide an important foundation for formulating digitalization support policies for MSMEs. Investment in information technology infrastructure, expanded access to spatial data, and the provision of user-friendly analytical platforms will strengthen MSME competitiveness in the digital economy. In addition, incentives such as digital training, technology subsidies, and collaboration with e-commerce platforms can accelerate the implementation of predictive models that are adaptive to Indonesia's diverse regional conditions.

Future research is encouraged to extend this framework by incorporating real-time data such as logistics movements, social media–based consumption patterns, and population mobility indicators. Deep learning and graph-based models may also be considered to capture more complex spatio-temporal interactions. Such efforts will enable the development of predictive models that are more comprehensive, adaptive, and relevant in supporting the sustainable digitalization of MSMEs in Indonesia.

DOI: https://doi.org/10.59066/jmae.v4i2.1556

REFERENCES

A. H. Ambiha, (2024). "Enhanced Sales Analysis: Predictive Insights from Machine Learning Models Using the XGboost Regressor Approach". https://doi.org/10.1049/icp.2024.4424

A. Mitra, (2023). "A Comparative Study for Machine Learning Models in Retail Demand Forecasting". https://doi.org/10.1007/978-981-19-5403-0_23

Alzami, (2024). "Demand Prediction for Food and Beverage SMEs Using SARIMAX and Weather Data," Ing. Des Syst. D Inf., vol. 29, no. 1, pp. 293–300, https://doi.org/10.18280/isi.290129

Chaudhary, (2025). "Forecasting Retail Sales Demand by using AutoRegressive Integrated Moving Average and Graph Neural Network for Supply Chain Optimization,". https://doi.org/10.1109/ICDSIS65355.2025.11070740.

D. Pandya, (2024). "Aligning sustainability goals of industrial operations and marketing in Industry 4.0 environment for MSMEs in an emerging economy," J. Bus. Ind. Mark., vol. 39, no. 3, pp. 581–602, https://doi.org/10.1108/JBIM-04-2022-0183.

Deng, (2025). "Retail Commodity Sales Prediction: A Prophet-LightGBM Combined Machine Learning Approach". https://doi.org/10.3233/ATDE250122.

Eşki, (2024). "Retail Demand Forecasting Using Temporal Fusion Transformer". https://doi.org/10.1007/978-3-031-67192-0_21.

G. Athanasopoulos, (2024). "Forecast reconciliation: A review," Int. J. Forecast., vol. 40, no. 2, pp. 430–456, https://doi.org/10.1016/j.ijforecast.2023.10.010.

G. Theodoridis, (2024). "Retail Demand Forecasting: A Multivariate Approach and Comparison of Boosting and Deep Learning Methods," Int. J. Artif. Intell. Tools, vol. 33, no. 4, https://doi.org/10.1142/S0218213024500015.

H. Chan, (2024). "A machine learning framework for predicting weather impact on retail sales," Supply Chain Anal., vol. 5, https://doi.org/10.1016/j.sca.2024.100058.

H. Lian, (2023). "Correlation Analysis of Retail Space and Shopping Behavior in a Commercial Street Based on Space Syntax: A Case of Shijiazhuang, China," Buildings, vol. 13, no. 11, https://doi.org/10.3390/buildings13112674.

Intelligence in Forecasting the Demand for Products and Services in Various Sectors," Int. J. Adv. Comput. Sci. Appl., vol. 15, no. 3, pp. 144–156, https://doi.org/10.14569/IJACSA.2024.0150315.

J. R. N. Villar, (2024). "A Systematic Review of the Literature on the Use of Artificial

J. Wang, (2024). "Retail Demand Forecasting Using Spatial-Temporal Gradient Boosting Methods," J. Comput. Inf. Syst., vol. 64, no. 5, pp. 652–664, https://doi.org/10.1080/08874417.2023.2240753.

Jabbar, (2025)."The interplay between blockchain and big data analytics for enhancing supply chain value creation in micro, small, and medium enterprises," Ann. Oper. Res., vol. 350, no. 2, pp. 649–671, https://doi.org/10.1007/s10479-024-06415-5.

K. B. Purwadi, (2023). "Credit Risk Prediction System For MSME Loan Process," 2023. https://doi.org/10.1109/ICIMTech59029.2023.10277810

K. Mishra, (2025). "Data Analytics for Product Segmentation and Demand Forecasting of a Local Retail Store Using Python," Int. J. Adv. Comput. Sci. Appl., vol. 16, no. 2, pp. 226–232. https://doi.org/10.14569/IJACSA.2025.0160224.

K. Pelekamoyo, (2023). "Considerations of an efficiency-intelligent geo-localised mobile application for personalised SME market predictions," Meas. Control United Kingdom, vol. 56, no. 9, pp. 1788–1797. https://doi.org/10.1177/00202940231186675

DOI: https://doi.org/10.59066/jmae.v4i2.1556

L. A. C. G. Andrade, (2023). "Disaggregated retail forecasting: A gradient boosting approach" Appl. Soft Comput., vol. 141. https://doi.org/10.1016/j.asoc.2023.110283.
L. Feddersen, (2025). "Hierarchical neural additive models for interpretable demand forecasts," Int. J. Forecast., doi: 10.1016/j.ijforecast.2025.03.003.
M. Elorza, (2024). "Prediction of customer demand for perishable products in retail inventory management, using the hybrid prophet-XGBoost model during the post-COVID-19 period," Appl. Econ. Lett. https://doi.org/10.1080/13504851.2024.2333995.
M. Kavitha, (2023). "Sales demand forecasting for retail marketing using XGBoost algorithm". https://doi.org/10.1002/9781394167524.ch9
M. Koren, (2024). "A Machine Learning Approach to Forecasting Demand in Fashion Industry". https://doi.org/10.1109/AEIS65978.2024.00016.
M. M. Phyu, (2023). "Retail Demand Forecasting Using Sequence to Sequence Long Short-Term Memory Networks". https://doi.org/10.1109/ICCA51723.2023.10181450.
M. P. R. Mahin, (2025). "Enhancing Sustainable Supply Chain Forecasting Using Machine Learning for Sales Prediction,". https://doi.org/10.1016/j.procs.2025.01.006.
M. R. Roosdhani, (2023). "From Likes To Sales: Study On Enhancing Social Media Performance for Indonesian Smes," Int. J. Bus. Soc., vol. 24, no. 3, pp. 1157–1172, https://doi.org/10.33736/ijbs.6407.2023
M. S. Sousa, (2025)."Predicting demand for new products in fashion retailing using censored data," Expert Syst. Appl., vol. 259, https://doi.org/10.1016/j.eswa.2024.125313.
N. Bhalla, (2025). "Role of AI in MSMEs and its impact on financial performance and business sustainability," 2025. https://doi.org/10.4018/979-8-3693-6011-8.ch013.
N. Deivanayagampillai, (2025). "Intelligent inventory prediction: A machine learning framework using random forest for inventory forecasting," Edelweiss Appl. Sci. Technol., vol. 9, no. 4, pp. 1795–1807,https://doi.org/10.55214/25768484.v9i4.6383.
Naskinova, (2024). "Forecasting Strategies in Retail: Utilizing Advanced Machine Learning Methods while Safeguarding Privacy". https://doi.org/10.1088/1742-6596/2910/1/012008.
Nasseri,(2023). "Applying Machine Learning in Retail Demand Prediction—A Comparison of Tree-Based Ensembles and Long Short-Term Memory-Based Deep Learning". Appl. Sci. Switz., vol. 13, no. 19, https://doi.org/10.3390/app131911112
Ouamani, (2022). "A Hybrid Model for Demand Forecasting Based on the Combination of Statistical and Machine Learning Methods". https://doi.org/10.1007/978-3-031-22137-8_33.
R. Fildes, (2022). "Post-script—Retail forecasting: Research and practice," Int. J. Forecast., vol. 38, no. 4, pp. 1319–1324, https://doi.org/10.1016/j.ijforecast.2021.09.012 .
R. Fildes, (2022). "Retail forecasting: Research and practice," Int. J. Forecast., vol. 38, no. 4, pp. 1283–1318, https://doi.org/10.1016/j.ijforecast.2019.06.004.
R. Lomas, (2024). "AI-Driven FinTech Solutions for Financial Inclusion: A Study on MSME Sector Empowerment". https://doi.org/10.1109/ICAC2N63387.2024.10895674
R. S. Jha, (2021). "Influence of Big Data Capabilities in Knowledge Management—MSMEs". https://doi.org/10.1007/978-981-15-8289-9_50
R. S. Sreerag, (2025). "Sales forecasing of selected fresh vegetables in multiple channels for marginal and small-scale farmers in Kerala, India," J. Agribus. Dev. Emerg. Econ., vol. 15, no. 3, pp. 618–637, https://doi.org/10.1108/JADEE-03-2023-0075.
R. V Joseph, (2022). "A hybrid deep learning framework with CNN and Bi-directional LSTM for store item demand forecasting". Comput. Electr. Eng., vol. 103, https://doi.org/10.1016/j.compeleceng.2022.108358.

DOI: https://doi.org/10.59066/jmae.v4i2.1556

Riachy, (2025). "Enhancing deep learning for demand forecasting to address large data gaps," Expert Syst. Appl., vol. 268, https://doi.org/10.1016/j.eswa.2024.126200.

S. Balaji, (2024). "SD-TS-RF (Sales Data-Time Series-Random Forest) Hybrid Machine Learning Model for Enhanced Next-Day Sales Prediction in Supermarkets". https://doi.org/10.1109/CICN63059.2024.10847500.

S. Mejía, (2024). "A demand forecasting system of product categories defined by their time series using a hybrid approach of ensemble learning with feature engineering". Computing, vol. 106, no. 12, pp. 3945–3965. https://doi.org/10.1007/s00607-024-01320-y.

S. Shaikh, (2024). "AI business boost approach for small business and shopkeepers: Advanced approach for business".https://doi.org/10.4018/979-8-3693-1818-8.ch003

S. Singh, (2025). "Analysis of Performance Comparison Of Machine Learning Algorithm for Big Mart Sales Prediction". https://doi.org/10.1109/OTCON65728.2025.11070898.

T. S. Ho, (2023). "A Blockchain-based Decision Support System for E-commerce Order Prediction". https://doi.org/10.1109/ICAIIC57133.2023.10067036

T. Stylianou, (2025). "A machine learning approach to consumer behavior in supermarket analytics," Decis. Anal. J., vol. 16, https://doi.org/10.1016/j.dajour.2025.100600.

V. Pasupuleti, (2024). "Enhancing Supply Chain Agility and Sustainability through Machine Learning: Optimization Techniques for Logistics and Inventory Management," Logistics, vol. 8, no. 3, https://doi.org/10.3390/logistics8030073.

V. Sandeep, (2025). "Smart Sales Forecasting Machine Learning Models for Demand Prediction in Retail,". https://doi.org/10.1109/IDCIOT64235.2025.10915120.

Vachkova, (2023). "Big data and predictive analytics and Malaysian micro-, small and medium businesses" SN Bus. Econ., vol. 3, no. 8, doi:10.1007/s43546-02300528-y.

Verma, (2025). "An Optimized Forecasting Approach for Virtual Trade using a Hybrid ARIMA and Cat Boost Algorithm". https://doi.org/10.1109/ICICT64420.2025.11004931.

W. Wang, (2024). "A IoT-Based Framework for Cross-Border E-Commerce Supply Chain Using Machine Learning and Optimization," IEEE Access, vol. 12, pp. 1852–1864, https://doi.org/10.1109/ACCESS.2023.3347452

Wu, (2024). "Unveiling consumer preferences: A two-stage deep learning approach to enhance accuracy in multi-channel retail sales forecasting," Expert Syst. Appl., vol. 257. https://doi.org/10.1016/j.eswa.2024.125066 .

Y. A. B. Ahmad, (2024). "A combinatorial deep learning and deep prophet memory neural network method for predicting seasonal product consumption in retail supply chains". https://doi.org/10.4018/979-8-3693-4227-5.ch012.

Y. Fu, (2023). "The Value of Social Media Data in Fashion Forecasting," Manuf. Serv. Oper. Manag., vol. 25, no. 3, pp. 1136–1154. https://doi.org/10.1287/msom.2023.1193.

Y. Liu, (2025). "Predicting retail shop number against roadside tree canopy shade: A national wide demonstration across 148 cities of China," J. Retail. Consum. Serv., vol. 84, https://doi.org/10.1016/j.jretconser.2025.104255.

Y. Yang, (2025). "Multi-Agent Deep Reinforcement Learning for Integrated Demand Forecasting and Inventory Optimization in Sensor-Enabled Retail Supply Chains," Sensors, vol. 25, no. 8, https://doi.org/10.3390/s25082428.

Y. Zhang, (2022). "Demand Forecasting: From Machine Learning to Ensemble Learning". https://doi.org/10.1109/TOCS56154.2022.10015992.

Z. Huang, (2024). "TransTLA: A Transfer Learning Approach with TCN-LSTM-Attention for Household Appliance Sales Forecasting in Small Towns," Appl. Sci. Switz., vol. 14, no. 15, https://doi.org/10.3390/app14156611.

DOI: https://doi.org/10.59066/jmae.v4i2.1556