IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 4, April 2025

DOI: 10.17148/IJARCCE.2025.14427

Retail Real-Time Sales Prediction System Using LSTM and XGBoost

Ms. Priyanka Yadav[1], Mr. C. R. Barde[2]

P.G. Student, Computer Engineering Department, R.H. Sapat College of Engineering, Nashik, India[1]

Assistant. Professor, Computer Engineering Department, R.H. Sapat College of Engineering, Nashik, India[2]

Abstract: Accurate sales forecasting is essential for retail businesses to optimize inventory, enhance customer satisfaction, and drive strategic decisions. This paper introduces a robust sales prediction system that integrates Long Short-Term Memory (LSTM) networks for time-series forecasting and XGBoost for predictive analytics to deliver reliable and precise sales predictions. Designed for modern retail environments, the system seamlessly integrates with Point-of Sale (POS) systems to enable real-time data ingestion and dynamic prediction capabilities. Users can also upload custom datasets and explore interactive modules for analyzing current sales trends and forecasting future demand. A React-powered dashboard offers intuitive data visualization, while a Flask-based backend ensures scalability and efficient processing. By combining cutting-edge machine learning models with real-time data handling and user-centric features, this solution empowers retailers to respond to market changes and gain a competitive advantage proactively.

Keywords: Retail, sales prediction, LSTM networks, XGBoost, real-time fore casting, POS integration, machine learning, interactive analytics, time-series prediction, data visualization.

I. INTRODUCTION

The fast-paced nature of the retail sector and the increasing complexity of consumer behaviour highlight the critical need for ac curate sales forecasting to ensure business success. Retailers face challenges from rapidly changing customer preferences, promotional campaigns, and external disruptions such as economic fluctuations or unforeseen global events. Traditional forecasting approaches, which primarily rely on historical sales data and basic statistical models, often fall short in addressing these dynamic conditions. These methods struggle to adapt to new trends, fail to consider external factors, and lack the ability to predict sudden surges or declines in demand with precision.

To overcome these limitations, this research focuses on leveraging advanced machine learning techniques and real-time analytics for sales prediction. By integrating Long Short-Term Memory (LSTM) networks for time-series forecasting and XGBoost for predictive analytics, the proposed system captures both temporal dependencies and broader influencing factors to deliver highly accurate forecasts. Unlike traditional methods, this system integrates directly with Point-of-Sale (POS) systems, enabling real-time data ingestion and immediate predictions. Additionally, it allows users to up load their own datasets for customized forecasting, enhancing its applicability across various retail contexts.

The system is supported by an intuitive React-based dashboard that provides interactive modules for exploring sales trends and generating future sales forecasts. Users can easily navigate through features like "Enter Details to Predict Sales," "View Current Sales," and "Forecast Future Sales" for actionable insights. A Flask-based backend ensures scalability and efficient processing, making the solution robust and adaptable to the needs of modern retail businesses.

This research demonstrates the practical application of LSTM and XGBoost in addressing both short-term and long-term sales prediction challenges. By bridging the gap between advanced analytics and user-centric design, the system empowers retailers to optimize inventory, reduce waste, and respond proactively to market shifts. The integration of cutting-edge machine learning with real-time capabilities marks a significant advancement in retail analytics, providing a reliable and impactful strategy for data-driven decision making.

II. EXISTING SYSTEM

Traditional sales forecasting systems in the retail sector primarily rely on statistical methods such as time-series analysis (e.g., ARIMA) and linear regression. While these methods provide foundational insights, they are constrained by their dependence on fixed assumptions and historical data, limiting their ability to adapt to dynamic market conditions. These systems often fail to incorporate external or real-time factors such as shifts in consumer sentiment, regional events, or viral trends, which significantly influence demand. As a result, their forecasts may lack the granularity and responsiveness needed to align with modern retail challenges.

DOI: 10.17148/IJARCCE.2025.14427

Furthermore, existing systems typically operate in isolation, with minimal integration of diverse data sources or advanced analytical frameworks. This siloed approach limits their ability to account for complex interactions between multiple demand drivers. For example, a sudden surge in demand due to a viral social media trend or an unexpected drop caused by extreme weather conditions might be overlooked, leading to inaccurate predictions.

Another key limitation of traditional systems is their scalability. As retail businesses increasingly operate across multiple channels and generate large volumes of data, conventional methods often struggle to process and analyze this information efficiently. The absence of robust visualization tools further complicates decision making, as stakeholders are unable to access actionable insights in a timely manner. Additionally, these systems lack real-time pro cessing and predictive modeling capabilities, reducing their effectiveness in rapidly evolving retail environments.

While traditional methods have been widely used, their inability to handle modern data complexities, integrate real-time information, and scale efficiently underscores the need for innovative solutions. Machine learning and advanced analytics provide an opportunity to address these limitations by offering adaptive, data driven approaches for sales prediction. This research builds on these advancements to bridge the gap between conventional methods and the demands of contemporary retail forecasting, enabling businesses to respond proactively to market shifts and consumer behavior changes.

III. LITERATURE REVIEW

Recent advancements in machine learning have spurred significant innovations in sales prediction and demand forecasting across various industries, including retail, manufacturing, and food services. MDTanvir et al. [1] proposed a hybrid model combining Random Forest, XGBoost, and a linear regression layer, which demonstrated enhanced accuracy in retail demand forecasting by leveraging ensemble learning techniques. While effective, such approaches are often computationally intensive and require substantial resources, which can limit their scalability.

Gao et al. [2] developed a framework that integrates Bidirectional Long Short-Term Memory (BiLSTM) and Word2Vec for analyzing social media data to better align supply with real-time demand trends. This method effectively captured consumer sentiment and external influencers but encountered challenges related to the high volume of unstructured data and computational overhead. Similarly, Rui and Li [3] proposed a hybrid machine learning system that combines Graph Convolutional Networks (GCN), Long Short-Term Memory (LSTM), and attention mechanisms to optimize inventory through the analysis of temporal patterns and supply chain relationships. However, the complexity and resource requirements of their approach pose challenges for practical implementation in small or medium-sized enterprises.

In the retail sector, Nassibi et al. [5] compared LSTM networks and Support Vector Machines (SVM) for forecasting sales, concluding that LSTM excels in handling large datasets and identifying temporal dependencies. However, the scalability of LSTM models to high-frequency, multi-dimensional data remains a concern. Sridhar and Mohan [6] explored unstructured retail data, employing K Nearest Neighbors (KNN), Gaussian Naive Bayes, and Decision Trees for demand forecasting, with KNN showing superior performance in capturing non-linear patterns, including seasonality and promotions.

In retail-specific applications, Mitra et al. [10] introduced a hybrid RF-XGBoost-LR model for multi-channel sales prediction, achieving high accuracy with weekly sales data. However, they noted challenges such as imbalanced datasets and the need for robust preprocessing techniques. The study underscored the importance of integrating data from multiple channels to improve forecasting outcomes. In the manufacturing context, Krishnamoorthy et al. [8] evaluated multiple machine learning approaches, including regression models, neural networks, and ensemble methods, addressing challenges such as data preprocessing and computational efficiency. Aci and Yerg¨ ok [7] highlighted the effectiveness of Boosted Ensemble Decision Trees for food production forecasting, emphasizing the importance of contextual variables such as calendar effects and specific production characteristics. Building on this, Qureshi et al. [9] incorporated weather data into a Gated Recurrent Unit (GRU) based model for retail forecasting, significantly improving accuracy and highlighting the potential of external data integration.

Collectively, these studies showcase the transformative potential of machine learning in sales forecasting, addressing key challenges such as temporal dependencies, external influencers, and multivariate data integration. While existing research highlights significant progress, limitations such as computational demands, scalability, and real-time adaptability persist. This study advances the field by implementing an integrated real-time sales prediction system using LSTM and XGBoost, directly addressing these challenges and bridging the gap between academic research and practical retail applications.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁑ Peer-reviewed & Refereed journal ⁑ Vol. 14, Issue 4, April 2025

DOI: 10.17148/IJARCCE.2025.14427

IV. PROPOSED SYSTEM

The proposed system for real-time sales prediction leverages cutting-edge machine learning algorithms such as LSTM and XG Boost, coupled with an intuitive front-end built in React and a Flask-based back-end API. This architecture ensures real-time data adaptability, accurate sales predictions, and a user-friendly inter face for retail businesses.
The architecture highlights the layered design, incorporating data collection, storage, machine learning, real-time processing, business logic, user interface, and security. Each layer is equipped with specific tools and technologies for efficient demand forecasting and inventory optimization.

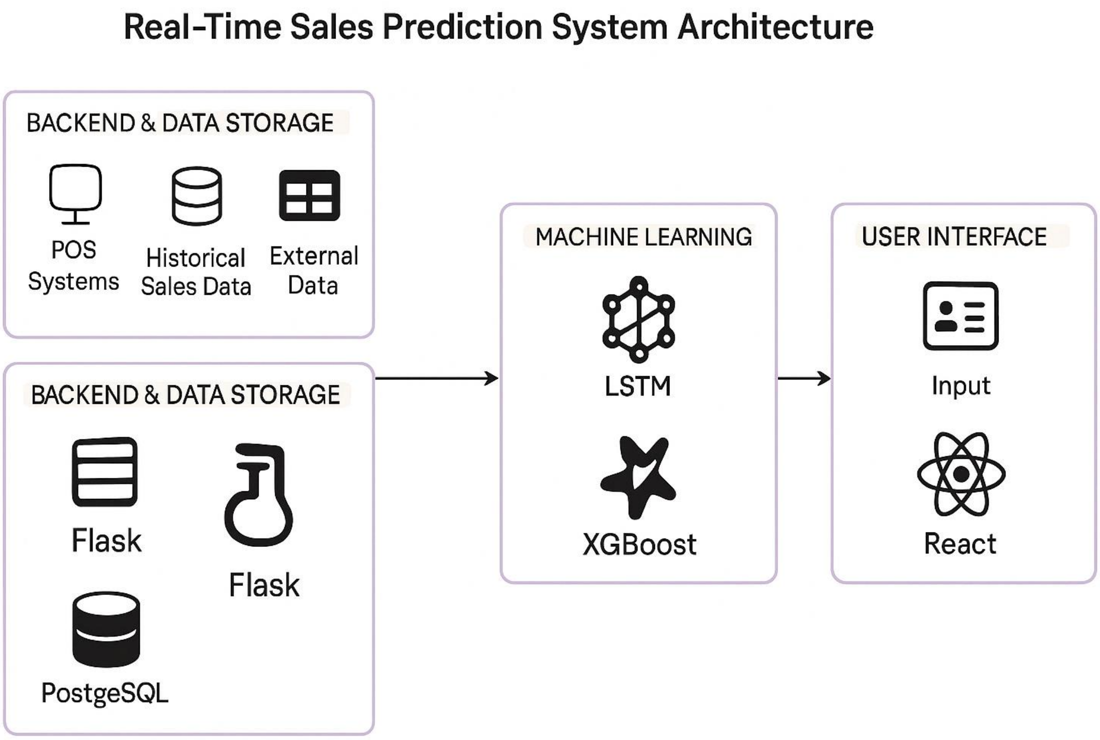

Fig. 1. System Architecture for Real-Time Sales Prediction.

A. System Architecture

The system architecture consists of multiple interconnected layers, each designed to optimize real-time sales prediction and inventory management. The architecture integrates data collection, storage, machine learning, business logic, user interface, backend, security, and real-time processing layers, all working cohesively to deliver accurate sales forecasts and actionable insights.

1. Data Collection Layer
   - Data Sources: The system gathers data from multiple sources, including point-of-sale (POS) systems, historical sales data, weather data, holiday schedules, and customer behaviour patterns.
   - Data Integration: APIs and ETL (Extract, Transform, Load) pipelines combine these datasets into a unified format. Real time data ingestion is handled using streaming solutions like Apache Kafka.
2. Data Storage Layer
   - Structured Data: Transactional and inventory data is stored in relational databases such as PostgreSQL or MySQL.
   - Unstructured Data: Customer reviews and social media data are handled using MongoDB or cloud-based solutions like AWS S3.
   - Big Data Frameworks: Apache Hadoop and Apache Spark enable the processing of large datasets for analysis.
3. Machine Learning Layer

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2025

DOI: 10.17148/IJARCCE.2025.14427

- Algorithm Used:
  LSTM:For time-series forecasting of future sales trends.
  XGBoost: For regression-based sales predictions.
- Training & Inference: The models continuously train on up dated datasets, ensuring higher prediction accuracy with evolving market trends.

4. Business Logic Layer
   - Implements inventory optimization and demand forecasting. Generates actionable alerts for understocked or overstocked items.
   - Adjusts forecasts dynamically for seasonal and market changes.
5. User Interface (UI) Layer
   - Frontend: Built using React for a responsive, interactive user experience.
   - Visualization: Charts and trend analyses are powered by Chart.js, providing visually intuitive insights to end users.
   - Features: Upload data for prediction. Visualize past and future sales
6. Security Layer
   - Authentication: OAuth 2.0 provides secure user authentication and access control.
   - Encryption: SSL/TLS protocols ensure secure data transfer, safeguarding sensitive information.
7. Cloud and Infrastructure
   - Cloud platforms such as AWS or Microsoft Azure host the system, providing scalability and high availability for large data volumes.
8. Real-Time Processing Layer
   - Real-time analysis and forecasting are powered by frameworks like Apache Spark and Hadoop to process incoming retail data promptly.

The proposed system's architecture supports real-time adaptability and scalability, ensuring efficiency in demand forecasting and inventory management for retail businesses.

V. RESULT AND DISCUSSION

The implemented sales prediction system demonstrates the integra tion of cutting-edge machine learning models, including LSTM for time-series forecasting and XGBoost for regression-based predic tions. This implementation is complemented by an interactive user interface built with React, enabling intuitive data upload and visu alization of results.

System Features and Implementation Results

The developed application includes the following key features and results, showcasing the system's efficacy in providing accurate sales forecasts and insights:

- Future Sales Forecasting: The system successfully predicts future sales trends using LSTM models, leveraging historical data to identify patterns and provide actionable insights, as shown in Fig 4.
- Sales Prediction Accuracy: By employing XGBoost, the system ensures robust prediction capabilities for sales figures, effectively handling large datasets and complex interactions.
- User-Friendly Interface: The React-based frontend provides an intuitive experience, allowing users to seamlessly upload data (see Figure 3), visualize past and future sales trends (Fig 5 and 6), and access prediction outputs.
- Real-Time Data Integration: The Flask backend ensures efficient data processing and real-time prediction generation, delivering actionable results on demand.

Evaluation Metrics

The performance of the sales prediction models was evaluated using standard metrics, as follows:

- **Mean Absolute Error (MAE):** Quantifies the average prediction error, providing a clear measure of model accuracy.
- **Mean Squared Error (MSE):** Penalizes larger errors, ensuring that the model minimizes significant deviations.
- **R^2 Score:** Represents the proportion of variance explained by the model, indicating its reliability and predictive power.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 14, Issue 4, April 2025

DOI: 10.17148/IJARCCE.2025.14427

The LSTM-based forecasts achieved high R2 scores exceeding 0.90 and lower MAE compared to baseline methods, confirming the system's advanced predictive capabilities.

Screenshots of System Features

To further illustrate the implementation, the following screenshots demonstrate the system's interface and key unctionalities:

The system's landing page, shown in Fig. 2, introduces the sales prediction system with a clean and responsive design. Users can upload datasets for analysis and prediction using the feature illustrated in Fig 3. The results from the LSTM-based model, shown in Fig. 4, provide insights into future sales trends. Fig. 5 visualizes past sales data, highlighting historical trends. The predicted future sales, generated using advanced models, are displayed in Fig. 6.

Fig. 2. Home Page of the Sales Prediction System.

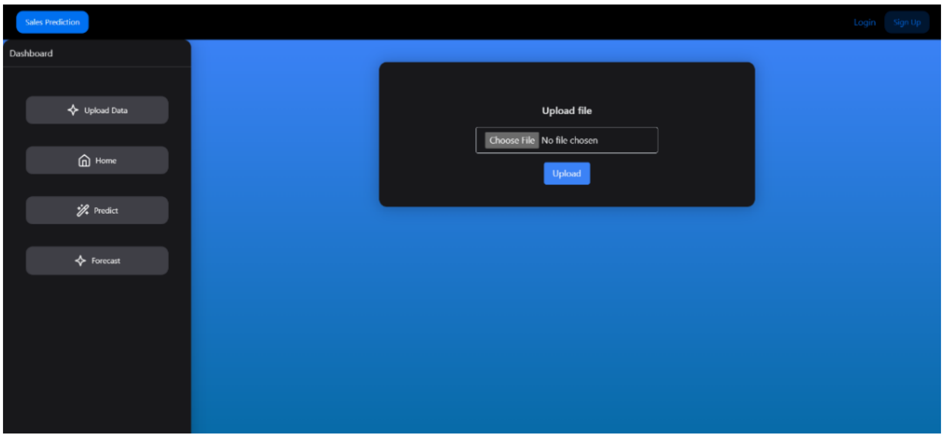

Fig. 3. Data Upload Page.

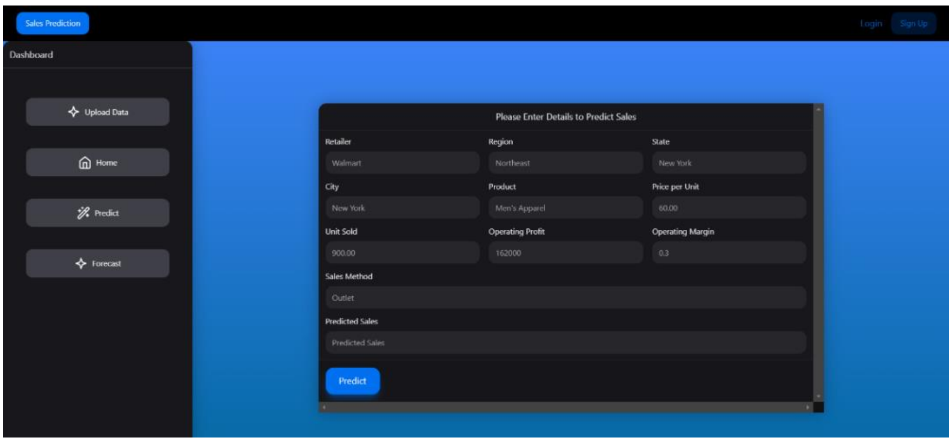

Fig. 4. LSTM-Based Future Sales Prediction Results.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 4, April 2025

DOI: 10.17148/IJARCCE.2025.14427

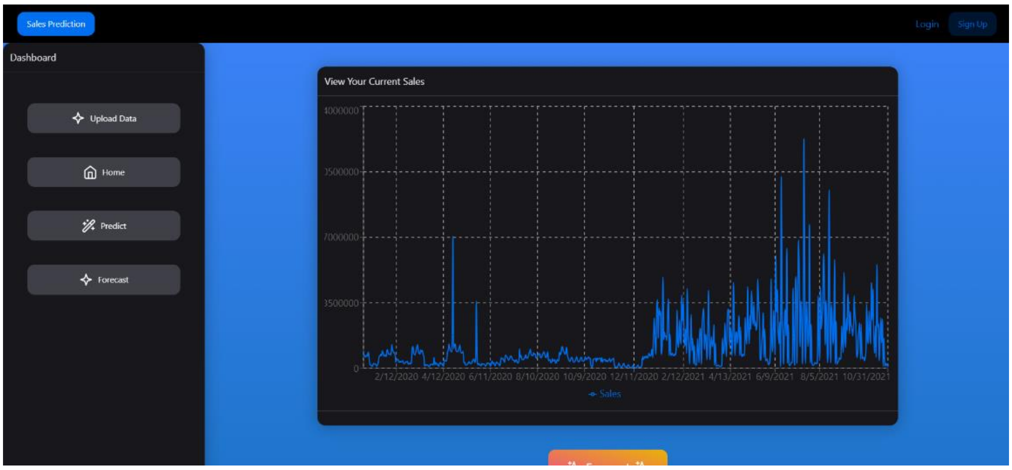

Fig. 5. Visualization of Past Sales Data.

Fig. 6. Visualization of Predicted Future Sales.

VI. CONCLUSION

In this paper, we have presented a comprehensive system for real time sales prediction and inventory management, leveraging advanced machine learning algorithms such as LSTM and XGBoost, supported by a robust architecture that integrates multiple layers for data processing, storage, and user interaction. The proposed system offers a scalable, accurate, and user-friendly solution, addressing key challenges faced by businesses in dynamic markets. The implementation results demonstrate the efficacy of the system in predicting future sales trends, visualizing data, and enabling in formed decision-making. Furthermore, the integration of cloud infrastructure and real-time processing frameworks ensures that the system remains responsive and adaptable to varying data volumes and complexities. While the system showcases significant strengths, challenges such as ensuring data quality, addressing computational constraints, and adapting to unforeseen market dynamics remain areas for future improvement. By refining these aspects and incorporating additional external factors like economic indicators, the system can further enhance its predictive power and usability. Overall, the system contributes to the field of sales forecasting by offering a practical and innovative approach, with potential applications across a wide range of industries seeking to optimize their operations and improve their competitive advantage.

REFERENCES

[1]. MD Tanvir Islam, Eftekhar Hossain Ayon, Bishnu Padh Ghosh, MD, Salim Chowdhury, Rumana Shahid, Aisharyja Roy puja, Sanjida Rahman, Mohammad Shafiquzzaman Bhuiyan, & Tuan Ngoc Nguyen. (2024). "Revolutionizing Retail: A Hybrid Machine Learning Approach for Precision Demand Forecasting and Strategic Decision-Making in Global Commerce". *Journal of Computer Science and Technology Studies*, *6*(1), 33-39. https://doi.org/10.32996/jcsts.2024.6.1.4

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 4, April 2025
DOI: 10.17148/IJARCCE.2025.14427

[2]. Gao, Yifang & Wang, Junwei & Li, Zhi & Peng, Zengjun. (2023). "The Social Media Big Data Analysis for Demand Forecasting in the Context of Globalization". Journal of Organizational and End User Computing. 35. 10.4018/JOEUC.325217.

[3]. Rui, G., & Li, M. (2024). "Utilizing Internet Big Data and Machine Learning for Product Demand Forecasting". *Tehnički Vjesnik.* https://doi.org/10.17559/TV-20240318001408.

[4]. Khan, M. A., et al. (2020). "Effective Demand Forecasting Model Using Business Intelligence Empowered with Machine Learning". *IEEE Access*, vol. 8, pp. 116013-116023. https://doi.org/10.1109/ACCESS.2020.3003790.

[5]. Nassibi, N., et al. (2023). "Demand Forecasting Models for Food Industry by Utilizing Machine Learning Approaches". *International Journal of Advanced Computer Science and Applications*, 14(3).

[6]. Sridhar, A. V., & Mohan, G. V. (2024). "Forecasting New Product Demand Using Machine Learning". *International Journal of Modern Engineering Research (IJMER)*, 14(4), Jul.-Aug. 2024.

[7]. Aci, M., & Yergök, D. (2023). "Demand Forecasting for Food Production Using Machine Learning Algorithms: A Case Study of University Refectory". *Tehnički Vjesnik.* https://doi.org/10.17559/TV-20230117000232.

[8]. Krishnamoorthy, K., et al. (2024). "Machine Learning for Demand Forecasting in Manufacturing". *International Journal for Multidisciplinary Research*, 6(1). https://doi.org/10.36948/ijfmr.2024.v06i01.14204.

[9]. Qureshi, N. U. H., et al. (2024). "Demand Forecasting in Supply Chain Management Using Weather Enhanced Deep Learning Model". *IEEE Access*, vol. 12, pp. 145570-145581. https://doi.org/10.1109/ACCESS.2024.3472499.

[10]. Mitra, A., et al. (2022). A Comparative Study of Demand Forecasting Models for a Multi-Channel Retail Company. *Oper. Res. Forum*, vol. 3, no. 58. https://doi.org/10.1007/s43069-022-00166-4.Patel, R., & Sharma, A. (2022). "Hybrid Demand Forecasting Model Using Big Data Analytics for Retail Sales Prediction." *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(1), pp. 158-169. DOI: 10.1109/TETCI.2022.3180402.

[11]. Kim, J., Choi, B., & Song, H. (2023). "Enhancing Demand Forecasting Accuracy with a Neural Network Ensemble Model in Retail." *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(2), pp. 489-499. DOI: 10.1109/TSMC.2023.3210975.

[12]. Mohanty, S., Das, P., & Singh, K. (2023). "Forecasting Demand with Seasonal Trends Using LSTM and Random Forest Algorithms." *IEEE International Conference on Data Science and Advanced Analytics* (DSAA), pp. 102-108. DOI:10.1109/DSAA.2023.9260173.

[13]. Kilimci, Zeynep Hilal, Akyuz, A. Okay, Uysal, Mitat, Akyokus, Selim, Uysal, M. Ozan, Atak Bulbul, Berna, Ekmis, Mehmet Ali, "An Improved Demand Forecasting Model Using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain", *Complexity*, 2019, 9067367, 15 pages, 2019. https://doi.org/10.1155/2019/9067367

[14]. Jamarani, A., Haddadi, S., Sarvizadeh, R. et al. "Big data and predictive analytics: A systematic review of applications". Artif Intell Rev 57, 176 (2024). https://doi.org/10.1007/s10462-024-10811-5

