ACADEMIA International Journal for Social Sciences
Volume 4, Issue 4, 2025 ISSN-L (Online): 3006-6638

Business Analytics and Machine Learning-Based Demand Forecasting and Inventory Optimization in the Grocery Supply Chain

MD. Muhibur Rahman Sabbir
rsamuhib@gmail.com
University of Skill Enrichment and Technology Bachelor of Business Administration

Shammi Akhter
shammirahman@gmail.com
Government College of Applied Human Science Art and Creative Studies

Mohammad Siamul Islam
siamuls23@yahoo.com
HUNAN UNIVERSITY School of Economics and Trade
Corresponding Author: * MD. Muhibur Rahman Sabbir rsamuhib@gmail.com
Received: 16-08-2025 **Revised:** 26-09-2025 **Accepted:** 25-10-2025 **Published:** 24-11-2025

ABSTRACT

This paper offers a detailed solution to the use of machine learning methods in demand forecasting and inventory optimization of the grocery supply chain. Random Forest (RF) and Extreme Gradient Boosting (XGBoost) models were used to estimate the demand of short-term products based on historical data of sales and inventory. Mean Absolute Error (MAE) was used to measure model performance to determine that RF delivered slightly more accurate forecasts than XGBoost. The projected demand was then used to optimise essential influent factors in inventory, such as reorder points and days of inventory, to be able to obtain sufficient in-store inventory, but with a minimal holding cost. The findings indicate that machine learning-based predictions may go a long way to improve inventory management, minimize stockouts, and make operations more efficient. These results provide practical suggestions to grocery stores that are interested in using the data-guided decision-making as a tool to improve their supply chain operations.

Keywords: *Demand Forecasting, Inventory Optimization, Random Forest, XGBoost, Grocery Supply Chain, Machine Learning, Predictive Analytics*

INTRODUCTION

The grocery supply chain is one of the most sensitive and complicated parts of the world economy because of the high perishability of the products, dynamically changing demand and the emerging demands of customers to deliver the products in time and be available. Smooth demand forecasting and inventory optimization have thus become the cornerstones of the performance of the supply chain, which directly affect profitability, customer satisfaction and sustainability. Nonlinear effects, seasonality, and volatility in demand data of a grocery business are, however, usually not adequately represented by traditional techniques: moving averages, exponential smoothing, and regression-based models. [1, 12].

New platforms have been created by the emergence of business analytics and machine learning (ML) to deal with these challenges. Through big historical and transactional data, the ML algorithms would be able to surface concealed trends, adjust to market dynamics, and present more precise and dynamic predictions. This development is a paradigm change in reactive decision making to proactive and predictive supply chain management [3, 7]. In that regard, business analytics will not only underpin decisions made on a basis of data but also combine statistical information in that of predictive intelligence to be able to plan inventory and procurement with more efficiency.

The recent years saw a rise in the use of machine learning-based tools like Random Forest (RF), Gradient Boosting, and XGBoost in organizations to increase supply chain responsiveness. The algorithms have proved to be superior in terms of non-linear relationship and high-dimensional data that are prominent in grocery and retail market. [10, 2]. An example of this is that the ensemble learning mechanism of RF is a less overfitting predictor of demands and a more consistent predictor than XGBoost due to its application of gradient boosting and regularization, which has outstanding scalability and accuracy. These models together with the integration into the business analytics frameworks allow supply chain professionals to forecast future demand, establish optimal reorder points, and ensure the right amount of inventory.

Grocer supply chain presents however unique analytical problems. Seasonality, promotions, product life cycles and consumer behavior are some of the factors that affect demand patterns. Also, inventory decisions are complicated by the perishability limits, reliability of suppliers as well as delays in logistics. [4, 14]. The problems require advanced models, which can utilize a wide range of data, such as past sales, supplier measures, lead time, and inventory to produce accurate demand projections and optimize replenishment policies.

The machine learning models have been found to be better than the classical methods in this field. Studies such as those by Jahin et al. [8] and Qureshi et al. [13] have highlighted the integration of big data and ML to enhance demand forecasting accuracy in retail and e-commerce environments. Similarly, Juma et al. [9] and Pasupuleti et al. [11] focused on predictive analytics and optimality based on sustainability in enhanced inventory management. These methods enable real-time decision-making by linking data analytics to operational objectives such as waste reduction, minimization of holding costs, and stockout prevention.

The issue of inventory management is important in the area of grocery retailing since even minor errors in the forecasting process can cause a significant amount of financial damages or even spoilage. A combination of the use of business analytics and ML techniques will offer a solid resolution of this issue and offer a statistical insight, a model of optimization, and a predictive algorithm of the whole supply chain [18, 19]. Grocery retailers will be able to balance the availability of products and the costs incurred by combining data-driven prediction and inventory management.

Thus, the current research aims at the creation of an integrated business analytics system with the help of machine learning to foretell the demand and optimal inventory performance within the grocery supply chain. Particularly, the models of Random Forest and XGBoost are used to process historical data and predict such main metrics as demand in the next period and days of inventory. Model performance is evaluated using standard measures of evaluation such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and the coefficient of determination (R2). The task is to create a strong predictive model that will not only enhance predictive accuracy but will also inform more effective inventory and replenishment measures and be able to make the supply chain more resilient and agile overall.

The rest of this paper will be organized in the following way: Section 2 will review recent literature on machine learning use in demand forecasting and inventory management. Section 3 provides the methodological framework, that is, data description, model development, and evaluation metrics. Section 4 receives and discusses the experimental results whereas Section 5 provides the conclusion of the work with the managerial implications and future research directions.

LITERATURE REVIEW

The recent literature proves that there is a growing interest in implementing machine learning alongside supply chain management and forecasting. Aamer et al. [1] conducted a review of the uses of data analytics in the supply chain systems, with a focus that the ML algorithms improve the accuracy of forecasting and minimize the uncertainty of its operations. On the same note, Punia and Shankar [12] came up with an idea of a deep learning-based decision support system that was much more effective in forecast accuracy than the traditional technique.

There are a number of studies, which examine the use of ML in optimization of multi-objective supply chain. Abdi et al. [2] created an extensive machine learning framework that deals with the multi-product and multi-period demand forecasting showing the usefulness of the hybrid optimization models. Douaioui et al. [7] critically reviewed supply chains forecasting deep learning models and were able to conclude that ensemble and hybrid ML models perform better than single-algorithm models in volatile demand environments.

Aichner and Santa [3] pointed out in the retail and fast-moving consumer goods (FMCG) industry that the use of ML is increasingly becoming a vital solution when it comes to enhancing the accuracy of demand forecasts in the short term and help firms to more accurately match the level of inventory with the consumer demand. Rakholia et al. [14] also proposed a solution based on data optimization of inventory using predictive analytics, demonstrating that the approach has better stock management and lower logistics expenses. Seyedan [15] went ahead to extend this discussion by coming up with predictive analytics of inventory forecasting that combine both structured and unstructured data sources.

In this case, Chowdhury et al. [6] analyzed the existing forecasting models in the e-commerce retail industry, including XGBoost and Random Forest as the top forecasting models on nonlinear and multi-variable data. Nassibi et al. [10] and Thejasree et al. [18] also supported the similar finding focusing on the superiority of ML in food supply chains because of its flexibility to the volatile consumption trends.

Supply chain intelligence is also taking a centre stage in frameworks that have been driven by big data. One of the studies was by Jahin et al. [8], who suggested a demand forecasting architecture based on big data and combined preprocessing and feature selection methods to enhance the efficiency of the ML models. Similarly, Juma et al. [9] also presented the ability of predictive analytics and IoT information to turn demand and inventory management into large retail networks. The authors of the article by Arvind et al. [4] presented an analogous model of smart warehousing on the basis of IoT and ML to optimize inventory accurately, demonstrating its efficiency in minimizing delays in the operation process and enhancing responsiveness of a warehouse.

Pasupuleti et al. [11] investigated the importance of machine learning in logistic and inventory optimization of a supply chain in order to ensure sustainability in the supply chain, found that data-driven optimization helps lower the cost and carbon footprint. Sharma et al. [16] and Stanelyte. The validity of the arguments that ML can be used to enhance inventory policies was also supported by [17] by connecting the models of demand prediction with the models of replenishment choices.

Altogether, these works form the solid background of implementing ML-based business analytics to grocery supply networks. In spite of these improvements, there are few studies that have concerned the creation of aligned frameworks of simultaneous demand prediction and optimizing inventory in the e-grocery industry. The given research paper fills this gap as RF and XGBoost are applied to real-life grocery data that is supposed to maximize forecast accuracy and optimize inventory management decisions based on the data-driven business analytics paradigm.

METHODOLOGY

The main aim of the study is to predict demand and optimize inventory on the grocer supply chain with machine learning. Random Forest (RF) and XGBoost (XGB) regression models were used to forecast demand and predicted demand was used to make decisions on inventory management. The process is made up of preprocessing of data, model development, assessment, and optimization of inventory.

Data Preprocessing

Before using the machine learning models, the data had been preprocessed thoroughly to assure the quality of data and model preparation. First, every numeric column was transformed to the right numeric type and missing values have been filled with zeroes to preserve the integrity of the data sets. Categorical features such as Category, Supplier Name, and Warehouse Location were encoded through the use of Label Encoding so as to enable suitability with the regression models. The input variables have been carefully chosen in terms of their importance to demand forecasting and inventory optimization, the demand predictive variable was Forecast Next 30d, and the variables pertaining to inventory optimization were Days of Inventory and Reorder Point. At last, the dataset was divided into training and testing segments with 80 percent of the data going to train the models and 20 percent used to test the predictive power of the models.

Random Forest Regression

Random Forest is an ensemble learning method based on decision trees. It reduces variance by averaging predictions across multiple trees. For a dataset with N samples and T decision trees, the prediction for the i-th sample is:

$$y_i = \frac{1}{T}\sum_{t=1}^{T} h_t(x_i)$$

In the Random Forest model, the prediction for a given sample x_i is obtained by aggregating the outputs of multiple decision trees. Specifically, if $h_t(x_i)$ represents the prediction of the t-th tree, the final aggregated prediction $\hat{y}_i$ is computed as the average of all individual tree predictions. The performance and behavior of the model are influenced by several hyperparameters, including the number of trees (*n estimators*), the maximum depth of each tree (*max _depth*), the minimum number of samples required to split an internal node (*min _samples split*), and the minimum number of samples required to form a leaf node (*min samples leaf*). Careful tuning of these hyperparameters ensures optimal predictive accuracy and prevents overfitting.

XGBoost Regression

XGBoost is a gradient boosting framework that sequentially builds decision trees to minimize a loss function. For regression, the objective function is:

$$L = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 + \sum_{k=1}^{K} \Omega(f_k)$$

In the XGBoost model, the objective function is designed to minimize both the prediction error and the complexity of the model. Here, y_i denotes the true value of the target variable, while $\hat{y}_i$ represents the predicted value. The term $\Omega(f_k)$ corresponds to the regularization function for the k-th tree, and K is the total number of trees in the ensemble. By incorporating this regularization term, the model is able to balance predictive accuracy with complexity, effectively preventing overfitting and improving generalization to unseen data.

Evaluation Metric

The Mean Absolute Error (MAE) was used to determine the performance of the forecasting models. MAE uses the mean size of the difference between the approximated values and the actual target values and gives a clear picture on the accuracy of the model and is not very sensitive to outliers. Formally, MAE is defined as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

where y_i represents the true value of the target variable, $\hat{y}_i$ is the predicted value, and N is the total number of observations in the dataset. Using MAE allows for straightforward interpretation of the average forecast error in the context of inventory and demand predictions.

Inventory Optimization

The predicted demand obtained from the Random Forest (RF) and XGBoost models forms the basis for optimizing inventory within the grocery supply chain. Inventory optimization focuses on maintaining adequate stock levels to meet customer demand while minimizing holding costs and reducing the risk of stockouts, aligning with the principles of business analytics in supply chain management. Two key metrics are calculated for this purpose: The Reorder Point and Days of Inventory. The Reorder Point is computed as

$$\text{Reorder\_Point} = \text{Safety\_Stock} + (\text{Avg Daily Sales} \times \text{Lead\_Time Days}),$$

ensuring that sufficient stock is available during supplier lead times. Meanwhile, Days of Inventory, defined as

$$\text{Days of Inventory} = \frac{\text{Quantity On Hand}}{\text{Avg Daily Sales}},$$

provides a measure of how long current stock can satisfy predicted demand. These calculations enable data-driven decisions for inventory management, directly supporting the goals of business analytics and machine learning-based demand forecasting in enhancing efficiency and responsiveness in the grocery supply chain.

The methodology integrates machine learning demand forecasting with practical inventory optimization. RF and XGBoost provide robust predictions, while the Mean Absolute Error (MAE) quantifies model performance. Inventory metrics derived from predicted demand support data-driven decision-making for reorder points and stock management, ensuring sufficient inventory levels while minimizing holding costs in the grocery supply chain.

RESULTS AND DISCUSSION

Demand Forecasting Results

The results of the Random Forest (RF) and the XGBoost (XGB) models were compared based on the mean absolute error (MAE) of the demand within 30 days of prediction. RF was the one that scored slightly lower on MAE (97.45) than XGB (109.66), which shows that RF was slightly more accurate in the demand prediction.

Table 1: MAE for Demand Forecasting Models

Model	MAE
Random Forest	97.45
XGBoost	109.66

Figure 1 and Figure 3 present the actual versus predicted demand for RF and XGB, respectively. It is observed that both models capture the overall demand trends, although RF predictions follow the actual values more closely. Feature importance plots (Figures 2 and 4) indicate that Avg Daily Sales, Lead Time Days, and Safety Stock are the most influential features, aligning with practical supply chain considerations [1, 2].

Inventory Optimization Results

The predicted demand from RF and XGB models was then used to optimize inventory by calculating Reorder Points and Days of Inventory. The inventory MAE is summarized in Table 2, showing that XGBoost achieved slightly better performance (MAE = 0.76) compared to RF (MAE = 0.95). Lower MAE indicates that XGB provides more precise inventory predictions, supporting efficient stock management and reducing the risk of stockouts.

Figure 5 illustrates the actual inventory days alongside predictions from RF and XGB models. The figure demonstrates that both models capture inventory trends effectively, with XGB predictions following the actual inventory levels more closely. This indicates

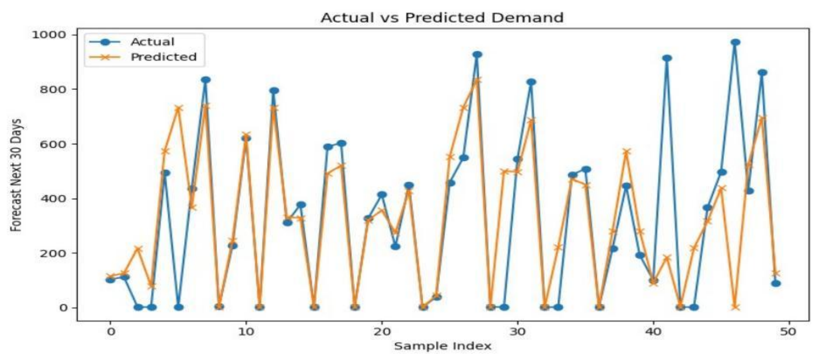

Figure 1: Random Forest: Actual vs Predicted Demand

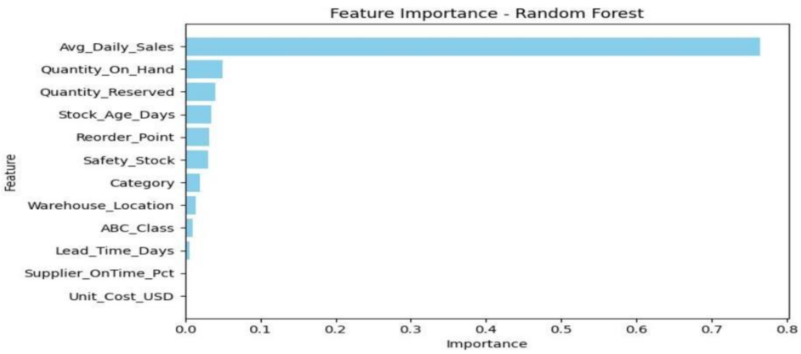

Figure 2: Random Forest: Feature Importance Table 2: MAE for Inventory Optimization Predictions

Model	MAE
Random Forest	0.95
XGBoost	0.76

that ensemble-based machine learning models can support data-driven inventory optimization decisions in grocery supply chains [4, 19, 14].

Overall, the results suggest that while RF slightly outperforms XGB for demand forecasting, XGB is better at optimization of inventory. The importance of daily sales, lead time, and safety stocks in demand forecasting as well as optimal inventory management underline the essential importance of the feature importance analysis in accordance with the best practices in the analytics of a supply chain.

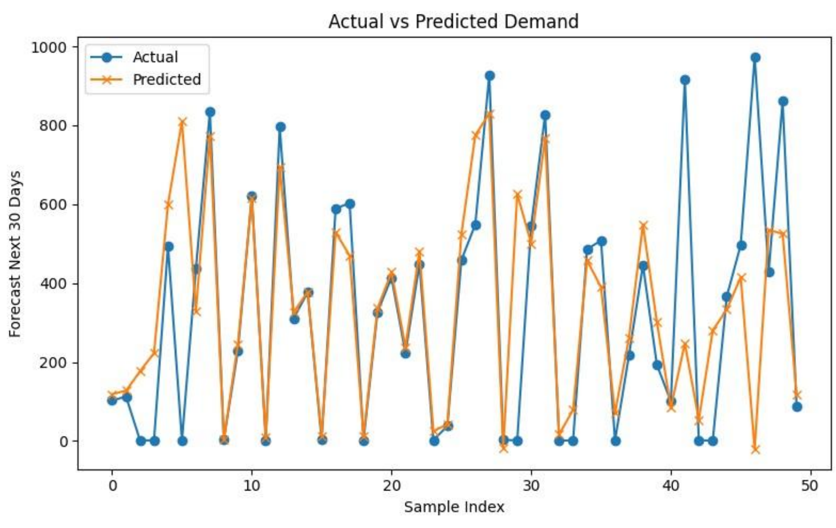

Figure 3: XGBoost: Actual vs Predicted Demand

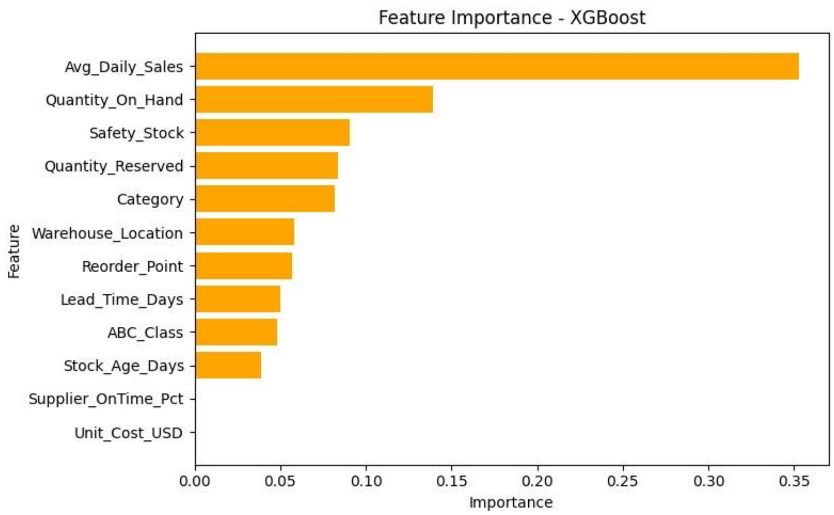

Figure 4: XGBoost: Feature Importance

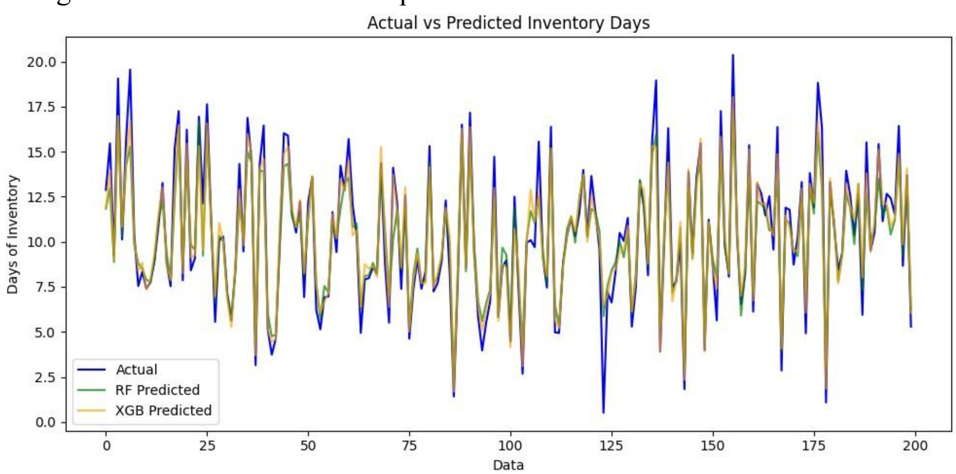

Figure 5: Actual vs Predicted Inventory Days using RF and XGBoost

CONCLUSION

This research proves how machine learning methods can be successfully integrated with inventory management strategies to improve the work of the supply chains in the grocery market segment. Through the use of the Random Forest (RF) and XGBoost (XGB) models, the study is able to give the correct demand forecasts in the past 30 days, which is used as the basis of making inventory optimization decisions. The comparison based on Mean Absolute Error (MAE) reveals that RF is more successful in demand forecast (MAE = 97.45), but XGB is more accurate in stock-related predictions (MAE = 0.76), which is the strength of complement of these models.

The importance of features analysis indicated that the variables Avg Daily Sales, the Days of Lead Time, and Safety Stock are important inputs to the formation of the demand and inventory, and they can be used by the managers to make their businesses focus on critical factors that positively influence the efficiency of the supply chain. The calculated inventory measures, such as Reorder Point and Days of Inventory, help companies to keep the optimal stock level, lower the holding costs, and decrease the risk of stockout.

On the whole, the presented methodology is a useful and evidence-based way of making decision-making in grocery supply chains more effective. RF and XGB together enable the business to precisely predict demand and at the same time manage inventory in the best possible way to achieve better operational efficiency, utilization of resources and responsiveness to market forces. The future study can build on this framework by adding real-time sales forecasts, multi-level supply chains and other machine learning models to further enhance accuracy of forecasts and inventory control.

Data Source

The dataset is publicly available on Kaggle: https://www.kaggle.com/datasets/mustofaahmad/ inventory-management-grocery-industry.

References

[1] Aamer, A., Eka Yani, L., & Alan Priyatna, I. (2020). Data analytics in the supply chain management: Review of machine learning applications in demand forecasting. *Operations and Supply Chain Management: An International Journal*, 14(1), 1–13.

[2] Abdi, F., Abolmakarem, S., Yazdi, A. K., Popescu, V., Birau, R., & Mitroi, A. T. (2025). A machine learning-based approach for multi-objective, multi-product, and multi-period supply chain optimization via demand forecasting. *IEEE Access*.

[3] Aichner, T., & Santa, V. (2023). Demand forecasting methods and the potential of machine learning in the FMCG retail industry. In *Serving the customer: The role of selling and sales* (pp. 215–252). Wiesbaden: Springer Fachmedien Wiesbaden.

[4] Arvind, V. R., Shrinidhi, R. M., Deepa, T., & Maheedhar, M. (2025). Intelligent warehousing: A machine learning and IoT framework for precision inventory optimization. *IEEE Access*.

[5] Chidiebube, I. N., Onyeka, N. C., Sunday, A. P., & Chiedu, E. O. (2025). A comparative analysis of machine learning models for inventory demand forecasting in a food manufacturing SME. *Indonesian Journal of Innovation Science and Knowledge*, 2(3), 35–48.

[6] Chowdhury, A. R., Paul, R., & Rozony, F. Z. (2025). A systematic review of demand forecasting models for retail e-commerce enhancing accuracy in inventory and delivery planning. *International Journal of Scientific Interdisciplinary Research*, 6(1), 1–27.

[7] Douaioui, K., Oucheikh, R., Benmoussa, O., & Mabrouki, C. (2024). Machine learning and deep learning models for demand forecasting in supply chain management: A critical review. *Applied System Innovation (ASI)*, 7(5).

[8] Jahin, M. A., Shovon, M. S. H., Shin, J., Ridoy, I. A., & Mridha, M. F. (2023). Big data–supply chain management framework for forecasting: Data preprocessing and machine learning techniques. *arXiv preprint arXiv:2307.12971*.

[9] Juma, B. K., Jain, R., Cooper, A. F., Zulu, R., Palaniappan, D., Parmar, K. J., & Premavathi, T. (2026). Big data in supply chain management demand forecasting and inventory optimization. In *Transforming sustainability and value creation with logistics and supply chain management* (pp. 1–28). IGI Global Scientific Publishing.

[10] Nassibi, N., Fasihuddin, H., & Hsairi, L. (2023). Demand forecasting models for food industry by utilizing machine learning approaches. *International Journal of Advanced Computer Science and Applications*, 14(3), 892–898.

[11] Pasupuleti, V., Thuraka, B., Kodete, C. S., & Malisetty, S. (2024). Enhancing supply chain agility and sustainability through machine learning: Optimization techniques for logistics and inventory management. *Logistics*, 8(3), 73.

[12] Punia, S., & Shankar, S. (2022). Predictive analytics for demand forecasting: A deep learning–based decision support system. *Knowledge-Based Systems*, 258, 109956.

[13] Qureshi, N. U. H., Javed, S., Javed, K., Naqvi, S. M. R., Raza, A., & Saeed, Z. (2024). Demand forecasting in supply chain management for Rossmann stores using weather enhanced deep learning model. *IEEE Access*.

[14] Rakholia, K. R., Chandraprabh, C., Ramesh, R., Rao, K., Punitha, S., & Kumar, M. (2024). Optimizing inventory management through demand forecasting: A datadriven approach for enhanced supply chain efficiency. *Guru Vimal (November 15, 2024)*.

[15] Seyedan, S. (2023). Development of predictive analytics for demand forecasting and inventory management in supply chain using machine learning techniques (Doctoral dissertation, Concordia University).

[16] Sharma, H., Sahu, N., Singhal, R. K., Garg, A., Singhal, R., & Tripathi, S. (2024, August). Data-driven forecasting and inventory optimization using machine learning models and methods. In *2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET)* (pp. 1–6). IEEE.

[17] Stanelyte˙, G. (2021). Inventory optimization in retail network by creating a demand prediction model (Doctoral dissertation, Vilniaus Gedimino technikos universitetas).

[18] Thejasree, P., Manikandan, N., Vimal, K. E. K., Sivakumar, K., & Krishnamachary, P. C. (2023). Applications of machine learning in supply chain management—A review. In *Industry 4.0 Technologies: Sustainable Manufacturing Supply Chains: Volume 1—Theory, Challenges, and Opportunity* (pp. 73–82).

[19] Yusof, Z. B. (2024). Analyzing the role of predictive analytics and machine learning techniques in optimizing inventory management and demand forecasting for ecommerce. *International Journal of Applied Machine Learning*, 4(11), 16–31.