Enhancing Supply Chain Efficiency through Advanced Machine Learning Models for Demand Forecasting and Queuing Analysis

DOI: 10.26573/2026.20.1.3
Volume 20, Number 1
January 2026, *pp.* 31-48

Anshu Rathi
Geetanjali Sharma
Banasthali Vidyapith
(anshurathi0569@gmail.com)
(geetanjali.bu@gmail.com)

Supply chain optimisation (SCO) is essential for enhancing operational efficiency, reducing costs, and ensuring sustainability. With the increasing complexity of global supply chains, traditional approaches probably fail to address dynamic challenges, necessitating advanced tools like machine learning (ML). While ML has transformed demand forecasting and bottleneck analysis, limited research integrates these aspects for comprehensive supply chain management, creating a significant research gap. This study evaluates the performance of Random Forest, Gradient Boosting, and Linear Regression models for demand forecasting and queuing analysis. We have analysed the data which have been collected from the Kaggle dataset. Model optimisation is performed using hyperparameter tuning, with performance validated through metrics including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 score. Results show that Random Forest achieved the lowest MAE (24.54) and RMSE (44.69) with the highest R^2 (0.86) for demand forecasting, followed by Gradient Boosting (MAE: 33.72, RMSE: 47.72, R^2: 0.84.

Keywords: Demand Forecasting, Queuing Delays, Machine Learning, Random Forest, Gradient Boosting, Supply Chain Optimisation.

1. Introduction

Supply chain optimization (SCO) is a critical aspect of modern business operations, focusing on improving the efficiency, performance, and sustainability of supply chains. Since world markets are constantly in stiff competition, efficient management of a supply chain becomes an operational excellence tool as well as a way of adapting to dynamic customer requirements (Geunes & Pardalos, 2007). It basically constitutes the application of strategic planning together with using advanced optimization methods, for which objectives include cost reduction, decreased lead times, improvement of customer satisfaction levels, as well as being eco-friendly.

Supply chain networks optimization enables better decision-making concerning key functions such as production, transportation, and stock management. Studies have shown that advmodelslgorithms, including genetic algorithms and mixed-

integer linear programming models, have the greatest impact on scheduling and resource allocation in distributed manufacturing systems when they are complex (Beheshtinia et al., 2021). Moreover, real-time optimization approaches, such as Dynamic Real-Time Supply Chain Optimization with Algorithmic Intelligence (DRT-SCAI), allow organizations to respond rapidly to disruptions, thus enhancing their resilience and predictive capabilities (Lalitha et al., 2024). Green supply chain design and closed-loop networks have been of top priority in reducing carbon footprints while maintaining service levels up (Nurjanni et al., 2017).

It is the effective management of the supply chain that directly affects financial performance by optimizing the use of resources and avoiding waste. According to research, supply chain performance is very much related to financial outcomes and mediates SCM in attaining cost reduction and profit maximization (Agus, 2013). In addition, hybrid approaches combining simulation with optimization tools are useful in decision-making processes for risk-prone scenarios (Nikolopoulou & Ierapetritou, 2012). Data analytics and ML have transformed SCM to make it possible to optimize organizations so that they can work according to the dynamic market demands. It also offers a way of extracting practical insights from complex data that help in decision-making functions involved in the supply chain.

Future research stresses transparency in algorithms and ethical AI practices to mitigate the associated challenges (Nzeako et al., 2024). Data analytics and machine learning are transforming supply chain management by enabling smarter, faster, and more accurate decisions. Their application in different areas, such as forecasting, risk management, logistics, and fraud detection, explains their critical role in helping global supply chains be effective and resilient.

Despite advancements in machine learning applications, research on integrating demand forecasting with queuing dynamics for comprehensive supply chain optimization remains sparse. Existing studies often neglect the interplay between predictive accuracy and real-time bottleneck management, limiting their practical utility. This study aims to bridge the gap by employing advanced machine learning models to forecast demand and predict queuing delays. The objective is to evaluate their predictive capabilities, identify critical bottleneck factors, and propose actionable strategies for optimizing inventory flow and service rates in supply chains.

The remainder of the paper is organized as follows. Section 2 presents the literature review. Section 3 explains the methodology, Exploratory Data Analysis, and Model development. The results and discussion are in Section 4. Section 5 has the conclusion.

2. Literature Review

The integration of advanced technologies, as summarised in Table 1, such as predictive analytics, machine learning, and real-time optimisation, is reshaping supply chain management, offering significant improvements in efficiency, decision-making, and risk management. Predictive models and machine learning algorithms enhance demand forecasting, inventory management, and fraud detection, while real-time data analytics facilitate rapid responses to disruptions. However, challenges related to data quality, integration complexities, and high computational demands remain significant barriers. Furthermore, the implementation of ethical AI and data

governance frameworks is crucial to ensure privacy, security, and regulatory compliance. Future research should focus on overcoming these challenges to further optimize supply chain operations.

Table 1 *Overview of Emerging Technologies and Their Applications in Supply Chain Management*

Study	Description	Benefits	Challenges	Tools/Technologies	References
Predictive Analytics	Use of predictive models to forecast demand, optimize inventory, and predict disruptions, improving overall Supply chain efficiency.	Improved accuracy in demand forecasts and reduced costs.	Data quality and integration with existing systems.	Predictive models, ML algorithms, Big Data	(Aljohani, 2023)
Machine Learning Applications	Use of algorithms like neural networks and support vector machines for demand forecasting, inventory optimization, and logistics planning.	Enhanced decision-making and operational efficiency.	Need for skilled personnel and potential algorithmic biases.	Neural Networks, SVM, LSTM	(Aamer et al., 2020), (He & He, 2024)
Automated Fraud Detection	Application of machine learning frameworks for Identifying and mitigating fraudulent activities within the supply chain ecosystem.	Increased transparency and Security operations.	Data sensitivity and computational resource needs.	Automated ML Frameworks, Anomaly Detection Models	(Lokanan & Maddhesia, 2024)
Real-time Optimization	Integration of Time data analytics for instant decision-making and adaptive responses to supply chain disruptions.	Faster response times and better agility in supply chain processes.	High computational requirements and cost of implementation.	IoT, Dynamic Optimization Tools	(Pasupuleti et al., 2024), (Brintrup et al., 2020)
Hybrid Algorithms and Deep Learning	Use of hybrid machine learning algorithms and deep learning techniques to optimize logistics, reduce costs, and improve decision-making accuracy.	Greater computational efficiency and improved forecasting accuracy.	Complexity in implementation and need for high computational power.	Hybrid ML Algorithms, Deep Learning	(Ghodake et al., 2024), (Tirkolaee et al., 2021)
Big Data Integration	Utilization of	Better insights and improved	High storage and processing	Big Data Platforms, Data Lakes	(Raviteja et al., 2022),

	large- scale data from diverse sources to improve forecasting, inventory management, and customer service.	customer service.	requirements.		Patnayak et al., 2023)
Advanced Predictive Analytics	Machine learning Applications in ecommerce SCM, Including real-time inventory management and dynamic customer experience personalization.	Enhanced responsiveness and customer satisfaction.	Cost of implementation for smaller businesses.	Predictive Analytics, Personalized AI Frameworks	(Tuli et al., 2024)
Ethical AI and Data Governance	Emphasis on ensuring data security, transparency,	Builds trust and ensures regulatory compliance.	Balancing privacy concerns with technological advancements.	Data Governance Frameworks, Ethical AI Models	(Nzeako et al., 2024)
Predictive Analytics for Risk Mitigation	Predictive models to anticipate supply chain risks and disruptions.	Enhanced resilience and proactive risk management.	Difficulty in capturing unforeseen risks.	Predictive Analytics, Anomaly Detection	(Aljohani, 2023)
Prescriptive Analytics in Smart Manufacturing	Optimization and simulation models to prescribe best actions for production and logistics.	Actionable insights and improved operational efficiency.	High computational complexity and model interpretability.	Optimization Models, Simulation, Heuristics	(Pascual et al. 2019)
Predictive Analytics in Finance	Forecasting customer behavior and financial outcomes using historical data.	Better investment decisions and fraud detection.	Model overfitting and bias in training data.	Regression, Decision Trees, Ensemble Models	Siegel, E**.** (2016).
Prescriptive Analytics in Logistics	Simulation and decision analysis to optimize delivery routes and resource allocation.	Reduced delays and improved delivery efficiency.	Sensitive to assumptions and scenario variability.	Simulation, Decision Trees, Heuristics	Tian, et al. (2023)
Predictive & Prescriptive Integration in SCM	Combined use of forecasting and optimization to guide inventory and queuing decisions.	Holistic decision-making and dynamic adaptability.	Integration complexity and real-time data requirements.	ML + Optimization, SimPy, Bayesian Optimization	Kumar & Garg (2018)

The applications of machine learning models in supply chain management are presented in Table 2, demonstrate their potential in enhancing various aspects of the supply chain, such as demand forecasting, inventory optimization, and supplier risk analysis. Models like Neural Networks (NN), Support Vector Machines (SVM), and Random Forest (RF) offer high accuracy in tasks like demand prediction and

disruption forecasting. Despite their strengths, challenges such as high computational costs, sensitivity to hyper parameters, and the risk of over fitting remain. Models like Long Short-Term Memory (LSTM) and Gradient Boosting Machines (GBM) excel in handling temporal and sequential data but require considerable computational resources. Meanwhile, simpler models like k-Nearest Neighbors (k-NN) and Decision Trees (DT) offer interpretability and efficiency but struggle with high-dimensional data or overfitting. Further research and advancements in these models are necessary to overcome these challenges and optimize their application in supply chain management. In Table 2 Machine learning models and their application in supply chain management are discussed.

Table 2 *Machine Learning Models and Their Applications in Supply Chain Management*

Machine Learning Model	Application in Supply Chain	Performance Metrics	Strengths	Challenges	References
Neural Networks	Demand forecasting, Inventory optimization, and transportation Management.	Accuracy: 85%-95% in demand forecasting	Handles complex patterns, robust for large datasets.	High computational cost and risk of over-fitting.	(Tirkolaee et al., 2021), (Pasupuleti et al., 2024)
Support Vector Machines	Supplier risk analysis, inventory classification, and demand prediction.	Accuracy: 80%-90% for classification tasks	Good for small and medium-sized datasets, effective in non-linear problems.	Computationally expensive for large datasets.	(Aamer et al., 2020), (Brintrup et al., 2020)
Long Short-Term Memory (LSTM)	Time series forecasting for demand, lead time prediction, and inventory turnover analysis.	Accuracy: 87%-96% In lead time predication	Excels in handling sequentialand temporal data.	High training time Sequential and temporal data.	(Pasupuleti et al., 2024), (He &He ,2024)
Gradient Boosting Machines (GBM)	Predicting order delays, inventory replenishment and logistics optimization	Accuracy: 85%-92% in order delay prediction	High accuracy and robust overfitting with appropriate tuning.	Can be computationally intensive, prone to diminishing returns with added trees	(Ghodake et al., 2024), (Brintrup et al., 2020)
k-Nearest Neighbors (k-NN)	Customer clustering, supplier segmentation, and inventory and inventory categorization.	Accuracy: 75%-85% for clustering and segmentation tasks	Imple to implement, effective for small datasets.	Not suitable for high-dimensional datasets, sensitive to noisy data.	(Chandrasekaran et al. 2021), (Agarwal et al., 2023)
Decision Trees (DT)	Supply chain decision-making for inventory control,	Accuracy: 78%-88% in decision-	Easy to interpret, low computational	Prone to over-fitting without	(Okoduwa et al., 2024), (Brintrup et al., 2020)

	supplier selection, and routing.	making tasks	cost.	pruning.	
XG Boost	Used	Accuracy: 88%-94% in logistics optimization	High performance And scalability for large datasets.	High training Complexity and Sensitivity to parameter tuning.	(Tirkolaee et al. 2021) (He & He, 2024)
K-Means Clustering	Supplier clustering, customer segmentation, and distribution network design.	Clustering Accuracy: 75%-85%	Efficient for large datasets, effective for clear segmentation	Prone to local minima, sensitive to initial cluster assignment.	(Tuli et al., 2024), (Patnayak et al., 2023)

3. Methodology

The dataset for this study was obtained from Kaggle. This dataset provides structured data related to various supply chain activities, including sales, provisioning, production, and distribution. To ensure the dataset's reliability, comprehensive data cleaning was performed by removing incomplete, duplicate, and inconsistent entries. Missing values were either imputed or excluded based on their impact on the analysis. Feature engineering was conducted to derive critical variables such as "Sales per customer," which served as a proxy for demand forecasting, and "Days for shipping (real)" and "Days for shipment (scheduled)," which were used as metrics to analyze shipping delays. Additional features like "Late_delivery_risk," a binary indicator highlighting shipment delays, and categorical variables such as "Shipping Mode," "Category Name," and "Customer City" were encoded to enhance compatibility with machine learning algorithms. Numerical data was standardized to improve model performance, ensuring that all variables contributed effectively to the analysis.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was conducted to identify significant patterns and relationships within the dataset, facilitating a deeper understanding of the supply chain dynamics. Two correlation heatmaps are generated to illustrate the among key features.

Figure 1 highlights the correlation among primary predictors such as "Sales per Customer," "Product Price," and "Late Delivery Risk." Notably, a strong positive correlation was observed between "Sales per Customer" and "Product Price" ($r = 0.78$), indicating that higher-priced products tend to generate higher sales per customer. Conversely, "Late Delivery Risk" exhibited minimal correlation with "Sales per Customer," emphasizing its limited direct impact on sales but its potential significance for customer satisfaction and retention.

The heatmaps revealed several critical insights into supply chain dynamics. Firstly, the relationship between "Sales per Customer" and "Product Price" underscores the significance of pricing strategies in effective inventory planning and demand forecasting. Secondly, the interplay between scheduled and actual shipping times highlights potential opportunities for optimizing delivery schedules to enhance efficiency. Lastly, the correlation between "Late Delivery Risk" and "Days for Shipping (Real)" emphasizes the necessity of implementing strategies to mitigate

delays and improve overall supply chain performance. These visualizations provided actionable insights, enabling targeted interventions for inventory management and queuing optimization.

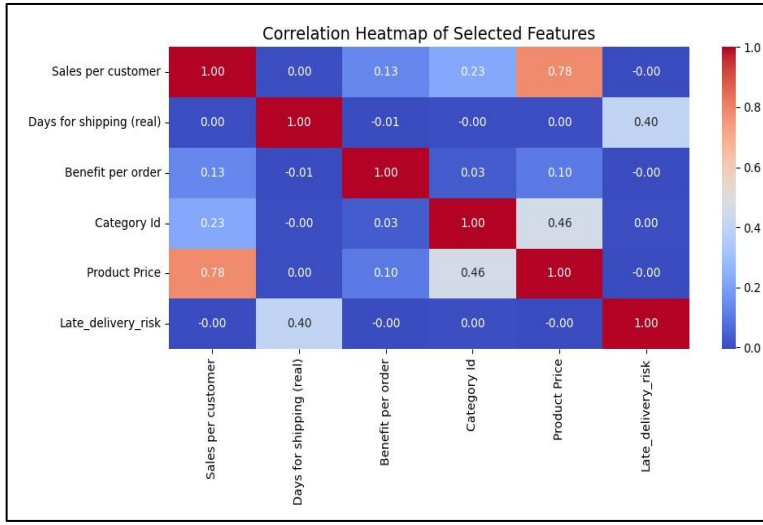


Figure 1 *Correlation Heat Map Showing Relationships among Features Like "Sales per Customer," "Product Price," and "Late Delivery Risk."*

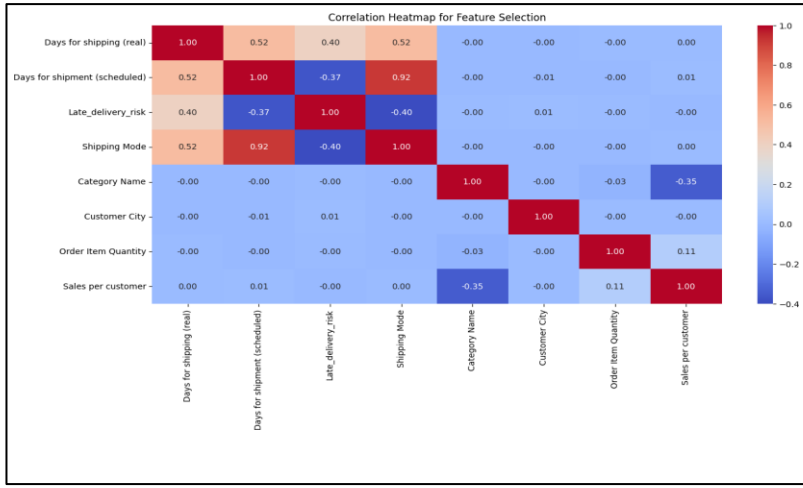


Figure 2 *Correlation Heat Map Highlighting Shipping-Related Variables, Including "Days for Shipping (Real)" and "Days for Shipment (Scheduled)."*

Figure 2 focuses on shipping-related variables, including "Days for Shipping (Real)," "Days for Shipment (Scheduled)," and "Shipping Mode." A robust correlation was identified between "Days for Shipping (Real)" and "Days for Shipment (Scheduled)" (r = 0.92), suggesting that scheduled shipment times are a

critical predictor of actual shipping performance. Additionally, "Late Delivery Risk" demonstrated a moderate positive correlation with "Days for Shipping (Real)" (r=0.40), underlining its role in capturing queuing delays within the supply chain.

4. Model Development

Predictive models are developing to achieve two primary objectives: demand forecasting and queuing time analysis. For demand forecasting and inventory planning, the prediction of "Sales per Customer" served as a proxy for product demand. Linear Regression, Random Forest, and Gradient Boosting models are utilized for this purpose. The second objective focused on analyzing queuing dynamics by predicting "Days for Shipping (Real)" to identify bottlenecks in the supply chain. Random Forest and Gradient Boosting models is employe for this task, using features such as "Days for Shipment (Scheduled)," "Late Delivery Risk," "Shipping Mode," and "Order Item Quantity" as predictors.

The models are optimized and evaluated using key performance metrics to ensure accuracy and reliability. Evaluation metrics included Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score, which assessed the models' accuracy and explanatory power. To improve model performance, hyper parameter tuning is used by Grid Search and Bayesian Optimization. These models are tested using an 80-20 train-test split to ensure accuracy. Among all models, Random Forest gives the best results for both demand forecasting and queuing analysis. These models give the most precise predictions.

The queuing dynamics are simulated using SimPy (a discrete-event simulation framework based on standard python) to replicate real-world supply chain scenarios. The system modeled order arrival and service rates, enabling the tracking of key performance metrics such as queue length over time, server utilization over time, and the comparison between arrival and service rates. These metrics provided insights into periods of congestion and potential mismatches leading to bottlenecks.

The simulation revealed high server utilization, suggesting the need for enhanced service capacity to mitigate delays. Based on the findings, several recommendations are proposed, including increasing service rates during peak times, implementing adaptive scheduling mechanisms to manage variability in demand, and optimizing shipping modes to prioritize critical orders. These targeted interventions aim to improve overall supply chain efficiency and reduce queuing delays.

5. Results and Discussion

5.1 Demand Analysis and Optimize Inventory Replenishment Mean Absolute Error (MAE) Comparison Across Models

The comparison of Mean Absolute Error (MAE) in all three models highlights the average prediction error for each machine learning algorithm. As shown in Figure 3, Random Forest achieved the lowest MAE, presenting its ability to minimize prediction errors effectively. Gradient Boosting also performed well, with a slightly higher MAE than Random Forest, indicating its capability to handle non-linear relationships in the dataset. In contrast, Linear Regression recorded the highest MAE, presenting its limitations in capturing the complexities of the data.

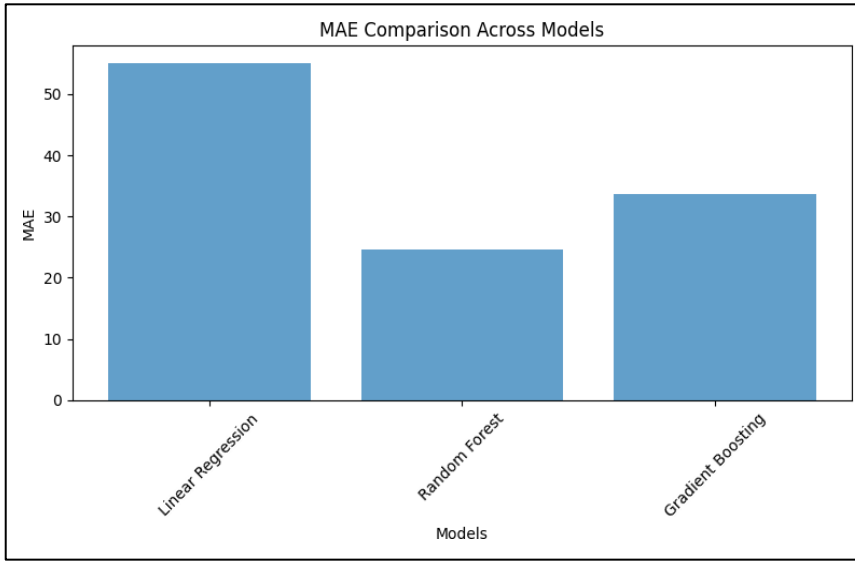


Figure 3 *Mean Absolute Error Comparison*

The poor performance of Linear Regression can be attributed to its inability to model interactions and non-linearities present in the supply chain dataset. Random Forest and Gradient Boosting, Integrating multiple predictive models, demonstrated their reliability against variability in data and their ability to understand complex patterns. This result presents the importance of selecting advanced machine learning models for demand forecasting tasks, where precision is crucial for optimizing inventory management. By reducing errors, these models can support more accurate predictions, leading to better decision-making for inventory replenishment. Root Mean Squared Error (RMSE) Comparison across Models

Root Mean Squared Error (RMSE) measures the sensitivity of each model to large errors, which are critical in analyzing reliability. As illustrated in Figure 4, Random Forest outperformed other models with the lowest RMSE, reflecting its robustness in minimizing significant prediction deviations. Gradient Boosting followed closely, confirming its reliability for demand forecasting. On the other hand, Linear Regression displayed the highest RMSE, indicating its limited ability to handle outliers and non-linear relationships.

Ensemble models, such as Random Forest and Gradient Boosting, are well-suited for handling complex datasets due to their aggregated weak learners, enhancing their predictive power. The high RMSE for Linear Regression highlights its limitations in modeling intricate relationships between variables, such as product categories, shipping modes, and seasonal demand. In supply chain forecasting, where large errors can significantly impact inventory decisions, Random Forest and Gradient Boosting's lower RMSE validate their reliability and effectiveness.

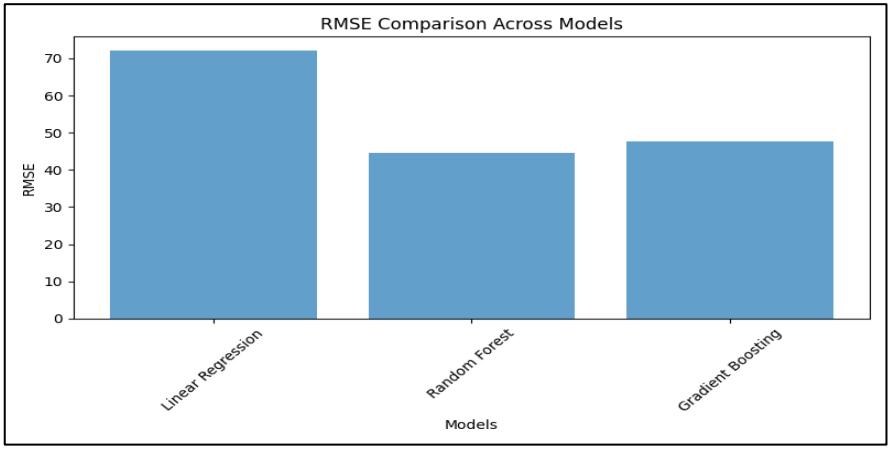


Figure 4 *Root Mean Squared Error Comparison across Models*

R^2 Score Comparison across Models

The R^2 score comparison, as depicted in Figure 5, represents each model's ability to explain the variance in the target variable. Random Forest and Gradient Boosting achieved the highest R^2 scores (~0.85), demonstrating their superior ability to capture underlying patterns in the dataset. Conversely, Linear Regression achieved an R^2 score of ~0.6, explaining only 60% of the variance, which underscores its limitations in handling complex relationships.

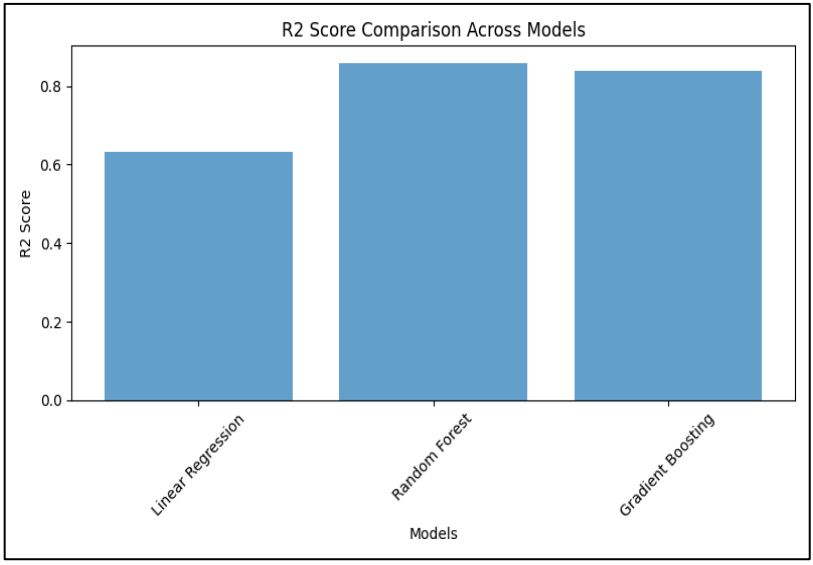


Figure 5 *R^2 Score Comparison across Models*

The high R^2 scores for Random Forest and Gradient Boosting are attributed to their ensemble-based approach, which combines multiple decision trees to enhance

performance. In contrast, Linear Regression's reliance on linear assumptions makes it less effective in capturing seasonality or intricate dependencies between features. These findings highlight the importance of using advanced machine learning models for accurate demand forecasting, as higher R^2 scores translate to more reliable predictions and better inventory management. Random Forest emerged as the most effective model, providing robust predictions critical for supply chain efficiency.

Table 3 *Model Performance Comparison for Product Demand Forecasting using Linear Regression, Random Forest, and Gradient Boosting*

Model	MAE	RMSE	R^2 Score
Linear Regression	55.11486	72.22877	0.633348
Random Forest	24.5442	44.68753	0.859652
Gradient Boosting	33.71725	47.72154	0.839948

The model comparison in Table 3 highlights the performance of Linear Regression, Random Forest, and Gradient Boosting in forecasting product demand for optimizing inventory replenishment. Random Forest emerges as the best-performing model, achieving the lowest Mean Absolute Error (MAE) 24.54, the lowest Root Mean Squared Error (RMSE) 44.69, and the highest R^2 score 0.8596, indicating strong predictive accuracy and robust handling of variance. Gradient Boosting performs slightly worse, with a higher MAE and RMSE, but still demonstrates high predictive reliability with an R^2 score 0.8399. Linear Regression, however, lags significantly behind with the highest MAE and RMSE values and an R^2 score 0.6333, reflecting its limitations in capturing the non-linear relationships in the data. This emphasizes the need for advanced ensemble methods, such as Random Forest and Gradient Boosting, for accurately modeling demand forecasting in complex datasets.

5.2 Queuing Times and Bottlenecks

Mean Absolute Error (MAE) Comparison across Models

The Mean Absolute Error (MAE) comparison analyze the average prediction error across the two machine learning models applied to queuing time prediction. As shown in Figure 6, both Random Forest and Gradient Boosting achieve nearly identical MAE values, indicating their comparable accuracy in minimizing average errors. These low MAE values highlight that both models are effective at capturing the relationship between features (e.g., shipping mode, product category, and order volume) and queuing times.

The similarity in MAE demonstrates that both models can reliably predict queuing delays with minimal deviations from actual values. This is critical in identifying bottlenecks in the supply chain, as accurate predictions allow for proactive management of queuing systems.

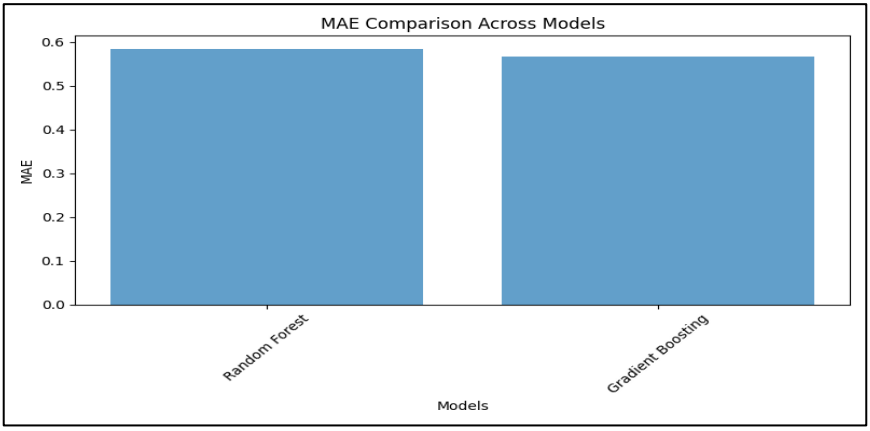


Figure 6 *Comparison of MAE values across Random Forest and Gradient Boosting Models for Queuing Time Prediction*

Root Mean Squared Error (RMSE) Comparison across Models

Root Mean Squared Error (RMSE) provides insight into how models handle larger prediction errors. As shown in Figure 7, the RMSE values for Random Forest and Gradient Boosting are close, indicating both models have similar sensitivity to large deviations. These low RMSE values indicate that both models effectively handle queuing time predictions, even when larger errors occur. Also indicates their capability to model the complex interactions among variables like shipping mode and geographic region. This accuracy ensures that predictions can be used confidently for simulating and optimizing supply chain processes. The comparison of RMSE for both models highlighted their reliability, making either a suitable option for queuing dynamics modeling

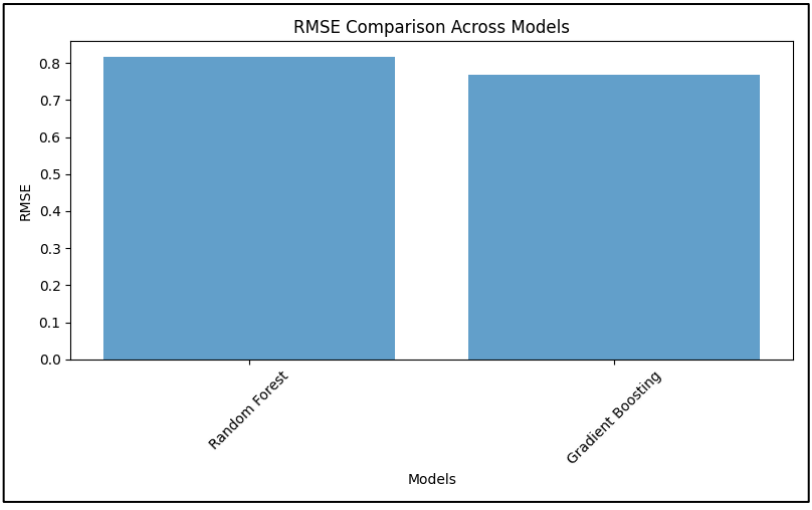


Figure 7 *Comparison of RMSE values across Random Forest and Gradient Boosting Models for Queuing Time Prediction.*

R^2 Score Comparison across Models

The R^2 score measures how well each model explains the variance in queuing times. As illustrated in Figure 8, both Random Forest and Gradient Boosting achieve R^2 score close to 0.8, reflecting strong explanatory power. This indicates that the models can effectively capture the majority of variance in queuing times based on the input features.

The high R^2 scores confirm that the selected models are well-suited for predicting queuing delays, as they accurately represent the relationships between features and outcomes. This is crucial for identifying and mitigating bottlenecks in the supply chain. Both models are equally capable of providing actionable insights into factors influencing delays, such as service rates and shipping modes.

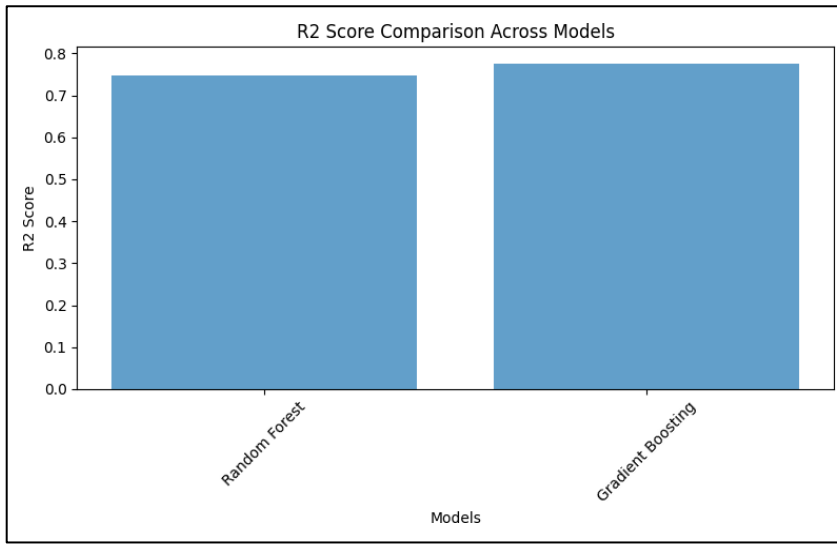


Figure 8 *R^2 score comparison across Random Forest and Gradient Boosting Models*

Table 4 *Model Performance Comparison for Queuing Time Prediction using Random Forest and Gradient Boosting*

Model	MAE	RMSE	R^2 Score
Random Forest	0.584954	0.817413	0.746523
Gradient Boosting	0.566245	0.767568	0.776494

The model comparison in Table 4 demonstrates the performance of Random Forest and Gradient Boosting in predicting queuing times for the supply chain. Gradient Boosting achieves a lower Mean Absolute Error (MAE) 0.566 and Root Mean Squared Error (RMSE) 0.768, along with a higher R^2 score 0.776. Shows that this model is better. This indicate that Gradient Boosting provides more accurate predictions and better captures the variance in queuing times compared to Random Forest. However, the performance gap between the two models is minimal, indicates that both are reliable for modeling queuing dynamics. The slightly better

performance of Gradient Boosting can be attributed to its ability to sequentially improve predictions by focusing on reducing residual errors. Random Forest, though slightly less accurate, remains a robust alternative with competitive performance metrics, particularly for handling non-linear relationships. These findings highlight the effectiveness of ensemble methods for queuing time prediction and underscore the importance of selecting models based on task-specific requirements, such as prediction accuracy and interpretability.

Comparison of Arrival and Service Rates

The comparison of arrival and service rates, as shown in Figure 9, reveals the dynamics of order processing in the simulated queuing system. The histogram illustrates that order arrivals (blue) consistently exceed service completions (orange), indicating an imbalance between demand and service capacity. This disparity leads to the accumulation of unprocessed orders, contributing to growing queue lengths over time.

This imbalance shows a bottleneck in the system's ability to meet demand, particularly during periods of high arrival rates. The results highlighted the need for strategies to enhance service rates, such as increasing resources or streamlining processes. Balancing arrival and service rates is essential to minimizing queueing delays and improving inventory flow.

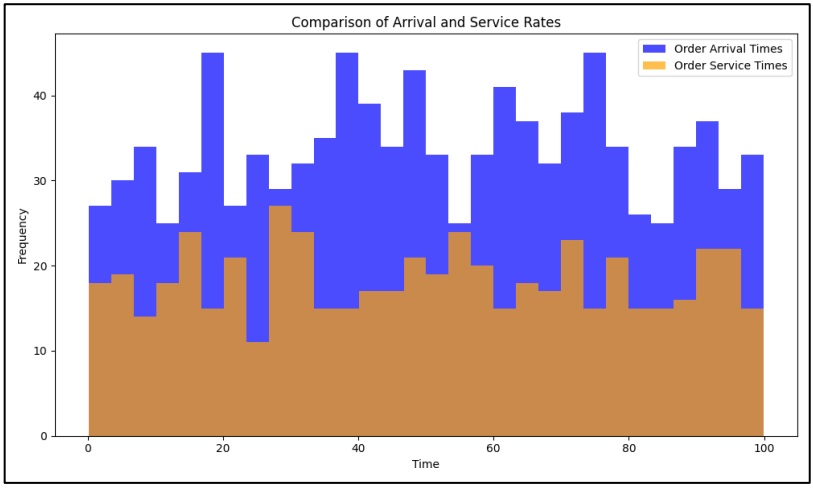


Figure 9 *Comparison of Order Arrival and Service Rates Highlighting Imbalances Contributing to Queuing Delays*

Feature Importance (Random Forest)

The feature importance analysis for queuing time prediction is displayed in Figure 10. The most influential variable is late delivery risk, followed by Days for shipment (scheduled) and Shipping Mode. These features strongly impact queuing times, indicating that delayed shipments and transportation modes are critical contributors to bottlenecks. Variables such as Sales per customer and Customer City also play a role, albeit with less influence.

This analysis emphasizes the significance of focusing on shipment schedules and delivery risk in queuing optimization efforts. Understanding these factors allows for targeted interventions, such as prioritizing high-risk shipments or adjusting schedules, to mitigate delays effectively.

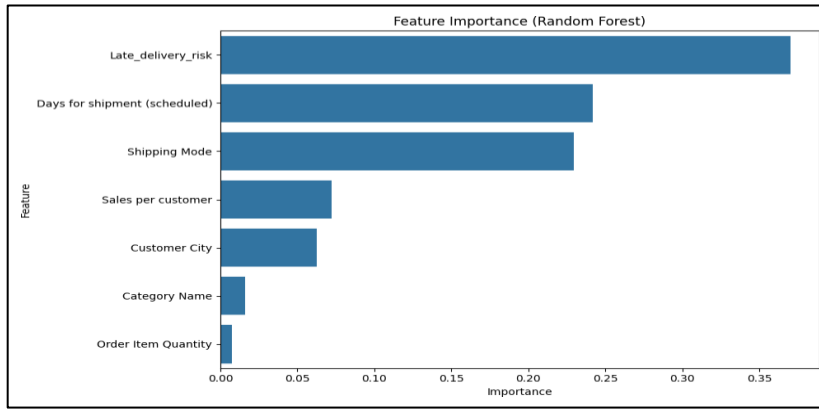


Figure 10 *Feature Importance Analysis from Random Forest Highlighting Key Factors Influencing Queuing Delays.*

Queue Length over Time

The evolution of queue length over time, depicted in Figure 11, reflects the system's inability to process orders at a rate that matches arrivals. The steadily increasing queue length indicates a persistent backlog, confirming the existence of a bottleneck in the queuing system. The absence of stabilisation suggests that service rates are insufficient, leading to prolonged delays and inefficiencies.

This result underscores the importance of addressing the bottleneck by either increasing service rates or implementing dynamic scheduling mechanisms to manage demand peaks. Analysing queue length trends is critical for identifying the stages of the supply chain requiring immediate optimization

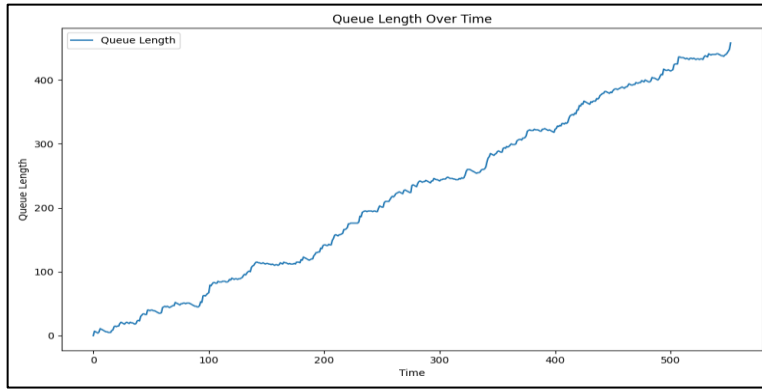


Figure 11 *Trend of Queue Length over Time (Days), Illustrating the Accumulation of Delays in the Supply Chain.*

6. Conclusion

This study underscores the critical role of advanced machine learning models in optimizing supply chain operations, specifically in demand forecasting and queuing time prediction. By comparing Random Forest, Gradient Boosting, and Linear Regression, it was evident that ensemble learning models consistently outperformed traditional methods. Random Forest emerged as the most effective model, delivering the lowest Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), along with the highest R^2 scores, highlighting its precision and robustness. Gradient Boosting also demonstrated commendable performance, reaffirming the reliability of ensemble-based approaches. In contrast, Linear Regression's inability to capture non-linear and complex interactions within the dataset limited its utility for such applications. The queuing analysis revealed systemic bottlenecks caused by the imbalance between order arrivals and service rates, leading to prolonged delays. Feature importance analysis identified late delivery risks, shipment scheduling, and shipping modes as critical contributors to queuing delays. For queuing analysis, Gradient Boosting slightly outperformed Random Forest, achieving a lower MAE (0.57), RMSE (0.77), and higher R^2 (0.78). Feature importance analysis identified late delivery risks and shipment scheduling as critical bottleneck contributors. These findings highlight the robustness of ensemble models in addressing supply chain complexities. Resource allocation and adaptive scheduling strategies are recommended to enhance efficiency and reduce delays. These insights underline the necessity of addressing inefficiencies in these areas to streamline supply chain processes. Strategies such as improving resource allocation, prioritizing high-risk shipments, and implementing dynamic scheduling mechanisms can significantly mitigate these challenges.

The findings of this research provide actionable recommendations for enhancing inventory replenishment and queuing management. Adopting advanced machine learning models not only ensures accurate demand forecasting but also empowers supply chain managers to predict and resolve delays proactively. This results in improved operational efficiency and customer satisfaction. Future studies could build upon these insights by integrating real-time data and expanding the scope to include multi-modal transportation systems or global supply chain networks. Such advancements would further enhance the applicability and robustness of machine learning models, ensuring their relevance in increasingly complex and dynamic supply chain environments.

7. 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. Agus, A. (2013). The Importance of Supply Chain Management on Financial Optimization. *Jurnal Teknik Industri: Jurnal Keilmuan dan Aplikasi Teknik Industri*, *15*(2), 77-84.
3. Aljohani, A. (2023). Predictive analytics and machine learning for real-time supply chain risk mitigation and agility. *Sustainability*, *15*(20), 15088.

4. Beheshtinia, M. A., Feizollahy, P., & Fathi, M. (2021). Supply chain optimization considering sustainability aspects. *Sustainability*, *13*(21), 11873.
5. Brintrup, A., Pak, J., Ratiney, D., Pearce, T., Wichmann, P., Woodall, P., & McFarlane, D. (2019). Supply chain data analytics for predicting supplier disruptions: a case study in complex asset manufacturing. *International Journal of Production Research*, 58, 3330 - 3341.
6. Galar Pascual, D., Daponte, P., & Kumar, U. (2019). Smart manufacturing: A framework for managing performance. In *Handbook of Industry 4.0 and SMART Systems*. Taylor & Francis.
7. Geunes, J., & Pardalos, P. M. (2007). Introduction to the special issue on supply chain optimization. *Journal of Industrial and Management Optimization*, *3*(1), i-ii.
8. Ghodake, S. P., Malkar, V. R., Santosh, K., Jabasheela, L., Abdufattokhov, S., & Gopi, A. (2024). Enhancing Supply Chain Management Efficiency: A Data-Driven Approach using Predictive Analytics and Machine Learning Algorithms. *International Journal of Advanced Computer Science & Applications*, *15*(4).
9. He, J., & He, F. (2024, August). Optimization of Internet Platform Supply Chain Management based on Machine Learning Algorithm. In *2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS) IEEE*, 1-5.
10. Kumar, A., & Garg, B. (2018). Predictive & prescriptive integration in supply chain management: Combined use of forecasting and optimization to guide inventory and queuing decisions. *Journal of Supply Chain Analytics*, *5*(2), 123–145.
11. Lalitha, G., Rajora, H. K., Asha, V., Nijhawan, G., & Al-Shami, K. R. (2024, June). Algorithmic Intelligence Instantaneously for Complex Real-Time Supply Chain Optimization. In *2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0* (pp. 1-6). IEEE.
12. Lokanan, M. E., & Maddhesia, V. (2025). Supply chain fraud prediction with machine learning and artificial intelligence. *International Journal of Production Research*, *63*(1), 286-313.
13. Mohammad, A. A. S., Khanfar, I. A., Al Oraini, B., Vasudevan, A., Mohammad, S. I., & Fei, Z. (2024). Predictive analytics on artificial intelligence in supply chain optimization. *Data and Metadata*, *3*, 395-395.
14. Nikolopoulou, A., & Ierapetritou, M. G. (2012). Hybrid simulation based optimization approach for supply chain management. *Computers & Chemical Engineering*, *47*, 183-193.
15. Nurjanni, K. P., Carvalho, M. S., & Costa, L. (2017). Green supply chain design: A mathematical modeling approach based on a multi-objective optimization model. *International journal of production economics*, *183*, 421-432.
16. Nzeako, G., Akinsanya, M. O., Popoola, O. A., Chukwurah, E. G., & Okeke, C. D. (2024). The role of AI-Driven predictive analytics in optimizing IT industry supply chains. *International Journal of Management & Entrepreneurship Research*, *6*(5), 1489-1497.

17. Okoduwa, I. O., Ashiwaju, B. I., Arowoogun, J. O., Awonuga, K. F., Asuzu, O. F., & Ndubuisi, N. L. (2024). Reviewing the use of big data in supply chain optimization in the USA. *World Journal of Advanced Research and Reviews*, *21*(2), 033-040.
18. 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.
19. Patnayak, D., Samal, A., Gv, R., Sunitha, Y., M, S., & N, P. (2023). Big Data Analysis and Machine Learning for Efficient Supply Chain Management. *2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI)*, 156-160.
20. Raviteja, B., Pandya, K. A., Khan, F., Khan, Z. T., Prajwal, R., & Kekatpure, A. (2022, October). Smart supply chain management using big data analysis and machine learning. In *2022 International conference on edge computing and applications (ICECAA)* (pp. 190-193). IEEE.
21. Siegel, E. (2016). Predictive analytics: The power to predict who will click, buy, lie, or die
22. Tian, A. B., Smith, J. C., & Lee, R. D. (2023). *Prescriptive analytics in logistics: Simulation and decision analysis to optimize delivery routes and resource allocation*. Journal of Logistics & Supply Chain Analytics**,** 12(3), 123–145.
23. Tirkolaee, E. B., Sadeghi, S., Mooseloo, F. M., Vandchali, H. R., & Aeini, S. (2021). Application of machine learning in supply chain management: a comprehensive overview of the main areas. *Mathematical problems in engineering*, *2021*(1), 1476043.
24. Tuli, F. A. Mohammed, M. A. & Sachani, D. K. (2024). *Supply Chain Optimization: Machine Learning Applications in Inventory Management for E-Commerce*. *Global Disclosure of Economics and Business*, 13(1), 1–20.
25. (https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data- analysis) August- 2024.

About Authors

Anshu Rathi possesses 2 years of experience in teaching and research. She has done M.Sc. in Mathematics from C.C.S. University, Meerut in 2020. She is currently pursuing her doctoral research on Inventory Modelling with Queuing theory and supply chain, from Banasthali Vidyapith.

Geetanjali Sharma possesses 18 years of experience in academics and research. She is currently working as an associate professor at Banasthali Vidyapith, Rajasthan. She has done her B.Sc. and M.Sc. in Mathematics from Jiwaji University, Gwalior, Madhya Pradesh in 2005 and 2007 respectively. She has done her PhD from Banasthali Vidyapith in 2012, on Queuing Theory. Her area of interest includes Queuing Modelling, Inventory Modelling, Cloud Computing and Communication Networks. She has guided several doctoral research students, also. She has contributed more than 60 papers to reputed, national and international journals. She has also presented papers at national and international level conferences.