Journal of Decision Science and Optimization
Vol. 1 (2025), No. 1, pp. 55-66
DOI: https://doi.org/10.55578/jdso.2506.003

ARTICLE

Leveraging Predictive Analytics and AI Techniques to Enhance the Efficiency in Supply Chain Management: A Case Study to Optimize Supply Chain Characteristics

Lavanya Samineni[1], Sai Saran Ogoti[1], Afshin Zahraee[2] and Lash Mapa[3,*]

[1] *College of Technology, Purdue University, Hammond, Indiana, 46323, United States; lsaminen@pnw.edu; sogoti@pnw.edu*

[2] *Construction Science and Organizational Leadership City, Purdue University, Hammond, Indiana, 46323, United States; azahraee@pnw.edu*

[3] *Mechanical Engineering Technology, Purdue University, Hammond,Indiana,46323, United States; Lmapa@pnw.edu*

[*]Corresponding author. Email: *Lmapa@pnw.edu*

Received: 06 June 2025, Accepted: 20 June 2025, Published: 27 June 2025

Abstract

This study explores applying predictive analytics and Machine learning techniques in optimizing supply chain management (SCM) within the retail industry. As retailers face increasing complexities due to fluctuating consumer demand and competitive pressures, integrating advanced data-driven methodologies becomes essential. This research employs machine learning algorithms and statistical modelling to forecast demand patterns, utilizing key features such as shipping, order status, and customer segment analysis to enhance inventory management and streamline logistics operations [1]. By analyzing historical sales data, including sales per customer and order profit per order, and leveraging real time insights, the study aims to improve operational efficiency and decision-making processes. Key performance indicators (KPIs) such as inventory turnover rates and late delivery risk will be evaluated to assess the impact of predictive analytics on SCM performance. Ultimately, this research highlights the transformative potential of leveraging predictive analytics in the retail sector, paving the way for more agile and responsive supply chain strategies that can significantly enhance competitive advantage over adversaries in the market [2].

Keywords: Predictive Analytics; Time Series Analysis; Supply Chain Management; Retail Industry

1. INTRODUCTION

In the contemporary business environment, supply chain management (SCM) is critical for enhancing operational efficiency and responsiveness. Predictive analytics, which employs historical data and statistical models to forecast future trends, is vital in optimizing supply chain operations [3]. By accurately predicting demand fluctuations, businesses can make informed decisions regarding inventory management and logistics. The retail industry, characterized by rapid changes in consumer preferences and competitive pressures, presents unique challenges that can be addressed well through predictive analytics. Retailers must manage complex inventory systems while ensuring timely deliveries to meet customer expectations [4]. By leveraging predictive analytics, retailers can optimize stock levels, reduce holding costs, and improve overall supply chain efficiency. With these factors, we have chosen the retail industry for this project to demonstrate the transformative potential of predictive analytics in enhancing supply chain characteristics such as Inventory management, Demand Forecasting and achieving a competitive edge. A potential Predictive analytics AI model that we aim to create will intake the historical time series data related to supply chain and transform the information to suggest the actions needed to minimize the possible supply chain inefficiencies.

https://doi.org/10.55578/jdso.2506.003

Hypothesis 1: Accurately forecasting the product demand based on customer location, product category along with customer segmentation based on purchase patterns, will optimize stock levels thereby leading to better customer satisfaction and efficient inventory management.

Hypothesis 2: Optimizing inventory stock levels and shipping modes for different order sizes and different locations will reduce inventory operational costs and minimize supply chain inefficiencies such as stockouts and late deliveries

2. LITERATURE REVIEW

In the evolving landscape of machine learning, forecasting techniques have become crucial tools for improving accuracy in demand prediction. These methods are typically categorized into three main groups: (1) time series analysis, (2) regression-based models, and (3) advanced supervised and unsupervised approaches. Time series techniques, such as ARIMA and Holt-Winters Exponential Smoothing, are widely used in retail and other industries for their ability to capture trends, seasonal patterns, and cyclic behaviors. By identifying these recurring patterns, businesses can better plan and allocate resources, making these methods foundational in supply chain forecasting [5].

Alongside series models, regression-based approaches provide a flexible framework for understanding the relationships between independent variables and their effects on demand [6]. They allow for a more comprehensive view of how various factors influence demand patterns. Meanwhile, advanced techniques like artificial neural networks (ANNs) and long short-term memory (LSTM) models stand out for their ability to manage complex and nonlinear data. These sophisticated methods often outperform traditional techniques in dynamic and unpredictable environments, making them indispensable for industries where demand is highly volatile and complex [1].

Demand uncertainty in inventory management can be addressed through methodologies that assume the demand distribution and its parameters are known at the point of decision making in each period. However, this assumption is often unrealistic, as decision-makers frequently lack prior knowledge about demand distribution and its temporal variations. Another approach involves a two-phase process: (a) an estimation or forecasting phase, and (b) an optimization phase. Initially, assumptions about historical data are made, and statistical methods are used to estimate parameters, which are then applied to determine optimal decision variables. However, data noise or inaccurate forecasting methods may lead to suboptimal outcomes during the optimization phase [4,7].

The third category emphasizes the application of distributional or parametric assumptions based on historical demand data, which can be estimated through machine learning and deep learning techniques [8]. These methods can be either parametric or nonparametric. Nonparametric forecasting methods do not rely on normality or other distributional assumptions and use raw data directly [9]. Recent advancements in artificial intelligence and data analytics have facilitated the incorporation of machine learning and deep learning approaches into time-series forecasting. These methods often leverage autoregressive (AR), moving average (MA), or autoregressive integrated moving average (ARIMA) techniques to establish linear functions of past observations [3].

Research by Ifraz et al., explored demand forecasting for spare parts in bus fleets, comparing various methods including regression-based, rule-based, tree-based, and artificial neural networks (ANN). Their findings indicated that ANN consistently outperformed other methods, achieving the highest accuracy rates and minimal deviation in demand forecasting [10]. Swaminathan and Venkitasubramony reviewed forecasting techniques in fashion product demand prediction, focusing on recent advancements in artificial intelligence and machine learning methods. They highlighted the importance of combining various techniques to improve forecast accuracy [11].

The challenge of addressing demand uncertainty without oversimplifying distributional assumptions was tackled by Huang et al. [12]. They contested the prevalent assumption of normally distributed demand, advocating for the application of nonparametric kernel density estimation (KDE) for short lead times. Their study examined the effectiveness of KDE, generalized autoregressive conditionally heteroscedastic (GARCH), and simple exponential smoothing (SES) within an order-up-to-level (OUTL) policy.The authors concluded that while nonparametric methods may struggle with small sample sizes, parametric techniques excel when the demand distribution is normal. Similarly, Zhang et al. integrated KDE and conditional GARCH methods to minimize loss functions and enhance empirical safety stock predictions within a newsvendor model framework [13].

An innovative approach by Liu et al. utilized a double parallel feed-forward neural network to determine stock levels for a newsvendor problem, outperforming traditional statistical forecasting methods such as Holt-Winters' triple exponential smoothing [14]. This approach successfully captured both stationary and nonstationary time series patterns. Babai et al. provided a thorough analysis of demand forecast error under stochastic lead times, emphasizing the importance of accounting for demand autocorrelation and lead time variability to enhance forecasting strategy selection for inventory control [2].

Recent studies have also highlighted the effectiveness of ensemble learning in demand forecasting and inventory management. Ensemble learning combines multiple forecasting models to enhance accuracy and robustness, addressing individual model biases and capturing diverse patterns demonstrated that partially combining predictors can yield comparable, if not superior, performance compared to using all predictors simultaneously [15,13]. Furthermore, Yang et al. emphasized the need for efficient deployment methods for ensemble deep learning while minimizing associated time and resource costs [12]. Mohammed and Kora discussed integrating traditional ensemble learning concepts into deep learning to optimize hyperparameter tuning, reinforcing the necessity of effective ensemble forecasting methods for improved inventory optimization solutions [16].

Despite the recognized benefits of ensemble learning, limited research has been conducted on the application of ensemble deep learning for demand forecasting in supply chain management and inventory optimization [17]. The fundamental task of combining basic predictors within ensemble learning remains vital. Techniques such as simple averages, weighted averages, Bayesian model averaging (BMA), and meta-learning methods have demonstrated satisfactory performance across classification and regression tasks [18]. Andrade and Cunha proposed a methodology for disaggregated demand forecasting in retail, employing XGBoost, a non-linear ensemble-based model, and a structural change correction method to account for shifts in consumer behavior. Their approach achieved superior accuracy metrics, reduced stockouts, and lowered inventory costs, showcasing the potential of ensemble methods in modern demand forecasting [19].

3. DATA COLLECTION

Data collection is in this study, particularly in selecting, analyzing, and processing data suitable for a potential predictive analytics model. For our research, we utilized the DataCo Smart Supply Chain for Big Data Analysis dataset, which is available on Kaggle. Initially, we considered the Superstore Sales Analysis dataset for our study on supply chain operations [20,21]. This dataset, available on GitHub, focuses on a retail entity with fictional data that deals with office supplies, technology products, and furniture. However, upon further evaluation, we found that the DataCo Smart Supply Chain for Big Data Analysis dataset is more aligned with our research objectives. Its comprehensive supply chain parameters provided deeper insights, making it more suitable for retail supply chain analysis and operational efficiency optimization [20]. We have performed data analysis on the Dataco Dataset, which contains the anonymized supply chain data of Electronics products, Apparel items and sporting goods over the span of 4 years (2015-2018). Some of the most notable variables present in the Dataco Dataset are mentioned in the following table.

Table 1. Key Variables

S.no	Category	Key Variables
1	Temporal Indicators	Date, Month, Year
2	Product Details	Product Name, Product Category
3	Geographic Features	Market, Region, State, City, Latitude, Longitude
4	Performance Metrics	Order Item Quantity, Sales, Profitability Indicators
5	Shipping Indicators	Shipping Mode, Delivery Status, Shipping
6	Customer Characteristics	Delay

7	Transactional Details	Customer ID, Customer Segment Order ID, Order Profit Per Order
8	Clustering Attributes	Latitude, Longitude, Sales, Cluster

4. METHODOLOGY

This paper majorly revolves around observing the forecast of demand, inventory management, customer segmentation, and late delivery risk within the retail supply chain. The methodology process begins with data collection where a four-year timed data is collected and further preprocessed to handle missing values, remove inconsistencies such as invalid categories, outliers etc..., and normalize the data for effective analysis. Following the preprocessing step, the feature engineering step is employed for extracting the main features through underlying patterns of the data. As the data used for the study is time-dependent, time series analysis is implemented using predictive analytics models like Prophet and LSTM.

Hypothesis 1 focuses on demand forecasting; the study explores the effectiveness of predictive models such as Prophet and LSTM. These models analyze product demand based on customer segmentation, considering factors like average product sale price and purchase frequency. Though these models are highly useful in handling time-dependent data, their performance differs based on the volatility of the available supply chain data and the granularity of the variables considered. For instance, high-demand products exhibited more predictable patterns, whereas low-demand products require more refined models to account for periodic sales. The results that can be derived through analysis highlight the importance of suitable model selection with the specific characteristics in collected data and the business context.

As Hypothesis 2, aims at optimizing inventory stock levels, the study examines the effects of order quantity, customer location, and shipping modes on overall shipping costs. The predictive analytical models are trained on refined data, including supplier and consumer order details in a time-series format, to validate model efficiency. The primary goal is to minimize stockouts and delayed deliveries by optimizing inventory management and reducing shipping costs. This can be achieved by leveraging predictive analytics to perform cost-benefit analyses, ensuring streamlined supply chain operations. By systematically evaluating the efficiency of different predictive models, the study identifies the optimal approach for enhancing inventory control and operational efficiency in retail businesses.

In summary, the methodology integrates data collection, preprocessing, feature engineering, and time series analysis using advanced predictive models. Using this methodology, we explore how analytical algorithms can work on large datasets to draw insights from the raw data and understand the importance of selecting appropriate variables and models to address the complexities of long-term market trends, particularly when analyzed across different regions. The results derived are analyzed to provide tailored recommendations for improving supply chain operations, that align with the proposed hypotheses.**DATA ANALYSIS**

For our study, we analyzed the Dataco Dataset (Kaggle), a comprehensive and anonymized supply chain dataset spanning four years of data (2015–2018) from the sporting goods, clothing, and electronics retail industry [20]. This data set originally comprised 180,519 rows and 53 columns. To refine the dataset further with our research objectives such as demand forecasting, customer segmentation, inventory management, and shipping analysis. We performed data preprocessing and feature engineering. Initially, the number of columns was reduced from 53 to 33 based on their relevance to the chosen hypothesis. Subsequently, after conducting feature engineering, the key features included are Shipping Delay, Order Month and Year (extracted from the order date for time-based analysis), Profit Margin (calculated for each order), Order Size Category (to categorize orders based on quantity), and Region-Market Combination (to facilitate regional analysis). Furthermore, we cleaned the dataset by removing rows with null values, negative values in relevant numeric columns, and inconsistent or ambiguous data (e.g., invalid categories, outliers, or incorrect formats). After these refinements, we obtained a final dataset consisting of 101,193 rows and 39 columns, which played a crucial role in our subsequent analyses and meaningful exploration of supply chain.

5.1. Data Analysis for Forecasting Demand and Managing Inventory

After a comprehensive analysis of the dataset related to demand forecasting and inventory management, the dataset includes information on product demand across various categories, marketwise demand trends, and seasonal variations. The first and primary thing analyzed in this study is the least demanded products which we believe provide critical insights for inventory optimization. The horizontal bar chart Figure 1 showcasing the ten least demanded products identifies items like the Garmin Forerunner 910XT GPS Watch and Titleist Club Glove Travel Cover as having minimal demand, with quantities ordered barely exceeding 30 units. This category also includes Fitness-related products such as Bow flex Select Tech 1090 and SOLE E25 Elliptical Dumbbells. Lower demanding products such as these, present a chance to streamline the existing inventory by phasing out underperforming items or reducing the inventory volume. This can reduce the carrying costs, while these resources will be rediverted to the higher demand products to maximize operational revenue [12].

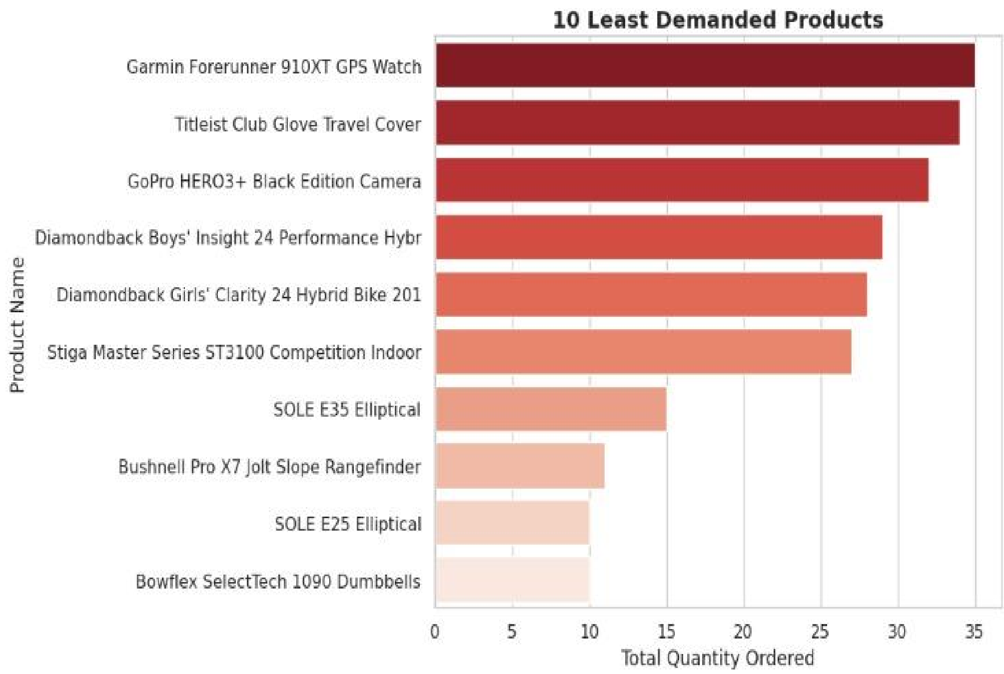

Figure 1. Illustration of Ten least demanded products

Following this, we also performed an analysis of the product demand based on its product category and its total order quantities, which also play a critical role in demand forecasting and inventory management. Here, based on the above (Figure 2) it is clear that, the High-demand categories like Consumer Electronics, Fitness Accessories, and Golf Apparel should be prioritized in inventory planning to ensure sufficient stock levels and avoid stockouts, especially during peak seasons. Conversely, low-demand categories such as CDs, DVDs, and Trade in products may require reevaluation or phase-outs to minimize carrying costs regional preferences and seasonal variations play a huge role in product demand [12]. For Example, Season-based summer products such as camping & hiking accessories see a higher demand spikes. Latin America sees higher demand in terms of soccer products, since they are actively being utilized higher in that region. By accessing such seasonal factors and localized factors, businesses can strategically tailor their decision to better fit customer's needs [12].

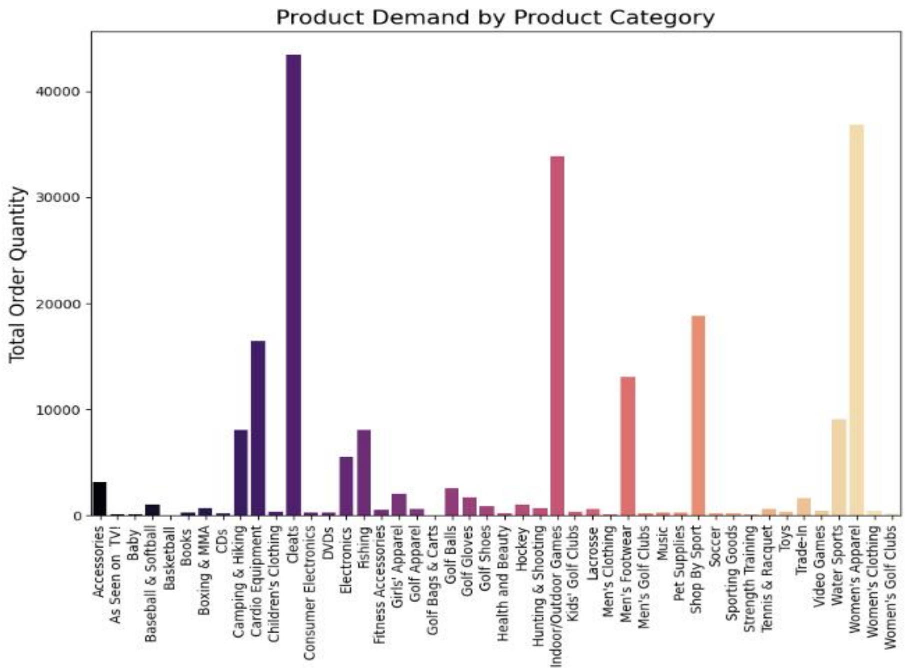

Figure 2. Product Demand by Product Category

Coming to the monthly demand trend for the Electronics category emphasizes the seasonal nature of demand. The graph shows frequent surges, most likely brought on by Christmas purchases, especially in the fourth quarter. Early 2016 and Late 2017 indicated that there was lower demand and a period of decline in terms of product demand. These kinds of low demand time periods can drive up the holding costs and can potentially cause overstocking. in a similar way businesses should anticipate high demand months and stockpile inventory to accommodate the upcoming demand. These seasonal fluctuations monitoring can be automated using analytics like time series or other algorithms.

The market demand for electronics shows how product popularity varies by area. Africa has the lowest demand for this category, whereas Latin America (LATAM) is the largest market for electronics, followed by Pacific Asia and Europe. These results highlight how crucial it is to modify marketing and inventory plans to meet the unique requirements of every area. (Figure 3), While the African market might need to reevaluate its product offers or distribution strategy, Latin America (LATAM), a region with significant demand, could benefit from priority inventory allocation and promotional efforts.

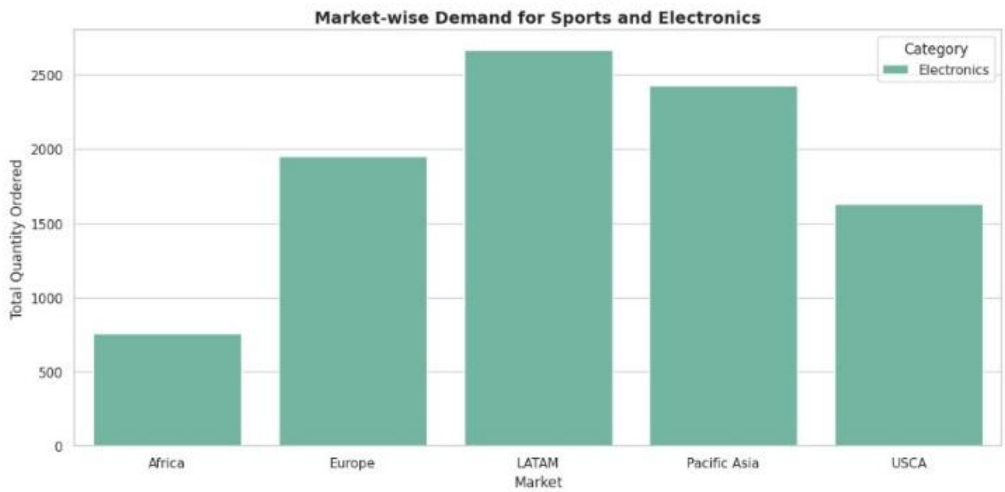

Figure 3. Market-wise Demand for Sports and Electronics

To understand the market of various regions, we utilized a time series analysis in drawing the yearly market-wise demand for the products. While Pacific Asia and USCA saw comparatively steady trends, the Latin America (LATAM) region saw a notable spike in demand in 2016 before a severe decrease. On the other hand, after 2016, markets such as Africa and Europe saw a decline in demand, indicating possible problems like market saturation or decreased interest in the product. These insights allow businesses to focus on markets with growth potential and investigate the causes of demand reduction in underperforming regions. The market trends are visible in the following Figure 3.

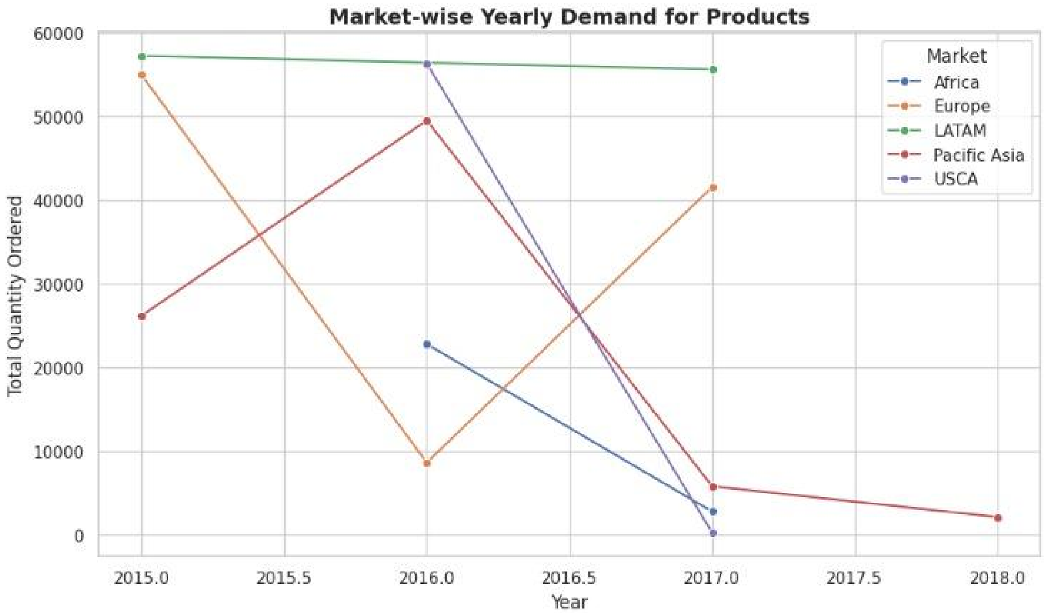

Figure 4. Market-wise Yearly Demand for Products

The Prophet algorithm provided comprehensive demand forecasts according to consumer location and product category to boost demand forecasting precision. The algorithm provides valuable data of demand and inventory through expected demand values (`yhat`) and their respective confidence bounds (`yhat_lower` and `yhat_upper`). The Prophet algorithm predicted a demand of 0.902948 for February 26th, 2018, which had a lower bound of -1.262204 and an upper bound of 3.018608. The expected demand value for 2018-03-02 stood at 0.910001 while the confidence interval ranged between 1.332201 and 3.025086. These narrow confidence intervals help businesses to maintain optimal inventory levels by stocking enough products during peak demand periods and preventing overstock during low demand times while mitigating the risk of overstocking.

However, a detailed analysis of the Prophet algorithm outputs displayed certain data inconsistencies in the chosen dataset, particularly in inventory predictions, as the results are indicated by extreme confidence intervals. For example, on February 21, 2018, the expected inventory level (yhat) was 0.319768, yet the confidence interval spanned from -3.907868 to 4.751218, which shows significant uncertainty. These discrepancies may arise due to missing values, outliers, or inconsistencies in historical inventory records, which can negatively impact model performance. While our dataset selection was appropriate for demand forecasting, the presence of missing values contributed to extreme variations in both demand and inventory predictions. Notably, demand forecasting yielded more reliable results compared to inventory predictions. In the end, we can surely say that this model works well with a well detailed dataset and helps in promoting operational efficiency and profitability by ensuring that demand and inventory levels are in line with market trends and supporting strategic decision making.

5.2. Data Analysis for Late Delivery Risk and CustomerSegmentation

The analysis of shipping delays across regions and modes reveals significant insights. Second Class shipping experiences the highest delays, averaging around 2 days in most regions, indicating inefficiencies that may require operational improvements. However, Standard Class and First-Class shipping perform better overall, with delays usually less than a day, making them more dependable choices. Same-day delivery, which frequently has delays of less than 0.5 days in all locations, is the most dependable shipping choice. Regionally, Western Europe and North America have the fewest delays, indicating superior logistical procedures, whereas South Asia and Eastern Asia have the longest, with Second Class delays lasting more than two days.

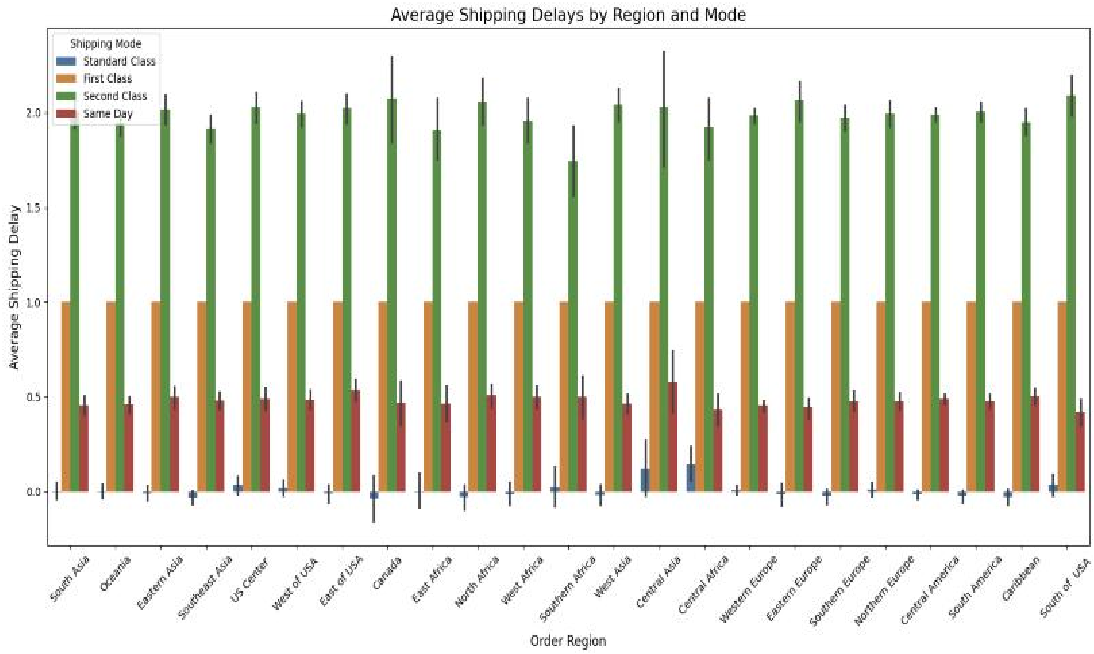

Figure 5. Average Shipping Delays by Region and Mode

Most orders are successfully delivered on time, according to the delivery status distribution. Still, a sizable percentage of deliveries roughly 10% to 15% occur late across all shipping methods and geographical areas. This suggests that several supply chain components continue to be inefficient, particularly for modes such as Second Class. Addressing these delays by optimizing processes for underperforming shipping modes or improving infrastructure in high-delay regions could enhance the overall reliability of delivery services.

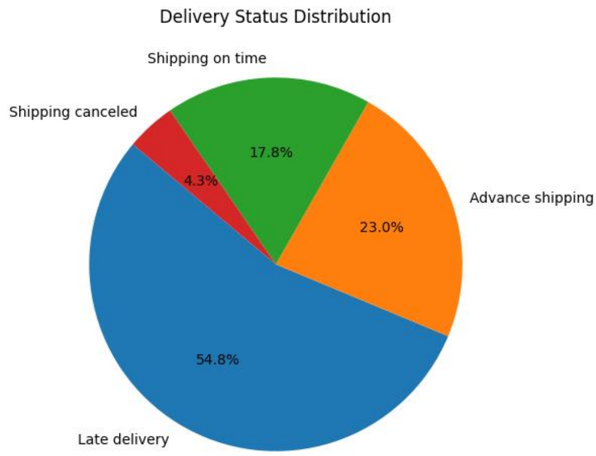

Figure 6. Delivery Status Distribution

According to the order size distribution, most orders fall into the small size group, usually containing amounts ranging from 0 to 5. Large orders (more than 20 products) and medium-sized orders (6–20 items) are less common but nonetheless important, particularly in areas with more demand. According to the analysis of high-demand products, a small number of items account for a disproportionate amount of overall order amounts, suggesting that if inventory is not well managed, there may be stockouts for these items. For instance, the top ten most popular products account for a significant portion of overall sales and might need to have their inventory replenished first to prevent supply interruptions.

5.2.1. Profitability vs. Shipping Delays

The relationship between shipping delays and profitability shows an interesting trend. Second, Class shipping delays have a detrimental effect on profitability, even though shorter delays often translate into larger earnings per order. Profit margins are frequently lowered for orders that are delayed by more than two days, particularly in areas like South and Eastern Asia where logistical difficulties are more severe. Conversely, modes like Same Day delivery maintain higher profitability due to faster turnaround times and better customer satisfaction.

5.2.2. Sales Trends by Product and Region

The analysis of sales trends reveals clear patterns by region. Better infrastructure and operational efficiency are the main drivers of the steady sales growth seen over time in regions like Western Europe and North America. On the other hand, due to delays and logistical challenges, sales growth is slower in regions like South Asia and Eastern Asia. There are also clear seasonal patterns in sales, with peaks in specific months, indicating chances for demand forecasting and inventory management to optimize profits during times of strong demand

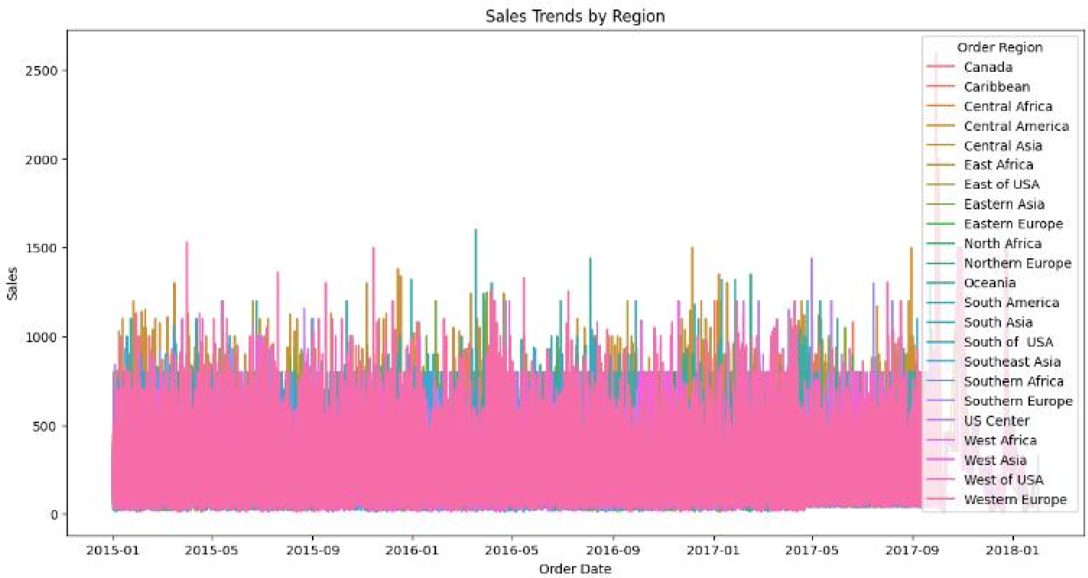

Figure 7. Sales Trends by Region

For further technical analysis, the Prophet algorithm was employed to observe its performance on this dataset, demonstrating its robustness and reliability in predicting late deliveries and segmenting customers effectively. The model gives a range of potential outcomes with a 95% confidence level, which comprise the date stamp (ds), anticipated late delivery risk (yhat), and prediction intervals (yhat_lower and yhat_upper). For instance, the model in this study forecasts a high probability of late deliveries on dates, like January 22, 2019, with a risk of almost 71%. The prediction intervals show that the real risk may vary greatly, demonstrating the model's capacity to take extreme situations and uncertainty into account and guarantee that firms are ready for a variety of outcomes. The Prophet algorithm has demonstrated strong performance in customer segmentation and late delivery risk prediction, making it a valuable tool for businesses seeking to identify high-risk periods, optimize resource allocation, and implement targeted mitigation strategies, such as streamlining delivery routes, increasing workforce capacity, or enhancing inventory control.

6. CONCLUSION

In summary, a thorough grasp of demand patterns across multiple dimensions, such as monthly, annual, and market-specific trends, has been made possible by the dataset analysis. An analysis of marketwise annual demand reveals notable oscillations, with Pacific Asia and USCA being relatively stable, while regions such as LATAM exhibit periods of high demand followed by declines. Product demand trends show recurrent increases during peak seasons, like the fourth quarter, and decreases during off seasons, especially in categories such as electronics. Businesses can use this information to align operational strategies and inventory management with seasonal and market-specific fluctuations.

The identification of low-demand products and shipping inefficiencies highlights critical areas for optimization. Low-demand items present opportunities for cost reduction through phasing out or minimizing stock, while shipping delay analysis underscores the need to improve underperforming delivery modes, such as second-class shipping in regions like South Asia and Eastern Asia. Simultaneously, the correlation between profitability and shipping modes emphasizes the importance of efficient logistics in maintaining customer satisfaction and profit margins.

Additionally, by offering comprehensive demand forecasts and confidence intervals, the Prophet algorithm provided a deeper understanding of data behavior across various data points, allowing businesses to identify patterns, trends, and anomalies more effectively. Rather than merely improving forecasting accuracy, Prophet enabled a more dynamic interpretation of demand fluctuations and delivery risks, making it a valuable tool for predictive decision-making. This capability allowed businesses to optimize inventory levels, anticipate peak demand periods, and refine supply chain strategies with greater confidence.

However, to further enhance projections and manage uncertainties, addressing dataset inconsistencies, such as missing or inconsistent data points is crucial. This underscores the importance of hybrid models to enhance predictive performance by integrating both forecasting and optimization aspects, offering greater adaptability and robustness for a deeper understanding of data points, helping to reduce extreme confidence intervals and enhance demand and inventory predictions [22]. By addressing these challenges, businesses can reduce holding costs, ensure on-time deliveries, and improve demand prediction reliability. Furthermore, Prophet's ability to predict late delivery risks with high precision empowers businesses to proactively mitigate risks through targeted strategies such as inventory management enhancements, workforce adjustments, and route optimizations. These insights not only enhance operational efficiency but also drive higher customer satisfaction and profitability in an increasingly competitive market.

7. FUTURE PROSPECTS

As we continue our study, we aim to develop a flexible and adaptive forecasting model using advanced techniques. With consumer demand constantly changing, markets shifting rapidly, and unexpected disruptions, traditional long-term forecasting may not be effective. To improve accuracy, we would like to identify the best forecasting period for demand and inventory management in largecap businesses operating across different regions, rather than relying only on years of past data. By analyzing various timeframes with AI-driven analytics, we aim to create a more precise and responsive model that helps businesses adapt quickly and make better decisions.

Supplementary Materials: The dataset that supported this research, DataCo Smart Supply Chain for Big Data Analysis, is publicly accessible on Kaggle and can be found at: https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis

Funding Statement: This research was not supported by any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgment: We would like to express our deepest gratitude to our respected mentors, Professor Afshin Zahraee and Professor Lash Mapa, for their invaluable guidance, encouragement, and unwavering support throughout the development of this research. Their insights and mentorship not only shaped our academic thinking but also helped us grow professionally. We are truly thankful for the time and knowledge they shared with us. We also extend our sincere thanks to the College of Technology at Purdue University for providing us with an intellectually stimulating environment to pursue this research. No external funding was received for this project; the work was self-directed and collaboratively completed by the authors.

Contribution: Lavanya Samineni and Sai Saran Ogoti equally contributed to the conceptualization, dataset selection, analysis, model implementation, and manuscript preparation. Both authors were actively involved in discussions, results interpretation, and revision of the content. Prof. Afshin Zahraee and Prof. Lash Mapa served as academic mentors, providing critical feedback, direction, and review support throughout the research process. All authors reviewed and approved the final manuscript.

Informed Consent Statement: Not applicable.

AI Tools Used: We used OpenAI's ChatGPT solely for refining grammar, enhancing clarity, and structuring the manuscript in line with academic standards. No core content, analysis, or ideas were generated using AI tools. We, the authors, take full responsibility for the accuracy, originality, and integrity of the entire manuscript.

Data Availability Statement: The dataset analyzed during the current study is publicly available and can be accessed through Kaggle at DataCo Smart Supply Chain for Big Data Analysis. The data used spans four years (2015–2018) and includes anonymized records from the retail sector, covering electronics, apparel, and sporting goods. No proprietary or confidential datasets were used in this study.

Conflict of Interest Statement: The authors declare no conflict of interest related to the research or publication of this article.

REFERENCES

1. Nasseri, M., Falatouri, T., Brandtner, P., & Darbanian, F. (2023). Applying Machine Learning in Retail Demand Prediction, A comparison of Tree-Based Ensembles and Long Short-Term Memory-Based Deep Learning. Applied Sciences, 13(19), 11112. https://doi.org/10.3390/app131911112.
2. M. Z. Babai, F. Dai et al., "Demand forecasting error under stochastic lead times," International Journal of Production Economics, vol. 231, p. 107878, 2021.
3. P. Choudhury, A. Sahu, and A. Zadeh, "Advances in time-series forecasting: A systematic review," Information Systems, vol. 113, p. 101887, 2022.
4. Lee, C., Lee, H., & Park, S. (2017). Demand forecasting in the supply chain: A systematic literature review.International Journal of Production Research, 55(24), 7407-7423.
5. Falatouri, T., Darbanian, F., Brandtner, P., & Udokwu, C. (2022). Predictive analytics for demand forecasting – A comparison of SARIMA and LSTM in retail SCM. Procedia Computer Science, 200, 993–1003. https://doi.org/10.1016/j.procs.2022.01.298.
6. G, S. F., & N, P. (2020). Machine learning in demand forecasting - A review. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3733548.
7. Marz, M., Mernik, M., & Dervishi, F. (2018). Demand forecasting in inventory problems: A systematic literature review. European Journal of Operational Research, 267(2), 327339.
8. Jha, R., Kumar, M., & Mehta, S. (2022). Machine learning in supply chain demand forecasting: A review. Journal of Intelligent Manufacturing, 33(4), 1047-1065.
9. Bontempi, G., H. Ben Taieb, et al. (2012). Machine learning strategies for time series forecasting. European Journal of Operational Research, 226(1), 1-16.
10. İfraz, B., Uçar, B., et al. (2020). Demand forecasting of spare parts in bus fleets: A comparative study. Journal of Manufacturing Systems, 56, 189-204.
11. Swaminathan, J. M., & Venkitasubramony, K. (2022). Demand forecasting techniques for fashion products: A review. European Journal of Marketing, 56(5), 1168-1192.
12. Huang, L., Yang, Y., et al. (2021). Nonparametric approaches to demand uncertainty in inventory management. Computers & Operations Research, 129, 105190.
13. Zhou, G., Zhao, Y., & Yan, Z. (2022). Partially combining predictors in ensemble learning. Pattern Recognition, 123, 108338.

14. Liu, Y., Wang, J., et al. (2023). Double parallel feed-forward neural network for demand forecasting. Operations Research Letters, 51(2), 213-220.
15. Zhang, C., Zhou, M., & Huang, Y. (2021). Ensemble learning in demand forecasting: A review. Forecasting, 3(1), 12.
16. Mohammed, S. B., & Kora, P. S. (2020). Ensemble deep learning methods for time series forecasting: A review. Journal of Big Data, 7(1), 1-17.
17. Wang, H., Zhang, Y., & Xu, L. (2022). Ensemble deep learning methods for demand forecasting in supply chain management. Decision Support Systems, 145, 113553.
18. Zhang, L., Zhao, Y., & Huang, L. (2020). Combining predictive models with metalearning: A review. Artificial Intelligence Review, 53(3), 1519-1541.
19. Andrade, A. A., & Cunha, J. F. (2019). Disaggregated demand forecasting in the retail industry. Journal of Retailing and Consumer Services, 51, 315-323.
20. DataCo. (2019). SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS. Kaggle. https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis.
21. WuCandice. (n.d.). Superstore-Sales-Analysis. GitHub. https://github.com/WuCandice/Superstore-Sales- Analysis?tab=readme-ov-file#readme.
22. MD Tanvir Islam et al. (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. Retrieved from https://alkindipublisher.com/index.php/jcsts/article/view/6510/5683.