FGS Press

Journal of Computer Allied Intelligence
ISSN: 2584-2676
Vol.03(02), April 2025, pp.28-36

Research Article

Machine Learning in Supply Chain Management

Dhruv Sharma[1], Nitesh[2], Abhirav Khanna[3] , Manish kumar[4], Mohit Tomar[5] and Vipin Kumar Sharma[6,*]

[1-5]Student, Department of Mechanical Engineering, Maharaja Agrasen Institute of Technology, Rohini Sector-22, Delhi-110086

[6]Assistant Professor, Department of Mechanical Engineering, Maharaja Agrasen Institute of Technology, Rohini Sector-22, Delhi-110086

[*]Corresponding Author: Vipin Kumar Sharma. Email: vipinksharma@mait.ac.in

Received: 05/03/2025; Revised: 31/03/2025; Accepted: 20/04/2025; Published: 30/04/2025.

DOI: https://doi.org/10.69996/jcai.2025010

Abstract: The study highlights the application of ML techniques like predictive analytics, optimization algorithms, and advanced demand forecasting in critical areas such as inventory management, supplier selection, logistics optimization, and predictive maintenance. By leveraging ML, businesses can anticipate customer demands with greater accuracy, minimize waste, and respond swiftly to potential disruptions. The findings demonstrate that ML not only enhances decision-making and operational efficiency but also fosters improved customer satisfaction and a stronger competitive edge. This research provides actionable insights into how organizations can harness ML to meet the dynamic demands of modern supply chains and navigate the complexities of a rapidly evolving business landscape.

Keywords: Machine Learning, Maintenance, Inventory, Supply Chain, Predictive Analytics

1.Introduction

Supply Chain Management (SCM) is a vital framework that connects various processes to ensure the efficient movement of goods and services, from raw materials to end customers. Its success lies in effectively integrating planning, sourcing, production, and distribution while adapting to challenges like material shortages, rising costs, and environmental concerns. Advanced technologies, such as machine learning, have further revolutionized SCM by enabling precise demand forecasting, efficient inventory management, and optimized logistics. These tools enhance flexibility, reduce waste, and allow businesses to respond swiftly to market changes and customer demands. While implementing machine learning poses challenges, such as the need for quality data and significant investments, its benefits—improved accuracy, reduced costs, and enhanced operational efficiency— make it indispensable in modern supply chains [1-2]. Ethical practices and a focus on sustainability further strengthen supply chains, fostering trust and loyalty among customers and stakeholders. By blending innovation with strategic planning, businesses can create adaptive, resilient, and future-ready supply chains that drive growth and success in an increasingly competitive market

This paper explains how machine learning (ML) is making supply chains better and more efficient. Supply chains involve different groups, like manufacturers, retailers, and customers, who need to work together to lower costs and meet actual customer needs. Problems like guessing wrong about customer demand and poor communication between groups are big challenges [3-5]. The paper talks about three types of ML: supervised, unsupervised, and reinforcement learning. These methods are used to solve real-world problems in supply chains. Examples include neural networks, decision trees, and support vector machines, which help with things like predicting demand, managing inventory, and planning better delivery routes. It gives examples of companies using ML to save time and money, such as improving warehouses, planning routes, and even helping with online orders. The document also looks to the future, saying ML will keep improving how supply chains work by making smarter decisions and helping people and machines work together. The paper shows how AI and ML can make supply chains faster and more accurate in many industries [6].

2.Related Works

The article "The Role of AI and ML in Revolutionizing Supply Chain Management" explains how artificial intelligence (AI) and machine learning (ML) are making supply chains better. These technologies help businesses handle a lot of data quickly, making smarter decisions like predicting how much of a product is needed, managing inventory, and planning deliveries. AI and ML make supply chains faster, cheaper, and more accurate by finding the best routes, solving problems early, and ensuring products are available. But there are challenges, like bad data, privacy concerns, and high costs, which need careful planning and help from people. The article also talks about how future supply chains could be smarter, able to learn and adjust on their own to meet customer needs and make businesses more efficient. The article "Machine Learning in Supply Chain Management: A Systematic Literature Review" examines the role of machine learning (ML) in improving supply chain management (SCM) [7-9]. It highlights the importance of digital transformation, known as Supply Chain 4.0, for enhancing efficiency and customer responsiveness.

The study identifies key applications of ML, including demand forecasting, inventory management, production planning, and transportation management [10]. By reviewing 40 articles published from 2010 to 2020, the authors aim to understand current ML uses in SCM and propose a taxonomy for better categorization. They also discuss challenges companies face in adopting ML solutions and suggest areas for future research. Ultimately, the paper emphasizes that investing in ML technologies is crucial for businesses seeking to optimize their supply chain processes and boost profitability. The paper looks at how conditional generative adversarial networks (CGANs) can be used in supply chain management (SCM) to make businesses run more efficiently and improve decision-making [11-12]. It focuses on how to choose supply chain partners dynamically by using CGANs to manage many decision factors even when there isn't much data available.

The research highlights the importance of machine learning in predicting inventory and purchasing needs, which helps maintain the right amount of stock [13]. Smart SCM systems can automate tasks and respond better to changes in the market. Successful SCM depends on good teamwork and sharing information among all parties involved, supported by modern technology [14]. The framework promotes ongoing improvement through feedback, allowing companies to adapt and enhance their performance. Overall, the study offers a solid way to use machine learning in SCM for a more flexible and cost-effective supply chain [15]. This research presents a review-based study on SCM4.0, the review includes bibliometric analysis, motives, impediments, and the impact of technologies on distinct SC processes.

3.Proposed Method

Data Cleaning and Modification The initial step in the data analysis process is to clean and prepare the dataset for further analysis. This involves identifying and handling any missing values (NaN), checking for inconsistencies, and ensuring that all the necessary data is complete and properly formatted. We began by reviewing the total number of rows and columns in the dataset to get an overview of the data structure. Once the shape of the data was confirmed, we focused on addressing missing or incomplete information. This could involve filling missing values with appropriate imputed data or removing any irrelevant or incomplete records. This thorough cleaning process ensures that we work with high-quality data, free from errors, and ready for analysis. Distribution of Sales per Customer To explore the spread of sales across customers, we created a histogram, which divides sales data into 30 intervals or bins. This allows us to understand the frequency of different sales amounts, providing a clear view of how customers' spending is distributed. In addition to the histogram, we overlaid a smooth Kernel Density Estimate (KDE) curve, which provides a more continuous representation of the data, helping us identify any trends or patterns in customer spending.

The KDE curve is particularly helpful for spotting areas where sales are concentrated, allowing for a more intuitive understanding of customer behaviour. This combined analysis—histogram and KDE curve—offers a clear, detailed picture of sales distribution and customer spending habits. Sales per Customer by Delivery Status Next, we examined how sales vary depending on the delivery status of an order. We used a box plot to display the range of sales per customer within each delivery status category, such as "Delivered On Time" and "Late Delivery." The box plot is a valuable tool because it not only shows the spread of sales but also highlights the median (or average) sales value, as well as any potential outliers. By comparing the different delivery status categories, we can quickly assess if there are significant differences in sales performance based on whether deliveries were on time or delayed. The box plot helps visualize this in an easy-to-understand manner and reveals any patterns in sales related to delivery performance.

Count of Delivery Status To further understand the distribution of delivery statuses, we used a count plot, which simply counts the number of occurrences of each delivery status category. This provides a quick and clear view of how often each delivery status occurs across the dataset. For instance, we can easily determine how many orders were delivered on time, how many were delayed, and how many-faced other issues. The count plot is a straightforward and effective way to compare the frequency of delivery statuses, helping us identify any trends or imbalances in the dataset. It serves as a foundation for further analysis of how delivery performance affects other factors like customer satisfaction or sales.

Sales vs Profit A critical aspect of understanding business performance is analyzing the relationship between sales and profit. To do this, we used a scatter plot, which plots each data point with sales on the x-axis and profit on the y-axis. By examining the pattern of points, we can determine whether higher sales are associated with higher profits. The scatter plot also helps to identify any anomalies or outliers—cases where sales are high but profit is low, or vice versa. The visualization allows us to examine trends and correlations between these two important metrics, providing valuable insights into the effectiveness of sales strategies and profit generation.

Heatmap for Correlation To understand how different numerical variables relate to each other, we used a heatmap to display the correlation matrix. The correlation matrix measures how strongly two variables are related, with values ranging from -1 (perfect negative correlation) to +1 (perfect positive correlation). A value of 0 indicates no correlation. The heatmap uses color to represent these correlations, with warm colors indicating strong positive relationships and cool colors indicating negative relationships. We specifically focused on the correlation between numerical columns such as sales, profit, and delivery time. The heatmap provides a quick visual summary of these relationships, allowing us to identify any strong dependencies or surprising patterns that could influence further analysis or decision-making

4. Results and Analysis

Data Cleaning and Modification The initial step in the data analysis process is to clean and prepare the dataset for further analysis. This involves identifying and handling any missing values (NaN), checking for inconsistencies, and ensuring that all the necessary data is complete and properly formatted. We began by reviewing the total number of rows and columns in the dataset to get an overview of the data structure. Once the shape of the data was confirmed, we focused on addressing missing or incomplete information. This could involve filling missing values with appropriate imputed data or removing any irrelevant or incomplete records. This thorough cleaning process ensures that we work with high-quality data, free from errors, and ready for analysis.

Distribution of Sales per Customer To explore the spread of sales across customers, we created a histogram, which divides sales data into 30 intervals or bins. This allows us to understand the frequency of different sales amounts, providing a clear view of how customers' spending is distributed. In addition to the histogram, we overlaid a smooth Kernel Density Estimate (KDE) curve, which provides a more continuous representation of the data, helping us identify any trends or patterns in customer spending. The KDE curve is particularly helpful for spotting areas where sales are concentrated, allowing for a more intuitive understanding of customer behaviour. This combined analysis—histogram and KDE curve—offers a clear, detailed picture of sales distribution and customer spending habits.

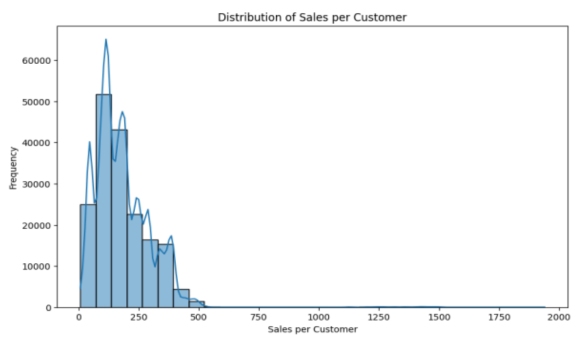

Figure 1: Distribution of sales per customer

Sales per Customer by Delivery Status Next, we examined how sales vary depending on the delivery status of an order. We used a box plot to display the range of sales per customer within each delivery status category, such as "Delivered On Time" and "Late Delivery." The box plot is a valuable tool because it not only shows the spread of sales but also highlights the median (or average) sales value, as well as any potential outliers. By comparing the different delivery status categories, we can quickly assess if there are significant differences in sales performance based on whether deliveries were on time or delayed. The box plot helps visualize this in an easy-to-understand manner and reveals any patterns in sales related to delivery performance.

Figure 2: Sales per customer by delivery status

Count of Delivery Status To further understand the distribution of delivery statuses, we used a count plot, which simply counts the number of occurrences of each delivery status category. This provides a quick and clear view of how often each delivery status occurs across the dataset. For instance, we can easily determine how many orders were delivered on time, how many were delayed, and how many faced other issues. The count plot is a straightforward and effective way to compare the frequency of delivery statuses, helping us identify any trends or imbalances in the dataset. It serves as a foundation for further analysis of how delivery performance affects other factors like customer satisfaction or sales.

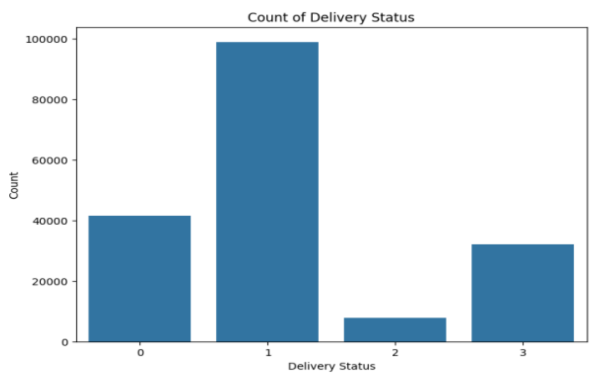

Figure 3: Delivery Status

Sales vs Profit A critical aspect of understanding business performance is analyzing the relationship between sales and profit. To do this, we used a scatter plot, which plots each data point with sales on the x-axis and profit on the y-axis. By examining the pattern of points, we can determine whether higher sales are associated with higher profits. The scatter plot also helps to identify any anomalies or outliers—cases where sales are high but profit is low, or vice versa. The visualization allows us to examine trends and correlations between these two important metrics, providing valuable insights into the effectiveness of sales strategies and profit generation.

Output:

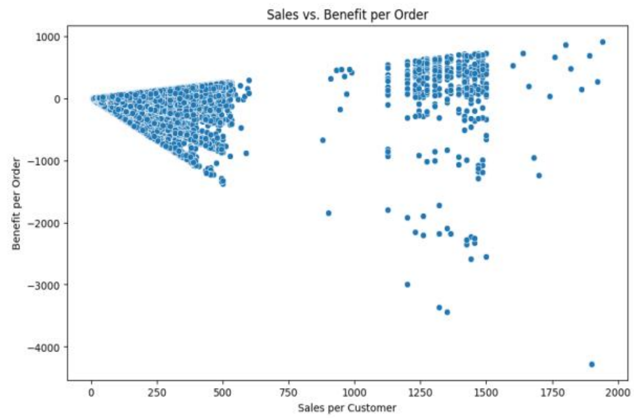

Figure 4: Sales vs. Benefits per order

Heatmap for Correlation To understand how different numerical variables relate to each other, we used a heatmap to display the correlation matrix. The correlation matrix measures how strongly two variables are related, with values ranging from -1 (perfect negative correlation) to +1 (perfect positive correlation). A value of 0 indicates no correlation. The heatmap uses colour to represent these correlations, with warm colours indicating strong positive relationships and cool colours indicating negative relationships. We specifically focused on the correlation between numerical columns such as sales, profit, and delivery time. The heatmap provides a quick visual summary of these relationships, allowing us to identify any strong dependencies or surprising patterns that could influence further analysis or decision-making.

Linear and Logistic Regression Models In our analysis, we used two types of regression models: linear and logistic. The linear regression model was applied to predict continuous outcomes, such as future sales figures, based on other variables in the dataset. It works by finding the best-fit line that describes the relationship between the dependent and independent variables. On the other hand, the logistic regression model was used for classification tasks, such as predicting whether a delivery will be late or on time. This model outputs probabilities that fall between 0 and 1, helping us classify outcomes into discrete categories. Both models are essential for understanding the factors that influence key business metrics and for making predictions about future trends based on historical data.

Table 1: Presenting the Classification Report

Linear Regression Mean Squared Error: 13874.121768697298

Logistic Regression Accuracy: 0.9748504320850875

Classification Report for Logistic Regression:

	precision	recall	f1-score	support
0	1.00	0.94	0.97	16307
1	0.96	1.00	0.98	19797
accuracy			0.97	36104
macro avg	0.98	0.97	0.97	36104
weighted avg	0.98	0.97	0.97	36104

Product Category Distribution To gain a better understanding of the product mix, we analysed the distribution of products across different categories. We created a bar chart that counts the number of orders in each product category. This allows us to see which product categories are the most popular or generate the most sales. The chart is displayed with sky-blue bars for visual clarity, and the xaxis labels (representing product categories) are rotated for better readability. This analysis helps identify trends in product demand and can inform inventory and marketing decisions.

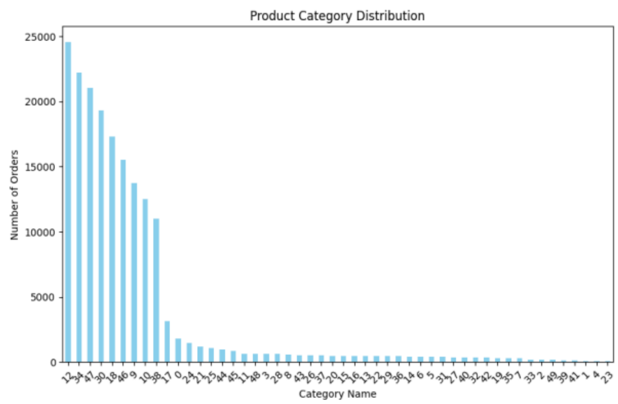

Figure 5: Product Category Distribution

Delivery Status Count (Pie Chart) We created a pie chart to visualize the distribution of delivery statuses, showing what percentage of deliveries were on time, late, or had other statuses. The pie chart provides a clear visual of the proportion of each delivery status, with each slice of the pie labeled with the percentage it represents. Custom colors (lightgreen, salmon, and gold) were used to make the chart visually appealing and easy to understand. This visualization is a simple yet effective way to display the relative frequencies of different delivery statuses in the dataset

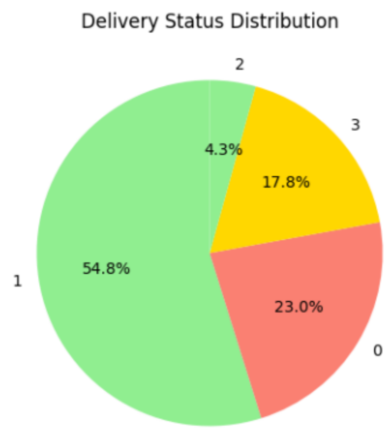

Figure 6: Delivery status Distribution

5.Conclusion

This work underscores the transformative potential of advanced technologies like machine learning and data analytics in revolutionizing supply chain management. The process began with meticulous data cleaning and preparation to ensure the dataset was accurate, complete, and ready for analysis. This step was crucial in laying the foundation for deriving meaningful insights and building predictive models. By addressing missing values and standardizing formats, the data was transformed into a reliable resource for advanced analyses. Exploratory visualizations provided a window into the dynamics of sales distribution, customer spending habits, and delivery performance. Tools like histograms and box plots revealed trends and patterns, offering a deeper understanding of consumer behavior and operational efficiency. Scatter plots illuminated the relationship between sales and profit, highlighting areas where performance could be optimized. These insights served as a guiding framework for further analysis and decision-making. The integration of machine learning models further elevated the analytical capabilities of this project. Linear regression was instrumental in predicting sales trends, while logistic regression and Random Forest Classifiers identified and classified late delivery risks with remarkable accuracy. The application of these models showcased how predictive analytics could improve planning and mitigate risks in supply chains.

Heatmaps added another layer of insight, illustrating correlations between key variables like delivery times, sales, and profits, which proved invaluable in identifying actionable areas for improvement. Time-series analyses, such as monthly sales trends and average delivery times, offered a clearer picture of operational fluctuations throughout the year. These insights are essential for forecasting and preparing for seasonal demands, ensuring a smoother flow of goods and services. The analysis of product categories and sales by region provided strategic insights into high-performing areas and customer preferences, enabling data-driven decision-making to enhance marketing and inventory strategies. This comprehensive exploration demonstrates how a data-driven approach to supply chain management can lead to significant improvements in efficiency, cost-effectiveness, and customer satisfaction. By harnessing the power of machine learning and visualization tools, businesses can not only address current challenges but also proactively prepare for future uncertainties. The findings of this project underscore the critical role of technological integration in creating smarter, more adaptive supply chains that can thrive in a competitive and ever-changing landscape.

Acknowledgment: Not Applicable.
Funding Statement: The author(s) received no specific funding for this study.
Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

[1] Makkar, "Applications of Machine Learning Techniques in Supply Chain Optimization," *System Reliability, Quality Control, Safety, Maintenance and Management ICICCT(2019*), pp.861-869, 2020

[2] Mallesham, "The Role of AI and ML in Revolutionizing Supply Chain Management," *International Journal of Scientific Research and Management (IJSRM),* vol. 10, no.06, pp.918-928, 2022

[3] Mohamed-Iliasse, "Machine Learning in Supply Chain Management: A Systematic Literature Review," *International Journal of Supply and Operation Management (IJSOM*), vol. 9, no.4, pp.398-416, 2022

[4] Lin, "An innovative machine learning model for supply chain management," *Journal of Innovation & Knowledge (JIK*), vol. 7, no.4, 2022

[5] Srinivasa Sai Abhijit Challapalli, "Optimizing Dallas-Fort Worth Bus Transportation System Using Any Logic," *Journal of Sensors, IoT & Health Sciences*, vol.2, no.4, pp.40-55, 2024.

[6] M.I. Mahraz, L. Benabbou and A. Berrado, "Machine learning in supply chain management: A systematic literature review," *Int. J. Supply Oper. Manag*, vol.2022, no.9, pp.398–416, 2022.

[7] E. Theodorou, E. Spiliotis and V. Assimakopoulos, "Optimizing inventory control through a data-driven and model-independent framework," *EURO J. Transp. Logist*, vol. 12, pp.100103, 2023.

[8] Srinivasa Sai Abhijit Challapalli, "Sentiment Analysis of the Twitter Dataset for the Prediction of Sentiments," *Journal of Sensors, IoT & Health Sciences*, vol.2, no.4, pp.1-15, 2024.

[9] M. Bertolini, D. Mezzogori, M. Neroni and F. Zammori, "Machine Learning for industrial applications: A comprehensive literature review," *Expert Syst. Appl.*, vol. 175, pp.114820, 2021.

[10] A.C. Odimarha, S.A. Ayodeji and E.A. Abaku, "Machine learning's influence on supply chain and logistics optimization in the oil and gas sector: A comprehensive analysis," *Comput. Sci. IT Res. J.*, vol. 5, pp.725–740, 2024.

[11] M. Elahi, S.O. Afolaranmi, J.L. Martinez Lastra and J.A. Perez Garcia, "A comprehensive literature review of the applications of AI techniques through the lifecycle of industrial equipment," *Discov. Artif. Intell*., vol. 43, 2023.

[12] B. Venkateswarlu and Rekha Gangula, "Exploring the Power and Practical Applications of K-Nearest Neighbours (KNN) in Machine Learning," *Journal of Computer Allied Intelligence*, vol.2, no.1, pp.8-15, 2024.

[13] B.Ashok Kumar, K.Vijayachandra, G.Naveen Kumar and V.N.Lakshmana Kumar, "Blockchain Technology Communication Technology Model for the IoT," *Journal of Computer Allied Intelligence*, vol.2, no.4, pp.20-35, 2024.

[14] G.K.Uyanık and N. Güler, "A study on multiple linear regression analysis," *Procedia Soc. Behav. Sci.*, vol. 106, pp.234–240, 2013.

[15] K.P. Sinaga and M.S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp.80716–80727, 2020.