SUPPLY CHAIN DELIVERY ANALYSIS AND PREDICTION

[1] K Srikanth, [2] D Venkata Akash Babu, [3] A Srijan, [4] M Charvini
[1]Assistant Professor, [234]Students
Department of CSE
Siddhartha Institute of Technology & Sciences, Narapally
n.ramya@siddhartha.org.in, 23tq1a0555@siddhartha.co.in, 23tq1a0556@siddhartha.co.in, 23tq1a0558@siddhartha.co.in,

Abstract

The Supply Chain Delivery Analysis and Prediction system is designed to analyze supply chain delivery data and predict delivery performance using data analytics and machine-learning techniques. In modern supply chains, timely delivery is essential for maintaining customer satisfaction, reducing operational costs, and ensuring smooth business operations. Delays can occur due to factors such as transportation problems, warehouse processing time, inventory availability, weather conditions, route distance, supplier performance, and order volume. The proposed system collects historical supply chain information such as order details, product category, supplier, warehouse, order date, shipping date, delivery date, transportation mode, distance, shipping cost, delivery status, and delay information. The collected data is preprocessed to handle missing values, duplicate records, inconsistent information, and categorical attributes. Exploratory Data Analysis is performed to identify delivery patterns, delay trends, supplier performance, transportation efficiency, and warehouse-related issues. Data visualization techniques are used to represent delivery times, delayed orders, supplier performance, transportation modes, and monthly delivery trends through charts and dashboards. Relevant features are selected and transformed into a suitable format for machine-learning models. Classification algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Gradient Boosting can be used to predict whether an order is likely to be delivered on time or delayed. Regression algorithms can also be applied to estimate the expected delivery duration when sufficient historical data is available. The trained models are evaluated using suitable metrics such as accuracy, precision, recall, F1-score, ROC-AUC, MAE, RMSE, and R^2, depending on the prediction objective. The system provides analytical reports and prediction results that can help supply chain managers identify potential delivery delays and take preventive actions. Overall, the Supply Chain Delivery Analysis and Prediction system combines data analysis, visualization, and machine learning to improve delivery planning, reduce delays, optimize logistics operations, and support data-driven decision-making in supply chain management.

I. Introduction

In modern business operations, an efficient supply chain is essential for ensuring that products reach customers at the right time and in the right condition. Delivery performance directly affects customer satisfaction, operational costs, inventory management, and overall business productivity. Delays in delivery can occur due to transportation issues, warehouse processing, inventory shortages, supplier problems, traffic conditions, weather, and inefficient routing. Traditional supply chain management methods often depend on manual tracking and historical reports, making it difficult to identify potential delays in advance. Therefore, analyzing historical delivery data and predicting future delivery performance can help organizations improve logistics efficiency and customer service.

The Supply Chain Delivery Analysis and Prediction system is developed to analyze historical supply chain and delivery information and predict possible delivery outcomes. The system maintains important details such as order ID, product category, supplier, warehouse, order date, shipping date, delivery date, transportation mode, distance, shipping cost, delivery status, and delay information. By storing these records in a structured format, the system provides a centralized approach for monitoring delivery activities. It also helps supply chain managers understand delivery patterns and identify factors that contribute to delays.

The system begins with the collection of historical supply chain data from company records or suitable datasets. The collected data may contain missing values, duplicate records, inconsistent dates, incorrect values, and categorical attributes such as transportation mode, supplier, and product category. Data preprocessing techniques are applied to clean and transform the data into a suitable format for analysis. Missing values and duplicate records are handled appropriately, while categorical variables can be encoded for machine-learning models. Date-related features such as delivery duration, shipping duration, and order processing time can also be derived from the available dates.

II. Literature Survey

The growth of e-commerce and global business has increased the importance of efficient supply chain and delivery management. Researchers have increasingly used data analytics and machine-learning techniques to study delivery times, transportation performance, logistics costs, inventory conditions, and potential delivery delays. Historical supply chain data can provide valuable information about factors that influence delivery performance, including supplier reliability, transportation mode, order volume, warehouse processing, distance, and seasonal conditions. These studies provide the foundation for developing intelligent systems that can analyze delivery performance and predict future delivery outcomes.

Research on supply chain analytics has shown that data-driven decision-making can improve logistics planning and operational efficiency. Analytics techniques can be used to examine historical orders, inventory levels, transportation activities, and supplier performance. By identifying patterns in historical data, organizations can understand the causes of delays and improve resource allocation. Such approaches demonstrate the importance of integrating data collection, preprocessing, analysis, and visualization into modern supply chain management systems.

Studies on delivery time prediction have investigated the use of machine-learning algorithms to estimate the time required for orders to reach customers. Factors such as order characteristics, shipping distance, transportation method, warehouse processing time, and historical delivery performance can be used as predictive features. Regression algorithms can estimate delivery duration, while classification algorithms can determine whether an order is likely to arrive on time or be delayed. These approaches provide useful information for logistics planning and customer communication.

Research on machine learning in logistics has explored algorithms such as Decision Trees, Random Forest, Support Vector Machines, Logistic Regression, Gradient

Boosting, and Neural Networks. Tree-based models are particularly useful for supply chain datasets because they can capture nonlinear relationships between multiple operational factors. Machine-learning models can identify relationships that may not be easily detected through traditional statistical analysis. Model evaluation using metrics such as accuracy, precision, recall, F1-score, MAE, RMSE, and R^2 helps determine the suitability of different algorithms for specific prediction tasks.

Several studies have focused on last-mile delivery optimization, which is one of the most important stages of the supply chain. Delivery performance in the last mile can be affected by traffic congestion, delivery distance, customer availability, vehicle capacity, weather, and route selection. Data-driven approaches can analyze historical delivery information and help organizations identify locations, routes, and time periods associated with frequent delays. Predictive analytics can therefore support better route planning and delivery scheduling.

Research involving IoT and real-time logistics monitoring has further improved supply chain visibility. GPS devices, sensors, and connected transportation systems can provide real-time information about vehicle location, temperature, speed, route conditions, and shipment status. Combining real-time information with historical data allows predictive systems to detect potential delays earlier. Such systems can provide alerts to logistics managers and enable them to take corrective actions before a delivery problem significantly affects customers.

Studies on supplier and transportation performance analysis have also demonstrated the importance of monitoring different supply chain participants. Supplier reliability, order processing time, transportation mode, shipping cost, and historical delivery performance can significantly influence overall delivery efficiency. Comparing these factors through analytical dashboards can help organizations identify reliable suppliers and inefficient logistics processes. This information can support decisions related to supplier selection, transportation planning, and warehouse operations.

Recent research has increasingly focused on Artificial Intelligence, deep learning, and predictive supply chain management. Advanced models can process large and complex datasets containing order, transportation, inventory, and environmental information. These models can be used to forecast demand, predict delivery delays, optimize routes, and improve inventory planning. However, complex machine-learning models may be difficult for supply chain managers to interpret, making model explainability and clear visualization important for practical implementation.

III. System Analysis

The **Supply Chain Delivery Analysis and Prediction** system is designed to analyze delivery operations and predict possible delivery delays using historical supply chain data. The system maintains important information such as order details, product category, supplier, warehouse, transportation mode, order date, shipping date, delivery date, distance, shipping cost, and delivery status. Historical delivery records are used to identify patterns and factors that influence delivery performance. The system first collects and organizes supply chain data for further processing and analysis. Data preprocessing is performed to handle missing values, duplicate records, inconsistent dates, and categorical attributes. Exploratory Data Analysis is conducted to understand delivery times, delayed orders, supplier performance, transportation efficiency, and warehouse processing patterns. Data visualization techniques are used to represent on-time deliveries, delayed deliveries, delivery duration, shipping costs, and monthly trends. Relevant features are selected and transformed into a suitable format for machine-learning algorithms. Classification models can be trained to predict whether an order is likely to be delivered on time or experience a delay. Regression models can also be used to estimate the expected number of days required for delivery. The trained models are evaluated using suitable performance metrics to identify an effective prediction model. Overall, the system converts supply chain delivery data into useful insights and predictions that can support better logistics planning and operational decision-making.

Existing System

In the existing system, organizations generally manage delivery information using spreadsheets, manual records, basic logistics software, or individual tracking systems. Order details, shipment information, transportation details, and delivery status are recorded during different stages of the supply chain. Traditional systems mainly focus on tracking current orders rather than analyzing historical delivery performance. Logistics staff may manually calculate delivery duration, delayed orders, shipping costs, and supplier performance. When the number of orders increases, maintaining and analyzing large amounts of delivery data becomes difficult and time-consuming. Historical delivery information may be stored but is not always effectively used to identify recurring causes of delays. Traditional systems generally have limited capabilities for predicting future delivery delays. It can also be difficult to compare the performance of suppliers, warehouses, transportation modes, and product categories. Generating detailed delivery reports and visualizations may require additional manual effort. Unexpected delays can affect customer satisfaction, inventory availability, and overall operational costs. The absence of predictive analytics makes it difficult for logistics teams to take corrective action before a delivery becomes delayed. Therefore, an intelligent system is required to combine delivery tracking, data analysis, visualization, and machine-learning-based prediction.

Disadvantages of Existing System

- Heavy dependence on manual delivery tracking.
- Difficult to manage large numbers of orders and shipment records.
- Greater possibility of data-entry and calculation errors.
- Limited analysis of historical delivery patterns.
- Limited ability to predict delivery delays in advance.
- Difficult to identify high-risk orders early.
- Limited visualization of delivery performance.
- Difficult to compare supplier and transportation performance.
- Historical delivery data may not be fully utilized.
- Generating detailed reports may require additional manual effort.
- Unexpected delivery delays can affect customer satisfaction.
- Limited support for proactive logistics planning.

Proposed System

The proposed **Supply Chain Delivery Analysis and Prediction** system provides an integrated platform for tracking, analyzing, and predicting supply chain delivery performance. The system stores information such as order details, product category, supplier, warehouse, transportation mode, order date, shipping date, delivery date, distance, shipping cost, and delivery status. Historical delivery records are analyzed to identify important factors associated with on-time and delayed deliveries. The collected data is first preprocessed by handling missing values, duplicate records, inconsistent dates, and categorical attributes. Exploratory Data Analysis is performed to understand delivery duration, delay frequency, supplier performance, transportation efficiency, and warehouse processing time. Visualization techniques are used to display delivery trends, delayed orders, shipping costs, supplier performance, and transportation statistics through dashboards. Relevant order and logistics features are selected and transformed into a suitable format for machine-learning models. Classification algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Gradient Boosting can be used to predict possible delivery delays. Regression algorithms can also be applied to estimate the expected delivery duration. The trained models are evaluated using appropriate performance metrics to identify the most effective prediction model. The system can identify high-risk orders and provide useful insights to logistics managers for taking preventive actions. Overall, the proposed system combines delivery tracking, data analysis, visualization, and machine learning to improve supply chain efficiency and reduce delivery delays.

Advantages of Proposed System

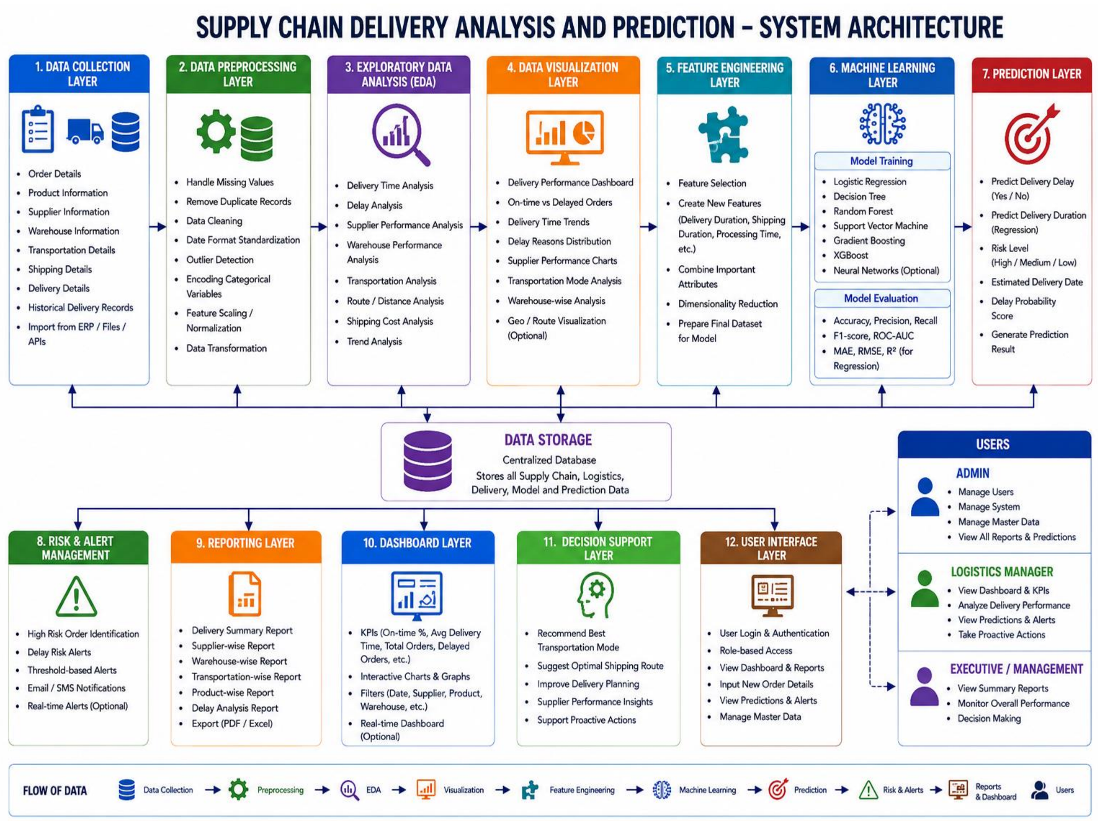
- Automates supply chain delivery tracking.
- Reduces manual effort in maintaining delivery records.
- Minimizes data-entry and calculation errors.
- Efficiently manages large amounts of order data.
- Predicts possible delivery delays in advance.
- Helps identify high-risk orders.
- Analyzes historical delivery patterns.
- Provides supplier-wise performance analysis.
- Provides transportation-wise delivery analysis.
- Provides warehouse and product-wise analysis.
- Displays delivery trends through charts and dashboards.
- Helps improve logistics planning and scheduling.
- Supports proactive action against potential delays.
- Helps improve customer satisfaction.
- Supports data-driven supply chain decision-making.
- Can be extended with real-time GPS and IoT tracking.

IV. Methodology

The methodology of the **Supply Chain Delivery Analysis and Prediction** system consists of several stages, beginning with data collection and ending with delivery prediction and reporting. First, historical supply chain data containing order, supplier, warehouse, transportation, shipping, and delivery information is collected. The dataset is examined to understand its attributes, data types, missing values, and target variable. Data preprocessing is performed to handle missing values, duplicate records, inconsistent dates, and categorical attributes. Date-based features such as shipping duration, delivery duration, and order processing time are derived from the available information. Categorical variables such as supplier, product category, warehouse, and transportation mode are converted into numerical representations when required. Exploratory Data Analysis is performed to understand delivery trends, delay patterns, supplier performance, and relationships between different logistics variables. Data visualization techniques are used to identify important patterns through charts, graphs, and dashboards. Relevant features such as shipping distance, transportation mode, supplier history, warehouse processing time, shipping cost, and order information are selected for prediction. The processed dataset is divided into training and testing datasets for developing and evaluating machine-learning models. Classification algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Gradient Boosting can be trained to predict delivery delays, while regression algorithms can be used to predict delivery duration. Finally, the best-performing model is selected based on suitable evaluation metrics, and the system displays delivery analysis, risk predictions, and performance reports through a user-friendly interface.

System Architecture

The Supply Chain Delivery Analysis and Prediction system architecture represents the complete flow of supply chain data from collection to delivery prediction and reporting.



The system starts with the Data Collection Layer, where order, product, supplier, warehouse, transportation, shipping, and delivery information is collected. The collected data is transferred to the Data Preprocessing Layer, where missing values, duplicate records, inconsistent dates, and categorical attributes are handled. The cleaned data is then passed to the Data Analysis Layer, where delivery duration, delay frequency, supplier performance, and transportation efficiency are analyzed. The Visualization Layer presents delivery trends, delayed orders, shipping costs, and logistics performance using charts, graphs, and dashboards. The Feature Engineering Layer selects and transforms relevant order and logistics attributes required for prediction. The processed data is provided to the Machine Learning Layer, where classification or regression models are trained using historical delivery records. The trained models are evaluated using suitable performance metrics to select the most effective prediction model. The selected model is connected to the Prediction Layer, which receives new order and logistics information and predicts possible delivery delays or expected delivery duration. The Risk Analysis Layer identifies high-risk orders and provides information that can support proactive logistics decisions. Finally, the User Interface Layer allows logistics managers and authorized users to monitor deliveries, analyze supply chain performance, and view prediction results, completing the flow **Data Collection → Preprocessing → Analysis → Visualization → Feature Engineering → Machine Learning → Model Evaluation → Prediction → Risk Analysis → Reporting → User Interface**.

V. Result and Output

VI. Conclusion

The SUPPLY CHAIN DELIVERY ANALYSIS AND PREDICTION system provides an efficient solution for monitoring, analyzing, and predicting supply chain delivery performance. The system maintains important information such as order details, product category, supplier, warehouse, transportation mode, shipping date, delivery date, shipping cost, and delivery status. By organizing historical delivery records in a structured manner, the system makes it easier to understand delivery patterns and logistics performance. Data preprocessing techniques improve the quality of the dataset by handling missing values, duplicate records, inconsistent dates, and categorical information. Exploratory Data Analysis and visualization provide useful insights into on-time deliveries, delayed orders, supplier performance, transportation efficiency, and delivery trends. Machine-learning algorithms can analyze historical delivery information and predict whether new orders are likely to be delivered on time or experience delays. The prediction results can help logistics managers identify high-risk orders and take preventive actions before delays occur. The dashboard provides clear charts, tables, key performance indicators, and prediction results for effective supply chain monitoring. The system can also support supplier comparison, transportation analysis, warehouse evaluation, and delivery performance assessment. By combining data analytics with machine learning, the system reduces manual analysis and supports faster, data-driven logistics decisions. In the future, the system can be enhanced with real-time GPS tracking, traffic and weather data, IoT sensors, automated customer notifications, route optimization, and advanced predictive models. Overall, the project demonstrates how data analytics and machine learning can improve delivery planning, reduce delays, optimize supply chain operations, and enhance customer satisfaction.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.
[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.
[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm
[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.
[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.
[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, “Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks,” in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.
[7] R. D. Kumar, V. N. S.Manaswini, “Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology,” in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.
[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.
[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.