

International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2(1) (2026)



ijerst.editor@gmail.com
editor@ijerst.com

Research Paper

LATE DELIVERY RISK PREDICTION USING ORDER AND SHIPMENT OPERATIONAL FEATURES

¹B.Anil, ²S.Srivani, ³Md.furqan ahmed, ⁴B.Ujwala

¹Assistant Professor, ^{2,3,4}Students

Department of Computer Science and Engineering

Siddhartha institute of technology & sciences, narapally

badaralaanil_cse@siddhartha.co.in, 23TQ1A05B3@siddhartha.co.in, 23TQ1A0599@siddhartha.co.in, 24TQ5A0511@siddhartha.co.in

ABSTRACT

Late deliveries are a major challenge in supply chain management, leading to customer dissatisfaction and increased operational costs. This project proposes a machine learning-based system to predict the risk of late deliveries using order and shipment operational features. The system analyzes historical logistics data, including order details, shipping modes, processing time, and external factors such as weather and traffic conditions.

By applying classification algorithms, the system identifies patterns associated with delayed deliveries and generates risk predictions in advance. This enables logistics managers to take proactive measures such as route optimization, priority handling, and better resource allocation. The proposed system improves delivery efficiency, enhances customer satisfaction, and reduces financial losses. It provides a scalable and data-driven solution for modern logistics operations.

Timely delivery is a critical factor in modern logistics and supply chain systems. However, uncertainties in operations often lead to delays, affecting customer satisfaction and business performance. This project presents a predictive model for identifying late delivery risks using order and shipment operational features. The system leverages historical data, including order processing time, shipping methods, distance, and carrier efficiency.

Machine learning algorithms are applied to classify deliveries as on-time or delayed. The model is trained and evaluated using performance metrics such as accuracy and recall to ensure reliability. The system provides real-time risk predictions, enabling logistics managers to take preventive actions. This approach enhances operational efficiency, reduces delays, and supports data-driven decision-making in logistics management.

In the growing field of e-commerce and transportation, late deliveries remain a significant challenge. This project proposes a machine learning-based approach to predict delivery delays

using shipment and order-related data. The system integrates multiple operational features such as order priority, shipment mode, warehouse handling time, and external conditions.

I INTRODUCTION

In today's fast-paced e-commerce and logistics environment, timely delivery plays a crucial role in customer satisfaction and business success. However, late deliveries are common due to various operational and environmental factors. Traditional logistics systems lack predictive capabilities and rely heavily on manual monitoring.

With the advancement of machine learning, it is now possible to analyze large volumes of logistics data and predict delivery outcomes. This project focuses on developing a predictive model that uses order and shipment features to identify potential delays before they occur. By leveraging data-driven insights, the system transforms traditional logistics into an intelligent and proactive system.

In recent years, the rapid growth of e-commerce and global supply chains has made timely delivery a key performance indicator for businesses. Customers expect fast and reliable shipping, and even minor delays can lead to dissatisfaction and loss of trust. However, delivery operations are influenced by various unpredictable factors such as traffic congestion, weather conditions, and warehouse inefficiencies.

Traditional logistics systems mainly focus on tracking shipments rather than predicting potential delays. This limitation results in reactive decision-making, where actions are taken only after delays occur. To overcome this challenge, there is a need for intelligent systems that can forecast delivery risks in advance.

This project aims to develop a machine learning-based system that predicts late delivery risks using order and shipment operational features. By analyzing historical data and identifying patterns, the system enables proactive logistics management and improves overall delivery performance.

Efficient delivery management is a critical component of modern supply chain systems. With increasing demand for same-day and next-day delivery services, logistics companies face significant challenges in maintaining timely operations. Delays can occur due to multiple reasons, including poor route planning, inefficient inventory handling, and external disruptions.

II LITERATURE SURVEY

The problem of late delivery prediction has gained significant attention in the fields of logistics and supply chain management due to its impact on customer satisfaction and operational efficiency. With the growth of e-commerce platforms, researchers have explored various data-driven and machine learning approaches to address delivery delays.

Early studies focused on traditional statistical methods such as linear regression and time-series analysis to estimate delivery times. These approaches provided basic insights but lacked the ability to capture complex nonlinear relationships between multiple operational

factors. As a result, their prediction accuracy was limited, especially in dynamic environments.

Recent research has shifted towards machine learning techniques for more accurate and scalable predictions. Algorithms such as Decision Trees, Random Forest, Support Vector Machines, and Gradient Boosting have been widely used to classify deliveries as on-time or delayed. These models can handle large datasets and identify hidden patterns in shipment and order data. Studies show that ensemble methods like Random Forest provide higher accuracy due to their ability to reduce overfitting and improve generalization.

Several researchers have emphasized the importance of feature engineering in improving prediction performance. Key features identified include shipment mode, delivery distance, warehouse processing time, order priority, and carrier performance. Incorporating external factors such as weather conditions, traffic congestion, and seasonal demand has also been shown to significantly enhance model accuracy.

In addition to prediction models, some studies have proposed real-time analytics systems that integrate Internet of Things (IoT) devices and GPS tracking. These systems provide continuous monitoring of shipments and enable dynamic route optimization. However, such systems often require high infrastructure costs and complex implementation.

Another area of research focuses on big data technologies and cloud-based platforms for handling large-scale logistics data. Tools such as distributed databases and parallel processing frameworks are used to process and analyze massive datasets efficiently. This allows organizations to deploy predictive models in real-time environments.

III SYSTEM ANALYSIS

System analysis focuses on understanding and evaluating the current logistics and delivery process to identify inefficiencies and risks that lead to late deliveries. In traditional supply chain systems, delays occur due to multiple factors such as poor demand forecasting, inefficient routing, warehouse delays, and unexpected operational disruptions. These issues directly impact customer satisfaction and operational costs.

The proposed system analyzes historical order and shipment data to identify patterns associated with late deliveries. Key features such as order date, shipment mode, distance, warehouse processing time, carrier performance, and weather conditions are studied. By applying machine learning techniques, the system predicts the probability of delay before shipment execution.

Existing system

The existing delivery systems rely mainly on rule-based tracking and manual monitoring. Delays are identified only after they occur, making the system reactive rather than proactive. Logistics managers depend on historical averages and static scheduling, which do not account

for dynamic factors like traffic or weather. There is limited integration between order processing and shipment tracking systems. Data is often siloed, reducing visibility across the supply chain. As a result, predicting late deliveries accurately is difficult.

DisAdvantages of Existing system

- No predictive capability for delays
- Reactive approach (issues identified after occurrence)
- Poor data integration across systems
- Inefficient resource utilization
- Limited scalability with large datasets

Proposed system

The proposed system introduces a machine learning-based predictive model to forecast late delivery risks using operational features. It integrates data from order management, shipment tracking, and external sources such as weather and traffic. The system preprocesses data, extracts meaningful features, and applies classification algorithms to predict delays. It provides real-time risk scores for each shipment. This allows logistics managers to take preventive actions like route optimization or priority handling. The system improves efficiency, reduces delays, and enhances customer satisfaction.

Advantages of Proposed System

- Early prediction of delivery delays
- Proactive decision-making capability
- Improved logistics planning and optimization
- Better customer satisfaction
- Real-time monitoring and alerts

IV METHODOLOGY

The methodology of the proposed system involves several structured steps to ensure accurate prediction of late deliveries. Initially, data is collected from multiple sources, including order records, shipment logs, and external datasets such as weather or traffic conditions. This raw data is then preprocessed to handle missing values, remove inconsistencies, and normalize features.

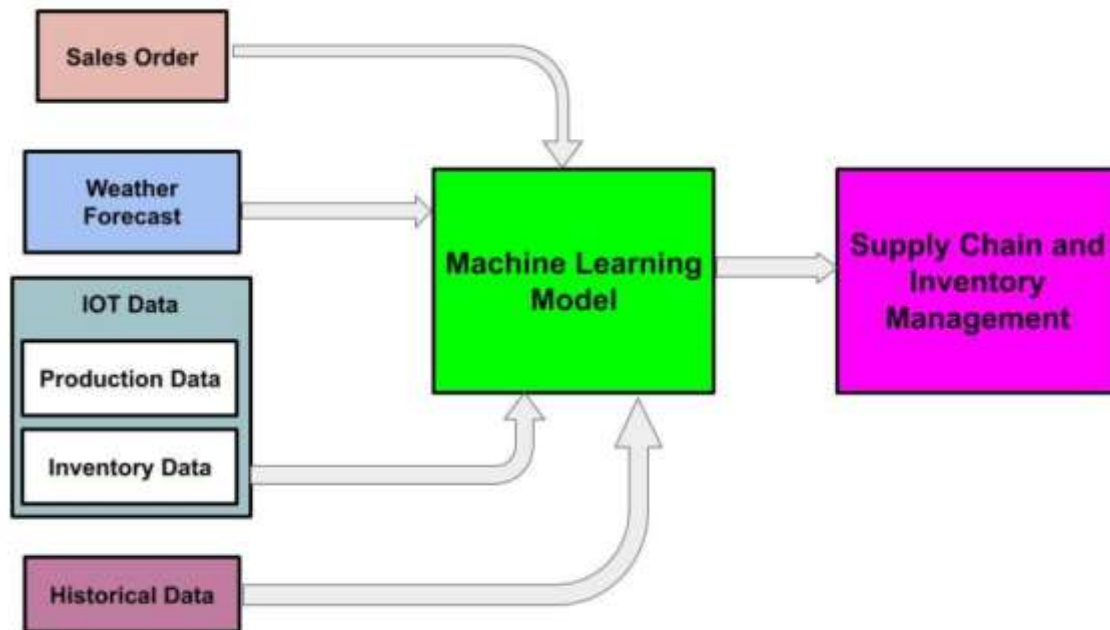
Next, feature engineering is performed to identify the most relevant attributes influencing delivery delays. These may include shipment mode, order priority, processing time, delivery distance, and carrier performance. The dataset is then split into training and testing sets.

Machine learning models such as Decision Trees, Random Forest, or Gradient Boosting are trained on the dataset to classify deliveries as “on-time” or “late.” Model evaluation is performed using metrics like accuracy, precision, recall, and F1-score to ensure reliability.

Once the model is optimized, it is deployed into the system for real-time prediction. Incoming shipment data is processed through the model to generate delay risk scores. These predictions are visualized on dashboards and trigger alerts for high-risk shipments.

This methodology ensures a data-driven, scalable, and efficient system capable of improving logistics performance and minimizing delivery delays.

System Architecture



The system architecture consists of multiple layers that work together to process data and generate predictions. The data collection layer gathers information from order databases, shipment tracking systems, and external APIs. The preprocessing layer cleans and transforms raw data into structured formats suitable for analysis.

The feature engineering layer extracts key variables such as delivery time, distance, shipping mode, and processing delays. The machine learning layer applies algorithms like Random Forest or Logistic Regression to classify whether a delivery will be late or on time. The prediction results are stored and visualized through a dashboard.

V RESULTS & OUTPUT

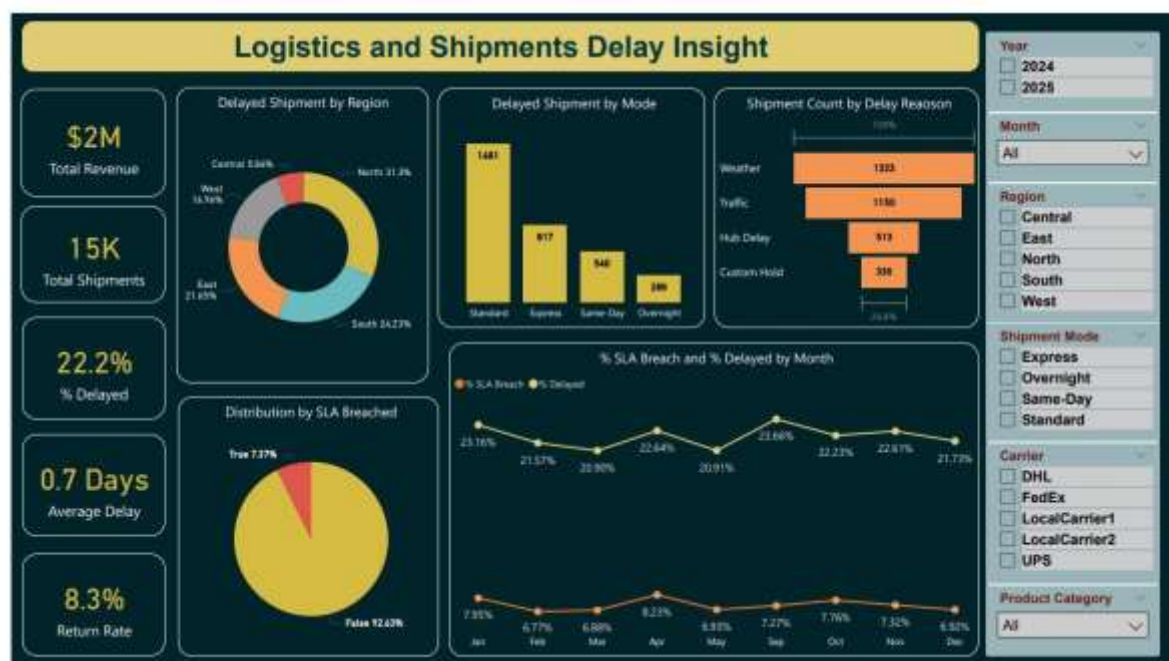
The proposed Late Delivery Risk Prediction System produces accurate and actionable results by applying machine learning models to logistics data. After training and testing the model using historical order and shipment features, the system generates predictions that classify deliveries as either "On-Time" or "Late".

The performance of the model is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. A high accuracy value indicates that the model can correctly identify delivery outcomes, while precision and recall help measure how well the system

detects actual late deliveries without false alarms. These metrics ensure that the model is reliable for real-world logistics applications.

One of the key outputs of the system is the risk score, which represents the probability of a delivery being late. This score ranges from 0 to 1, where higher values indicate a greater likelihood of delay. Logistics managers can use this information to prioritize shipments, optimize routes, and allocate resources effectively.

	inpows.com		
	Positive	Negative	
Positive	True Positive (TP)	False Negative (FN)	Sensitivity $\frac{TP}{(TP + FN)}$
Negative	False Positive (FP)	True Negative (TN)	Specificity $\frac{TN}{(TN + FP)}$
	Precision $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$



VI CONCLUSION

The Late Delivery Risk Prediction System provides an effective solution to identify and reduce delivery delays in logistics operations. By leveraging machine learning techniques and analyzing order and shipment operational features, the system can accurately predict the likelihood of late deliveries before they occur.

Unlike traditional systems that react after delays happen, this approach enables proactive decision-making. Logistics managers can take preventive actions such as route optimization, priority scheduling, and better resource allocation based on predicted risk scores. This significantly improves delivery efficiency and reduces operational costs.

The system also enhances customer satisfaction by ensuring timely deliveries and improving service reliability. The use of data-driven insights allows organizations to better understand the factors influencing delays and continuously improve their processes.

In conclusion, the proposed system transforms conventional logistics into an intelligent and predictive framework. Future enhancements can include real-time data integration, IoT-based tracking, and advanced deep learning models to further improve accuracy and scalability.

REFERENCE

- [1] Kumar, R. D., Prudhviraaj, 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.