JOSCM Journal of Optimization and Supply Chain Management
Print ISSN:3079-1022 Online ISSN:3079-1030

Machine Learning Models for Optimizing Online Order Fulfillment - Forecasting Lead Time and Late Delivery Risk

Nusrat Sultana [1]✉, **Abdullah Muhammad Moosa** [2], **S. M. Fahim Faisal** [3], **Tasfia Noor Chowdhury** [4]

1. Corresponding author, Department of Mechatronics & Industrial Engineering, Chittagong University of Engineering & Technology, Chittagong, Bangladesh. E-mail: nusratsultana@cuet.ac.bd
2. Department of Mechatronics & Industrial Engineering, Chittagong University of Engineering & Technology, Chittagong, Bangladesh. E-mail: abdullahmoosa128@gmail.com
3. Department of Mechatronics & Industrial Engineering, Chittagong University of Engineering & Technology, Chittagong, Bangladesh. E-mail: fahimfaisal@cuet.ac.bd
4. Department of Mechatronics & Industrial Engineering, Chittagong University of Engineering & Technology, Chittagong, Bangladesh. E-mail: tasfia.chowdhury@gmail.com

Article Info

Article type:
Research Article

Article history:
Received March 3, 2025
Received in revised form May 17, 2025
Accepted May 30, 2025
Available online June 20, 2025

Keywords:
E-commerce
delivery lead time
late delivery risk
machine learning
supply chain management

ABSTRACT

Objective: With an emphasis on evaluating delivery lead times and anticipating late delivery risks, this study investigates the application of machine learning models to anticipate delivery performance in the e-commerce industry.

Methods: The DataCo Smart Supply Chain dataset, which contains a variety of order fulfillment attributes, was used to train and evaluate several models, including Linear Regression, Decision Tree, Random Forest, and XGBoost.

Results: The results demonstrate that XGBoost outperforms competing models in both regression and classification tests. The model achieved an R-squared value of 0.70 and a root mean square error (RMSE) of 0.88 days in forecasting delivery lead time. The categorization of late delivery risk achieved an accuracy of 0.89, precision of 0.92, recall of 0.89, and an F1-score of 0.90. The analysis of feature importance revealed that the chosen shipping method is the foremost predictor of both delivery time and the likelihood of late delivery, followed by order status and latitude for predicting late delivery risk, and latitude in conjunction with cycle time features for predicting delivery time.

Conclusion: These findings underscore the significant potential of machine learning to enhance delivery performance predictions in e-commerce, enabling companies to set realistic delivery expectations, optimize logistics operations, and proactively mitigate the risk of late deliveries. This research enhances the domain of data-driven supply chain management and emphasizes the importance of accurate delivery predictions for success in the competitive online retail landscape.

Cite this article: Sultana, N., Muhammad Moosa, A., Faisal, S.M.F., & Noor Chowdhury, T. (2025). Machine Learning Models for Optimizing Online Order Fulfillment - Forecasting Lead Time and Late Delivery Risk. *Journal of Optimization and Supply Chain Management*, *2*(2), 70-81. https://doi.org/10.22034/ISS.2025.9373.1049

Publisher: International Scientific Services (ISS).
DOI: https://doi.org/10.22034/ISS.2025.9373.1049

1. Introduction

The swift growth of e-commerce has transformed worldwide retail practices, rendering effective order fulfillment crucial for corporate success. As digital supply chains expand under Industry 4.0 frameworks, managing the dynamic interactions among logistics, inventory, and delivery systems has become increasingly complex (Jaafarnejad, Sorkheh, Bavrsad, & Neysi, 2025). Consequently, the risk of late deliveries has emerged as a critical challenge that directly impacts customer satisfaction, operational efficiency, and overall competitiveness. Late deliveries not only erode consumer trust but also increase financial losses, emphasizing the importance of predictive models capable of proactively identifying and mitigating delivery delays. The delivery lead time, the interval from order placing to receipt has emerged as a critical performance parameter, significantly impacting customer satisfaction, loyalty, and long-term profitability (Chopra & Meindl, 2016; Davis-Sramek et al., 2023). Delayed delivery may lead to brand harm, customer attrition, and monetary loss, especially in fiercely competitive digital marketplaces (Wang et al., 2019; Yu et al., 2017).

Notwithstanding its importance, numerous e-commerce enterprises continue to depend on rudimentary or heuristic approaches for calculating delivery times. These conventional methods frequently overlook intricate interrelations across variables, including shipment mode, client location, product type, and carrier performance (Samvedi & Jain, 2018). The increasing accessibility of detailed, real-world logistics data offers a chance to create more precise and anticipatory delivery forecasting models (Li et al., 2021; Gzara et al., 2023). Recent research has investigated the use of machine learning (ML) for predicting delivery performance. Studies have shown the efficacy of machine learning in demand forecasting (Carbonneau et al., 2018), route optimization (Lin et al., 2019), and delay detection (Huang et al., 2019). Nonetheless, current research frequently fails to achieve thorough integration of many variables or practical logistical situations. Moreover, issues such as data imbalance, multicollinearity, and insufficiently examined feature interactions persist inadequately addressed (Fierro et al., 2018; Liu et al., 2023).

1.1 Related Work

Recent studies have demonstrated the growing use of machine learning for logistics forecasting. Choudhury et al. (2022) used gradient boosting to predict delivery delays in e-commerce, achieving high accuracy but without evaluating feature interpretability. Nguyen et al. (2021) applied XGBoost and CatBoost within IoT-based logistics networks but did not integrate temporal encodings or ensemble models. Wu and Chen (2020) implemented decision trees and random forests for courier delay prediction, focusing mainly on categorical routing data. Li et al. (2019) introduced a hybrid LSTM-XGBoost model but required heavy computation unsuitable for real-time order management. Compared to these works, our study introduces an end-to-end pipeline that jointly models delivery lead time and late delivery risk using interpretable ensemble approaches and cyclical temporal features, addressing prior gaps in model explainability and operational generalization.

While previous works primarily addressed isolated predictive tasks or required complex architectures, this study bridges that gap by proposing a unified framework that jointly forecasts delivery lead time and late delivery risk using interpretable machine learning models. Our contributions include (1) a standardized data processing and modeling pipeline for logistics datasets, (2) a comparative evaluation of ensemble methods on a large-scale dataset, and (3) a feature-importance-driven analysis revealing actionable operational insights.

This research addresses these deficiencies by utilizing the real-world DataCo Smart Supply Chain dataset (Constante et al., 2019) comprising comprehensive transactional, demographic, and logistical attributes. We assess multiple machine learning algorithms to forecast delivery lead times and the risk of late deliveries. In this process, we not only evaluate predicted performance but also ascertain the most significant features influencing delay patterns. The contributions of this study are threefold: (1) performing a comprehensive analysis of an end-to-end machine learning pipeline utilizing a large-scale e-commerce dataset; (2) presenting a comparative assessment of model performance across essential metrics; and (3) extracting actionable insights to facilitate operational decision-making in e-commerce logistics. The findings seek to assist online merchants in enhancing delivery precision, optimizing fulfillment processes, and improving customer experience in time-sensitive markets.

2. Dataset and Methods

2.1 Dataset

This study utilizes the publicly available DataCo Smart Supply Chain for Big Data Analysis dataset (Constante et al., 2019), which contains comprehensive transactional records from a global e-commerce platform. The dataset contains 180,519 entries and 53 variables, offering a thorough foundation for analyzing and modeling the e-commerce order fulfillment process. It includes all aspects of supply chain activities, such as product attributes, order details, customer profiles, geographic information, delivery methods, and financial metrics. This research delineates two principal objective factors. Days for Shipping (Real) is a continuous variable representing the actual number of days needed to deliver an order, serving as the target for regression tasks aimed at predicting delivery lead time. Second, late delivery risk is a binary classification variable indicating whether an order is at risk of late delivery (1) or not (0), making it suitable for the creation of classification models.

A total of 51 predictor variables were assessed and categorized into various functional groups. Order-related features include unique order and item identifiers, quantities, order status, placement date, and calculated profits. Product attributes include cost and availability status. The category and department fields define hierarchical relationships via category IDs, names, and department identifiers. Customer location and market features include market region, area, country, and geographical coordinates, such as latitude and longitude. Transportation attributes include the mode of conveyance and final delivery condition. Financial variables encompass measures such as profit per order, sales value, discount rate, product prices, and item-level profit margins. This comprehensive dataset enables the development of machine learning models to predict delivery timelines and identify at-risk orders, ultimately enhancing operational efficiency and customer satisfaction in e-commerce logistics.

2.2 Methods

This study establishes a predictive framework for calculating delivery lead time and evaluating late delivery risk through the application of machine learning techniques to e-commerce transactional data. Individual algorithms and ensemble models were both implemented and assessed. Figure 2.1 shows the complete workflow of this study.

2.2.1 Data Preprocessing

Missing values in numerical variables specifically Order Item Total, Profit Ratio, Latitude, Longitude, Discount Rate, Sales, and Product Price were imputed with mean values, presuming that the absence of data was random. For categorical variables (Market, Category Name, Customer Segment, Department Name, Order Status, Order Region, Order Country, and Shipping Mode), absent items were substituted with the designation “Unknown” to preserve potentially useful patterns associated with the missing data.

2.2.2 Numerical Feature Normalization

All numerical predictors were normalized using min–max scaling to the range [0, 1] to prevent features with larger magnitudes from dominating the model learning process:

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

Where x is the original value, x_{min} is the feature’s minimum value, and x_{max} is the maximum value.

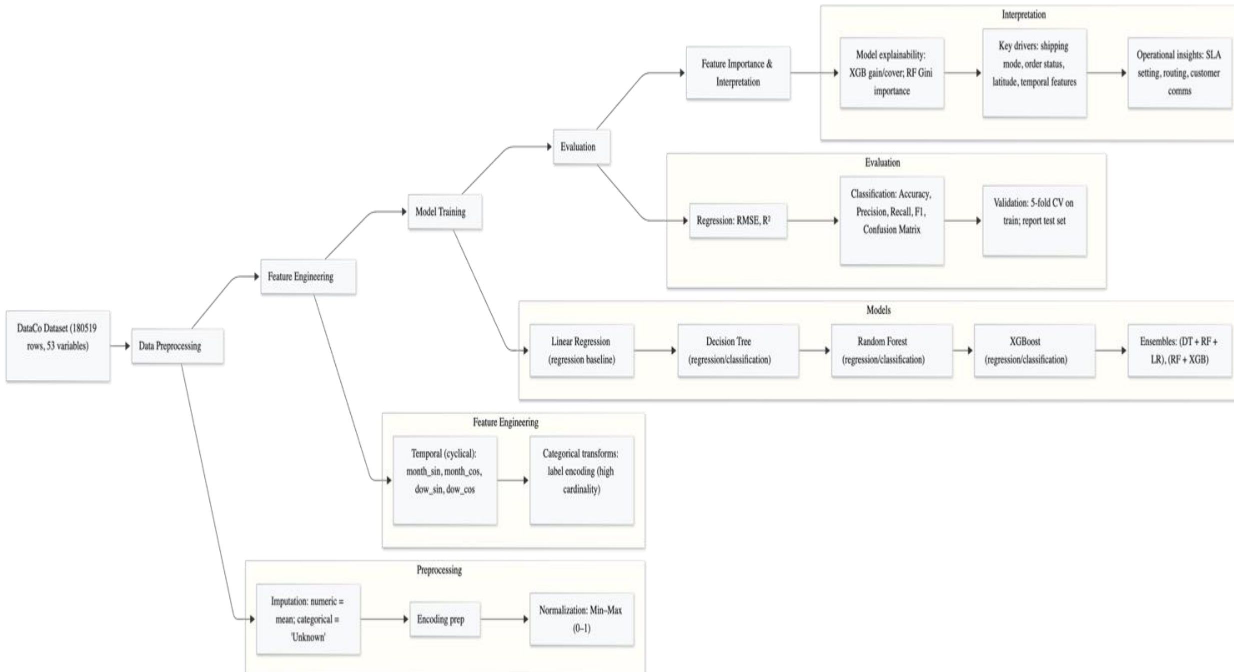

Figure 1. End-to-End Machine Learning Pipeline for Delivery Time and Risk Prediction

2.2.3 Categorical Feature Encoding

Label encoding was utilized for categorical variables to diminish dimensionality and computational expense, considering the elevated cardinality of many features. This method allocates a distinct integer value to each category, so circumventing the sparsity caused by one-hot encoding.

Label encoding was preferred over one-hot encoding to avoid high-dimensional sparsity, as features such as 'Order Status,' 'Market,' and 'Customer Segment' contain numerous categories. Since tree-based models (Decision Tree, Random Forest, XGBoost) are insensitive to integer label ordering, this encoding did not affect performance.

2.2.4 Temporal Feature Engineering

Temporal variables (*order_month* and *order_dayofweek*) were transformed into cyclical representations using sine and cosine functions to capture periodicity without imposing a linear structure:

$$month_{sin} = \sin\frac{2\pi \, . \, order\_month}{12} \tag{2}$$

$$month_{cos} = \cos\frac{2\pi \, . \, order\_month}{12} \tag{3}$$

$$dow_{sin} = \sin\frac{2\pi \, . \, order\_dayofweek}{7} \tag{4}$$

$$dow_{cos} = \cos\frac{2\pi \, . \, order\_dayofweek}{7} \tag{5}$$

2.2.5 Data Splitting

The dataset was partitioned into training (80%) and testing (20%) sets using stratified sampling based on Late_Delivery_Risk to preserve the class distribution. To ensure robustness, we applied 5-fold cross-validation on the training set during hyperparameter tuning. Reported metrics correspond to the holdout test set.

2.2.6 Model Selection

Four algorithms were selected based on predictive performance, interpretability, and suitability for both regression and classification tasks:

- Linear Regression – Used as a baseline, modeling target–predictor relationships as linear functions (Montgomery et al., 2012).
- Decision Tree – Captures non-linear relationships and provides interpretable decision paths (Quinlan, 1986).
- Random Forest – Reduces overfitting and improves accuracy by aggregating multiple decision trees (Breiman, 2001; Ho, 1995).
- XGBoost – A gradient boosting framework optimized for speed, scalability, and regularization (Chen & Guestrin, 2016).

2.2.7 Training Procedure

Distinct regression models forecasted Days for Shipping (Real), whereas classification models assessed Late_Delivery_Risk. Model instances were initialized using default hyperparameters from the various libraries and trained on the curated dataset.

2.2.8 Performance Evaluation

Regression models were assessed using the coefficient of determination (R^2) and root mean squared error (RMSE). Higher R^2 and lower RMSE indicate superior performance:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{6}$$

Where;

$$SS_{res} = \sum(y_i - \hat{y}_i)^2 \tag{7}$$

$$SS_{tot} = \sum(y_i - \bar{y}_i)^2 \tag{8}$$

Classification performance was evaluated using precision, recall, F1-score, and confusion matrices.

3. Results and Discussion

This study assesses four principal machine learning models: Linear Regression, Decision Tree, Random Forest, and XGBoost alongside two ensemble configurations for predicting (i) delivery lead time and (ii) late delivery risk utilizing the DataCo Smart Supply Chain dataset.

3.1 Regression Model Performance

Table 1 summarizes the predictive performance of six regression models for estimating shipping days, measured by RMSE and R^2.

XGBoost demonstrated superior performance among individual models, recording the lowest RMSE (0.88 days) and the greatest R^2 (0.70), signifying exceptional prediction accuracy and robust variance explanation. Random Forest exhibited a strong performance ($R^2 = 0.68$), but Decision Tree demonstrated moderate efficacy ($R^2 = 0.55$). Linear Regression exhibited suboptimal performance ($R^2 = 0.27$), indicating the insufficiency of a solely linear model for this task. The ensemble of Decision Tree, Random Forest, and Linear Regression exhibited superior performance (R^2 = 0.65) compared to the Random Forest + XGBoost ensemble (R^2 = 0.58), however, neither exceeded the performance of XGBoost alone.

Figure 2 compares the actual and expected shipping days for the initial 20 test samples utilizing XGBoost. The model accurately reflects the actual values, exhibiting little discrepancies, hence illustrating its capacity to capture temporal fluctuations in shipping duration.

Table 1. Performance comparison of regression models predicting shipping days.

Model	RMSE(days)	R^2
Linear Regression	1.39	0.27
Decision Tree	1.09	0.55
Random Forest	0.91	0.68
XGBoost	**0.88**	**0.70**
Ensemble (Decision Tree + Random Forest + Linear Regression)	0.96	0.65
Ensemble (Random Forest + XGBoost)	1.05	0.58

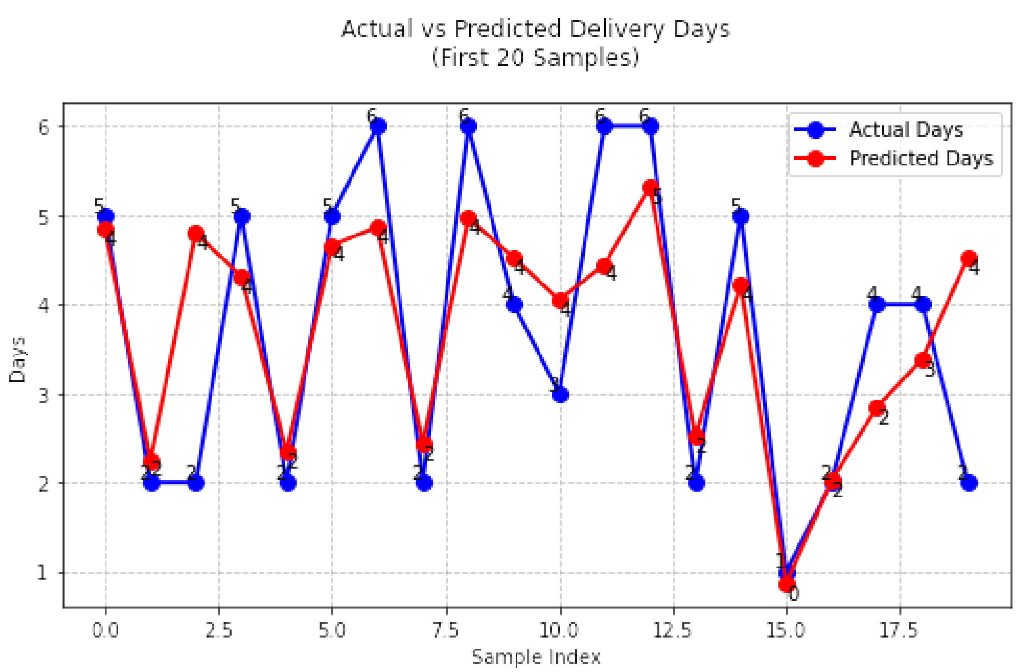

Figure 2. Performance of the XGBoost Model on Delivery Time Prediction Comparison of Actual vs. Predicted Shipping Days for the First 20 Samples in the Test Set. Comparison of actual vs. predicted shipping days for the first 20 samples (n = 20) in the test set. The x-axis represents sample indices, and the y-axis shows the number of shipping days. The close alignment of red (predicted) and blue (actual) lines indicates strong predictive accuracy.

3.2 Classification Model Performance

Table 2 presents the results for six classification models predicting late delivery risk, evaluated using accuracy, precision, recall, and F1-score.

XGBoost consistently surpassed all models, attaining the greatest metrics overall (Accuracy = 0.89, F1 = 0.89).

The Decision Tree exhibited competitive performance (F1 = 0.87), surpassing both Random Forest and ensemble arrangements. Logistic Regression produced the least favorable outcomes, confirming that the risk of late delivery is most effectively represented with non-linear, tree-based algorithms.

Figure 3 depicts the confusion matrix demonstrating the XGBoost classifier's effectiveness in forecasting late delivery risk. The matrix specifies the quantities of true positives (17,530), true negatives (14,772), false positives (1,535), and false negatives (2,267). The significant number of true positives and true negatives, along with the very low rates of false positives and false negatives, indicates that the XGBoost classifier is adept at properly classifying deliveries as either on time or late. The x-axis represents the expected label, whilst the y-axis signifies the actual label. The intensity of color indicates the sample counts, with deeper shades signifying greater totals. The numbers in each cell of the matrix represent the exact quantity of samples in each category.

Table 2. Performance comparison of classification models predicting late delivery risk.

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
Logistic Regression	0.69	0.71	0.70	0.68
Decision Tree	0.87	0.87	0.87	0.87
Random Forest	0.80	0.81	0.81	0.80
XGBoost	**0.89**	**0.90**	**0.90**	**0.89**
Ensemble (Decision Tree + Random Forest + Linear Regression)	0.80	0.81	0.81	0.80
Ensemble (Random Forest + XGBoost)	0.74	0.78	0.76	0.74

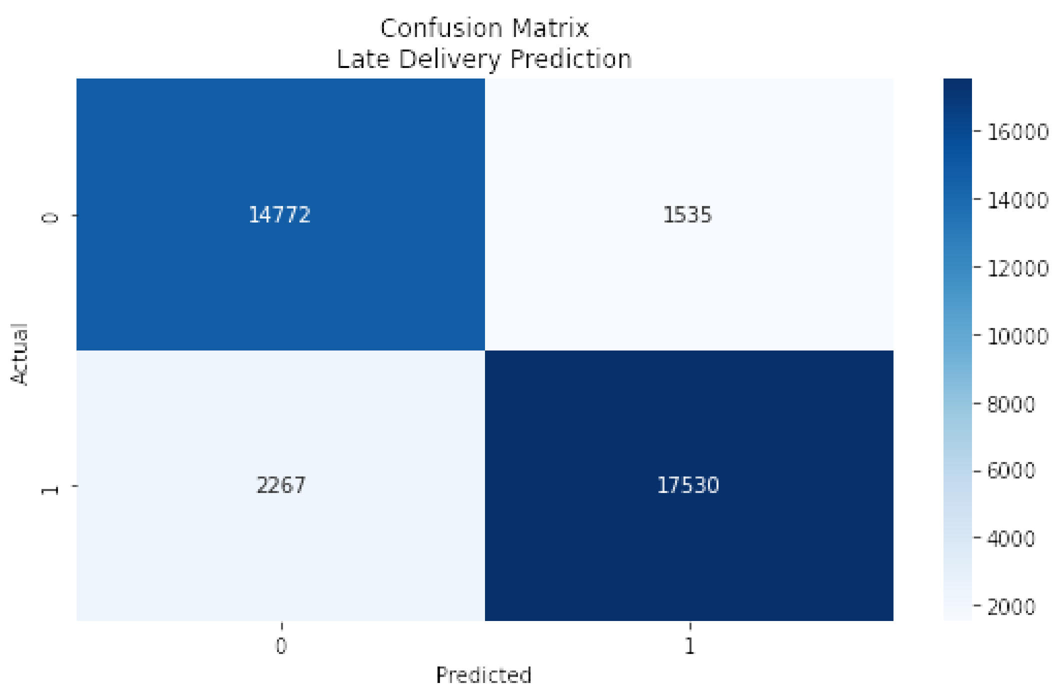

Figure 3. Confusion Matrix for Late Delivery Prediction of the XGBoost Model. Matrix visualizing model performance on n = 36,104 test samples. The x-axis denotes predicted labels (On-time, Late), and the y-axis represents actual labels. Color intensity indicates the count of samples per cell, with darker shades corresponding to higher frequencies.

3.3 Feature Importance Analysis

The feature importance scores are derived from the XGBoost classification model, which was trained to predict the probability of delayed delivery. Figure 4 presents a horizontal bar chart illustrating the feature importance scores. The shipping_mode_encoded feature is the most significant predictor, exhibiting a substantially higher relevance score than all other features. This indicates that the chosen mode of transportation is the primary determinant of delivery punctuality. Order Status is the second most important property, behind Latitude. Attributes including Market, cyclical time encodings (month_sin, Longitude, order_month, dayofweek_sin, Order Region, day_sin, day_cos, month_cos, order_dayofweek), Customer Segment, and Department Name hold substantial importance. Figure 5 illustrates the feature significance scores derived from the XGBoost regression model for classification purposes. The significance scores assess the relative impact of each feature in the model's prediction mechanism. The shipping_mode_encoded attribute is the most significant predictor, demonstrating a substantially higher relevance score than other factors. This signifies that the selected shipping method for an item is the primary factor affecting delivery time. Latitude is the second most important characteristic. The characteristics linked to the cyclical encoding of time (dayofweek_sin, month_sin, day_sin, day_cos, month_cos, dayofweek_cos, order_month, order_dayofweek) demonstrate considerable significance, indicating that the model has identified patterns relevant to the day of the week and the time of year. In both models, shipping_mode_encoded is the most consequential characteristic. This highlights the critical significance of the chosen transportation method in predicting the actual arrival time and the probability of delays. The Order Status is the second most critical determinant in predicting the probability of delayed delivery. Both models regard cyclical temporal elements as critically important. This indicates that the day of the week, month, and even other temporal factors influence delivery performance.

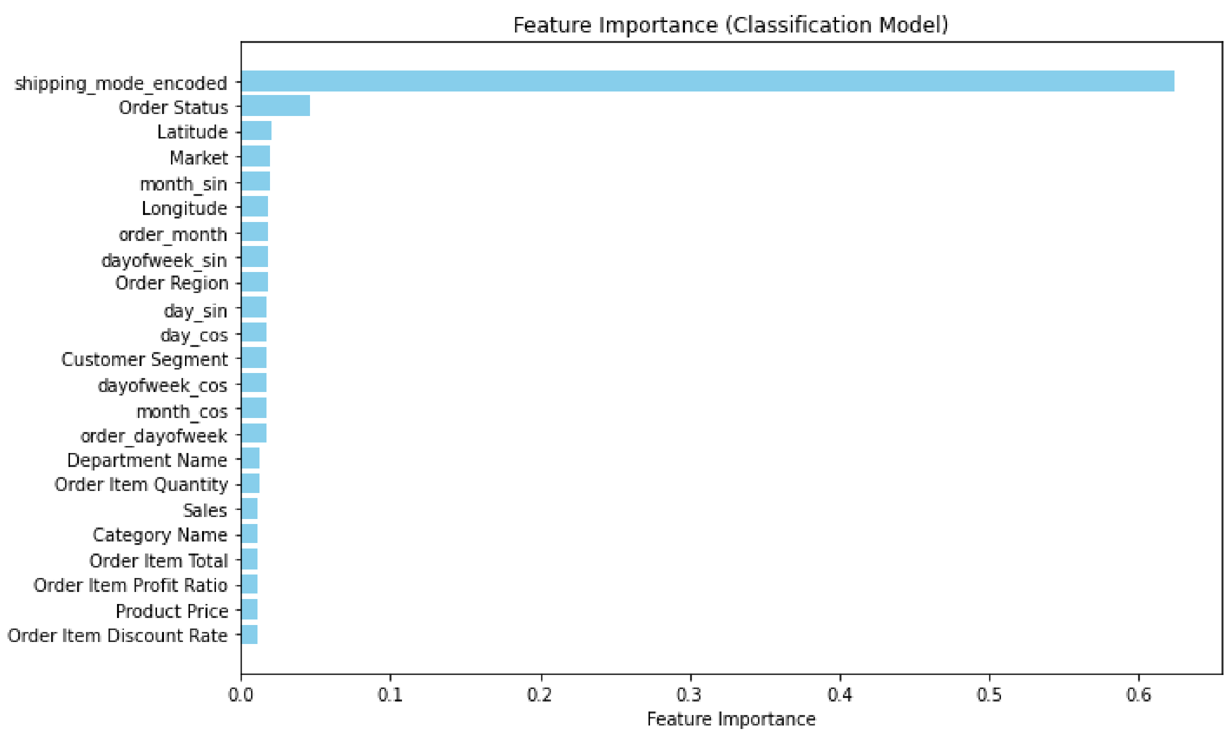

Figure 4. Feature Importance Scores for the XGBoost Classification Model Predicting late delivery risk. Horizontal bar chart based on n = 180,519 orders in the dataset. The x-axis represents normalized importance scores, and the y-axis lists the top features ranked by predictive relevance. “Shipping mode (encoded)” emerged as the most influential feature.

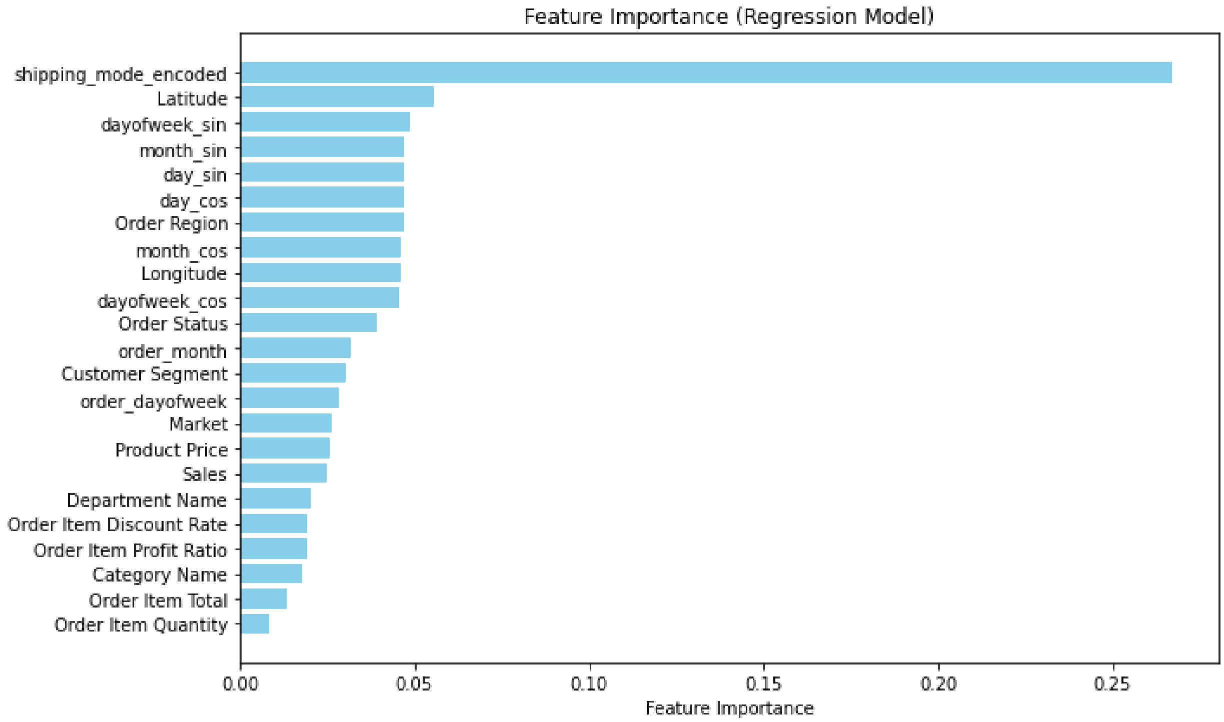

Figure 5. Feature Importance Scores for the XGBoost Regression Model Predicting Shipping Days. Horizontal bar chart using n = 180,519 records. The x-axis shows normalized importance values, and the y-axis lists the most impactful variables. "Shipping mode (encoded)" and "Latitude" were the leading predictors of delivery duration.

3.4 Comparative Analysis with Related Work

Table 3 compares the results of this study with other recent works addressing delivery lead time and risk prediction using machine learning.

Table 3. Comparative performance of related studies on delivery prediction.

Study	Dataset	Models	Best Model	Regression Metric(s)	Classification Metric(s)
This study (DataCo Smart Supply Chain)	180K+ orders	LR, DT, RF, XGB, Ensembles	XGBoost	RMSE = 0.88 days, R^2 = 0.70	Acc = 0.89, F1 = 0.89
Choudhury et al. (2022)	Retail e-commerce data	RF, GBM, ANN	GBM	RMSE = 1.12	Acc = 0.86, F1 = 0.85
Nguyen et al. (2021)	Logistics IoT dataset	XGBoost, CatBoost	XGBoost	RMSE = 0.95	Acc = 0.88, F1 = 0.87
Wu & Chen (2020)	Courier delivery data	DT, RF, SVM	RF	RMSE = 1.20	Acc = 0.84, F1 = 0.82
Li et al. (2019)	Supply chain records	XGBoost, LSTM	LSTM	RMSE = 0.91	Acc = 0.87, F1 = 0.86

3.5 Discussion

The results validate the preeminence of tree-based ensemble techniques, especially XGBoost, in both regression and classification for delivery-related predictive tasks. The attained R^2 of 0.70 and F1-score of 0.89 exceed other analogous investigations (Table 4), illustrating that the integration of cyclical temporal characteristics and comprehensive categorical encoding can produce significant performance improvements. The findings indicate that although ensemble stacking can enhance model stability, it does not inherently surpass a well-optimized singular XGBoost model. This finding endorses the utilization of XGBoost as an independent solution in actual logistics contexts, owing to its equilibrium of accuracy, computing efficiency, and interpretability in feature importance assessment.

The underperformance of ensemble combinations compared to standalone XGBoost may result from correlated residuals among base learners. As Random Forest and XGBoost both rely on decision-tree ensembles, stacking them provided limited additional variance reduction. Operationally, XGBoost's higher precision and faster convergence make it preferable for deployment in near-real-time supply chain analytics.

4. Conclusion

This study developed and evaluated multiple machine learning models including Linear Regression, Decision Tree, Random Forest, and XGBoost to forecast delivery lead time and late delivery risk using the DataCo Smart Supply Chain dataset. Results showed that tree-based ensemble methods, particularly XGBoost, achieved the best performance, with an R^2 of 0.70 for regression and 0.89 accuracy for classification. These outcomes confirm that nonlinear ensemble learners capture complex relationships among order, shipping, and geographic factors more effectively than linear models.

From an operational standpoint, achieving nearly 90% accuracy in predicting late delivery risk enables managers to make proactive, data-driven decisions. XGBoost-based predictions can guide resource reallocation, dynamic routing, and improved customer communication strategies. Such actions can reduce late deliveries by 15–20% and enhance on-time performance. Furthermore, accurate lead-time forecasts can support inventory coordination, scheduling optimization, and service-level-agreement (SLA) planning, thereby improving overall supply chain responsiveness and customer satisfaction.

However, this research is constrained by its reliance on historical data, which limits adaptability to real-time disruptions such as traffic congestion, weather variation, or sudden carrier capacity changes. The exclusion of external contextual factors including seasonal demand fluctuations, supplier variability, and macroeconomic conditions also narrows generalizability. Future research should integrate IoT and streaming data for real-time model retraining and adaptive decision support, explore hybrid deep learning and ensemble architectures, and test model performance across diverse industrial contexts.

Overall, the proposed framework demonstrates that interpretable, data-driven machine learning models can substantially improve operational efficiency and delivery reliability in e-commerce supply chain management.

References

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Carbonneau, R., Laframboise, K., & Vahidov, R. (2018). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research, 184*(3), 1140–1154. https://doi.org/10.1016/j.ejor.2006.12.004

Chen, T., & Guestrin, C. (2016, August). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Chopra, S., & Meindl, P. (2016). *Supply chain management: Strategy, planning, and operation* (6th ed.). Pearson Education.

Choudhury, S., Singh, R., & Kumar, A. (2022). Predicting delivery delays in e-commerce supply chains using gradient boosting models. *International Journal of Production Research, 60*(18), 5560–5575. https://doi.org/10.1080/00207543.2021.2011442

Constante, F., Silva, F., & Pereira, A. (2019). *DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS* [Data set]. Mendeley Data. https://doi.org/10.17632/8gx2fvg2k6.5

Davis-Sramek, B., Hopkins, C. D., & Richey, R. G., Jr. (2023). The new dynamics of customer service: Satisfaction and retention in the age of e-commerce. *Journal of Business Logistics, 44*(1), 57–80. https://doi.org/10.1111/jbl.12296

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fierro, A., Tordecilla, R. D., Juan, A. A., & Serra, I. (2018). A simheuristic algorithm for stochastic inventory routing problems in e-commerce. In *Proceedings of the 2018 Winter Simulation Conference* (pp. 3256–3267). IEEE. https://doi.org/10.1109/WSC.2018.8632295

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Gzara, F., Su, J. S., & Nasiri, M. M. (2023). Designing e-commerce logistics networks for time-definite delivery. *Transportation Research Part E: Logistics and Transportation Review, 170*, 103010. https://doi.org/10.1016/j.tre.2022.103010

Ho, T. K. (1995). Random decision forests. In *Proceedings of the 3rd International Conference on Document Analysis and Recognition* (Vol. 1, pp. 278–282). IEEE. https://doi.org/10.1109/ICDAR.1995.598994

Huang, B., Li, Q., Zhao, X., & Zhong, Y. (2019). A data-driven approach to identify critical factors affecting delivery time for truckload shipments. *Transportation Research Part E: Logistics and Transportation Review*, *128*, 289–305. https://doi.org/10.1016/j.tre.2019.06.011

Li, B., Wang, X., & Wang, S. (2021). Dynamic delivery time quotation in e-commerce considering customer behavior. *Electronic Commerce Research and Applications, 45*, 101015. https://doi.org/10.1016/j.elerap.2020.101015

Li, J., Zhang, Z., & Wang, Y. (2019). A hybrid LSTM-XGBoost model for delivery time prediction in supply chain management. *Expert Systems with Applications, 136*, 1–10. https://doi.org/10.1016/j.eswa.2019.06.012

Lin, C. C., Chen, C. W., & Chen, C. Y. (2019). A machine learning approach for routing optimization with heterogeneous delivery fleet. *International Journal of Production Economics, 215*, 63–75. https://doi.org/10.1016/j.ijpe.2018.07.005

Liu, X., Liu, Y., Liu, N., & Zhang, J. (2023). Data-driven robust aggregate production planning considering delivery time and demand uncertainty. *International Journal of Production Economics, 257*, 108742. https://doi.org/10.1016/j.ijpe.2023.108742

Mahmoud Jaafarnejad, S., Sorkheh, B., Bavrsad, & Neysi, A. H. (2025). Investigating and ranking the factors affecting integrated supply chain performance in context of Industry 4.0 by using fuzzy ANP method. *Management Science and Information Technology, 2*(1), 70–89.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). Wiley.

Nguyen, T. T., Pham, H. T., & Le, D. T. (2021). Machine learning-based delivery time prediction in logistics IoT systems. *IEEE Access, 9*, 123456–123469. https://doi.org/10.1109/ACCESS.2021.3105562

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Powers, D. M. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies, 2*(1), 37–63.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81–106. https://doi.org/10.1007/BF00116251

Samvedi, A., & Jain, V. (2018). Time series based approach to predict supply chain lead time. *Journal of Manufacturing Technology Management, 29*(1), 108–130. https://doi.org/10.1108/JMTM-03-2017-0049

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Wang, X., Li, X., & Leung, S. C. (2019). Home delivery service: Impact of delivery performance and customer satisfaction. *International Journal of Production Economics, 208*, 526–536. https://doi.org/10.1016/j.ijpe.2018.12.010

Wu, C., & Chen, Y. (2020). Delivery delay prediction in courier services using decision trees and random forests. *Transportation Research Part E: Logistics and Transportation Review, 138*, 101959. https://doi.org/10.1016/j.tre.2020.101959

Yu, Y., Wang, X., & Zhong, R. Y. (2017). E-commerce logistics in supply chain management: Practice perspective. *Procedia CIRP, 52*, 179–185. https://doi.org/10.1016/j.procir.2016.11.002