Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 6, No. 3, Juni 2025, Page. 1443-1456
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4689

Comparative Analysis of Supervised Learning Algorithms for Delivery Status Prediction in Big Data Supply Chain Management

Riri Damayanti Apnena*[1], Gerinata Ginting[2], Ari Sudrajat[3], Hussain Md Mehedul Islam[4]

[1]Industrial Mechanics and Design, Politeknik TEDC, Indonesia
[2]Computerized Accounting, Politeknik TEDC, Indonesia
[3]Informatics Engineering, Politeknik TEDC, Indonesia
[4]Software Engineer, The Matlab Inc., United States

Email: [1]riri.apnena@poltektedc.ac.id

Received : May 5, 2025; Revised : Jun 5, 2025; Accepted : Jun 9, 2025; Published : Jun 23, 2025

Abstract

This study addresses the problem of predicting delivery status in supply chain data, a critical task for optimizing logistics and operations. The dataset, which includes multiple features like order details, product specifications, and customer information, was pre-processed using oversampling to address class imbalance, ensuring that the model could handle rare cases of late or canceled deliveries. The data cleaning process involved handling missing values, removing irrelevant columns, and transforming categorical variables into numerical formats. After pre-processing and cleaning, five machine learning models were applied: Logistic Regression, Random Forest, SVM, K-Nearest Neighbors (KNN), and XGBoost. Each model was evaluated using metrics such as accuracy, precision, recall, and F1-score. The results showed that XGBoost outperformed the other models, achieving the highest accuracy and providing the most reliable predictions for the delivery status. This makes XGBoost the best choice for supply chain data analysis in this context. This study contributes to the growing application of machine learning in supply chain optimization by identifying XGBoost as a robust model for delivery status prediction in large datasets. For future research, exploring hybrid models and advanced feature engineering techniques could further improve prediction accuracy and address additional challenges in supply chain optimization, especially in the context of real-time data processing and dynamic supply chain environments.

Keywords : *Delivery Status Prediction, Handling Big Data, Machine Learning Models, Supervised Learning, Supply Chain Analysis, XGBoost.*

1. INTRODUCTION

In modern supply chain management systems, timely and accurate decision making is a major challenge, especially when faced with large and complex data volumes [1], [2]. Many variables in supply chain data such as product type, delivery time, logistics costs, and vendor performance are interrelated and affect overall operational efficiency [3], [4]. On the other hand, the complexity of the relationship between these variables makes manual analysis less effective in identifying important patterns that can be used as a basis for decision making [4], [5], [6]. Therefore, a machine learning-based approach is needed that is able to process data efficiently and produce accurate predictions [7], [8], [9]. Comparison between supervised learning models is important to determine the most optimal method in supporting strategic decisions in supply chain management [10].

Based on the problem analysis, there is an urgency in decision making in a complex supply chain system [2], [3], [11]. To answer this need, a machine learning-based approach can be used as an adaptive and data-oriented solution. With its ability to recognize complex patterns and make predictions automatically, machine learning can help optimize processes such as demand forecasting, vendor

Vol. 6, No. 3, Juni 2025, Page. 1443-1456
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4689

performance assessments, and delivery time estimates [12], [13], [14]. The model training process is carried out by utilizing historical supply chain data that has been cleaned and engineered for features, resulting in a predictive model that is able to support operational and strategic decision making [15], [16]. This approach not only improves decision accuracy but also enables higher efficiency in resource and logistics management.

To support this research, several related studies have proposed the use of similar models. However, there are several limitations identified in these studies, which raises the urgency for the submission of this research to address these issues. Such as Kang et al. (2025) [17], where the authors propose the use of supervised machine learning to assess supplier performance and risk profiles in supply chain management. Several classification models are used to analyze and predict supplier performance based on historical data. The models used can improve supplier selection efficiency and provide more accurate recommendations to reduce risks in the supply chain. This approach successfully optimizes supply chain performance by improving supplier selection consistency. The study has limitations in the quality of the data used. The high reliance on clean and complete historical data makes the prediction results susceptible to data bias or data deficiencies. In addition, this model only considers certain supplier features, so it may miss other important variables that affect overall performance.

Sani et al. (2023) [18] proposes the use of Light Gradient Boosting Machine (LightGBM) optimized with Bayesian approach to predict risks in supply chain, especially backorder risk. The Bayesian optimized model shows very high prediction accuracy in predicting backorder risk, with good computational efficiency. This approach can provide deeper insight into potential disruptions in supply chain. Although this model shows good results, Bayesian optimization requires large computational resources and long training time, especially on large datasets. In addition, this model is limited to backorder risk modeling, so it does not consider other external factors that may affect risk management in supply chain.

Kiran et al. (2025) [4] proposes the use of various machine learning algorithms such as Long Short-Term Memory (LSTM), Support Vector Machine (SVM), Genetic Algorithm (GA), and Reinforcement Learning (RL) to improve decision making in supply chain management. LSTM achieves 94.3% accuracy, SVM achieves 87.8%, while GA and RL improve delivery efficiency and resource optimization in the supply chain system. A hybrid model combining these techniques provides better results in real-time decision making and resource management. Combining different algorithmic techniques can cause problems in the consistency of results and managing complex models. In addition, LSTM requires very large and precise time-series data to achieve optimal performance, which is sometimes difficult to achieve in dynamic real-world applications.

The main contribution of this study lies in the comparative application of supervised learning models in decision making in supply chain management using big data, which distinguishes it from previous studies that used smaller datasets. Although previous studies have successfully applied supervised learning models to large data, they have not utilized the full potential of big data that can cover larger volumes, variations, and speeds of information. This study proposes a more scalable and efficient approach to processing and analyzing big data to support strategic decisions in the supply chain, which allows the model to identify more complex patterns and provide more accurate predictions in a dynamic and evolving context. By utilizing big data, this study aims to overcome the challenges of previous models and provide more adaptive and relevant solutions in global supply chain management.

2. METHOD

This research begins with collecting raw data, which is then processed through data processing to ensure it is ready for use. Next, data preparation and data cleaning are performed to over-sampling and inconsistencies. The cleaned data is then initialized with labels and attributes to make it recognizable by

Vol. 6, No. 3, Juni 2025, Page. 1443-1456
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4689

machine learning models. The following stage involves initializing models using various algorithms such as Logistic Regression, Random Forest Classifier, Support Vector Machine, K-Nearest Neighbors, and Xtreme Gradient Boosting. The data is split into three parts: 80% for training, 10% for validation, and 10% for testing. The trained models are then evaluated using a confusion matrix to measure their performance, followed by a comparison of the best models to determine the most effective algorithm for prediction or classification tasks.

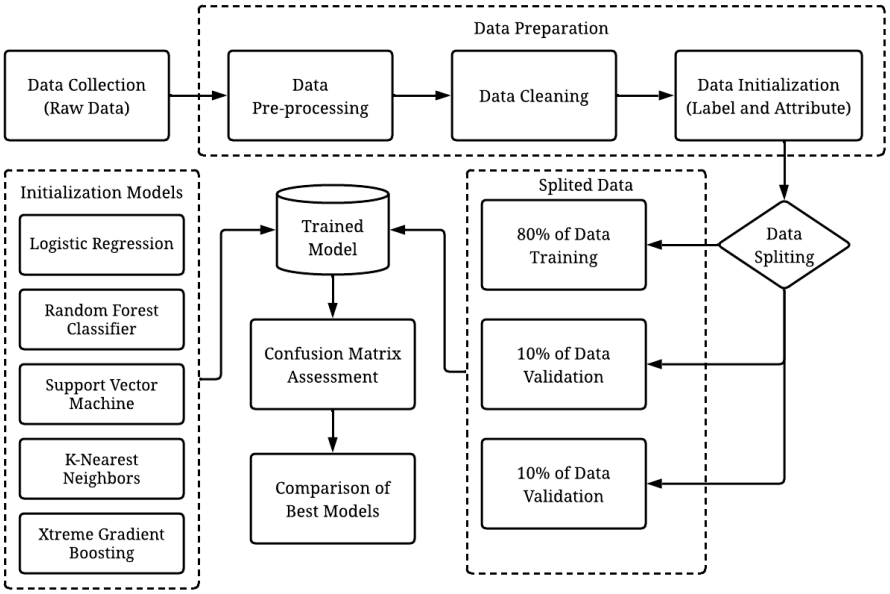

Figure 1. Proposed Scheme

2.1. Pre-processing (Oversampling)

The target variable in the dataset demonstrates a clear class imbalance. The Late delivery category contains 98,977 samples, while Shipping canceled has only 7,754 samples. This imbalance can negatively impact the performance of supervised learning models, as they tend to favor majority classes during training, leading to poor generalization for minority classes.

Table 1. Pre-processed Class Distribution

Class Label	Raw Count	Processed Count	Sampling Status
Late delivery	98,977	98,977	Stable
Advance shipping	41,592	98,977	Oversampled
Shipping on time	32,196	98,977	Oversampled
Shipping canceled	7,754	98,977	Oversampled

As seen in table 1, this study applies an oversampling technique. Oversampling increases the number of samples in the minority classes to match the size of the majority class. This is done without discarding valuable data from the majority class, which is particularly important in this context, where the largest class (Late delivery) contains useful patterns for accurate prediction.

Among various oversampling methods, this study adopts a random oversampling approach by duplicating existing samples from underrepresented classes such as Shipping canceled and Shipping on time [19]. This helps ensure that the model is exposed to a balanced representation of all target categories during training, ultimately improving its ability to learn from all classes and make more reliable and fair predictions in the supply chain decision-making process [20].

Vol. 6, No. 3, Juni 2025, Page. 1443-1456
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4689

2.2. Data Cleaning

The data cleaning process aims to ensure the quality, consistency, and usability of the dataset before applying machine learning models [21]. Based on the referenced source code, several key cleaning procedures were applied systematically. The dataset was initially inspected for missing values using statistical functions and visualized through heatmaps to identify sparsity. Columns with a high percentage of missing data or those deemed non-contributive were removed entirely to reduce dimensionality and enhance overall data quality. Meanwhile, rows with minor missing entries were either imputed or removed based on their relevance and impact on the target variable.

To eliminate redundancy and irrelevant information, non-informative columns such as 'ID', 'Product Status', and other identifiers with no predictive value were dropped. This step was guided by correlation analysis and domain-specific knowledge of supply chain operations. Furthermore, categorical attributes including 'Warehouse_block', 'Mode_of_Shipment', and 'Product_importance' were converted into numerical formats using encoding techniques such as Label Encoding or One-Hot Encoding to ensure compatibility with machine learning algorithms.

Numerical features like 'Cost_of_the_Product' and 'Discount_offered' were assessed for outliers using statistical visualizations such as boxplots. Extreme values that could potentially distort the model's learning capability were either capped or excluded. Temporal variables such as 'Order Date' and 'Shipping Date' were cleaned and transformed into derived metrics like shipping duration by computing the time difference between dates. Any inconsistencies or unrealistic timestamps were identified and corrected during this process. Lastly, all data types were standardized—for instance, converting string values to datetime objects and floats to integers—to ensure consistency and compatibility throughout the modeling pipeline. An overview of the dataset before and after the cleaning process as seen in Table 2, highlighting the structural improvements resulting from these preprocessing steps. This cleaning process is crucial because it directly impacts the model's performance. Clean and relevant data helps the model learn patterns more accurately and reduces the risk of overfitting or mispredictions due to biased or malformed data.

Table 2. Dataset Overview: Before and After Data Cleaning

Aspect	Before Cleaning	After Cleaning
Number of Features	53	35
Missing Values	Present in several columns	Handled or removed
Redundant Columns	Included (e.g., 'ID', 'Product Status')	Dropped
Categorical Features	Raw text labels	Encoded numerically
Date Columns Format	String/Unprocessed	Parsed and transformed
Outliers in Numeric Columns	Detected in 'Cost_of_the_Product', 'Discount_offered'	Removed or capped
Data Type Inconsistencies	Mixed types	Standardized

2.3. Supervised Learning Model Based on Machine Learning Algorithm

In this study, multiple machine learning algorithms are applied to predict the Delivery Status based on various order and product-related features. The models considered include:

2.3.1. Logistic Regression

Logistic Regression is a statistical method commonly used for binary classification tasks [22]. In this context, the model is applied to predict the probability of a shipment being late or on time. The model calculates the probability of an event occurring based on a linear combination of the features.

Vol. 6, No. 3, Juni 2025, Page. 1443-1456
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4689

Despite its relative simplicity, Logistic Regression is used as a baseline model due to its efficiency and ability to provide clear interpretations, especially when the relationship between features and targets is nearly linear.

2.3.2. Random Forest Classifier

Random Forest is an ensemble algorithm based on decision trees [23], [24]. It combines multiple decision trees to improve classification accuracy and reduce overfitting. Each tree in the forest is trained on a random subset of the data, and the final prediction is made based on majority voting. Random Forest is particularly useful in this context because of its ability to handle large datasets and capture complex non-linear relationships between features and target variables.

2.3.3. Support Vector Machine

Support Vector Machine (SVM) is a powerful classification algorithm that finds the optimal hyperplane to separate different classes in a feature space [25], [26], [27]. SVM is very effective when the data is not linearly separable by using a kernel function to transform the data into a higher-dimensional space. In this study, SVM is applied to classify Delivery Status and is expected to provide good performance on both linearly and non-linearly separable data.

2.3.4. K-Nearest Neighbors

K-Nearest Neighbors (KNN) is a non-parametric algorithm that classifies data based on its proximity to other data in a feature space [28]. Prediction is done by considering the majority class among the k-nearest neighbors of the test point. KNN is simple, intuitive, and often used for classification tasks where the decision boundary is highly non-linear. However, its performance can degrade on high-dimensional data or large datasets due to its reliance on distance calculations.

2.3.5. Extreme Gradient Boosting

XGBoost (Extreme Gradient Boosting) is a highly efficient and scalable gradient boosting algorithm for classification tasks [29], [30]. It builds an ensemble of decision trees sequentially, with each tree trying to correct the errors made by the previous tree. XGBoost has become popular due to its high accuracy, speed, and ability to handle missing values and large datasets. It is very effective in predicting outcomes based on complex non-linear patterns and is expected to perform best on these datasets.

2.4. Confusion Matrix

Confusion matrix is an important tool in evaluating the performance of classification models, allowing to analyze how well the model predicts different classes [31], [32], [33]. It shows the distribution of correct and incorrect predictions in the relevant categories, providing a clearer picture of the types of errors made by the model [34], [35]. In the context of this study, confusion matrix is used to evaluate the performance of different machine learning models applied to predict Shipment Status.

Accuracy is the most commonly used evaluation metric in classification models. Accuracy measures the extent to which a model can predict the correct label from all the data tested. This metric is calculated by comparing the number of correct predictions (for both positive and negative classes) to the total number of data tested. The accuracy assessment can be seen in equation (1).

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Data} \quad (1)$$

Precision is a metric that measures the extent to which a model's positive predictions are actually relevant. It shows how many of all the positive predictions made by the model are actually correct. High

Vol. 6, No. 3, Juni 2025, Page. 1443-1456
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4689

precision indicates that the model makes few errors in predicting the positive class. This is very useful when we want to minimize the number of errors in predicting the positive class (e.g., a delivery delay that is incorrectly predicted on time). The precision assessment can be seen in equation (2).

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2)$$

Recall, or Sensitivity, is a metric that measures the extent to which a model can recognize all positive class examples in the data. It indicates how many of all positive class examples the model successfully recognized. High recall indicates that the model was able to find most of the positive class examples, but does not guarantee that the positive predictions were correct (it is more about sensitivity to minority classes). The recall assessment can be seen in equation (3).

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (3)$$

F1-Score is a metric that combines precision and recall into one number, which is useful when we need a balance between the two. F1-Score provides a more complete picture of the model's performance, especially on imbalanced datasets, where the model may be very good in one metric (such as recall) but bad in another metric (such as precision). A higher F1-Score indicates that the model has a good balance between precision and recall, and it is often used as a primary metric in imbalanced classification problems. The f1-score assessment can be seen in equation (3).

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision \times Recall} \quad (4)$$

3. RESULT

In this chapter, we will discuss the results of the analysis and evaluation carried out after applying the machine learning method to the processed dataset. The results obtained include a description of the cleaned data, analysis of the model used, and model performance based on evaluation using various relevant metrics. All experiments were conducted to evaluate the model's ability to predict the desired outcome, in this case the prediction of the shipping status and profit of each order. Before going into the model analysis, the first part of this chapter will review the description of the dataset that has been prepared after the data cleaning process, which includes various features, values contained in the dataset, and explanations related to each feature used in the analysis.

3.1. Selection Importance Feature

Of the total 35 features that have gone through the data cleaning and transformation process, not all features are used directly in training the prediction model. As a step for efficiency and accuracy improvement, a feature selection process was carried out using a statistical test approach (F-value and P-value) to determine the attributes that have the most influence on the target variable, namely Delivery Status. Several features that have proven significant in influencing delivery status prediction include: Order Id, Order Item Discount, Order Item Cardprod Id, Shipping Date, Order Date, Order Customer Id, Order Profit Per Order, Market, Order Region, Order State, Order Item Total, Department Name, Product Card Id, Customer Id, Product Category Id, Product Image, Category Name, Product Name, Product Price, Sales per Customer, Benefit per Order, Order Zipcode, Order Item Id, Order City, and Customer Segment. These features were selected because they have high F-Value values and significant P-Value (≤ 0.05), indicating that their presence statistically contributes to variations in the prediction target. Further information regarding the F-Value and P-Value of each important feature can be seen in Table 3.

Vol. 6, No. 3, Juni 2025, Page. 1443-1456
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4689

Table 3. Selected Importance Feature

Feature	F-Value	P-Value	Range (Sample)	Explanation
Order Id	1165.17	0.000	Unique ID (e.g., 5961)	Unique identifier for each order in the system.
Order Item Discount	57166.13	0.000	0 – 120000	Discount applied to the item in the order.
Order Item Cardprod Id	12782.97	0.000	Numeric code	ID of the product from the product catalog.
Shipping Date	142.65	0.000	Date	The date the product was shipped.
Order Date	128.46	0.000	Date	The date the order was placed.
Order Customer Id	673.46	0.000	Customer ID (e.g., 69)	Identifier for the customer who placed the order.
Order Profit Per Order	13782.67	0.000	0 – 20000+	Profit earned per individual order.
Market	240.91	0.000	Market code (e.g., 3)	Geographic market where the product is sold.
Order Region	140.52	0.000	Region code (e.g., 8)	Geographical region of order distribution.
Order State	27.94	0.000	State code (e.g., 11)	State or province where the order was shipped.
Order Item Total	481682.35	0.000	0 – 500000+	Total cost of all items in the order.
Department Name	524.09	0.000	Department code	Department associated with the product.
Product Card Id	12782.97	0.000	Product ID	Unique identifier of each product.
Customer Id	673.46	0.000	Unique customer ID	Identifier for a customer in the system.
Product Category Id	10095.78	0.000	Product category ID	Classification of the product category.
Product Image	37751.72	0.000	Image code/file name	Visual representation of the product.
Category Name	26066.33	0.000	Product category name	Specific category assigned to a product.
Product Name	37751.72	0.000	Product name	Name or label of the purchased item.
Product Price	116680.12	0.000	Unit price	The price per product unit.
Sales per Customer	481682.35	0.000	Customer sales total	Total amount spent by a single customer.
Benefit per Order	13782.67	0.000	Profit margin	Average profit gained per order transaction.
Order Zipcode	4.20	0.040	Zip/postal code	Geographical code for delivery location.
Order Item Id	1133.74	0.000	Item ID within order	Unique code for each item in an order.

Vol. 6, No. 3, Juni 2025, Page. 1443-1456
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4689

Order City	8.76	0.003	City name	City where the product is shipped.
Customer Segment	4.27	0.039	0–3 (segment types)	Segment classification such as consumer, corporate, or small business.

From the selected features presented in Table 3, a correlation matrix can be derived to better understand the relationships among these influential variables, as illustrated in Figure 2.

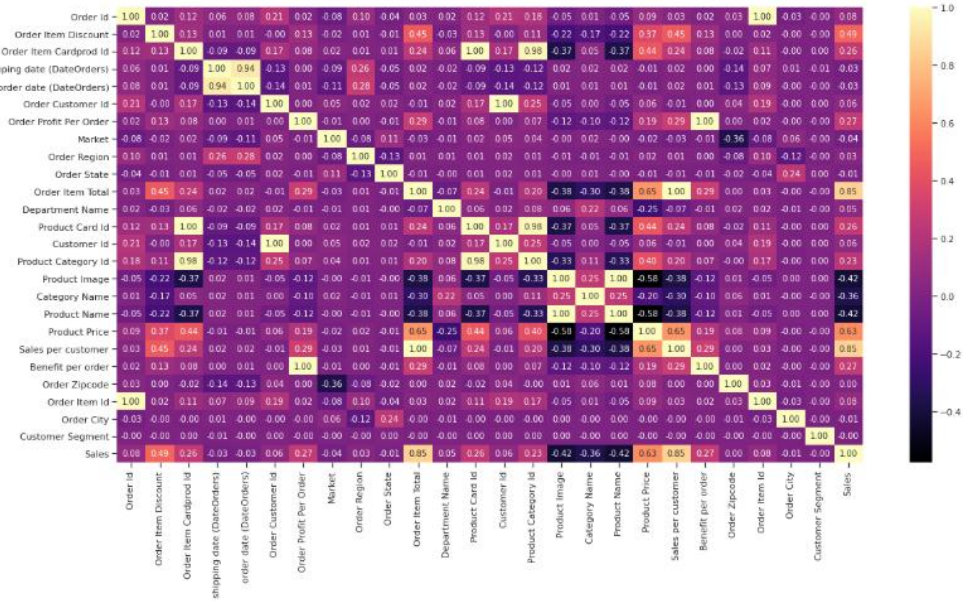

Figure 2. Correlation Analysis

3.2. Model Assessment

The evaluation of model performance was conducted using four key classification metrics: accuracy, precision, recall, and F1-score. These metrics were applied consistently across all models to ensure objective and comparable assessment results. The outcomes of each model's performance are presented in Table 4, highlighting the effectiveness of each algorithm in predicting delivery status across the dataset.

Table 4. Model Assessment

Model Classifier	Without Oversampling				With Oversampling			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.6907	0.70	0.68	0.69	0.7853	0.80	0.78	0.79
Random Forest	0.8927	0.90	0.89	0.895	0.9995	0.99	1.00	0.995
SVM	0.6719	0.68	0.67	0.675	0.7853	0.79	0.78	0.785
K-Nearest Neighbors	0.7423	0.74	0.75	0.745	0.7806	0.78	0.79	0.785
XGBoost	0.8907	0.89	0.89	0.89	0.9997	0.99	1.00	1.00

The evaluation of each classifier, as presented in Table 4, was performed using a consistent set of training parameters. These parameters, which were carefully selected for each model, are outlined in

Vol. 6, No. 3, Juni 2025, Page. 1443-1456
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4689

Table 5. The models were trained on the dataset using these specified parameters, and their performance was measured based on accuracy, precision, recall, and F1-score. To provide a clear visual comparison of the model performance across different metrics, a graphical representation of accuracy, precision, recall, and F1-score for each classifier can be seen in Figure 3. This figure highlights the performance of each model under both standard and oversampling conditions, allowing for an easy comparison of how each model fares in predicting the delivery status categories.

Table 5. Model Parameter

Model	Parameter
Logistic Regression	penalty='l2', solver='lbfgs', max_iter=1000
Random Forest	n_estimators=100, max_depth=None, min_samples_split=2, min_samples_leaf=1, random_state=42
SVM	C=1.0, kernel='rbf', gamma='scale', decision_function_shape='ovr', random_state=42
K-Nearest Neighbors (KNN)	n_neighbors=5, weights='uniform', algorithm='auto', metric='minkowski'
XGBoost	n_estimators=100, learning_rate=0.1, max_depth=6, subsample=0.8, colsample_bytree=0.8, random_state=42

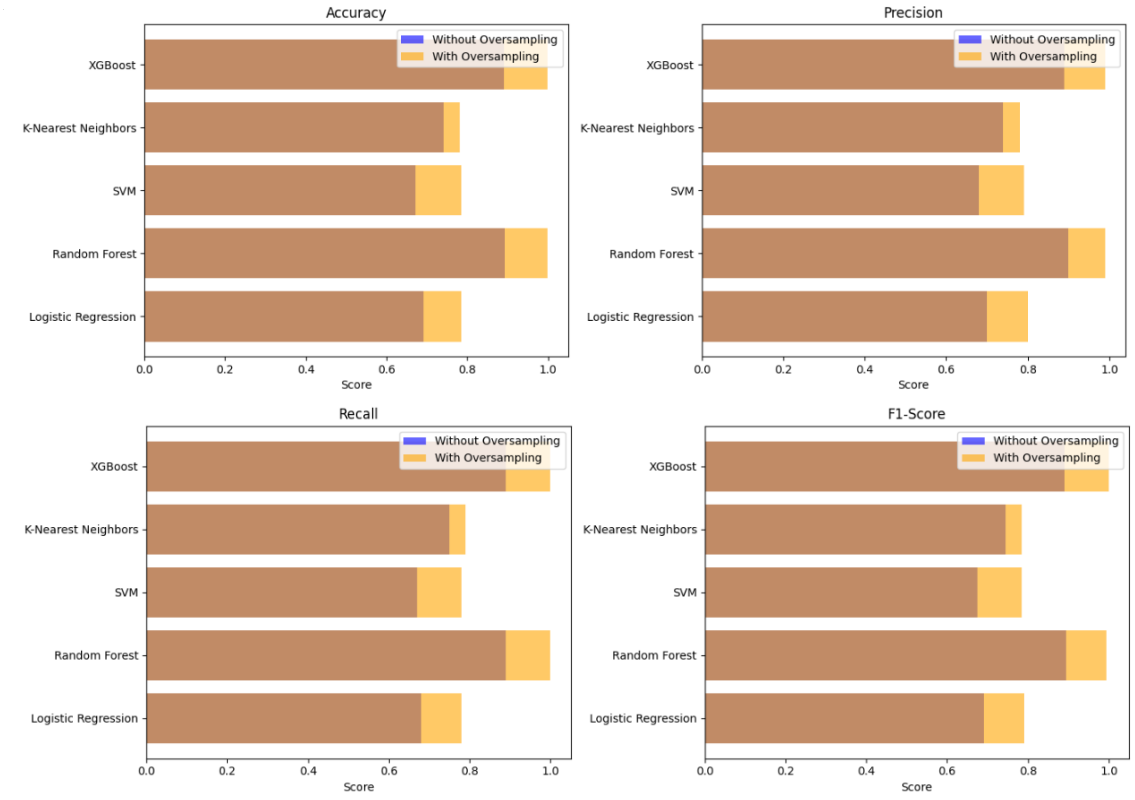

Figure 3. Performance Model Comparison

4. DISCUSSIONS

In this study, five machine learning models were developed and evaluated to predict delivery status in the supply chain dataset. The models built included Logistic Regression, Random Forest Classifier, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and XGBoost. These models were selected due to their popularity and proven effectiveness in classification tasks. The models were

Vol. 6, No. 3, Juni 2025, Page. 1443-1456
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4689

assessed based on multiple metrics, such as accuracy, precision, recall, and F1-score, with the goal of determining the best-performing model for predicting delivery status.

Among the models evaluated, XGBoost emerged as the best-performing model for this task, achieving an accuracy of 99.97% with oversampling. XGBoost showed exceptional performance across all metrics, including precision, recall, and F1-score, making it a reliable choice for predicting delivery status in supply chain applications. Its ability to handle complex relationships and its robustness to overfitting contributed to its top ranking. The model's effectiveness can be attributed to its ensemble approach, where multiple decision trees are trained to make predictions, thus improving generalization.

The second-best model was Random Forest, with an accuracy of 99.95% under oversampling conditions. Random Forest demonstrated a strong performance with a balance of high precision and recall, offering an alternative to XGBoost. While slightly trailing behind XGBoost in terms of accuracy, XGBoost proved to be a powerful tool for handling imbalanced datasets and complex feature interactions, making it suitable for supply chain prediction tasks. Following these, KNN and SVM were also evaluated, with their performance ranking lower, though they still provided valuable insights into the impact of different machine learning approaches on supply chain delivery prediction.

These findings align with previous studies that have explored the use of machine learning in supply chain optimization. Kang et al. [17] used supervised learning to evaluate supplier performance and risks, but their approach was constrained by data quality and focused only on limited features. Our study extends this by utilizing a broader feature set and applying models to multiclass delivery outcomes. Sani et al. [18] demonstrated the strength of LightGBM with Bayesian optimization for backorder risk prediction, highlighting the effectiveness of tree-based models; however, their scope was narrow and computationally intensive. In contrast, our use of XGBoost achieved similar high performance while maintaining model efficiency and generalizability. Meanwhile, Kiran et al. [4] investigated various algorithms including LSTM and hybrid methods for supply chain decision making. Although they achieved good results in specific scenarios, their reliance on time-series data and complex model integration can limit applicability in diverse, real-world supply chain environments. Our research contributes by offering a scalable solution using ensemble methods with proven robustness in large, non-time-series datasets.

5. CONCLUSION

This study focused on predicting the delivery status in a supply chain dataset. The dataset was pre-processed and cleaned to ensure high-quality, reliable data for modeling. The data cleaning process involved handling missing values, encoding categorical features, and removing irrelevant columns, which resulted in a dataset with 38 features and a target label, "Delivery Status," that includes categories such as "Late delivery," "Advance shipping," "Shipping on time," and "Shipping canceled." After preprocessing, five machine learning models were built and evaluated: Logistic Regression, Random Forest Classifier, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and XGBoost. Among the five models, XGBoost achieved the best performance, with an accuracy of 99.97% when oversampling was applied. XGBoost followed closely, offering a balanced performance with high precision and recall. These models were evaluated based on key metrics such as accuracy, precision, recall, and F1-score, with Random Forest showing the most consistent and reliable results. The findings highlight the effectiveness of ensemble models in handling complex, imbalanced datasets commonly found in supply chain operations.

For future research, further exploration into feature engineering and tuning hyperparameters could improve model performance even more. Additionally, experimenting with other machine learning techniques, such as deep learning or hybrid models, could offer insights into enhancing predictive accuracy.

Vol. 6, No. 3, Juni 2025, Page. 1443-1456
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4689

REFERENCES

[1] N. Zhao, J. Hong, and K. H. Lau, "Impact of supply chain digitalization on supply chain resilience and performance: A multi-mediation model," Int J Prod Econ, vol. 259, p. 108817, May 2023, doi: 10.1016/j.ijpe.2023.108817.

[2] M. K. Lim, Y. Li, C. Wang, and M.-L. Tseng, "A literature review of blockchain technology applications in supply chains: A comprehensive analysis of themes, methodologies and industries," Comput Ind Eng, vol. 154, p. 107133, Apr. 2021, doi: 10.1016/j.cie.2021.107133.

[3] B. Rolf, I. Jackson, M. Müller, S. Lang, T. Reggelin, and D. Ivanov, "A review on reinforcement learning algorithms and applications in supply chain management," Int J Prod Res, vol. 61, no. 20, pp. 7151–7179, Oct. 2023, doi: 10.1080/00207543.2022.2140221.

[4] Kiran Kumar Reddy Penubaka, "Optimizing Decision-Making in Supply Chain Management Using Machine Learning and Mathematical Modeling Techniques," Journal of Information Systems Engineering and Management, vol. 10, no. 11s, pp. 574–586, Feb. 2025, doi: 10.52783/jisem.v10i11s.1654.

[5] Y. Popova and I. Sproge, "Decision-making within smart city: Waste sorting," Sustainability (Switzerland), vol. 13, no. 19, Oct. 2021, doi: 10.3390/su131910586.

[6] S. Setyani, I. Abu Hanifah, and I. I. Ismawati, "The Role of Budget Decision Making as A Mediation of Accounting Information Systems and Organizational Culture on The Performance of Government Agencies," Journal of Applied Business, Taxation and Economics Research, vol. 1, no. 3, pp. 311–324, Feb. 2022, doi: 10.54408/jabter.v1i3.59.

[7] N. R. D. Cahyo, C. A. Sari, E. H. Rachmawanto, C. Jatmoko, R. R. A. Al-Jawry, and M. A. Alkhafaji, "A Comparison of Multi Class Support Vector Machine vs Deep Convolutional Neural Network for Brain Tumor Classification," in 2023 International Seminar on Application for Technology of Information and Communication (iSemantic), IEEE, Sep. 2023, pp. 358–363. doi: 10.1109/iSemantic59612.2023.10295336.

[8] M. M. I. Al-Ghiffary, C. A. Sari, E. H. Rachmawanto, N. M. Yacoob, N. R. D. Cahyo, and R. R. Ali, "Milkfish Freshness Classification Using Convolutional Neural Networks Based on Resnet50 Architecture," Advance Sustainable Science Engineering and Technology, vol. 5, no. 3, p. 0230304, Oct. 2023, doi: 10.26877/asset.v5i3.17017.

[9] F. Farhan, C. A. Sari, E. H. Rachmawanto, and N. R. D. Cahyo, "Mangrove Tree Species Classification Based on Leaf, Stem, and Seed Characteristics Using Convolutional Neural Networks with K-Folds Cross Validation Optimalization," Advance Sustainable Science Engineering and Technology, vol. 5, no. 3, p. 02303011, Oct. 2023, doi: 10.26877/asset.v5i3.17188.

[10] A. Wieland, "Dancing the Supply Chain: Toward Transformative Supply Chain Management," Journal of Supply Chain Management, vol. 57, no. 1, pp. 58–73, Jan. 2021, doi: 10.1111/jscm.12248.

[11] L. Bednarski, S. Roscoe, C. Blome, and M. C. Schleper, "Geopolitical disruptions in global supply chains: a state-of-the-art literature review," Production Planning & Control, pp. 1–27, Dec. 2023, doi: 10.1080/09537287.2023.2286283.

[12] C. Irawan, E. H. Rachmawanto, and H. P. Hadi, "An Ensemble Learning Layer for Wayang Recognition using CNN-based ResNet-50 and LSTM," Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 2025.

[13] E. H. Rachmawanto, C. A. Sari, and F. O. Isinkaye, "A good result of brain tumor classification based on simple convolutional neural network architecture," Telkomnika (Telecommunication Computing Electronics and Control), vol. 22, no. 3, pp. 711–719, Jun. 2024, doi: 10.12928/TELKOMNIKA.v22i3.25863.

[14] R. R. Ali et al., "Learning Architecture for Brain Tumor Classification Based on Deep Convolutional Neural Network: Classic and ResNet50," Diagnostics, vol. 15, no. 5, p. 624, Mar. 2025, doi: 10.3390/diagnostics15050624.

[15] M. M. I. Al-Ghiffary, N. R. D. Cahyo, E. H. Rachmawanto, C. Irawan, and N. Hendriyanto, "Adaptive deep learning based on FaceNet convolutional neural network for facial expression recognition," Journal of Soft Computing, vol. 05, no. 03, pp. 271–280, 2024, doi: https://doi.org/10.52465/joscex.v5i3.450.

Vol. 6, No. 3, Juni 2025, Page. 1443-1456
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4689

[16] N. R. D. Cahyo and M. M. I. Al-Ghiffary, "An Image Processing Study: Image Enhancement, Image Segmentation, and Image Classification using Milkfish Freshness Images," IJECAR) International Journal of Engineering Computing Advanced Research, vol. 1, no. 1, pp. 11–22, 2024.
[17] P. S. Kang and B. Bhawna, "Enhancing supply chain resilience through supervised machine learning: supplier performance analysis and risk profiling for a multi-class classification problem," Business Process Management Journal, Jan. 2025, doi: 10.1108/BPMJ-03-2024-0174.
[18] S. Sani, H. Xia, J. Milisavljevic-Syed, and K. Salonitis, "Supply Chain 4.0: A Machine Learning-Based Bayesian-Optimized LightGBM Model for Predicting Supply Chain Risk," Machines, vol. 11, no. 9, p. 888, Sep. 2023, doi: 10.3390/machines11090888.
[19] A. J. Albert, R. Murugan, and T. Sripriya, "Diagnosis of heart disease using oversampling methods and decision tree classifier in cardiology," Research on Biomedical Engineering, vol. 39, no. 1, pp. 99–113, Dec. 2022, doi: 10.1007/s42600-022-00253-9.
[20] U. Hasanah, A. M. Soleh, and K. Sadik, "Effect of Random Under sampling, Oversampling, and SMOTE on the Performance of Cardiovascular Disease Prediction Models," Jurnal Matematika, Statistika dan Komputasi, vol. 21, no. 1, pp. 88–102, Sep. 2024, doi: 10.20956/j.v21i1.35552.
[21] P. Li, X. Rao, J. Blase, Y. Zhang, X. Chu, and C. Zhang, "CleanML: A Study for Evaluating the Impact of Data Cleaning on ML Classification Tasks," in 2021 IEEE 37th International Conference on Data Engineering (ICDE), IEEE, Apr. 2021, pp. 13–24. doi: 10.1109/ICDE51399.2021.00009.
[22] A. D. Amirruddin, F. M. Muharam, M. H. Ismail, N. P. Tan, and M. F. Ismail, "Synthetic Minority Over-sampling TEchnique (SMOTE) and Logistic Model Tree (LMT)-Adaptive Boosting algorithms for classifying imbalanced datasets of nutrient and chlorophyll sufficiency levels of oil palm (Elaeis guineensis) using spectroradiometers and unmanned aerial vehicles," Comput Electron Agric, vol. 193, p. 106646, Feb. 2022, doi: 10.1016/j.compag.2021.106646.
[23] M. A. Rasyidi, T. Bariyah, Y. I. Riskajaya, and A. D. Septyani, "Classification of handwritten javanese script using random forest algorithm," Bulletin of Electrical Engineering and Informatics, vol. 10, no. 3, pp. 1308–1315, Jun. 2021, doi: 10.11591/eei.v10i3.3036.
[24] E. H. Rachmawanto, D. R. I. M. Setiadi, N. Rijati, A. Susanto, I. U. W. Mulyono, and H. Rahmalan, "Attribute Selection Analysis for the Random Forest Classification in Unbalanced Diabetes Dataset," in 2021 International Seminar on Application for Technology of Information and Communication (iSemantic), 2021, pp. 82–86. doi: 10.1109/iSemantic52711.2021.9573181.
[25] S. S. Kavitha and N. Kaulgud, "Quantum machine learning for support vector machine classification," Evol Intell, 2022, doi: 10.1007/s12065-022-00756-5.
[26] N. Bayati et al., "Locating high-impedance faults in DC microgrid clusters using support vector machines," Appl Energy, vol. 308, Feb. 2022, doi: 10.1016/j.apenergy.2021.118338.
[27] C. Umam, L. B. Handoko, and F. O. Isinkaye, "Performance Analysis of Support Vector Classification and Random Forest in Phishing Email Classification," Scientific Journal of Informatics, vol. 11, no. 2, pp. 367–374, May 2024, doi: 10.15294/sji.v11i2.3301.
[28] R. Widadi, B. Arifwidodo, K. Masykuroh, and A. Saputra, "Klasifikasi Tingkat Kesegaran Ikan Nila Menggunakan K-Nearest Neighbor Berdasarkan Fitur Statistis Piksel Citra Mata Ikan," JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 7, no. 1, p. 242, Jan. 2023, doi: 10.30865/mib.v7i1.5196.
[29] A. Farzipour, R. Elmi, and H. Nasiri, "Detection of Monkeypox Cases Based on Symptoms Using XGBoost and Shapley Additive Explanations Methods," Diagnostics, vol. 13, no. 14, p. 2391, Jul. 2023, doi: 10.3390/diagnostics13142391.
[30] J. Ou et al., "Coupling UAV Hyperspectral and LiDAR Data for Mangrove Classification Using XGBoost in China's Pinglu Canal Estuary," Forests, vol. 14, no. 9, p. 1838, Sep. 2023, doi: 10.3390/f14091838.
[31] S. C. Kim and Y. S. Cho, "Predictive System Implementation to Improve the Accuracy of Urine Self-Diagnosis with Smartphones: Application of a Confusion Matrix-Based Learning Model through RGB Semiquantitative Analysis," Sensors, vol. 22, no. 14, Jul. 2022, doi: 10.3390/s22145445.
[32] I. Markoulidakis and G. Markoulidakis, "Probabilistic Confusion Matrix: A Novel Method for

Vol. 6, No. 3, Juni 2025, Page. 1443-1456
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4689

Machine Learning Algorithm Generalized Performance Analysis," Technologies (Basel), vol. 12, no. 7, p. 113, Jul. 2024, doi: 10.3390/technologies12070113.

[33] I. Markoulidakis, I. Rallis, I. Georgoulas, G. Kopsiaftis, A. Doulamis, and N. Doulamis, "Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem," Technologies (Basel), vol. 9, no. 4, Dec. 2021, doi: 10.3390/technologies9040081.

[34] I. P. Kamila, C. A. Sari, E. H. Rachmawanto, and N. R. D. Cahyo, "A Good Evaluation Based on Confusion Matrix for Lung Diseases Classification using Convolutional Neural Networks," Advance Sustainable Science, Engineering and Technology, vol. 6, no. 1, p. 0240102, Dec. 2023, doi: 10.26877/asset.v6i1.17330.

[35] N. E. W. Nugroho and A. Harjoko, "Transliteration of Hiragana and Katakana Handwritten Characters Using CNN-SVM," IJCCS (Indonesian Journal of Computing and Cybernetics Systems), vol. 15, no. 3, p. 221, Jul. 2021, doi: 10.22146/ijccs.66062.

Vol. 6, No. 3, Juni 2025, Page. 1443-1456
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4689