VERİMLİLİK DERGİSİ
JOURNAL OF PRODUCTIVITY

Evaluation of Machine Learning and Ensemble Learning Models for Classification Using Delivery Data

İrem Karakaya[1]

ABSTRACT

Purpose: This study aims to evaluate the performance of various machine learning and ensemble learning models in classifying delivery times using Amazon delivery data. Fast deliveries' role in providing a competitive advantage and boosting customer loyalty highlights the importance of this study.

***Methodology*:** The research employs a dataset of 43,739 delivery records with 15 features. Data preprocessing steps include handling missing values, encoding categorical variables, calculating geospatial distances, and normalizing data. Advanced machine learning techniques (e.g., KNN, SVM, Logistic Regression) and ensemble methods (e.g., ExtraTrees, AdaBoost) were systematically compared based on accuracy, precision, recall, and F-score.

Findings: Ensemble learning models, particularly those using SVM, NB, and LDA as base models and ET as the meta model, achieved the highest accuracy (99.89%) and F-score (99.89%). These results underscore the potential of such models to optimize logistics operations, reduce delays, and enhance customer satisfaction.

Originality: This study demonstrates the effectiveness of machine and ensemble learning methods on complex logistics data, contributing to optimizing logistics efficiency and enhancing customer satisfaction. Additionally, the application of ensemble learning methods on complex and large-scale logistics data structures is unique in terms of its contribution to the literature. The proposed framework offers a scalable solution for real-time predictive modeling and logistics optimization.

Keywords*:* Machine Learning, Ensemble Learning, Logistics Optimization, E-Commerce Logistics.
JEL Codes: C45, L81, L91.

Teslimat Verileri Kullanılarak Makine Öğrenimi ve Topluluk Öğrenme Modelleri ile Sınıflandırma Performansının Değerlendirilmesi

ÖZET

Amaç: Bu çalışma, Amazon teslimat verilerini kullanarak çeşitli makine öğrenimi ve topluluk öğrenme modellerinin teslimat sürelerini sınıflandırma performansını değerlendirmeyi amaçlamaktadır. Hızlı teslimatların rekabet avantajı sağlamadaki ve müşteri sadakatini artırmadaki rolü, bu çalışmanın önemini vurgulamaktadır.

***Yöntem*:** Araştırmada, 15 özelliğe sahip 43.739 teslimat kaydından oluşan bir veri seti kullanılmaktadır. Veri ön işleme adımları, eksik değerlerin işlenmesi, kategorik değişkenlerin kodlanması, coğrafi mesafelerin hesaplanması ve verilerin normalleştirilmesini içermektedir. Gelişmiş makine öğrenimi teknikleri (örneğin, KNN, SVM, Lojistik Regresyon) ve topluluk yöntemleri (örneğin, ExtraTrees, AdaBoost), doğruluk, hassasiyet, geri çağırma ve F-skoru gibi metrikler temel alınarak sistematik bir şekilde karşılaştırılmıştır.

Bulgular: Topluluk öğrenme modelleri, özellikle temel model olarak SVM, NB ve LDA ile üst model olarak ET kullanıldığında en yüksek doğruluk (%99.89) ve F-skoru (%99.89) değerlerine ulaşmıştır. Bu sonuçlar, bu tür modellerin lojistik operasyonlarını optimize etme, gecikmeleri azaltma ve müşteri memnuniyetini artırma potansiyelini vurgulamaktadır.

Özgünlük: Bu çalışma, makine ve topluluk öğrenme yöntemlerinin karmaşık lojistik verilerdeki etkinliğini göstererek, lojistik verimliliğin optimize edilmesine ve müşteri memnuniyetinin artırılmasına katkı sağlamaktadır. Ayrıca, karmaşık ve geniş ölçekli lojistik veri yapıları üzerinde topluluk öğrenme yöntemlerinin uygulanmasının literatüre yaptığı katkı açısından benzersizdir. Önerilen çerçeve, gerçek zamanlı tahmin modelleme ve lojistik optimizasyonu için ölçeklenebilir bir çözüm sunmaktadır.

Anahtar Kelimeler*:* Makine Öğrenimi, Topluluk Öğrenme, Lojistik Optimizasyonu, E-Ticaret Lojistiği.
JEL Kodları: C45, L81, L91.

[1]Bartın Üniversitesi, Bartın Meslek Yüksekokulu, Yönetim ve Organizasyon Bölümü, Bartın, Türkiye

Corresponding Author: İrem Karakaya, isahmutoglu@bartin.edu.tr
DOI: 10.51551/verimlilik.1526436
Research Article | Submitted: 01.08.2024 | Accepted: 07.01.2025
Cite: Karakaya, İ. (2025). “Evaluation of Machine Learning and Ensemble Learning Models for Classification Using Delivery Data”, *Verimlilik Dergisi,* Productivity for Logistics (SI), xxx-xxx.

1. INTRODUCTION

In today's rapidly digitalizing world, the e-commerce sector plays a pivotal role in transforming retail sales and necessitating the optimization of logistics processes. With the growth of the e-commerce industry, accurately and swiftly predicting delivery times has become critically important for customer satisfaction and business efficiency. The complexity of logistics operations and increasing customer demands require the analysis of large datasets that traditional methods cannot manage (Tsang et al., 2021). Major e-commerce platforms place significant emphasis on optimizing the accuracy and efficiency of delivery times to enhance customer satisfaction and gain a competitive edge. Machine learning algorithms are crucial in making logistics processes more efficient due to their ability to make effective predictions on large and complex datasets. In this context, machine learning algorithms offer revolutionary solutions in logistics and supply chain management through their capabilities to process large datasets and model complex relationships (Tsolaki et al., 2023).

Accurately predicting delivery times in e-commerce logistics poses various challenges. These challenges arise from the complexity of logistics operations, which involve numerous variables such as traffic conditions, weather disruptions, warehouse processing times, and last-mile delivery inefficiencies (Eskandaripour and Boldsaikhan, 2023). Last-mile delivery, often regarded as the most resource-intensive segment, is further complicated by urban congestion and unpredictable customer availability. These challenges highlight the need for robust predictive models capable of handling large-scale, heterogeneous datasets. Traditional forecasting methods struggle to account for these dynamic and often unpredictable factors, leading to delays and inaccuracies in delivery estimations. Inaccurate predictions result in customer dissatisfaction, increased operational costs, and a loss of competitive advantage in the highly competitive e-commerce environment. Therefore, finding reliable solutions to improve the accuracy of delivery time predictions is critical for enhancing customer satisfaction, optimizing resources, and ensuring the overall efficiency of logistics processes.

In this study, a series of machine learning models were tested to classify delivery times using Amazon delivery data. The obtained results were compared to determine which model provided the most effective performance in this field. At the end of the study, a comprehensive evaluation of the effectiveness of the analysis methods and selected machine learning models is presented. This aims to identify the most suitable and effective machine learning model for optimizing delivery times in the retail sector. Additionally, this research is intended to make significant contributions to the utilization of datasets and the optimization of machine learning models in the fields of retail product logistics and last-mile transportation. The findings underscore the potential for advanced data-driven approaches to transform logistics efficiency and customer satisfaction.

Despite the existence of various studies in the current literature on predicting delivery times in e-commerce logistics, previous studies have predominantly been conducted using smaller datasets and limited model combinations. For instance, studies by Kazan and Karakoca (2019) and Khiari and Olaverri-Monreal (2020) used relatively small datasets with limited model diversity. This study addresses these limitations by leveraging a large-scale dataset from Amazon, incorporating diverse features, and applying a comprehensive range of machine learning and ensemble learning models. Research employing large-scale datasets and a wide range of algorithms is relatively scarce. This study seeks to fill this gap by conducting a comprehensive analysis of large-scale Amazon delivery data through the application of machine learning and ensemble learning methods. The objective is to address the limitations of traditional models in accurately predicting delivery times, particularly when faced with the complexities of large-scale, dynamic logistics data. By improving delivery time predictions, e-commerce platforms can gain a competitive edge and better meet growing customer demands. Key research questions include:

1) How can machine learning and ensemble learning models improve the accuracy of delivery time predictions?
2) Which model performs best across various logistics scenarios?
3) Which is the most effective model when the performance of methods is systematically compared with the aim of improving the efficiency of logistics processes?

In this study, the methodologies for big data analysis were meticulously selected in alignment with established successful applications in the literature. As highlighted by Bruni et al. (2023) and Salari et al. (2022), essential data preprocessing steps, including imputing missing values, transforming categorical variables, and feature extraction, were systematically implemented. For the deployment of machine learning models, widely adopted algorithms such as KNN, SVM, Naive Bayes, and Random Forest were utilized. Furthermore, ensemble learning methods were employed through the use of bagging, boosting, and stacking techniques, following the recommendations of prior Research (Karakaya et al., 2022). The effectiveness of these models was evaluated based on key performance metrics such as accuracy, precision, and F-score to ensure the robustness of the predictions. These methodologies are intended to enhance the efficiency of logistics processes.

1.1. Contributions

This study demonstrates the effectiveness of machine learning and ensemble learning models in classifying delivery times using real-world data from Amazon, addressing a critical gap in the literature where large-scale logistics datasets have been underexplored. By leveraging advanced preprocessing techniques and conducting a comparative analysis of 12 machine learning and ensemble learning models, the study provides a robust framework for improving delivery time predictions. Key contributions include:

Practical Applications: The study highlights how ensemble learning methods, specifically combining SVM, NB, and LDA with ET, can significantly enhance efficiency, reduce costs, and improve customer satisfaction in e-commerce logistics.

Methodological Insights: The research showcases advanced feature engineering techniques, such as timestamp transformations and geospatial distance calculations, to optimize data for machine learning applications.

Comprehensive Evaluation: The findings offer a detailed comparative analysis of multiple models, establishing ensemble learning methods as superior for complex logistics data.

Generalizability: While focusing on Amazon delivery data, the study's methodology and insights are applicable to other logistics and supply chain scenarios, including last-mile delivery, inventory management, and network optimization.

This research sets a foundation for the broader adoption of machine learning and ensemble learning models in logistics, providing actionable insights for both academic and industrial applications. By addressing the challenges associated with complex, large-scale logistics data, the study contributes to enhancing operational efficiency and meeting the growing demands of e-commerce.

1.2. Organization

This study is organized as follows: Section 1 provides the motivation, purpose, contrubutions, and significance of the study. Section 2 summarizes previous research on machine learning and ensemble learning in logistics. Section 3 describes the dataset, preprocessing steps, and details of the machine learning and ensemble learning models used. Section 4 presents the performance metrics and results of the models, including accuracy, precision, recall, and F-score. Section 5 analyzes the findings, compares model performances, and discusses the implications for logistics optimization. Finally, Section 6 summarizes the study's contributions, results, and potential future work.

2. LITERATURE REWIEV

Leveraging machine learning and ensemble learning methods to support areas such as delivery times and customer satisfaction holds significant importance in modern logistics and retail sectors. Studies in these fields provide data-driven approaches to enhance operational efficiency, improve customer satisfaction, and optimize logistics processes. This section presents summaries of literature, showcasing the key findings and results of various studies using different datasets and methods. These summaries offer valuable insights into current trends and successful applications in the literature.

Kazan and Karakoca (2019) used to classify categories with machine learning algorithms, product information from an e-commerce website was analyzed. Two different feature extraction techniques, TF-IDF and CountVectorizer, were compared during the data preprocessing phase, and six categories were classified using various classifiers (Random Forest, Decision Tree, Naive Bayes, Logistic Regression, SVM, ANN). According to the results, the SVM and MLP algorithms showed the highest performance with an accuracy rate of 97%. Yüce and Kabak (2021) applied machine learning algorithms to estimate production time in work centers related to four different processes in a manufacturing facility. The comparison of artificial neural networks, support vector regression, and gradient boosting algorithms revealed that the gradient boosting model achieved the highest success rate. The results demonstrated that selecting the right algorithm for production time estimation provides significant advantages in terms of cost and time.

Alnahhal et al. (2021) investigate the dynamic prediction of whether customer orders will arrive in the next delivery week using machine learning. Real data from December 2014 to August 2016 was used. Predictions using methods such as moving averages, simple linear regression, and logistic regression achieved an accuracy rate of 93%. The results are utilized to reduce waiting times at the consolidation center and lower transportation costs. (Erkmen et al., 2022) utilized the Support Vector Machine model with sequential and periodic look-back approaches to predict delivery times. Analysis with data obtained from Kaggle showed that using look-back approaches reduced prediction error by 59.12%. Khiari and Olaverri-Monreal (2020) applied various boosting algorithms to predict delivery times using seven months of data from a postal service company in Austria. Algorithms like Light Gradient Boosting and CatBoost outperformed other methods with high accuracy and efficiency. These approaches increased operational efficiency by ensuring accurate prediction of delivery times.

Lochbrunner and Witschel (2022) developed three different models combining machine learning models with human knowledge to predict delivery times using shipping data from a large retailer. Using the XGBoost regression algorithm and SHAP explanatory package, it was found that pure machine learning models performed better than human-machine combinations. However, both approaches had specific weaknesses and areas for improvement. (Rokoss et al., 2024) analyzed the data of two German manufacturing companies, using machine learning approaches to predict delivery times in small batch production companies. Predictions with machine learning models such as XGBoost could accurately predict delivery times early, effectively reducing manual efforts. These approaches provided significant results in enhancing the efficiency of the production process. Salari et al. (2022) applied tree-based models like quantile regression forests to predict delivery times and manage customer promises in online retail using JD.com data. The proposed methods increased prediction accuracy by over 40% compared to existing methods and boosted sales volume by 3.7% to 6.1%. This approach significantly improved customer satisfaction and operational efficiency.

Bruni et al. (2023) developed a machine learning-based optimization approach for last-mile delivery and third-party logistics services and tested it with real data from Italy. The proposed method provided high performance in a shorter time than existing heuristic methods and effectively optimized logistics processes. Chu et al. (2023) developed a data-driven approach combining machine learning and capacity-constrained vehicle routing optimization to improve the last-mile delivery performance of online food delivery platforms. Analyses using multi-source real data showed that the proposed method performed approximately 5% better than other methods. Sheng Liu (2021) developed a framework integrating travel time predictors with order assignment optimization to improve last-mile delivery performance using two months of data from a food delivery service provider in China. Analyses using machine learning and robust optimization tools enhanced the accuracy and efficiency of order assignment decisions. This method significantly improved the timely performance of last-mile delivery services. Gore et al. (2023) optimized digital marketing strategies in the food delivery business using ensemble learning methods supported by various algorithms. Decision trees, nearest neighbors, and Naive Bayes algorithms, along with ensemble learning methods such as Random Forest, Gradient Boosting, and XGBoost, were used. The results showed that these methods significantly improved the accuracy of marketing strategies, enhancing customer satisfaction.

Deshmukh et al. (2024) used data analysis and machine learning techniques to increase the delivery efficiency of electronic products on e-commerce platforms. He conducted an analysis with over 100,000 transaction data from January 2019 to December 2019, proposing route optimization and increasing logistics capacities during peak periods. Zaghloul et al. (2024) compared machine learning and deep learning methods to predict customer satisfaction in online retail using over 100,000 order data from a major retailer. The Random Forest model showed the best performance with a 92% accuracy rate and identified delivery time and order accuracy as the most influential factors on customer satisfaction. The summary of methods and key findings of the literature reviews is presented in Table 1.

In the literature, reviews on predicting delivery times in e-commerce logistics rely on small datasets. For example, Kazan and Karakoca (2019) and Khiari and Olaverri-Monreal (2020) used limited timeframes and small datasets in their analyses. While studies like those by Salari et al. (2022) and Deshmukh et al. (2024) have focused on larger datasets, such research remains in the minority. This study addresses these gaps by conducting a comprehensive analysis using large-scale Amazon delivery data. Unlike prior research, this approach incorporates broader timeframes and diverse logistics scenarios, aiming to contribute to more accurate predictions of delivery times and advancing the current state of e-commerce logistics research.

Table 1. Details of reviews

Paper	*Method*	*Key Findings*	*Dataset*
Kazan and Karakoca (2019)	Machine learning classification with TF-IDF & CountVectorizer	SVM and MLP achieved 97% accuracy	E-commerce product data
(Yüce and Kabak (2021)	ANN, SVM, Gradient Boosting	Gradient Boosting had the highest success rate	Manufacturing facility data
Alnahhal et al. (2021)	Dynamic prediction using machine learning	Logistic regression achieved 93% accuracy	Data from Dec 2014 to Aug 2016
Erkmen et al. (2022)	SVM with look-back approaches	Look-back approaches reduced error by 59.12%	Kaggle delivery data
Khiari and Olaverri-Monreal (2020)	Boosting algorithms for delivery time prediction	CatBoost outperformed others with high accuracy	7 months of postal service data
Lochbrunner and Witschel (2022)	XGBoost regression, SHAP explanation	ML models outperformed human-machine combos	Retailer shipping data
Rokoss et al. (2024)	XGBoost for small batch production	XGBoost accurately predicted delivery times	Data from 2 manufacturing companies
Salari et al. (2022)	Tree-based models for delivery time prediction	Prediction accuracy improved by over 40%	JD.com delivery data
Bruni et al. (2023)	ML-based optimization for last-mile delivery	High performance in a shorter time	Real data from Italy
Chu et al. (2023)	Data-driven vehicle routing optimization	Improved last-mile delivery by 5%	Multi-source real data
Sheng Liu (2021)	ML for travel time and order assignment	Enhanced order assignment accuracy	2 months of food delivery data
Gore et al. (2023)	Ensemble learning for marketing strategies	Improved marketing strategy accuracy	Digital marketing data
Deshmukh et al. (2024)	ML for delivery efficiency of electronic products	Increased delivery efficiency during peak periods	100,000 transaction data

3. PROPOSED MODELS

This section encompasses the dataset characteristics, the suggested method and model, the experimental setup, the obtained results and discussion.

3.1. Dataset

The dataset is provided as a comma-separated values (CSV) file containing Amazon's delivery data. This dataset offers a detailed overview of the company's last-mile logistics operations. Each row consists of delivery data from various cities, including information on order details, delivery agents, weather and traffic conditions, and delivery performance metrics, as outlined in Table 2. The dataset enables researchers and analysts to explore the factors influencing delivery efficiency, identify optimization opportunities, and examine the impact of various variables on the overall customer experience.

The dataset is utilized to classify delivery times, optimize delivery efficiency, and identify potential improvement areas aimed at enhancing customer satisfaction through the application of machine learning and ensemble learning methods. By employing machine learning algorithms and ensemble learning techniques, the relationships among various features within the dataset will be analyzed, and models with the highest accuracy rates will be selected to evaluate the performance of logistics operations. This approach contributes to the development of data-driven strategies to improve delivery efficiency and customer experience.

Machine learning and ensemble learning models are applied to the preprocessed dataset. Initially, the features related to each order delivery are considered. Subsequently, the processes of handling missing data and eliminating insignificant data with very few instances are carried out. Following this, feature extraction and the application of machine learning and ensemble learning models are performed. Feature extraction plays a critical role in transforming raw attributes into meaningful predictors for machine learning models. For instance, geospatial data (latitude and longitude of store and drop locations) is processed using the Haversine formula to calculate the actual distance between points. This transformation provides a single numerical feature representing the delivery distance, which directly correlates with delivery time and significantly improves model performance. Similarly, temporal attributes, such as 'Order_Date' and 'Order_Time,' are converted into timestamp values to facilitate numerical processing. Additionally, the time difference between 'Order_Time' and 'Pickup_Time' is calculated to represent delays in warehouse operations. These extracted features enhance the predictive capacity of the models, enabling more accurate delivery time classifications. Subsequently, various machine learning and ensemble learning methods are systematically applied to identify the most effective model for improving the efficiency of logistics processes. The data processing workflow for machine and ensemble learning is illustrated in Figure 1.

Table 2. Details of dataset

Feature Name	*Description*	*Value Range*
Order_ID	Unique identifier assigned to each order.	Unique values
Agent_Age	Age of the delivery agent.	Numerical values (15-50)
Agent_Rating	Rating assigned to the delivery agent based on performance.	Numerical values (0-5)
Store_Latitude	Latitude coordinate of the store location.	Numerical values (-30.90 - 30.91 degrees)
Store_Longitude	Longitude coordinate of the store location.	Numerical values (-88.37 - 88.43 degrees)
Drop_Latitude	Latitude coordinate of the delivery destination.	Numerical values (-30.90 - 30.91 degrees)
Drop_Longitude	Longitude coordinate of the delivery destination.	Numerical values (-88.37 - 88.43 degrees)
Order_Date	Date when the order was placed.	Date values (2022-02-11 - 2024-07-03)
Order_Time	Time when the order was placed.	Time values (00:00:00 – 23:55:00)
Pickup_Time	Time when the order was picked up from the store.	Time values (00:00:00 – 23:55:00)
Weather	Weather conditions during the delivery period.	Categorical values (e.g., clear, rainy, stormy)
Traffic	Traffic conditions encountered during the delivery.	Categorical values (e.g., light, moderate, heavy)
Vehicle	Type of vehicle used for the delivery.	Categorical values (e.g., bike, car, van)
Area	Geographic area or zone where the delivery took place.	Categorical values (e.g., metropolitan, urban, semi-urban, other)
Delivery Time	Total time taken to complete the delivery.	Numerical values (10 – 270 minutes)
Category	Classification of the delivered item.	Categorical values (e.g., clothing, electronics, sports, cosmetics, toys)

3.2. Preprocessing

In this study, machine learning and ensemble learning models are proposed using an Amazon delivery dataset. The dataset consists of 43,739 different delivery records and 15 features, containing the delivery data of Amazon's e-commerce system.

The data preprocessing phase involves six distinct steps: filling missing values, removing missing values, converting categorical data, extracting meaningful features, creating classes, and normalization. The preprocessing steps were chosen to address the specific characteristics of the dataset and enhance model performance. Missing values were imputed using the mean method to preserve the data's overall distribution. While other techniques such as median or mode imputation were considered, the mean method was deemed most appropriate given the low percentage of missing values (0.123%). Similarly, categorical variables were encoded using a combination of label encoding and one-hot encoding to ensure compatibility with machine learning algorithms. The data preprocessing phase involves six distinct steps:

1. Filling Missing Values: Missing values were filled using the mean method to prevent any bias in the model training. This technique was chosen because it preserves the overall distribution of the data, which is crucial when working with large-scale datasets. In the original dataset, there are 54 missing values for the “Agent_Rating” attribute. These missing cells are filled by calculating the mean of the other “Agent_Rating” values. Given the total number of records in the dataset, the ratio of missing values to the total data is 0.123%. Therefore, completing the missing values based on the mean has a minimal impact on the overall structure of the dataset.

2. Removing Missing Values: In the original dataset, some records have many cells corresponding to different attributes marked as “NaN”. This results in incomplete information for those records and can cause issues with the performance of machine learning models. Thus, such records are removed. Out of the total 43,739 records in the dataset, only 91 records contain “NaN” values, which constitute merely 0.208% of the total dataset. Due to the very low ratio of missing values, removing these records has a minimal impact on data integrity.

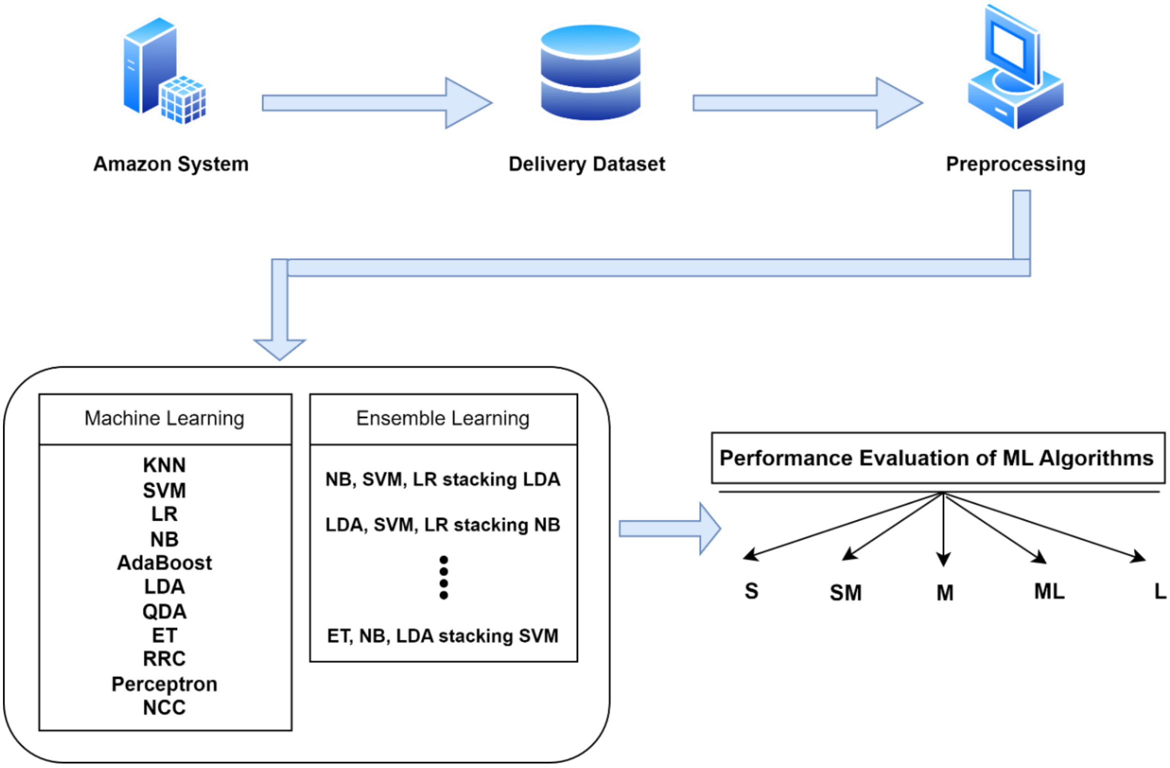

Figure 1. Data processing for machine and ensemble learning

3. Converting Categorical Data: Categorical data were converted into numerical representations using label encoding and one-hot encoding techniques. This conversion was necessary because most machine learning models work with numerical inputs, and categorical features need to be transformed into a format that can be interpreted by the model. The dataset contains categorical values for the attributes “Weather,” “Traffic,” “Area,” “Vehicle,” and “Category.” These categorical values are converted into numerical representations. For example, the values high, jam, low, and medium in the “Traffic” column are represented by 1, 2, 3, and 4, respectively.

4. Extracting Meaningful Features: The attributes “Order_Date,” “Order_Time,” and “Pickup_Time” consist of date and time values. To express datetime values as numerical values, they are converted into timestamp values. A timestamp typically represents the number of seconds since January 1, 1970. This conversion is particularly useful in data analysis and machine learning models dealing with time data (Dyreson and Snodgrass 1993)**.** For instance, for a record with “Order_Date” as 2022-03-19, “Order_Time” as 11:30:00, and “Pickup_Time” as 11:45:00, the combined “Order_Datetime” value represents March 19, 2022, at 11:30:00 and is converted into a timestamp value of 1,647,691,800 seconds. Similarly, the “Pickup_Datetime” value of March 19, 2022, at 11:45:00 is converted into a timestamp value of 1,647,692,700 seconds. These timestamp values facilitate the numerical processing and analysis of time data in data analysis and machine learning models. Similarly, distance data is obtained using the latitude and longitude values in the “Store_Latitude,” “Store_Longitude,” “Drop_Latitude,” and “Drop_Longitude” columns. The Haversine formula is used to convert the latitude-longitude information of the store and drop locations into distance data. This formula gives the distance along a straight line passing through the center of the Earth between two points and represents the shortest distance between two points on the Earth's surface (Winarno, Hadikurniawati, and Rosso 2017)**.** The values of longitude and latitude are determined using Equation 1. Subsequently, the intersection of the axis (c) is calculated as described in Equation 2. The final step in the Haversine method involves calculating the actual distance between two points using Equation 3.

$$a = \sin^2\left(\frac{\Delta\phi}{2}\right) + \cos(\phi_1)\cdot\cos(\phi_2)\cdot\sin^2\left(\frac{\Delta\lambda}{2}\right) \tag{1}$$

$$c = 2 \cdot \backslash\, \mathrm{atan}\, 2(\sqrt{a}, \sqrt{1-a}) \tag{2}$$

$$d = R \cdot c \tag{3}$$

5. Creating Classes: In classification algorithms, each feature must belong to a specific class. The Amazon delivery dataset's delivery times, which range from 10 to 270 minutes, are divided into five classes: Short (S), Short-Medium (SM), Medium (M), Medium-Long (ML), and Long (L). This division ensures that each class has equal width. Consequently, delivery times are classified as follows: 10-58 minutes (S), 59-106 minutes (SM), 107-154 minutes (M), 155-202 minutes (ML), and 203-270 minutes (L). The "Delivery_Class" column, created in this manner, is added to the dataset. The purpose of creating these classes is to enhance the performance of machine learning models and enable more precise predictions of delivery times. By using equal-width classes, the duration range represented by each class is balanced, thereby minimizing data imbalance during the model training process.

6. Normalization: Machine learning models tend to bias towards higher value data; hence, it is necessary to represent the data on a specific scale. Normalization is performed to ensure that each data point has the same scale and importance. In this study, the min-max normalization technique is chosen because it is simple, flexible, and intuitive. Min-max normalization scales the values of features (feature columns) in the dataset to the range [0, 1] (Patro and Sahu 2015). The new value of each data point is calculated according to Equation 4.

$$X_n = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{4}$$

After the preprocessing steps, the dataset is input into machine learning algorithms, and the results are calculated according to different machine learning metrics.

3.3. Details of Proposed Machine Learning and Ensemble Learning Models

The application of machine learning and ensemble learning methods on a dataset containing Amazon delivery information aims to classify delivery times, optimize delivery efficiency, and identify potential improvement areas to enhance customer satisfaction. The selected models (e.g., KNN, SVM, Logistic Regression) were chosen based on their well-documented performance in classification tasks involving structured data. For instance, SVM is known for handling high-dimensional data well, while Random Forest is effective in preventing overfitting. These models were chosen to cover a broad spectrum of machine learning techniques, ensuring that the best possible model for the dataset is identified. By employing machine learning algorithms and ensemble learning techniques, the relationships among various features within the dataset will be analyzed, and models with the highest accuracy will be selected to evaluate the performance of logistics operations. This approach will make significant contributions to the development of data-driven strategies for improving delivery efficiency and customer experience.

Using machine learning and ensemble learning algorithms on delivery data allows for more accurate and precise predictions of delivery times, contributing to optimized logistics efficiency and increased customer satisfaction. These methods predict potential delays and issues based on historical data, identify bottlenecks in operational processes, and enable proactive measures. The use of accuracy, precision, recall, and F-score as evaluation metrics ensures a comprehensive assessment of model performance. F-score, in particular, is useful in this context as it balances both precision and recall, which are critical in ensuring timely and accurate delivery predictions in logistics operations. Additionally, they support data-driven decision-making processes, aiding in more effective resource utilization and reducing logistics costs. Consequently, continuous improvement of delivery processes and the implementation of innovative solutions become possible.

In machine learning, models are developed using classification techniques such as K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Logistic Regression (LR), Naive Bayes (NB), AdaBoost, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), ExtraTrees (ET), Ridge Regression Classifier (RRC), Perceptron, and Nearest Centroid Classifier (NCC). These models are trained to obtain results, categorizing data based on specific features and offering various advantages in different situations. KNN is effective for small datasets due to its simplicity and understandability (Guo et al., 2003: 988). SVM performs well with high-dimensional datasets and aims to find the best separating hyperplane (Wang and Hu, 2005). Logistic Regression provides fast and effective results, particularly in binary classification problems. Naive Bayes is a quick and computationally easy model under the assumption of independence. AdaBoost enhances accuracy by sequentially boosting weak learners. LDA separates data with linear combinations that provide maximum separation (Jelodar et al. 2019). QDA is effective when class boundaries are not linear (Ghojogh and Crowley 2019). ExtraTrees reduces the risk of overfitting by increasing model diversity (Ahmad, Reynolds, and Rezgui 2018). RRC prevents overfitting using regularization (He et al. 2014). Perceptron is quick and effective in simple linear separation problems (Gallant, 1990). NCC quickly classifies based on distance to each class's centroid (Sharma and Paliwal 2010). The selection of the most suitable model for specific data types and problems directly impacts the model's success. The performance of classification algorithms depends on various factors such as the quality of the training dataset, feature engineering, and model optimization techniques.

In machine learning, methods can be combined within a logical framework to create ensemble learning models. Bagging, stacking, and boosting form the three fundamental structures of ensemble learning. Firstly, in the bagging method, the dataset is usually divided into test and training groups in a 70/30 ratio. Specific numbers of bags are created by randomly and repeatedly sampling from the training data. Each bag is trained using well-known models. During decision-making, outputs are evaluated by averaging or voting. Similar to bagging, the boosting process also involves data splitting and random sampling. However, in boosting, each sample is independently trained and produces outputs like in bagging, giving each model an equal chance of success (Karakaya et al., 2022).

In the boosting process, three classifier sets are created simultaneously. Like bagging, the first and second classifiers are trained with various randomly selected segments of the dataset. The third classifier is trained on the data where the first and second classifiers fail. These three classifiers are then combined using the majority voting technique. On the other hand, the stacking method makes deci. P8sions based on the percentage of the feature area each classifier succeeds in. The outputs of the classifiers are combined with another classifier to make the final decision (Polikar, 2012: 8).

Ensemble learning using stacking is illustrated in Figure 2. Ensemble learning techniques such as stacking and bagging were chosen for their ability to combine the strengths of individual models. For example, stacking integrates diverse base models, capturing complementary patterns in the data. These techniques help in improving the robustness of predictions and reducing overfitting, which is particularly important when working with complex datasets like the Amazon delivery data. Compared to single classifiers such as SVM or Logistic Regression, ensemble methods show higher accuracy and resilience to overfitting in this study, as evidenced by their superior performance metrics. Four different base classifier examples are provided. Depending on the model design, more or fewer classifiers can be used. A new example is evaluated by each classifier for classification. The results of each classifier are then evaluated by a new meta-classifier. Based on the meta-classifier's result, the example data is labeled with a class tag (Sagi and Rokach 2018). The preference for using the stacking method in the proposed model is due to its ability to combine the strengths of different machine learning models, enhancing overall performance. Using stacking to classify delivery times based on the Amazon delivery dataset allows each model to capture different features, resulting in more precise and accurate predictions. Additionally, the stacking method increases model diversity, reducing the risk of overfitting and enhancing generalization capability. This approach improves the classification of delivery times and makes logistics operations more efficient.

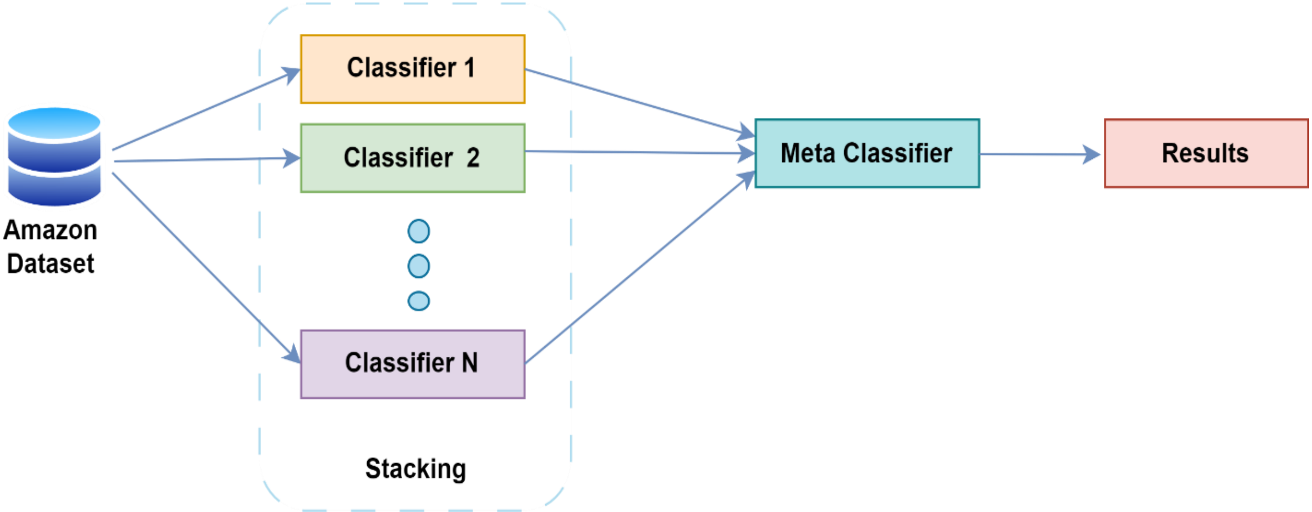

Figure 2. Stacking process in the ensemble learning

One of the ensemble learning methods is the voting classifier. The structure of a voting classifier consists of a machine learning model that evaluates predictions both hard and soft. In hard voting, the prediction with the most votes wins. In soft voting, the probabilities produced by each machine learning model are considered, and the class with the highest weighted average probability wins.

3.4. Experimental Design

The experimental studies of the proposed method were conducted on a computer with an AMD Ryzen 7 4800H processor running at 2.9 GHz, 16 GB RAM, and Windows 10 operating system, using Python 3.x. Stacking-based ensemble learning models, where machine and ensemble learning models such as KNN, SVM, LR, NB, AdaBoost, LDA, QDA, ET, RRC, Perceptron, and NCC serve as base and meta classifiers, were trained and tested on the Amazon delivery dataset. As a result of these processes, the models with the highest performance were identified as ensemble learning models. Each model was compared using the parameters of accuracy, precision, recall, and F-score (Reddy and Karthikeyan 2022). These metrics are obtained from the confusion matrix of the model's output, which includes four states: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) (Vujović 2021). The first metric evaluated using these states is accuracy, which indicates the ratio of correct predictions to the total number of predictions made. The accuracy formula is provided in Equation 5.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

Precision, which is the ratio of correctly predicted positive results to all predicted positive results, is given by Equation 6.

$$Precision = \frac{TP}{TP+FP} \tag{6}$$

Recall, defined as the ratio of correctly predicted positive results to all actual positive results, is shown in Equation 7.

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

The F-score, representing the weighted harmonic mean of precision and recall, is calculated using Equation 8.

$$Fscore = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{8}$$

4. RESULTS and DISCUSSION

In this study, various machine learning and ensemble learning models are evaluated for classifying delivery statuses using the Amazon delivery dataset. Performance metrics such as precision, recall, accuracy, and F-score are presented in Table 3 for each machine learning model and in Table 4 for each ensemble learning model. The results indicate that ensemble methods generally outperform individual classifiers.

Among the machine learning models, ET demonstrated superior performance, achieving the highest accuracy (0.978389) and F-score (0.978315), effectively capturing the delivery status. LDA also proved to be a robust choice with high accuracy (0.974189) and precision (0.974753). SVM performed well among individual classifiers, with an accuracy of 0.971592 and a high F-score of 0.971414. LR and NB also demonstrated good performance, with accuracies of 0.960825 and 0.962199, respectively, highlighting their effectiveness for this classification problem.

In contrast, models such as KNN and RRC exhibited lower performance metrics, revealing limitations in handling the complexity of the dataset. Perceptron and NCC also had relatively low accuracy and F-scores, indicating that these models are not well-suited for this particular application. The superior performance of the ET model can be attributed to its ability to handle large and complex datasets by reducing the risk of overfitting through randomized tree ensembles. SVM's effectiveness, on the other hand, stems from its capacity to handle high-dimensional data and find optimal hyperplanes for classification. In contrast, models like KNN and NCC, which rely heavily on proximity measures, struggle with the dataset's complexity and variability, leading to lower performance metrics.

Tablo 3. Machine learning model results

Model	*Precision*	*Recall*	*Accuracy*	*F-score*
KNN	0.691618	0.687743	0.687743	0.687265
SVM	0.971792	0.971592	0.971592	0.971414
LR	0.961722	0.960825	0.960825	0.960278
NB	0.962634	0.962199	0.962199	0.962118
AdaBoost	0.679882	0.803436	0.803436	0.727570
LDA	0.974753	0.974189	0.974189	0.973957
QDA	0.955301	0.954486	0.954486	0.954418
ET	0.978672	0.978389	0.978389	0.978315
RRC	0.689599	0.633906	0.633906	0.576764
Perceptron	0.714537	0.616953	0.616953	0.603055
NCC	0.523986	0.517297	0.517297	0.512587

The results demonstrate that ensemble learning approaches yield successful performance. When using NB, SVM, and LR as base models, and ET as the meta model, the highest accuracy (0.994196) and F-score (0.994201) were achieved. Similarly, high performance was observed when using LDA, LR, and ET as base models, and SVM as the meta model (accuracy: 0.994425, F-score: 0.994422). Notably, the highest performance (accuracy: 0.998931, F-score: 0.998930) was achieved with SVM, NB, and LDA as base models, and ET as the meta model. This result indicates that the ET model has a strong classification capacity when combined with other models. Conversely, some combinations showed relatively lower performance. For instance, when ET, SVM, and LDA were used as base models, and NB was selected as the meta model, both accuracy and F-score were recorded at 0.982971. This suggests that certain meta models may not perform optimally with specific combinations of base models. ET-based meta models and SVM combinations, in particular, stand out in enhancing classification accuracy.

Tablo 4. Ensemble learning (stacking) model results

Base Model - Meta Model	*Precision*	*Recall*	*Accuracy*	*F-Score*
Base Model (NB, SVM, LR) - Meta Model(LDA)	0.972463	0.972279	0.972279	0.972176
Base Model (NB, SVM, LR) - Meta Model(QDA)	0.966922	0.963650	0.963650	0.964250
Base Model (NB, SVM, LR) - Meta Model(ET)	0.994230	0.994196	0.994196	0.994201
Base Model (LDA, SVM, LR) - Meta Model(NB)	0.974985	0.973807	0.973807	0.973940
Base Model (QDA, SVM, ET) - Meta Model(LR)	0.990145	0.990149	0.990149	0.990145
Base Model (LDA, LR, ET) - Meta Model(SVM)	0.994425	0.994425	0.994425	0.994422
Base Model (SVM, NB, LDA) - Meta Model(ET)	0.998932	0.998931	0.998931	0.998930
Base Model (ET, NB, LDA) - Meta Model(SVM)	0.995343	0.995342	0.995342	0.995340
Base Model (ET, SVM, LDA) - Meta Model(NB)	0.983242	0.982971	0.982971	0.982985
Base Model (ET, SVM, NB) - Meta Model(LDA)	0.981309	0.981214	0.981214	0.981180

According to Tables 3 and 4, the ET model achieved the highest accuracy (0.978389) and F-score (0.978315). SVM and LDA also demonstrated high performance, with accuracy values of 0.971592 and 0.974189, respectively. Conversely, ensemble learning models, particularly when using SVM, NB, and LDA as base models and ET as the meta model, exhibited superior performance with the highest accuracy (0.998931) and F-score (0.998930). The practical significance of these findings lies in the ability of the top-performing models to predict delivery times more accurately than traditional approaches. This allows e-commerce companies to enhance operational efficiency by better managing delivery schedules and minimizing delays. Furthermore, the ability to predict delivery times with such precision offers companies a competitive edge, as faster and more reliable delivery services directly impact customer satisfaction. These models can be applied in real-time logistics operations, enabling proactive responses to potential delays, improving route planning, and ultimately optimizing the entire delivery process. This indicates that ensemble learning approaches are more effective than individual machine learning models. Ensemble learning provides better performance due to the generalization capability achieved by combining various models. In this context, ensemble learning models, especially those with ET-based meta models and SVM combinations, are significantly superior in enhancing classification accuracy. These results underscore the necessity of employing ensemble learning techniques for complex datasets.

Tablo 5. Sensitivity analysis data based on train/test split

Model	*Split*	*Precision*	*Recall*	*Accuracy*	*F-score*
ET	60/40	0.978448	0.978179	0.978179	0.978079
	70/30	0.978672	0.978389	0.978389	0.978315
	80/20	0.979297	0.979038	0.979038	0.978971
	90/10	0.978436	0.978236	0.978236	0.978180
Base Model (SVM, NB, LDA) - Meta Model (ET)	60/40	0.995134	0.995132	0.995132	0.995132
	70/30	0.998932	0.998931	0.998931	0.998930
	80/20	0.993364	0.993356	0.993356	0.993355
	90/10	0.994969	0.994960	0.994960	0.994961

The results presented in Table 5 illustrate how the performance metrics of the models vary with different split ratios. The experimental studies involve dividing the dataset into training and testing subsets at various split ratios to evaluate the sensitivity of model performance. This study employed a sensitivity analysis approach by dividing the dataset into training and testing subsets at multiple split ratios (60/40, 70/30, 80/20, and 90/10). This method allowed for the evaluation of model performance across different data distributions, providing insights into the robustness and generalizability of the proposed models. The top-performing models, particularly the ensemble learning approach using SVM, NB, LDA, and ET, offer significant practical usability for e-commerce companies. These models can be integrated into logistics management systems to predict delivery times with high accuracy, enabling more precise scheduling and efficient resource allocation. By identifying potential delays and optimizing delivery routes, these models contribute to reducing operational costs and enhancing customer satisfaction. The ET model consistently achieved high accuracy and F-score values across all split ratios. Specifically, the highest accuracy (0.979038) and F-score (0.978971) were attained with the 80/20 split ratio. The ET model demonstrated similarly high performance with other split ratios, indicating better generalization when a larger portion of the dataset is used for training. The ensemble learning model combining Base models (SVM, NB, LDA) and the Meta model (ET) showed the best performance across all split ratios. Notably, the highest accuracy (0.998931) and F-score (0.998930) were achieved with the 70/30 split ratio. These results indicate that ensemble learning models maintain high performance regardless of the train/test split ratio. In conclusion, while train/test split ratios can impact the performance of machine learning models, ensemble learning models demonstrate more stable performance. This underscores the ability of ensemble learning techniques to provide consistent results across different portions of the dataset.

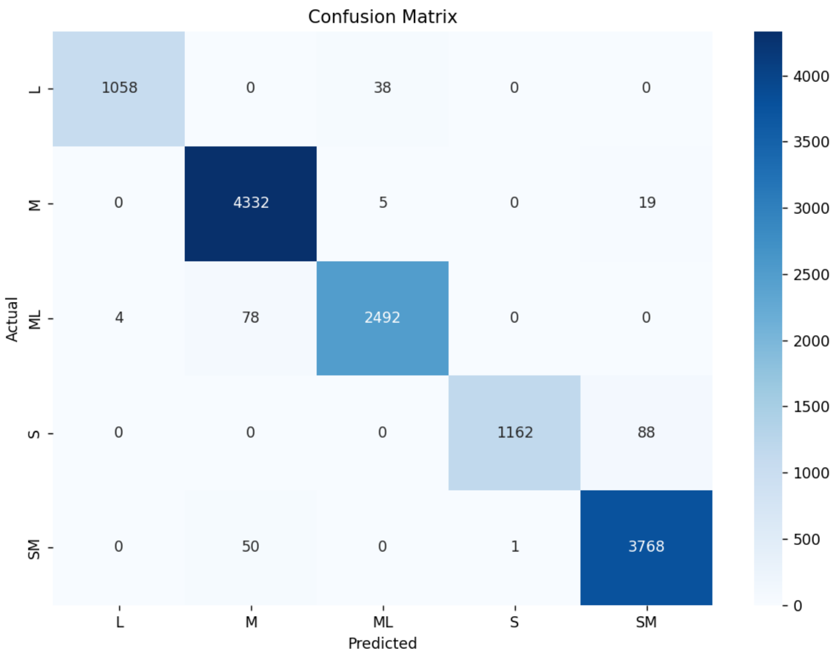

Figure 3. Confusion matrix for ET

Figure 3 presents the confusion matrix, illustrating the performance of the ET model in classifying delivery times. The delivery times are categorized into five classes: S, SM, M, ML, and L. Overall, the model exhibits a high classification accuracy.

- In the S class, there were 1162 correct classifications, with 88 instances misclassified as SM.
- In the SM class, there were 3768 correct classifications, with 50 instances misclassified as M and 1 as S.
- In the M class, there were 4332 correct classifications, with 5 instances misclassified as ML and 19 as SM.
- In the ML class, there were 2492 correct classifications, with 4 instances misclassified as L and 78 as M.
- In the L class, there were 1058 correct classifications, with 38 instances misclassified as ML.

These results indicate that the model generally distinguishes delivery times successfully, although some confusion exists, particularly between the M and ML classes. This confusion might stem from the indistinct boundaries between these classes. Overall, the ET model is effective in accurately classifying delivery times.

Figure 4 presents the confusion matrix for the ensemble learning model using SVM, NB, and LDA as Base Models and ET as the Meta Model, illustrating its performance in classifying delivery times.

- In the S class, 1243 instances were correctly classified, with 7 instances misclassified as SM.
- In the SM class, 3818 instances were correctly classified, with only 1 instance misclassified as S.
- In the M class, 4355 instances were correctly classified, with only 1 instance misclassified as ML.
- In the ML class, 2570 instances were correctly classified, with 4 instances misclassified as M.

- In the L class, 1093 instances were correctly classified, with only 3 instances misclassified as ML.

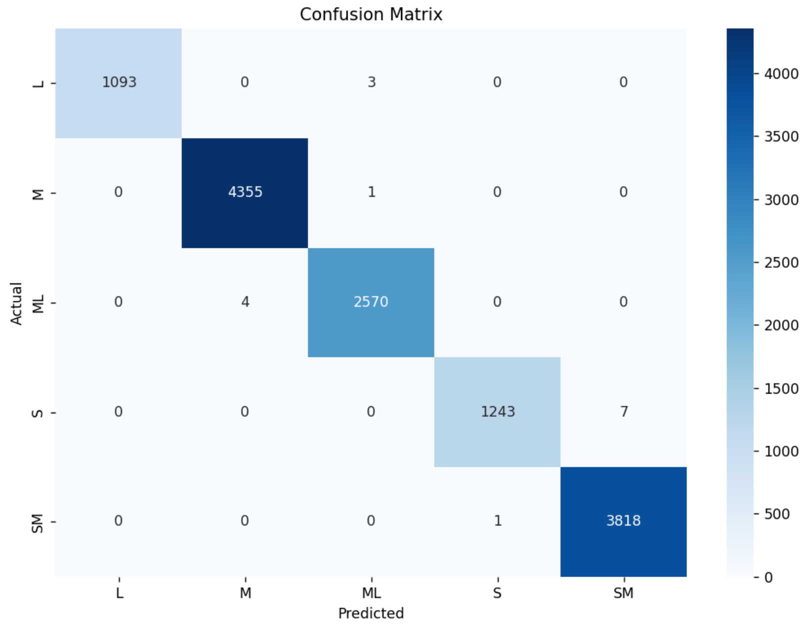

Figure 4. Confusion matrix for Base Model (SVM, NB, LDA) - Meta Model (ET)

These results indicate that the ensemble learning model is highly effective in accurately classifying delivery times, demonstrating superior performance. The minimal misclassifications, particularly between the M and ML classes, highlight the model's ability to effectively distinguish delivery times. Overall, the model exhibits a high level of accuracy in delivery time classification.

The results of this study highlight the significant potential of machine learning and ensemble learning models in transforming logistics operations, particularly in e-commerce. The high accuracy and reliability demonstrated by the ensemble learning models, such as the stacking method combining SVM, NB, and ET, offer robust solutions for predicting delivery times. These predictions are critical for optimizing key logistics functions, including route planning, resource allocation, and last-mile delivery. For instance, the ability to accurately forecast delivery times enables logistics companies to minimize delays, reduce fuel consumption, and allocate resources more effectively, thereby improving overall operational efficiency.

One of the most practical implications of these findings is the enhancement of customer satisfaction. Timely and reliable delivery is a cornerstone of customer loyalty in the competitive e-commerce market. By leveraging the predictive power of ensemble learning models, logistics companies can proactively identify potential delays and take corrective actions, such as rerouting deliveries or deploying additional resources. This proactive approach not only ensures timely deliveries but also fosters trust and reliability, which are crucial for sustaining customer relationships in the long term.

Beyond immediate operational benefits, these results also contribute to the broader logistics field by demonstrating the scalability and adaptability of ensemble learning methods for large-scale, dynamic datasets. Unlike traditional forecasting models, which often struggle with the complexities of modern logistics, the proposed models handle diverse variables such as traffic, weather, and geospatial data with high precision. Real-world applications could include integrating these models into logistics management systems for real-time decision-making. For example, warehouse operations can utilize these predictions to streamline inventory flow and optimize loading processes, while last-mile delivery teams can use them to enhance delivery route accuracy and reduce delivery windows.

Overall, the study underscores the transformative potential of machine learning and ensemble learning models in addressing logistical challenges. By improving prediction accuracy and operational efficiency, these models not only offer a competitive edge for e-commerce platforms but also pave the way for data-driven innovations in logistics and supply chain management.

5. CONCLUSION

In this study, we evaluated the performance of various machine learning and ensemble learning models in classifying delivery times using the Amazon delivery dataset. The results demonstrate that ensemble learning methods outperform individual machine learning models. Notably, the ensemble learning model using SVM, NB, and LDA as Base Models and ET as the Meta Model achieved the highest accuracy and F-score values across all split ratios. These findings indicate that ensemble learning approaches provide more consistent and superior performance on complex datasets. In practice, the application of top-performing models has the potential to revolutionize e-commerce logistics by providing highly accurate delivery time predictions. These models offer solutions that can be implemented in operational systems to improve customer satisfaction, optimize delivery routes, and reduce costs, making them highly valuable tools for modern logistics management.

Data preprocessing steps included filling and removing missing values, converting categorical data, extracting meaningful features, creating classes, and normalizing the dataset to make it suitable for machine learning and ensemble learning algorithms. The careful and accurate execution of these steps positively impacted the models' performance. Additionally, performance metrics such as precision, recall, accuracy, and F-score were effectively utilized to evaluate the models' accuracy and classification success. This study contributes to the field by demonstrating the effectiveness of ensemble learning methods in improving the prediction accuracy of e-commerce delivery times, a relatively underexplored area in large-scale logistics datasets. Unlike previous studies that primarily focused on smaller datasets, this research addresses the scalability of machine learning models, offering a more robust solution for real-world logistics management.

In conclusion, our study demonstrates that applying machine learning and ensemble learning models to Amazon delivery data allows for more accurate and precise prediction of delivery times. This contributes to optimizing logistics efficiency and enhancing customer satisfaction. The study aimed to answer two key research questions: (1) How can machine learning and ensemble learning models improve the accuracy of delivery time predictions? (2) Which model performs best across various logistics scenarios? (3) Which is the most effective model when the performance of methods is systematically compared with the aim of improving the efficiency of logistics processes? The results clearly show that ensemble learning models, particularly those involving ET, SVM, and NB, significantly enhance prediction accuracy. Additionally, the evaluation of different models underscores the superiority of ensemble methods in handling the complexities of large-scale logistics data, further validating their practical use in improving e-commerce delivery operations.

Despite the promising results, the study has certain limitations. The dataset used, although large, is confined to Amazon's delivery data, which may limit the generalizability of the findings to other logistics operations. Future studies could explore the application of these models across diverse datasets from different industries. Additionally, integrating real-time data and optimizing algorithms for specific logistical challenges, such as route optimization and vehicle management, could further improve model performance.

Future work has the potential to improve model performance by using expanded and diversified datasets. By exploring combinations of different machine learning and ensemble learning algorithms, innovative approaches that provide higher accuracy and efficiency in logistics operations can be developed. Applying these approaches to other logistics and supply chain processes could significantly enhance operational efficiency. Such data-driven strategies will continue to play a critical role in supply chain management practices, and further development of these methods could lead to significant innovations in the industry.

Conflict of Interest
No potential conflict of interest was declared by the author.

Funding
Any specific grant has not been received from funding agencies in the public, commercial, or not-for-profit sectors.

Compliance with Ethical Standards
It was declared by the author that the tools and methods used in the study do not require the permission of the Ethics Committee.

Ethical Statement
It was declared by the author that scientific and ethical principles have been followed in this study and all the sources used have been properly cited.

REFERENCES

Ahmad, M.W., Reynolds, J. and Rezgui, Y. (2018). "Predictive Modelling for Solar Thermal Energy Systems: A Comparison of Support Vector Regression, Random Forest, Extra Trees and Regression Trees", *Journal of Cleaner Production,* 203, 810-821. https://doi.org/10.1016/j.jclepro.2018.08.207

Alnahhal, M., Ahrens, D. and Salah, B. (2021). "Dynamic Lead-Time Forecasting Using Machine Learning in a Make-to-Order Supply Chain", *Applied Sciences*, 11(21), 10105. https://doi.org/10.3390/app112110105

Bruni, M.E., Fadda, E., Fedorov, S. and Perboli, G. (2023). "A Machine Learning Optimization Approach for Last-Mile Delivery and Third-Party Logistics", *Computers & Operations Research,* 157, 106262. https://doi.org/10.1016/j.cor.2023.106262

Chu, H., Zhang, W., Bai, P. and Chen, Y. (2023). "Data-Driven Optimization for Last-Mile Delivery", *Complex and Intelligent Systems,* 9(3), 2271-2284. https://doi.org/10.1007/s40747-021-00293-1

Deshmukh, Y., Patil, J., Pawar, P., Pardeshi, P. and Patil, H. (2024). "E-Commerce Product Delivery Analysis", *IRE Journals*, 7(12), 177-181.

Dyreson, C.E. and Snodgrass, R.T. (1993). "Timestamp Semantics and Representation", *Information Systems* 18(3), 143-166. https://doi.org/10.1016/0306-4379(93)90034-X

Erkmen, O.E., Nigiz, E., Z. Sarı, Z.D., Arlı, H.Ş. and Akay, M.F. (2022). "Delivery Time Prediction Using Support Vector Machine Combined with Look-Back Approach", *International Joint Conference on Engineering, Science and Artificial Intelligence-IJCESAI 2022*, 33-38.

Eskandaripour, H. and Boldsaikhan, E. (2023). "Last-Mile Drone Delivery: Past, Present, and Future", *Drones,* 7(2), 77. https://doi.org/10.3390/drones7020077

Gallant, S.I. (1990). "Perceptron-Based Learning Algorithms", *IEEE Transactions on Neural Networks*, 1(2), 179-191.

Ghojogh, B., and Crowley, M. (2019). "Linear and Quadratic Discriminant Analysis: Tutorial", arXiv:1906.02590. https://doi.org/10.48550/arXiv.1906.02590

Gore, S., Mishra, P.K. and Gore, S. (2023). "Improvisation of Food Delivery Business by Leveraging Ensemble Learning with Various Algorithms", *International Conference on Self Sustainable Artificial Intelligence Systems, ICSSAS 2023*, 221-229. https://doi.org/10.1109/ICSSAS57918.2023.10331669

Guo, G., Wang, H., Bell, D., Bi, Y. and Greer, K. (2003). "KNN Model-Based Approach in Classification", *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE. OTM 2003. Lecture Notes in Computer Science*, (Editors: R. Meersman, Z. Tari, D.C. Schmidt), 2888, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-39964-3_62

He, J., Ding, L., Jiang, L., and Ma, L. (2014). "Kernel Ridge Regression Classification", *Proceedings of the International Joint Conference on Neural Networks,* 2263-2267. https://doi.org/10.1109/IJCNN.2014.6889396

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y. and Zhao, L. (2019). "Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey", *Multimedia Tools and Applications,* 78(11), 15169-15211. https://doi.org/10.1007/s11042-018-6894-4

Karakaya, A., Ulu, A. and Akleylek, S. (2022a). "GOALALERT: A Novel Real-Time Technical Team Alert Approach Using Machine Learning on an IoT-Based System in Sports", *Microprocessors and Microsystems,* 93, 104606. https://doi.org/10.1016/j.micpro.2022.104606

Kazan, S. and Karakoca, H. (2019). "Makine Öğrenmesi İle Ürün Kategorisi Sınıflandırma", *Sakarya University Journal of Computer and Information Sciences,* 2(1), 18-27. https://doi.org/10.35377/saucis.02.01.523139

Khiari, J., and Olaverri-Monreal, C. (2020). "Boosting Algorithms for Delivery Time Prediction in Transportation Logistics", *IEEE International Conference on Data Mining Workshops, ICDMW 2020*, 251-258. https://doi.org/10.1109/ICDMW51313.2020.00043

Lochbrunner, M., and Witschel, H.F. (2022). "Combining Machine Learning with Human Knowledge for Delivery Time Estimations", *CEUR Workshop Proceedings,* 3121.

Patro, S. and Sahu, K.K. (2015). "Normalization: A Preprocessing Stage." ArXiv Preprint ArXiv:1503.06462.

Polikar, R. (2012). "Ensemble Learning", *Ensemble Machine Learning: Methods and Applications*, (Editors: C. Zhang and Y. Ma), 1-34, Springer, New York.

Reddy, B.H., and Karthikeyan, P. R. (2022). "Classification of Fire and Smoke Images Using Decision Tree Algorithm in Comparison with Logistic Regression to Measure Accuracy, Precision, Recall, F-Score." *14th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)*. 1-5, IEEE.

Rokoss, A., Syberg, M., Tomidei, L., Hülsing, C., Deuse, J. and Schmidt, M. (2024). "Case Study on Delivery Time Determination Using a Machine Learning Approach in Small Batch Production Companies", *Journal of Intelligent Manufacturing,* 35, 3937-3958. https://doi.org/10.1007/s10845-023-02290-2

Sagi, O., and Rokach, L. (2018). “Ensemble Learning: A Survey”, *WIRES Data Mining and Knowledge Discovery,* 8(4), 1-18. https://doi.org/10.1002/widm.1249

Salari, N., Liu, S. and Max Shen, Z-J. (2022). “Real-Time Delivery Time Forecasting and Promising in Online Retailing: When Will Your Package Arrive?”, *Manufacturing and Service Operations Management,* 24(3), 1421-1436. https://doi.org/10.1287/msom.2022.1081

Sharma, A., and Paliwal, K.K. (2010). “Improved Nearest Centroid Classifier with Shrunken Distance Measure for Null LDA Method on Cancer Classification Problem”, *Electronics Letters,* 46(18), 1251-1252. https://doi.org/10.1049/el.2010.1927

Sheng L., He, L., and Max Shen, Z-J. (2021). “On-Time Last-Mile Delivery: Order Assignment with Travel-Time Predictors”, *Management Science,* 67(7), 4095-4119. https://doi.org/10.1287/mnsc.2020.3741

Tsang, Y.P., Wu, C-H., Lam, H.Y., Choy, K.L. and Ho, G.T.S. (2021). “Integrating Internet of Things and Multi-Temperature Delivery Planning for Perishable Food E-Commerce Logistics: A Model and Application”, *International Journal of Production Research,* 59(5), 1534-1556.

Tsolaki, K., Vafeiadis, T., Nizamis, A., Ioannidis, D. and Tzovaras, D. (2023). “Utilizing Machine Learning on Freight Transportation and Logistics Applications: A Review”, *ICT Express,* 9(3), 284-295.

Wang, H. and Hu, D. (2005). “Comparison of SVM and LS-SVM for Regression”, *2005 International Conference on Neural Networks and Brain*, 279-283. https://doi.org/10.1109/ICNNB.2005.1614615

Winarno, E., Hadikurniawati, W. And Rosso, R.N. (2017). “Location Based Service for Presence System Using Haversine Method”, *2017 International Conference on Innovative and Creative Information Technology (ICITech)*, Salatiga, Indonesia, 1-4. https://doi.org/10.1109/INNOCIT.2017.8319153

Yüce T. and Kabak M. (2021). “Makine Öğrenmesi Algoritmaları Ile Detay Üretim Alanları İçin İş Merkezi Kırılımında Üretim Süresi Tahminleme”, *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi*,37(1), 47-60.

Zaghloul, M., Barakat, S. and Rezk, A. (2024). “Predicting E-Commerce Customer Satisfaction: Traditional Machine Learning vs. Deep Learning Approaches”, *Journal of Retailing and Consumer Services*, 79, 103865. https://doi.org/10.1016/j.jretconser.2024.103865

Željko, V. (2021). “Classification Model Evaluation Metrics”, *International Journal of Advanced Computer Science and Applications,* 12(6), 599-606. https://doi.org/10.14569/IJACSA.2021.0120670