Vol. 11, No. 1, April 2026, Page. 69-77
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

Predicting Successful Timely Order Completion in E-Commerce Using XGBoost and CatBoost Models

Muhammad Risqi Nuryana[1], Muhammad Nur Aulia Rahman[2], Ridwan Satrio Hadikusuma[3]

[1]Teknik Elektro, Fakultas Biosains Teknologi dan Inovasi, Universitas Katolik Atma Jaya, Jakarta
[2]Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional Indonesia, Jakarta
[3]Teknik Elektro, Fakultas Teknik, Universitas Pamulang, Tangerang Selatan
Email: [1]muhamma.12025004200@student.atmajaya.ac.id,
[2]muhammadnurauliarahman.2017@student.unas.ac.id, [3]dosen03333@unpam.ac.id

Abstrak

Keberhasilan penyelesaian pesanan tepat waktu merupakan aspek penting dalam operasional e-commerce karena kinerja pengiriman berpengaruh langsung terhadap kepuasan pelanggan dan kualitas layanan. Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor utama yang memengaruhi ketepatan waktu pengiriman serta mengevaluasi kinerja model XGBoost dan CatBoost dalam melakukan prediksi. Analisis dilakukan menggunakan dataset transaksi e-commerce publik yang mencakup karakteristik produk, waktu pemesanan, dan informasi pengiriman. Tahapan pra-pemrosesan data meliputi penanganan missing values, normalisasi fitur, serta feature engineering berbasis waktu. Optimasi hyperparameter dilakukan menggunakan framework Optuna untuk meningkatkan performa model. Hasil penelitian menunjukkan bahwa product_length_cm, product_weight_g, dan order_purchase_timestamp merupakan fitur yang paling berpengaruh terhadap keberhasilan pengiriman. Model CatBoost lebih sensitif terhadap karakteristik fisik produk, sedangkan XGBoost lebih menekankan pola pemesanan berbasis waktu, khususnya pada skala bulanan. Selain itu, penerapan Optuna terbukti mampu meningkatkan akurasi prediksi kedua model. Temuan ini dapat membantu platform e-commerce dalam mengoptimalkan manajemen persediaan, strategi logistik, dan penjadwalan pengiriman. Penelitian selanjutnya disarankan untuk mempertimbangkan faktor eksternal seperti kondisi cuaca, kepadatan pengiriman, dan lalu lintas guna meningkatkan akurasi prediksi dan efisiensi operasional.

Kata Kunci: **e-commerce, prediksi pengiriman, XGBoost, CatBoost, Optuna**

Abstract

The successful completion of orders on time is a critical aspect of e-commerce operations, as delivery performance directly affects customer satisfaction and overall service quality. This study aims to identify the key factors influencing delivery timeliness and to evaluate the predictive performance of XGBoost and CatBoost models. The analysis is conducted using a public e-commerce transaction dataset that includes product characteristics, order timestamps, and delivery-related information. Data preprocessing involves handling missing values, feature normalization, and time-based feature engineering. Hyperparameter optimization is performed using the Optuna framework to enhance model performance. The results indicate that product_length_cm, product_weight_g, and order_purchase_timestamp are the most influential features in predicting delivery success. CatBoost demonstrates higher sensitivity to physical product characteristics, while XGBoost places greater emphasis on temporal purchasing patterns, particularly at the monthly level. Furthermore, hyperparameter tuning using Optuna significantly improves the predictive accuracy of both models. These findings provide practical insights for e-commerce platforms in optimizing inventory management, logistics planning, and delivery scheduling. Future research is encouraged to incorporate external factors such as weather conditions, traffic congestion, and delivery density to further improve prediction accuracy and operational efficiency..

Keywords: **e-commerce, delivery prediction, XGBoost, CatBoost, Optuna**

This work is an open access article and licensed under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0)

1. INTRODUCTION

Previous studies indicate that delivery delays are not always perceived as a critical issue by customers when other factors such as product quality, secure packaging, and competitive pricing are

adequately maintained [1]. While these findings provide important insights into customer preferences, they are largely perception-based and do not sufficiently explain the operational factors that determine delivery timeliness from a logistics perspective. In particular, such studies rarely utilize transaction-level data to model delivery performance quantitatively.

A Journal tittled "Study on Analysis of Factors Influencing Delivery Time Preference & Customer Repurchase Intention for E-Tailing Websites," highlights an important finding: delivery delays are often not considered a major problem by customers. This occurs when other factors such as maintaining product quality during delivery, packaging security, and competitive pricing are also considered and influence their preferences [2], [14] . Therefore, to address the complexity of challenges in the context of delivery services in Indonesia, research utilizing artificial intelligence (AI) is needed to thoroughly analyze the factors that shape these delivery time preferences.

However, such research efforts face a major obstacle: the difficulty of finding high-quality open data sources in Indonesia [3]. Even when researchers take the initiative to collect data independently [4] , the data labeling process often proves to be costly and highly complex [5]. Another relevant study, conducted by Zhang et al [6], highlights an important trend in efforts to address delivery issues: the use of Artificial Intelligence (AI) with a focus on analyzing user behavior on the internet. Currently, various digital platforms allow users to easily voice their opinions and experiences on various matters [7].

Given the challenges in data limitations for AI research [8], this recorded online behavior can be a valuable source of data for building robust AI models [9]. In the context of predictive modeling, machine learning algorithms such as XGBoost and CatBoost have proven their effectiveness in many applications. Their main advantages lie in their ability to manage large and complex datasets, as well as capture complex nonlinear interactions between various features [10]. However, optimizing these algorithmic models requires a deep understanding of the appropriate hyperparameters. This is where the Optuna function, which serves as an automatic hyperparameter optimization framework, plays a crucial role by efficiently searching for the best hyperparameter combination to improve model performance [11].

The urgency of this research arises from the increasing need for e-commerce companies to improve the accuracy of order fulfillment predictions in dynamic logistics environments. Although machine learning models such as XGBoost and CatBoost have been widely applied for prediction tasks in logistics and supply chain management, limited studies have specifically compared the sensitivity of these two models to temporal features (such as delivery time patterns, order intervals, and seasonal demand fluctuations) versus physical features (such as package weight, dimensions, and shipping distance) in imbalanced e-commerce delivery datasets [12]. This research gap is important because feature characteristics significantly influence model performance and prediction reliability. By addressing this gap, the study aims to identify which algorithm performs more robustly under different feature conditions, enabling companies to better plan logistics operations, reduce the risk of delivery delays, and improve customer satisfaction. In addition, the findings are expected to provide practical insights for developing more efficient and adaptive logistics strategies in response to fluctuating market demand and operational challenges [13].

2. RESEARCH METHODS

The research methodology follows a structured workflow, as illustrated in figure 1. Each stage of the process is explained in detail in the following subsections, starting from data collection, data preprocessing, feature engineering, model training with hyperparameter optimization, and model evaluation.

p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

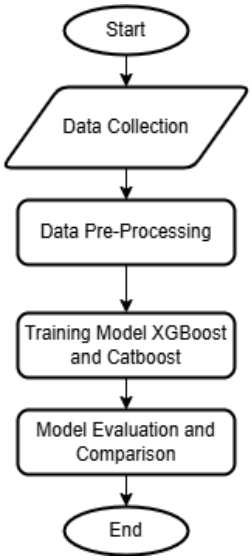

Figure 1. Research Flow

A. Data Collection

The dataset used in this study was obtained from a publicly available e-commerce transaction dataset on Kaggle. It contains transaction records related to customer orders, product characteristics, seller information, and delivery timestamps, making it suitable for analyzing factors affecting delivery timeliness. After data cleaning and feature selection, the final dataset consisted of 99,441 samples with 11 main predictive features, including both temporal and physical attributes.

The target variable was classified into two categories: on-time delivery and delayed delivery. The class distribution was imbalanced, where 93% of orders were successfully delivered on time, while only 7% experienced delays. This imbalance reflects real-world logistics conditions in e-commerce operations and requires careful model evaluation to avoid biased prediction results.

B. Data Preprocessing

Data preprocessing involved handling missing values, normalization, and encoding. Missing values in numerical features were imputed using median values to reduce the influence of outliers and improve model robustness. The feature product_category_name was removed due to its high proportion of missing values, which could introduce noise and reduce model reliability.

Numerical features were normalized to ensure consistent scaling during model training, while categorical identifiers such as seller_id were encoded appropriately for compatibility with the machine learning models.

C. Feature Engineering and Selection

Feature engineering was performed by extracting temporal information from the order_purchase_timestamp variable, including day of the week, month, year, quarter start, and month start indicators. These features were used to capture seasonal demand patterns and operational delivery trends. In addition, physical product features such as product_length_cm, product_width_cm, product_weight_g, and shipping distance were retained because they directly influence logistics handling and transportation efficiency. Features unrelated to delivery operations, such as payment-related variables, were excluded based on domain knowledge and preliminary statistical observations regarding relevance and completeness.

D. Model Training and Hyperparameter Optimization

Two machine learning classification models were employed in this study: XGBoost and CatBoost. These models were selected because of their strong performance in handling structured tabular data and

nonlinear feature interactions. The dataset was divided using an 80:20 train-test split, where 80% of the data were used for training and 20% for testing. A stratified split approach was applied to preserve the original class distribution due to the imbalanced nature of the dataset.

To improve model performance, hyperparameter optimization was conducted using the Optuna framework. Optuna automatically searches for the optimal hyperparameter combinations efficiently, reducing manual tuning efforts and improving prediction accuracy. In addition, stratified k-fold cross-validation was applied during optimization to ensure model stability and reduce overfitting risk.

E. Model Evaluation

Model performance was evaluated using accuracy, precision, recall, and F1-score to ensure balanced performance, particularly given the imbalanced nature of the dataset.

3. RESULT AND DISCUSSION

A. Analysis of Baseline Data and Performance

The graph in Figure 4 below provides an overview of the distribution of late deliveries in e-commerce as well as the significance of various features or variables that contribute to the delay. On the left side of the graph, there is a pie chart showing the proportion of late and not late shipments. Based on this visualization, it is known that only 7% of the total shipments are late, marked in blue, while 93% of the shipments successfully arrive on time, marked in red. This shows that the majority of orders in the analyzed e-commerce system were successfully delivered according to the specified time.

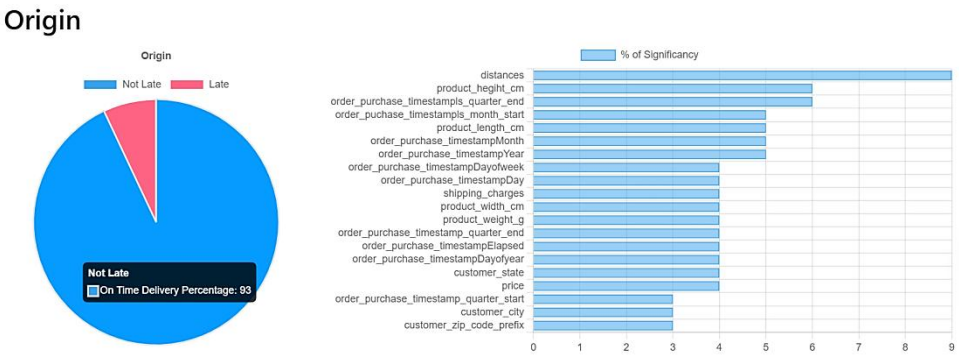

Figure 2. Initial Data Classification Results

On the right side of the graph, there is a bar chart that illustrates the significance level of each feature or variable in determining the delay or timeliness of delivery. Each bar represents the percentage contribution of a particular feature to delay prediction and can be clearly "order_purchase_timestamp_quarter_start" has a significance of 5%, indicating that the purchase time at the beginning of the quarter has an influence on the delay, although not very large. Meanwhile, the feature with the highest significance level is "distances" with a value of 9%, which means that the distance of product delivery is a major factor that can affect the possibility of orders arriving on time or experiencing delays.

Table 1. Feature Significance Results on Original Data

Feature	Significance (%)
Distances	9
order_purchase_tiestamp_quarter_start	5
order_purchase_timestampDayofweek	4
order_purchase_timestampls_month_start	4
product_length_cm	5

p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

product_width_cm	4
order_purchase_timestampsElapsed	4
order_purchase_timestampMonth	5
order_purchase_timestampYear	5
product_weight_g	6
Seller_id	5

Several features have a significance of 5%, namely product_length_cm, order_purchase_timestampMonth, order_purchase_timestampYear, and seller_id. The product_length_cm feature indicates that product length is an important factor in on-time delivery, possibly due to storage or transportation limitations for certain product sizes. The order_purchase_timestampMonth and order_purchase_timestampYear features suggest that the time of order purchase on a monthly and yearly scale can affect delays, possibly related to spikes in demand during certain periods such as holidays or major promotions. Meanwhile, the seller_id feature indicates that variations among sellers contribute to delivery speed, which may be influenced by each seller's operational efficiency in processing orders.

The highest significance percentage, 9%, was found in the distance feature. This may be due to variations in the number of orders processed by logistics systems at different times. Additionally, product width also plays a considerable role, possibly related to packaging and shipping constraints. The feature with the lowest significance level, order_purchase_timestamp_quarter_start at 3%, still contributes to delay prediction but has a smaller impact than other features. This suggests that purchases made at the beginning of a quarter have a more limited effect on order delays.

B. XGBoost Model Optimization Results with GridSearchCV

The XGBoost graph in Figure 5, which is optimized using GridSearchCV, aims to predict whether the delivery will be late or not, while evaluating how significant the contribution of each feature (variable) is to the prediction result. In this graph, there are two main parts that provide important information. First, the left part of the graph displays the distribution of the prediction results, which are divided into two categories: late and not late. From the graph, it can be seen that the percentage of late predictions is 6%, represented by the blue color, while the percentage of not late predictions is 94%, represented by the red color. This indicates that most shipments are predicted to arrive on time, with only a small percentage (6%) potentially experiencing delays.

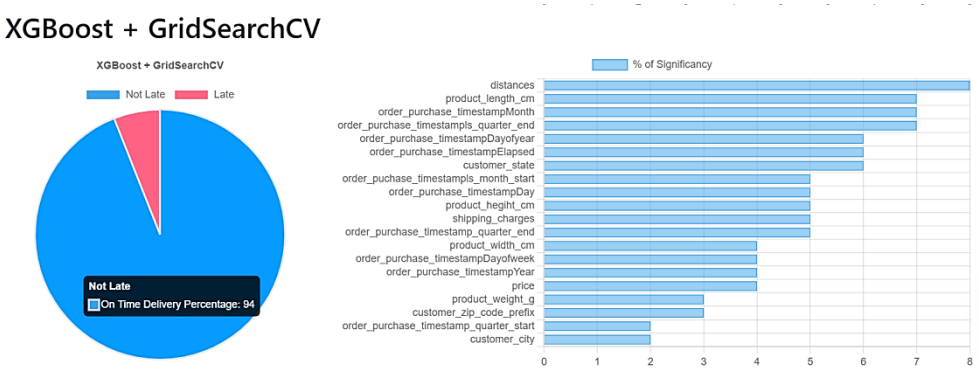

Figure 3. Classification of Data Significance Using the XGBoost + GridSearchCV Method

In addition this graph not only presents data, but also reveals the level of significance (importance) of each feature in influencing the model's prediction results. On the right side of the visualization, we can see how much each variable contributes to the decisions made by the model. For example, the "order_purchase_timestamp_quarter_start" feature has a relatively low significance level of only 2%. This indicates that this feature has a minimal impact on predicting delivery delays. Conversely, the "distances" feature has the highest significance level, reaching 8%. This figure indicates that delivery distance is the most influential factor in determining whether an order will be late or not. This

p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

significance level basically reflects how often and how important the feature is used by the XGBoost model in making predictive decisions.

Table 2. Feature Significance Results on XGBoost

Feature	Significance (%)
distances	8
order_purchase_tiestamp_quarter_start	2
order_purchase_timestampDayofweek	4
order_purchase_timestampls_month_start	5
product_length_cm	7
product_width_cm	4
order_purchase_timestampsElapsed	6
order_purchase_timestampMonth	7
order_purchase_timestampYear	4
product_weight_g	8
Seller_id	6

Based on Table 2, the most significant feature in influencing delivery timeliness is product_weight_g (8%), by product_length_cm and order_purchase_timestampMonth (7%). This shows that product weight and length, as well as order time on a monthly scale, greatly affect delivery delays. Other features such as order_purchase_timestampsElapsed and seller_id (6%) also have considerable influence, suggesting that order processing efficiency and seller characteristics contribute. Meanwhile, features with lower significance such as order_purchase_timestampDayofweek, order_purchase_timestampYear, and order_purchase_timestamp_quarter_start (2%), indicate that time factors based on day, year, and quarter start have less influence.

C. CatBoost Model Optimization Result With Optuna Framework

The CatBoost graph optimized using Optuna is used to predict whether a shipment will be late or not, as well as evaluate the significance of each feature on the prediction results. In Figure 6, the prediction distribution shows 8% late (red color) and 92% not late (blue color), indicating that most of the shipments are predicted to arrive on time.

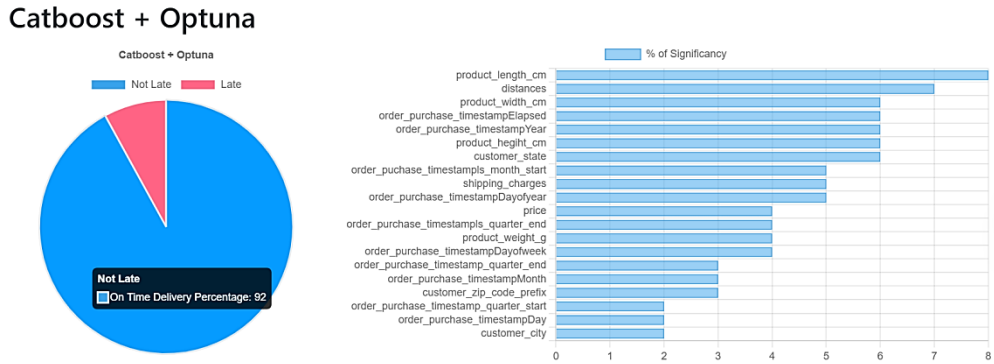

Figure 4. Classification of Data Significance Using the Catboost + Optuna Method

On the right side of the graph, the significance of the features is shown, with "order_purchase_timestamp_quarter_start" contributing 2%, while "product_weight_g" and "distances" have the highest contribution of 7%. This shows that product weight and product delivery distance are the most influential factors in determining delays, while order time has less influence.

Table 3. Feature Significance Results on CatBoost

Feature	Significance (%)
distances	7
order_purchase_tiestamp_quarter_start	7
order_purchase_timestampDayofweek	4
order_purchase_timestampls_month_start	5
product_length_cm	8
product_width_cm	6
order_purchase_timestampsElapsed	6
order_purchase_timestampMonth	3
order_purchase_timestampYear	6
product_weight_g	7
Seller_id	5

Based on Table 3, the features that have the greatest influence on delivery timeliness are product_length_cm (8%) and product_weight_g (7%), indicating that product dimensions and weight play an important role in determining delivery speed. Other factors such product_width_cm, order_purchase_timestampYear, and seller_id (6% each) also have significant contributions, indicating that order processing time and seller characteristics play a role in the delay or timeliness of delivery. The order_purchase_timestampsIs_month_start (5%) and order_purchase_timestampMonth (3%) features indicate that the timing of orders in the context of the beginning of the month or a particular month also has an impact on predicting delivery success. Meanwhile, order_purchase_timestampDayofweek (4%) and order_purchase_timestamp_quarter_start (2%) have a smaller effect, indicating that the day of the week and the beginning of the quarter factors are less significant.

D. Comparison of XGBoost and CatBoost Model Effectiveness

Based on the comparison results of XGBoost and CatBoost models in predicting the success of on-time order completion in e-commerce, there are some differences in the significance level of the features used as listed in Table 4.4. Both models show that product_length_cm and product_weight_g are the dominant factors, but CatBoost gives higher weight to product_length_cm (8%) than XGBoost (7%), while the opposite is true for product_weight_g, where XGBoost is higher (8%) than CatBoost (7%). The “distances” in the XGboost model has a significance of 8% which is 1% greater than CatBoost which is only 7%.

Table 4. Comparison of XGBoost and CatBoost Model Effectiveness

Feature	Significance XGBoost (%)	Significance CatBoost (%)
Distances	8	7
order_purchase_tiestamp_quarter_start	2	2
order_purchase_timestampDayofweek	4	4
order_purchase_timestampls_month_start	5	5
product_length_cm	7	8
product_width_cm	4	6
order_purchase_timestampsElapsed	6	6
order_purchase_timestampMonth	7	3
order_purchase_timestampYear	4	6
product_weight_g	8	7
Seller_id	6	5

Another difference is seen in product_width_cm, where CatBoost rates it more significantly (6%) than XGBoost (4%). In addition, order_purchase_timestampYear also has a higher weight on CatBoost (6%) than XGBoost (4%). In contrast, XGBoost gives more weight to order_purchase_timestampMonth (7%), while CatBoost only values this feature at 3%. Other features such as order_purchase_timestampDayofweek, order_purchase_timestampsElapsed, and order_purchase_timestampsIs_month_start have the same weight in both models, indicating that the purchase time factor still has a uniform influence in order timeliness prediction. From these results, it can be concluded that XGBoost emphasizes more on monthly order timing patterns, while CatBoost is more sensitive to the physical characteristics of the product in determining the success of on-time delivery.

The difference in feature sensitivity between XGBoost and CatBoost can be explained by the characteristics of each algorithm in processing data patterns. CatBoost tends to be more sensitive to physical product dimensions such as product_length_cm and product_width_cm because its ordered boosting mechanism is more effective in capturing complex nonlinear relationships between product size, packaging requirements, warehouse handling, and transportation constraints. Larger product dimensions generally require more storage space and special delivery arrangements, which directly affect delivery timeliness. In contrast, XGBoost shows higher sensitivity to temporal features such as order_purchase_timestampMonth because its gradient boosting structure performs well in identifying repetitive monthly purchasing patterns, seasonal demand fluctuations, and delivery workload peaks. This indicates that CatBoost is more suitable for analyzing operational constraints related to physical handling, while XGBoost is more effective in capturing demand patterns related to time-based logistics variations.

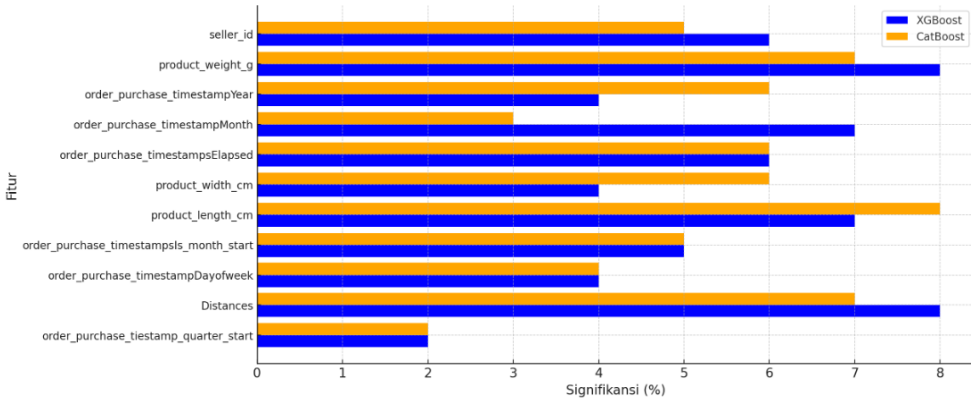

Figure 5. Comparison of Feature Significance in XGBoost and CatBoost Models

4. CONCLUSION

Based on the results of this study, the main factors affecting delivery timeliness in e-commerce are product dimensions, product weight, shipping distance, and purchasing time patterns. Features such as product_length_cm, product_width_cm, product_weight_g, distances, and order_purchase_timestampMonth have the highest influence on whether an order is delivered on time or delayed. This shows that both physical product characteristics and order timing play an important role in logistics performance, especially in warehouse handling, transportation processes, and delivery scheduling. Comparison between XGBoost and CatBoost shows that each model has different strengths in predicting delivery performance. XGBoost is more sensitive to temporal features, especially monthly purchasing patterns and changes in order volume, making it more suitable for identifying seasonal demand and delivery workload fluctuations. On the other hand, CatBoost gives more attention to physical product characteristics such as length, width, and weight, which are directly related to packaging, storage, and shipping constraints. This means that the selection of the prediction model should also consider the operational conditions of the logistics system, not only the final accuracy value.

In addition, the use of the Optuna framework helps improve the performance of both models through more efficient hyperparameter tuning. The evaluation using accuracy, precision, recall, F1-score, and AUC-ROC also provides better measurement results, especially because the dataset is

imbalanced, with 93% on-time deliveries and only 7% delayed deliveries. Overall, this study can support e-commerce companies in improving logistics planning, inventory management, and delivery scheduling to reduce delays and improve service quality.

REFERENCES

[1] N. Vasic, M. Kilibarda, and T. Kaurin, "The Influence of Online Shopping Determinants on Customer Satisfaction in the Serbian Market," *J. Theor. Appl. Electron. Commer. Res.*, vol. 14, no. 2, pp. 0–0, May 2019, doi: 10.4067/S0718-18762019000200107.

[2] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 50, Jun. 2018, doi: 10.35314/isi.v3i1.335.

[3] A. Erfina and M. F. Al-shufi, "ANALISIS SENTIMEN APLIKASI JASA KURIR DI PLAY STORE MENGGUNAKAN ALGORITMA NAIVE BAYES," *J. Sist. Inf. Dan Inform. Simika*, vol. 5, no. 2, pp. 103–110, Aug. 2022, doi: 10.47080/simika.v5i2.1789.

[4] L. Yao, C. Mao, and Y. Luo, "Graph Convolutional Networks for Text Classification," *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 01, pp. 7370–7377, Jul. 2019, doi: 10.1609/aaai.v33i01.33017370.

[5] S. Amaria, K. Guidedi, W. Lazarre, and Kolyang, "A Survey on Multimedia Ontologies for a Semantic Annotation of Cinematographic Resources for the Web of Data," *Acadlore Trans. AI Mach. Learn.*, vol. 1, no. 1, pp. 2–10, Nov. 2022, doi: 10.56578/ataiml010102.

[6] C. Zhang, Q. Li, and X. Cheng, "Text Sentiment Classification Based on Feature Fusion," *Rev. Intell. Artif.*, vol. 34, no. 4, pp. 515–520, Sep. 2020, doi: 10.18280/ria.340418.

[7] Z. Wang, S.-B. Ho, and E. Cambria, "A review of emotion sensing: categorization models and algorithms," *Multimed. Tools Appl.*, vol. 79, no. 47–48, pp. 35553–35582, Dec. 2020, doi: 10.1007/s11042-019-08328-z.

[8] W. Chen, X. Zheng, H. Zhou, and Z. Li, "Evaluation of Logistics Service Quality: Sentiment Analysis of Comment Text Based on Multi-Level Graph Neural Network," *Trait. Signal*, vol. 38, no. 6, pp. 1853–1860, Dec. 2021, doi: 10.18280/ts.380630.

[9] K. Sattar, Q. Umer, D. G. Vasbieva, S. Chung, Z. Latif, and C. Lee, "A Multi-Layer Network for Aspect-Based Cross-Lingual Sentiment Classification," *IEEE Access*, vol. 9, pp. 133961–133973, 2021, doi: 10.1109/ACCESS.2021.3116053.

[10] E. H. Muktafin, K. Kusrini, and E. T. Luthfi, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," *J. Eksplora Inform.*, vol. 10, no. 1, pp. 32–42, Sep. 2020, doi: 10.30864/eksplora.v10i1.390.

[11] I. K. Susanto, "Analisis Sentimen dan Topic Modelling Pada Pembelajaran Online di Indonesia Melalui Twitter," *JOINTECS J. Inf. Technol. Comput. Sci.*, vol. 6, no. 2, p. 85, May 2021, doi: 10.31328/jointecs.v6i2.2350.

[12] V. A. Flores, P. A. Permatasari, and L. Jasa, "Penerapan Web Scraping Sebagai Media Pencarian dan Menyimpan Artikel Ilmiah Secara Otomatis Berdasarkan Keyword," *Maj. Ilm. Teknol. Elektro*, vol. 19, no. 2, p. 157, Dec. 2020, doi: 10.24843/MITE.2020.v19i02.P06.

[13] S. Juniarsih, E. F. Ripanti, and E. E. Pratama, "Implementasi Naive Bayes Classifier pada Opinion Mining Berdasarkan Tweets Masyarakat Terkait Kinerja Presiden dalam Aspek Ekonomi," *J. Sist. Dan Teknol. Inf. Justin*, vol. 8, no. 3, p. 239, Jul. 2020, doi: 10.26418/justin.v8i3.39118.

[14] P. Chaturvedi and D. Agnihotri, "A Study on Analysis of Factors Influencing Delivery Time Preference & Customer Repurchase Intention for E-Tailing Websites," *Int. J. Manag. Stud.*, vol. VI, no. 1(6), p. 57, Jan. 2019, doi: 10.18843/ijms/v6i1(6)/07.

[15] M. Niazkar *et al.*, "Applications of XGBoost in water resources engineering: A systematic literature review (Dec 2018–May 2023)," *Environ. Model. Softw.*, vol. 174, p. 105971, Mar. 2024, doi: 10.1016/j.envsoft.2024.105971.

[16] F. Zhou *et al.*, "Fire Prediction Based on CatBoost Algorithm," *Math. Probl. Eng.*, vol. 2021, pp. 1–9, Jul. 2021, doi: 10.1155/2021/1929137.