OPEN ACCESS
VOL. 8, NO. 2, PP. 62–74, MAY 2026

ISSN : 2622-8106 (ONLINE)
DOI: 10.20895/INISTA.V8I2.2122

Journal of Informatics, Information System, Software Engineering and Applications (INISTA)

Classification of Package Delivery Duration Using Decision Tree Algorithm

Nava Azahra[1,*], Febriana Permatasari[2], Nalendra Cahaya Heraditya[3], Wahyu Dhani Prayoga[4], Muhammad Arifin[5]

[1,2,3,4,5]Department of Information Systems, Muria Kudus University
Lingkar Utara UMK, Gondangmanis, Bae, Kudus-59327 Jawa Tengah-Indonesia
[1]nicenava1405@gmail.com
[2]febrianaperr@gmail.com
[3]nalendracahaya10@gmail.com
[4]dhaniwahyu354@gmail.com
[5]arifin.m@umk.ac.id

Received: 29-04-2026; Revised: 22-05-2026; Accepted: 10-06-2026.

Abstract

This study aims to classify package delivery duration using the Decision Tree algorithm based on historical logistics data. The dataset used was obtained from PT Idenative and consists of package delivery data from January–February 2026, including attributes such as the number of items, destination location, delivery status, package weight, and delivery time. The data preprocessing stage was carried out through data cleaning, data transformation, delivery duration categorization, and categorical attribute encoding. Delivery duration was calculated based on the difference between Create Time and Delivered Time and then categorized into four classes: Fast, Normal, Slow, and Very Slow. The classification model was developed using the Decision Tree algorithm due to its strong interpretability in identifying data classification patterns. The dataset was divided into training and testing data with an 80:20 ratio. Model evaluation was conducted using a confusion matrix, accuracy, precision, recall, and F1-score metrics. The testing results showed that the model achieved an accuracy of 40.72%, a macro precision of 0.62, a macro recall of 0.35, and a macro F1-score of 0.34. The findings indicate that the Decision Tree algorithm was able to identify classification patterns based on logistics attributes such as the number of items and destination location, although the model performance remained limited due to data imbalance and similarities in characteristics among classes. This study demonstrates that the Decision Tree algorithm can be applied to support operational analysis in the logistics sector and may be further improved through the use of ensemble methods and data balancing techniques.

Keywords: *Classification, Data Mining, Decision Tree, Delivery Duration, Logistics*

Corresponding Author:
*Nava Azahra
Department of Information Systems, Muria Kudus University
Email: nicenava1405@gmail.com

I. INTRODUCTION

The increasing use of logistics services, in line with the growth of digital transactions and the rising demand for fast delivery, indicates that delivery timeliness has become one of the primary indicators of service quality. Timely delivery not only affects customer satisfaction but also plays an important role in maintaining customer trust and ensuring the sustainability of logistics businesses [1]. Uncertainty in delivery time not only impacts customer satisfaction but also creates challenges in distribution planning, operational scheduling, and objective performance evaluation of logistics services. Furthermore, delivery delays may lead to operational losses and reduce the overall efficiency of logistics systems [2]. Therefore, an approach capable of accurately predicting package delivery time is required to support improvements in service quality.

The rapid growth of logistics activities and package delivery services has generated a large amount of historical delivery data. However, these data are often underutilized in generating predictive insights, particularly for predicting delivery time. In fact, historical data have significant potential to support data-driven decision-making. In delivery operations, package delivery time is influenced by various factors such as package weight, destination location, shipping cost, and the number of items. Therefore, a data-driven approach is required to identify patterns that can support more accurate delivery time predictions [3]. In this study, delivery time prediction was not performed as a numerical estimation problem but rather as a classification problem. Delivery duration was calculated as the difference between Create Time and Delivered Time, which was then transformed into four categories: Fast, Normal, Slow, and Very Slow. This classification approach was selected because it simplifies the interpretation of prediction results, making them easier to implement in operational decision-making processes. Data Mining approaches, particularly classification methods, have been widely used to support predictive analytics and data-driven decision-making. The Decision Tree algorithm is one of the most commonly used methods due to its ability to handle both categorical and numerical attributes, as well as its capability to generate rule-based models that are easy to understand and interpret [4].

Several previous studies have demonstrated that the Decision Tree algorithm performs well in various classification tasks. For instance, the algorithm has been used to classify best-selling staple goods with an accuracy of approximately 85%, thereby supporting inventory management and business strategies [5]. Another study showed that Decision Tree could classify animal feed sales data into popular and unpopular categories through the construction of a decision tree based on dominant attributes [6]. Furthermore, the application of Decision Tree in vehicle inspection classification achieved an accuracy of 94.12%, indicating excellent performance in developing rule-based classification models [7]. Similarly, in flood-related rainfall classification, the algorithm achieved an accuracy of 83.33%, demonstrating its effectiveness in classification-based prediction tasks [8].

On the other hand, studies comparing Decision Tree and Random Forest have shown that Random Forest generally achieves higher accuracy than a single Decision Tree model, but at the cost of increased model complexity and lower interpretability [9]. Compared to other methods, the main advantage of Decision Tree lies in its ability to generate models that are easy to understand, transparent, and suitable for analyzing decision-making patterns. Nevertheless, this algorithm also has certain limitations, such as sensitivity to imbalanced data and the potential for overfitting if data processing is not properly performed [10].

Although the Decision Tree algorithm has demonstrated strong performance across various classification domains, its application in predicting package delivery time in the logistics sector remains limited. Most previous studies have focused on general classification problems, such as best-selling versus non-best-selling products, pass versus fail outcomes, or accepted versus rejected results. In contrast, classification based on delivery duration categories as prediction targets has received relatively little attention. This indicates a research gap in the application of Decision Tree within the logistics domain, particularly in modeling delivery time based on categorized duration classes (Fast, Normal, Slow, and Very Slow) using historical delivery data [11, 12].

Although the Decision Tree algorithm has been widely used in classification tasks, most existing studies focus on general prediction problems and do not specifically address delivery duration classification in the logistics field. In addition, many studies emphasize model accuracy without providing interpretable insights into the factors affecting delivery performance. Therefore, there is a need for a study that not only applies classification methods to delivery duration but also explores key attributes contributing to delivery delays using real logistics data.

Based on the aforementioned problems, this study aims to develop a classification model capable of predicting package delivery time using the Decision Tree algorithm based on logistics attributes such as delivery status, pickup type, sender and recipient locations, package weight, number of items, and package value. The main contribution of this study lies in the application of the Decision Tree algorithm to classify delivery duration using real logistics data, as well as in identifying key factors affecting delivery performance through an interpretable model. The resulting model, in the form of a decision tree, is expected not only to provide prediction results but also to offer insights into the factors influencing delivery duration across different categories (Fast, Normal, Slow, and Very Slow).

Furthermore, based on the model evaluation results, which achieved an accuracy of 40.72%, this study demonstrates that the Decision Tree algorithm is capable of identifying package delivery duration patterns based on logistics attributes. Therefore, it can support operational decision-making and package delivery service evaluation, although further improvements are still required to enhance classification performance.

II. RESEARCH METHOD

This study employed a Data Mining approach to classify package delivery duration based on historical data. This approach was selected because it is capable of extracting hidden patterns from large datasets to generate predictive information [13]. In accordance with the research objectives stated in the introduction, the method used in this study was the Decision Tree algorithm, which was chosen due to its ability to produce an interpretable model in the form of decision rules. This characteristic makes the model easy to understand and applicable for operational decision-making in the logistics field. In addition, the algorithm can effectively handle both numerical and categorical data.

Data processing and model development were carried out using the Python programming language through the Google Colab platform, with the support of libraries such as Pandas, NumPy, and Scikit-learn. In the Decision Tree algorithm, the selection of the best attribute is performed using the concepts of entropy and information gain to optimally construct the decision tree structure [11].

This study was conducted through several systematic stages following the data processing workflow in Data Mining, starting from data collection to model evaluation. These stages were intended to ensure that the data were properly processed in order to produce an optimal classification model. The overall research workflow is illustrated in Figure 1.

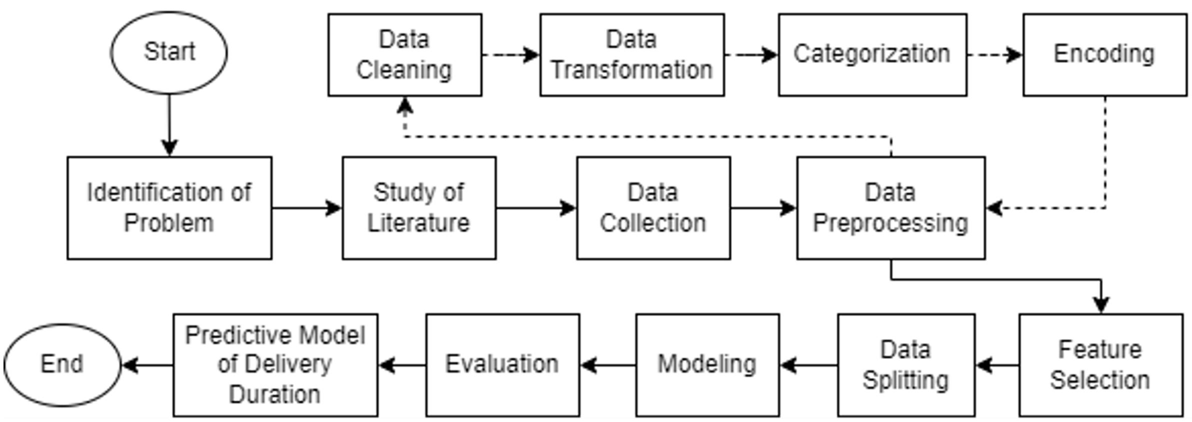

Fig. 1. Research Flow.

Figure 1 presents the research workflow, beginning with problem identification, followed by data processing, and ending with the development and evaluation of the delivery duration prediction model. After the problem identification stage, the study proceeded to the data collection stage. The collected data were then processed through several systematic stages, including preprocessing, feature preparation, modeling, and finally evaluation. Each stage was carried out sequentially to ensure data quality and model accuracy. The stages are described as follows:

A. *Data Collection*

The dataset used in this study consisted of historical package delivery data obtained from PT Idenative, covering the period from January to February 2026. The dataset represents real operational data from logistics activities, making it highly relevant for analyzing delivery performance and classification tasks. The data contain various attributes related to delivery characteristics, including package weight, sender and recipient locations, delivery status, estimated delivery time, and actual delivery time.

At this stage, an initial data exploration process was conducted to understand the structure and characteristics of the dataset. This process included identifying the number of records, examining attribute types (categorical and numerical), and analyzing data distribution across different variables. Understanding these aspects is essential to ensure that the dataset is suitable for further processing and modeling [14].

In addition, attribute selection was performed to determine which variables were relevant to the research objectives. Attributes that had direct or indirect relationships with delivery duration were retained, while irrelevant or redundant attributes were excluded. This step helps improve model performance and reduce computational complexity.

The data collection stage plays an important role in ensuring that the dataset used is valid, representative, and aligned with the research objective of classifying package delivery duration based on historical logistics data. After the data were collected and understood, the next step was data preprocessing to improve data quality and prepare the dataset for modeling.

B. *Data Preprocessing*

Data preprocessing is a fundamental stage in preparing raw data before entering the modeling phase. In real-world datasets, particularly in the logistics domain, data often contain inconsistencies, missing values, and noise that may negatively affect model performance. Therefore, preprocessing is required to improve data quality and ensure that the dataset is suitable for classification using the Decision Tree algorithm.

Data preprocessing in this study was carried out through several sequential steps, as illustrated in Figure 2, including data cleaning, data transformation, categorization, and encoding. Each process was performed in sequence to ensure that the data were accurate, consistent, and ready for analysis.

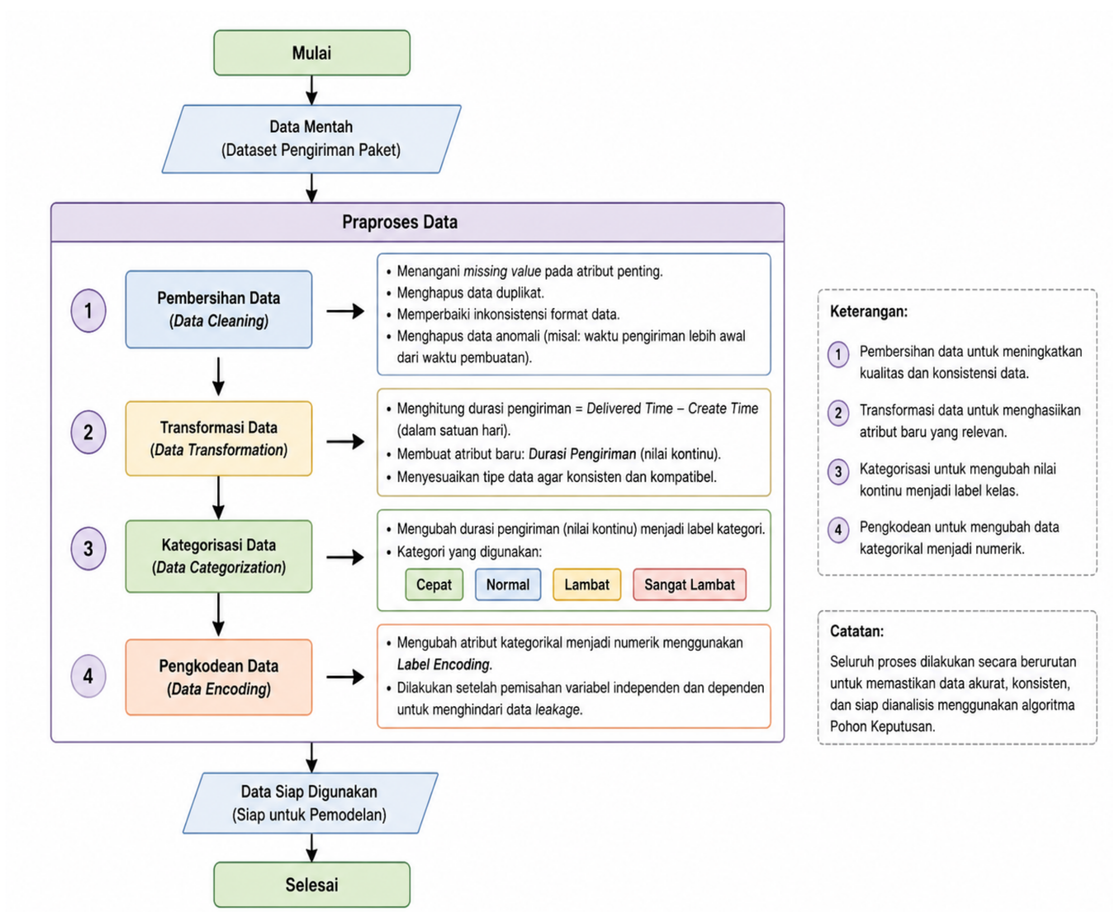

Fig. 2. Research Workflow.

1) Data Cleaning

Data cleaning is the initial step in preprocessing, aimed at improving the overall quality of the dataset. In this study, several cleaning techniques were applied to ensure data reliability. First, missing values were identified and handled appropriately. Records with incomplete essential attributes, such as delivery timestamps, were either corrected (when possible) or removed to avoid bias in duration calculations. Second, duplicate data entries were detected and eliminated to prevent redundancy that could affect the model learning process.

In addition, inconsistencies in data formats were corrected, particularly for timestamp attributes such as Create Time and Delivered Time. The standardization of these formats is essential to ensure accurate delivery duration calculations. Any anomalies, such as invalid or illogical timestamps (e.g., delivery time occurring earlier than package creation time), were also identified and addressed. Through this process, the dataset became cleaner, more consistent, and more reliable for subsequent analysis.

2) Data Transformation

After the cleaning process, the dataset underwent a transformation process to generate new variables relevant to the research objectives. The main transformation performed in this study was the calculation of delivery duration. Delivery duration was calculated as the difference between Create Time and Delivered Time, expressed in days. This newly generated attribute represents the actual time required for each package to be delivered. The transformation process converts raw timestamp data into meaningful information that can be directly utilized for classification.

In addition, the transformation process may include restructuring or simplifying certain attributes to improve compatibility with machine learning algorithms. Numerical values may be normalized or converted into categorical representations when necessary, depending on the modeling requirements. This stage is essential because it bridges the gap between raw data and meaningful features that can be processed by the Decision Tree algorithm.

3) Categorization

After the transformation stage, delivery duration values were categorized into four distinct classes: Fast, Normal, Slow, and Very Slow. This categorization was based on predefined duration intervals that represent different levels of delivery performance. The purpose of this step was to transform continuous numerical values into discrete categorical labels, which served as the target variable in the classification process. Thus, the prediction problem was converted from a regression task into a classification task, making it more suitable for the Decision Tree algorithm [15].

Categorization also simplifies the interpretation of results. Instead of predicting exact delivery times, the model classifies deliveries into easily understandable categories, which are more meaningful for decision-making in logistics management. In addition, this step helps identify patterns related to delivery performance, such as the proportion of delayed deliveries or the frequency of fast deliveries. These insights are highly valuable for improving operational efficiency and service quality.

4) Encoding

Categorical features were transformed into numerical representations using encoding techniques so that they could be processed by machine learning algorithms. The dataset was first separated into independent variables (features) and the dependent variable (target), and encoding was applied only to the feature variables [16]. This step is important to prevent data leakage and ensure that the model learns from valid input data. All preprocessing steps were implemented using Python on the Google Colab platform, utilizing libraries such as Pandas and Scikit-learn. After the preprocessing stage was completed, the next step was feature selection to identify the most relevant attributes for the classification model.

C. Feature Selection

Feature selection was performed to identify and retain the most relevant attributes that contribute to delivery duration prediction. This process helps reduce data dimensionality, eliminate irrelevant features, and improve model efficiency. By selecting only important features, the model becomes easier to interpret and may achieve better performance in classification tasks. After the relevant features had been selected, the dataset was then divided into training and testing datasets to support the modeling and evaluation processes.

D. Data Splitting

After the preprocessing stage was completed, the dataset was divided into training and testing datasets. The data split was performed using an 80:20 ratio with the assistance of the `train_test_split` function from the Scikit-learn library. The training subset was used to build and train the Decision Tree model, while the testing subset was used to evaluate the model's performance on previously unseen data [1]. This separation allows the evaluation results to reflect the model's ability to generalize rather than merely memorize the training data. By applying this approach, the model can be assessed objectively based on its predictive performance.

E. Modeling (Decision Tree)

The analysis model used in this study is a classification model based on the Decision Tree algorithm. At this stage, a classification model is built using the Decision Tree algorithm by utilizing the attributes that have been selected as predictor variables. The process of forming the decision tree is carried out by selecting the best attributes based on entropy and information gain values.

a. At this stage, a classification model is built using the Decision Tree Algorithm by utilizing the selected attributes as predictor variables [17]. The process of constructing the decision tree involves selecting the best attributes based on entropy and information gain values.

   Entropy is used to measure the level of uncertainty or impurity within a dataset. A higher entropy value indicates that the data is more heterogeneous, while a lower entropy value indicates that the data is more homogeneous. The entropy calculation is expressed in Equation 1.

$$\text{Entropy}(S) = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{1}$$

   Explanation:

   $\text{Entropy}(S)$: Represents the level of uncertainty or impurity within the dataset S.
   n: Denotes the total number of distinct classes present in the dataset.
   p_i: Represents the proportion, or equivalently the probability, of instances in dataset S that belong to class i.
   $\log_2$: Refers to the logarithm of the probability p_i computed with base 2.

b. To identify the most appropriate attribute for data partitioning, the concept of information gain is employed. Information gain quantifies the decrease in entropy resulting from dividing the dataset according to a specific attribute. The attribute that yields the highest information gain is chosen as the decision node, as it ensures the most effective separation of the data [11]. The information gain is calculated using Equation 2.

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \cdot \text{Entropy}(S_v) \tag{2}$$

   Explanation:

   $\text{Gain}(S, A)$: A measure of the effectiveness of attribute A in classifying the dataset S.
   $\text{Entropy}(S)$: The entropy of the original dataset before splitting.
   S_v: Subset of dataset S for which attribute A has value v.
   $|S|$: The total number of data samples in dataset S.
   $\frac{|S_v|}{|S|}$: The proportion of subset S_v relative to the entire dataset.
   $\text{Entropy}(S_v)$: The entropy of subset S_v after splitting.
   $\text{Values}(A)$: The set of all possible values of attribute A.

Based on the calculations in Equation 1 and Equation 2, the tree formation process is carried out iteratively until all data can be classified into the appropriate classes [18]. The final result is a decision tree structure that can be translated into easily understandable classification rules. After the model is successfully built, the final stage is model evaluation to assess its performance in classifying delivery duration.

F. Evaluation

The constructed model is evaluated to measure its performance in classifying the data. Performance evaluation is conducted using several metrics, namely accuracy, precision, and recall, along with a confusion matrix to analyze the distribution of predicted results against actual data [19].

a. Accuracy measures the overall correctness of the model and is calculated using Equation 3.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}} \times 100\% \tag{3}$$

Explanation:

Accuracy: The proportion of correctly classified data instances compared to the total number of data instances.

TP (True Positive): Refers to the count of positive instances that are accurately classified as positive.

TN (True Negative): Denotes the number of negative instances that are correctly identified as negative.

FP (False Positive): Represents the number of negative instances that are erroneously classified as positive.

FN (False Negative): Indicates the number of positive instances that are incorrectly classified as negative.

b. Precision measures the accuracy of positive predictions made by the model. A high precision value indicates that the model makes fewer errors in predicting positive classes. Precision is calculated using Equation 4.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

Explanation:

Precision: Represents the proportion of predicted positive instances that are correctly classified.

TP (True Positive): Denotes the number of instances accurately identified as belonging to the positive class.

FP (False Positive): Refers to the number of instances from the negative class that are incorrectly classified as positive.

c. Recall measures the model's ability to correctly identify all relevant instances within a specific class. It is calculated using Equation 5.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

Explanation:

TP (True Positive): The count of instances that are accurately classified as belonging to the positive class.

FN (False Negative): The number of positive instances incorrectly classified as negative.

d. F1-Score measures the balance between precision and recall in evaluating the performance of a classification model. It is particularly useful when dealing with imbalanced data, as it considers both false positives and false negatives. It is calculated using Equation 6.

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

Explanation:

F1-Score: The harmonic mean of precision and recall, used to measure the balance between precision and recall.

Precision: The accuracy of positive predictions made by the model.

Recall: The ability of the model to correctly identify positive instances.

Based on Equation 3, Equation 4, Equation 5, and Equation 6, the model's performance can be thoroughly analyzed to assess the accuracy and reliability of the model in classifying package delivery durations. Based on all the stages carried out, this study produces a classification model for package delivery duration using the Decision Tree algorithm, developed through a systematic and structured data processing pipeline. Each stage, from data collection, cleaning, transformation, to evaluation, is integrated to ensure the generation of an optimal model. The resulting model is then used to analyze classification performance, which will be discussed further in the results and discussion section.

III. RESULTS AND DISCUSSION

This section presents the results of the data processing and modeling stages, including data transformation, class distribution, decision tree construction, and model evaluation. The analysis aims to assess the performance of the Decision Tree algorithm in classifying delivery duration based on logistics attributes. The results are discussed to identify patterns, evaluate model effectiveness, and understand factors influencing delivery performance.

A. Results of Data Transformation and Distribution

Based on the data transformation process, delivery duration was calculated as the difference between the shipping time (Create Time) and the delivery completion time (Delivered Time) in days. The duration was then categorized into four classes: Fast, Normal, Slow, and Very Slow. The distribution of delivery duration classes is presented in Table I.

TABLE I. DISTRIBUTION OF DELIVERY DURATION CLASSES.

No	Duration Category	Sample Size	Percentage (%)
1	Fast	189	2,47
2	Normal	2.813	36,77
3	Slow	3.004	39,27
4	Very Slow	1.644	21,49
	Total	7.650	100,00

The data presented in Table I were obtained from the transformation and categorization processes of delivery duration using a historical package delivery dataset consisting of 7,650 records. Delivery duration was calculated based on the difference between Create Time and Delivered Time in days and was subsequently grouped into four categories: Fast, Normal, Slow, and Very Slow. The number of records in each category was then calculated to determine the class distribution within the dataset. The percentage of each category was obtained by comparing the number of records in each class with the total number of delivery records.

Based on the results, the dataset was primarily dominated by the Slow category with 3,004 records (39.27%) and the Normal category with 2,813 records (36.77%). Meanwhile, the Very Slow category consisted of 1,644 records (21.49%), whereas the Fast category contained only 189 records (2.47%), making it the class with the smallest number of samples.

This distribution indicates that although the data are relatively balanced between the Slow and Normal categories, there is still a significant class imbalance, particularly in the Fast category. This imbalance may affect the performance of the classification model because the Decision Tree algorithm tends to learn patterns from majority classes more effectively than from minority classes. As a result, the model may encounter difficulties in recognizing patterns associated with categories that contain limited data, thereby affecting prediction results.

The categorization of delivery duration was performed based on predefined delivery time intervals according to the characteristics of the data distribution within the logistics dataset. This categorization was intended to simplify the classification process and provide clearer interpretation of package delivery performance. Furthermore, the class distribution imbalance requires attention because it may influence model evaluation results, particularly recall and F1-score values for minority classes.

The classification model was developed using the Decision Tree algorithm based on selected logistics attributes as predictor variables. The resulting decision tree is presented in Figure 3. The model constructs a hierarchical structure by recursively selecting the most informative attributes to split the data, allowing the classification process to be carried out step by step [20]. This structure enables the model to represent decision rules that explain how different logistics factors influence delivery duration.

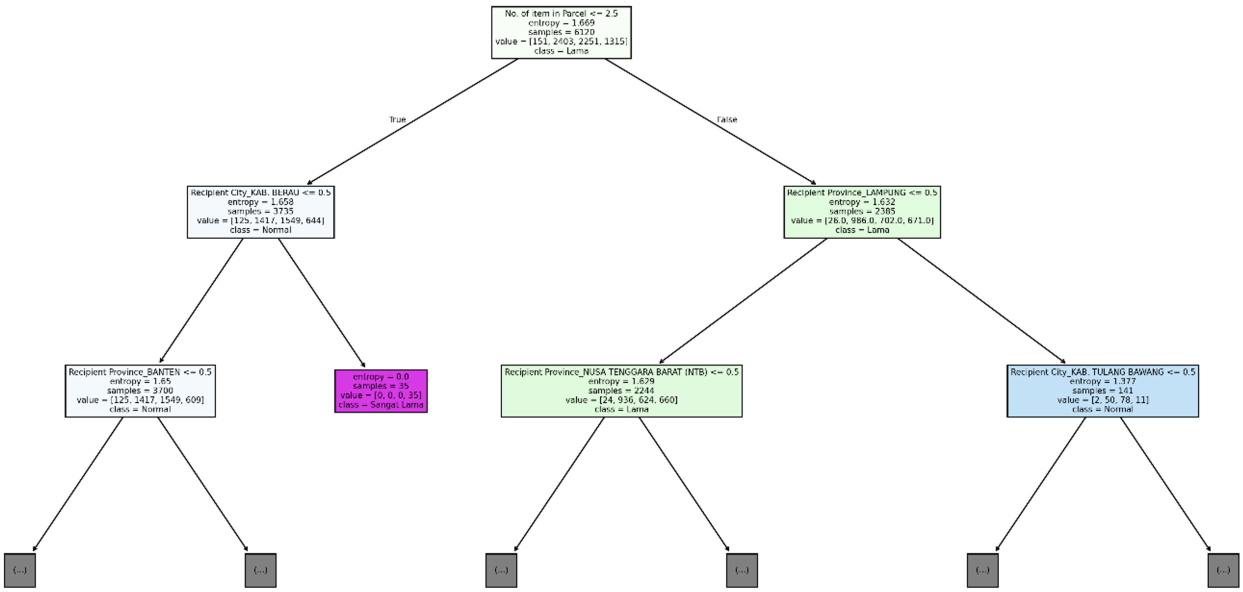

Fig. 3. Decision Tree Model.

Figure 3 illustrates the structure of the Decision Tree model constructed based on the selected logistics attributes. The model forms a hierarchical tree by recursively splitting the dataset using attributes such as the number of items (No. of item in Parcel), destination location (Recipient City and Recipient Province), and other relevant logistics variables. These attributes are selected based on their ability to reduce uncertainty in the data, allowing the model to distinguish between different delivery duration categories.

The classification process is performed sequentially through a series of decision rules at each node. At the upper levels of the tree, the model uses the most influential attributes to divide the data into broader groups. For instance, the model may initially split the data based on the number of items in the parcel, which reflects shipment complexity, and then further refine the classification using destination-related attributes such as city and province. This step-by-step splitting process continues until the data is classified into the final categories of delivery duration [12]. The resulting decision tree demonstrates that attributes related to shipment characteristics and destination location play a significant role in determining delivery time. In particular, the number of items and destination region are identified as key factors influencing whether a shipment falls into the Fast, Normal, Slow, or Very Slow category. This indicates that both operational factors and geographical distribution contribute to variations in delivery performance.

Furthermore, it should be noted that the visualization shown in Figure 3 represents a simplified view of the decision tree, where only a limited depth is displayed for clarity and readability. The full model used for evaluation has a deeper structure, allowing it to capture more complex patterns in the data. This approach ensures that the model remains interpretable in the figure while maintaining its predictive capability during the evaluation process. Overall, the Decision Tree model is able to generate interpretable classification rules that describe the relationship between logistics attributes and delivery duration. These rules can be used to support decision-making in logistics operations, particularly in understanding factors that influence delivery delays and optimizing shipment processes.

Model performance is assessed using a confusion matrix as well as accuracy, precision, recall, and F1-score metrics. The outcomes of this evaluation are presented in Table II and Table III.

B. *Model Evaluation*

Model evaluation is conducted using a confusion matrix as well as accuracy, precision, recall, and F1-score metrics. The results of the model evaluation are shown in Table II.

In addition to the numerical representation shown in Table II, the confusion matrix is also visualized in Figure 4 to provide a clearer illustration of the model's classification performance across each delivery

TABLE II. CONFUSION MATRIX.

Actual / Prediction	Fast	Normal	Slow	Very Slow
Fast	12	21	5	0
Normal	5	397	160	0
Slow	1	395	205	0
Very Slow	0	163	157	9

duration category. This visualization facilitates the identification of both correct prediction patterns and classification errors made by the model. The values along the main diagonal of the matrix indicate the number of records that were correctly classified by the model, whereas values outside the diagonal represent misclassifications among categories.

Based on Table II and Figure 4, the model demonstrates relatively better performance in classifying the Normal category compared to other categories, as indicated by the higher number of correctly predicted instances for this class. However, a considerable number of misclassifications still occur between the Normal and Slow categories. This suggests that these two categories possess similar data characteristics, making it difficult for the model to distinguish between them optimally.

In addition, the Fast and Very Slow categories have relatively lower numbers of correctly classified instances compared to their actual number of records. This condition indicates that the model still faces challenges in recognizing patterns within minority classes, particularly due to the imbalanced data distribution. The confusion matrix visualization also reveals that most prediction errors tend to occur between adjacent classes, indicating an overlap in characteristics among delivery duration categories.

Overall, Figure 4 demonstrates that the Decision Tree algorithm is capable of learning classification patterns from logistics data; however, its performance remains suboptimal in consistently distinguishing all delivery duration categories. The findings from the confusion matrix are consistent with the evaluation results of accuracy, recall, and F1-score, which indicate that the model still has limitations in consistently classifying all delivery duration categories.

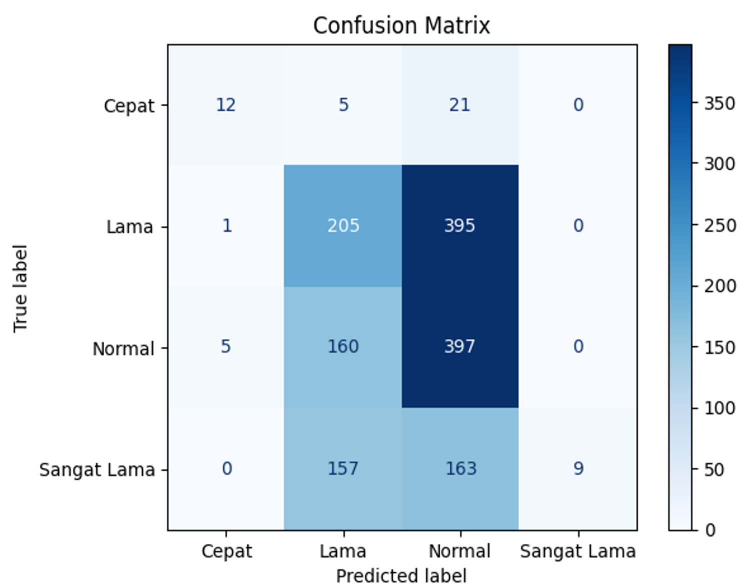

Fig. 4. Confusion Matrix Visualization.

The confusion matrix in Table II and Figure 4 provides a detailed view of how the model classifies each category. For the Normal category, the model shows relatively good performance with 397 correctly classified instances. However, there are still significant misclassifications, particularly where Normal data is predicted as Slow (160 instances), indicating that these two classes share similar characteristics. For the Slow category, although 205 instances are correctly classified, a large number of data points (395 instances) are misclassified as Normal. This suggests that the model has difficulty distinguishing between Slow and Normal categories due to overlapping patterns in the data.

In the Very Slow category, the model performs poorly, with only 9 correctly classified instances, while most data are incorrectly predicted as Normal (163 instances) and Slow (157 instances). This indicates that the model struggles to identify this category effectively, possibly due to insufficient distinguishing features or data distribution issues. Similarly, for the Fast category, only 12 instances are correctly classified, while a significant portion is misclassified as Normal (21 instances). This reflects the challenge of predicting minority classes, as the number of samples in this category is relatively small compared to others.

TABLE III. MODEL EVALUATION RESULTS.

Metrics	**Accuracy**	**Precision (Macro)**	**Recall (Macro)**	**F1-Score (Macro)**
Results	40,72%	0,62	0,35	0,34

Based on the evaluation results presented in Table III, the Decision Tree model achieved an accuracy of 40.72%, indicating that the model was capable of performing package delivery duration classification, although its performance remained at a moderate level. The model's ability to perform classification can be observed from the presence of correctly predicted instances in each category within the confusion matrix, particularly in the Normal and Slow categories, which exhibited relatively higher numbers of correctly classified instances compared to other categories.

The macro precision value of 0.62 indicates that the model performed reasonably well in generating positive predictions that matched the actual classes. This suggests that when the model predicted a particular delivery duration category, most of those predictions were consistent with the actual conditions. In addition, the model was able to establish classification patterns based on the logistics attributes used as predictor variables.

However, the macro recall value of 0.35 indicates that the model still has limitations in consistently identifying all actual instances across categories. This condition suggests that although the model successfully performed classification, its ability to identify all instances within certain classes remains suboptimal. The F1-score value of 0.34 also indicates that the balance between precision and recall is still relatively low, particularly due to the imbalanced data distribution and similarities in characteristics among classes, such as the Normal and Slow categories.

Overall, the testing results demonstrate that the Decision Tree algorithm was successfully applied to classify package delivery duration and was capable of generating interpretable decision patterns. Nevertheless, the model's performance still requires improvement to achieve more accurate and consistent classification results across all delivery duration categories.

C. *Discussion*

The results of this study indicate that the Decision Tree algorithm was capable of developing a classification model for predicting package delivery duration based on logistics attributes. However, the model performance remained at a moderate level, with an accuracy of 40.72%. These findings suggest that relationships exist between logistics attributes, such as the number of items, destination location, and delivery characteristics, and delivery duration; however, these relationships are not sufficiently strong to produce highly accurate predictions. This indicates that delivery duration is influenced not only by the attributes used in this study but also by other factors, such as operational conditions, travel distance, and external variables that were not included in the dataset. Furthermore, although the data distribution was relatively balanced, the Fast category still contained a significantly smaller number of samples. This condition made it more difficult for the model to effectively learn patterns within this class, as reflected by the relatively low recall value.

Based on the confusion matrix results, the model tended to misclassify records into adjacent categories, particularly between the Slow and Normal classes. This finding indicates that the boundaries between these categories possess similar characteristics, making them difficult for the model to distinguish. These results are consistent with previous studies stating that the Decision Tree algorithm has advantages in interpretability but also limitations in handling complex data patterns and unclear class boundaries [1]. Furthermore, based on feature importance analysis obtained from the Decision Tree model, the number of items (Number of Items in Package) and destination location (Recipient City and Recipient Province) were identified as the most influential factors in determining delivery duration. This finding is consistent with the structure of the decision tree, where these attributes frequently appeared at higher levels of the tree.

This suggests that deliveries involving larger numbers of items and more complex destination areas tend to require longer delivery times, thereby increasing the likelihood of delays. In addition, model performance was affected by class imbalance, particularly in the Fast category, which contained considerably fewer samples. As a result, the model tended to be biased toward majority classes, affecting its ability to accurately classify minority classes. To address these limitations, future studies may apply more robust methods such as Random Forest or Gradient Boosting, which have better capabilities for handling complex data patterns and reducing overfitting. Moreover, data balancing techniques such as SMOTE may also be employed to improve classification performance on minority classes. For logistics cases involving more complex route optimization and delivery time considerations, methods such as Genetic Algorithms and the A* algorithm may also be considered as potential approaches for future development.

IV. CONCLUSION

This study successfully developed a classification model for predicting package delivery duration using the Decision Tree algorithm based on logistics attributes. The model was able to identify patterns between delivery characteristics and delivery duration while generating interpretable decision rules. However, the model achieved only moderate performance, with an accuracy of 40.72%, indicating that it was not yet optimal in classifying all delivery duration categories. The results showed that attributes such as the number of items and destination location were significant factors influencing delivery duration. Nevertheless, the complexity of the delivery process and limitations of patterns. Future studies are recommended to improve model performance by applying ensemble methods, addressing class imbalance, and incorporating additional variables related to operational and environmental factors.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to PT Idenative for providing the dataset used in this research. The authors also thank the academic supervisors for their valuable guidance, support, and constructive feedback throughout the completion of this study.

REFERENCES

[1] A. H. Arayadiba, E. K. Nurnawati, N. Rahmawati, and I. Setiawan, "Prediksi tingkat risiko keterlambatan pengiriman barang menggunakan algoritma decision tree," *Jurnal SCRIPT*, vol. 13, 2025.

[2] I. Supriadi, A. Mauluddin, I. Ibrahim, and P. Sagala, "Implementasi algoritma simple hill climbing sebagai optimasi travelling salesman problem dalam penentuan rute pengiriman barang (studi kasus: Jasa pengiriman barang j&t)," *INFORMASI (Jurnal Informatika dan Sistem Informasi)*, vol. 15, no. 1, 2024. doi:10.37424/informasi.v15i1.220.

[3] A. Ramdani, C. D. Sofyan, F. Ramdani, M. F. A. Tama, and M. A. Rachmatsyah, "Algoritma klasifikasi data mining untuk memprediksi masyarakat dalam menerima bantuan sosial," *Jurnal Ilmiah Sistem Informasi*, vol. 1, no. 2, 2022. doi:10.51903/juisi.v1i2.363.

[4] P. W. Sari, G. Giwangga, N. D. Fitriyanti, A. A. Putra, and H. Irawan, "Implementasi algoritma c4.5 untuk klasifikasi data insentif karyawan pt. adhi cakra utama mulia," *Jurnal Komputer Antartika*, vol. 2, pp. 95–101, 2024. doi:10.70052/jka.v2i3.318.

[5] M. A. Melfia, K. A. Ramadhini, M. A. Maulana, A. L. Dasiva, A. A. Briantoro, and B. O. Lubis, "Penerapan data mining untuk klasifikasi data penjualan sembako terlaris dengan algoritma c4.5," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 3, 2025. doi:10.36040/jati.v9i3.13450.

[6] D. M. Musa *et al.*, "Penerapan data mining untuk klasifikasi data penjualan pakan ternak terlaris dengan algoritma c4.5," *Jurnal Teknologi Informatika dan Komputer*, vol. 10, pp. 172–186, Mar. 2024.

[7] R. R. Andarista and A. Jananto, "Penerapan data mining algoritma c4.5 untuk klasifikasi hasil pengujian kendaraan bermotor," *Jurnal Tekno Kompak*, vol. 16, no. 2, 2022. doi:10.33365/jtk.v16i2.1525.

[8] I. Risnawati *et al.*, "Klasifikasi data mining untuk mengestimasi potensi curah hujan berdampak banjir daerah menggunakan algoritma c4.5," *Jurnal INSAN (Journal of Information Systems Management Innovation)*, vol. 3, no. 2, 2023. [Online]. Available: https://jurnal.bsi.ac.id/index.php/jinsan/article/view/3050. Accessed: Mar. 20, 2026.

[9] A. Riansah, O. Nurdiawan, and R. Herdiana, "Penerapan algoritma random forest dan decision tree untuk meningkatkan akurasi klasifikasi penjualan pada toko bangunan," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 3, 2025. doi:10.36040/jati.v9i3.13622.

[10] F. G. Falentina, G. Y. Y. Wabdaron, D. Andiyani, M. S. Wondiwoi, and H. Sutejo, "Analisis penerapan data mining untuk klasifikasi penjualan makanan terlaris menggunakan algoritma decision tree (c4.5)," *Jurnal Ilmiah Sistem Informasi*, vol. 4, pp. 382–395, Jan. 2026. doi:10.51903/f2jegk76.

[11] F. Gorunescu, *Data Mining: Concepts, Models, and Techniques*, vol. 12. Berlin, Germany: Springer, 2011. doi:10.1007/978-3-642-19721-5.

[12] P. Jayadi, W. S. Bhagawan, and J. D. Aldida, "Class balancing and parameter tuning of machine learning models for enhancing aphrodisiac herbal plant classification," *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, vol. 7, pp. 183–192, May 2025. doi:10.20895/inista.v7i2.1832.

[13] P. W. Rahayu *et al.*, *Buku Ajar Data Mining*. Jambi, Indonesia: PT. Sonpedia Publishing Indonesia, 2024. [Online]. Available: https://www.researchgate.net/publication/377415198.

[14] I. Akbar, I. S. Samad, R. Rahmat, and S. Rosmiana, "Data mining analysis of k-means algorithm and decision tree for early detection of students at risk of dropping out," *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, vol. 7, pp. 148–162, May 2025. doi:10.20895/inista.v7i2.1630.

[15] M. Arifin, G. W. Bhawika, M. M. Habibi, W. Firdaus, D. E. Agustinova, and R. Rahim, "Application of the cluster classification data mining method to child illiteracy in indonesia," *Library Philosophy and Practice*, vol. 2021, pp. 1–8, 2021.

[16] J. N. Permana, R. Goejantoro, and S. Prangga, "Comparison of c4.5 algorithm and naive bayes for prediction of student study timeliness (case study: Departement of statistics mulawarman university)," *Jurnal Eksponensial*, vol. 13, Nov. 2022.

[17] D. Avianto and A. P. Wibowo, "Pembentukan pohon keputusan untuk penerima bantuan beras miskin menggunakan algoritma decision tree c4.5," *Jurnal Ilmiah NERO*, vol. 9, no. 1, 2024. [Online]. Available: https://journal.trunojoyo.ac.id/nero/article/view/28020. Accessed: Apr. 24, 2026.

[18] Z. Fatah and R. Atreji, "Klasifikasi penyakit gagal jantung mengunakan metode decision tree," *Jurnal Ilmiah Sistem Informasi*, vol. 4, pp. 844–854, Nov. 2025. doi:10.51903/ana03n66.

[19] S. Helmiyah and R. Pramestiawan, "Analisis komparatif algoritma machine learning dengan metrik akurasi, presisi, recall, dan f1-score pada dataset kacang kering," *Jurnal Ilmu Komputer dan Teknologi*, vol. 6, pp. 152–159, Oct. 2025. doi:10.35960/ikomti.v6i3.2031.

[20] F. A. K. Faridian, L. Riyandari, and S. Sulistiyasni, "Public sentiment analysis against identitas kependudukan digital application using decision trees," *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, vol. 8, no. 1, pp. 1–11, 2025.