MORFAI JOURNAL Multidisciplinary Output Research For Actual and International Issue
Multidisciplinary Output Research For Actual and International Issue (MORFAI Journal) ISSN (e): 2808-6635
Volumes 6 No. 3 (2026)

COMPARATIVE ANALYSIS OF TREE-BASED ALGORITHMS FOR CUSTOMER SATISFACTION CLASSIFICATION IN THE LOGISTICS INDUSTRY: A CASE STUDY OF JNE AND J&T EXPRESS

Kevin Benedicta[1], Rudi Nurdiansyah[2*]
[1,2]Universitas Negeri Malang, Malang, Indonesia
E-mail: kevin.benedicta.2205166@students.um.ac.id[1], rudi.nurdiansyah.ft@um.ac.id[2*]

Received	: 10 February 2026	Accepted	: 10 March 2026
Revised	: 20 February 2026	Published	: 18 March 2026

Abstract

The rapid growth of Indonesia's e-commerce sector has intensified competition within the logistics industry, positioning customer satisfaction as a critical determinant of competitive advantage. However, the multidimensional and non-linear nature of service quality complicates traditional statistical analysis. This study aims to compare the performance of three tree-based machine learning algorithms (Decision Tree, Random Forest, and Gradient Boosting) in classifying customer satisfaction for JNE and J&T Express, while identifying the key service quality dimensions driving satisfaction. Using a validated dataset of 408 respondents, individual service indicators are modeled as predictive features. Hyperparameter tuning is conducted through 500-iteration Randomized Search with 5-fold cross-validation. The results show that the Decision Tree achieves the highest performance for the JNE dataset with an accuracy of 78.05%, precision of 79.16%, recall of 78.05%, and an F1-score of 77.84%. In contrast, Gradient Boosting outperforms other models for the J&T Express dataset with an accuracy of 81.71%, precision of 81.69%, recall of 81.71%, and an F1-score of 81.37%. Furthermore, Feature Importance analysis consistently identifies Shipping Cost as the dominant predictor of satisfaction. These findings highlight the efficacy of tree-based machine learning in decoding complex satisfaction patterns, offering actionable, data-driven insights for logistics service providers.

Keywords: Customer Satisfaction, Decision Tree, Gradient Boosting, Random Forest, Service Quality.

INTRODUCTION

The rapid expansion of the e-commerce sector in Indonesia has significantly transformed the logistics landscape, creating a competitive environment for expedition service providers (Syihabudin et al., 2022). This growth is evidenced by Bank Indonesia (2025) data, which reveals a consistent increase in e-commerce transaction value from Rp. 205.5 trillion in 2019 to Rp. 487.01 trillion in 2024. As the primary bridge between sellers and buyers, logistics companies play a pivotal role in sustaining this digital supply chain ecosystem. This surge in demand has driven significant growth for shipping services acting as key partners on e-commerce platforms (Pranida & Kurniawardhani, 2022). Within this competitive market, JNE and J&T Express have emerged as two dominant market leaders, competing aggressively in terms of service quality, speed, and pricing strategies. According to Top Brand Awards (2024), JNE and J&T are the top 2 brands in the courier service subcategory in Indonesia in 2024.

Consequently, customer satisfaction has become the primary differentiator and a critical determinant of business success and loyalty in the modern market landscape (Kullolli et al., 2024). However, analyzing customer satisfaction in the logistics sector is increasingly complex due to the subjective and multidimensional nature of service quality. This complexity arises because customer perception is not driven by a single isolated factor, but rather by a combination of various service dimensions. In the context of logistics, these dimensions typically include reliability, responsiveness, and assurance (Tjendana & Pranitasari, 2024). Despite this complexity, customer satisfaction is traditionally measured using survey-based instruments analyzed through conventional statistical methods. For instance, Prasetyo et al., (2023) used linear regression and descriptive analysis to examine the relationship between service quality and customer satisfaction. While such methods can identify general trends, they often assume linear relationships between variables. To address this limitation, machine learning classification techniques offer a more robust analytical framework. Unlike traditional statistics, machine learning algorithms are designed to model complex, non-linear interactions between variables without relying on rigid assumptions (Barri et al., 2022; Dhillon et al., 2022). In particular, Tree-based algorithms are widely employed to address complex classification problems (Uddin & Lu, 2024). To ensure a comprehensive evaluation, this study benchmarks three distinct algorithms, each selected for its specific analytical strength. Decision Tree serves as the baseline for interpretable decision structures Mienye & Jere (2024), Random Forest improves robustness and reduces variance through ensemble learning (Akbari et al., 2024), and Gradient Boosting optimizes accuracy through sequential model refinement (Albostami et al., 2024). Nevertheless, empirical applications of advanced ensemble-based classification models within the Indonesian logistics satisfaction context remain limited. The existing study by Rahadian (2022) applied a C4.5 Decision Tree model using aggregated service dimension inputs, which may obscure nuanced indicator-level effects and reduce model generalization capability. Moreover, prior research deployed a single-tree model without explicitly addressing the issue of class imbalance, a common phenomenon where satisfied customers disproportionately outnumber dissatisfied ones. Failing to address this imbalance can lead to biased classification models that favor the majority class, thereby yielding misleading accuracy metrics (Khan et al., 2023).

To mitigate this problem, resampling techniques are commonly applied to rebalance the class distribution before model training. One of the most widely used approaches is the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic samples for the minority class by interpolating existing observations in the feature space. Compared to simple oversampling methods that replicate minority instances, SMOTE generates synthetic samples in the feature space, which encourages the classifier to learn broader and more general decision regions instead of highly specific ones that may lead to overfitting (Chawla et al., 2002). Therefore, incorporating SMOTE becomes essential to ensure that machine learning models can fairly capture patterns associated with both satisfied and dissatisfied customers. Addressing this gap, the present study advances the literature in three significant ways. First, it adopts an item-level modeling strategy by utilizing all individual service indicators as predictive features rather than aggregated dimension-level averages, thereby preserving informational granularity. Second, it implements a comparative benchmarking framework across three distinct tree-based algorithms under hyperparameter-optimized configurations to evaluate predictive stability and robustness. Third, it incorporates class imbalance handling using SMOTE to enhance model fairness and generalization capability. Through this integrated framework, the study not only strengthens predictive performance but also enhances interpretative depth via aggregated feature importance analysis. Consequently, this research contributes methodologically to service analytics literature while offering practical insights for strategic decision-making in the logistics industry.

THEORETICAL AND CONCEPTUAL FRAMEWORK

Customer Satisfaction

Customer satisfaction is conceptualized as a psychological state resulting from a consumer's evaluation of their service experience (Gam et al., 2023). It is determined by the gap between the service perceived by the customer and their initial expectations (Yahya et al., 2023). In the competitive landscape, customer satisfaction acts as a critical determinant for business survival and sustainability rather than just a transactional outcome (Masyhuri, 2022). Theoretically, satisfaction functions as a mediator that translates service quality into long-term customer trust and loyalty (Bhaskar & Subramanyam, 2021). Furthermore, satisfaction is considered cumulative in nature, as it is shaped by repeated service encounters rather than a single interaction. Consequently, for service-oriented industries, sustained customer satisfaction plays a pivotal role in strengthening market positioning and long-term competitiveness.

Service Quality Dimensions

Service quality is defined as a customer's overall evaluation of the excellence of a service (An et al., 2023). In contrast to the quality of physical products, which can be objectively measured, service quality is inherently intangible, subjective, and difficult to standardize due to its experiential nature (Álvarez-García et al., 2019). In the context of the logistics industry, service quality extends beyond the physical transportation of goods to encompass the entire customer experience, including order processing, tracking transparency, delivery accuracy, and post-delivery services (Hui et al., 2025). Consequently, service quality has been conceptualized through multiple evaluative dimensions that reflect specific attributes used by customers to assess service performance (Li et al., 2023).

Machine Learning Approaches

Numerous studies on customer satisfaction have relied on conventional statistical techniques. For example, Nuralam et al., (2024) employed Structural Equation Modeling to model latent constructs of service quality. While conventional statistical techniques are effective for testing established theories, customer satisfaction itself is inherently a multidimensional construct driven by various service quality factors and perceived value (Liu, 2024). The formation of satisfaction often involves complex and dynamic interactions among these attributes that traditional linear models may overlook. As noted by Dukalang et al. (2025), conventional statistical approaches generally assume linear relationships and face limitations in representing these nonlinear interactions. Consequently, when relationships are constrained within linear functional forms, the resulting models may fail to capture nonlinear patterns and interaction effects, potentially limiting predictive performance and explanatory depth. In contrast, Machine Learning (ML) offers a data-driven modeling paradigm that does not impose a predefined functional form on the data (Razavi, 2021). Machine Learning is designed to learn patterns directly from data, allowing it to automatically detect and model intricate, non-linear dependencies that traditional models often fail to capture (Guo et al., 2025). This capability makes Machine Learning particularly advantageous in analyzing large heterogeneous datasets where relationships between variables are dynamic and multifaceted (Kamm et al., 2023). Within the domain of service analytics, Machine Learning has been increasingly adopted to transform satisfaction analysis from a descriptive task into a predictive one. For instance, Ahmad et al. (2019) demonstrated the efficacy of machine learning by developing a predictive model to forecast customer churn, allowing companies to take proactive retention measures. Similarly, in the logistics and online delivery industry, Madani & Alshraideh (2021) employed tree-based algorithms to predict consumer purchasing decisions based on delivery time and service quality. These studies collectively demonstrate the effectiveness of data-driven models in capturing complex service-related patterns and enhancing predictive performance.

Tree-Based Classification Algorithms

Tree-Based algorithms are one of the most dominant approaches in Machine Learning classification tasks. As highlighted by Mienye & Jere (2024), Tree-Based algorithms have gained significant popularity across diverse domains due to their simplicity and inherent interpretability, which allow researchers to easily visualize and understand the decision-making process. The fundamental building block of this approach is the Decision Tree, a non-parametric method that recursively partitions the feature space into hierarchical structures based on the most discriminative attributes. However, single decision trees often suffer from high variance and limited generalization capabilities when applied to unseen data. To address these limitations, ensemble learning techniques such as Random Forest and Gradient Boosting are developed. Random Forest employs a parallel “bagging” mechanism to reduces variance, whereas Gradient Boosting utilizes a sequential approach to iteratively correct residual errors. As noted by Lundberg et al. (2020), Decision Tree, Random Forest, and Gradient Boosting are currently recognized as widely adopted algorithms for solving non-linear predictive tasks. Consequently, this study performs a comparative analysis of these three distinct algorithms to evaluate their efficacy in predicting customer satisfaction.

METHOD

Research Design

This research employs a quantitative method using a supervised machine learning classification approach. Figure 1 presents the research framework, which is structured into four major stages: Data, Data Preparation, Modeling & Evaluation, and Interpretation & Insight.

The first stage consists of validated JNE and J&T Express customer satisfaction datasets. These datasets contain multidimensional service quality indicators representing Warranty, Friendliness, Delivery Accuracy, Responsiveness, and Shipping Cost.

The second stage structures the raw data into a format suitable for predictive modeling. This stage ensures that the analytical structure aligns with supervised learning requirements.

The third stage applies three tree-based machine learning algorithms to capture nonlinear patterns between service quality attributes and customer satisfaction. Model performance evaluation and comparison enable identification of the most reliable predictive approach.

Finally, the analytical outcomes are interpreted to generate meaningful insights. Model results are examined to identify the dominant factors influencing customer satisfaction, and these findings are synthesized into practical and managerial implications.

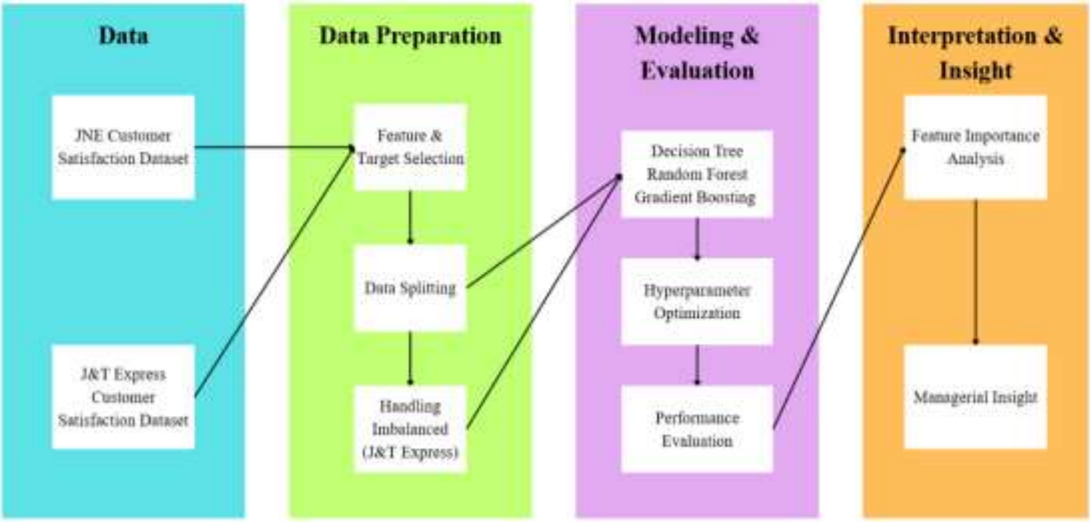

Figure 1: Research Framework

Following the research framework illustrated in Figure 1, the study is implemented through a series of systematic procedures. Initially, problem identification and a review of relevant literature are conducted to establish the theoretical foundation of the study. Subsequently, validated datasets are acquired and examined. During the preprocessing stage, data cleaning is conducted to eliminate missing or duplicate observations. The selected satisfaction indicator is transformed into a binary classification variable. Feature and target separation is performed, followed by an 80:20 data splitting for training and testing the models. For the datasets exhibiting class imbalance, resampling techniques are applied. Model development is conducted independently for three tree-based machine learning algorithms: Decision Tree, Random Forest, and Gradient Boosting. Hyperparameter tuning is performed using Randomized Search Cross-Validation to ensure optimal model configuration. Each optimized model is evaluated on the testing dataset using accuracy, precision, recall, and F1-score. Finally, the best-performing model for each dataset is selected. An aggregated feature importance analysis is then conducted to identify the primary drivers of customer satisfaction. The study concludes by translating analytical findings into managerial insights.

Data Source and Dataset Characteristics

This study utilizes a dataset from Rahadian (2022), which originally aimed to model customer satisfaction classification for JNE and J&T Express regular services. The original data collection is conducted via online questionnaires using accidental sampling, resulting in 408 eligible respondent observations. All features within this dataset are measured using a 5-point Likert scale, ranging from 1 for “Strongly Dissatisfied” to 5 for "Strongly Satisfied”. The dataset is formulated into six main variables, comprising five independent predictor variables based on service quality and one dependent target variable representing service satisfaction status.Given that this dataset has already undergone instrument validation, the restructured dataset is suitable for advanced re-analysis. Utilizing this optimized dataset enables the current study to provide a more comprehensive assessment of model stability and generalization capability by comparing three distinct tree-based algorithms.The current study strategically diverges from the original data aggregation approach. While the previous research aggregated the data by calculating the average scores for all five service dimensions and the customer satisfaction variable, this study utilizes all individual indicators across the five dimensions as distinct predictive features. This item-level granularity allows the models to capture specific, nuanced patterns that might be lost through averaging.

Research Variables

The research variables in this study are adopted from Rahadian (2022). In the original study, customer satisfaction is conceptualized as the dependent variable measured using three indicators. Meanwhile, the independent variables consist of five service quality dimensions, namely Warranty, Friendliness, Delivery Accuracy, Responsiveness, and Shipping Cost, each represented by several observable indicators reflecting specific aspects of logistics service performance. While the conceptual measurement structure remains consistent with the original study, this research introduces an analytical adjustment at the modeling stage. Supervised machine learning classification requires a single, clearly defined target variable to ensure unambiguous optimization and stable model training. Therefore, instead of averaging the three satisfaction indicators into a composite score, this study selects a single representative indicator as the classification target. To ensure that this selection is methodologically grounded, a Mutual Information (MI) analysis is conducted to examine the informational dependency between service quality attributes and each satisfaction indicator. As presented in Table 1, the second indicator consistently demonstrates the highest mean Mutual Information score across both datasets, with 0.071 for JNE and 0.037 for J&T Express. In addition, this indicator also exhibits the largest number of informative features, with 12 informative predictors for JNE and 9 for J&T Express, indicating that more service quality attributes contribute meaningful information to the prediction of this indicator. These results suggest that the second indicator maintains stronger and more distributed informational relationships with the predictor variables compared to the other indicators. Consequently, this indicator is selected as the most suitable target variable for the supervised classification model.

Table 1. Mutual Information Analysis

Dataset	Target Indicator	Highest MI Score	Mean MI Score	Informative Feature (>0)	Selected as Target
JNE	1	0.142	0.052	11	No
	2	0.126	0.071	12	Yes
	3	0.093	0.043	11	No
J&T Express	1	0.064	0.028	8	No
	2	0.085	0.037	9	Yes
	3	0.035	0.016	7	No

Data Preprocessing

Data Preprocessing is a crucial stage to ensure the dataset is suitable for algorithmic processing. The first stage begins with data cleaning to ensure the absence of missing values or duplicated data within the observations. Once the data is confirmed clean, a transformation is performed on the target variable. Because the initial target variable is measured using a 5-point Likert scale, the data is transformed into a binary classification format to simplify predictive modeling. The response values are converted into two main classes (1 for “Satisfied” and 0 for "Dissatisfied”) based on a predefined threshold. The detailed target labeling rule is presented in Table 2.

Table 2. Target Labeling Rule

Original Value	New Binary Class	Numerical Representation
1 (Very Dissatisfied), 2 (Dissatisfied), 3 (Neutral)	Dissatisfied	0
4 (Satisfied), 5 (Very Satisfied)	Satisfied	1

In this study, the neutral midpoint response on the 5-point Likert scale is categorized as “Dissatisfied”. This decision is grounded in recent findings by Fiorentino (2025), who highlights the existence of a “zone of indifference”, where neutral satisfaction levels do not generate meaningful loyalty or positive behavioral outcomes. Customers in this zone exhibit behavioral tendencies that closely resemble those of dissatisfied customers, including higher switching vulnerability. Therefore, classifying neutral responses as “Dissatisfied” ensures that the binary labeling more accurately reflects actual defection risk and behavioral reality. Following the completion of the target labeling, the dataset is divided into two main components. All individual indicators from the five service quality dimensions are allocated as the feature matrix (X). Meanwhile, the binarized customer satisfaction indicator is isolated as the target vector (y). To evaluate the generalization capability of the predictive models on unseen data, the dataset is then partitioned using the Train-Test Split Method. The splitting proportion details for each partition are outlined in Table 3.

Table 3. Data Splitting Proportion

Partition	Function	Quantity	Proportion
Training Set	Model training	326	80%
Testing Set	Model performance evaluation	82	20%

This proportion is selected to balance model learning capacity and generalization evaluation. Considering the relatively limited sample size of the study, allocating a larger portion of data to the training set is essential to ensure that the model has sufficient observations to effectively capture underlying patterns and relationships within the dataset. At the same time, reserving 20% of the data for testing set provides an adequate and unbiased evaluation of model performance on unseen data. Following the data splitting process, the class distribution is observed for each logistics provider. The dataset for JNE exhibits a naturally balanced distribution between the “Satisfied” and “Dissatisfied” categories, therefore, no resampling intervention is required for the JNE predictive models. Conversely, the J&T Express dataset demonstrated a slight class imbalance, where the number of “Satisfied” customers dominated the observations. To prevent the algorithm from developing a predictive bias toward the majority class when training on J&T Express data, this study applied the Synthetic Minority Over-sampling Technique. Crucially, the application of SMOTE is performed exclusively on the J&T training set. This strategic isolation strictly prevents data leakage and ensures that the testing set retains its original, real-world distribution for a valid and unbiased final performance evaluation. The class distribution changes before and after the application of SMOTE on the J&T training set are illustrated in Figure 2.

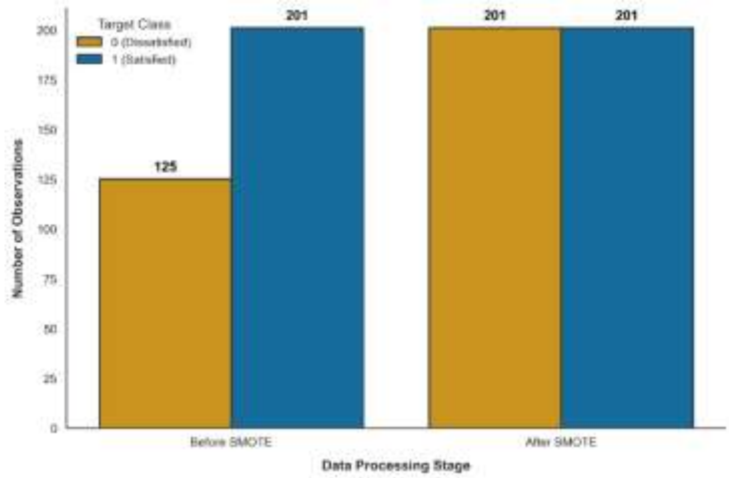

Figure 2: Class Distribution Changes on J&T Training Set

Prior to resampling, the dataset exhibits class imbalance, with 201 observations in the “Satisfied” class and 125 observations in the "Dissatisfied” class. After the application of SMOTE, the number of “Dissatisfied” observations increased from 125 to 201, resulting in a perfectly balanced distribution where both classes contain 201 observations. This means that 76 synthetic minority samples are generated to balance the class distribution, enabling the machine learning models to learn more representative patterns for both classes and reducing the risk of biased prediction.

Hyperparameter Tuning and Model Development

Following the data preprocessing stage, the balanced dataset is utilized to train three distinct tree-based machine learning algorithms. To ensure optimal predictive performance and mitigate the risk of overfitting, hyperparameter tuning is rigorously conducted for each model rather than relying on default algorithm configurations. The hyperparameter optimization process is executed using Randomized Search Cross-Validation (RandomizedSearchCV). Previous studies have shown that RandomizedSearchCV can efficiently explore large hyperparameter spaces and often identify competitive configurations faster than exhaustive grid search, especially when only a subset of hyperparameters significantly influences model performance (Subaşı, 2024). In this study, the search process is conducted over 500 randomized iterations, allowing extensive exploration of the hyperparameter space while maintaining computational feasibility. Each sampled configuration is evaluated using 5-fold cross-validation on the training dataset. By iteratively training and validating the models on different subsets of the data, this technique ensures that the selected hyperparameters yield robust and generalizable performance. For the single Decision Tree model, the tuning process focused on optimizing the split quality measure *‘criterion’* alongside key structural hyperparameters to control tree complexity and prevent overfitting. These structural hyperparameters included the maximum tree depth *‘max_depth’*, the minimum number of samples required to split an internal node *‘min_samples_split’*, and the minimum number of samples required to form a leaf node *‘min_samples_leaf’*. For the ensemble models, the optimization space is appropriately expanded. In the Random Forest algorithm, the number of independent trees *‘n_estimators’* is tuned to maximize the benefits of bootstrap aggregating. Additionally, individual tree constraints *‘max_depth’*, *‘min_samples_leaf’*, and the maximum number of features considered for the best split *‘max_features’* are optimized to ensure sufficient model diversity and limit individual tree variance. Similarly, for the Gradient Boosting Tree, optimization focused on the sequential ensemble configuration. This included the total number of boosting stages *'n_estimators'*, tree complexity boundaries *'max_depth'* and *'min_samples_leaf'*, and crucially, the learning rate *'learning_rate'*, which dictates the step size at each iteration while the model sequentially minimizes residual errors. The hyperparameter combination that produced the highest cross-validated scoring metric for each respective algorithm is dynamically selected as the final optimized model. These optimal values are empirically derived from the Randomized Search procedure conducted exclusively on the training dataset, rather than predetermined from prior studies, ensuring that the model configuration remains fully data-driven. The final optimal hyperparameters identified for each algorithm are detailed in Table 4.

Table 4. Optimal Hyperparameters

Algorithm	Hyperparameter	Optimal Value (JNE)	Optimal Value (J&T)
Decision Tree	'criterion'	'entropy'	'gini'
	'max_depth'	4	7
	'min_samples_leaf'	33	10
	'min_samples_split'	55	35
Random Forest	'n_estimators'	173	97
	'max_depth'	5	21
	'min_samples_leaf'	20	2
	'max_features'	'sqrt'	'log2'
Gradient Boosting	'n_estimators'	243	182
	'max_depth'	2	3
	'min_samples_leaf'	27	35
	'learning_rate'	0.036	0.097

Consequently, the predictive models are explicitly constructed and developed utilizing these exact optimal hyperparameters obtained from the RandomizedSearchCV. These refined models are established as the definitive algorithms to be deployed for the subsequent predictive performance evaluation on the unseen testing set.

RESULTS AND DISCUSSION

Predictive Performance Evaluation

Following the hyperparameter tuning and model development phases, the predictive capabilities of the optimized models are evaluated using the unseen 20% testing set. To provide a comprehensive assessment of model robustness, performance is measured across four evaluation metrics: accuracy, precision, recall, and the F1-Score. The comparative performance results for both the JNE and J&T datasets are systematically presented in Table 5 and Table 6. For the JNE dataset, the single Decision Tree model achieved the highest performance with an accuracy of 78.05%, precision of 79.16%, recall of 78.05%, and an F1-score of 77.84%. These values slightly exceed those of the Random Forest and Gradient Boosting models, both of which recorded identical accuracy scores of 76.83%, with marginally lower precision and F1-scores. The performance difference of 1.22% in accuracy indicates only a modest margin of superiority. Although ensemble methods are generally expected to outperform single-tree models, the superior performance of the Decision Tree in this case suggests that the JNE dataset exhibits relatively clear and well-defined classification boundaries. The naturally balanced class distribution may have allowed the simpler tree structure to effectively capture the primary decision rules without requiring additional complexity from bagging or boosting mechanisms.

Table 5. Performance Comparison for JNE Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	78.05	79.16	78.05	77.84
Random Forest	76.83	76.85	76.83	76.83
Gradient Boosting	76.83	77.23	76.83	76.74

In contrast, the J&T Express dataset exhibits a clear preference for ensemble learning, specifically the Gradient Boosting algorithm, which achieved the highest overall performance with an accuracy of 81.71%, precision of 81.69%, recall of 81.71%, and an F1-score of 81.37%. This performance slightly surpasses the Random Forest model, which obtained an accuracy of 80.49% and an F1-score of 80.21%, while the Decision Tree model recorded considerably lower performance with an accuracy of 70.73%. The performance gap of more than 10% in accuracy between Gradient Boosting and Decision Tree highlights the increased dataset complexity in the J&T case. The superior performance of Gradient Boosting can be attributed to its sequential learning mechanism, which iteratively corrects previous classification errors. Given that the J&T Express dataset initially exhibits class imbalance and required SMOTE augmentation, the boosted ensemble appears better suited to capture the more complex decision boundaries introduced by synthetic minority samples. To further validate these findings, the Confusion Matrices for the best performing model of each dataset are presented in Figure 3.

Table 6. Performance Comparison for J&T Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	70.73	75.41	70.73	70.98
Random Forest	80.49	80.35	80.49	80.21
Gradient Boosting	81.71	81.69	81.71	81.37

The confusion matrix for the JNE dataset model indicates a relatively balanced classification performance. The model correctly identified 36 dissatisfied customers (True Negatives) and 28 satisfied customers (True Positives). While only 5 False Positives are generated, suggesting that the model may rarely misclassify dissatisfied customers as satisfied, a higher number of False Negatives (13 cases) is observed. From a managerial perspective, the relatively low False Positive rate is beneficial, as it reduces the risk of overlooking genuinely dissatisfied customers who require immediate service recovery interventions.

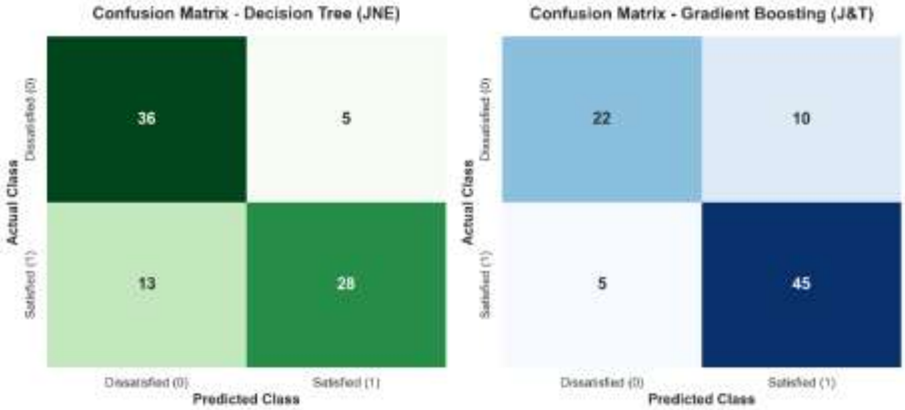

Figure 3: Confusion Matrix for JNE and J&T Best Model

For the J&T Express dataset, the Gradient Boosting model demonstrated strong robustness in handling the real-world imbalanced test distribution. The model correctly identified 22 out of 32 dissatisfied customers. Although 10 False Positives are recorded, the model maintained a low number of False Negatives (5 cases), indicating effective recognition of satisfied customers. These results suggest that the SMOTE-augmented Gradient Boosting model successfully learned representative patterns of the minority class without substantially compromising performance on the majority class.

Feature Importance Analysis

To translate the predictive algorithm into actionable managerial intelligence, an aggregated feature importance analysis is conducted. The individual feature importance scores extracted from the models are grouped based on their theoretical service quality dimensions to identify the primary drivers of customer satisfaction. Feature Importance for the JNE and J&T models is shown in Figure 4 and Figure 5. For the JNE dataset, the Decision Tree model demonstrates a highly concentrated importance distribution. Shipping Cost overwhelmingly dominates the classification process with an aggregated importance score of 0.733, substantially exceeding other service dimensions such as Delivery Accuracy and Responsiveness. Notably, the dimensions of Friendliness and Warranty received an importance score of zero in the single-tree model. This outcome does not imply that these attributes are irrelevant in practice, rather, it reflects the greedy splitting mechanism inherent in Decision Trees, where only the most informative predictors are selected to construct hierarchical splits. Once Shipping Cost and key operational dimensions sufficiently reduce impurity, additional attributes may not contribute incremental discriminatory power, resulting in zero importance values. The ensemble models for JNE further substantiate this pattern while offering a more distributed contribution across dimensions. In both models, Shipping Cost remains the dominant predictor, confirming its structural importance in determining customer satisfaction. However, unlike the single Decision Tree, all other service dimensions retain non-zero importance scores, suggesting that ensemble learning mechanisms can capture secondary patterns that a single tree may overlook. This cross-model consistency strengthens the robustness of the conclusion that price perception is the primary driver of customer satisfaction in the JNE context.

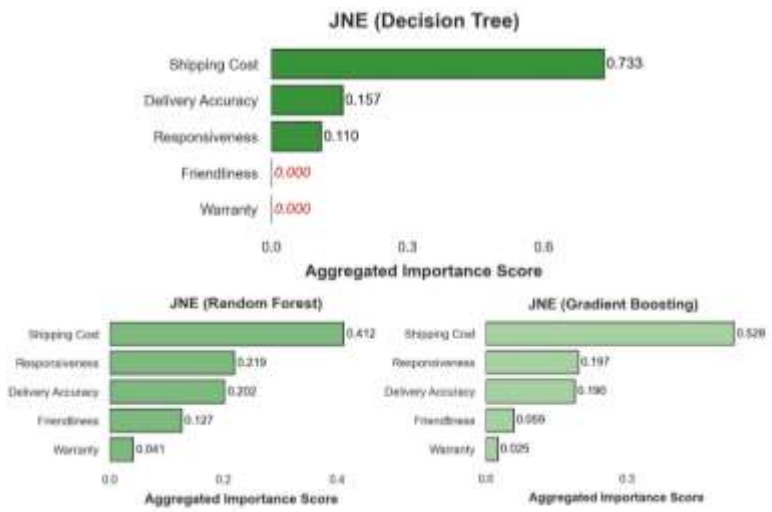

Figure 4: JNE Models Feature Importance

In contrast, the J&T dataset exhibits a more balanced importance distribution, particularly in the best-performing Gradient Boosting model. Although Shipping Cost remains the most influential factor, the relative contributions of Responsiveness, Delivery Accuracy, Friendliness, and Warranty are more evenly distributed. This indicates that customer satisfaction in J&T Express is influenced by a broader combination of service quality dimensions rather than being predominantly price driven.

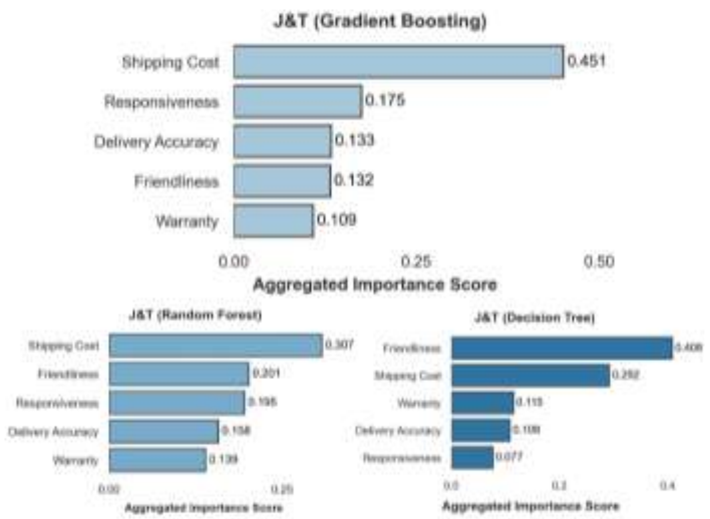

Figure 5: J&T Models Feature Importance

From a managerial perspective, these findings imply that JNE customers appear highly price-sensitive, with satisfaction decisions strongly anchored in cost-related evaluations. Conversely, J&T Express customers assess satisfaction using a more holistic evaluation framework, where operational performance and interpersonal service quality play substantial roles alongside pricing considerations. The consistency of Shipping Cost as the dominant factor across all models and datasets also enhances the external validity of the findings, as it demonstrates stability across different algorithmic structures. The dominance of Shipping Cost in the present model is consistent with the findings of Do et al. (2023), who empirically demonstrate that price fairness constitutes a significant determinant of customer satisfaction within logistics service quality frameworks.

Overall, the feature importance results not only validate the predictive dominance of specific service dimensions but also provide strategic insights into how customer satisfaction drivers differ between logistics providers. The alignment between the best-performing models and their importance structures reinforces the credibility of the analytical framework and supports the reliability of the study's conclusion.

CONCLUSION

This study demonstrates that model performance is strongly influenced by dataset characteristics. For the naturally balanced JNE dataset, the Decision Tree model achieved the highest predictive performance with an accuracy of 78.05% and an F1-score of 77.84%, outperforming the ensemble-based alternatives. In contrast, for the initially imbalanced J&T Express dataset, the Gradient Boosting model delivered superior results with an accuracy of 81.71% and an F1-score of 81.37%, confirming its effectiveness in handling more complex classification boundaries following SMOTE resampling. Feature importance analysis further shows that Shipping Cost consistently emerges as the dominant predictor across models and datasets. However, JNE satisfaction appears predominantly price-driven, whereas J&T Express satisfaction reflects a more multidimensional structure involving Responsiveness, Delivery Accuracy, Friendliness, and Warranty. These findings confirm the effectiveness of tree-based machine learning for modeling non-linear satisfaction patterns while providing actionable managerial insights. For future research, several extensions are recommended. First, incorporating larger and more diverse datasets from additional logistics providers would improve model generalizability. Second, future studies may explore advanced ensemble or deep learning architectures to assess potential performance improvements. Third, integrating behavioral data such as repurchase behavior or complaint records could enhance predictive depth beyond perception-based survey measures. Such developments would further strengthen the analytical robustness and practical applicability of machine learning in service quality research.

AI Declaration

The authors declare that AI-assisted tools including ChatGPT, Gemini, and grammar-checking software are utilized solely for language refinement, structural editing, and clarity enhancement during the manuscript preparation process. All research design decisions, data preprocessing procedures, model development, coding implementation, statistical evaluation, result interpretation, and analytical conclusions are independently carried out by the authors. The AI tools are not used to generate experimental results, conduct data analysis, or formulate scientific findings.

REFERENCES

Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, 6(1), Article 28. https://doi.org/10.1186/s40537-019-0191-6

Akbari, P., Zamani, M., & Mostafaei, A. (2024). Machine learning prediction of mechanical properties in metal additive manufacturing. *Additive Manufacturing*, 91, Article 104320. https://doi.org/10.1016/j.addma.2024.104320

Albostami, A. S., Al-Hamd, R. K. S., & Al-Matwari, A. A. (2024). Data-Driven Predictive Modeling of Steel Slag Concrete Strength for Sustainable Construction. *Buildings*, 14(8), Article 2476. https://doi.org/10.3390/buildings14082476

Álvarez-García, J., González-Vázquez, E., Del Río-Rama, M. de la C., & Durán-Sánchez, A. (2019). Quality in customer service and its relationship with satisfaction: An innovation and competitiveness tool in sport and health centers. *International Journal of Environmental Research and Public Health*, 16(20), Article 3942. https://doi.org/10.3390/ijerph16203942

An, S., Lee, P., & Shin, C. H. (2023). Effects of Servicescapes on Interaction Quality, Service Quality, and Behavioral Intention in a Healthcare Setting. *Healthcare (Switzerland)*, *11*(18), 1–16. https://doi.org/10.3390/healthcare11182498

Bank Indonesia. (2025). *Sinergi jadi energi: Saat UMKM naik kelas lewat digitalisasi*. https://www.bi.go.id/id/publikasi/ruang-media/cerita-bi/Pages/Sinergi-Jadi-Energi-Saat-UMKM-Naik-Kelas-Lewat-Digitalisasi.aspx

Barri, E. Y., Farber, S., Jahanshahi, H., & Beyazit, E. (2022). Understanding transit ridership in an equity context through a comparison of statistical and machine learning algorithms. *Journal of Transport Geography, 105*, Article 103482. https://doi.org/10.1016/j.jtrangeo.2022.103482

Bhaskar, K., & Subramanyam, P. (2021). Customer satisfaction on service quality dimensions. *Journal of Management and Science*, *11*(3), 35–39. https://doi.org/10.26524/jms.11.21

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16,* 321–357. https://doi.org/10.1613/jair.953

Dhillon, S. K., Ganggayah, M. D., Sinnadurai, S., Lio, P., & Taib, N. A. (2022). Theory and Practice of Integrating Machine Learning and Conventional Statistics in Medical Data Analysis. *Diagnostics*, 12(10), Article 2526. https://doi.org/10.3390/diagnostics12102526

Do, Q. H., Kim, T. Y., & Wang, X. (2023). Effects of logistics service quality and price fairness on customer repurchase intention: The moderating role of cross-border e-commerce experiences. *Journal of Retailing and Consumer Services*, *70*, Article 103165. https://doi.org/10.1016/j.jretconser.2022.103165

Dukalang, H. H., Otok, B. W., & Purhadi. (2025). Modified partial least square structural equation model with multivariate adaptive regression spline: Parameter estimation technique and applications. *MethodsX*, *14*, Article 103381. https://doi.org/10.1016/j.mex.2025.103381

Fiorentino, A. (2025). Beyond boundaries: a multidisciplinary approach to understand the relationship between customer satisfaction and behavior in services. *Italian Journal of Marketing*, 2025(1), 51–69. https://doi.org/10.1007/s43039-025-00108-y

Gam, A., Bahri-Ammari, N., & Soliman, M. (2023). Intelligent Customer Experience and Behavioral Brand Loyalty: Empirical Evidence from the Commerce Environment. *International Journal of Human-Computer Interaction*, *39*(10), 2090–2101. https://doi.org/10.1080/10447318.2022.2074120

Guo, L., Burke, M. G., & Griggs, W. M. (2025). A new framework to predict and visualize technology acceptance: A case study of shared autonomous vehicles. *Technological Forecasting and Social Change*, *212*, Article 123960. https://doi.org/10.1016/j.techfore.2024.123960

Hui, G., Al Mamun, A., Reza, M. N. H., & Hussain, W. M. H. W. (2025). An empirical study on logistic service quality, customer satisfaction, and cross-border repurchase intention. *Heliyon*, *11*(1), e41156. https://doi.org/10.1016/j.heliyon.2024.e41156

Kamm, S., Veekati, S. S., Müller, T., Jazdi, N., & Weyrich, M. (2023). A survey on machine learning based analysis of heterogeneous data in industrial automation. *Computers in Industry*, *149*, Article 103930. https://doi.org/10.1016/j.compind.2023.103930

Khan, A. A., Chaudhari, O., & Chandra, R. (2024). A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation. *Expert Systems with Applications*, *244*, Article 122778. https://doi.org/10.1016/j.eswa.2023.122778

Kullolli, T., Trebicka, B., & Fortuzi, S. (2024). Understanding customer satisfaction factors: A logistic regression analysis. *Journal of Educational and Social Research*, *14*(2), 218–231. https://doi.org/10.36941/jesr-2024-0038

Li, X., Xu, M., Zeng, W., Tse, Y. K., & Chan, H. K. (2023). Exploring customer concerns on service quality under the COVID-19 crisis: A social media analytics study from the retail industry. *Journal of Retailing and Consumer Services*, *70*, Article 103157. https://doi.org/10.1016/j.jretconser.2022.103157

Liu, J. (2024). Influence of service quality and customer perceived value on customer loyalty with customer satisfaction as a moderating factor: A study based on private elderly care services in China. *Journal of Infrastructure, Policy and Development*, *8*(6), 1–19. https://doi.org/10.24294/jipd.v8i6.6316

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Madani, B., & Alshraideh, H. (2021). Predicting consumer purchasing decisions in the online food delivery industry. *Computer Science & Information Technology (CS & IT)*, *11*(15), 103–117. https://doi.org/10.5121/csit.2021.111510

Masyhuri, M. (2022). Key Drivers of Customer Satisfaction on the E-Commerce Business. *East Asian Journal of Multidisciplinary Research*, *1*(4), 657–670. https://doi.org/10.55927/eajmr.v1i4.405

Mienye, I. D., & Jere, N. (2024). A survey of decision trees: Concepts, algorithms, and applications. *IEEE Access*, *12*, 86716–86727. https://doi.org/10.1109/ACCESS.2024.3416838

Nuralam, I. P., Yudiono, N., Fahmi, M. R. A., Yuliaji, E. S., & Hidayat, T. (2024). Perceived ease of use, perceived usefulness, and customer satisfaction as driving factors on repurchase intention: The perspective of the e-commerce market in Indonesia. *Cogent Business & Management*, *11*(1), Article 2413376. https://doi.org/10.1080/23311975.2024.2413376

Pranida, S. Z., & Kurniawardhani, A. (2022). Sentiment Analysis of Expedition Customer Satisfaction using BiGRU and BiLSTM. *Indonesian Journal of Artificial Intelligence and Data Mining*, *5*(1), 44. https://doi.org/10.24014/ijaidm.v5i1.17361

Prasetyo, D., Sugiharto, Daengs, A., Istanti, E., & Karnain, B. (2023). Service quality relationship with customer satisfaction in PDAM. *International Journal of Economics and Management Research*, *2*(1), 88–102. https://ijemr.politeknikpratama.ac.id/index.php/ijemr

Rahadian, R. E. (2022). *Model klasifikasi kepuasan pelanggan pada layanan reguler jasa ekspedisi dengan menggunakan metode decision tree (studi kasus pada JNE dan J&T)* [Unpublished bachelor's thesis]. Universitas Negeri Malang.

Razavi, S. (2021). Deep learning, explained: Fundamentals, explainability, and bridgeability to process-based modelling. *Environmental Modelling & Software*, *144*, Article 105159. https://doi.org/10.1016/j.envsoft.2021.105159

Subaşı, N. (2024). Comprehensive Analysis of Grid and Randomized Search on Dataset Performance. *European Journal of Engineering and Applied Sciences*, *7*(2), 77–83. https://doi.org/10.55581/ejeas.1581494

Syihabudin, Kurnia, M. R., & Najmudin. (2022). Impact of The Covid-19 Pandemic on Financial Performance of Logistics and Delivery Companies Listed on ISSI. *Ekonomi, Keuangan, Investasi Dan Syariah (EKUITAS)*, *3*(3), 461–467. https://doi.org/10.47065/ekuitas.v3i3.1197

Tjendana, Y., & Pranitasari, D. (2024). *Determinants Of SERVQUAL Dimensions on Customer Satisfaction and Loyalty in Maritime Transportation and Logistics Services*. *9*(3), 227–240.

Top Brand Award. (2025). *Top Brand Index*. https://www.topbrand-award.com/top-brand-index/

Uddin, S., & Lu, H. (2024). Confirming the statistically significant superiority of tree-based machine learning algorithms over their counterparts for tabular data. *PLOS ONE*, *19*(4), Article e0301541. https://doi.org/10.1371/journal.pone.0301541

Yahya, A. A. A., Pratama, R. C., & Farahdiansari, A. P. (2023). Analysis of Service Quality to Customer Satisfaction with CSI and Servqual Methods. *Journal of Information System, Technology and Engineering*, *1*(4), 104–119. http://publications.lib.chalmers.se/records/fulltext/245180/245180.pdf%0Ahttps://hdl.handle.net/20.500.12380/245180%0Ahttp://dx.doi.org/10.1016/j.jsames.2011.03.003%0Ahttps://doi.org/10.1016/j.gr.2017.08.001%0Ahttp://dx.doi.org/10.1016/j.precamres.2014.12