Received 25 April 2025, accepted 02 August 2026, date of publication 30 September 2026
www.computingonline.net / computing@computingonline.net

Print ISSN 1727-6209
Online ISSN 2312-5381
DOI 10.47839/ijc.25.3.4855

Enhancing Fraud Detection in Digital Transactions through a Modified CART Algorithm

ASRIANDA[1], NAUFAL BACHRI[2], SUJACKA RETNO[1]
[1]Department of Informatics, Universitas Malikussaleh, 24355, Indonesia
[2]Department of Management Science, Universitas Malikussaleh, 24355, Indonesia

Corresponding author: Asrianda (e-mail: asrianda@unimal.ac.id).

ABSTRACT Online payment fraud has become a major challenge in the banking and e-commerce sectors due to the rapid growth of digital transactions and increasingly sophisticated cybercrime techniques. Although the Classification and Regression Trees (CART) algorithm has been widely applied to fraud detection, its performance may deteriorate when dealing with imbalanced datasets. This study proposes CART-SQ (Classification and Regression Trees Simple Quadrat), which applies an equivalent quadratic formulation of the Gini impurity during attribute split evaluation while preserving the mathematical properties of the conventional CART algorithm. The proposed method was evaluated using an imbalanced financial transaction dataset and compared with the conventional CART and C4.5 algorithms under two train–test partition scenarios. The experimental results demonstrate that CART-SQ achieved the best overall classification performance, with accuracies of 98.25% and 98.20%, while also attaining the highest precision and F1-score for the fraud class. These findings indicate that the proposed equivalent quadratic formulation provides an effective alternative for improving decision tree performance in fraud detection tasks involving imbalanced datasets.

KEYWORDS fraud detection; online payment; fraud; CART; CART-SQ; C4.5; imbalanced data.

I. INTRODUCTION

Online payment fraud is one of the major challenges faced by the banking and e-commerce sectors, as cybercriminals continuously develop more sophisticated methods [1, 2]. With the rapid advancement of financial technology and e-commerce, fraud techniques have become increasingly varied and difficult to detect. Online payment fraud has become a primary concern, significantly impacting the economy and causing substantial losses to involved parties [3]. In recent years, cybercriminals have innovated with new techniques that are more difficult to trace and detect. Therefore, effective preventive measures are essential to minimize further losses, especially with the growing reliance on online transactions. Fraud detection methods must evolve to keep up with the continuously changing tactics used by fraudsters. Machine learning algorithms, such as Classification and Regression Trees (CART), have made notable contributions in identifying fraudulent transactions. CART has the potential to predict and detect suspicious activities within digital transaction data [4]. The performance of the CART algorithm in fraud detection is largely determined by the impurity evaluation criterion used to select split nodes. Therefore, enhancing the impurity evaluation mechanism represents a promising approach to improving the discriminative capability of CART for fraud detection.

A widely used approach for fraud detection is the Classification and Regression Trees (CART) algorithm. CART is recognized for its simplicity, interpretability, and effectiveness in classification tasks. Although CART has been widely applied, challenges remain, particularly when dealing with imbalanced datasets [5] and noisy data. While CART can produce accurate predictions, its performance is often hindered by data imbalance [6], where the abundance of legitimate transactions causes the model to focus primarily on the majority class and neglect the minority fraud class [7, 8]. This imbalance reduces the effectiveness of fraud detection. Additionally, many datasets used for training are of poor quality, suffer from incomplete information, or fail to capture rapidly shifting fraud trends [9]. Therefore, improving the attribute selection mechanism becomes important to enhance classification performance on imbalanced datasets. Poor data quality, incompleteness, and the presence of noise [10] remain significant obstacles in building effective fraud detection systems.

Previous studies have primarily focused on improving classification performance through preprocessing techniques to address the issue of data imbalance [11]. These approaches generally retain the original Gini-based splitting criterion, which may limit the ability to discriminate informative attributes in highly imbalanced datasets. The modus operandi

of fraudsters is rapidly evolving [12], making approaches that rely on fixed patterns less effective in the long term [13]. Although some models may initially succeed in detecting fraudulent activities, cybercriminals tend to adapt and exploit new vulnerabilities to bypass detection systems. Therefore, fraud detection algorithms should incorporate a robust attribute selection mechanism capable of preserving classification performance in datasets exhibiting complex characteristics and highly imbalanced class distributions.

Although the CART algorithm has been widely adopted for classification tasks because of its simplicity and interpretability, several limitations restrict its effectiveness in fraud detection. One of its main weaknesses is its sensitivity to minor variations in the training data, which may generate substantially different tree structures and reduce the stability of the resulting model [14]. This behavior often leads to inconsistent prediction outcomes, particularly when the dataset contains noisy observations [15, 16]. Another limitation arises from the conventional Gini Index, which serves as the impurity criterion during node splitting. In datasets characterized by complex data patterns and imbalanced class distributions, the conventional criterion may not adequately distinguish among alternative split candidates, thereby limiting the quality of the resulting decision tree [17, 18]. These limitations highlight the need for a more discriminative impurity evaluation strategy to strengthen the robustness of the CART algorithm in fraud classification.

To address this issue, the present study introduces a Quadratic Gini formulation as a replacement for the conventional Gini Index in the CART node-splitting process. The proposed impurity criterion is designed to improve the evaluation of candidate splits, enabling the tree construction procedure to identify more informative partitions while preserving the interpretability of the CART model. By enhancing the impurity evaluation mechanism, the proposed approach is expected to provide more reliable classification performance for fraud detection, particularly when applied to datasets containing complex patterns, noisy observations, and imbalanced class distributions [19].

Building on the limitations identified in previous studies, this research presents a modified Classification and Regression Trees (CART) algorithm that incorporates a Quadratic Gini formulation into the node-splitting process. Rather than relying on the conventional Gini Index, the proposed impurity criterion is designed to provide a more discriminative evaluation of candidate split attributes, leading to more informative partitioning during tree construction. This modification is intended to improve the robustness of the resulting decision tree while maintaining reliable classification performance when confronted with complex data structures, imbalanced class distributions, and noisy observations. Through this refinement, the study seeks to advance the application of decision tree methods for fraud detection in digital payment systems by introducing a more effective impurity evaluation mechanism.

II. RELATED WORK

Fraud detection research has progressively evolved from relying on single classification algorithms toward more adaptive approaches capable of addressing the increasing complexity of digital transactions. Modern fraud detection systems are expected not only to achieve high classification accuracy but also to identify continuously evolving fraudulent behaviors while maintaining reliable performance under dynamic operating conditions [20, 21]. In addition, practical deployment requires models that preserve interpretability and support real-time decision-making, particularly in financial environments where rapid responses are essential. Supervised learning has become the predominant paradigm for enhancing fraud detection by improving the identification of diverse and evolving fraud characteristics across digital transaction datasets [22]. To further improve predictive performance, recent studies have increasingly adopted ensemble learning strategies that combine multiple classifiers to enhance robustness and generalization. Nevertheless, comparatively little attention has been devoted to improving the attribute evaluation mechanism within decision tree algorithms, despite its fundamental role in determining tree structure, split quality, and the model's ability to distinguish fraudulent transactions from legitimate ones.

The increasing complexity of machine learning models has made interpretability a critical requirement for fraud detection systems in the financial sector, where transparency is essential for supporting reliable and accountable decision-making processes [23]. To preserve transparency without compromising classification performance, inherently interpretable models remain an attractive solution. Among these, the Classification and Regression Tree (CART) algorithm has been widely adopted because it produces an intuitive decision tree structure that facilitates model interpretation, auditing, validation, and regulatory compliance. In the CART algorithm, the Gini index is employed as the attribute evaluation criterion to determine the optimal split at each decision node [24]. Although the Gini index is computationally efficient and mathematically straightforward, its attribute evaluation mechanism focuses on minimizing local node impurity during the recursive partitioning process. Consequently, the resulting tree structure may not be sufficiently discriminative when applied to fraud datasets characterized by complex patterns, noisy observations, and severe class imbalance [25]. These limitations suggest that improving the attribute evaluation criterion within the CART algorithm represents a promising research direction for enhancing the model's discriminative capability while preserving the interpretability that is essential for financial fraud detection applications.

Numerous studies have sought to improve the performance of the CART algorithm through a variety of approaches, including ensemble learning [26], class imbalance handling techniques [27], tree pruning [28] and feature engineering [29]. Although these approaches have demonstrated substantial improvements in classification accuracy, they primarily optimize components external to the attribute selection mechanism that governs tree construction. In contrast, comparatively little attention has been devoted to enhancing the attribute evaluation function, which constitutes the core of the decision tree induction process. The effectiveness of this evaluation function directly influences the selection of the optimal splitting attribute [30], the resulting tree structure, and the model's ability to distinguish fraudulent transactions from legitimate ones. This observation highlights the need for a more discriminative impurity evaluation criterion to improve the predictive performance of the CART algorithm while preserving its inherent interpretability.

The selection of the optimal attribute at each decision node plays a fundamental role in determining the overall structure and complexity of a decision tree [31]. The effectiveness of the attribute evaluation criterion also influences the model's generalization capability and its classification stability when applied to previously unseen data. An evaluation criterion that provides more discriminative attribute selection is more likely to produce a compact tree structure without compromising classification performance. In financial fraud detection, where transaction patterns evolve continuously and datasets are often characterized by noise and class imbalance, the attribute evaluation mechanism should not only minimize impurity at individual nodes but also improve the overall quality of the partitioning process. These challenges highlight the need for adaptive attribute evaluation functions that can enhance the accuracy, robustness, and generalization capability of the CART algorithm while preserving its inherent interpretability.

III. MATERIAL AND METHODS

The method proposed in this study, referred to as CART Simplified Quadratic (CART-SQ), extends the conventional Classification and Regression Tree (CART) algorithm by incorporating a simplified quadratic Gini formulation based on the squared proportions of class distributions. The proposed formulation is designed to enhance the attribute evaluation process at each decision node by reducing the influence of majority-class dominance during attribute selection in imbalanced datasets. Consequently, the proposed modification is intended to strengthen the discriminative capability of the model, leading to a more representative decision tree structure while preserving the inherent interpretability of the original CART algorithm. This improvement is expected to support a more reliable classification process for datasets characterized by class imbalance.

A. MODIFICATION OF THE GINI INDEX

The simplification of the Gini index [32] is performed using a quadratic calculation [33] based on the class proportions. The simplification process is as follows:

$$Gini = 1 - \sum_{i=1}^{n} p_i^2 \quad (1)$$

In a binary classification case such as fraud detection, there are two classes: Fraud (label = 1) and Non-Fraud (label = 0). Let p represent the proportion of data belonging to the fraud class, then $1-p$ denotes the proportion of the non-fraud class. The Gini impurity at a CART node is calculated using the formula derived from Equation (1):

$$Gini(p) = 1 - (p^2 + (1-p)^2)$$
$$Gini(p) = 1 - \left(p^2 + (p^2 - 2p + 1)\right)$$
$$Gini(p) = 1 - (2p^2 - 2p + 1)$$
$$Gini(p) = -2p^2 + 2p$$
$$Gini(p) = 2p(1-p)$$

The conventional Gini impurity is defined based on the squared proportions of the class distributions. For binary classification, this expression can be reformulated into the algebraically equivalent quadratic form, $2p(1-p)$, without altering either the impurity values or the attribute selection criterion. In the proposed CART-SQ algorithm, this quadratic representation is consistently employed throughout the impurity evaluation process, providing a more compact computational formulation while preserving the theoretical properties and decision-making behavior of the conventional Gini criterion.

The class proportion in a node is denoted by p_i representing the fraction of data in the node that belongs to class i, and n is the number of classes in the node.

1. Simplifying the Quadratic Calculation
   Consider a binary classification with classes A and B. The simplification proceeds as follows:

$$Gini = 1 - (p_A^2 - p_B^2)$$

Given that there are only two classes in the node:

$$p_A + p_B = 1$$

Expressing one proportion in terms of the other:

$$p_B = 1 - p_A$$

Substitute p_B into the Gini formula:

$$Gini = 1 - (p_A^2 + (1 - p_A)^2)$$

Compute the square of $(1 - p_A)$:

$$(1 - p_A)^2 = 1 - 2p_A + p_A^2$$

Now substitute into the Gini formula:

$$Gini = 1 - (p_A^2 + 1 - 2p_A + p_A^2)$$

2. Combine and Simplify

$$Gini = 1 - (2p_A^2 - 2p_A + 1)$$
$$Gini = 1 - 2p_A^2 + 2p_A - 1$$
$$Gini = 2p_A(1 - p_A)$$

Thus, the simplified Gini index for the binary case becomes:

$$Gini = 2p_A(1 - p_A) \quad (2)$$

To derive the Gini index formula, consider two classes—class A and class B—within the left subset. The proportion of class A in the left subset is denoted as $P_{L.A}$, which represents the dominance of class A in that subset. The proportion of class B in the same subset is calculated as $P_{L.B}$ = 1-$P_{L.A}$, since the total proportion in a subset must equal 1. These proportions are essential in evaluating the Gini index: the greater the dominance of a single class, the lower the impurity.

The Gini index for the left subset is given by:

$$Gini_{l;eft} = 1(p_{L.A}^2 + (1 - p_{L.A})^2)$$

Calculating the expression:

$$Gini_{left} = 1 - (p_{L.A}^2 + 1 - 2p_{L.A} + p_{L.A}^2)$$
$$Gini_{left} = 1 - (2p_{L.A}^2 - 2p_{L.A} + 1)$$
$$Gini_{left} = 2p_{L.A}(1 - P_{L.A})$$

Similarly, the Gini index for the right subset is:

$$Gini_{right} = 2p_{R.A}(1 - P_{R.A})$$

These calculations are then used to compute the overall Gini split using the weighted average formula.

Equation (2) shows that, for binary classification, the Gini impurity can be computed using the proportion of a single class because the proportion of the complementary class is determined by $1 - p_A$. This algebraically equivalent quadratic formulation preserves the impurity characteristics of the conventional Gini criterion while providing a compact representation for impurity computation. In the proposed CART-SQ algorithm, the quadratic formulation is consistently applied to evaluate the impurity of the parent node and child nodes before calculating the weighted Gini split and impurity reduction ($\Delta Gini$). As in the conventional Gini criterion, the impurity approaches zero when $p$approaches 0 or 1, indicating that the node is highly homogeneous. Conversely, the impurity reaches its maximum when $p = 0.5$, corresponding to a balanced class distribution within the node [34].

B. CART (CLASSIFICATION AND REGRESSION TREE)

The CART algorithm consists of three main components: feature selection, tree construction, and tree pruning. It applies a binary recursive splitting method to divide the dataset into two subsets [35]. Node splitting in CART is based on the Gini index, which is used to identify the best attribute for dividing the data by minimizing impurity in the resulting subsets. The lower the Gini index after a split, the better the quality of the split, as it indicates more homogeneous groups [37]. The quadratic expression in Equation (2) serves as the basis for impurity evaluation during the subsequent split selection procedure.

In the CART algorithm, the Gini split is calculated to choose the best feature that separates the data into two subsets (left and right) with the lowest impurity. Let N be the total number of data points in the current node, and N_L and N_R be the number of data points in the left and right subsets, respectively. The Gini index for each subset is calculated based on the class proportions within them. This calculation is aimed at determining the purity or homogeneity of the resulting subsets. The optimal split is the one that minimizes the Gini index for both child nodes.

1. Calculating Gini Split

$$Gini_{left} = 2p_{L.A}(1 - P_{L.A})$$
$$Gini_{right} = 2p_{R.A}(1 - P_{R.A})$$

The Gini impurity of the left node ($Gini_{left}$) and the right node ($Gini_{right}$) measures the impurity of the data after the split. These values are calculated based on the proportion of samples belonging to a particular class in each node. $P_{L.A}$ and $P_{R.A}$ represent the class proportions in the left and right nodes, respectively, while $1-P_{L.A}$ and $1-P_{R.A}$ represent the proportions not belonging to the class. A dominant class proportion close to 1 results in low impurity, whereas proportions near 0.5 lead to maximum impurity. This calculation helps the CART algorithm choose the optimal split to improve classification accuracy.

2. Calculating the Overall Gini Split

$$Gini_{split} = \frac{N_L}{N}\ x\ Gini_{left} + \frac{N_R}{N}\ x\ Gini_{right}$$

Here, N_L and N_R are the number of data points in the left and right subsets, respectively, and $N=N_L+N_R$ is the total number of samples in the parent node. $Gini_{left}$ and $Gini_{right}$ represent the impurity of the left and right child nodes.

3. Calculating Gini for the Parent Node
The Gini impurity of the parent node ($Gini_{parent}$) is calculated before the data is split. It is based on the class proportions in the original dataset:

$$Gini_{parent} = 2_{PA}(1 - P_A)$$

Here, P_A represents the proportion of samples belonging to the target class in the parent node, and $1-P_A$ represents the remaining proportion. Both values are used to compute the impurity of the parent node and assess its initial purity before splitting.

4. Selecting the Best Split

$$\Delta Gini = Gini_{parent} - Gini_{split} \quad (3)$$

Equation (3) measures the reduction in impurity after splitting the parent node into two child nodes. The value of ΔGini indicates how effectively the split reduces impurity. The higher the value, the better the split. If $Gini_{split}$ is lower than $Gini_{parent}$, it means that the attribute used in the split has improved the data purity. This helps in selecting the most effective attribute for constructing an accurate decision tree. In the CART algorithm, the best split is identified by minimizing $Gini_{split}$; the greater the impurity reduction after the split, the more optimal the separation.

C. PSEUDOCODE

Algorithm 1 outlines the complete workflow of the proposed CART-SQ algorithm. The process begins by loading the training dataset and constructing the root node. Similar to the conventional CART algorithm, CART-SQ preserves the recursive tree induction mechanism while employing the simplified quadratic Gini formulation as the impurity evaluation criterion during attribute selection. For each candidate attribute, the class proportions of the left and right child nodes are calculated to determine the node impurity using Equation (2). The weighted Gini Split and the corresponding impurity reduction (ΔGini) are subsequently computed according to Equation (3). The attribute yielding the highest ΔGini value is selected as the optimal splitting attribute. This procedure is recursively applied to each child node until the stopping criterion is satisfied, after which tree pruning is performed to improve the generalization capability of the resulting decision tree.

Algorithm 1. Proposed CART-SQ Algorithm

Input :
Dataset D ← {(x1,y1), (x2,y2), ..., (xn,yn)}

Output :
Decision Tree T

1. *Load dataset D.*
2. *Create the root node containing all training samples.*
3. *If all samples belong to the same class,*
   create a leaf node and terminate.
4. *Compute the parent node impurity:*
   $GiniParent \leftarrow 2PA(1 - PA)$
5. *For each candidate attribute Ai do*
6. *Split the data into Left and Right subsets.*
7. *Compute the class proportions:*
   $PL,A \leftarrow FraudLeft / NLeft$
   $PR,A \leftarrow FraudRight / NRight$
8. *Compute the child node impurities:*
   $GiniLeft \leftarrow 2PL,A(1 - PL,A)$
   $GiniRight \leftarrow 2PR,A(1 - PR,A)$
9. *Compute the weighted Gini Split:*
   $GiniSplit \leftarrow (NLeft/N) \times GiniLeft + (NRight/N) \times GiniRight$
10. *Compute the impurity reduction:*
    $\Delta Gini \leftarrow GiniParent - GiniSplit$
11. *End For*
12. *Select the attribute with the maximum ΔGini.*
13. *Partition the current node using the selected attribute.*
14. *Recursively repeat Steps 3–13 for each child node.*
15. *Apply tree pruning.*
16. *Return the final CART-SQ decision tree.*

As illustrated in Algorithm 1, the primary distinction between the proposed CART-SQ algorithm and the conventional CART algorithm lies in the impurity evaluation stage during attribute selection. CART-SQ employs an equivalent quadratic formulation of the Gini impurity to evaluate the purity of each candidate node. The remaining procedures, including recursive tree construction, attribute selection based on the maximum impurity reduction (ΔGini), and tree pruning, follow the standard CART framework. Consequently, the proposed approach preserves the fundamental characteristics of the original CART algorithm while introducing a more compact formulation for impurity evaluation during the attribute selection process in decision tree induction.

D. DATASET

This study uses a pre-processed financial transaction dataset that contains information related to customer activity. The dataset is utilized to distinguish between normal and fraudulent transactions. It consists of 3,142 data entries, with 3,020 transactions labeled as non-fraudulent (isFraud = 0) and 122 transactions labeled as fraudulent (isFraud = 1).

Figure 1 illustrates the attributes present in the dataset, which are used to detect fraudulent transactions. The step attribute represents the time in hours since the start of data recording and is used to track the chronology of transactions. The type attribute indicates the type of transaction, such as payment, transfer, cash_out, or others. The amount reflects the transaction value in monetary units. The oldbalanceOrg is the account balance of the sender before the transaction, while newbalanceOrig is the sender's account balance after the transaction is completed. The isFraud attribute serves as the target label, marking whether a transaction is fraudulent or not.

step	type	amount	oldbalanceOrg	newbalanceOrig	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
1	TRANSFER	367768.40	0.00	0.00	370763.10	16518.36	0.0	0.0
5	CASH_IN	596406.73	1850448.20	2446854.94	1615874.57	962737.60	0.0	0.0
4	CASH_IN	209292.82	7674233.58	7883526.40	4430219.90	4775319.66	0.0	0.0
8	CASH_OUT	47249.07	0.00	0.00	1040812.95	1930535.01	0.0	0.0
11	TRANSFER	1504521.82	3582.00	0.00	453618.07	3426464.19	0.0	0.0

Figure 1. Dataset Attributes.

These attributes form the basis for model evaluation and classification in detecting potential fraud in financial transactions. The dataset reveals a significant class imbalance, with non-fraudulent transactions far outnumbering fraudulent ones. Therefore, an effective classification method is required to handle this imbalance properly.

E. EXPERIMENTAL SETUP

The performance of the proposed CART-SQ algorithm was evaluated against two decision tree algorithms, namely the conventional CART and C4.5. The conventional CART was selected because CART-SQ was developed based on the CART framework, whereas C4.5 was adopted as the baseline method due to its use of the Gain Ratio criterion for attribute selection and its capability to handle both categorical and continuous attributes [38]. To ensure a fair and unbiased evaluation, all algorithms were implemented using the same dataset, preprocessing procedures, and experimental configuration.

The dataset was partitioned into training and testing sets using the hold-out validation approach while preserving the class distribution in both subsets through stratified sampling. All algorithms were trained and evaluated using identical data partitions to ensure that the observed performance differences were attributable to the algorithms themselves rather than variations in the experimental data. Model performance was assessed using Accuracy, Precision, Recall, and F1-score. These evaluation metrics provide a comprehensive assessment of the classification performance, particularly in measuring the model's ability to identify minority-class instances in an imbalanced dataset.

IV. RESULT AND DISCUSSION

A. DISTRIBUTION OF FRAUD CASES

An evaluation of fraud case distribution in the dataset reveals that unauthorized activities are not evenly distributed but are instead concentrated within specific transaction patterns, typically involving the transfer of funds out of a user's account. Fraudulent transactions tend to occur when a substantial amount of money is moved from the source account, followed by a significant reduction in balance. The majority of

transactions in the system are recorded as legitimate and show no indication of fraud.

This imbalance suggests that perpetrators prefer transaction channels that enable immediate fund transfers, thereby minimizing the chances of being detected in real time. Transactions involving unidirectional interactions or direct fund transfers from the source account are typically not exploited in fraudulent schemes.

Based on the transaction distribution in Table 1, most records represent legitimate transactions without any signs of fraud. The most frequent transaction type is *payment* with 1,130 instances, followed by *cash_out* (988 transactions) and *cash_in* (599 transactions), all of which are categorized as non-fraudulent. While *transfer* transactions appear commonly as normal with 272 cases, a notable 60 transactions of this type were classified as fraudulent. Similarly, *cash_out* is the most frequently used transaction type in fraudulent cases, totaling 62 instances. The *debit* type appeared 31 times, all of which were non-fraudulent.

Table 1. Distribution of Transaction Types by Fraud Status

type	isFlaggedFraud	isFraud	Count
payment	0.0	0.0	1130
cash_out	0.0	0.0	988
cash_in	0.0	0.0	599
transfer	0.0	0.0	272
cash_out	0.0	1.0	62
transfer	0.0	1.0	60
debit	0.0	0.0	31

The data suggest that fraud is predominantly associated with *cash_out* and *transfer* transactions. This is likely due to their nature, which involves direct fund transfers to other accounts, making them easier to exploit and harder to trace once completed. In contrast, the absence of fraudulent activity in *payment*, *cash_in*, and *debit* transactions may reflect tighter verification mechanisms or lower attractiveness to fraudsters.

As illustrated in Figure 2, the majority of transactions fall under the categories of *payment*, *cash_out*, and *cash_in*, all of which are recorded as non-fraudulent. The *payment* type dominates the total number of transactions, followed by *cash_out* and *cash_in*. Interestingly, fraud cases were only identified in the *cash_out* and *transfer* transaction types, with 62 and 60 fraudulent instances, respectively. Other transaction types showed no signs of fraudulent activity, suggesting a pattern in which fraudsters tend to exploit these two specific types. The *debit* transaction type had a relatively low number of occurrences and contained no fraudulent cases.

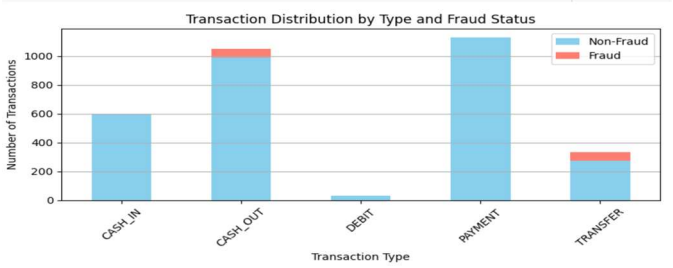

Figure 2. Distribution of Transactions by Type and Fraud Status.

Figure 3 presents the distribution of transactions across various time ranges, grouped into four categories: <1 hour, 1–<5 hours, 5–<7 hours, and 7–<11 hours from the start of data recording. Non-fraudulent *payment* transactions consistently dominate, particularly within the 5–<7 hour range, reaching a total of 819 transactions. *Cash_out* and *cash_in* transactions also show high frequency within this time range, yet remain within the non-fraudulent category.

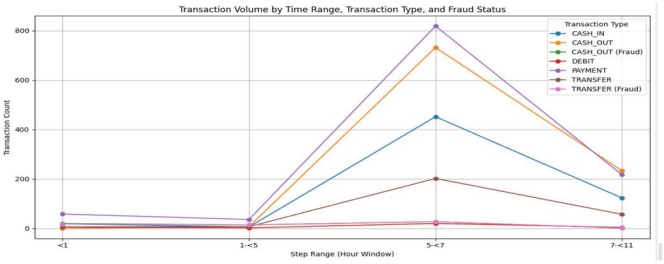

Figure 3. Transaction Distribution by Type, Time Range, and Fraud Status.

Fraudulent transactions-specifically *cash_out* and *transfer*—begin to appear in the earlier time windows (<1 and 1–<5 hours), with a noticeable increase during the 5–<7 hour period. This indicates that fraudulent activity tends to be active early and peaks within the mid-range of the observed time frame. Fraudulent *transfer* transactions follow a similar pattern to *cash_out*, suggesting a temporal trend shared between these two methods.

Overall, this distribution highlights that the majority of fraud cases are concentrated within specific time intervals, making these periods critical for early detection of suspicious activity.

As illustrated in Figure 4, non-fraudulent transactions are most frequent within the IDR 100,000 to less than 500,000 range, accounting for 1,108 transactions. This range also includes 22 fraudulent transactions, indicating a relatively high level of trust associated with mid-value transactions. The range between IDR 10,000 and less than 50,000 also shows significant activity, with 737 non-fraudulent transactions. However, it also contains a notable number of fraud cases—48 transactions—suggesting that this range should be closely monitored.

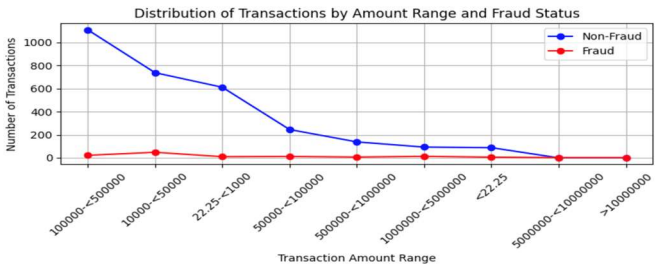

Figure 4. Distribution of Fraud and Non-Fraud Transactions by Transaction Amount Range.

Smaller transaction ranges, such as IDR 22.25 to less than 1,000 and below IDR 22.25, also exhibit a considerable number of non-fraudulent activities, totaling 611 and 88 transactions respectively. Nevertheless, these ranges include 10 and 6 fraudulent transactions, respectively, indicating that low-value transactions are not exempt from fraud risk.

Although transactions exceeding IDR 5,000,000 up to 10,000,000 and above 10,000,000 are relatively rare—each accounting for 2 fraudulent cases and no non-fraud cases—fraud is still present in these high-value ranges. This implies that while large transactions occur less frequently, they carry

substantial security risks and require stricter monitoring.

Overall, fraudulent transactions occur across all transaction value ranges, both low and high. However, there is a notable tendency for fraud to concentrate in low-to-mid-value transactions. Therefore, anomaly detection systems should prioritize monitoring medium and small-value transactions with high frequency.

B. IMPACT OF ALGORITHM MODIFICATION ON CLASSIFICATION PERFORMANCE

The classification report presents the evaluation results of the model's performance on test data, using key metrics such as accuracy, precision, recall, and F1-score. These metrics reflect how well the model identifies and separates target classes, particularly in binary classification tasks. *Precision* indicates the proportion of correctly predicted positive cases, while *recall* measures the model's ability to capture all actual positive cases. The *F1-score* is the harmonic mean of precision and recall, offering a balanced overview of the model's performance.

In the context of fraud classification, the classification report plays a crucial role in assessing a model's ability to accurately detect fraud cases without generating excessive false positives in the non-fraud class. It not only provides a numerical performance summary but also serves as a key reference in determining the most appropriate model for practical implementation.

As shown in Table 2, using an 80:20 data split, the classification performance is notably strong, particularly at a tree depth of 7. The standard CART model achieved an accuracy of 0.9762, while the modified CART-SQ (Classification and Regression Trees Simple Quadrat) slightly outperformed it with an accuracy of 0.9777. The C4.5 algorithm yielded an accuracy of 0.9763, representing a slight improvement over the conventional CART. Nevertheless, its overall performance remained lower than that of the proposed CART-SQ algorithm.

Table 2. Performance with 80:20 Data Split

Metric	CART		CART-SQ		C4.5	
	Non-Fraud	Fraud	Non-Fraud	Fraud	Non-Fraud	Fraud
Accuracy		0.9762		0.9777		0.9763
precision	0.98	0.82	0.98	0.88	0.98	0.85
recall	0.99	0.70	0.99	0.67	0.99	0.58
F1-score	0.99	0.75	0.99	0.76	0.99	0.69

For the non-fraud class, both models demonstrated nearly identical performance, with precision at 0.98, recall at 0.99, and an F1-score of 0.99. This indicates that CART, CART-SQ, and C4.5 are equally effective in identifying legitimate transactions. However, more notable differences appear in the fraud class. The CART model yielded a precision of 0.82 and recall of 0.70, resulting in an F1-score of 0.75. CART-SQ improved the fraud precision to 0.88, though its recall slightly declined to 0.67, maintaining a comparable F1-score of 0.76. Despite achieving competitive overall accuracy, C4.5 was less effective at detecting fraudulent transactions than both the conventional CART and the proposed CART-SQ algorithm, as evidenced by its lower Recall and F1-score values.

These results indicate that while both models exhibit high overall accuracy, CART-SQ provides the best balance between classification accuracy and fraud detection performance. Compared with the conventional CART and C4.5 algorithms, the proposed CART-SQ achieved the highest precision in detecting fraudulent transactions while maintaining a competitive F1-score. These results suggest that the proposed Simple Quadrat formulation influences the split selection process, leading to more selective decision boundaries that reduce false-positive predictions without compromising overall classification performance.

The decline in recall, alongside an increase in precision, suggests that CART-SQ adopts a more conservative approach in classifying fraud, potentially reducing false positives. The lower recall achieved by C4.5 indicates that the algorithm is less effective in identifying minority-class instances in an imbalanced dataset. In fraud detection applications, the trade-off between precision and recall requires careful consideration, as minimizing the misclassification of fraudulent transactions is essential for improving the reliability of the detection system.

Based on the results shown in Table 3, evaluated at a tree depth of 7, the CART, CART-SQ and C4.5 algorithms demonstrate different levels of performance in detecting fraudulent transactions. The overall accuracy increased from 0.9724 for CART to 0.9820 in CART-SQ, while C4.5 achieved an accuracy of 0.9758. Among the three algorithms, CART-SQ produced the highest classification accuracy.

For the non-fraud class, all three algorithms demonstrated nearly identical performance, achieving a precision of 0.98 and an F1-score of 0.99. CART-SQ slightly outperformed both CART and C4.5 by attaining a recall of 1.00, compared with 0.99 achieved by the other two algorithms. These results indicate that all models classified legitimate transactions with a high degree of consistency, while CART-SQ correctly identified all non-fraud instances in the test set.

Table 3. Performance with 70:30 Data Split

Metric	CART		CART-SQ		C4.5	
	Non-Fraud	Fraud	Non-Fraud	Fraud	Non-Fraud	Fraud
Accuracy		0.9724		0.9820		0.9758
precision	0.98	0.79	0.98	0.91	0.98	0.86
recall	0.99	0.53	1.00	0.67	0.99	0.56
F1-score	0.99	0.64	0.99	0.77	0.99	0.68

More pronounced differences were observed in the fraud class. The standard CART model achieved a precision of 0.79, a recall of 0.53, and an F1-score of 0.64. In comparison, CART-SQ substantially improved the classification performance, increasing precision to 0.91, recall to 0.67, and the F1-score to 0.77. Meanwhile, C4.5 achieved a precision of 0.86, a recall of 0.56, and an F1-score of 0.68. Although C4.5 produced higher precision than the conventional CART model, its recall and F1-score remained lower than those of CART-SQ, indicating that the proposed method provided a more balanced performance in detecting fraudulent transactions.

The superior performance of CART-SQ under the 70:30 data split further demonstrates the effectiveness of the proposed quadratic formulation in improving classification performance on imbalanced datasets. Its higher precision indicates a reduction in false-positive predictions, while the improved recall shows that the model successfully identified a greater proportion of fraudulent transactions than both the conventional CART and C4.5 algorithms. Furthermore, CART-SQ achieved the highest F1-score, indicating a better balance between precision and recall and confirming its effectiveness in fraud detection under class-imbalanced conditions.

Although the recall for the fraud class remained lower than that of the non-fraud class, CART-SQ provided a more effective balance between detecting fraudulent transactions and minimizing classification errors. In contrast, the lower recall achieved by C4.5 indicates that the algorithm was less effective in identifying minority-class instances within the imbalanced dataset. These findings suggest that CART-SQ is a more suitable classification model for fraud detection, where the ability to accurately recognize minority-class instances is critical for achieving reliable predictive performance.

C. CLASSIFICATION PERFORMANCE OF CART, CART-SQ, AND C4.5 AT DIFFERENT TREE DEPTHS

Evaluating classification performance at different tree depths provides a comprehensive assessment of the proposed CART-SQ algorithm. As the tree depth increased from 1 to 20, all three algorithms exhibited progressive improvements in classification accuracy before converging to a relatively stable performance level. Nevertheless, noticeable differences were observed in both the rate of performance improvement and the highest classification accuracy attained by CART, CART-SQ, and C4.5, demonstrating the influence of the splitting criterion on the predictive capability of each model.

As presented in Table 4, CART-SQ consistently outperformed both the conventional CART and C4.5 algorithms across most tree depths. Using the 80:20 data split, CART-SQ began to surpass the conventional CART model at a tree depth of 4 and maintained its superior performance through the maximum tree depth evaluated. At a tree depth of 7, CART-SQ achieved a classification accuracy of 0.9777, slightly exceeding the accuracies of 0.9762 and 0.9763 obtained by the conventional CART and C4.5 algorithms, respectively. As the tree depth increased, the classification accuracy of CART-SQ continued to improve, reaching its highest value of 0.9825 at tree depths ranging from 15 to 20. In contrast, the classification accuracy of the conventional CART and C4.5 algorithms stabilized at 0.9777 and 0.9763, respectively

Table 4. Comparative Classification Performance of CART, CART-SQ, and C4.5

Tree Depth	Split Data 80:20			Split Data 70:30		
	CART-SQ	CART	C4.5	CART-SQ	CART	C4.5
1	0.9475	0.9475	0.9542	0.9544	0.9544	0.9547
2	0.9539	0.9475	0.9668	0.9544	0.9544	0.9642
3	0.9571	0.9682	0.9731	0.9682	0.9692	0.9726
4	0.9730	0.9618	0.9700	0.9671	0.9639	0.9726
5	0.9777	0.9714	0.9747	0.9735	0.9714	0.9758
6	0.9777	0.9777	0.9763	0.9756	0.9714	0.9758
7	0.9777	0.9762	0.9763	0.9820	0.9724	0.9758
8	0.9793	0.9762	0.9763	0.9820	0.9724	0.9758
9	0.9793	0.9746	0.9763	0.9820	0.9714	0.9758
10	0.9793	0.9762	0.9763	0.9809	0.9714	0.9758
11	0.9809	0.9777	0.9763	0.9809	0.9703	0.9758
12	0.9809	0.9777	0.9763	0.9809	0.9703	0.9758
13	0.9809	0.9777	0.9763	0.9767	0.9703	0.9758
14	0.9809	0.9777	0.9763	0.9767	0.9692	0.9758
15	0.9825	0.9777	0.9763	0.9767	0.9692	0.9758
16	0.9825	0.9777	0.9763	0.9767	0.9692	0.9758
17	0.9825	0.9777	0.9763	0.9767	0.9692	0.9758
18	0.9825	0.9777	0.9763	0.9767	0.9692	0.9758
19	0.9825	0.9777	0.9763	0.9767	0.9692	0.9758
20	0.9825	0.9777	0.9763	0.9767	0.9692	0.9758

A similar trend was observed with the 70:30 data split. Although the conventional CART and C4.5 algorithms achieved slightly higher classification accuracy at several shallow tree depths, CART-SQ consistently delivered the best performance once the tree depth reached 7. The proposed method attained a maximum classification accuracy of 0.9820 at tree depths ranging from 7 to 10, whereas the conventional CART achieved a maximum accuracy of 0.9724 and C4.5 remained stable at 0.9758. These results indicate that the proposed quadratic formulation enables CART-SQ to maintain robust classification performance as the tree structure becomes more complex. Consequently, CART-SQ consistently achieved higher classification accuracy than both benchmark algorithms throughout the evaluation.

V. CONCLUSION

The experimental results indicate that CART-SQ provides a better balance between precision and recall than both the conventional CART and C4.5 algorithms. The proposed method identifies fraudulent transactions more reliably while reducing the number of false-positive predictions. Although the recall for the fraud class remains constrained by the inherent class imbalance, the proposed equivalent quadratic formulation consistently improves decision tree performance across different experimental settings. These findings suggest that the proposed formulation represents an effective alternative for enhancing decision tree classification on imbalanced datasets.

Future research may extend the proposed approach to multiclass classification problems and evaluate its performance on larger financial transaction datasets. In addition, integrating the proposed equivalent quadratic formulation with fuzzy decision tree models represents a promising direction for investigating its generalization capability in more complex classification scenarios.

References

[1] S. K. Hashemi, S. L. Mirtaheri, and S. Greco, “Fraud Detection in Banking Data by Machine Learning Techniques,” *IEEE Access*, vol. 11, pp. 3034–3043, 2023, https://doi.org/10.1109/ACCESS.2022.3232287.

[2] M. T. Aras and M. A. Guvensan, "A Multi-Modal Profiling Fraud-Detection System for Capturing Suspicious Airline Ticket Activities," *Appl. Sci.*, vol. 13, no. 24, p. 13121, 2023, https://doi.org/10.3390/app132413121.

[3] S. Jiang, R. Dong, J. Wang, and M. Xia, "Credit Card Fraud Detection Based on Unsupervised Attentional Anomaly Detection Network," *Systems*, vol. 11, no. 6, p. 305, 2023, https://doi.org/10.3390/systems11060305.

[4] M. Akouhar, et al., "Enhancing credit card fraud detection: the impact of oversampling rates and ensemble methods with diverse feature selection," Radioelectronic and Computer Systems, v. 2025, n. 1, p. 85-101, 2025. https://doi.org/10.32620/reks.2025.1.06.

[5] A. Asrianda, H. Mawengkang, P. Sihombing, and M. K. M. Nasution, "Optimization of marketing campaigns using a modified ID3 decision tree algorithm," *East.-Eur. J. Enterp. Technol.*, vol. 2, no. 13 (134), pp. 58–70, 2025, https://doi.org/10.15587/1729-4061.2025.327158.

[6] Y. Feng, M. Zhou, and X. Tong, "Imbalanced classification: a paradigm-based review," *Stat. Anal. Data Min. ASA Data Sci. J.*, vol. 14, no. 5, pp. 383–406, 2021, https://doi.org/10.1002/sam.11538.

[7] M. H. Bahar and H. Saad, "Decision Tree Induction Using Evolutionary Algorithms: ASurvey," *Int. J. Comput. Digit. Syst.*, vol. 15, no. 1, pp. 99–113, 2024, https://doi.org/10.12785/ijcds/150109.

[8] F. Bollwein and S. Westphal, "A branch & bound algorithm to determine optimal bivariate splits for oblique decision tree induction," *Appl. Intell.*, vol. 51, no. 10, pp. 7552–7572, 2021, https://doi.org/10.1007/s10489-021-02281-x.

[9] A. Singh, R. K. Ranjan, and A. Tiwari, "Credit Card Fraud Detection under Extreme Imbalanced Data: A Comparative Study of Data-level Algorithms," *J. Exp. Theor. Artif. Intell.*, vol. 34, no. 4, pp. 571–598, 2022, https://doi.org/10.1080/0952813X.2021.1907795.

[10] R. Nasfi, G. De Tré, and A. Bronselaer, "Improving Data Cleaning by Learning From Unstructured Textual Data," *IEEE Access*, vol. 13, pp. 36470–36491, 2025, https://doi.org/10.1109/ACCESS.2025.3543953.

[11] E. Rendón, R. Alejo, C. Castorena, F. J. Isidro-Ortega, and E. E. Granda-Gutiérrez, "Data Sampling Methods to Deal With the Big Data Multi-Class Imbalance Problem," *Appl. Sci.*, vol. 10, no. 4, p. 1276, 2020, https://doi.org/10.3390/app10041276.

[12] M. Alamri and M. Ykhlef, "Survey of Credit Card Anomaly and Fraud Detection Using Sampling Techniques," *Electronics*, vol. 11, no. 23, p. 4003, 2022, https://doi.org/10.3390/electronics11234003.

[13] H. Zhou, M. Zhang, L. Pang, and J.-H. Li, "Abnormal Detection of Cash-Out Groups in IoT Based Payment," *Sensors*, vol. 21, no. 22, p. 7507, 2021, https://doi.org/10.3390/s21227507.

[14] M. Esteve, N. Mollá-Campello, J. J. Rodríguez-Sala, and A. Rabasa, "The Effects of Abrupt Changing Data in CART Inference Models," in *Trends and Applications in Information Systems and Technologies*, vol. 1366, Á. Rocha, H. Adeli, G. Dzemyda, F. Moreira, and A. M. Ramalho Correia, Eds., in Advances in Intelligent Systems and Computing, vol. 1366. , Cham: Springer International Publishing, 2021, pp. 214–223. https://doi.org/10.1007/978-3-030-72651-5_21.

[15] A. Gottard, G. Vannucci, L. Grilli, and C. Rampichini, "Mixed-effect models with trees," *Adv. Data Anal. Classif.*, vol. 17, no. 2, pp. 431–461, 2023, https://doi.org/10.1007/s11634-022-00509-3.

[16] V. Blanco, A. Japón, and J. Puerto, "Robust optimal classification trees under noisy labels," *Adv. Data Anal. Classif.*, vol. 16, no. 1, pp. 155–179, 2022, https://doi.org/10.1007/s11634-021-00467-2.

[17] S. Lee, C. Lee, K. G. Mun, and D. Kim, "Decision Tree Algorithm Considering Distances Between Classes," *IEEE Access*, vol. 10, pp. 69750–69756, 2022, https://doi.org/10.1109/ACCESS.2022.3187172.

[18] X. Zhao and X. Nie, "Splitting Choice and Computational Complexity Analysis of Decision Trees," *Entropy*, vol. 23, no. 10, p. 1241, 2021, https://doi.org/10.3390/e23101241.

[19] M. Han, A. Li, Z. Gao, D. Mu, and S. Liu, "Hybrid Sampling and Dynamic Weighting-Based Classification Method for Multi-Class Imbalanced Data Stream," *Appl. Sci.*, vol. 13, no. 10, p. 5924, 2023, https://doi.org/10.3390/app13105924.

[20] K. A. Kumar, A. D., I. Chauhan, "Enhanced Credit Card Fraud Detection Using iForest Classifier of Ensemble Learning with Automated Hyperparameter Tuning", International Journal of Education and Management Engineering (IJEME), Vol.15, No.1, pp. 52-60, 2025. https://doi.org/10.5815/ijeme.2025.01.05.

[21] F. A. Almarshad, M. Zakariah, G. A. Gashgari, and T. Vaiyapuri, "RABEM: risk-adaptive Bayesian ensemble model for fraud detection," *Sci. Rep.*, vol. 15, no. 1, p. 36796, 2025, https://doi.org/10.1038/s41598-025-20651-0.

[22] J. Nicholls, A. Kuppa, and N.-A. Le-Khac, "Financial Cybercrime: A Comprehensive Survey of Deep Learning Approaches to Tackle the Evolving Financial Crime Landscape," *IEEE Access*, vol. 9, pp. 163965–163986, 2021, https://doi.org/10.1109/ACCESS.2021.3134076.

[23] S. Arshad, S. Latif, A. Salman, and S. Irfan, "Increasing Profitability and Confidence by using Interpretable Model for Investment Decisions," *Research Square*, 2023, https://doi.org/10.21203/rs.3.rs-3800581/v1.

[24] J. Han, "System Optimization of Talent Life Cycle Management Platform Based on Decision Tree Model," *J. Math.*, vol. 2022, no. 1, p. 2231112, 2022, https://doi.org/10.1155/2022/2231112.

[25] G. Petrides and W. Verbeke, "Cost-sensitive ensemble learning: a unifying framework," *Data Min. Knowl. Discov.*, vol. 36, no. 1, pp. 1–28, 2022, https://doi.org/10.1007/s10618-021-00790-4.

[26] Pradeep B. M., Sudeep J., Shivashankara S., Pavithra D. R., Ananth G. S., "Enhancing ATM Card Fraud Detection in Nigeria: A High-Performance Model with AI-Based Spending Pattern Analysis and Biometric Authentication", International Journal of Education and Management Engineering (IJEME), Vol.16, No.3, pp. 15-28, 2026. https://doi.org/10.5815/ijeme.2026.03.02.

[27] C. Pimsarn, T. Boongoen, N. Iam-On, N. Naik, and L. Yang, "Strengthening intrusion detection system for adversarial attacks: improved handling of imbalance classification problem," *Complex Intell. Syst.*, vol. 8, no. 6, pp. 4863–4880, 2022, https://doi.org/10.1007/s40747-022-00739-0.

[28] Z. Xu, G. Zhu, C. Yuan, and Y. Huang, "One-Stage Tree: end-to-end tree builder and pruner," *Mach. Learn.*, vol. 111, no. 5, pp. 1959–1985, 2022, https://doi.org/10.1007/s10994-021-06094-4.

[29] J. Soyemi, J. R. Olasina, "A Hybrid Stacked Ensemble Framework for Fraud Detection in Nigerian Financial Ecosystems: Evaluation with Localized Synthetic Data", International Journal of Wireless and Microwave Technologies(IJWMT), Vol.16, No.4, pp. 233-250, 2026. https://doi.org/10.5815/ijwmt.2026.04.13.

[30] G. Kaur, M. Kaur, P. Rattan, M. Kumar, "Feature Selection for Fraud Detection: Improving Machine Learning Capabilities on Portable Wallet", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.18, No.2, pp. 55-72, 2026. https://doi.org/10.5815/ijieeb.2026.02.04.

[31] F. Aaboub, H. Chamlal, and T. Ouaderhman, "Statistical analysis of various splitting criteria for decision trees," *J. Algorithms Comput. Technol.*, vol. 17, p. 17483026231198181, 2023, https://doi.org/10.1177/17483026231198181.

[32] A. Makolo, T. Adeboye, "Credit Card Fraud Detection System Using Machine Learning", International Journal of Information Technology and Computer Science (IJITCS), Vol.13, No.4, pp.24-37, 2021. https://doi.org/10.5815/ijitcs.2021.04.03.

[33] B. Reid O'Connor and S. Norton, "Exploring the challenges of learning quadratic equations and reflecting upon curriculum structure and implementation," *Math. Educ. Res. J.*, vol. 36, no. 1, pp. 151–176, 2024, https://doi.org/10.1007/s13394-022-00434-w.

[34] G. Zhang and A. Gionis, "Regularized impurity reduction: accurate decision trees with complexity guarantees," *Data Min. Knowl. Discov.*, vol. 37, no. 1, pp. 434–475, 2023, https://doi.org/10.1007/s10618-022-00884-7.

[35] S. Yu, X. Li, H. Wang, X. Zhang, and S. Chen, "C_CART: An instance confidence-based decision tree algorithm for classification," *Intell. Data Anal.*, vol. 25, no. 4, pp. 929–948, 2021, https://doi.org/10.3233/IDA-205361.

[36] H. Song, X. Wang, W. Tian, L. Shi, and S. Li, "Study on urban residents' travel mode choice based on the CART-Apriori method," *Sci. Rep.*, vol. 16, no. 1, p. 6270, 2026, https://doi.org/10.1038/s41598-026-37216-4.

[37] S. Tangirala, "Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm*," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 2, 2020, https://doi.org/10.14569/IJACSA.2020.0110277.

[38] Y. Liu and S. Yang, "Application of Decision Tree-Based Classification Algorithm on Content Marketing," *J. Math.*, vol. 2022, no. 1, p. 6469054, 2022, https://doi.org/10.1155/2022/6469054.

ASRIANDA *is currently a Lecturer at the Department of Informatics, Universitas Malikussaleh. His research interests include decision tree algorithms, classification techniques, and fraud detection in digital transactions. Inspired by the challenges in imbalanced data and real-world predictive modeling, he actively develops adaptive machine learning solutions for practical applications.*

NAUFAL BACHRI *is Marketing Management Lecturer at the Department of Management, Faculty of Economics and Business, Universitas Malikussaleh, Lhokseumawe, Indonesia. His research interests focus on service marketing and Islamic marketing.*

SUJACKA RETNO *received his Master's Degree Informatics from Universitas Sumatera Utara, Indonesia. He is currently a Lecturer at the Department of Informatics Engineering Universitas Malikussaleh. His research interests focus on algorithm development and machine learning optimization, with particular attention to improving model performance for intelligent systems.*