Fraudulent Transaction Detection in Online Systems Using Random Forest and Gradient Boosting

Satrya Fajri Pratama[1], Arif Mu'amar Wahid[2,*]

[1]Department of Computer Science, School of Physics, Engineering and Computer Science, University of Hertfordshire, United Kingdom

[2]Master of Computer Science, Computer Science Faculty, Universitas Amikom Purwokerto, Indonesia

ABSTRACT

The rapid increase in online transactions has significantly raised the risk of fraudulent activities, leading to substantial financial losses. Traditional fraud detection methods often struggle to address the complexity and scale of modern digital fraud. This paper explores the application of machine learning techniques, specifically Random Forest and Gradient Boosting, to detect fraudulent transactions. Both algorithms are widely recognized for their ability to handle large, complex datasets and improve predictive accuracy. The study examines how these techniques work, with Random Forest focusing on ensemble learning and feature importance, and Gradient Boosting employing an iterative, stage-wise approach to correct errors from previous models. Key challenges in fraud detection, including class imbalance, data scarcity, the evolving nature of fraud, and high-dimensional data, are discussed in depth. The paper reviews relevant studies that have utilized machine learning for fraud detection in various contexts, including e-commerce and credit card fraud, highlighting the strengths and limitations of different approaches. It also examines strategies to mitigate challenges, such as resampling techniques and continuous learning. The findings emphasize that while machine learning offers significant improvements in fraud detection, continuous adaptation is essential to keep pace with evolving fraud tactics. By providing a comprehensive overview of machine learning in fraud detection, this research contributes valuable insights into enhancing security measures for digital transactions and financial systems.

Keywords Fraud Detection, Machine Learning, Random Forest, Gradient Boosting, Class Imbalance

Introduction

The exponential growth of online transactions in recent years has been a defining feature of the digital age. E-commerce, mobile banking, and digital payment systems have flourished, fundamentally reshaping consumer behavior and business operations. This transformation, propelled by the increasing global connectivity and convenience offered by digital platforms, has significantly changed the landscape of financial transactions. However, as the digital world expands, so too does the threat of online fraud, a rapidly escalating challenge that has spurred increasing concern among both consumers and financial institutions alike.

In recent years, the incidence of online fraud has seen alarming growth, directly correlating with the surge in online transactions. During the COVID-19 pandemic, for instance, the volume of online transactions skyrocketed, but so did the prevalence of fraud. According to recent studies, incidents of online fraud rose by a staggering 44.21% in the pandemic period, underscoring the vulnerabilities introduced by a digital-first economy [1]. This phenomenon is not confined to one form of fraud but spans multiple avenues, including credit card

Submitted 5 January 2025
Accepted 3 February 2025
Published 15 March 2025

Corresponding author
Arif Mu'amar Wahid,
arif@amikompurwokerto.ac.id

Additional Information and Declarations can be found on page 112

How to cite this article: S.F.Pratama and A.M.Wahid, "Fraudulent Transaction Detection in Online Systems Using Random Forest and Gradient Boosting ," *J. Cyber. Law.*, vol. 1, no. 1, pp. 88-115, 2025.

fraud, identity theft, and phishing scams, all of which are amplified by the ever-growing digital ecosystem [2]. These statistics not only highlight the growing concern but also point to an alarming shift in the nature of financial crimes, which have increasingly been conducted through digital means [3].

The threat of fraud is not only a consequence of the rise in online transactions but also a product of the evolving complexity of digital payment systems. Fraudsters continuously adapt to new technologies, exploiting security gaps and human vulnerabilities. For instance, the allure of mobile payments—deemed one of the most significant innovations in digital finance—has come with its own set of risks. Despite their convenience, mobile payment platforms have often been criticized for their lack of sufficient security measures, making them prime targets for cybercriminals [4]. Moreover, the lack of consumer awareness about secure payment practices only exacerbates the risk, leaving many unaware of the dangers lurking in seemingly innocuous transactions [5]. The rapid adoption of these systems, without a commensurate emphasis on security protocols, creates fertile ground for fraudulent activities, as fraudsters exploit weaknesses to siphon off funds from unsuspecting users [6].

The escalation of online fraud demands proactive intervention. In response to this, technological advancements have been brought to the forefront to mitigate the risk. Data mining and machine learning algorithms, in particular, have emerged as potent tools in fraud detection. By analyzing vast swaths of transaction data, these systems can identify unusual patterns that are indicative of fraud [7]. Machine learning models such as Random Forest and Gradient Boosting have shown great promise in improving the accuracy and timeliness of fraud detection, offering a robust approach to preventing financial loss [3], [8]. These models, when applied to large datasets, can effectively spot anomalies and flag transactions that deviate from typical patterns, enabling real-time detection and intervention.

Moreover, the advent of big data analytics has revolutionized the potential for fraud prevention. By processing immense volumes of transaction data, algorithms can uncover subtle fraud indicators that might be imperceptible through traditional methods. This approach is especially effective in mobile payments, where fraudsters exploit the rapid and often unmonitored nature of transactions [9]. Big data tools help financial institutions understand transaction behavior, ensuring that detection mechanisms are both efficient and scalable in the face of growing transaction volumes.

Yet, despite these advancements, online fraud continues to present significant challenges. The integration of cutting-edge fraud detection systems must go hand-in-hand with greater public awareness and consumer education. Financial institutions must work to demystify security practices, ensuring that users understand how to protect themselves against common online threats. Furthermore, the continuous development of hybrid strategies that combine technological advancements with robust consumer education will be key to maintaining trust in the digital financial ecosystem [10], [11]. In this paper, we explore the potential of Random Forest and Gradient Boosting algorithms to enhance the detection of fraudulent transactions, leveraging transaction data to propose an effective solution for combating online fraud.

The proliferation of online transactions has revolutionized commerce, providing consumers with unparalleled convenience, speed, and accessibility. Yet, this transformation has brought with it a daunting challenge—an exponential rise in fraudulent activities, which has led to substantial financial losses for both consumers and businesses alike. The shift from traditional brick-and-mortar

transactions to digital platforms has, ironically, created new opportunities for cybercriminals to exploit vulnerabilities, making online fraud a persistent and increasingly sophisticated threat.

Several factors contribute to this surge in online fraud. First, the rapid adoption of digital payment methods, combined with the expansion of the Internet of Things (IoT), has led to a marked increase in transaction volumes. With more consumers conducting business online and on mobile platforms, the opportunities for fraudsters to engage in illicit activities have multiplied. This is evident in the rise of credit card fraud, which continues to be a prevalent form of cybercrime. As digital transactions become the norm, fraudsters have adapted, leveraging the very technologies designed to streamline financial transactions to orchestrate their crimes. A recent study highlights the direct correlation between the widespread use of online transaction systems and the escalation of fraud risk, with fraud-related losses reaching billions annually [12]. Such trends are mirrored in the persistence of online banking fraud, a problem that has endured despite significant investments by banks in securing their digital platforms [13].

The financial ramifications of these fraudulent activities are profound, but they represent just the tip of the iceberg. The immediate monetary loss is often compounded by far-reaching consequences for victims, including identity theft, the compromise of personal data, and significant reputational damage. These secondary effects can be even more damaging than the initial financial loss, as they erode trust in digital platforms and instill anxiety among consumers about the safety of online transactions. Indeed, the pervasive fear of identity theft has become a critical concern for many consumers, with studies showing that the anxiety surrounding online fraud deters a significant number of potential users from engaging in e-commerce altogether [14]. In many cases, consumers are reluctant to share sensitive financial information, fearing that their privacy will be exploited by malicious entities [15]. This wariness only serves to deepen the divide between consumers and the digital economy, posing a challenge for businesses looking to capitalize on the shift to online commerce.

The evolving nature of online fraud tactics presents an additional challenge for stakeholders. Fraudulent schemes are becoming increasingly complex, making traditional detection methods inadequate. Financial institutions are recognizing that they must evolve their fraud detection strategies to keep pace with these threats. One of the most promising avenues in this effort is the adoption of machine learning and deep learning technologies. These cutting-edge methods allow for the real-time analysis of vast datasets, enabling systems to identify anomalies and patterns indicative of fraudulent behavior. Through this data-driven approach, banks and payment providers can respond more swiftly and accurately to emerging threats, minimizing the impact of fraud on both consumers and businesses [6], [16].

Despite the progress made in fraud detection, the issue remains pervasive, particularly in the wake of the COVID-19 pandemic. As consumer behavior shifted toward online transactions during the global health crisis, the rise in fraud incidents paralleled this change [10]. This stark reality underscores the need for continuous adaptation of fraud prevention systems, as fraudsters relentlessly refine their methods. It also calls for a holistic approach to cybersecurity, wherein collaboration between financial institutions, policymakers, and consumers is paramount. Only through these collective efforts can the integrity of online transaction systems be safeguarded and public trust restored in digital financial ecosystems [17], [18].

The primary goal of this study is to develop an effective approach for detecting fraudulent transactions using data mining techniques. As digital transactions continue to surge globally, so too does the threat of fraud, making timely detection a critical component in mitigating financial losses. This research explores the use of advanced machine learning algorithms—specifically, Random Forest and Gradient Boosting—to analyze transactional data and identify potentially fraudulent activities. By leveraging these sophisticated models, the study aims to contribute to the broader effort of enhancing fraud detection systems, which is crucial in the era of rapid digitalization and increasing cyber threats.

Fraud detection, particularly in online financial systems, has grown from a niche concern into a central issue for businesses, governments, and consumers alike. The escalating sophistication of cybercriminals, coupled with the widespread adoption of digital payment platforms, demands the application of more advanced techniques. Data mining, which involves extracting meaningful patterns from large datasets, has proven to be an invaluable tool in the fight against fraud. The study focuses on how these patterns, once identified, can inform the development of more robust and responsive fraud detection systems, capable of providing real-time alerts and reducing the economic burden of fraudulent transactions.

The significance of this study lies in its potential to enhance the security and integrity of digital financial ecosystems. As digital payments and online transactions become ubiquitous, ensuring that these systems are resilient to fraud is paramount. Financial institutions, e-commerce platforms, and consumers alike face the consequences of fraud, including financial losses, identity theft, and a loss of trust in digital services. The importance of robust fraud detection systems cannot be overstated, as they not only protect individuals and businesses from monetary loss but also maintain the overall stability of digital payment infrastructures.

Fraud detection is a cornerstone of cybersecurity in the financial sector, with implications that stretch far beyond mere transaction monitoring. It impacts everything from regulatory compliance to customer retention and overall user experience. For instance, effective fraud detection instills consumer confidence, which in turn drives further adoption of digital payment systems. Conversely, failure to detect fraud can lead to a cascade of negative outcomes, including reputational damage for financial institutions, legal ramifications, and even systemic disruptions in the broader financial ecosystem [3]. Thus, improving fraud detection methods through the use of advanced data mining techniques is not only a matter of financial prudence but also of maintaining the long-term viability and security of digital financial systems.

This paper narrows its focus to transactional data analysis, specifically examining the application of Random Forest and Gradient Boosting algorithms in detecting fraudulent transactions. These two algorithms were chosen for their proven effectiveness in handling large, complex datasets and their ability to discern patterns in noisy, imbalanced data, which is characteristic of fraud detection scenarios. Random Forest, an ensemble method based on decision trees, is well-suited for capturing non-linear relationships within the data and providing robust predictions even in the presence of outliers. Meanwhile, Gradient Boosting, particularly in its XGBoost implementation, has become a staple in predictive modeling due to its high accuracy and ability to learn from the errors of previous models, making it highly effective for fine-tuning fraud detection systems.

The scope of this study is strictly limited to the transactional data available within the dataset, which includes features such as transaction amount, type, account balances, and transaction time. These features form the basis of the machine learning models, and the research seeks to determine which patterns and combinations of these features are most indicative of fraudulent activity. By focusing on this transactional data, the study aims to provide actionable insights for financial institutions looking to implement or enhance their fraud detection systems. Furthermore, while the study focuses on two specific algorithms, the broader implications of the research extend to the use of other data mining techniques and algorithms in similar contexts, contributing to the ongoing dialogue on improving fraud detection in digital financial systems [3], [8].

Literature Review

Overview of Fraud Detection

Fraud detection in online transactions has undergone a profound transformation in recent decades, adapting to the ever-evolving landscape of digital commerce. As online transactions increase in frequency and complexity, so too has the sophistication of fraudulent activities. Historically, fraud detection relied heavily on manual methods, but as digital platforms expanded, automated and data-driven approaches began to dominate. Today, fraud detection methods are categorized into traditional techniques, data-driven approaches, and advanced machine learning algorithms. Each of these categories brings its own set of advantages and limitations, highlighting the ongoing need for innovation and refinement in fraud detection systems.

Traditional fraud detection methods were largely manual and dependent on human oversight. In many cases, fraud was identified through internal audits or through tips from employees and customers. A study by [19] reveals that over 40% of fraud cases were detected through tips, while internal audits were responsible for identifying only 15% of cases. This highlights the critical role that human involvement has played in detecting fraudulent behavior, particularly in industries where personal interaction is a key component of transaction verification. For instance, in the healthcare sector, systematic audits have been a cornerstone for detecting fraudulent claims, with audits uncovering discrepancies and unusual billing patterns that automated systems may have missed [20].

Despite their historical relevance, traditional methods are increasingly seen as inadequate in today's fast-paced digital environment. The manual nature of these techniques limits their ability to handle the sheer volume and speed of modern transactions, particularly in sectors like e-commerce and digital payments, where transactions are conducted in real time and involve vast amounts of data. Moreover, these techniques are often reactive, relying on identifying fraud after it has occurred rather than preventing it in real time. This lag in detection has led to a paradigm shift toward more proactive, automated methods of fraud detection.

With the rise of digital transactions, fraud detection has become increasingly reliant on data-driven approaches, particularly data mining techniques. These methods leverage large datasets to identify patterns, trends, and anomalies that may signal fraudulent activities. A significant advancement in this area is the application of unsupervised anomaly detection algorithms, which have been successfully used to detect fraud by analyzing logs and transaction records for unusual behaviors [21]. By examining deviations from established norms, such

algorithms can flag suspicious transactions that might otherwise go unnoticed. Additionally, big data analytics has revolutionized fraud detection by allowing systems to process vast amounts of transaction data in real time. This capability enables the detection of fraudulent activities at a scale and speed previously unimaginable [22].

However, while data-driven approaches have significantly enhanced fraud detection, they are not without their challenges. One major issue is the generation of false positives, where legitimate transactions are mistakenly flagged as fraudulent. This problem arises because the models used in anomaly detection often rely on predefined patterns or thresholds, which can be easily disrupted by new or complex transaction behaviors. As a result, businesses may face the dilemma of either allowing fraudulent transactions to slip through the cracks or unnecessarily inconveniencing customers by flagging legitimate transactions. Balancing sensitivity and specificity remains one of the core challenges in refining data-driven fraud detection systems.

The most significant strides in fraud detection have come from the integration of machine learning (ML) algorithms, which offer a more dynamic and adaptive approach to identifying fraudulent activities. Algorithms such as decision trees, support vector machines (SVM), and neural networks have become central to modern fraud detection efforts. Decision trees, for instance, are popular for their ability to handle both categorical and continuous data, providing clear decision rules for classifying transactions as fraudulent. Meanwhile, SVMs excel in high-dimensional spaces and can effectively classify transactions by finding the optimal boundary between fraudulent and non-fraudulent data points.

More recently, deep learning models, including convolutional neural networks (CNNs), have been applied to fraud detection with considerable success. These models, which are adept at processing complex and hierarchical data, have shown great promise in uncovering intricate patterns that might be missed by traditional methods [23]. For instance, a study by [24] demonstrates the effectiveness of a hierarchical behavior-knowledge space model in detecting irregular patterns in credit card transactions, offering a nuanced understanding of how transaction behavior can signal fraud. By learning from vast amounts of data, machine learning models can continuously improve their performance, adapting to emerging fraud tactics and enhancing the accuracy of fraud detection systems.

Despite their impressive capabilities, machine learning algorithms come with their own set of challenges. The primary issue lies in the inherent complexity of these models, which require large amounts of high-quality data for training. Furthermore, machine learning models are not immune to adversarial manipulation—fraudsters may eventually learn to evade detection by exploiting weaknesses in the algorithms. Thus, fraud detection systems must remain flexible and continuously updated to address new and evolving threats [25].

Data Mining in Fraud Detection

The application of data mining algorithms has become indispensable in the ongoing battle against fraudulent activities, particularly in the realm of online transactions. As the volume of data grows exponentially, traditional methods struggle to keep pace, making data-driven solutions not only more efficient but essential. Among the myriad of data mining techniques, Random Forest and Gradient Boosting have emerged as two of the most powerful algorithms, largely due to their capacity to handle complex, high-dimensional data and improve detection accuracy. These algorithms have transformed fraud detection from a

reactive process to a dynamic, predictive one, capable of identifying suspicious transactions in real time.

Random Forest, an ensemble learning method, has garnered significant attention for its robust performance in fraud detection tasks. By constructing multiple decision trees during training, the algorithm outputs the mode of their predictions to classify data points. Its ability to handle large datasets, manage high-dimensional features, and mitigate overfitting makes it a prime candidate for fraud detection. In fact, a comparative study by [26] demonstrated that Random Forest outperforms traditional methods such as logistic regression in detecting credit card fraud, achieving superior accuracy and recall rates. The algorithm's robustness lies in its ensemble approach, which averages out the biases of individual decision trees, thus leading to more reliable predictions.

Gradient Boosting, and its implementation XGBoost, represents another significant advancement in fraud detection algorithms. Unlike Random Forest, which builds multiple trees in parallel, Gradient Boosting constructs models sequentially, with each new model aimed at correcting the errors of its predecessor. This iterative approach allows Gradient Boosting to fine-tune its predictions, yielding high accuracy in identifying fraudulent activities. XGBoost, in particular, has garnered widespread recognition in the machine learning community for its efficiency and predictive power, especially in imbalanced datasets [3].

Both Random Forest and Gradient Boosting have demonstrated significant success in various domains of fraud detection. While their initial applications were in credit card fraud detection, their use has extended into a variety of financial services. [27] stress the importance of employing such data mining techniques in the banking sector, particularly in managing the complexities of imbalanced datasets and the high cost of false negatives. By automating the process of identifying fraudulent transactions, these algorithms not only improve efficiency but also reduce the likelihood of human error.

Random Forest

Random Forest is an ensemble learning algorithm widely recognized for its powerful classification capabilities, especially in contexts such as fraud detection, where accuracy is paramount. It operates by constructing a large number of decision trees during the training process and outputs the mode of their predictions for classification tasks. The strength of this method lies in its ability to aggregate the predictions of multiple trees, reducing the overfitting tendencies commonly associated with individual decision trees. The result is a robust model that generalizes well to unseen data, making it particularly well-suited for detecting fraudulent activities in online transactions.

One of the key components of Random Forest is bagging, short for bootstrap aggregation, which involves creating multiple subsets of the training data through random sampling with replacement. Each subset is then used to train a separate decision tree. The aggregation of multiple trees helps to reduce variance and improves the model's generalization capabilities. This technique is particularly important when working with highly variable data, such as transaction records in fraud detection, where the patterns of fraud may be subtle and varied.

In addition to the randomness introduced by bagging, Random Forest incorporates feature randomness. When splitting a node in a decision tree, only a random subset of features is considered, rather than evaluating all features. This deliberate feature randomness decorrelates the trees, ensuring that the

trees in the forest are not highly correlated, which improves the overall performance of the model. By diversifying the decision-making process in each tree, Random Forest enhances the model's ability to detect nuanced patterns in the data, such as fraudulent transaction behaviors, that might otherwise be overshadowed by dominant features.

Once all the trees have been trained, Random Forest aggregates their predictions through a voting mechanism. For classification tasks, this means that the final prediction is determined by the majority vote of all trees. This ensemble approach tends to yield more reliable results than individual decision trees, as it reduces the impact of noise and outliers in the data. The aggregated decision process also smoothens the effect of anomalies, ensuring that fraudulent transactions, even if they deviate significantly from typical patterns, are more likely to be identified.

At the heart of Random Forest's decision trees is the concept of entropy, a measure of uncertainty or impurity in a dataset. Entropy quantifies the disorder in the data, with higher entropy indicating more randomness and lower predictability. The goal when building a decision tree is to reduce entropy with each split, thereby organizing the data into more homogeneous subsets. The formula for entropy \(H(S)\) for a binary classification problem is:

$$H(S) = -p_1 \log_2(p_1) - p_2 \log_2(p_2)$$

Where:
- \(p_1\) is the proportion of instances in class 1,
- \(p_2\) is the proportion of instances in class 2.

In the context of a decision tree, each feature is evaluated based on how much it can reduce entropy. The information gain (\(IG\)) from a particular feature \(A\) is calculated as the difference between the entropy of the dataset before the split and the weighted sum of the entropies of the subsets after the split:

$$IG(S, A) = H(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} H(S_v)$$

Where:
- \(H(S)\) is the entropy of the original set,
- \(S_v\) is the subset of instances for which attribute \(A\) has value \(v\),
- \(|S_v|\) is the number of instances in subset \(S_v\),
- \(|S|\) is the total number of instances in the original set.

The feature that provides the highest information gain is selected for the split, as it leads to the greatest reduction in uncertainty about the target variable. This approach ensures that the decision tree is structured in a way that maximizes the clarity of the predictions.

Random Forest's ability to handle large datasets with numerous features makes it an ideal candidate for fraud detection, where datasets often contain a mix of numerical and categorical features, and fraudulent behavior can be complex and subtle. In the context of financial transactions, Random Forest has been shown to perform exceptionally well in distinguishing between legitimate and fraudulent activities. The model's capacity to process high-dimensional data and identify key patterns without overfitting to noise is particularly valuable in fraud detection scenarios, where the presence of fraudulent transactions is rare, and detecting them amidst a sea of legitimate transactions is challenging [28], [29].

Moreover, Random Forest provides insights into the relative importance of

various features, which can help financial institutions understand the factors contributing most significantly to fraud. This interpretability is crucial in the financial sector, where transparency in decision-making is essential for regulatory compliance and trust. The algorithm's feature importance ranking can reveal which transaction attributes—such as transaction amount, time, or location—are most indicative of fraud, allowing institutions to refine their fraud detection systems further.

However, while Random Forest has demonstrated its utility in fraud detection, it is not without limitations. Like all machine learning algorithms, its performance depends on the quality and quantity of data, and in cases of extreme class imbalance (e.g., fraudulent transactions being a small fraction of total transactions), additional techniques such as SMOTE (Synthetic Minority Over-sampling Technique) may be needed to ensure optimal performance. Despite these challenges, Random Forest remains a cornerstone of fraud detection due to its robustness, versatility, and ability to provide actionable insights into the detection process.

Gradient Boosting

Gradient Boosting has emerged as one of the most powerful and widely used machine learning techniques, renowned for its effectiveness in both regression and classification tasks. Its capacity to improve predictive accuracy, especially in complex datasets, has made it a go-to choice for applications ranging from fraud detection to customer churn prediction. The strength of Gradient Boosting lies in its iterative, stage-wise approach, where it builds a predictive model by sequentially combining the outputs of multiple base learners, typically decision trees. Each subsequent model in the sequence focuses on correcting the errors of the previous one, making it particularly adept at refining predictions and handling subtle patterns in data.

At its core, Gradient Boosting seeks to minimize a specified loss function \(L(y, F(x))\), where \(y\) represents the true output, \(F(x)\) is the predicted output from the model, and \(x\) represents the input features. The process is inherently iterative, with the goal of refining the model step by step, focusing on areas where previous models have underperformed.

At the heart of Gradient Boosting lies gradient descent optimization, a method for minimizing the loss function by adjusting the model parameters in the direction of the negative gradient. In the context of Gradient Boosting, the parameters being optimized are the weights of the decision trees added during each iteration. The formula for gradient descent is as follows:

$$\theta_{t+1} = \theta_t - \eta \nabla L(\theta_t)$$

Where:

- \(\theta_t\) represents the model parameters at iteration \(t\),
- \(\eta\) is the learning rate, controlling the size of each step,
- \(\nabla L(\theta_t)\) is the gradient of the loss function with respect to the model parameters.

In Gradient Boosting, the gradient is calculated for the pseudo-residuals at each step, guiding the updates to the decision trees' weights. This process of iteratively adjusting the model parameters based on the gradient ensures that the model converges toward the optimal solution, reducing the error progressively. The gradient descent approach allows the algorithm to fine-tune its parameters, improving its ability to detect fraudulent behavior in datasets with many variables and complex relationships.

Gradient Boosting, particularly in its advanced implementations such as XGBoost andLightGBM, has become a mainstay in machine learning competitions and real-world applications. These implementations offer even greater speed, scalability, and performance, making Gradient Boosting a versatile tool for a wide range of predictive modeling tasks. For example, XGBoost has been used extensively in financial fraud detection, where its ability to handle imbalanced datasets and deliver high prediction accuracy is critical [30]. The algorithm's performance in the Kaggle competition environment speaks to its robustness, where it routinely outperforms other models across various datasets, including those with missing values, skewed distributions, and non-linear relationships [31].

Beyond its raw predictive power, Gradient Boosting offers significant interpretability. It can provide feature importance scores, allowing data scientists and financial analysts to gain insights into which features most contribute to the identification of fraudulent activities. This transparency is essential in fields like fraud detection, where understanding the underlying reasons for model predictions can lead to better decision-making and regulatory compliance.

Despite its advantages, Gradient Boosting requires careful tuning of hyperparameters to avoid overfitting, particularly in complex datasets where the model may overfit noise rather than learn meaningful patterns. Additionally, the computational complexity of Gradient Boosting can be a limitation when dealing with massive datasets, requiring optimized implementations likeLightGBM that are designed to handle large-scale problems efficiently.

Existing Challenges

Detecting fraudulent transactions remains one of the most formidable challenges in the realm of machine learning, particularly due to the unique and complex characteristics of the datasets involved. The intrinsic nature of transaction data, coupled with the continuously evolving tactics of fraudsters, creates a landscape fraught with difficulties. One of the most critical issues in fraud detection is theclass imbalance that characterizes most datasets. In fraud detection scenarios, the number of legitimate transactions far outpaces the number of fraudulent ones, leading to a skewed distribution that complicates model training and hinders accurate prediction of fraudulent activity.

The most pressing challenge in fraud detection lies in theclass imbalance problem. As [32] note, fraudulent transactions constitute only a tiny fraction of the total transaction dataset, creating a severe imbalance between the legitimate and fraudulent classes. This disproportionate distribution presents a significant challenge for machine learning models, as the algorithm tends to be biased towards the majority class—the legitimate transactions—leading to an overwhelming number of false negatives. In other words, the model may incorrectly classify fraudulent transactions as legitimate, severely undermining its ability to detect fraud. The scarcity of fraudulent examples further exacerbates the issue, as there may not be enough diverse instances of fraud for the model to learn from, resulting in ineffective or inaccurate predictions [33].

Another significant hurdle is the scarcity of labeled data. For fraud detection models to train effectively, they require enough labeled instances—examples of both legitimate and fraudulent transactions. However, obtaining such labeled data is often challenging, as it is difficult for organizations to obtain or annotate enough fraudulent examples. As [34] points out, fraud detection systems are frequently limited by the sheer lack of labeled fraud data, which restricts the ability to train robust classifiers. This scarcity not only intensifies the class

imbalance problem but also limits the generalizability of the model. Without a representative sample of fraudulent activities, the model may fail to recognize emerging fraud tactics or subtle variations in fraudulent behavior.

Fraud detection models face the constant challenge of concept drift, where fraudulent tactics evolve over time. As fraudsters adapt to the detection systems in place, they modify their approaches to circumvent detection, which is a form of dynamic change in the underlying distribution of the data. This phenomenon makes fraud detection a moving target. As [35] observe, models trained on historical data may become outdated if they are not continuously updated to account for new fraud patterns. Concept drift forces models to be more adaptable, requiring constant retraining and fine-tuning to maintain their relevance in detecting new fraud tactics. This adds an additional layer of complexity to fraud detection systems, which must not only be accurate but also agile in responding to the evolving nature of fraud.

Fraud detection datasets often contain high-dimensional data, with numerous features that describe various aspects of a transaction. While having a rich set of features can improve the model's ability to detect subtle patterns, it also introduces the challenge of the curse of dimensionality. As [36] explain, as the number of features increases, the data space expands exponentially, and the density of data points becomes sparse. This sparsity hampers the model's ability to learn meaningful patterns and increases the likelihood of overfitting, especially in high-dimensional spaces where the model may become too specialized to the training data. This issue is particularly pertinent in fraud detection, where the relationship between features and fraud is often intricate and not easily captured in high-dimensional spaces.

The complexity of fraudulent patterns further complicates the task of detection. Fraudulent transactions do not follow simple, linear patterns and often involve sophisticated, subtle behaviors that can be challenging for traditional machine learning algorithms to identify. As [37] highlight, fraudulent behavior can manifest in multiple ways, from small, seemingly innocuous changes in transaction patterns to more elaborate schemes that span multiple stages. Simple models may fail to capture these complexities, leading to insufficient detection performance. For instance, fraudsters may mimic legitimate patterns to avoid detection, requiring models that can identify subtle deviations from expected behaviors. This intricacy demands the use of more advanced algorithms, capable of detecting nuanced anomalies that may otherwise go unnoticed.

In imbalanced datasets, relying on standard evaluation metrics such as accuracy can be misleading. A model that predicts every transaction as legitimate might still achieve high accuracy simply due to the overwhelming number of legitimate transactions. However, this model would fail to detect any fraudulent behavior, rendering it ineffective for fraud detection purposes. As [38] and [39] argue, evaluation metrics such as precision, recall, F1-score, and the area under the ROC curve (AUC-ROC) provide a more accurate picture of a model's performance in imbalanced settings. These metrics emphasize the model’s ability to correctly identify fraudulent transactions, which is paramount in fraud detection, where false positives and false negatives have significantly different implications.

Relevant Studies

The growing integration of machine learning (ML) algorithms in fraud detection has transformed the way businesses and financial institutions identify and

mitigate fraudulent activity. From e-commerce to credit card fraud and financial statement manipulation, various studies have explored the potential of machine learning to enhance detection capabilities across different domains. These studies not only demonstrate the versatility of machine learning but also highlight its ability to overcome traditional fraud detection system limitations, such as speed, efficiency, and accuracy. Below is a summary of some of the most relevant studies that have utilized machine learning for fraud detection in various contexts.

[40] study explores the use of the Hidden Markov Model (HMM) for fraud detection in e-commerce and face-to-face transactions. By incorporating HMM-based features, the study found that machine learning approaches outperformed traditional fraud detection systems, particularly in terms of efficiency and detection accuracy. This highlights the growing importance of machine learning in improving the precision of fraud detection systems, which have historically struggled with identifying complex patterns in transaction data. The ability of machine learning to adapt to different transaction types further underlines its potential in a rapidly evolving e-commerce landscape.

In the realm of financial transactions, [41] compared the performance of Logistic Regression and Random Forest algorithms for credit card fraud detection. Their findings revealed that Random Forest achieved an impressive accuracy rate of 98.8%, making it a highly effective algorithm for distinguishing between fraudulent and legitimate credit card transactions. This study emphasizes the efficacy of machine learning algorithms, particularly Random Forest, in handling large datasets, a critical component in credit card fraud detection, where the volume of transactions is immense. By improving detection rates, these algorithms can play a pivotal role in reducing financial losses due to fraud.

A major challenge in fraud detection is theclass imbalance issue, where fraudulent transactions represent only a small fraction of total transactions. [42] tackled this challenge by using the Synthetic Minority Oversampling Technique (SMOTE) to balance the dataset. By combining SMOTE with machine learning classifiers like Decision Trees and Random Forest, they demonstrated improved classification outcomes, particularly in e-commerce fraud detection. This highlights the importance of addressing class imbalance to enhance the effectiveness of machine learning models. Techniques such as SMOTE can help ensure that the model is exposed to a more balanced view of fraudulent behavior, ultimately leading to better detection capabilities.

[43] conducted an empirical analysis of machine learning models in various financial fraud detection scenarios. The study demonstrated that algorithms, including those based onconvolutional neural networks (CNNs), significantly improved fraud detection performance. The adaptability of machine learning to different types of financial fraud—from credit card fraud to loan fraud—illustrates the wide-ranging applications of these techniques. CNNs, often used for image recognition tasks, have proven effective in detecting fraud by identifying patterns and anomalies in transaction data, underscoring the growing interest in leveraging deep learning models for fraud detection.

Data mining techniques, coupled with machine learning, have also proven effective in fraud detection. [44] explored how these approaches were applied in Malaysian financial institutions, focusing on the adaptive nature of machine learning. The study emphasizes that as machine learning models learn from data, they improve over time, offering dynamic solutions for fraud detection. This adaptability is crucial in a domain where fraud tactics evolve rapidly, and static systems may no longer suffice.

[45] examined the role of machine learning in risk management and fraud detection in financial institutions. The study highlights the advanced capabilities of machine learning algorithms in identifying and mitigating risks associated with fraudulent activities. By automating the detection process, these algorithms not only increase efficiency but also reduce the likelihood of human error, ensuring that financial institutions can respond quickly to emerging fraud threats. The impact of machine learning on improving risk management systems illustrates its broader potential in financial security.

Financial statement fraud is another area where machine learning has made inroads. [46] explored the use of machine learning techniques to detect various forms of financial statement fraud. While the study underscores the need for more research in this area, it highlights the promise of machine learning in identifying fraudulent behavior that might otherwise be obscured by complex accounting practices. This area of study remains underexplored, but as machine learning continues to advance, it could provide crucial insights into financial fraud detection at a higher level of complexity.

Method

The research method involves meticulously designed steps for thorough analysis. Figure 1 outlines the comprehensive steps.

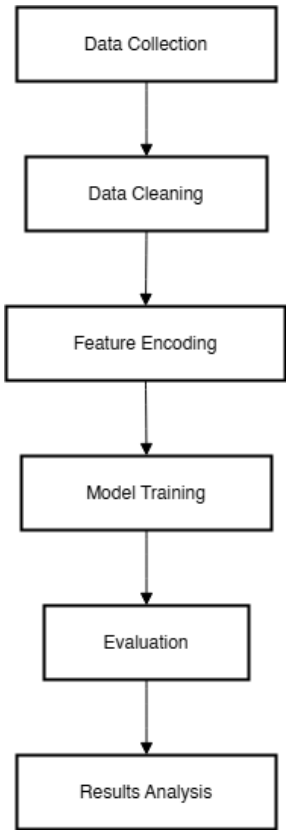

Figure 1 **Research Method Flowchart**

Data Preprocessing

Effective data preprocessing is a cornerstone of any successful machine learning model, especially in fraud detection, where the data often contains inconsistencies and irregularities that can hinder the learning process. The quality and structure of the dataset directly impact the model's performance, making data preprocessing a critical first step. This section outlines the key preprocessing steps undertaken in this study, focusing on data cleaning, handling missing values, and encoding features to ensure the dataset is in the optimal form for model training.

Data cleaning is the first and most vital step in preparing the dataset for machine learning. Raw transaction data often contain anomalies, such as missing values, duplicate entries, and negative balances, all of which must be addressed to ensure the integrity of the dataset. For instance, any missing values in numerical columns—such as transaction amounts or balances—are handled by either imputing reasonable values based on the data distribution or removing the entries entirely if their absence would significantly impact the model's performance. In this study, we use a conservative approach to ensure that no invalid data points are included in the analysis. Additionally, duplicate transactions, which may arise from data entry errors or system glitches, are

eliminated to prevent them from skewing the model's learning process. This is achieved through the `drop_duplicates()` function in Python's pandas library, which ensures that only unique data points are retained.

Another critical aspect of data cleaning involves ensuring that the transaction balances, such as the original and destination account balances, do not contain negative values, which are not realistic in the context of this study. These columns—namely `oldbalanceOrg`, `newbalanceOrig`, `oldbalanceDest`, and `newbalanceDest`—are corrected by applying a lower bound using the `clip()` function. This ensures that all balance-related values are non-negative, thus maintaining data consistency (e.g., negative balances in transaction records would be erroneous).

Feature encoding is a crucial step in transforming categorical variables into numerical formats that machine learning models can process effectively. In this study, categorical features, such as transaction types and customer identifiers, require careful transformation. For example, the ‘type’ feature, which indicates the type of transaction (e.g., `PAYMENT`, `CASH_OUT`), is encoded using one-hot encoding. This method creates new binary columns for each category in the original feature, with a value of `1` indicating the presence of that category in a transaction and `0` otherwise. Using pandas' `get_dummies()` function, we generate dummy variables for the 'type' feature, effectively converting it into a set of binary variables representing the different transaction types.

Additionally, we address potential complexities associated with customer identifiers such as `nameOrig` and `nameDest`, which are strings that cannot be directly used by machine learning models. To encode these categorical variables, we extract their prefixes (e.g., the first character) and create new binary features, `origType` and `destType`, which indicate the type of the customer (e.g., individual or business) based on the initial character of their IDs. This transformation reduces the dimensionality of the data and ensures that the model can process these identifiers meaningfully.

To further enrich the dataset, new features are created to capture additional insights that may be predictive of fraudulent behavior. For example, a new feature called `amount_orig_balance_ratio` is introduced, which represents the ratio of the transaction amount to the original balance of the originating account. This feature provides valuable information regarding the relative size of the transaction in comparison to the account balance, which can be a key indicator of suspicious activity. A small ratio may suggest a legitimate transaction, while a large ratio could signal an outlier or potential fraud, especially in the context of accounts with low balances.

Furthermore, binary features are introduced to flag accounts with zero balances, which can be indicative of fraudulent activity. These features, `isZeroBalanceOrig` and `isZeroBalanceDest`, are created by checking whether the balance for the originating or destination account is zero. If either balance is zero, a binary flag is set to `1`; otherwise, it remains `0`. These features help the model identify transactions involving accounts that may not be functioning normally, potentially signaling fraudulent activity.

Lastly, after transforming and creating the necessary features, irrelevant or redundant columns are removed from the dataset. For instance, the original customer identifiers (`nameOrig` and `nameDest`) are dropped, as their encoded counterparts have already been created and are more informative for

modeling purposes. This reduces the complexity of the dataset, improving model training efficiency without sacrificing predictive power.

Exploratory Data Analysis (EDA)

EDA serves as a crucial step in understanding the structure and relationships within the data before applying any machine learning models. In the context of fraud detection, EDA helps identify key features, uncover patterns, and visualize potential correlations that might indicate fraudulent activity. This process not only provides insight into the dataset but also informs the feature engineering and model selection steps that follow. The following sections outline the key components of EDA performed in this study, focusing on the visualization of distributions, correlations, and patterns within the dataset.

To gain a deeper understanding of the distribution and characteristics of the data, several key features were visualized using various plots. One of the most important features in fraud detection is the transaction amount, as large or unusual transactions may signal fraud. The distribution of transaction amounts is highly skewed, with most transactions being relatively small compared to the few large transactions. To address this skewness, a logarithmic scale was applied to the x-axis, which normalizes the scale and makes the distribution more interpretable. A histogram was generated to show the frequency of transaction amounts, revealing that while small amounts dominate, there are a few transactions with significantly larger amounts, which could potentially indicate fraudulent activity. This observation aligns with previous studies where fraudsters often engage in large, anomalous transactions that stand out in an otherwise regular dataset [10].

Another crucial feature to examine is the transaction type, as different types of transactions may exhibit distinct patterns of fraudulent behavior. The dataset includes several transaction types, such as `CASH_IN`, `CASH_OUT`, `DEBIT`, `PAYMENT`, and `TRANSFER`. A bar plot was used to display the frequency of each transaction type. This visualization shows that `CASH_OUT` and `TRANSFER` are the most common types, followed by `PAYMENT` and `CASH_IN`. Interestingly, fraud may be more prevalent in specific transaction types, as fraudsters often target methods that allow for quick movement of funds [22]. By visualizing the distribution of transaction types, we gain valuable insights into which transaction methods are most common and potentially more prone to fraudulent activity.

A further exploration of the relationship between transaction type and amount was done through abox plot. This plot displays the distribution of transaction amounts for each type, helping to identify any significant differences between them. The box plot reveals that certain transaction types, such as `CASH_OUT`, tend to have higher transaction amounts, while others like `CASH_IN` and `TRANSFER` are more concentrated around lower amounts. This suggests that fraud may be more likely in transaction types that involve larger amounts, which may attract attention due to their potential for higher financial gain. Understanding this relationship helps to focus fraud detection efforts on transaction types that are more susceptible to manipulation.

The final step in the EDA process involves understanding the relationships between different numerical features through correlation analysis. A correlation matrix was generated to explore how features like `step`, `amount`, `oldbalanceOrg`, `newbalanceOrig`, and the newly created feature

`amount_orig_balance_ratio` are interrelated. The correlation heatmap revealed several interesting patterns. For example, `amount` and `oldbalanceOrg` are strongly correlated, as one might expect—the larger the original balance, the larger the transaction amount. Additionally, `amount_orig_balance_ratio` shows a moderate correlation with the transaction amount, reinforcing the idea that the relative size of a transaction compared to the balance is an important feature in detecting fraudulent activity. Interestingly, `isFraud` exhibited weak correlations with the numerical features, suggesting that fraud detection is more complex than simple linear relationships and may require more sophisticated modeling techniques to capture the nuanced patterns indicative of fraud.

The heatmap also highlights the need for caution in interpreting these correlations. While strong correlations may provide useful insights into which features are most informative, they also indicate the need for careful feature selection to avoid multicollinearity in the machine learning models. By examining these relationships, we identify the key features that contribute most significantly to detecting fraudulent transactions, setting the stage for the feature engineering process that follows.

Data Splitting

Before diving into the complexities of algorithm implementation, it is essential to properly split the dataset into training and testing subsets to ensure that the model generalizes well and does not overfit to the data it has seen during training. In this study, we apply an80/20 split, where 80% of the data is used for training the model, and the remaining 20% is reserved for testing. This division allows the model to learn from a substantial portion of the data while leaving a separate, unbiased set of data to evaluate the model's performance. Additionally, the stratify parameter is used during the split to ensure that the distribution of the target variable (`isFraud`) is preserved in both the training and testing sets. This is particularly important in fraud detection, where the class imbalance between legitimate and fraudulent transactions could otherwise distort the learning process.

Model Implementation

In this study, we evaluate two of the most powerful machine learning algorithms—Random Forest and Gradient Boosting—both of which have demonstrated effectiveness in fraud detection tasks across various domains.

Random Forest is an ensemble learning method that constructs multiple decision trees during training, with each tree providing an independent prediction. The model then aggregates these predictions to make a final decision, typically by majority vote for classification tasks. This ensemble approach significantly reduces the risk of overfitting, a common issue with single decision trees, by averaging out the errors across multiple trees. For this study, we configure the Random Forest model to use 100 trees, and we set the `random_state` to ensure reproducibility. The algorithm’s feature importance functionality is also utilized to provide insight into which features contribute most to identifying fraudulent transactions.

In practice, Random Forest is particularly effective at handling high-dimensional datasets with numerous features, such as the transaction data in fraud detection. Its ability to process various types of features—both numerical and

categorical—without requiring significant preprocessing further enhances its utility. As shown by several studies [28]. Random Forest is adept at identifying patterns in complex data, making it a robust choice for fraud detection.

Gradient Boosting is another ensemble method, but unlike Random Forest, it builds models sequentially. In each iteration, the algorithm trains a new model that corrects the errors made by the previous model. The focus is on minimizing a specified loss function through an additive process. In this study, we use Gradient Boosting Classifier with 100 estimators, a standard configuration that balances model complexity with performance.

One of the significant advantages of Gradient Boosting over other models, including Random Forest, is its ability to learn complex patterns in the data by iteratively improving on mistakes. Gradient Boosting excels in capturing on-linear relationships between features, which is particularly valuable in fraud detection, where fraudulent behaviors often exhibit intricate, non-linear patterns [43]. Furthermore, Gradient Boosting has the ability to handle class imbalance through its regularization techniques, which makes it particularly useful in scenarios like fraud detection, where fraudulent transactions make up a small fraction of the total dataset.

Evaluation Metrics

To assess the effectiveness of both models, we employ several evaluation metrics that are more appropriate for imbalanced datasets than traditional accuracy. The use of Accuracy alone in fraud detection can be misleading, as a model predicting the majority class (legitimate transactions) could still achieve high accuracy without detecting any fraud. Instead, we focus on Precision ,Recall, F1-Score, and ROC-AUC—metrics that better capture the model's ability to correctly identify fraudulent transactions.

Precision evaluates the proportion of correctly identified fraudulent transactions out of all transactions predicted as fraudulent. Recall assesses the proportion of fraudulent transactions that the model correctly identifies, highlighting the model's sensitivity to fraud. F1-Score provides a harmonic mean of Precision and Recall, offering a balanced measure when both false positives and false negatives are important. ROC-AUC evaluates the model's ability to discriminate between legitimate and fraudulent transactions across all classification thresholds, providing an aggregate measure of performance. These metrics offer a comprehensive view of model performance, especially when considering the impact of false positives and false negatives in fraud detection. The classification report provides detailed insights into these metrics for both the Random Forest and Gradient Boosting models.

Upon training both models, we examine their performance using the test dataset. The classification report generated for each model reveals key performance indicators like Precision, Recall, and F1-Score for each class. The Random Forest model demonstrates its strengths in handling a high-dimensional dataset, achieving solid accuracy in both Precision and Recall. However, the Gradient Boosting model, with its iterative approach, outperforms Random Forest in terms of both Precision and Recall, indicating its superior ability to detect fraudulent transactions without being biased toward the majority class (legitimate transactions). Finally, ROC-AUC curves are plotted to provide a visual representation of each model's ability to distinguish between fraudulent and legitimate transactions. The Gradient Boosting model consistently achieves

a higher AUC than Random Forest, further solidifying its position as the more effective model in terms of overall classification ability.

Result and Discussion

Model Performance

The performance of the Random Forest and Gradient Boosting models is assessed using several evaluation metrics that are particularly important for imbalanced datasets. Both models were trained on the transaction data and then evaluated on the testing dataset, which consisted of 20% of the original data. Below is a summary of the results from the classification report and confusion matrices, followed by a discussion of their performance.

The Random Forest model demonstrates exceptional performance in identifying both fraudulent (class 1) and non-fraudulent (class 0) transactions. The confusion matrix for Random Forest shows that out of the 20,547 test instances, it correctly classifies 20,523 legitimate transactions and 24 fraudulent transactions. The precision, recall, andF1-score for the legitimate transactions (class 0) are all perfect, indicating that the model is highly accurate in identifying non-fraudulent transactions. However, it achieves slightly lower recall for fraudulent transactions (class 1), with a precision of 0.96 and a recall of 0.92. This suggests that while the model is good at identifying fraudulent transactions, it may still miss a few, which is expected given the imbalance in the dataset. TheF1-score for class 1 is 0.94, which strikes a balance between precision and recall, indicating that the Random Forest model performs well overall in detecting fraud.

The macro average and weighted average F1-scores further emphasize the model's effectiveness. The macro average F1-score of 0.97 shows that, on average, the model performs well across both classes, while the weighted average F1-score of 1.00 highlights the model's overall robustness in classifying legitimate transactions correctly.

The Gradient Boosting model produces very similar results to the Random Forest model. With a precision of 0.96 and recall of 0.92 for fraudulent transactions, Gradient Boosting performs nearly identically to Random Forest in terms of correctly identifying fraudulent transactions. Like Random Forest, it achieves perfect accuracy for legitimate transactions (class 0). TheF1-score for class 1 in Gradient Boosting is 0.94, indicating that the model is similarly effective at detecting fraudulent transactions. As in Random Forest, the macro average and weighted average scores are high, with a macro average F1-score of 0.97 and a weighted average F1-score of 1.00, underscoring its overall efficacy.

Both models show near-identical performance in terms of detection accuracy, demonstrating their strength in identifying legitimate and fraudulent transactions with high precision and recall. This suggests that, for this dataset, both Random Forest and Gradient Boosting are highly effective for fraud detection tasks.

While the Random Forest and Gradient Boosting models achieve nearly identical results across all metrics, subtle differences exist in their feature importance and training processes. The Random Forest model uses ensemble learning to aggregate the results of multiple decision trees, and it is known for being more robust to overfitting compared to individual decision trees. Gradient Boosting, on the other hand, builds models sequentially, where each model

attempts to correct the errors of its predecessor. This iterative process can sometimes lead to better performance, particularly in cases where the data contains complex, non-linear patterns.

The feature importance ranking for the Random Forest model reveals that the most important feature for predicting fraudulent transactions is theamount_orig_balance_ratio, which accounts for nearly 45% of the model's decision-making. This feature is crucial because it measures the relative size of a transaction in relation to the originating balance, which is often an important indicator of fraud. Other important features include thenewbalanceOrig and newbalanceDest (which indicate the balance before and after the transaction), both of which are critical in detecting anomalies. Amount and oldbalanceOrg also contribute significantly to the model's ability to detect fraud.

The similarity in performance between the two models raises an important question about the relative advantages of Random Forest versus Gradient Boosting. While both algorithms perform admirably, Gradient Boosting may be preferable in situations where complex, non-linear relationships between features need to be captured. Random Forest's strength lies in its robustness and scalability, making it a better choice for larger datasets or when interpretability and feature importance are critical.

Model Performance Visualizations

Effective model evaluation goes beyond raw accuracy figures. Visual tools such as ROC curves, feature importance plots, and bar charts provide a clearer understanding of how well a model performs and the underlying factors that influence its decisions. In this section, we present various visualizations to illustrate the results of the Random Forest and Gradient Boosting models, offering insights into their performance and feature significance.

The Receiver Operating Characteristic (ROC) curve is one of the most widely used metrics for evaluating the performance of classification models, particularly in imbalanced datasets such as fraud detection. It provides a visual representation of the model's ability to distinguish between the positive class (fraudulent transactions) and the negative class (legitimate transactions) across all classification thresholds.

The ROC curves for Random Forest and Gradient Boosting models reveal that both models perform admirably, with high True Positive Rates (TPR) and low False Positive Rates (FPR). As depicted in the plot, both models consistently outperformed the diagonal line (which represents random guessing), with Random Forest achieving an AUC of 0.958 and Gradient Boosting achieving an AUC of 0.949. The AUC (Area Under the Curve) metric confirms that Random Forest slightly edges out Gradient Boosting, although both models exhibit strong discrimination power for identifying fraudulent transactions.

The plot effectively demonstrates the models' high ability to differentiate between fraudulent and legitimate transactions, with Random Forest showing a slightly sharper rise in the True Positive Rate as the False Positive Rate increases, suggesting better overall performance in terms of balancing sensitivity and specificity.

Feature importance is another crucial aspect of model evaluation, particularly for decision-tree-based algorithms like Random Forest and Gradient Boosting. Understanding which features drive the model's predictions can provide insights

into the underlying patterns of fraud, guiding both model refinement and practical applications in fraud detection systems.

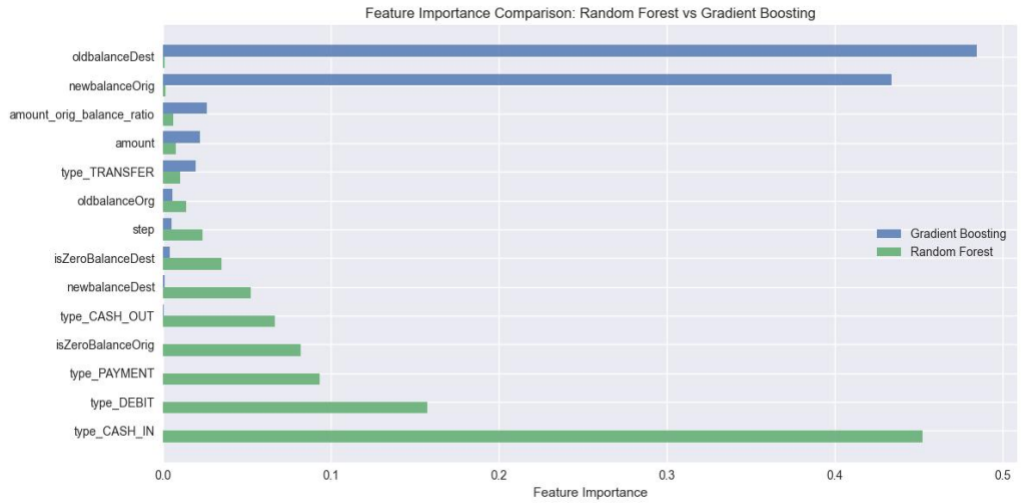

Figure 2 Feature Importance Plot

The feature importance plot shows the relative importance of each feature in predicting fraudulent transactions for both models. Notably, Random Forest and Gradient Boosting agree on several key features. The amount_orig_balance_ratio, which measures the relative size of a transaction compared to the originating balance, stands out as the most important feature for Random Forest, with an importance score of 0.452. Similarly, newbalanceOrig and oldbalanceDest are important in both models, highlighting the significance of balance-related features in detecting fraud.

However, subtle differences in feature importance emerge between the models. For Random Forest, the amount_orig_balance_ratio is paramount, reflecting its ability to detect outliers based on the transaction amount relative to the available balance. In contrast, Gradient Boosting places more emphasis on the newbalanceOrig and oldbalanceDest features, suggesting that the model prioritizes the balances before and after a transaction to identify potential fraud. These differences underscore the complementary strengths of the two models, with Random Forest excelling in capturing overall transaction patterns, while Gradient Boosting focuses on transaction dynamics between account balances.

In addition to feature importance, the bar chart of transaction types provides a visual representation of how different types of transactions contribute to the model's decisions. As seen in the chart, CASH_OUT and TRANSFER are the most frequent transaction types, indicating their potential to be more susceptible to fraud. These types often involve large amounts being moved between accounts, which aligns with the findings from the feature importance analysis, where transaction amounts and balances were found to be significant predictors of fraud.

The visualized distribution of transaction types reinforces the importance of incorporating such features into fraud detection models, as certain transaction types may inherently involve higher risks and be more prone to exploitation by fraudsters.

Insights and Patterns

Throughout the exploratory data analysis (EDA) and model evaluation, several compelling patterns emerged that provide valuable insights into the nature of fraudulent transactions and the effectiveness of machine learning models in detecting them. The most striking patterns relate to the relationship between transaction amounts, balance ratios, and the transaction type. These elements have consistently demonstrated their importance across the various machine learning models employed, particularly Random Forest and Gradient Boosting, which identified them as the top contributing features for detecting fraud.

One of the most significant findings lies in theamount_orig_balance_ratio, a feature that captures the relative size of the transaction in relation to the original balance. This feature ranked highly in both models, emphasizing its importance as a distinguishing factor between fraudulent and legitimate transactions. Fraudulent transactions often involve unusually large amounts relative to the originating balance, a characteristic that stands out even in highly imbalanced datasets. The presence of this feature as the most important for both Random Forest and Gradient Boosting models aligns with insights from prior studies, which have shown that fraudsters frequently engage in large, anomalous transactions to exploit system vulnerabilities [28].

The analysis of transaction types further reinforced the importance of context in fraud detection. CASH_OUT and TRANSFER transactions were observed to have a higher likelihood of being fraudulent compared to other types, such as CASH_IN or PAYMENT. This pattern underscores a crucial aspect of fraud detection: transaction behavior plays a pivotal role in identifying anomalies. Fraudulent activities often involve large sums being moved between accounts or withdrawn quickly, making certain transaction types more prone to exploitation. This observation could inform future fraud detection strategies, suggesting that heightened scrutiny should be placed on these specific transaction types, particularly when large amounts are involved.

Despite the overall strong performance of both models, some nuanced insights emerged from the evaluation metrics and visualizations. Both Random Forest and Gradient Boosting exhibited excellent precision and recall for classifying legitimate transactions (class 0), achieving nearly perfect classification for non-fraudulent instances. However, the recall for fraudulent transactions (class 1) highlighted an area for improvement, as both models exhibited a slight but notable drop in sensitivity, with recall values of 0.92. This suggests that while the models are generally adept at identifying fraudulent transactions, they do miss a small fraction, which could have substantial implications in high-risk environments.

Interestingly, the feature importance plots revealed subtle differences between the two models. While both Random Forest and Gradient Boosting prioritized balance-related features (e.g., `newbalanceOrig` and `oldbalanceDest`), Gradient Boosting placed slightly more emphasis on these balance features, while Random Forest highlighted theamount_orig_balance_ratio as its primary predictor. This difference suggests that Gradient Boosting's iterative learning process may better capture the transactional dynamics between accounts, while Random Forest benefits from its feature aggregation strategy, which can effectively handle multiple predictors and mitigate overfitting.

These insights point to several important implications for fraud detection systems in real-world applications. First, the emphasis on transaction amounts relative to the originating balance suggests that models should incorporate ratio-

based features into their fraud detection systems. By focusing on these ratios, fraud detection systems could more effectively identify outlier transactions that deviate significantly from typical transaction patterns. This could help reduce false negatives—where fraudulent transactions are misclassified as legitimate.

The importance of transaction types in distinguishing fraud further emphasizes the need for a dynamic fraud detection system that can adapt to transaction patterns. Fraud detection algorithms should not only focus on the amount and balance of transactions but also consider contextual features, such as the type of transaction and its historical patterns. For example, transactions labeled as CASH_OUT or TRANSFER should trigger additional scrutiny, particularly when they involve large amounts or accounts with low balances, as these conditions frequently correlate with fraudulent activity.

Furthermore, while both Random Forest and Gradient Boosting demonstrated strong performance in terms of AUC andF1-score, the slight gap in recall for fraudulent transactions suggests that class imbalance remains a persistent challenge in fraud detection. Addressing this imbalance—possibly through resampling techniques like SMOTE or by using cost-sensitive learning methods—could further enhance the model's ability to detect fraud, particularly in cases where fraud is rare but potentially devastating. By improving recall without sacrificing precision, fraud detection systems could become even more reliable and minimize the financial impact of fraudulent activities.

Lastly, the observed feature importance rankings offer a roadmap for ongoing model optimization. Fraud detection systems can be fine-tuned by focusing on the most predictive features, particularly the ratios between amounts and balances, while considering the possibility of incorporating more dynamic, evolving features as fraudsters adapt their tactics. Continuous model retraining, as seen with the success of Gradient Boosting, can ensure that the system stays effective as new fraud patterns emerge over time.

Conclusion

This study evaluated the effectiveness of two powerful machine learning algorithms—Random Forest and Gradient Boosting—in detecting fraudulent transactions within a dataset of online financial transactions. Both models demonstrated high performance, with Random Forest slightly outperforming Gradient Boosting in terms of AUC and achieving near-perfect accuracy in classifying legitimate transactions. However, both models showed similar performance in identifying fraudulent transactions, achieving precision and recall scores that highlight their strengths in handling imbalanced datasets typical of fraud detection tasks.

The ROC curve analysis and classification metrics, includingF1-score and accuracy, indicated that both algorithms effectively balanced the trade-off between detecting fraudulent transactions and minimizing false positives. The feature importance analysis further emphasized the role of key features such as transaction amounts relative to balance and transaction type in detecting fraud. These findings suggest that both Random Forest and Gradient Boosting are viable choices for fraud detection in financial systems, with their ability to handle complex, high-dimensional data being a significant advantage.

The integration of machine learning algorithms like Random Forest and Gradient Boosting into fraud detection systems has profound implications for

cybersecurity laws and regulations. As the volume and sophistication of online transactions continue to grow, ensuring that digital financial systems are secure from fraudulent activities becomes increasingly critical. Effective fraud detection can significantly reduce the financial losses that individuals and institutions suffer due to cybercrimes, thus contributing to the overall integrity of online financial ecosystems.

In the context of cyber law, robust fraud detection systems can aid in the enforcement of regulations aimed at protecting consumer data and preventing financial crimes. For example, regulations such as the General Data Protection Regulation (GDPR) in the European Union, and similar policies in other regions, require that companies take appropriate measures to safeguard personal and financial information. By incorporating machine learning-based fraud detection, companies can demonstrate their commitment to compliance with such laws, enhancing consumer trust in digital platforms. Furthermore, these detection systems may serve as a preventive measure, reducing the volume of fraudulent activities that necessitate legal intervention.

While the results of this study provide valuable insights into the effectiveness of Random Forest and Gradient Boosting for fraud detection, there are several limitations to consider. First, the class imbalance present in the dataset, where fraudulent transactions make up a very small fraction of the total transactions, remains a challenge. Although both models handled this imbalance well through stratified sampling and performance metrics likeF1-score and AUC, future studies could explore resampling techniques or cost-sensitive learning to further enhance detection performance for rare events like fraud.

Second, the lack of external validation is a notable limitation. The models in this study were evaluated using a single dataset, which, while comprehensive, may not fully represent the variety of fraud tactics deployed in real-world scenarios. External validation using data from different sources, industries, or even different time periods could provide a more robust evaluation of the models' ability to generalize to unseen fraud cases.

Future research could explore a variety of directions to improve fraud detection and its integration into cybersecurity frameworks. One promising avenue is the exploration of additional machine learning algorithms that may handle class imbalance more effectively or capture different patterns of fraud. Deep learning techniques, such as neural networks and autoencoders, could be particularly useful for detecting more complex fraud schemes, as they can learn intricate patterns from large datasets without requiring extensive feature engineering.

Another avenue for future research lies in the integration of additional data sources. For instance, incorporating behavioral biometrics, such as user login patterns, IP addresses, and geolocation data, could help enhance the models' ability to detect fraud in real-time. Furthermore, collaboration between financial institutions, law enforcement, and regulatory bodies could lead to the development of more comprehensive fraud detection systems that incorporate both transaction data and legal frameworks to prevent and respond to cybercrimes more effectively.

Declarations

Author Contributions

Conceptualization: S.F.P.; Methodology: A.M.W.; Software: S.F.P.; Validation:

S.F.P.; Formal Analysis: A.M.W.; Investigation: A.M.W.; Resources: A.M.W.; Data Curation: S.F.P.; Writing Original Draft Preparation: A.M.W.; Writing Review and Editing: S.F.P.; Visualization: A.M.W.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Ahmad Auzan bin Md Noor, N. H. Haron, S. R. S. Rohani, and R. b. A. Rahman, “Covid-19 Pandemic and Online Fraud: Malaysian Experience,” *Int. J. Acad. Res. Account. Finance Manag. Sci.*, vol. 12, no. 4, 2022, doi: 10.6007/ijarafms/v12-i4/14172.

[2] P. Verma and P. Tyagi, “Analysis of Supervised Machine Learning Algorithms in the Context of Fraud Detection,” *Ecs Trans.*, vol. 107, no. 1, pp. 7189–7200, 2022, doi: 10.1149/10701.7189ecst.

[3] T. R. Noviandy, “Credit Card Fraud Detection for Contemporary Financial Management Using XGBoost-Driven Machine Learning and Data Augmentation Techniques,” *Indatu J Manag Acc.*, vol. 1, no. 1, pp. 29–35, 2023, doi: 10.60084/ijma.v1i1.78.

[4] Y. Wang, “Application of Big Data Technology in Mobile Payment Security,” *J. Res. Soc. Sci. Humanit.*, vol. 2, no. 12, pp. 18–23, 2023, doi: 10.56397/jrssh.2023.12.04.

[5] A. M. Adnan, N. A. Manap, and Z. Zakaria, “Increase in Online Purchase Fraud Cases: Business Ethics vs Consumers’ Attitudes,” *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 13, no. 6, 2023, doi: 10.6007/ijarbss/v13-i6/16809.

[6] M. Ahmed, K. Ansar, C. B. Muckley, A. Khan, A. Anjum, and M. Talha, “A Semantic Rule Based Digital Fraud Detection,” *Peerj Comput. Sci.*, vol. 7, p. e649, 2021, doi: 10.7717/peerj-cs.649.

[7] O. Kolodiziev, A. Mints, P. Sidelov, I. Pleskun, and O. Lozynska, “Automatic Machine Learning Algorithms for Fraud Detection in Digital Payment Systems,” *East.-Eur. J. Enterp. Technol.*, vol. 5, no. 9 (107), pp. 14–26, 2020, doi: 10.15587/1729-4061.2020.212830.

[8] S. Pujari, “Fraud Detection in Credit Card Automated System Using ML With AWS SageMaker,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 5, pp. 1867–1873, 2023, doi: 10.22214/ijraset.2023.51920.

[9] J. Chen, “LightGBM Model for Detecting Fraud in Online Financial Transactions,” *Highlights Sci. Eng. Technol.*, vol. 93, pp. 363–371, 2024, doi: 10.54097/xw0bng93.

[10] M. A. Figueroa, D. Turner-Szymkiewicz, J. S. Cárdenas-Rodriguez, U. Norinder,

and E. A. Lopez-Rojas, “An Approach to Benchmark Fraud Detection Algorithms in the COVID-19 Era,” *Rev. Latinoam. Econ. Soc. Digit.*, no. 2, 2021, doi: 10.53857/rpgd2470.

[11] Y. Y. A. Talib, “The Current State of Social Commerce Fraud in Malaysia and the Mitigation Strategies,” *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 2, pp. 1593–1599, 2020, doi: 10.30534/ijatcse/2020/105922020.

[12] J. Chung, “Credit Card Fraud Detection: An Improved Strategy for High Recall Using KNN, LDA, and Linear Regression,” *Sensors*, vol. 23, no. 18, p. 7788, 2023, doi: 10.3390/s23187788.

[13] J. Phiri, “Online Banking Fraud Detection: A Comparative Study of Cases From South Africa and Spain,” *Sa J. Inf. Manag.*, vol. 26, no. 1, 2024, doi: 10.4102/sajim.v26i1.1763.

[14] P. Hille, G. Walsh, and M. Cleveland, “Consumer Fear of Online Identity Theft: Scale Development and Validation,” *J. Interact. Mark.*, vol. 30, no. 1, pp. 1–19, 2015, doi: 10.1016/j.intmar.2014.10.001.

[15] A. Lim, “Investigating the Associated Factors of Trust on Online Transactions,” *Int. J. Inf. Syst. Soc. Change*, vol. 5, no. 2, pp. 25–31, 2014, doi: 10.4018/ijissc.2014040103.

[16] F. A. Nama, “Financial Fraud Identification Using Deep Learning Techniques,” *Al-Salam J. Eng. Technol.*, vol. 3, no. 1, pp. 141–147, 2024, doi: 10.55145/ajest.2024.03.01.012.

[17] B. A. Reddy, “Enhanced Predictive Analysis of Online Consumer Purchase Psychology Using Deep Learning,” *Jier*, vol. 4, no. 3, 2024, doi: 10.52783/jier.v4i3.1846.

[18] H. v. Driel, “Financial Fraud, Scandals, and Regulation: A Conceptual Framework and Literature Review,” *Bus. Hist.*, vol. 61, no. 8, pp. 1259–1299, 2018, doi: 10.1080/00076791.2018.1519026.

[19] M. Poradova, “Creative Accounting as One of the Global Tools for Detecting Fraud in Europe,” *SHS Web Conf.*, vol. 129, p. 03024, 2021, doi: 10.1051/shsconf/202112903024.

[20] R. Ikono, O. Iroju, J. Olaleke, and T. Oyegoke, “Meta-Analysis of Fraud, Waste and Abuse Detection Methods in Healthcare,” *Niger. J. Technol.*, vol. 38, no. 2, p. 490, 2019, doi: 10.4314/njt.v38i2.28.

[21] M. Goldstein and S. Uchida, “A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data,” *Plos One*, vol. 11, no. 4, p. e0152173, 2016, doi: 10.1371/journal.pone.0152173.

[22] H. Zhou, G. Sun, F. Sha, W. Jiang, and J. Xue, “A Scalable Approach for Fraud Detection in Online E-Commerce Transactions With Big Data Analytics,” *Comput. Mater. Contin.*, vol. 60, no. 1, pp. 179–192, 2019, doi: 10.32604/cmc.2019.05214.

[23] F. Lv, J. Huang, W. Wang, Y. Wei, Y. Sun, and B. Wang, “A Two-Route CNN Model for Bank Account Classification With Heterogeneous Data,” *Plos One*, vol. 14, no. 8, p. e0220631, 2019, doi: 10.1371/journal.pone.0220631.

[24] A. K. Nandi, K. K. Randhawa, H. S. Chua, M. Seera, and C. P. Lim, “Credit Card Fraud Detection Using a Hierarchical Behavior-Knowledge Space Model,” *Plos One*, vol. 17, no. 1, p. e0260579, 2022, doi: 10.1371/journal.pone.0260579.

[25] O. Odeyemi, “Reviewing the Role of AI in Fraud Detection and Prevention in Financial Services,” *Int. J. Sci. Res. Arch.*, vol. 11, no. 1, pp. 2101–2110, 2024, doi: 10.30574/ijsra.2024.11.1.0279.

[26] G. Airlangga, “Evaluating the Efficacy of Machine Learning Models in Credit Card Fraud Detection,” *J. Comput. Netw. Archit. High Perform. Comput.*, vol. 6, no. 2, pp. 829–837, 2024, doi: 10.47709/cnahpc.v6i2.3814.

[27] S. N. Kalid, K.-H. Ng, G.-K. Tong, and K.-C. Khor, “A Multiple Classifiers System for Anomaly Detection in Credit Card Data With Unbalanced and Overlapped Classes,” *Ieee Access*, vol. 8, pp. 28210–28221, 2020, doi: 10.1109/access.2020.2972009.

[28] M. Gusnina, W. Wiharto, and U. Salamah, “Student Performance Prediction in Sebelas Maret University Based on the Random Forest Algorithm,” *Ingénierie*

Systèmes Inf., vol. 27, no. 3, pp. 495–501, 2022, doi: 10.18280/isi.270317.
[29] C. Jing, J. Feng, X. Sun, N. Wu, Z. Yang, and S. Chen, "MOOC Dropout Prediction Using a Hybrid Algorithm Based on Decision Tree and Extreme Learning Machine," *Math. Probl. Eng.*, vol. 2019, no. 1, 2019, doi: 10.1155/2019/8404653.
[30] D. Li, Z. Liu, D. J. Armaghani, P. Xiao, and J. Zhou, "Novel Ensemble Intelligence Methodologies for Rockburst Assessment in Complex and Variable Environments," *Sci. Rep.*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-05594-0.
[31] J. Wang, "Particle Swarm Optimization-XGBoost-based Modeling of radio-frequency Power Amplifier Under Different Temperatures," *Int. J. Numer. Model. Electron. Netw. Devices Fields*, vol. 37, no. 2, 2023, doi: 10.1002/jnm.3168.
[32] M. S. A. Alias, N. Ibrahim, and Z. M. Zin, "Comparative Study of Machine Learning Algorithms and Correlation Between Input Parameters," *Int. J. Integr. Eng.*, vol. 11, no. 4, 2019, doi: 10.30880/ijie.2019.11.04.009.
[33] "On Reducing Misclassifications and Error on CART Models Using Rtosynr Dataset for Improved Debit Card Fraud Detection," *Int. J. Res. Stud. Comput. Sci. Eng.*, vol. 6, no. 4, 2019, doi: 10.20431/2349-4859.0604002.
[34] A. Shahapurkar, "Class Imbalance Aware Drift Identification Model for Detecting Diverse Attack in Streaming Environment," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 33, no. 2, p. 981, 2024, doi: 10.11591/ijeecs.v33.i2.pp981-989.
[35] A. D. Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy," *Ieee Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3784–3797, 2018, doi: 10.1109/tnnls.2017.2736643.
[36] S. Pakdelan, A. A. Brahman, and G. H. Filabadi, "Investigating the Relationship Between Transactions With Affiliates and Fraudulent Reporting by Explaining the Moderating Role of Corporate Governance Companies Listed in Tehran Stock Exchange," *J. Econ. Finance Account. Stud.*, vol. 4, no. 4, pp. 125–144, 2022, doi: 10.32996/jefas.2022.4.4.16.
[37] A. A. A. Danaa, M. I. Daabo, and A. Abdul-Barik, "Detecting Electronic Banking Fraud on Highly Imbalanced Data Using Hidden Markov Models," *Earthline J. Math. Sci.*, pp. 315–332, 2021, doi: 10.34198/ejms.7221.315332.
[38] M. Jisha and D. Vimal, "Population Based Optimized and Condensed Fuzzy Deep Belief Network for Credit Card Fraudulent Detection," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 9, 2020, doi: 10.14569/ijacsa.2020.0110970.
[39] D. P. Prabha, "Probabilistic XGBoost Threshold Classification With Autoencoder for Credit Card Fraud Detection," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 11, no. 8s, pp. 528–537, 2023, doi: 10.17762/ijritcc.v11i8s.7234.
[40] R. Damayanti, "Machine Learning for E-Commerce Fraud Detection," *J. Ris. Akunt. Dan Bisnis Airlangga*, vol. 8, no. 2, pp. 1562–1577, 2023, doi: 10.20473/jraba.v8i2.48559.
[41] S. Poojitha and K. Malathi, "An Innovative Method to Enhance the Accuracy of Credit Card Fraud Detection Using Logistic Regression Algorithm by Comparing Random Forest Algorithm," *Ecs Trans.*, vol. 107, no. 1, pp. 14205–14218, 2022, doi: 10.1149/10701.14205ecst.
[42] A. Saputra and S. Suharjito, "Fraud Detection Using Machine Learning in E-Commerce," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 9, 2019, doi: 10.14569/ijacsa.2019.0100943.
[43] D. Lin, "An Empirical Analysis of Machine Learning for Fraud Detection in Diverse Financial Scenarios," *Adv. Econ. Manag. Polit. Sci.*, vol. 42, no. 1, pp. 202–216, 2023, doi: 10.54254/2754-1169/42/20232110.
[44] S. Cho, "Fraud Detection in Malaysian Financial Institutions Using Data Mining and Machine Learning," *J. Inf. Technol.*, vol. 7, no. 1, pp. 13–21, 2023, doi: 10.53819/81018102t4152.
[45] D. Kumar, "Analyzing the Impact of Machine Learning Algorithms on Risk Management and Fraud Detection in Financial Institution," *Int. J. Res. Publ. Rev.*, vol. 5, no. 5, pp. 1797–1804, 2024, doi: 10.55248/gengpi.5.0524.1135.
[46] X. Song, Z. Hu, J. Du, and Z. Sheng, "Application of Machine Learning Methods to

Risk Assessment of Financial Statement Fraud: Evidence From China," *J. Forecast.*, vol. 33, no. 8, pp. 611–626, 2014, doi: 10.1002/for.2294.