THE ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 5. Issue 2 (2025)
https://doi.org/10.62019/abbdm.v5i2.332

Optimization of Fraud Detection Models for Safeguarding Customer Transactions

Muhammad Wajahat Ali, Waseemullah, Muhammad Qasim Memon, Muhammad Faraz Hyder, Aasma Memon

Chronicle

Article history
Received: Feb 19, 2025
Received in the revised format: March 28, 2025
Accepted: April 28, 2025
Available online: May 18, 2025

Muhammad Wajahat Ali & Waseemullah are currently affiliated with Department of Computer Science and Information Technology, NED University of Engineering and Technology Pakistan.
Email: ds.wajahat@gmail.com
Email: waseemu@cloud.neduet.edu.pk

Muhammad Qasim Memon is currently affiliated with the Department of Information and Computing, Computer Sciences, University of Sufism and Modern Sciences Bhitshah Pakistan.
Email: memon_kasim@usms.edu.pk

Muhammad Faraz Hyder is currently affiliated with the Department of Software Engineer, NED University of Engineering and Technology, Pakistan.
Email: farazh@neduet.edu.pk

Aasma Memon is currently affiliated with the Department of Business Administration, University of Sufism and Modern Sciences Bhitshah, Pakistan.
Email: kaasma.bjut@gmail.com

Corresponding Author*

Abstract

Fraud detection has become a critical task in various industries, particularly in financial transactions, as fraudulent activities continue to evolve and pose significant economic threats. In this study, the application of machine learning algorithms for the detection of fraudulent transactions in highly imbalanced datasets is explored. The fast-paced development of online transactions has prompted the creation of strong fraud-detection systems to protect customer transactions. This study delves into optimizing fraud-detection models through sophisticated machine-learning methods and ensemble techniques. We compared the performance of different classifiers, such as Logistic Regression, Random Forest, Decision Tree, Support Vector Classifier, and ensemble classifiers like Bagging, AdaBoost, and Gradient Boosting, in terms of accuracy, precision, recall, F1-Score, ROC-AUC, and log loss. Our findings show that ensemble classifiers, especially Bagging, AdaBoost, and Random Forest, perform near-perfect classification with AUC scores of 1.00, better than conventional classifiers. Further, we tackle the issue of class imbalance and underscore the significance of model generalization using stratified K-fold cross-validation. The results demonstrate that ensemble methods not only improve detection accuracy but also offer stable generalization, rendering them very effective for practical fraud detection purposes.

Keywords: Fraud detection, Machine Learning, Fraudulent activities, Ensemble Methods, Adversarial Attacks.

INTRODUCTION

The growth of electronic financial transactions has resulted in a staggering rise in fraudulent activities, which presents serious challenges to businesses, financial institutions, and consumers. Not only do fraudulent transactions cause heavy financial losses, but they also erode customer confidence and the integrity of electronic financial systems. To counter this emerging threat, fraud-detection models are essential for detecting and preventing fraudulent activities in real time. Nevertheless, standard fraud detection methods have high false positive rates, are computationally expensive, and cannot handle dynamic changes in fraud patterns. The optimization of fraud models is vital for improving accuracy, reducing false alarms, and providing

uninterrupted security for customer transactions (Al-Hashedi & Magalingam, 2021). The financial industry has undergone a transformation thanks to the growing usage of digital products, which provides businesses and customers with a degree of convenience and flexibility that was previously unattainable. Nevertheless, this shift has resulted in several problems, including electoral fraud (Wei et al., 2013). In a 2022 report, it was estimated that global e-commerce losses because of online payment fraud amounted to $41 billion and are set to increase to $48 billion in 2023 (Statista, n.d). Also, a report showed that 60% of e-commerce merchants and 53% of retailers saw an increase in overall levels of fraud. Fraudulent chargebacks were identified as the quickest-growing type of fraud by retailers, while e-commerce merchants identified identity theft as the quickest-growing threat (LexisNexis Risk Solutions, 2024). These figures highlight the huge monetary effect of online fraud on the retail sector, highlighting the necessity for effective fraud detection and prevention strategies. Several studies adopted the development of machine learning (ML) in the field of fraud detection presents a chance to circumvent these limitations, therein existing studies were fed historical data to find trends and connections, enabling real-time fraud detection (Ali et al., 2022).

Fraud detection has been a primary focus area for the banking, e-commerce, and financial technology (FinTech) sector for years. Rule-based approaches are commonly used in traditional fraud detection systems; wherein static heuristics are employed to identify suspicious transactions. These static methods are inadequate in detecting sophisticated and evolving fraud patterns. To overcome these drawbacks, machine learning (ML) and deep learning (DL) are being used with growing frequency for fraud detection. These models can process large-scale transactional data, discover hidden patterns, and identify anomalies with higher accuracy (Sadgali et al., 2019). Current innovations in detecting fraud involve ensemble learning, anomaly detection, and hybrid models, which blend supervised and unsupervised methods. New technologies like explainable AI (XAI), federated learning, and adversarial machine learning are also being researched to make fraud detection systems stronger and more transparent. Nevertheless, challenges persist such as performing real-time processing, dealing with imbalanced data sets, and meeting privacy requirements in strict regulations such as General Data Protection Regulation (GDPR).

These challenges still need to be fixed as machine learning advances. An example of an issue where the model is skewed, and fraudulent transactions are overlooked due to supposedly lower fraud rates than legitimate ones. Additionally, models need to be adjusted to account for new trends without needing a lot of iterations because idea drift affects the fraud domain. Our research aims to address these problems by increasing the precision and effectiveness of the models used in machine learning applications for fraud detection (Raghavan & El Gayar, 2019). Therefore, improving the capacity to identify fraud through machine learning approaches is the primary goal of this research work. The main factors contributing to our framework include enhanced feature engineering, robust cross-validation using K-Fold stratified validation, hyperparameter optimization with Grid Search, and handling of unbalanced data with methods like SMOTE (Chawla et al., 2002). Consequently, higher recall and precision because of these factors guarantee that fraudulent transactions are detected more successfully. In conclusion, our framework shows notable gains in both classification accuracy and the capacity to detect fraud, creating a new benchmark in the field. To achieve this, our paper attempts to

investigate and enhance fraud detection models to protect customer transactions from security threats. The major contributions of this research are:

- Comparing state-of-the-art machine learning and ensemble methods for detecting fraud.
- Presenting an improved hybrid approach with enhanced accuracy and minimized false positives.
- Mitigating problems such as class imbalance, adversarial attacks, and real-time detection through adaptive learning methods.
- Our framework provides solutions for greater transparency and trust in fraud detection outcomes.

The rest of the paper is organized as follows: Section 2 describes the related work. Section 3 explains the methodology in the current study. Section 4 presents experimental results. Section 5 elaborates the discussion regarding the insights of our framework. In the final section, concluding remarks are presented.

LITERATURE REVIEW

Several studies have investigated fraud detection models to enhance efficiency as well as precision. Studies have indicated that random forests, support vector machines (SVMs), deep neural networks (DNNs), and ensemble models perform better than classic algorithms in identifying fraudulent transactions. For instance, hybrid models that incorporate decision trees and gradient boosting have shown enhanced fraud detection rates while keeping false positives low (Kanksha et al., 2021). Moreover, research employing unsupervised anomaly detection methods, including autoencoders and generative adversarial networks (GANs), has also yielded encouraging results in detecting new fraud patterns without the need for large, labeled datasets (Hilal et al., 2022).

Despite these developments, fraudsters are constantly adapting their methods, using AI-facilitated attacks and adversarial methods to evade detection mechanisms (Kumar et al., 2024). The occurrence of imbalanced datasets, where fraudulent transactions are much smaller compared to genuine transactions, makes model training and assessment even more difficult. In addition to this, privacy issues and regulatory restrictions hinder institutions from exchanging transaction information for joint fraud detection (Makki et al., 2019).

Providing privacy-preserving fraud detection systems compliant with international data protection laws. The recent research examined various machine learning (ML) and deep learning models for fraud detection, and these models were assessed using accuracy, F1 Sore, and AUC-ROC (Kamuangu, 2024). According to this research, the accuracy of supervised learning models is 93% for Support Vector Machines (SVM), 94% for Decision Trees, and 95% for Gradient Boosting (GBM). On the other hand, SVM scores 91.29% accuracy, Decision Trees 96.35%, and our GBM 97.85.

Additionally, the growing number of digital transactions has been accompanied by an equivalent growth in fraudulent transactions, a situation that demands the creation of effective fraud detection frameworks. This literature review discusses recent innovations in fraud detection models based on machine learning approaches, imbalance alleviation methods, and novel frameworks for protecting customer transactions. The works reviewed are highly related to the context of maximizing fraud detection models. Recent research was conducted (Makki et al., 2019), who tackle the issue of unbalanced datasets in fraud detection, which is a

widespread problem with many legitimate transactions vastly outnumbering fraudulent ones (Al-dahasi et al., 2025). Their research presents a machine learning framework that incorporates imbalance reduction methods, including Synthetic Minority Oversampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN), to improve the identification of fraudulent financial transactions. The authors show that optimizing these methods enhances the precision and recall of fraud detection models to make them more effective when used in real-world. The research is closely related to optimizing fraud detection models, in that it accentuates the role of solving the problem of imbalance in data so as to gain accurate predictions.

In the same vein, Yan et al. (2024) concentrate on maximizing credit card fraud detection through adaptive model optimization. The authors adjust machine learning models dynamically in line with changing fraud patterns through the integration of real-time data and adaptive learning algorithms, minimizing the false positives and the false negatives by a wide margin. This work highlights the need for ongoing model optimization to stay ahead of the sophisticated methods used by fraudsters.

Amarnadh and Moparthi (2023) offer an extensive review of online payment fraud detection with machine learning methods. This research compares the performance of different algorithms, such as Random Forest, Gradient Boosting, and Neural Networks, in identifying fraudulent transactions. The authors highlight the importance of feature engineering and hyperparameter tuning in achieving optimal model performance. The results show that ensemble methods, which involve combining multiple models, provide the best accuracy in fraud detection. This research fits into the aim of maximizing fraud detection models through the proof of concept of advanced machine learning.

Chatterjee et al. (2024) discuss the use of digital twin technology for credit card fraud detection. A digital twin is a computerized replica of a physical system that can be used to simulate real-world conditions. The authors suggest a digital twin framework that constantly tracks transaction data and detects anomalies that point to fraud. This method provides the capability for real-time detection and mitigation of fraud. The research emphasizes the possibilities offered by digital twin technology to transform fraud detection by offering a dynamic and adaptive solution. This new technique is particularly pertinent to fraud detection model optimization, as it presents a novel paradigm for the protection of customer transactions.

Yadav et al. (2024) examine machine learning approaches and API implementations to visualize fraud detection within customer transactions (Yadav et al., 2024). Their contribution is centered on creating user-friendly interfaces that facilitate stakeholders to effectively monitor and analyze transactional data. Through the combination of machine learning models with visualization components, authors are offering a complete solution to detect fraud. This research highlights the need for the integration of technical and practical considerations to improve fraud detection systems, and it is a good contribution to the area.

Arshad et al. (2023) introduce a new ensemble approach for improving the security of Internet of Things (IoT) devices against botnet attacks. Although their research focuses mainly on IoT security, the new ensemble approach that integrates multiple machine learning models is very useful for fraud detection. The authors show that ensemble techniques perform better than single models in identifying malicious behavior, and their potential to improve fraud detection systems. The research offers

useful insights into the application of ensemble methods for protecting customer transactions. Huang et al. (2024) introduced a sophisticated blockchain-based system for tracking fraudulent claims in the network applications. The authors introduced a detection model for network access data tampering attacks based on blockchain technology. It solves the emerging issue of data integrity in network access by utilizing the tamper-proof and decentralized characteristics of blockchain. The model is intended to improve security through detecting and preventing unauthorized changes to network access data.

In a nutshell, the above-mentioned research overall identifies the value of improving fraud detection models via more sophisticated machine learning methods, data imbalance methods, and cutting-edge frameworks. In contrast, our framework provides the necessity for the correction of data imbalance for better performance by fraud detection models. Further, the viability of adaptive algorithms and ensemble algorithms to improve the performance of a model is also improvised that delineates the relevance of visualization and easy-to-use interfaces in making fraud detection systems more actionable and accessible. These findings form a solid basis in our framework that identifies the value of improving the fraud detection models using optimization for customer transaction protection. With an improvement in the strides made in above-mentioned studies, our current framework seeks to provide a solid and effective solution for detecting and prevention against fraudulent activity.

METHODOLOGY

The most important methods for developing, adjusting, and refining machine learning models for fraud detection are compiled in this research. Given the large number of classifiers used in this work, a detailed discussion of the methods for model selection data pre-processing, feature set generation, and performance evaluation metrics are required. It frequently occurs in industries including banking, e-commerce, and insurance, and it is crucial to spot unusual and fraudulent transactions. The first and most crucial step in creating machine learning models is data pre-processing. The final model's efficiency in this work environment is determined by the format and quality of the data. Data sub setting, completeness, and formatting are the goals of pre-processing to maximize the data performance on machine learning models. Additionally, feature extraction and selection are necessary to improve the sub models' accuracy and efficiency. In addition to increasing the computational cost, such structures with superfluous or redundant information might reduce the model's efficiency. Figure 1 shows a representation of transaction values, emphasizing both the peaks and troughs, which might be useful when conducting trend analysis, anomaly, or pattern analysis for transactional data.

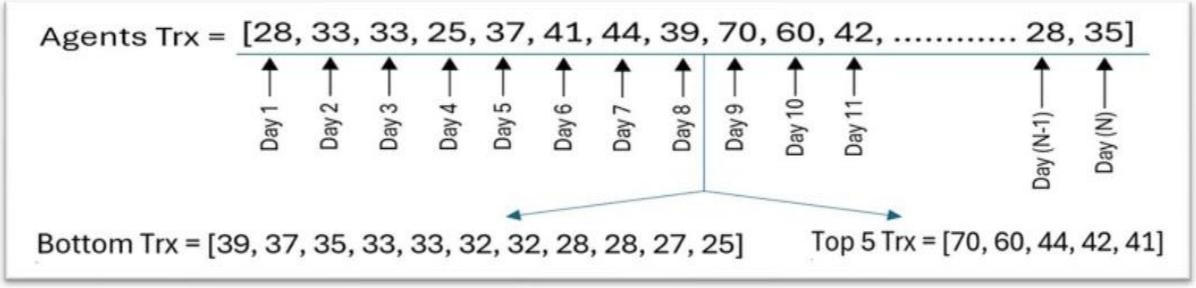

Figure 1.
depicts a dataset representing a set of transaction data spread across a set of days. Description is given as follows:

- Agents Trx: It is a series of transaction values from Day 1 to Day N. The values change, reflecting different transaction volumes or amounts over time.
- Days: The days are numbered in sequence from Day 1 to Day N, implying a chronology for the transactions.
- Bottom Trx: This list indicates the lowest transaction amounts, listed in descending order. It identifies the smallest transactions made within the period.
- Top 5 Trx: This list notes the top five transaction amounts, in descending order. It highlights the highest transactions within the period under observation.

Figure 2 is a thorough examination of recent transactions and information about identity, highlighting possible identity theft. It recognizes patterns and irregularities which could signify identity theft. In reviewing transaction dates, intervals, and matching sets, it allows us to deduce possible cases of fraudulence and assist us in creating approaches to prevent discovering and preventing identity theft by carrying out in-depth transaction and identity analysis.

Figure 2.
Identity Theft, comprises the following essential elements as follows:

- Recent 5 Transactions Dates: Compiles the last five transaction dates, creating a timeline of recent activity.
- Recent 5 Transactions Dates Difference: Indicates the differences (in days) between these transactions, the mean and maximum differences. This facilitates understanding of the frequency and pattern of the transactions.
- Agent Legal Name: Indicates the legal name of the agent who conducted the transactions, which is "Muhammad Wajahat Ali.
- Beneficiary Emails: Gives the email addresses of two beneficiaries who are related to the transactions.
- Matching Sequence: Refers to a sequence of matching names between the agent and beneficiaries, pointing towards possible connections or identity theft red flags.
- Length of Sequence: Refers to the number of matching elements in the sequence, here being 3.

Figure 3 presents an organized flow of the processes for data analysis and fraud detection, indicating the steps from data preparation to machine learning application for fraud detection. It is a visual map to comprehend the systematic process utilized in the study, making it transparent and reproducible analysis. Figure 3

is splitted into various main stages; User Input: The process starts with user input, which probably involves the choice of datasets and parameters for analysis. Data Preparation: This phase is where the datasets are prepared for analysis. It encompasses: Data Set 1 From: Indicates the origin of Data Set 1. Data Set 2 Starting: Identifies the origin of Data Set 2. Start to End Date: Identifies the duration for analysis. Date to (Ending Date - Spiking Duration): Aligns the end date by considering any spiking duration, thereby ensuring correct analysis. Analysis: This step is the computation of several parameters of both data sets: Calculated Parameters for Data Set 1: Metrics or characteristics computed from Data Set 1. Calculated Parameters for Data Set 2: Metrics or characteristics computed from Data Set 2. Difference Computation: A comparison of parameters of both datasets to find the differences or abnormalities. ML Algorithm: The data that is processed is then input into a machine learning algorithm to categorize transactions as either 1) Genuine: Transactions found to be genuine, and 2) Suspicious: Transactions determined to be possibly fraudulent.

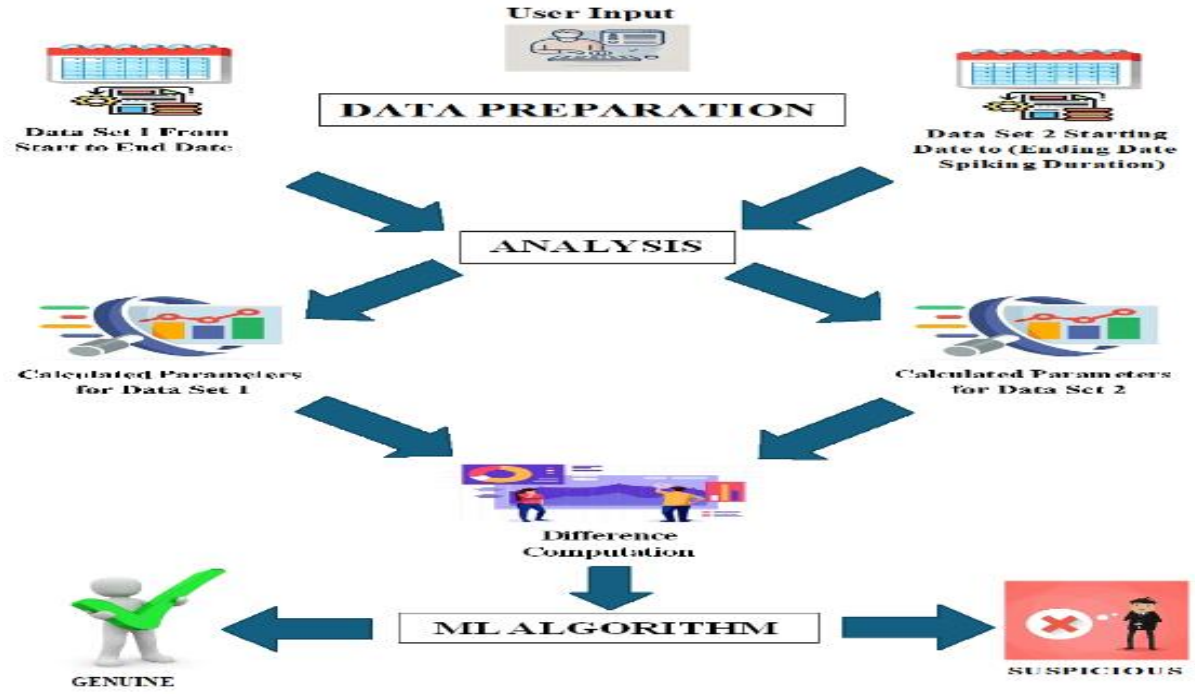

Figure 3.

METHODOLOGY

The model parameters will optimize the machine learning model's accuracy and generalizability. The Grid Search and Random Search techniques were used to hyperparameter tune each of the models in this investigation. If so, these methods look across the hyperparameter space to find the ideal values to utilize. The models' performances are evaluated using a variety of evaluation criteria, such as Accuracy, Precision, Recall, F1-Score, ROC-AUC, MCC, and others. These are the metrics that determine and evaluate the model, thus enables them to differentiate between authentic and fraudulent transactions.

RESULTS

The evaluation criteria used by the machine learning classifiers to guard against fraud will be compared in this section. The classifiers chosen for this study were assessed

based on their capacity to estimate fraudulent transactions when unbalanced data was present. KNN, Ridge Classifier, Random Forest, Decision Tree, Voting Classifier, Support Vector Classifier, Logistic Regression, and Gaussian Naive Bayes were taken into consideration. Fraud detection typically deals with data that is unbalanced, suggesting that there are far more genuine transactions than fraudulent ones. The process of identifying fraud strains is difficult due to the complexity and variety of fraud's nature and forms. The classifiers employed in this work fall into two general categories: single classification models and combination techniques. Two instances of ensemble approaches that are presumed due to the idea that efficiency might be achieved by mixing multiple exclusive models are Random Forest and Voting Classifier. These models integrate the expectations from several base models, from which one is chosen when deciding.

An ensemble technique is highly helpful when dealing with noisy and unbalanced data, such as in fraud detection single classifiers, on the other hand, rely on the output of a single classifier algorithm. Even if they are not as effective as the ensemble approaches combined, they can nevertheless identify fraud regardless of the intricacy of the issue.

Given the skewness of the datasets utilized in fraud detection, the evaluation measures included in this study are especially crucial for evaluating model performance. When it comes to predicting 1% of fraudulent transactions, a model that simply predicts the majority class (legal transactions) may achieve an accuracy of over 99% but fall well short. This statistic shows the proportion of all correctly predicted transactions, including both legitimate and fraudulent ones. However, it is not always the optimal measure in the case of imbalance because a model can obtain high accuracy just by forecasting the dataset's mode.

Recall and accuracy metrics are crucial for identifying fraud. Precision is a measure of the proportion of accurate positive predictions (in this example, fraudulent transactions). F1 stands for F1 score, which gives the harmonic mean of our data set's precision and recall if there is inequality between the two classes. Since both high precision and high recall are crucial for fraud detection, F1 is a more sensible way to gauge a model's effectiveness. The confusion matrix separates the model's predictions into four categories: False Positives (FP), False Negatives (FN), True Positives (TP), and True Negatives (TN) are examples of common parameters.

Table 1 indicates the test set performance of various classifiers so that Random Forest and Decision Tree record the maximum accuracy and F1-Score on the test set, with Random Forest being slightly more stable. Logistic Regression also performs well, particularly on the test set, pointing towards good generalization. Decision Tree and KNN demonstrate overfitting tendencies, with high performance in training but drastic drops in test performance.

Ridge Classifier and Naive Bayes perform poorly, with poor accuracy and recall, and hence are not fit for this task. SVM and Voting Classifier offer balanced results but are beaten by Random Forest and Logistic Regression. Random Forest is the best classifier, with high accuracy, precision, recall, and F1-Score on both training and test data. Logistic Regression also shows good generalization and is a good alternative. The models such as Decision Tree and KNN, in spite of classifying well when trained, get overfitted, while those of Ridge Classifier and Naive Bayes don't perform too well overall.

Table 1.
Accuracies for Different Classifiers

Classifiers	Train Accuracy	Test Accuracy	Train Precision	Test Precision	Train Recall	Test Recall	Train F1 Score	Test F1 Score
Random Forest	100.00%	97.89%	100.00%	97.47%	100.00%	97.89%	100.00%	97.60%
Decision Tree	100.00%	96.35%	100.00%	97.13%	100.00%	96.35%	100.00%	96.70%
Voting Classifier	96.64%	94.52%	96.65%	97.05%	96.64%	94.52%	96.64%	95.62%
Support Vector	95.19%	91.29%	95.25%	96.84%	95.19%	91.29%	95.19%	93.71%
Logistic Regression	86.75%	89.89%	87.03%	97.13%	86.75%	89.89%	86.73%	92.96%
KNN Classifier	96.67%	89.61%	96.88%	96.42%	96.67%	89.61%	96.67%	92.66%
Ridge Classifier	78.31%	74.30%	78.65%	96.37%	78.31%	74.30%	78.25%	83.33%
Naive Bayes	65.99%	40.03%	73.76%	97.11%	65.99%	40.03%	62.96%	54.82%

Table 2 indicates the performances of classifiers via K-Fold cross validation. Performance indicates Decision Tree and Random Forest as the best classifiers with the highest accuracy and stability. Ridge, Voting, and Logistic Regression models present good alternatives, whereas SVC and KNN perform moderately. Gaussian Naive Bayes is not suggested because of its low performance. These results emphasize the necessity of model choice based on cross-validation measures for generalization confidence.

Table 2.
K-Fold Cross Validations Results

Classifiers Accuracy	Cross-Validation	Cross-Validation Precision	Cross-Validation Recall	Cross-Validation F1 Score
Decision Tree Classifier	98.36%	98.22%	98.36%	97.97%
Random Forest Classifier	97.85%	97.90%	97.85%	96.90%
Ridge Classifier	97.72%	95.50%	97.72%	96.60%
Voting Classifier	97.72%	95.50%	97.72%	96.60%
Support Vector Classifier	97.64%	95.50%	97.64%	96.56%
Logistic Regression	97.56%	95.92%	97.56%	96.59%
KNN Classifier	96.97%	97.23%	96.97%	97.09%
Gaussian Naive Bayes	59.33%	95.57%	59.33%	72.59%

Table 3 presents ensemble classifier results such that Bagging Classifier and Random Forest Classifier are the best ensemble algorithms, with the highest accuracy and stability. Soft Voting Classifier is a good substitute, while Gradient Boosting and AdaBoost are moderately stable but less accurate. These findings illustrate the power of ensemble algorithms, especially Bagging and Random Forest, in terms of high-performance classification.

Table 3.
Ensembled Classifiers Results

Classifier	Cross-Validation Accuracy	Cross-Validation Precision	Cross-Validation Recall	Cross-Validation F1 Score
Bagging Classifier	97.19%	97.19%	97.19%	97.19%
Random Forest Classifier	97.19%	96.83%	97.19%	96.99%
Soft Voting Classifier	96.91%	96.91%	96.91%	96.91%
Gradient Boosting Classifier	95.08%	96.93%	95.08%	95.91%
AdaBoost Classifier	95.08%	96.17%	95.08%	95.61%

Random Forest Classifier and Bagging Classifier are the best ensemble methods, being the most accurate and robust. Hard Voting Classifier is a robust alternative, with Gradient Boosting, Soft Voting, and AdaBoost being effective but slightly less accurate. These findings establish the efficiency of ensemble methods, especially Bagging and Random Forest, in realizing high performance when classifying data that is stratified.

Table 4.
Ensembled Classifiers with K Fold Stratified Results

	Cross-Validation Accuracy	Cross-Validation Precision	Cross-Validation Recall	Cross-Validation F1 Score
Bagging Classifier	98.40%	98.19%	98.40%	98.22%
Random Forest Classifier	98.31%	98.12%	98.31%	97.93%
Gradient Boosting Classifier	97.72%	97.56%	97.72%	97.64%
AdaBoost Classifier	97.01%	97.44%	97.01%	97.21%
Hard Voting Classifier	97.85%	97.90%	97.85%	96.90%
Soft Voting Classifier	97.72%	95.50%	97.72%	96.60%

Figure 4 shows the confusion metrices of the classifiers. Results suggest that Random Forest classifier and Bagging classifier are the best ensemble methods, being the most accurate and robust. Hard Voting Classifier is a robust alternative, with Gradient Boosting, Soft Voting, and AdaBoost being effective but slightly less accurate. These findings establish the efficiency of ensemble methods, especially Bagging and Random Forest, in realizing high performance when classifying data that is stratified.

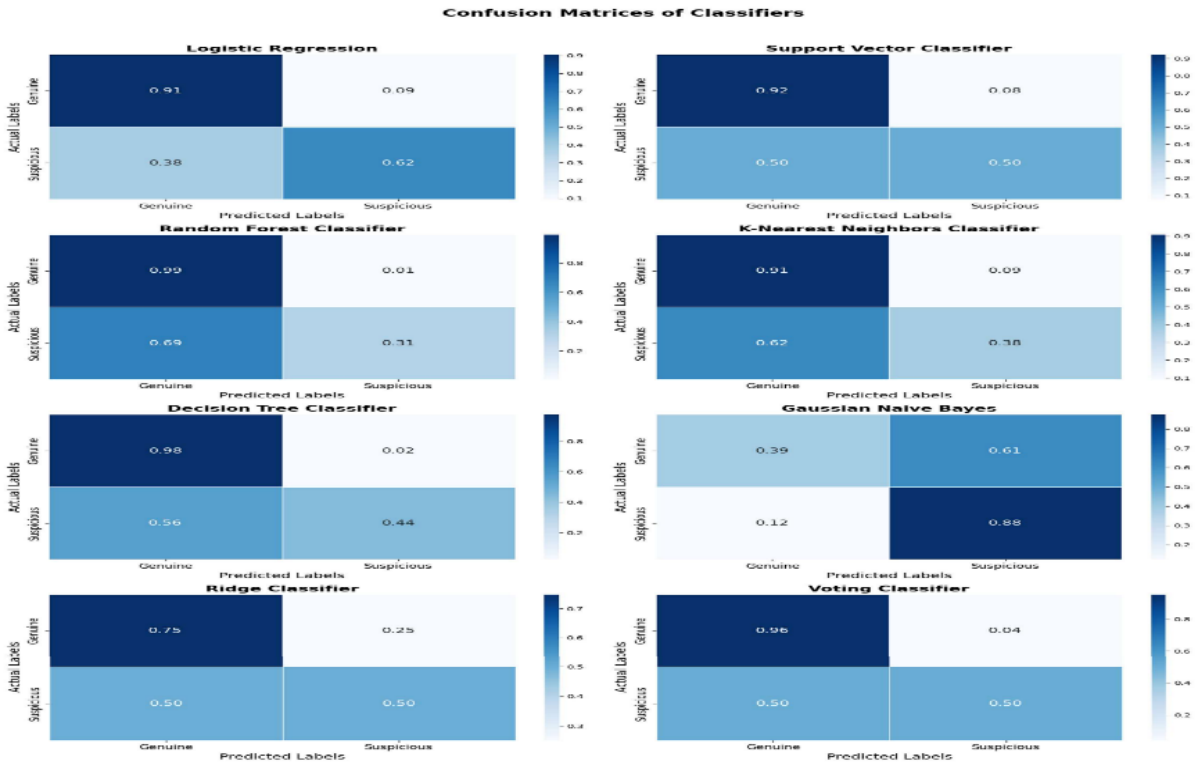

Figure 4.
Confusion Metrices of Ensemble and Classifiers.

Figure 5 shows the confusion metrices of simple classifiers. The results show that the Random Forest and Decision Tree perform well for real cases but are poor for suspicious cases. Gaussian Naive Bayes does the best in identifying suspicious cases but performs poorly for real cases. Support Vector, Logistic Regression, and KNN are equally good with balanced performance. Voting Classifier has good results for real cases but moderate performance for suspicious cases.

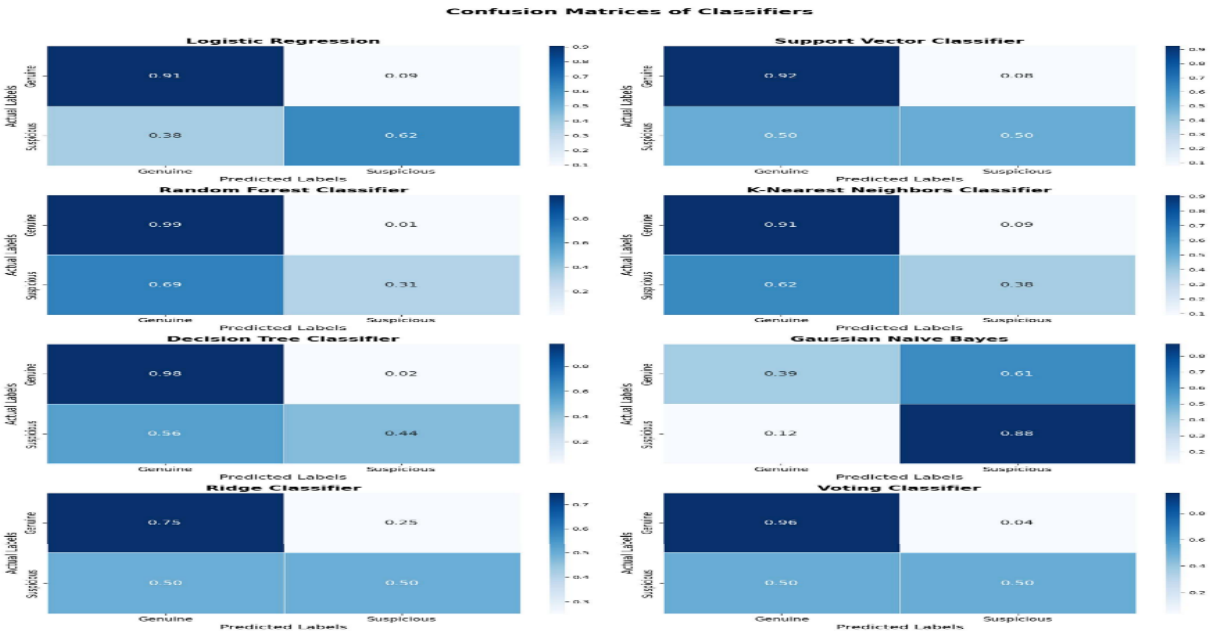

Figure 5.
Confusion Matrices of Simple ML Classifier

Figure 6 shows the training vs testing Accuracies for Simple ML Classifiers. Results suggest that Random Forest Classifier and Voting Classifier have the smallest difference between training and testing accuracies, reflecting good generalization. Logistic Regression also reflects good generalization with a moderate decline in testing accuracy. Decision Tree Classifier, KNN, and SVC have very high declines in testing accuracy relative to training accuracy, reflecting overfitting. Gaussian Naive Bayes and Ridge Classifier also reflect overfitting but to a lesser degree. Logistic Regression and Voting Classifier have equal training and testing performance; thus, they are good options.

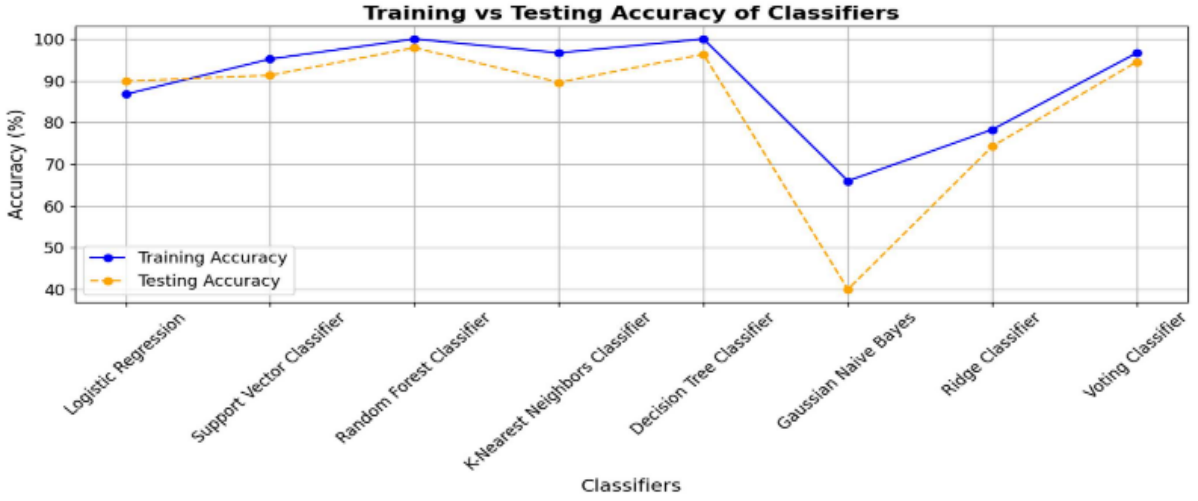

Figure 6.
Training vs Testing Accuracies of Simple ML Classifiers

Figure 7 depicts the Training vs Testing Log Loss for ML Classifiers. Results suggest that Random Forest Classifier has the best performance, with low log loss and high generalization. Logistic Regression achieves a good tradeoff between training and test performance. Decision Tree, KNN, and SVC models are overfitting and need regularization or tuning to enhance generalization. Gaussian Naive Bayes and Ridge Classifier perform less well with poor generalization power.

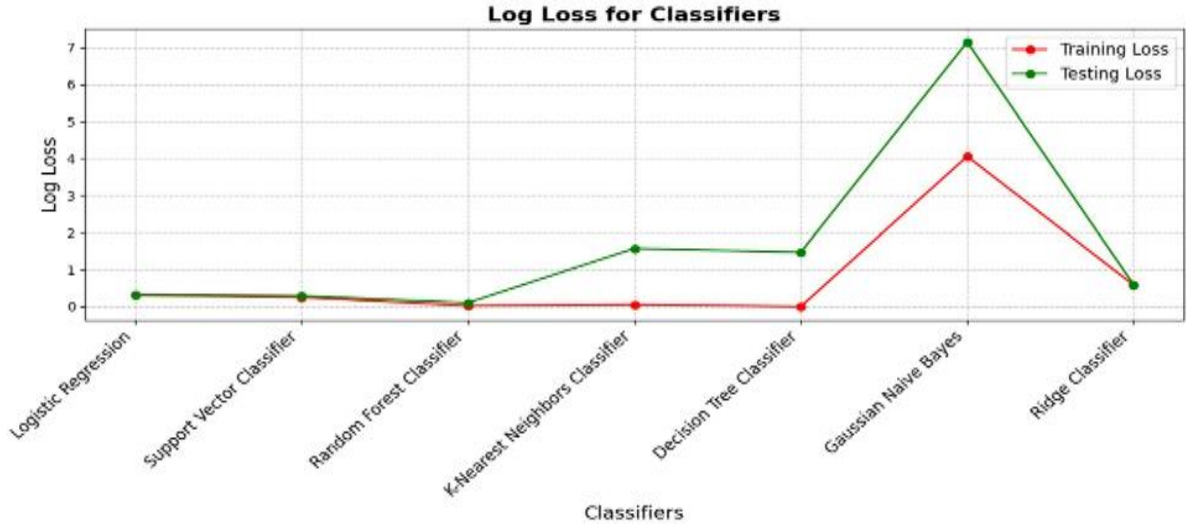

Figure 7.
Training vs Testing Log Loss of ML Classifiers

Figure 8 shows the Receiver Operating Characteristic (ROC) curves for different machine learning classifiers and their respective Area Under the Curve (AUC) values. The ROC curve and AUC are utilized to compare the performance of classification models, where higher values of AUC reflect better performance. Logistic Regression and Random Forest Classifier possess the highest AUC values (0.81), reflecting superior classification performance. Support Vector Classifier is second with an AUC of 0.79. Gaussian Naive Bayes performs moderately with an AUC of 0.74. KNN Classifier, Decision Tree Classifier, and Ridge Classifier possess the lowest AUC values (0.64, 0.61, and 0.64, respectively), reflecting poor classification capability. Thus, Logistic Regression and Random Forest Classifier are the top-performing models, with high AUC values reflecting high classification capability. Support Vector Classifier also works well, albeit lower than the best ones.

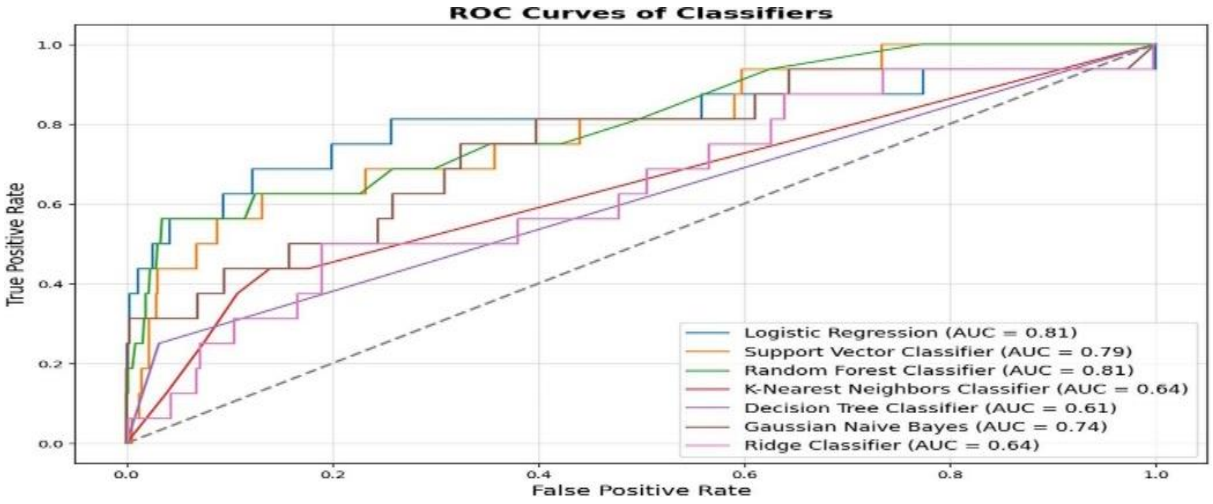

Figure 8.
ROC Curve of simple ML Classifiers

Figure 9 shows the ROC Curve for Ensembled Classifiers with Stratified K-Fold. Gradient Boosting Classifier comes close with an AUC of 0.99, which is indicative of near-perfect performance. Bagging Classifier, AdaBoost Classifier, Random Forest Classifier, and Soft Voting Classifier are the top-performing ensemble models, having perfect AUC scores and demonstrated high classification capability.

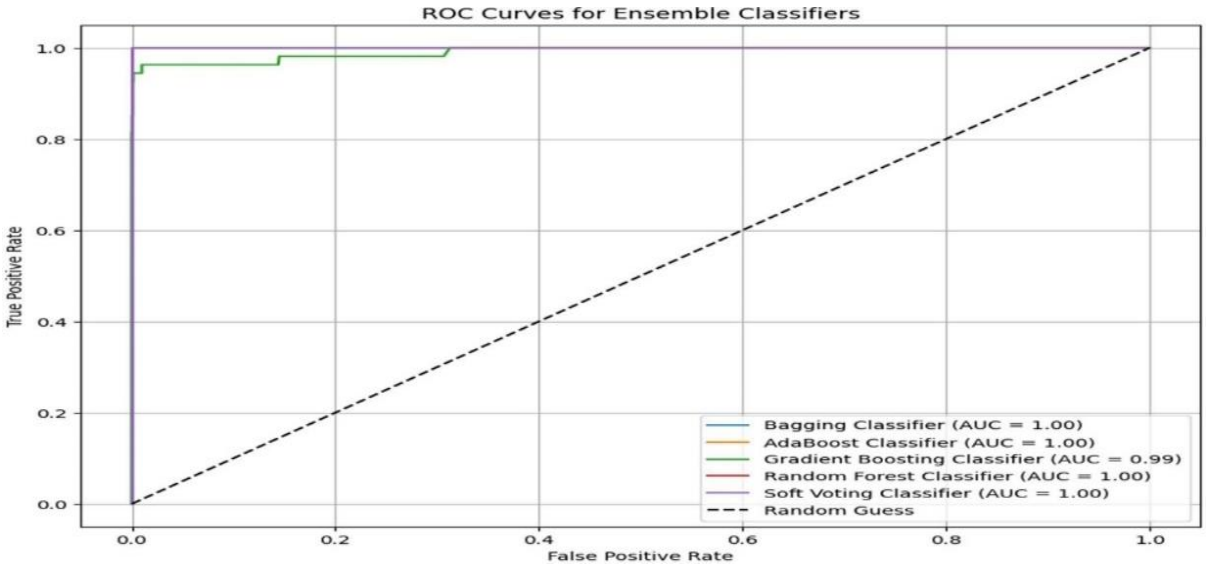

Figure 9.
ROC Curve of Ensembled Classifiers with Stratified K Fold

DISCUSSION

In general, every classifier has advantages and disadvantages with relation to the ability to identify falsification. Random Forest's resistance to overfitting and its capacity to manage huge datasets with numerous features are two of its main advantages. In an area where transparency is crucial, fraud detection, decision trees' high interpretability and ease of comprehension are a big plus. Additionally, they are prone to overfitting, particularly in cases when the data is noisy, or the models are deep. Models that perform badly on unknown data may result from this. By integrating the advantages of several models, the Voting classifier lessens the drawbacks of each model separately. However, this method has the drawback of being computationally costly and possibly unsuitable for situations where class differentiation is challenging or for very large data sets. Logistic regression analysis works well on huge datasets because of its speed, adaptability, and the ease with which computations may be completed on a computer. However, it may occasionally struggle to handle intricate and non-linear connections, and it performs poorly in circumstances where the data is extremely unbalanced, like in fraud detection.

Even though KNN is very easy to use and intuitive, especially when working with large data sets, it can be very time-consuming and computationally demanding. The study discovered that the K value and distance measure have a significant impact on KNN performance and should be appropriately tempered. Regularization is used to mitigate the overfitting risk that existed in the Ridge Classifier. Even so, it functions well when multicollinearity is present, but not when the relationship is nonlinear. It can seldom perform successfully if it is present in the data. Bayes is a quick and simple classifier to use, it requires feature independence, which is not applicable when the data is not orthogonal. The effectiveness of several classifiers for fraud detection have been investigated in this research, thus, Random Forest Bagging Classifier, AdaBoost

Classifier, Random Forest Classifier, and Soft Voting Classifier are the top-performing ensemble models.

CONCLUSION

Our Framework provides assessment of ML models based on several performance metrics, such as accuracy. The precision, recall, F1 score, and cross-validation levels for several 12% of the machine learning classifiers' performance in detecting fraudulent transactions were also examined in this exploratory work. Random Forest exhibited somewhat better accuracy than the other models on all evaluation metrics, including perfect accuracy and recall, equally assessed F-1 scores, and high train and test accuracy. Accordingly, Random Forest is a reliable tool for handling complicated data sets in fraud detection applications. Despite being somewhat less accurate than Random Forest, these models performed admirably in identifying fraudulent transactions. K-Nearest KNN and Logistic regression models yielded lower accuracy-recall values than those previously discussed. Future studies might investigate the use of under sampling or oversampling techniques (like SMOTE) to rectify the imbalance in fraud detection datasets. Examining deep learning-based methods like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) may yield better results, particularly when it comes to seeing complex connections and patterns in the data that traditional classifiers could miss. By increasing the signal-to-noise ratio in the data, a thorough analysis of feature engineering methods such as domain-specific feature extraction or dimensionality reduction using Principal Component Analysis (PCA) might improve model performance even further.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the contributor of research and their different affiliations.
Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.
Availability of data and material: In the approach, the data sources for the variables are stated.
Authors' contributions: Each author participated equally to the creation of this work.
Conflicts of Interests: The authors declare no conflict of interest.
Consent to Participate: Yes
Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCES

Al-dahasi, E. M., Alsheikh, R. K., Khan, F. A., & Jeon, G. (2025). Optimizing fraud detection in financial transactions with machine learning and imbalance mitigation. *Expert Systems*, *42*(2), e13682.

Al-Hashedi, K. G., & Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. *Computer Science Review*, *40*, 100402.

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial fraud detection based on machine learning: A systematic literature review. *Applied Sciences*, *12*(19), 9637.

Amarnadh, V., & Moparthi, N. R. (2023). Comprehensive review of different artificial intelligence-based methods for credit risk assessment in data science. *Intelligent Decision Technologies*, *17*(4), 1265-1282.

Arshad, A., Jabeen, M., Ubaid, S., Raza, A., Abualigah, L., Aldiabat, K., & Jia, H. (2023). A novel ensemble method for enhancing Internet of Things device security against botnet attacks. *Decision Analytics Journal*, *8*, 100307.

Chatterjee, P., Das, D., & Rawat, D. B. (2024). Digital twin for credit card fraud detection: Opportunities, challenges, and fraud detection advancements. *Future Generation Computer Systems*, *157*, 1-15.

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

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: A review of anomaly detection techniques and recent advances. *Expert Systems with Applications*, *193*, 116429.

Huang, C., Nong, L., Nong, Y., Lu, Y., Chen, Z., & Li, Z. (2024). Detection model for network access data tampering attacks with blockchain technology. *Intelligent Decision Technologies*, *18*(3), 1-3.

Kamuangu, P. (2024). A review on financial fraud detection using AI and machine learning. *Journal of Economics, Finance and Accounting Studies*, *6*(1), 67-77.

Kanksha, Bhaskar, A., Pande, S., Malik, R., & Khamparia, A. (2021). An intelligent unsupervised technique for fraud detection in health care systems. *Intelligent Decision Technologies*, *15*(1), 127-139.

Kumar, S., Dwivedi, M., Kumar, M., & Gill, S. S. (2024). A comprehensive review of vulnerabilities and AI-enabled defense against DDoS attacks for securing cloud services. *Computer Science Review*, *53*, 100661.

LexisNexis Risk Solutions. (2024). *North American Ecommerce and Retail Companies Face a $3.00 Total Cost for Each Dollar Lost to Fraud, According to True Cost of Fraud Study from LexisNexis Risk Solutions.* https://risk.lexisnexis.com/about-us/press-room/press-release/20240327-tcof-retail-ecommerce?utm_source=chatgpt.com

Makki, S., Assaghir, Z., Taher, Y., Haque, R., Hacid, M. S., & Zeineddine, H. (2019). An experimental study with imbalanced classification approaches for credit card fraud detection. *IEEE Access*, *7*, 93010-93022.

Raghavan, P., & El Gayar, N. (2019, December 11). Fraud detection using machine learning and deep learning. In *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)* (pp. 334-339). IEEE.

Sadgali, I., Sael, N., & Benabbou, F. (2019). Performance of machine learning techniques in the detection of financial frauds. *Procedia Computer Science*, *148*, 45-54.

Statista. (n.d.). *Value of e-commerce losses to online payment fraud worldwide in 2023 and 2024, with forecasts for 2029*.https://www.statista.com/statistics/1273177/ecommerce-payment-fraud-losses-globally/#statisticContainer

Wei, W., Li, J., Cao, L., Ou, Y., & Chen, J. (2013). Effective detection of sophisticated online banking fraud on extremely imbalanced data. *World Wide Web*, *16*(4), 449-475.

Yadav, R. A., Logofatu, D., Mim, S. S., & Ray, J. K. (2024, June 21). Effective machine learning techniques and API realizations for visualizing fraud detection in customer transactions. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 301-315). Springer Nature Switzerland.

Yan, C., Wang, J., Zou, Y., Weng, Y., Zhao, Y., & Li, Z. (2024, July 5). Enhancing credit card fraud detection through adaptive model optimization. In *2024 IEEE 7th International Conference on Big Data and Artificial Intelligence (BDAI)* (pp. 49-54). IEEE.