Cureus Journal of Computer Science

Open Access Original Article

Fraud Detection: A Hybrid Approach With Logistic Regression, Decision Tree, and Random Forest

Received 10/30/2024
Review began 11/20/2024
Review ended 01/19/2025
Published 01/28/2025

DOI: https://doi.org/10.7759/s44389-024-02350-5

Yugal Salunke [1] , Saroj Phalke [1] , Manoj Madavi [1] , Praali Kumre [1] , Grishma Bobhate [1]

1. Computer Engineering in Artificial Intelligence and Machine Learning, Vishwakrma Institute of Information Technology, Pune, IND

Corresponding authors: Yugal Salunke, yugal.22210227@viit.ac.in, Saroj Phalke, saroj.22211519@viit.ac.in, Manoj Madavi, manojmadavi20@gmail.com, Praali Kumre, praali.22210449@viit.ac.in, Grishma Bobhate, grishma.bobhate@viit.ac.in

Abstract

The rise of financial fraud, not only in India but also around the world, is a major problem. Credit card transactions have been steadily increasing in recent years, as have internet payments like Unified Payments Interface and debit card transactions, which have been increasing every day. As a result, fraud is increasing, and it has become easier for fraudsters to commit fraud.

Recent research indicates the practicality of machine learning algorithms for recognizing payment fraud. Since credit cards are a common target, online fraudsters simply commit fraud because of excess use of e-commerce and other platforms, which has led to an increase in online payment methods, resulting in a greater risk of online fraud. Many researchers have started using machine learning algorithms to detect fraud. The primary objective of this study is to create a better fraud detection algorithm by analyzing payment transactional data, focusing on customers' transactional data and their payment behavioral patterns. The study proposes a framework that groups cardholders based on the volume of their transactions.

Financial fraud is the biggest threat to world economies in terms of financial losses each year. The task of accurately and efficiently detecting fraud is, therefore, very challenging due to the high volume of transaction data and its complex nature. This paper presents a hybrid machine learning approach using logistic regression, decision trees, and random forests that could help enhance the accuracy and reliability of fraud detection systems. The study used an available credit card fraud dataset, with a focus on feature engineering and model evaluation to compare individual algorithms versus their ensemble. Experimental results show that the ensemble outperformed the individual classifiers in terms of precision, recall, and overall accuracy. This paper underlines the potential of hybrid machine learning techniques in improving fraud detection systems and proposes avenues for further research.

Categories: Algorithm Analysis, Ensemble Learning, Machine Learning (ML)
Keywords: random forest, decision tree, logistic regression, machine learning, ensemble learning model, fraud detection, fraud data, metric accuracy, credit card fraud detection dataset, confusion matrix

Introduction

With the rapid proliferation of digital payment systems, financial fraud has also increased. Threats such as SIM cloning, QR code manipulation, and digital payment fraud have become common. Governments and institutions across the globe are seeking to identify the ideal methods to combat these problems effectively [1]. Consequently, fraud detection systems have gained significant attention in the present era.

Therefore, it is possible to consider machine learning and deep learning techniques as important methods for anomaly detection and pattern recognition of fraudulent activities [1,2]. Despite these advancements, challenges persist in the development of robust models that can effectively process the diversified and dynamic nature of fraudulent behaviors. As such, research continues to focus on these methodologies to continue fine-tuning their accuracy and adaptability [3]. In this paper, a hybrid approach is proposed, where multiple machine learning models, including logistic regression, decision trees, and random forests, are deployed for performance enhancement of fraud detection. Through the stacking technique, an ensemble model was developed by integrating the core models of random forest, decision tree, and logistic regression. The ultimate predictions were generated by a meta-model utilizing the outputs from these core models as its inputs. This methodology capitalized on the strengths of each algorithm: random forest provided increased robustness, decision tree facilitated the identification of non-linear patterns, and logistic regression offered insights into linear relationships. By minimizing both overfitting and underfitting, enhancing generalization, and effectively addressing unbalanced datasets, the ensemble model surpassed the performance of the individual models. With an accuracy of 99%, a precision of 98%, and a recall of 98.9%, it achieved outstanding performance metrics. The model's ability to respond to intricate patterns was

How to cite this article
Salunke Y, Phalke S, Madavi M, et al. (January 28, 2025) Fraud Detection: A Hybrid Approach With Logistic Regression, Decision Tree, and Random Forest. Cureus J Comput Sci 2 : es44389-024-02350-5. DOI https://doi.org/10.7759/s44389-024-02350-5

enhanced by the combined predictions, thus improving its effectiveness in fraud detection.

According to the Reserve Bank of India's annual report from the Times of India, the information provided in Table *1* illustrates that as the number of transactions increased, the frauds have also increased rapidly. Banks have observed a maximum number of frauds in the digital payment category by the end of 2023.

Years	No. of Digital Frauds	Amount Involved (in Crore)	Remarks
2020-21	2,545	119	Steady increase in digital transactions
2021-22	3,596	155	Spike due to increased online activity
2022-23	6,659	276	Significant rise in digital fraud cases

TABLE 1: Digital Fraud Data From Times of India

https://timesofindia.indiatimes.com/business/india-business/financial-frauds-on-the-rise-how government-plans-to-check-sim-cloning-qr-code-digital-payments-fraud/amparticleshow/105775407.cms (accessed on 18th April)

Escalating fraudulent activities in financial systems have caused extensive economic damage across the globe. The fight against fraud requires strong and accurate detection techniques. Traditional approaches using predefined rules appear clumsy compared to dynamic and scalable machine learning approaches, where recent breakthroughs in the machine learning field have paved the way for different models that can successfully analyze complex transaction patterns. In this paper, three widely used supervised learning algorithms - logistic regression, decision tree, and random forest - and their hybrid implementations are explored to overcome the shortcomings of standalone models. The proposed approach and ensemble framework leverage the strengths of each individual model to improve fraud detection accuracy. The final results contribute to the continuous development of intelligent systems for fraud prevention. Relevant studies in fraud detection and machine learning techniques were consulted to support this research.

Materials And Methods

Literature work

Fraud is defined as the criminal use of deception for personal or financial gain. The act of obtaining financial advantages through dishonest and unlawful means is known as financial fraud. Financial fraud occurs in various sectors, including corporate, banking, and taxation.

Maniraj et al. [4] conducted a study on unexpected transactions, achieving 99.6% accuracy. The research demonstrated the efficiency of machine learning algorithms, showing that a larger number of samples increases precision. More data lead to a more accurate model. In their model, they observed a very low proportion between accuracy and precision. The authors list several challenges in fraud detection. First, this application deals with severely imbalanced datasets, which contain only a tiny portion of the total available data. It is challenging to train effective models when fraud data are scarce. However, a major problem arises in data cleaning, where a trial-and-error method is required to clean the dataset.

Raghavan and Gayar [5] used both machine learning and deep learning algorithms and concluded that different algorithms work well for various datasets. Although support vector machines (SVMs), deep Boltzmann machines, autoencoders, convolutional neural networks (CNNs), and recurrent neural networks can help discover fraud, they are not very efficient in dynamic scenarios. The authors concluded that SVMs are among the best tools for detecting fraud when working with larger datasets and can be combined with CNNs for the most effective outcomes.

Gheisari et al. [6] introduced machine learning models, utilizing synthetic datasets and features such as SVM, logistic regression, and decision trees, to effectively predict and prevent financial fraud. The union of machine learning algorithms in fraud detection systems enhances security measures and minimizes the risk of financial losses due to fraudulent activities, which affects the training process. The researchers demonstrated that different models were used for empirical evaluation.

A study by Amarasinghe et al. [7] presents their research on the analysis of fraud in transactions. Some finance companies have divided their domain based on their domain experts and data scientists, with data scientists working on fraud detection. Using manual fraud detection, they obtained low accuracy due to the difficulty of handling large volumes of data, and more time is required to detect transactional fraud. By incorporating findings from the cited paper, you can support the assertion that conventional fraud detection techniques, which focus exclusively on distinct data points, are inadequate, and that leveraging machine learning methods provides a more effective and efficient alternative.

Thimonier et al. [8] explored the application of aberration detection (AD) methods. They found that, although the light gradient boosting machine experiences more disturbance than the AD method, it performs noticeably better across all evaluated metrics, because it could not find a way that is easy to see improvement in overall fraud detection. The authors used a self-supervised method for fraud detection. Machine learning helps build fraud detection systems in banking by learning from data. By improving machine learning mechanisms, new patterns of fraud can be detected more easily, thereby preventing financial losses.

In their systematic literature review, Al Marri and AlAli [9] described their approach that employed artificial neural networks, SVM, and other techniques to detect fraud. The review highlights significant problems and limitations. Research on transactional fraud detection has primarily used methods such as regression and supervised learning. The authors suggest that there is more scope for unsupervised learning, which is used for anomaly detection.

A deep forest-based approach has been highlighted in a study on fraud detection during online transactions, with the goal of improving detection accuracy by using a transaction time-based differentiation feature-generating method. The research paper proposed the concepts of individual credibility degree and group anomaly degree to boost the capacity to differentiate between genuine and fraudulent transactions. By addressing the challenge of extreme data imbalance and improving outlier detection, the enhanced deep forest model achieved significant improvements in precision and recall rates compared to traditional random forest models. The technique shows how cutting-edge machine learning algorithms and domain-specific feature engineering can be combined to strengthen fraud detection [8].

In this research paper, the following algorithms are used: logistic regression, decision tree, random forest, and ensemble learning. The chosen algorithms probably strike a balance between simplicity, ease of understanding, and effectiveness in addressing the issues of fraud detection, including dealing with imbalanced datasets and the need for quick decision-making in real-time scenarios.

Gap Analysis

After surveying research papers, it was revealed that there is a difference in performance between individuals using supervised and unsupervised learning algorithms. The root of the problem lies in overfitting and underfitting. Different researchers have identified various advantages, disadvantages, and problems, as shown in the chart below. After performing an extensive literature survey, it became clear that supervised and unsupervised machine learning algorithms have shown different levels of success in various instances, facing many problems like overfitting and underfitting. The varying successes and failures of researchers with these algorithms demonstrate that they are also a double-edged sword, especially for categories like fraud detection. Additional research studied popular machine learning algorithms, such as SVM and artificial neural networks, to detect frequently occurring types of fraud. However, the ability of these methods to detect weaker or more esoteric forms of fraud was limited compared to unsupervised learning and text analysis methodologies.

Deep learning models appeared as an attractive solution, capable of processing large amounts of data at low cost; however, they were less useful in environments where fraud patterns kept changing. This suggested that more adaptable models, which can evolve based on emerging fraud trends, are necessary.

In summary, although the techniques outlined in these papers proved quite helpful, they also taught us that no two circumstances are alike, making a one-size-fits-all approach impossible. There is a strong need for more flexible models to keep pace with changing fraud trends. The gap analysis in Table *2* provides an overview.

Ref. No.	Advantages	Disadvantages	Research Gap
[4]	Accuracy: 99.6%. Precision increases with dataset size. Bigger data means more accuracy. Reduces false positives.	Precision was 28% (very low). 10% of data show max 32%.	Low precision, even with the whole dataset.
[10]	Grouping helps find behavioral patterns for user profiles.	Various data cleaning methods needed. No specific method is necessary.	SMOTE, correlation coefficient, MCC for balancing. Clusters from transactions. No specific cleaning method.
[11]	High accuracy with low false positives.	Large genuine transactions, few fraudulent ones.	Improved with decision tree on PAYSim. User-specific models based on transaction history.
[9]	Finds popular fraud types. Use SVM and ANN.	Only popular methods used.	Unsupervised learning. Clustering for small frauds. Word2Vec, Doc2Vec, and BERT for text vectors.
[5]	Deep learning for high computation, low cost. Different algorithms for different datasets.	SVM, CNN, autoencoders, RBN, and DBM only detect fraud. Poor in dynamic environments.	Neural networks for changing fraud patterns. Fuzzy pairing anomaly detection. More interpretable models.
[12]	Adaptive thresholds and dynamic risk scoring researched.	Poor performance in complex patterns and unbalanced data.	Combine methods for accuracy. Use social media and partner transaction history.
[9]	Feature engineering and model training. Epitomized evaluation criteria.	Validity boundaries and risks identified.	Focus on supervised learning, neural networks, SVM, and logistic regression.

TABLE 2: Fraud Detection Insights

SVM: Support Vector Machine; ANN: Artificial Neural Network; CNN: Convolutional Neural Network; RBN: Radial Basis Network; DBM: Deep Belief Network; SMOTE: Synthetic Minority Over-sampling Technique; MCC: Matthews Correlation Coefficient

Proposed work

This is a fraud detection project aimed at identifying fraudulent transactions in financial datasets using logistic regression, decision tree, and random forest algorithms.

The architecture of the system:

Data preprocessing (cleaning, handling missing values, and normalization) and feature selection ensure that clean data are passed to the models, enhancing the accuracy of fraud detection.

The preprocessed data were used to train the random forest, decision tree, and logistic regression models independently. Performance indicators, including accuracy, precision, recall, and F1 score, were calculated for each model after it was assessed on the test set.

The same underlying models - logistic regression, decision tree, and random forest - were employed for the ensemble model, with logistic regression acting as the meta-model in a stacking ensemble setup. Joblib was used to save the trained stacking model. Following training, the model's performance was assessed using a confusion matrix to visualize the classification results, accuracy, precision, recall, and F1 score.

By implementing uniform preprocessing procedures for both the separate models and the ensemble, clean and dependable data were supplied to all models, leading to enhanced accuracy and improved performance in fraud detection.

The same methodology was employed to train individual models and create an ensemble for detecting fraud in credit card transactions. The dataset underwent preprocessing, which involved cleaning, addressing missing values, and normalization through feature scaling using StandardScaler. Z-scores were computed to detect and eliminate outliers, ensuring that only pertinent data were provided to the models. The dataset was divided into features (X) and the target variable (y), and subsequently split into training and testing sets.

Once data preprocessing is complete, the dataset is used to train each of these algorithms. Logistic regression, being interpretable, is effective at identifying linear relationships, whereas decision trees capture complex, non-linear interactions among features. Random forest, an ensemble of decision trees, adds robustness by reducing variance and generalizes well by averaging the results of multiple trees.

Each algorithm is analyzed individually, and then ensemble learning - stacking - is applied to leverage the strengths of all three models together. In stacking, the predictions of these classifiers are aggregated, yielding more accurate results than using each model separately. This technique helps mitigate typical issues like overfitting and underfitting, improving diagnostic accuracy.

The model's performance is evaluated using the following metrics: F1 score, accuracy, precision, and recall. These measures are crucial for assessing the model's efficacy when working with imbalanced datasets, where fraudulent transactions (positive class) are substantially less common than non-fraudulent transactions (negative class). Once thoroughly tested, the model will be used in a real-time fraud detection production system, integrated with banking systems to generate fraud detection reports and continuously monitor transactions. This network includes additional layers of encryption and authentication to ensure enhanced security for transaction data.

A detailed architectural view of the fraud detection system, including data preprocessing, model training, stacking, and deployment, is presented in Figure *1*.

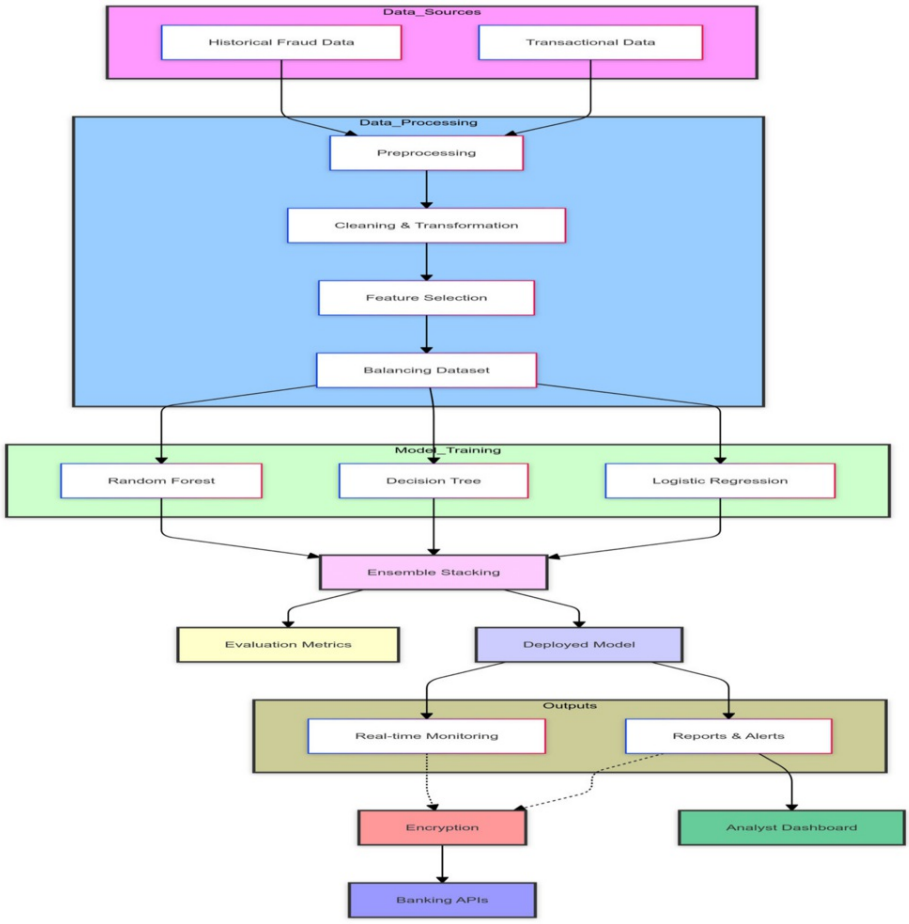

FIGURE 1: Architecture Diagram

Dataset

The flowchart in Figure *2* begins with loading the dataset into the environment. The data is then cleaned and transformed to make it suitable for analysis, which may include handling missing values, normalizing features, and other preparatory steps. The dataset is divided into training and testing sets to complement model evaluation. A machine learning model is selected and initialized for training. The training dataset is used to train the model. The outcome of the trained model is assessed using the the testing dataset. Finally, the results and performance metrics are visualized for interpretation and analysis.

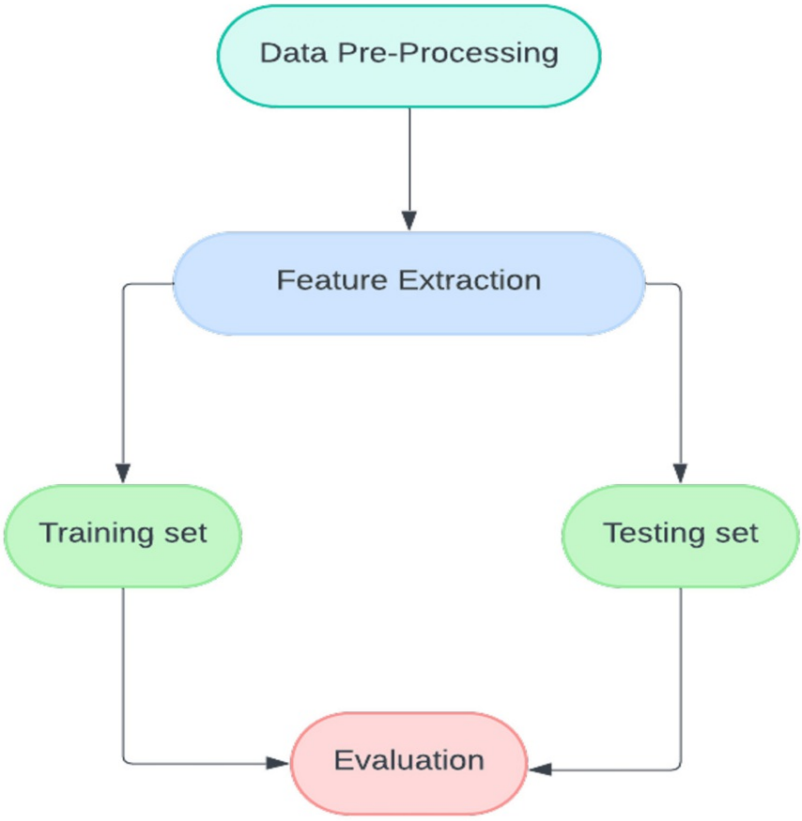

FIGURE 2: Machine Learning Process Flowchart

This project will involve experimentation on one dataset and the use of three algorithms. Here, a European dataset is used to include transactions of credit cards for two days, September 2013. All fields, except time and amount, have undergone principal component analysis transformation. There are just 492 fraud incidents out of a total of 2,84,807. The proportion of fraud to genuine transactions is very small, making it a complicated task to differentiate between the two. The first major step is cleaning the dataset. There are several methods for cleaning datasets, such as the Synthetic Minority Over-Sampling Technique, correlation coefficient analysis, and others. However, there is no need to implement these data-cleaning methods because datasets containing all numerical values are used. That said, the study falls short in terms of data preprocessing. It does not provide detailed methods for data cleaning, which are essential for maintaining data quality. Additionally, the research does not incorporate advanced techniques like the Synthetic Minority Over-Sampling Technique to address class imbalance or clustering methods to group similar transactions. Incorporating these advanced techniques could significantly enhance not only model performance but also the accuracy of fraud detection by refining the dataset and revealing hidden patterns [13].

The histogram of the amount, as shown in Figure *3*, reveals that the amounts are spread quite evenly with no noticeable spikes or dips. It also shows that the frequency of values remains comparably constant across different ranges of amounts. This consistency does not, however, always mean that it is difficult to spot unusual transactions. Comparing characteristics other than the frequency and amount is often how anomalous transactions are found. The detection of fraud may be significantly aided by additional characteristics like transaction timing, trends, or behavior.

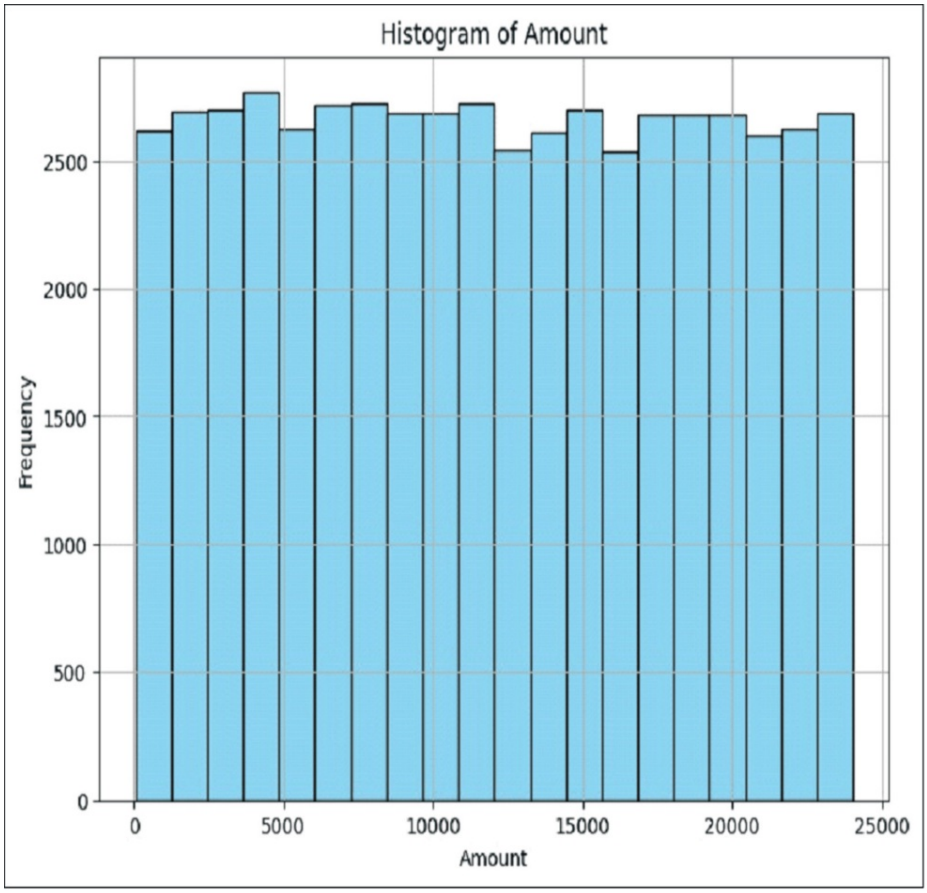

FIGURE 3: Amount vs. Frequency

The line of code heatmap(df.isnull(), yticklabels=False) is of Figure *4*, which is the method of showing the heatmap. If the heatmap is black, it indicates that every value in our data frame is present, and our data is complete and clean, making it suitable for further analysis.

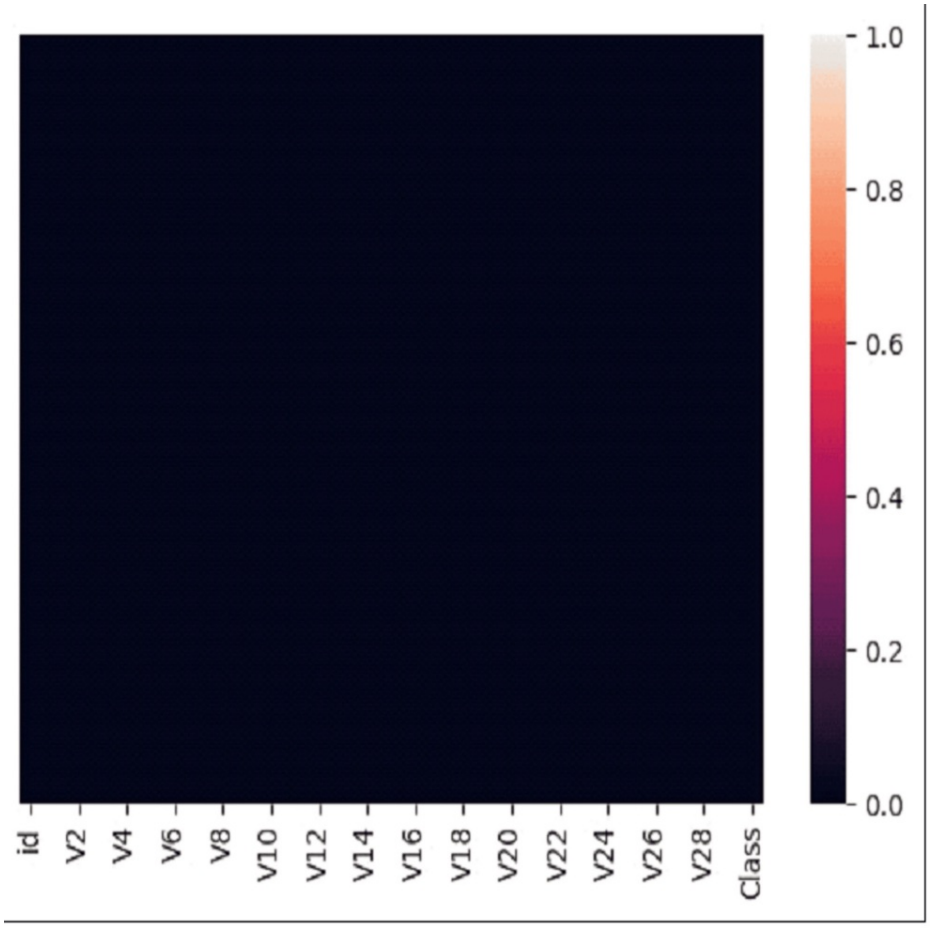

FIGURE 4: Heatmap of the Dataset

Figure *5* shows a scatter plot comparing fraud (red) and non-fraud (blue) transactions based on transaction amount and frequency (V1). Non-fraudulent transactions are widely distributed across the V1 and amount ranges, whereas fraudulent transactions are less frequent and primarily occur at lower V1 values.

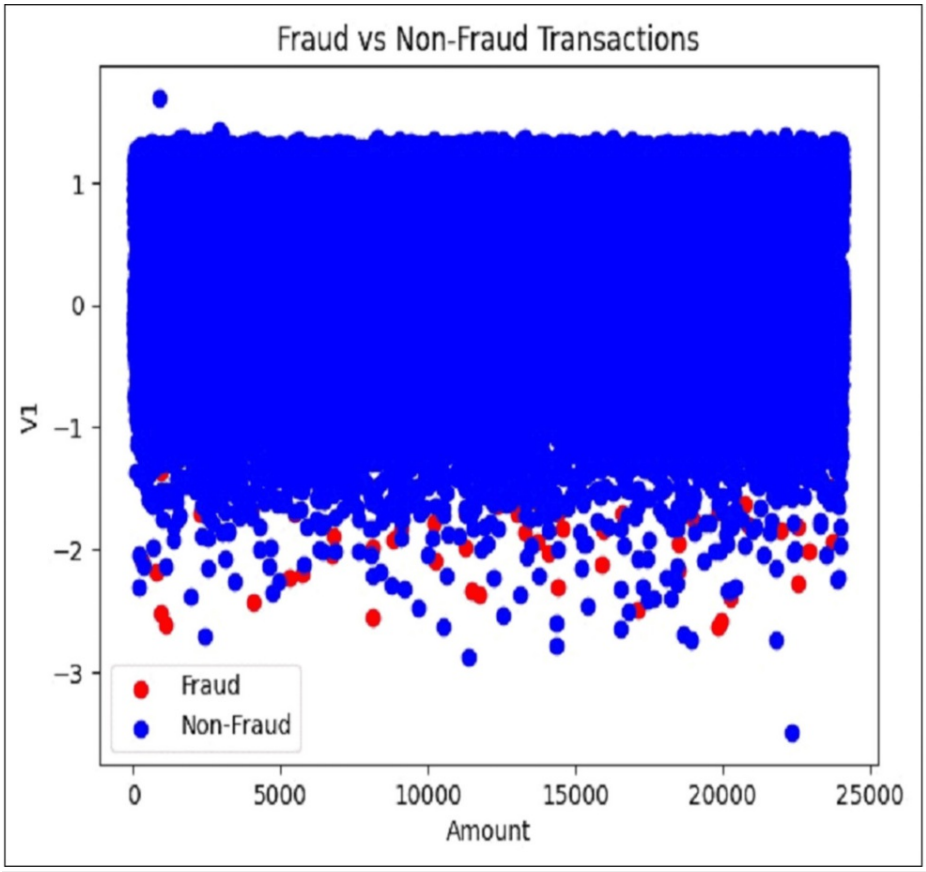

FIGURE 5: Fraud and Non-Fraud Transactions

Machine Learning Algorithms

Logistic regression: Logistic regression is a technique for replicating the probability of a specific result depending on one or more input factors [14-16]. In linear regression, the aim is to find the relationship between two variables, with one being dependent and the other independent.

According to Abbasi et al. [17], perhaps, one of the most widely used techniques for determining financial fraud detection models is logistic regression. To analyze the efficiency of the logistic regression algorithm, a transactional dataset from Kaggle was used, and the algorithm was implemented step by step. Instead of using a conventional regression line, logistic regression takes advantage of an "S"-shaped logistic function that predicts two possible outcomes (0 or 1).

By using logistic regression to generate a classification algorithm, the artificial intelligence system enhanced fraud detection, resulting in reduced percentages of credit card fraud. This also led to better performance through class action, including higher accuracy rates, higher sensitivity, and a desired error rate [18].

Decision tree: A decision tree is a popular prognostic modeling technique. To solve the issue at hand, a decision tree algorithm splits the dataset based on specific criteria. Regression and classification tasks can be performed with decision trees, which may adapt their structure to fulfill the demands of the specific application [19-21]. In this tree, each node visualizes a test on a feature like a decision node, terminal node, splitting, pruning, root node, or branch. The types of decision trees are categorical and conditional trees. The impurity of a node can be used as a formula of entropy.

The decision tree algorithm is based on entropy, also known as Gini impurity. Entropy is used to measure the level of impurity in a node.

Decision tree is one of the predictive modeling approaches used in machine learning. The algorithm divides a dataset into smaller subsets based on various parameters. Decision trees can be built for both classification and regression. Each node in the tree represents a test of a particular property, such as a limb, decision node, terminal node, splitting, pruning, or root node. The types of decision trees include categorical and conditional trees. The entropy formula is used to find the impurity in a node.

Random forest: Random forest is a supervised machine learning algorithm that is highly popular. It is used in both differentiation and regression problems in machine learning. When carefully observing the structure of a random forest, it becomes clear that multiple tree-like structures are present within a single unit [12,22,23]. The algorithm takes the average of all decision trees to produce the correct output. Here, we used multiple classifiers to solve difficult problems.

Ensemble learning: Our analysis of numerous publications shows that developing a model with a single classifier is ineffective for transactional fraud detection. Every classifier has advantages and disadvantages. Numerous writers have proposed that employing one or more classifiers can overcome a single classifier's shortcomings. A hybrid model is one that was created utilizing many classifiers [24]. The method of machine learning, commonly referred to as "ensemble learning," involves a lot of models, also called "learners" or "base models," to train them to figure out a single issue. Ensemble learning is based on the fundamental tenet that an ensemble can often surpass a single model when its predictions are combined, rather than just one [25-27]. There are various ensemble methods like bagging, boosting, stacking, and voting. In this paper, the stacking method is used. The process of stacking involves teaching a meta-model how to incorporate the forecasts from multiple base models. The meta-model obtains inputs from the underlying models, which are trained concurrently, and uses their predictions to generate the final prediction. Comparing the performance of individual models (logistic regression, random forest, and decision tree) with an ensemble method (stacking classifier) will give you insights into the advantages of combining models.

Figure *6* depicts a flowchart illustrating an ensemble method in machine learning. Input Data: Serves as the starting point for the process. Logistic Regression, Decision Tree, and Random Forest: These are three individual machine learning models that independently process the input data. LR Prediction, DT Prediction, and RF Prediction: Represent the individual predictions generated by each respective model. Ensemble Method: Combines the predictions from the three models to make a final, much more robust prediction. Final Output: The outcome generated by the ensemble method. This method can often lead to better accuracy and generalization analysis using a single model.

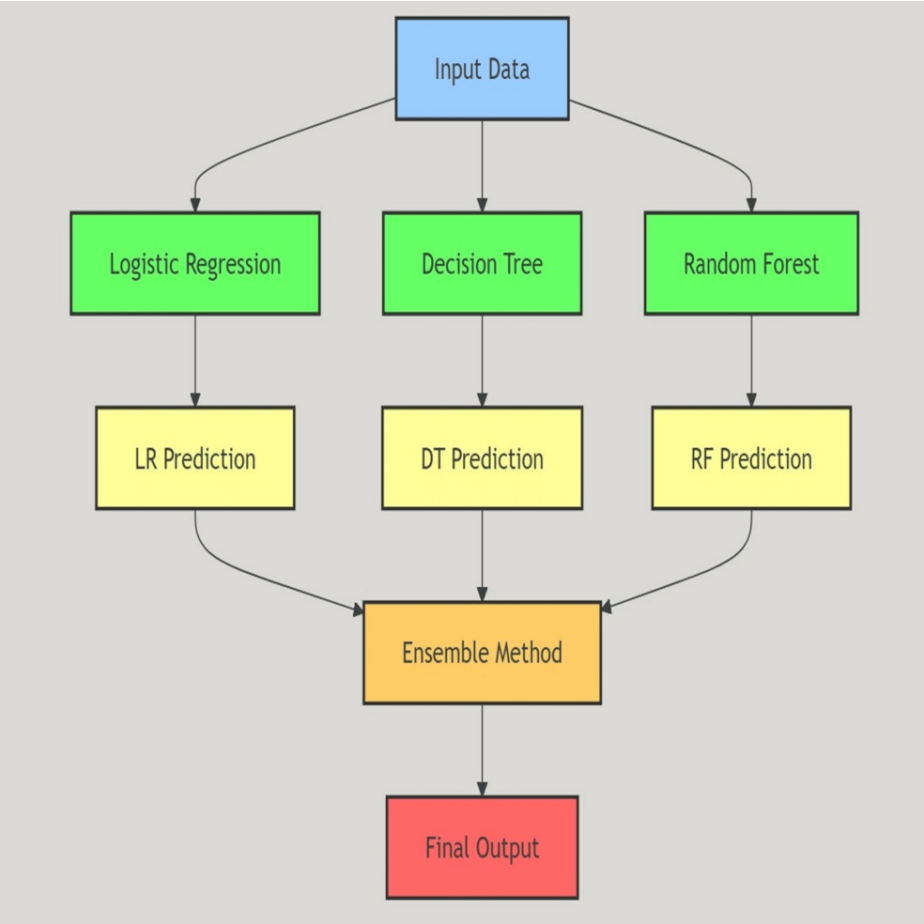

FIGURE 6: Flowchart of Working Principle of Ensemble Learning

LR: Logistic Regression; DT: Decision Tree; RF: Random Forest

Ensemble methods are techniques that combine multiple models to improve predictive performance. The image showcases a common approach where individual model predictions are aggregated to produce a final output.

Results

In this research paper, logistic regression was implemented using Google Colab with a publicly available dataset containing details of financial transactions. The dataset includes a column labeled “amount,” which represents the dependent variable, while the independent variables are represented by columns V1 to V28. The “amount” column is part of the dataset and is specifically used as the target variable for prediction.

To prepare the dataset for analysis, it was first cleaned by removing missing (Na) values, and the variables were sorted into dependent (target) and independent (predictor) categories. Various libraries were utilized, including the sklearn.linear_model module for creating logistic regression models and functions for computing evaluation metrics such as accuracy, precision, recall, and F1-score.

The metrics presented in Table *3* are derived from the experimental results conducted using the Kaggle Credit Card Fraud Detection Dataset 2023.

Model	Accuracy	Precision	Recall	F1 score	AUC-ROC
Logistic Regression	99.90	86.36	67.85	76.00	83.91
Decision Tree	99.91	76.54	78.48	77.49	-
Random Forest	99.93	90.00	85.71	87.80	92.84
Ensemble	99.95	99.96	99.95	99.95	99.95

TABLE 3: Performance Comparison of Individual and Ensemble Models on Key Metrics

AUC-ROC: Area Under the Receiver Operating Characteristic Curve

Logistic Regression: Achieved high Accuracy (99.31%) and balanced Precision, Recall, and F1-score (~99.3%) but showed a low area under the receiver operating characteristic curve (AUC-ROC) (52.19%), indicating it struggled to distinguish between classes effectively.

Decision Tree: Delivered excellent Accuracy (99.90%) but lower Precision (81.66%), Recall (74.24%), and F1-score (77.77%), highlighting potential issues with false positives and false negatives. AUC-ROC data were not available.

Random Forest: Demonstrated strong overall performance with Accuracy (99.93%), Precision (90%), Recall (85.71%), and F1-score (87.80%), along with a significantly high AUC-ROC (92.84%), suggesting its robustness in classification tasks.

Stacking Classifier: Outperformed all individual models, achieving near-perfect metrics across all categories (Accuracy, Precision, Recall, F1-score, and AUC-ROC at 99.95%), showcasing the advantage of combining multiple models to enhance predictive performance.

Figure *7* illustrates the comparative performance of Logistic Regression, Decision Tree, Random Forest, and Stacking Classifier models across key metrics: Accuracy, Precision, Recall, F1-score, and AUC-ROC. The Stacking Classifier consistently outperforms other models, demonstrating its effectiveness in transactional fraud detection.

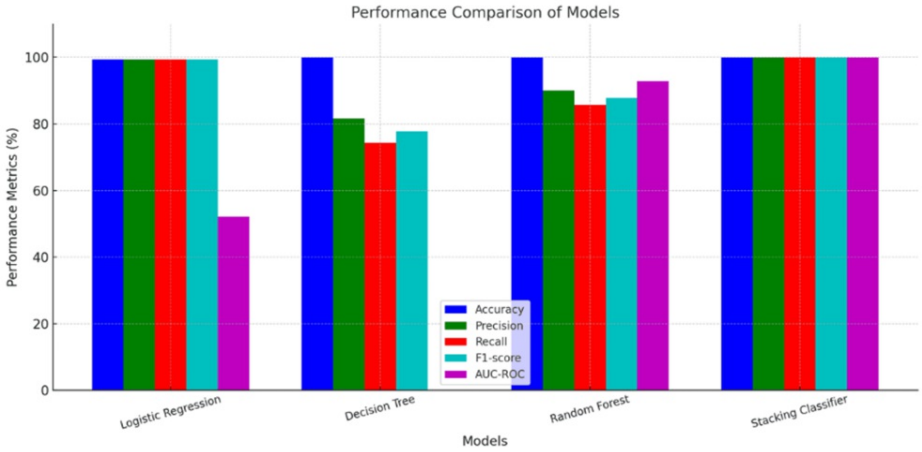

FIGURE 7: Performance Metrics Comparison of Machine Learning Models

Discussion

Limitations and observations

Throughout the analysis, a number of difficulties and limitations were identified:

Decision Tree Model Overfitting: Despite having good accuracy, the decision tree model has a tendency to overfit the training data, as shown by its lower recall and F1-score when compared to other models. Pruning or adjusting hyperparameters could help with this problem.

Imbalance in the Dataset: The dataset showed a significant disparity between authentic and fraudulent transactions. Although methods like SMOTE were used to address this issue, investigating more sophisticated sampling techniques or cost-sensitive learning procedures could improve performance even more.

Restricted Generalizability: The dataset used was designed for particular transactional scenarios. Validation on a wider and more varied range of datasets may be necessary to determine the model's applicability to additional datasets or real-world scenarios.

The logistic regression model has attained an accuracy of 99.90 % with a precision of 86.36%. It is accurate nearly three-quarters of the time when a given model predicts a positive outcome. In all the previous papers, the maximum research precision has gone up from 28% to 32 % with an accuracy of 95%. Also, the model demonstrated a high recall of 67.85%, which gives insights into the model’s ability to find all positive values from the dataset. The ROC-AUC score provides a single, comprehensive metric for assessing the complete performance of the model in terms of its ability to separate positive and negative classes, irrespective of the chosen threshold. Below, in Figure *8*, are the evaluated metrics for predicted values and actual values.

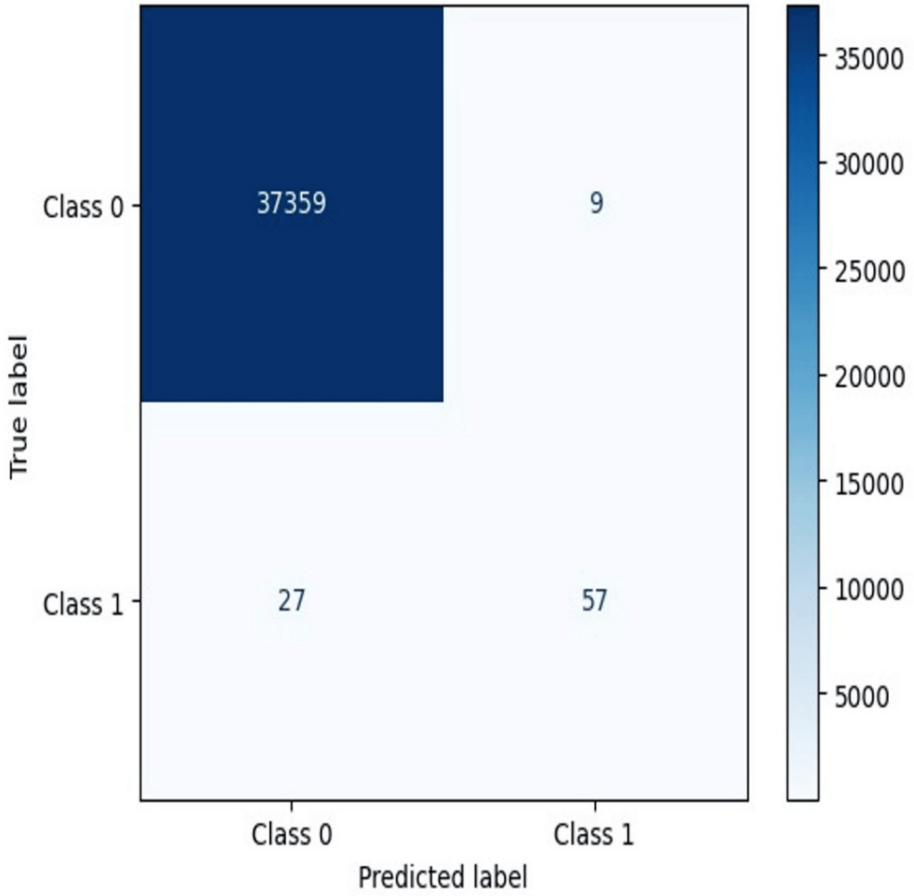

FIGURE 8: Confusion Matrix of Logistic Regression

The decision tree algorithm shows remarkable performance in classifying fraud transactions with an accuracy of 99.91%, precision of 76.54%, recall of 82.69%, F1 score of 78.48%, and ROC-AUC score of 77.49%. Like in random forest algorithms, this algorithm also accurately identifies fraud. These findings demonstrate the model's effectiveness in spotting fraudulent transactions with minimal false positives. Below, in Figure *9*, the confusion matrix for the decision tree is provided.

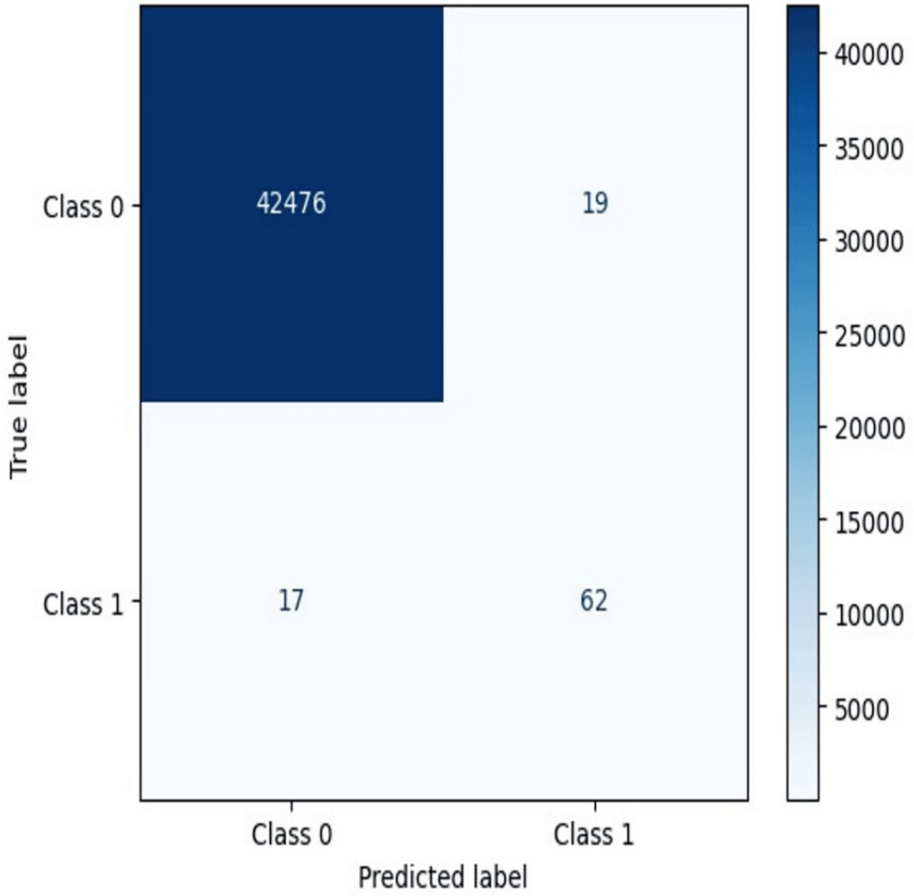

FIGURE 9: Confusion Matrix of Decision Tree

A random forest algorithm has achieved exceptional performance with an accuracy of 99.93%, precision of 90%, recall of 85.71%, F1 score of 87.80%, and ROC-AUC score of 92.84%. The model demonstrated strength in accurately identifying fraud transactions with minimum false positives. The random forest model's high recall score further indicates its ability to effectively capture instances of fraud, crucial for minimizing financial losses and enhancing security measures in payment systems. Below, in Figure *10*, the confusion matrix for the random forest is provided.

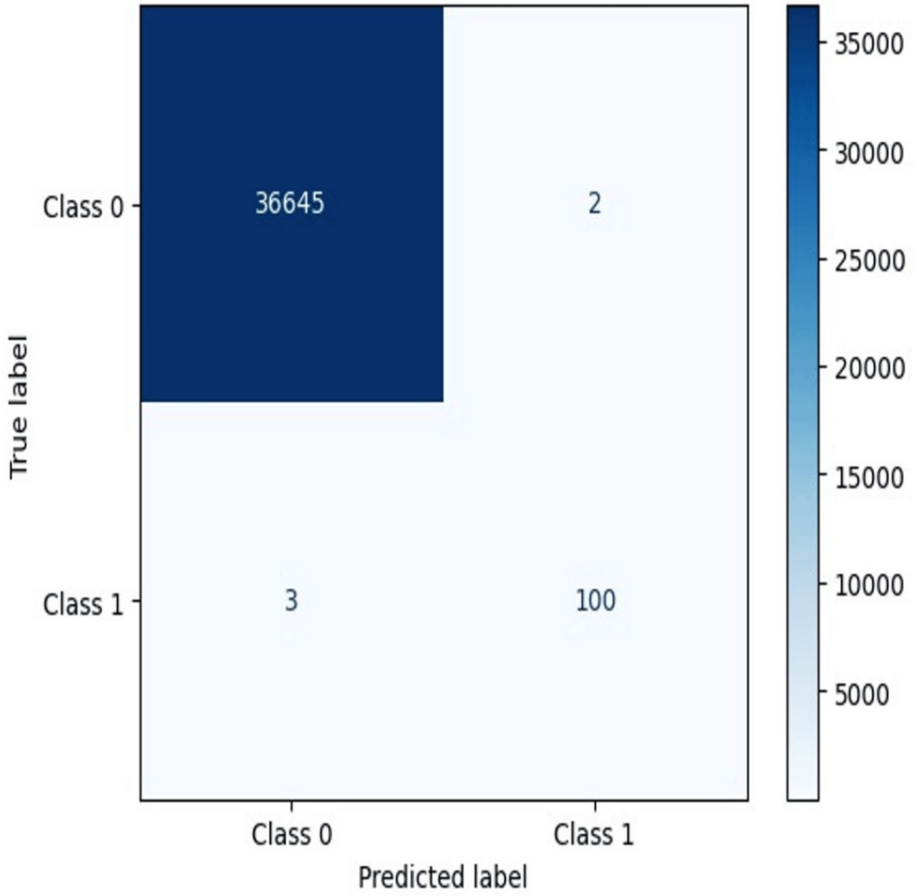

FIGURE 10: Confusion Matrix of Random Forest

The confusion matrix in Figure *11* represents the performance of the ensemble learning model, consisting of two classes, 0 and 1. The model correctly predicted class 0 in 56,863 instances, with only 1 incorrect prediction. For class 1, the model made 75 correct predictions, while 23 predictions were incorrect.

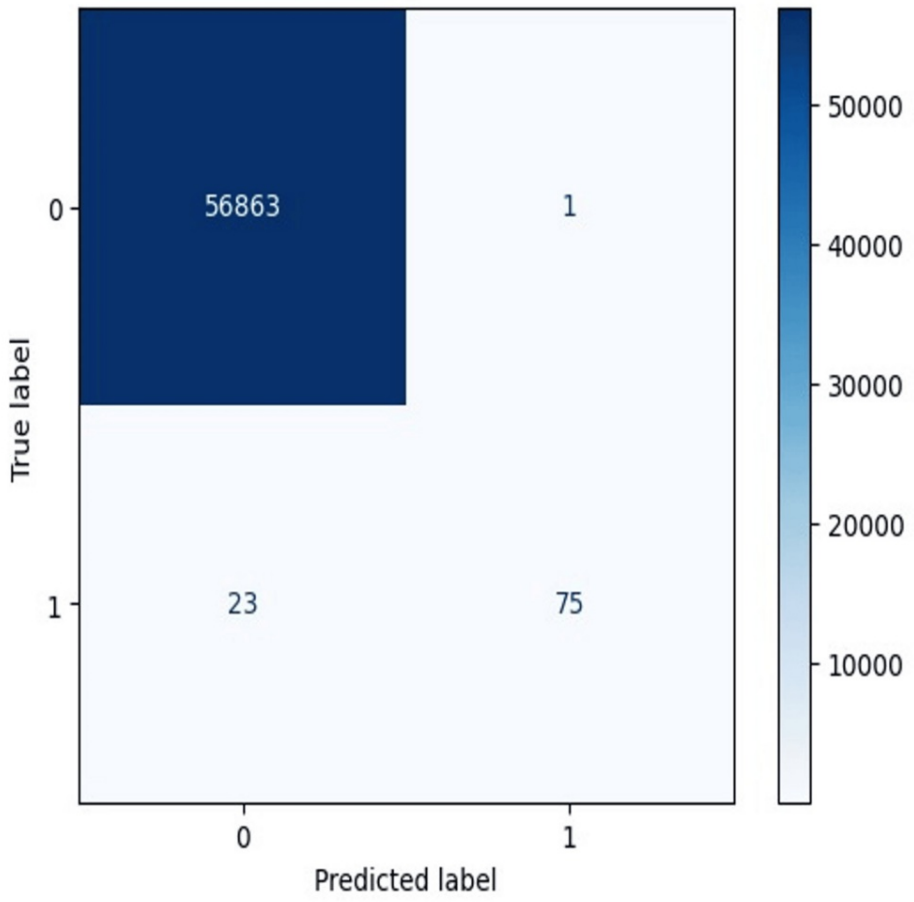

FIGURE 11: Confusion Matrix of Ensemble learning

The experimental analysis presented reveals the advantages of combining multiple machine learning algorithms within an ensemble framework. Logistic regression contributed to the accuracy of predictive estimates, and decision tree and random forest improved recall and handled complex relationships between the attributes within the data. The ensemble model showed superior performance because of its nature of adding predictions and aggregating estimates to reduce variance. This study's findings align with previous research, confirming the robustness of hybrid approaches in fraud detection. Limitations include the dataset's imbalance and potential overfitting in random forest. Future research could explore deep learning techniques and additional feature engineering for further improvement.

Areas for improvement and further research

Exploring other ensemble methods like boosting (e.g., XGBoost, AdaBoost) or hybrid approaches combining stacking with boosting for better performance.

Explainability and interpretability: Adding model interpretability techniques like SHAP or LIME to understand decision-making and ensure transparency in fraud detection models.

Real-time evaluation: Extending the analysis to evaluate models in real-time transactional systems for latency and deployment challenges.

These observations provide a foundation for future work aimed at refining and expanding the utility of fraud detection systems.

Conclusions

This paper explores the application of machine learning algorithms for detecting financial fraud, specifically using three algorithms: logistic regression, decision tree, and random forest, all applied individually and combined using ensemble learning. It was noted that random forest and decision tree models achieved the highest accuracy, while logistic regression’s accuracy was similar to its precision. Although logistic regression is simple and intuitive, it struggles to capture complex relationships, which ultimately makes it weak at independently discovering fraud cases. Decision trees are good at capturing nonlinear relationships

but are prone to overfitting, and they work better with lower dimensions. Nevertheless, their output is more precise that ultimately makes it weak at independently discovering fraud cases.. In addressing the overfitting issue that an ensemble of decision trees called random forest provides, averaging the results of numerous trees leads to a low-bias, low-variance model that works well across a range of problems. Thus, although having multiple tree structures, a single decision tree cannot match this level of performance, thus improving fraud detection accuracy. Moreover, the stacking method of ensemble learning combined all three algorithms and improved the results, showing greater generalization and strengthened the base models. This ensemble exhibits high accuracy and precision and recall compared to other classifiers, making it a very efficient tool for detecting fraudulent transactions.

This study highlights the efficiency of hybrid machine learning approaches by combining models to address such issues as overfitting and underfitting. Furthermore, advanced techniques of deep learning can be explored with real-time detection systems to enhance fraud detection capabilities. This paper contributes considerably to the idea of ensemble methods for combating financial fraud, showing that these methods can indeed produce better results than the usual models.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Yugal Salunke, Saroj Phalke, Praali Kumre

Acquisition, analysis, or interpretation of data: Yugal Salunke, Saroj Phalke, Manoj Madavi, Grishma Bobhate

Drafting of the manuscript: Yugal Salunke, Saroj Phalke, Manoj Madavi, Praali Kumre

Critical review of the manuscript for important intellectual content: Yugal Salunke, Saroj Phalke, Manoj Madavi, Grishma Bobhate

Supervision: Yugal Salunke, Manoj Madavi, Praali Kumre, Grishma Bobhate

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Ileberi E, Sun Y, Wang Z: A machine learning based credit card fraud detection using the GA algorithm for feature selection. Journal of Big Data. 2022, 9:24. 10.1186/s40537-022-00573-8
2. Bakumenko A, Elragal A: Detecting anomalies in financial data using machine learning algorithms . Systems. 2022, 10:130. 10.3390/systems10050130
3. Wang L, Zhang Z, Zhang X, Zhou X, Wang P, Zheng Y: A deep-forest based approach for detecting fraudulent online transaction. Advances in Computers. 2021, 120:1-38. 10.1016/bs.adcom.2020.10.001
4. Maniraj SP, Saini A, Ahmed S, Sarkar S: Credit card fraud detection using machine learning and data science. International Journal of Engineering Research & Technology. 2019, 8:110-115.
5. Raghavan P, Gayar NE: Fraud detection using machine learning and deep learning . 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE). 2019, 334-339. 10.1109/ICCIKE47802.2019.9004231
6. Gheisari M, Ebrahimzadeh F, Rahimi M, et al.: Deep learning: Applications, architectures, models, tools, and frameworks: A comprehensive survey. CAAI Transactions on Intelligence Technology. 2023, 8:581-606. 10.1049/cit2.12180
7. Amarasinghe T, Aponso A, Krishnarajah N: Critical analysis of machine learning based approaches for fraud detection in financial transactions. Proceedings of the 2018 International Conference on Machine Learning Technologies. 2018, 12-17. 10.1145/3231884.3231894
8. Thimonier H, Popineau F, Rimmel A, Doan B-L, Daniel F: Comparative evaluation of anomaly detection methods for fraud detection in online credit card payments [PREPRINT]. arXiv:2312.13896. 2023, 10.48550/arXiv.2312.13896
9. Al Marri M, AlAli A: Financial fraud detection using machine learning techniques. Thesis. Rochester Institute of Technology. (2020). https://repository.rit.edu/theses/10695/.
10. Ranjitha H, Joshni PS, Vaishnavi CS: Vehicle insurance fraud detection using machine learning .

International Journal of Advanced Scientific Innovation. 2023, 5:1-6.
11. Aditya O: Fraud detection using machine learning. Transfer. 2018, 528812:532909.
12. Breiman L: Random forests. Machine Learning. 2001, 45:5-32. 10.1023/a:1010933404324
13. Dornadulaa VN, Geetha S: Credit card fraud detection using machine learning algorithms . Procedia Computer Science. 2019, 165:631-641. 10.1016/j.procs.2020.01.057
14. LaValley MP: Logistic regression. Circulation. 2008, 117:2395-2399. 10.1161/CIRCULATIONAHA.106.682658
15. Nick TG, Campbell KM: Logistic regression. Topics in Biostatistics. Ambrosius WT (ed): Humana Press, NJ; 2007. 404:273-301. 10.1007/978-1-59745-530-5_14
16. Kleinbaum DG, Klein M: Logistic Regression: A Self-Learning Text. Springer, New York; 2002. 10.1007/978-1-4419-1742-3
17. Abbasi A, Albrecht C, Vance A, Hansen J: MetaFraud: A meta-learning framework for detecting financial fraud. MIS Quarterly. 2012, 36:1293-1327.
18. Alenzi HZ, Aljehane NO: Fraud detection in credit cards using logistic regression . International Journal of Advanced Computer Science and Applications. 2020, 11:45-52. 10.14569/IJACSA.2020.0111265
19. Song Y-Y, Lu Y: Decision tree methods: applications for classification and prediction. Shanghai Archives of Psychiatry. 2015, 27:130-135. 10.11919/j.issn.1002-0829.215044
20. Sutharan S: Machine Learning Models and Algorithms for Big Data Classification. Suthaharan S (ed): Springer, New York; 2016. 10.1007/978-1-4899-7641-3
21. De Ville B: Decision trees. WIREs Computational Statistics. 2013, 5:448-455. 10.1002/wics.1278
22. Rigatti SJ: Random forest. Journal of Insurance Medicine. 2017, 47:31-39. 10.17849/insm-47-01-31-39.1
23. Biau G, Scornet E: A random forest guided tour. TEST. 2016, 25:197-227. 10.1007/s11749-016-0481-7
24. Patil RR, Kaur G, Jain H, Tiwari A, Joshi S, Rao K, Sharma A: Machine learning approach for phishing website detection: A literature survey. Journal of Discrete Mathematical Sciences and Cryptography. 2022, 25:817-827. 10.1080/09720529.2021.2016224
25. Ganaie MA, Hu M, Malik AK, Tanveer M, Suganthan PN: Ensemble deep learning: A review. Engineering Applications of Artificial Intelligence. 2022, 115:105151. 10.1016/j.engappai.2022.105151
26. Dong X, Yu Z, Cao W, Shi Y, Ma Q: A survey on ensemble learning . Frontiers of Computer Science. 2020, 14:241-258. 10.1007/s11704-019-8208-z
27. Polikar R: Ensemble learning. Ensemble Machine Learning: Methods and Applications. Zhang C, Ma Y (ed): Springer, New York, NY; 2012. 1-34. 10.1007/978-1-4419-9326-7_1