Journal of Posthumanism

2025
Volume: 4, No: 3, pp. 2157–2184
ISSN: 2634-3576 (Print) | ISSN 2634-3584 (Online)
posthumanism.co.uk

DOI: https://doi.org/10.63332/joph.v4i3.3593

Intelligent Fraud Detection: Applying Advanced Analytics and Cybersecurity Insights in U.S. Finance

Mohammad Shahidullah[1], Hammed Esa[2], Md Abdur Rob[3], Md Bayzid Kamal[4], Md Mohaimin Rashid[5], Md Fakhrul Hasan Bhuiyan[6], Md Shayakh Alam[7], Durga Shahi[8]

Abstract

Fraud detection in financial transactions is a major and crucial problem that does not cease to exist, mainly because of the enormous imbalance in the datasets obtained and the very high requirement for an accurate distinction between legitimate and fraudulent activities. In the following study, we assess the performance of three common machine learning models: Logistic Regression, Random Forest, and Gradient Boosting, for the detection of fraud, using a real-data set of transactions (284807 of which only 0.173% are labelled as fraudulent). The models were thoroughly evaluated with respect to critical metrics of performance including the precision, recall, F1-score and Area Under the Receiver Operating Characteristic Curve (AUC) to try to understand which of the models may be appropriate for dealing with class imbalance and false positives. Of the analyzed models, Random Forest was the best, with AUC being equal to 0.98, being superior to Logistic Regression (AUC = 0.97) and equal to Gradient Boosting (AUC = 0.98), while enabling more superior recall (0.88) and precision (0.44). This implies a higher capacity of detecting fraud cases without compromising the rate of false alarm too much. Feature importance analysis further noted that V14, V10, and V4 features were most predictive and most responsible for model classifying accuracy. Furthermore, calibration analysis revealed that Random Forest was the most reliable in estimating probabilities, outputs closely conformed to the ideal calibration curve implicating better reliability in practical applications. These findings suggest the effectiveness of the ensemble machine learning models especially Random Forest in promoting the efficacy of fraud detection systems. The study supports future research on real-time deployment and integration with deep learning methods to enhance the strength of fraud detection in the constantly changing financial spaces.

Keywords: *Fraud Detection, Machine Learning, Financial Crime, Random Forest, Gradient Boosting, Logistic Regression, Precision-Recall Analysis.*

Introduction

With the rise of digital finance, the manner in which consumers and institutions handle transactions has changed. In addition, this rapid shift has made the financial industry a target for ever-more advanced online crime. Out of these problems, credit card fraud is a major issue and can badly affect the trustworthiness of online financial services (Pazarbasioglu et al., 2020). Not only do crimes like transaction fraud, account takeovers, and identity theft cost a lot of money, but they also lower trust in digital banking and payment services. Reports suggest that the

[1] Department of Business Administration International American University, Email: shahidbd2004@gmail.com
[2] Department of Business Administration International American University, Email: hammedesa@gmail.com
[3] Department of Economics Ohio University, Athens, Email: marob.sust2014@gmail.com
[4] Department of Business Analytics Brooklyn College, CUNY (City University of New York), Email: bayzidkamal181@gmail.com
[5] Department of Business Administration and International American University, Email: neeloyr26@gmail.com
[6] Department of Information Studies Trine University, Email: fakhrulbcef@gmail.com
[7] Master of Engineering Management Trine University, Email: shayakhalam23@gmail.com
[8] Department of Business Administration Westcliff University, Email: d.shahi.1396@westcliff.edu

financial sector in the U.S. loses billions of dollars to fraud every year, and global losses may rise sharply unless proper countermeasures are put in place. The growing threat points out the urgent importance of strong, adapting, and fast fraud detection systems to safeguard both customers and organizations.

A key difficulty in fraud detection is that the data is very biased toward one category called the majority class. (Al-Dahasi et al., 2024) mention that fewer than 0.2% of transactions in the real world are fraudulent. Traditional classification algorithms have a hard time with uneven classes because they mainly go after the majority and often ignore infrequent but important fraudulent examples. Simple 'if-then' logic-based systems aren't adaptable to new fraud patterns and often rely on people to update them to work well. Machine learning (ML) approaches, on the other hand, have shown notable capability in this area by making use of patterns in the data and evolving learning processes to spot fraud with more precision and efficiency (Mohan et al., 2022). Modern fraud detection systems need the ability to spot complex, non-linear patterns and hidden anomalies, and ML models can help with this.

Although many machine learning techniques are now used for fraud detection, there is still a major lack of research on how popular models stack up in terms of performance and understanding when working with very uneven financial datasets. Specifically, there is not much research looking at how models like Logistic Regression, Random Forest, and Gradient Boosting compare with each other in terms of accuracy, probability calibration, and feature interpretability. Paying attention to these factors is very important, because the chosen model's performance affects how well fraud is caught and how much financial institutions can trust the model in important decisions.

This research gap is addressed by this study, which evaluates and compares three of the most popular machine learning models in terms of their performance in detecting fraudulent financial transactions: Logistic Regression, Random Forest, and Gradient Boosting on a real-life dataset of financial transactions with over 56,962 records, with about 1% labeled fraudulent. All the models selected were based on their methodological strengths. Logistic Regression is a linear model which is simple and easy to interpret; Random Forest is an ensemble technique that can handle high dimensional data as well as a non-linear pattern; and Gradient Boosting is a technique- capable of making strong predictions and learning complicated interactions among features iteratively. The models are evaluated by bench mark evaluation metrics such as precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC) which gives a complete understanding of the classification performance especially class imbalances.

A key problem when performing fraud detection is achieving high precision and recall simultaneously. Again, high recall is essential to capture most fraud cases possible, while high precision is also necessary to minimise false positives, which can cause declines in actual customer transactions and poor customer satisfaction. Furthermore, this work explores model calibration that helps ascertain the accuracy of predicted probabilities, given that working with financial issues enshrined with sensitivity is crucial. We propose to examine the calibration performance of each model to assess which model provides the most accurate probability estimates of fraudulent behaviour to financial institutions.

The aim of this study is twofold: (1) methodological – to assess and compare the cut-off point and other performance measures of Logistic Regression, Random Forest, and Gradient Boosting in handling imbalanced data to classify fraud cases, and (2) to reveal which features have the most significant impact on the likelihood of fraud. Therefore, through performance measures and interpretability results, this study will equip financial institutions with knowledge of advanced analytics solutions to fraud. Besides, the findings show the advantages and disadvantages of each model and inform the subsequent research and development of improving artificial intelligence in achieving financial security.

This study makes the following major contributions:

1. An extensive comparative study of three popular machine learning algorithm on big and imbalanced financial transaction data with focus on the model effectiveness in fraud detection.
2. A comparison of calibration of each model, providing understanding to reliability and trustworthiness of fraud predictions, which is essential for real application deployment.
3. A study of feature importance to determine how transaction attributes most signify fraudulent activity, hence contributing to model explainability and decision transparency.
4. A hands-on exploration of the tradeoffs between detection accuracy, interpretability, and computational efficiency in the hope of helping financial institutions choose suitable models for fraud mitigation.

Literature Review

Fraud Detection Techniques

Fraud detection is one of the most significant topics of interests in financial security research due to potential impacts on consumer and institutions by fraudulent activities leading to significant fiscal losses (Hilal et al., 2022). In the past, fraud detection was limited to rule-based systems where a certain specific value and the existing model are used as a start of identification. As for the advantages of these systems, they are quite simple and globally interpretable. However, they must be more flexible since fraudsters commonly change their approach (Nesvijevskaia et al., 2021). This rigidity has created the need for more dynamic fraud detection methods like statistical analysis and more recent and more effective Machine Learning techniques for detecting fraud from changing transactional data.

More recently, a highly effective approach for fraud prevention, machine learning involves learning from data without human input (Bello et al., 2024). Compared to the conventional Anti-fraud rule-based system, which can be a continuous system that can learn new types of frauds, the machine learning approach is ideal for real-time anti-fraud detection systems (Bello et al., 2023). Training methods which include supervised training of the models with predefined inputs have had some success in this area. Logistic regression, random forest, and gradient boosting models have been identified as more suitable for using imbalanced and non-linear financial transactions data, with very high accuracy in fraud determination. The application of machine learning in the detection of fraud recommendations

Logistic Regression

Logistic Regression is a threadbare linear classification algorithm and a go-to baseline model for many fraud detection research works because of its simplicity and explainability. Logistic Regression is one of those algorithms that directly model probabilities for binary outcomes so it is quite helpful when deciding between fraudulent and legitimate transactions (Thétard, 2021). Logistics regression have been proven to work well for typical fraud detection tasks Nevertheless it can be outperformed with more complex relations in the transaction data (Thétard, 2021). However, because of the simplicity and relatively low computational complexity of implementation, Logistic Regression represents a good starting model against which more complex solutions may be compared.

Random Forest

Random Forest, a type of ensemble learning, has contributed to the area for strong reasons, including robustness and the likelihood of feature importance analysis (Zhao et al., 2022). Based on the decision trees technique, Random Forest reduces the probability of over-learning the data by generating many trees and providing the outcomes averaged from the trees as the prediction outputs. As a result, the prediction will be more generalised. Some papers show that Random Forest is useful in working with imbalanced data because it may provide higher importance values for the minority class, such as fraudulent transactions (Ghaleb et al., 2023). Moreover, its feature importance analysis is helpful in learning the most important attributes that help identify patterns with features related to fraud.

Gradient Boosting

Gradient Boosting is one of the ensemble learning strategies for constructing successive models that are iteratively established to minimise errors made by the previous models and has proven to be effective at handling imbalanced fraud datasets. By incorporating decision trees containing the most ambiguous data samples, Gradient Boosting enjoys high constructive accuracy rates of precision and recall, especially in fraud detection applications where missing a single fraudulent case might be expensive (Vichare, 2024). Scholars also established that models like XGBoost and LightGBM, known to belong to the Gradient Boosting family, effectively capture nonlinear relationships in large datasets. However, they are more time-consuming than other method (Yoon et al., 2023). However, Gradient Boosting can be easily interpreted, although after creating several layers or rounds of learning, the model resembles a 'black box' (Bodria et al., 2023).

Challenges in Fraud Detection

Fraud detection emerged as more difficult given that most financial transaction data had significantly few fraud cases (Hilal et al., 2022). Models favouring the biggest class will arise, compromising the recognition of fraud cases. Such synthetic sample-creating strategies, such as the Synthetic Minority Over-sampling Technique (SMOTE), are widely used to solve this problem by making the model sensitive to fraud (Ghaleb et al., 2023).

Another issue is to provide an equal trade-off of precision and recall in fraud detection. Few false alarms are generated, so legitimate buyers are not inconvenienced, while most frauds are detected by maximising recall (Rodríguez Vaquero, 2023). This trade-off is fundamental in practice as false positives turn customers away, and false negatives cost a lot of money. Model calibration is also important when producing probability estimates so that managers can easily determine the risk of fraud related to a particular transaction (Vanini et al., 2023). This paper emphasises the need for calibration because the flawed models could give out a wrong probability, which hampers with fraud-fighting techniques.

Gap Analysis

Although a large body of research has addressed the use of machine learning models for fraud detection, limited work has been conducted to compare and contrast the results of Logistic Regression, Random Forest, and Gradient Boosting on a single dataset. To address this gap in the existing literature, this research seeks to assess the efficiency of these models based on commonly used accuracy measures like precision, recall, F1-score, AUC and calibration with an actual credit card transaction dataset with imbalanced class distribution. Moreover, this research offers feature importance analysis in the context of the Random Forest model and gives helpful interpretability results for feature selection in practice for financial institutions.

Therefore, the literature also points to the necessity of developing versatile, sophisticated approaches to deal with such specificities as imbalanced data and the trade-off between precision and recall. Random Forest and Gradient Boosting have been claimed to manage such concerns, but comparative analysis is sparing. While previous studies explore specific models for fraud detection in financial organisations, this research attempts to advance prior work and systematically assess these models on a massive actual dataset, thus providing crucial information for practising applying to improve the performance of the fraud detection systems in the financial context.

Methodology

This study adopts a systematic and data-driven approach to evaluate the effectiveness of advanced machine learning techniques for detecting fraudulent transactions within U.S. financial institutions. Given the highly imbalanced nature of financial transaction data and the need for both high precision and high recall, a robust methodology was designed to address the dual challenges of classification accuracy and probability calibration. This section details the dataset used, preprocessing steps, model selection rationale, evaluation metrics, and model interpretability strategies implemented in the study.

Dataset Description

The data set employed in this study from Kaggle as Credit Card Fraud Detection dataset (https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud/data) has a public domain credit card with 284807 records and 30 binary and numerical features with 1 class indicating a fraudulent transaction (Class 1) or non-fraudulent transaction (Class 0). The features named V1 to V28 were developed from a PCA-transformed data set to obfuscate the data for privacy while preserving transaction patterns. Additionally, the dataset includes two non-anonymised features: Time is the

number of transactions between the first transaction and after the first encounter, and Amount is the value for each particular transaction. The pre-processed dataset contains fraudulent transactions that comprise only 0.173% of the whole data set, leading to a class imbalance challenge for accurate fraud prediction.

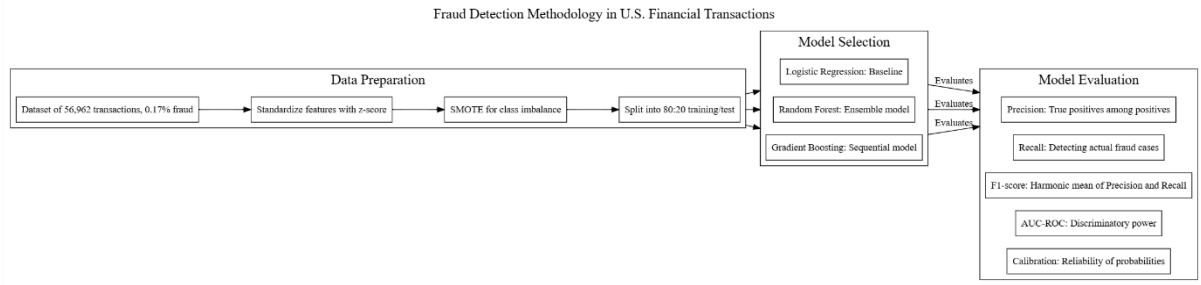

Figure 1: Proposed Methodology Diagram

Data Preprocessing

Due to the large amount of irrelevant and noisy information in the given dataset and even more due to the significant overlap between positive and negative classes, several preprocessing steps were applied to improve the quality and applicability of the gathered data to machine learning algorithms. Initially, missing values were inspected, but the given dataset did not include any, so the models could be trained without dealing with the missing values.

1. **Log Transformation for Skewed Features**

   While many columns in the dataset are Principal Component Analyzed and consequently normalized, the Amount column is positively skewed. Taking the log of each value makes the distribution more normal and reduces the spread, thereby making model learning more effective.

$$X' = \log(1 + X)$$

   Where, X is the original transaction amount and X' is log-transformed value.

2. **Correlation Analysis and Redundancy Removal**

   It may be observed that some anonymized PCA results show a lot of correlation between them. If the correlation between many features is very high, it may result in redundancy, multicollinearity, and overfitting in logistic regression. We apply the Pearson correlation coefficient to find and remove such features.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

   Where $\rho_{X,Y}$ represents pearson correlation coefficient between features X and Y, $cov(X,Y)$ is the covariance and σ_X, σ_Y are standard deviations of X and Y.

3. **SMOTE**

To address the problem of class imbalance, synthetic data for the minority class namely fraudulent transactions were created using the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE is an effective synthetic data generation technique that synthesises new samples from existing cases by determining feature space proximity (Endres et al., 2022); hence, it enhances the train dataset representation of fraudulent transactions without cloned samples. This method is widely used in fraud detection since, along with increasing the proportion of instances from the minority class, it enables models to learn from them more efficiently, thus increasing overall sensitivity to fraud. The pre-processed dataset, balanced through SMOTE by following,

$$x_{new} = x_i + \lambda \cdot (x_n - x_i)$$

Where, x_i represents the minority class instance, x_n determines one of its k-nearest neighbors and $\lambda \in [0,1]$ denotes random number.

After that the dataset is splitted into training and test sets using an 80:20 ratio. This split ratio was taken because the models were trained with a good number of data while the remaining part was set aside to test the efficacy of the models.

Model Selection and Configuration

Three machine learning algorithms were selected for this study because they are popular and known for their performance in fraud detection. Both models were chosen based on their features and overall performance capabilities in modelling fraud in transactions.

1. Logistic Regression was chosen as a baseline model because the model is reasonably simple, easy to interpret, and fast at binary classification processes. Although Logistic Regression cannot handle interactions and non-parametric features in data, the error analysis of classification performance is quite simple in this study. Default hyperparameters were set for Logistic Regression to maintain a fair model without much tweaking. Whether or not a transaction is regarded as fraud in Logistic Regression is estimated by calculating the probability that it falls into the positive class. Logistic (sigmoid) function is used to calculate this probability:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}}$$

Where $P(Y = 1|X)$ is the probability of the transaction being fraudulent, β_0 is the intercept (bias), $\beta_1, \beta_2, \ldots, \beta_n$ are the model coefficients (weights) and $X_1, X_2, \ldots, X_n$ are the input features.

The approach allows the model to produce calibrated probabilities that are important for setting thresholds and scoring risks in fraud detection.

2. Random Forest was selected as the main model because its ensemble approach solves the problems of class imbalance and data being non-linear. It accomplishes this by using separate decision trees made from bootstrap samples of the training data and merging all the predictions to improve accuracy and avoid excessive learning from the data. In the case of

data that has an imbalanced number of samples, this method is effective as it identifies complex features and borders that linear models often can't handle.

By picking 100 decision trees, the balance between the complexity of the model and its computation speed was achieved. Besides, to reduce the impact of class unbalance, the class_weight parameter was chosen as "balanced", so that more weight would be applied to the minority class (fraudulent transactions) while training.

For a binary classification task, the Random Forest model predicts by taking the response from the majority of its decision trees.

$$\hat{y} = mode(h_1(X), h_2(X), \dots, h_T(X))$$

Where $h_1(X)$ is the prediction of the t-th decision tree for input feature vector X, T is the total number of trees (in this study, T=100) and $\hat{y}$ is the final predicted class (0 for legitimate, 1 for fraud).

Apart from classifying the data, Random Forest provides scores of feature importance to identify the main features in detecting fraudulent transactions, as analyzed in this study. Being able to interpret the data helps financial companies identify the reasons behind fraud in transactions.

3. A Gradient Boosting machine learning technique was used in the analysis since it sequentially improves predictive models by focusing on minor errors and identifying behavior signs of fraud. Random Forest builds many trees in parallel, but Gradient Boosting creates the trees in order, trying to fix the mistakes the previous ones made. Using the additive model, it is possible to examine the complicated connections usually found in unbalanced and messy financial fraud data.

   To ensure that the model works well on different data and avoid overfitting, the learning rate of 0.1 was set to control the effect of every new tree. At the same time, the structure of each tree was capped at three levels, so the learners do not become too powerful but still learn how features are connected. By boosting the minority class, the model was trained to pay more attention to catching fraud cases, which are fewer in number than legitimate cases.

   The model builds the final prediction function $F_M(X)$ using a stage-wise additive approach:

$$F_M(X) = \sum_{m=1}^{M} \gamma_m h_m(X)$$

   Where M is the total number of trees, $h_m(X)$ is the m-th weak learner (typically a decision tree), γ_m is the learning rate (in this study, 0.1) and $F_M(X)$ is the final prediction score for feature vector X.

Evaluation Metrics

Various evaluation metrics were employed to evaluate and compare the different models because each possessed its speciality in revealing multiple aspects of the extent and efficiency of the models employed to reduce fraudulence. Specifically, the large number of examples of the majority class makes it inadequate if the accuracy criterion is used, an indicator of the performance of a classifier alone – this is because the accuracy would be very high simply because the majority of the total number of examples belongs to the majority class. Hence, besides accuracy, Precision, Recall and F1-Score were computed to get a better comparative analysis of the performance of each model in fraud detection.

Recall on the other hand measures the number of actually fraudulent transactions captured through the algorithm, out of all actual fraudulent transactions. This metric is essential in fraud detection since low values of precision mean many false alarms, which is disruptive to legitimate users. However, recall gauges the percentage of real fraudulent cases accurately detected by the model and its capacity to respond to fraudulent activities. The desire to minimise loss is why high Recall values are desirable for the maximum coverage of fraud cases.

Using precision and recall, the F1-score was utilised to offer a single measure that balances false positive and false negative results. Its usefulness becomes evident in a scenario where it is essential to maximise recall while minimising the number of false positives, making this metric a good measure of the global performance of a model that performs fraud detection.

ROC-AUC

The study also measured the performance of each model using the overall Area Under the curve of the Receiver Operating Characteristic (ROC-AUC), which determines the ability of the model to give different thresholds of two classes. A high AUC-ROC denotes strong discrimination capability, which is of enormous importance in fraud detection since it defines the ability of the model to distinguish between the fraudulent and genuine, fully and partially.

Furthermore, calibration analysis was done to assess the accuracy of the probabilities yielded by each model. Observing the distribution of actual outcomes to predicted probabilities or odds, calibration curves were drawn, and the ideal model has a line with a regression of 1. In fraud detection, the probability estimates are deployed in decision-making, and therefore model calibration is crucial to prevent over-reliance on these estimates. The calibration analysis added more decision-making criteria on which models gave reliable probabilities in applying models for practical use.

Results

Exploratory Data Analysis (EDA)

Class Distribution (Fraud vs non-fraud)

The class distribution plot illustrates a severe imbalance in the dataset, with fraudulent transactions accounting for only 0.173% of the total transactions. This imbalance underscores the

challenge of detecting fraud, as models tend to be biased toward the majority class. Techniques like SMOTE were employed to counter this, allowing the models to learn effectively from the minority (fraud) class without being overwhelmed by legitimate transactions (as shown in **Figure 2**).

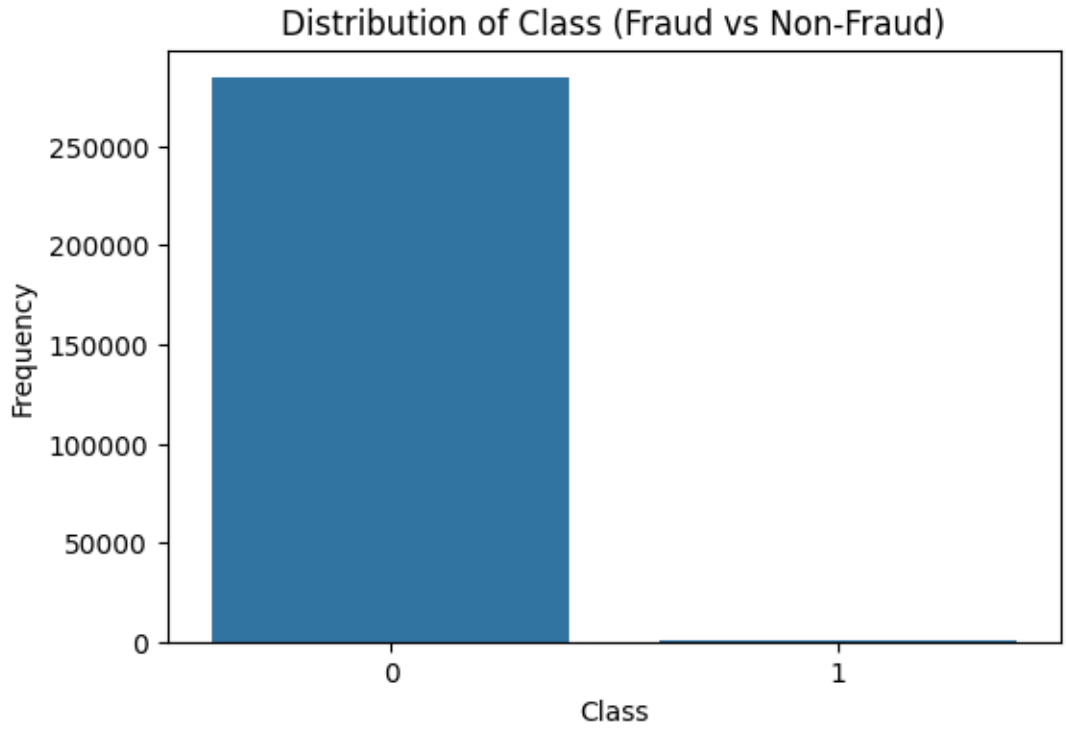

Figure 2: Class Distribution (Fraud vs non-fraud)

And to overcome the huge difference between classes in the old data, this study applied SMOTE, resulting in a balanced dataset made up of 284,314 samples from each class. Because of this process, model training includes a similar amount of fraudulent transactions and legitimate ones. From **Figure 3** it can be seen that the dataset has been properly balanced. This step is necessary to ensure the model doesn't favor the majority class and also helps it perform better in catching fraud.

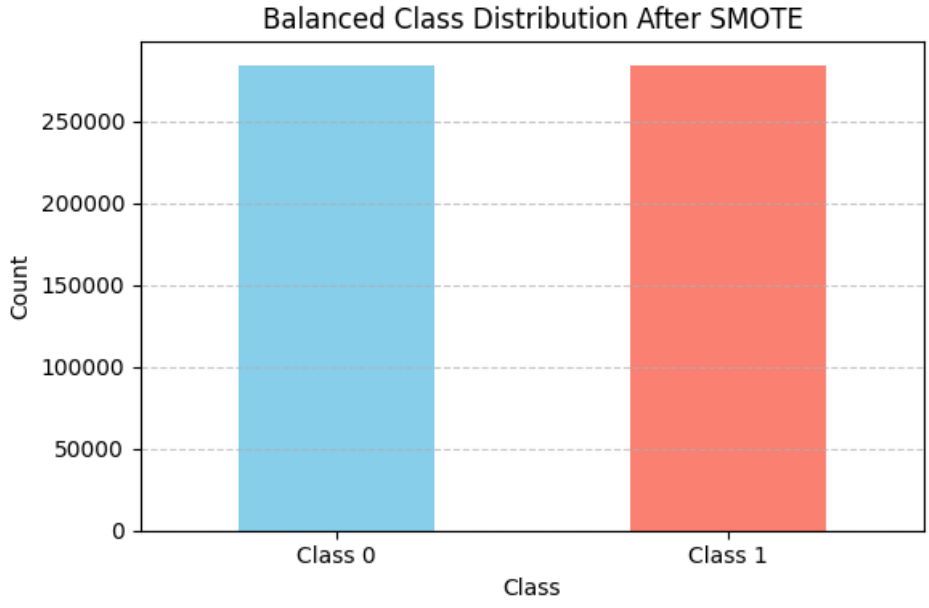

Figure 3. Class distribution after SMOTE

Transaction Amount Distribution

The transaction amount distribution shows a highly skewed pattern, with the majority of transactions involving smaller amounts and a few outliers with very high values (as shown in **Figure 4**). This skew can impact model training by emphasising low-amount transactions, potentially overlooking fraud patterns in high-value transactions. To address this, all features were standardised to ensure that the Amount feature did not dominate model predictions.

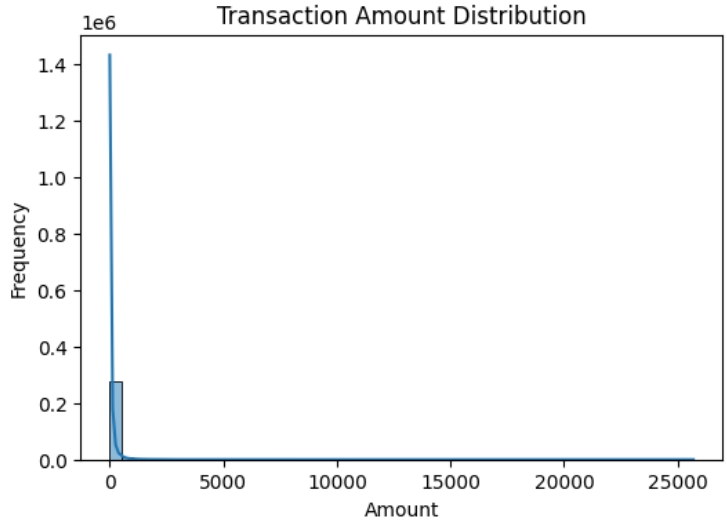

Figure 4: Transaction Amount Distribution

Transaction Amount Over Time

The scatter plot of transaction amounts over time shows no discernible temporal pattern in fraud occurrences, indicating that fraudulent transactions are spread randomly across the dataset (as shown in **Figure 5**). This lack of time-based clustering suggests that fraud can occur at any time, making it crucial for the models to detect fraud without relying on temporal dependencies. This insight reinforces the importance of feature-based, rather than time-dependent, detection methods.

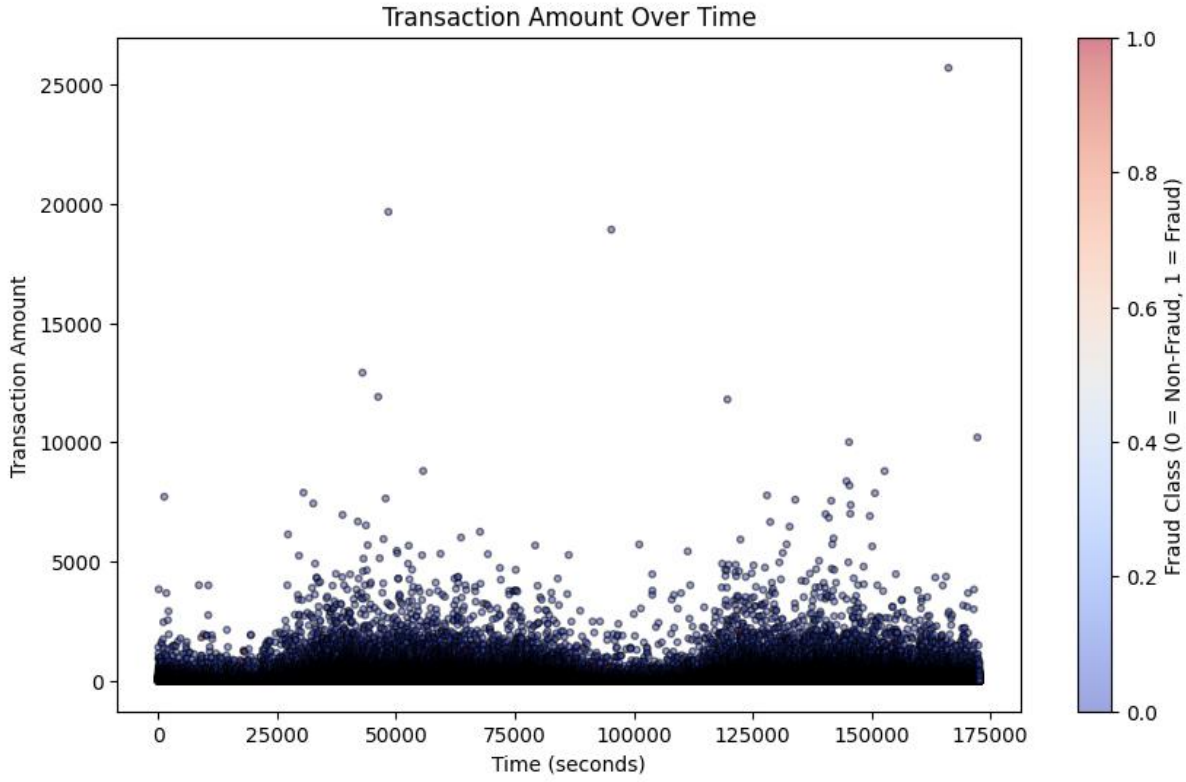

Figure 5: Transaction Amount Over Time

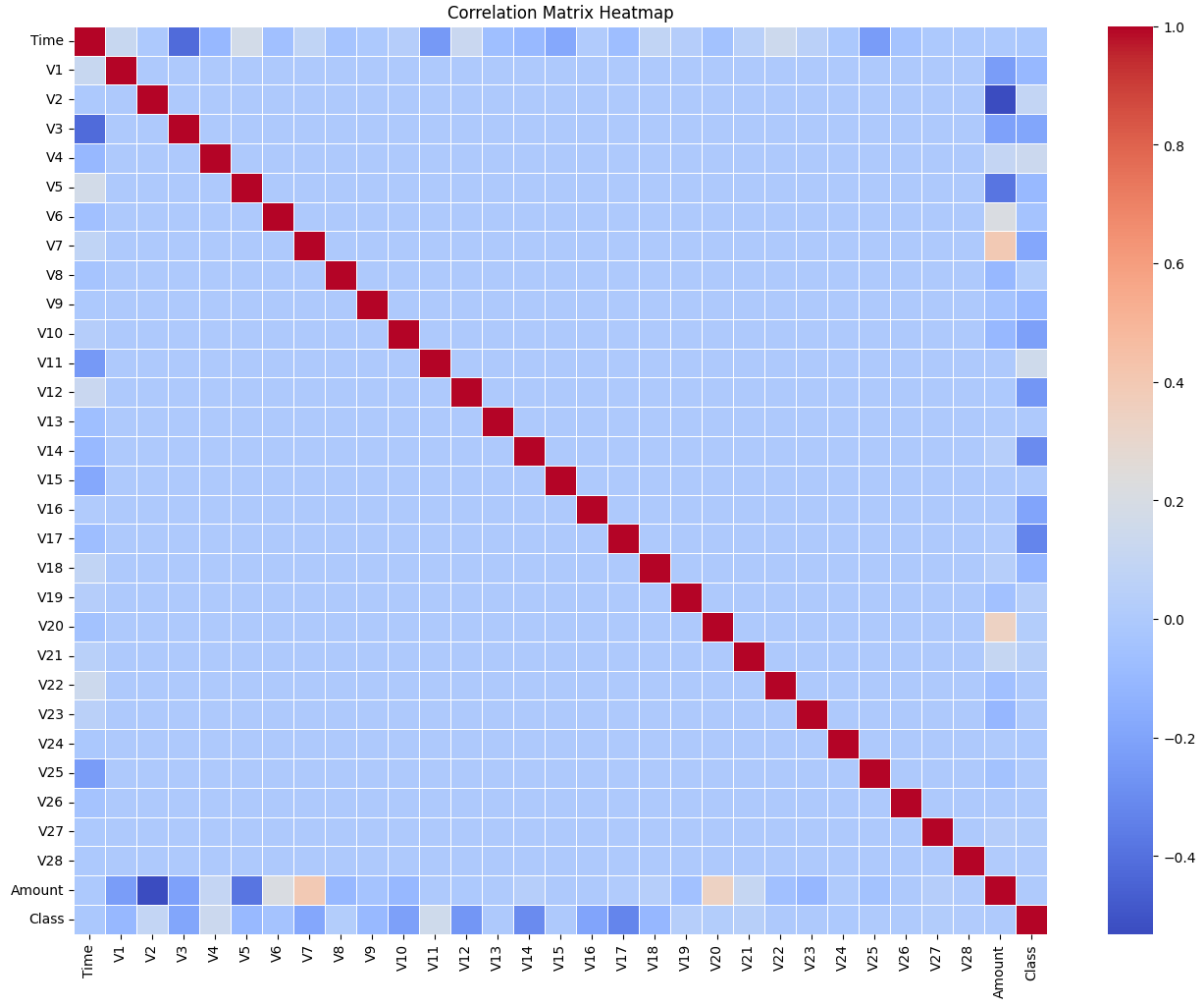

Figure 6: Correlation Heatmap

The correlation matrix heatmap displays the pairwise correlations among the features in the dataset, with values ranging from -1 (strong negative correlation) to 1 (strong positive correlation). The heatmap primarily shows low correlations across most feature pairs, reflecting the effectiveness of Principal Component Analysis (PCA) in reducing multicollinearity and ensuring features are mainly independent. Notably, the features V2, V4, V11, and V19 exhibit slight positive or negative correlations with the Class label, indicating some predictive potential for fraud detection (as shown in **Figure 6**). The Amount feature also shows mild correlations with certain variables, hinting at its relevance in fraud detection. These insights suggest that only a few features carry inherent associations with the target variable, while the rest are minimally correlated, which can improve model performance by reducing noise and focusing on the most relevant attributes. This sparse correlation structure aligns well with a machine-learning approach for fraud detection.

Pair Plot of Highly Correlated Features

The pair plot of highly correlated features, including V10, V12, V14, and V17, reveals distinct fraudulent and non-fraudulent transaction clusters. These clusters indicate that these feature pairs are effective predictors of fraud. Such patterns can be particularly useful in model training, as they allow the models to distinguish between fraud and legitimate transactions based on these correlated feature combinations, improving detection accuracy (as shown in **Figure 7**).

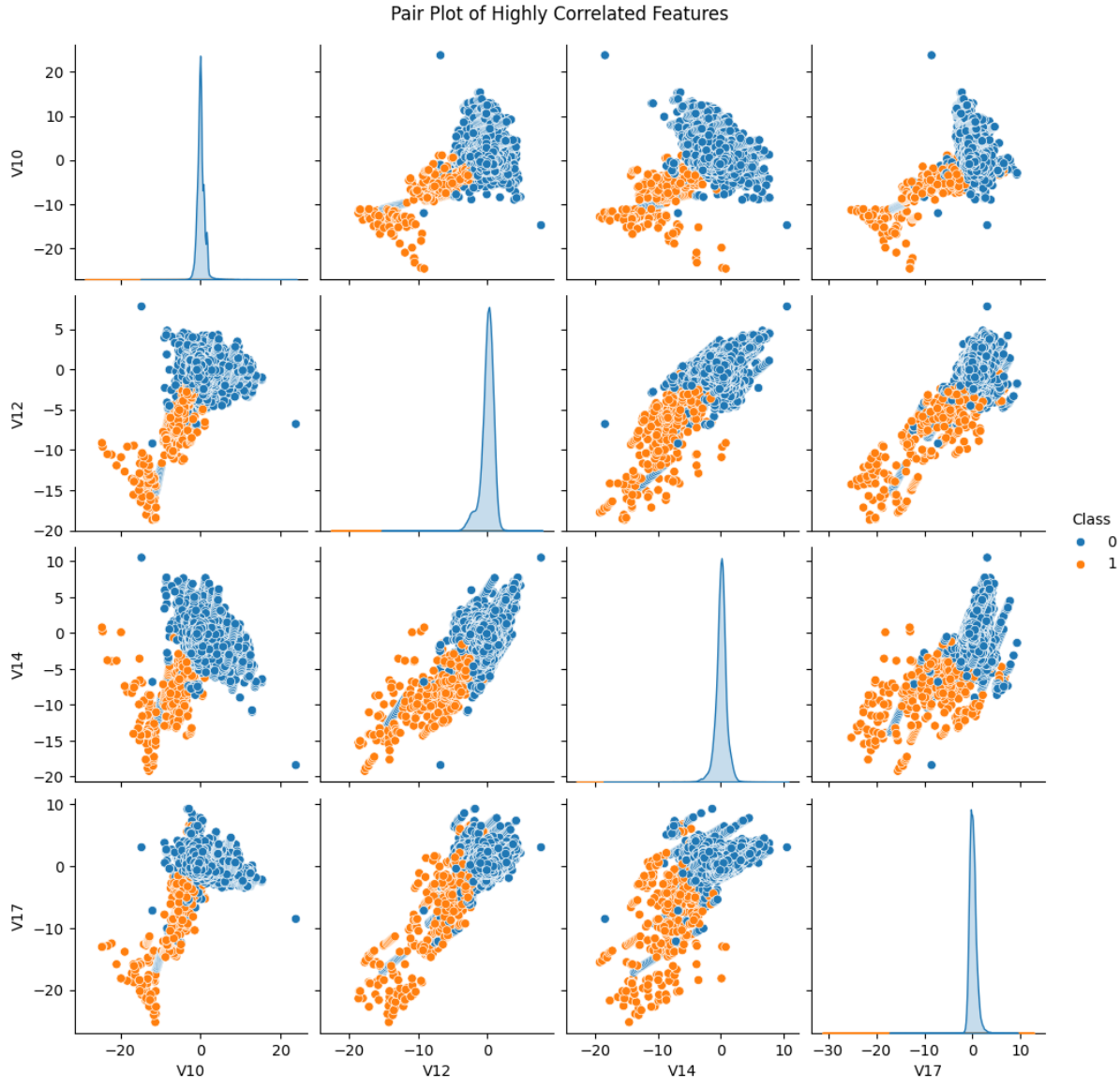

Figure 7: Pair Plot of Highly Correlated Features

Logistic Regression

Logistic Regression struggled to balance false positives and true positives, with a high rate of false positives and limited detection of true fraud cases. This result reflects the model's limitations in effectively differentiating between classes (as shown in **Figure 8**).

```
Evaluating LogisticRegression on Test Data:

Confusion Matrix:
[[55406  1458]
 [    8    90]]
```

Figure 8: Confusion Matrix Logistic Regression

The precision for the fraud class was only 0.06, while recall was relatively high at 0.92. This indicates that, although the model captured most fraud cases, its low precision led to numerous false positives, reducing its practical effectiveness (as shown in **Figure 9**).

```
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.97      0.99     56864
           1       0.06      0.92      0.11        98

    accuracy                           0.97     56962
   macro avg       0.53      0.95      0.55     56962
weighted avg       1.00      0.97      0.99     56962
```

Figure 9: Classification Report

Based on the results shown in the precision-recall curve, precision falls drastically as it increases recall, which implies that Logistic Regression is not accurate when used with a high recall value. This trade-off makes the model less effective when specific precision is necessary for the model's outcome (as shown in **Figure 10**).

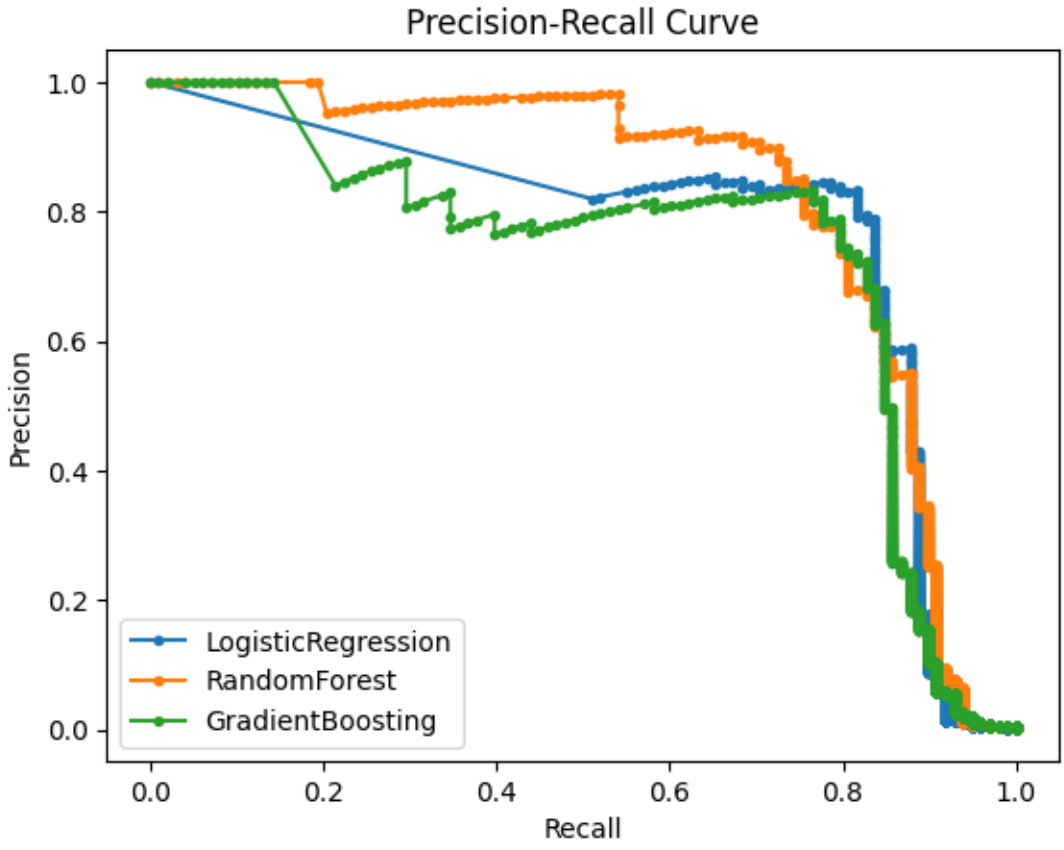

Figure 10: Precision-Recall Curve

Logistic Regression has a moderate to good ability to sort patients according to the risk of developing MACE based on the AUC score of 0.97. However, this performance is quite low compared to other models as it is linear and cannot capture the various complexities in the fraud pattern (as shown in **Figure 11**).

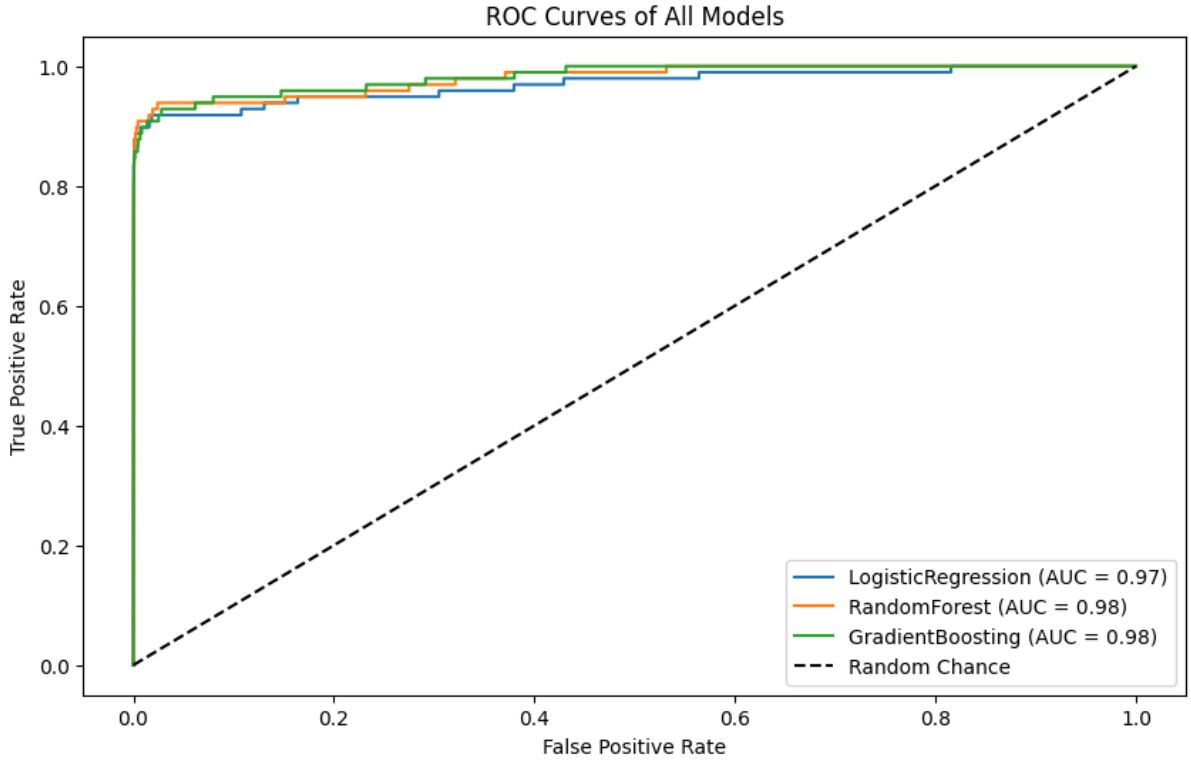

Figure 1: ROC Curve

Random Forest

When analysing the results, Random Forest showed a reasonable ratio between true positives and false negatives, securing 86 of the 98 fraud cases. This result shows that the proposed method can differentiate between fraudulent transaction (as shown in **Figure 12**).

```
Confusion Matrix:
[[56754   110]
 [   12    86]]
```

Figure 12: Confusion Matrix Random Forest

Random Forest obtained a precision of 0.44 and a recall of 0.88, which suggests a low false positive rate and a high fraud detection rate, respectively. This balance is essential for practical research where both measures are helpful (as shown in **Figure 13**).

```
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     56864
           1       0.44      0.88      0.59        98

    accuracy                           1.00     56962
   macro avg       0.72      0.94      0.79     56962
weighted avg       1.00      1.00      1.00     56962

ROC AUC Score: 0.9799910168766438
```

Figure 3: Classification Report

The model shows high accuracy and recall values, demonstrating how it balances finding more fraudulent cases and reducing false positives. This characteristic is valuable in fraud detection as false positives are expensive (as shown in **Figure 14**).

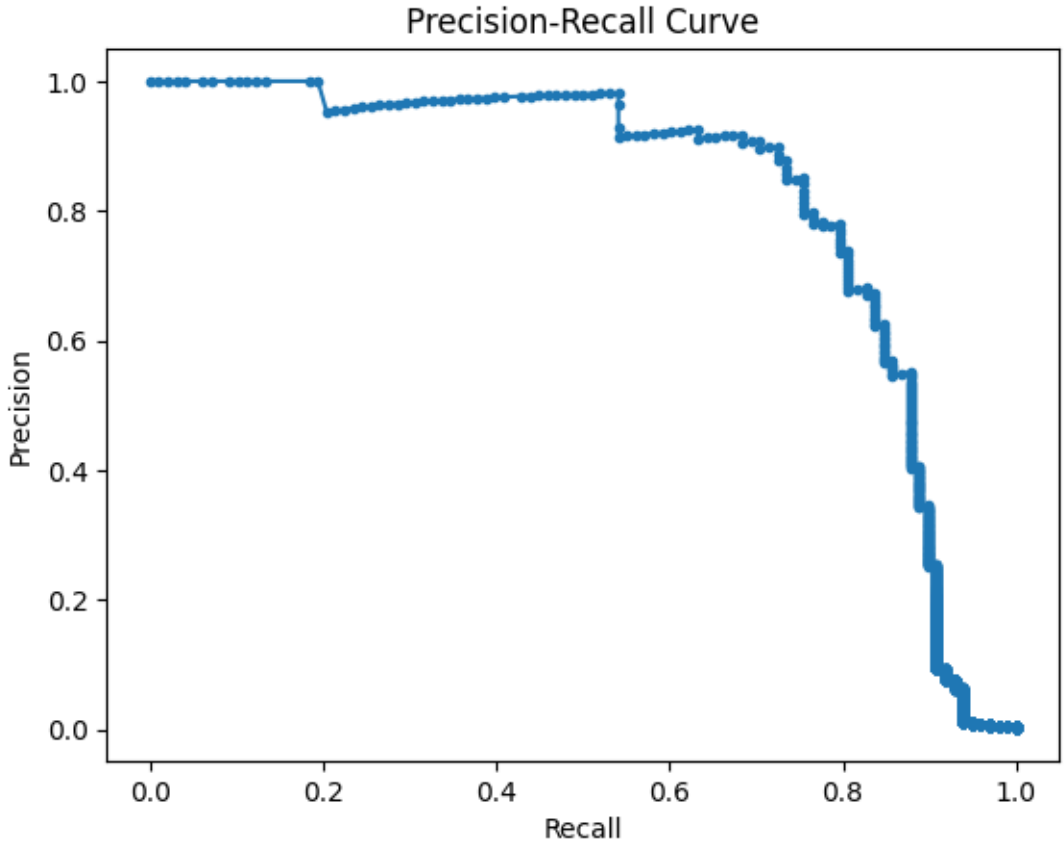

Figure 14: Precision Recall Curve

Random Forest achieved an AUC score of 0.98, which shows more efficiency in discriminating fraudulent and non-fraudulent transactional data. This level of performance further validates its use in real-world fraud identifying scenarios (as shown in **Figure 15**).

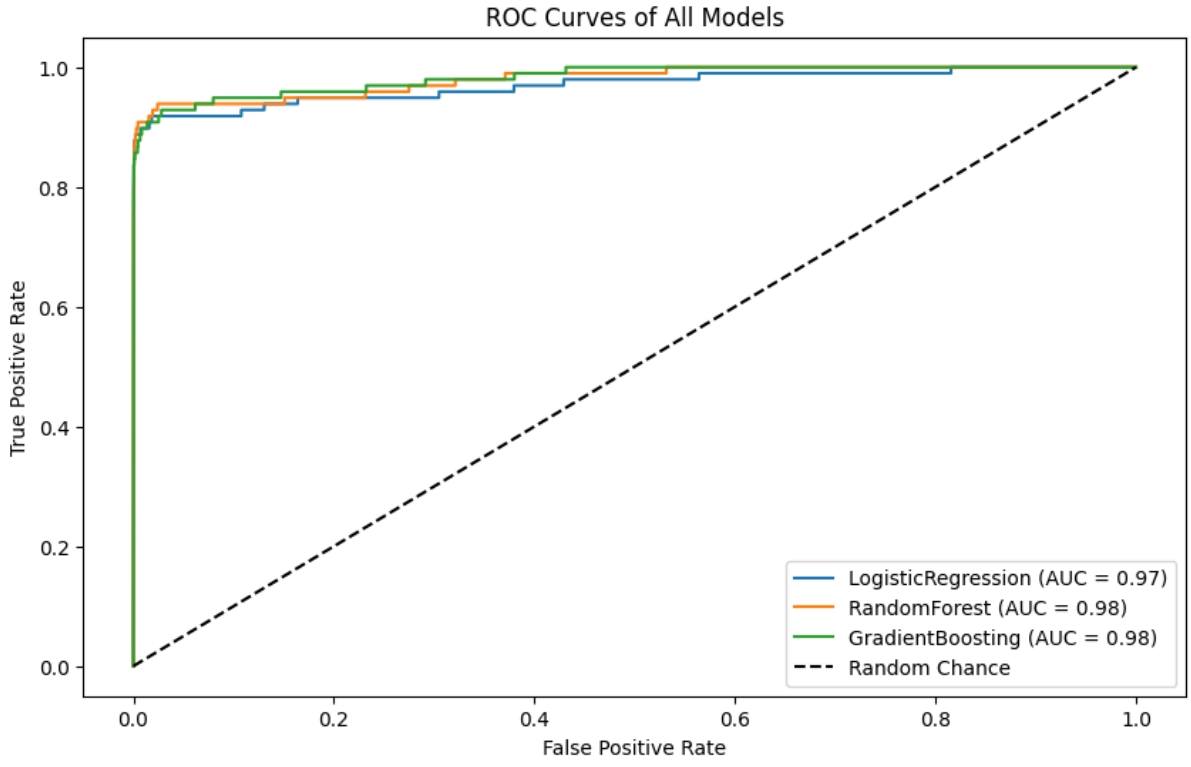

Figure 15: ROC Curve

The feature importance plot shows that the most significant feature for fraud prediction is V14, followed by V10 V4. This is helpful in interpretability and implies that these features encompass important variables of fraudulent actions that constitute the basis for targeted fraud prevention (as shown in **Figure 16**).

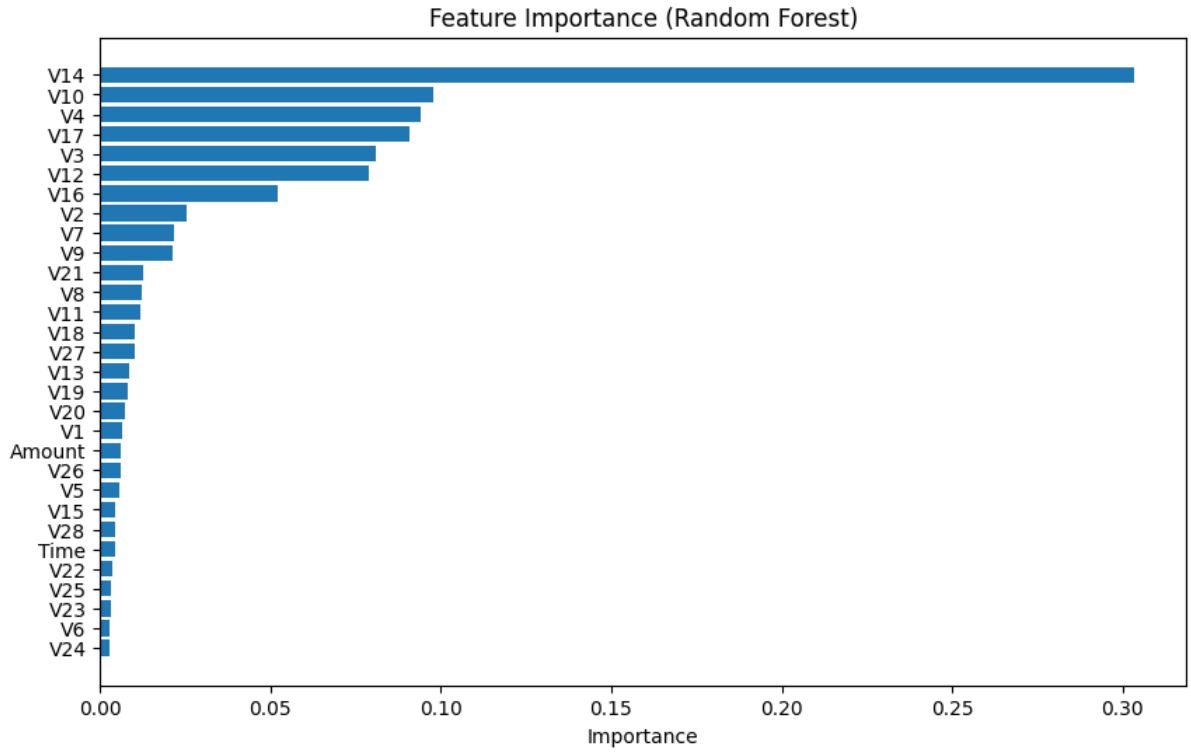

Figure 16: Feature Importance (Random Forest)

Gradient Boosting

Gradient Boosting was slightly better than Logistic Regression in achieving a moderate, balanced identification of fraud cases. However, it did not outperform Random Forest in true positive detection and false positive reduction (**Figure 17**).

```
Evaluating GradientBoosting on Test Data:

Confusion Matrix:
[[56144   720]
 [   10    88]]
```

Figure 17: Confusion Matrix Gradient Boosting

The model reached a recall of 0.90, but only 0.11 precision; that means the model can identify most of the fraud cases but was alarmingly wrong. This performance illustrates the price Gradient Boosting pays for its sensitivity-oriented focus (**Figure 18**).

```
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.99      0.99     56864
           1       0.11      0.90      0.19        98

    accuracy                           0.99     56962
   macro avg       0.55      0.94      0.59     56962
weighted avg       1.00      0.99      0.99     56962

ROC AUC Score: 0.9823
```

Figure 18: Classification Report

Gradient boosting has the AUC score of 0.98 which show the similar discriminative performance to the random forest algorithm. However, the model's lower precision reduces the practicality in scenarios requiring the utmost precision to reduce operational costs **(Figure 19**).

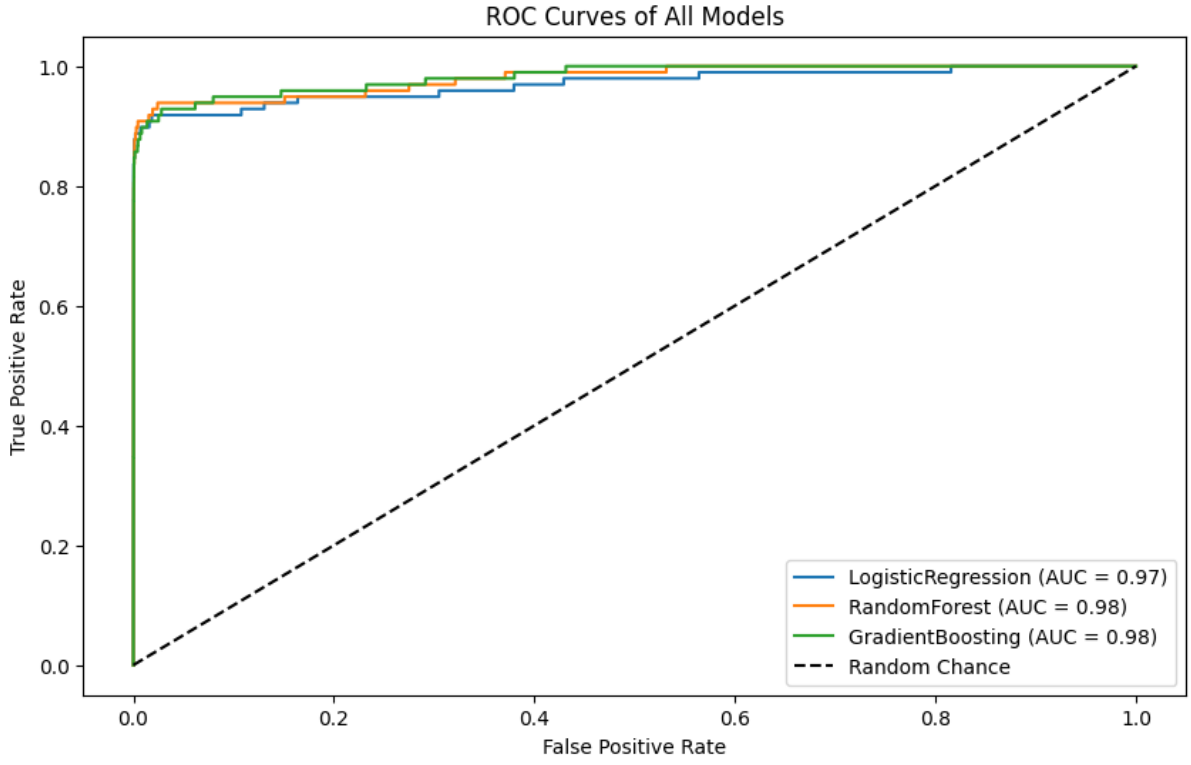

Figure 19: ROC Curve

Confusion Matrix

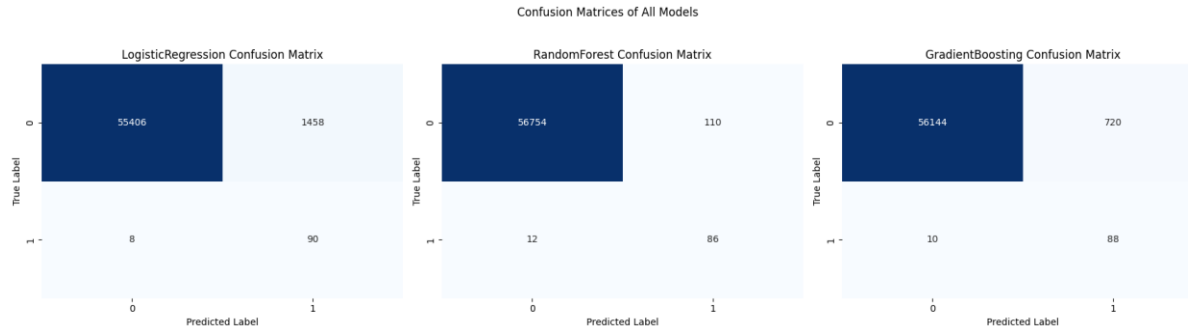

Figure 20: Confusion Matrice of Models

The confusion matrices for Logistic Regression, Random Forest, and Gradient Boosting show different trends for fraud detection. Since Logistic Regression is susceptible to false positives, the algorithm misidentifies 1,458 legitimate transactions as fraud but learns 90 true fraud cases (as shown in **Figure 20**). The Random Forest model records slightly higher Co asteroids at 110 and now false negatives at 12, making it the most accurate of the three models. False negatives with Gradient Boosting reduce to 10 missed fraud cases and false positives are higher. Gradient Boosting is 720, which shows that Gradient Boosting is sensitive with a higher false positive than Random Forest.

Calibration Plot for All Models

The calibration plot displays the relative frequency of instances for each model's given probability estimates (as shown in **Figure 21**). Analysing the probability confirmation, Random Forest shows the most balanced position concerning the ideal calibration line. This reliability makes a Random Forest very suitable for applications that require decision-makers to make decisions based on high fraud probability rates. Logistic Regression and Gradient Boosting are exposed to deviations that highlight too-confident probability estimations, which leads to misrepresentations of actual real-world fraud cases in the reality of fraud detection systems.

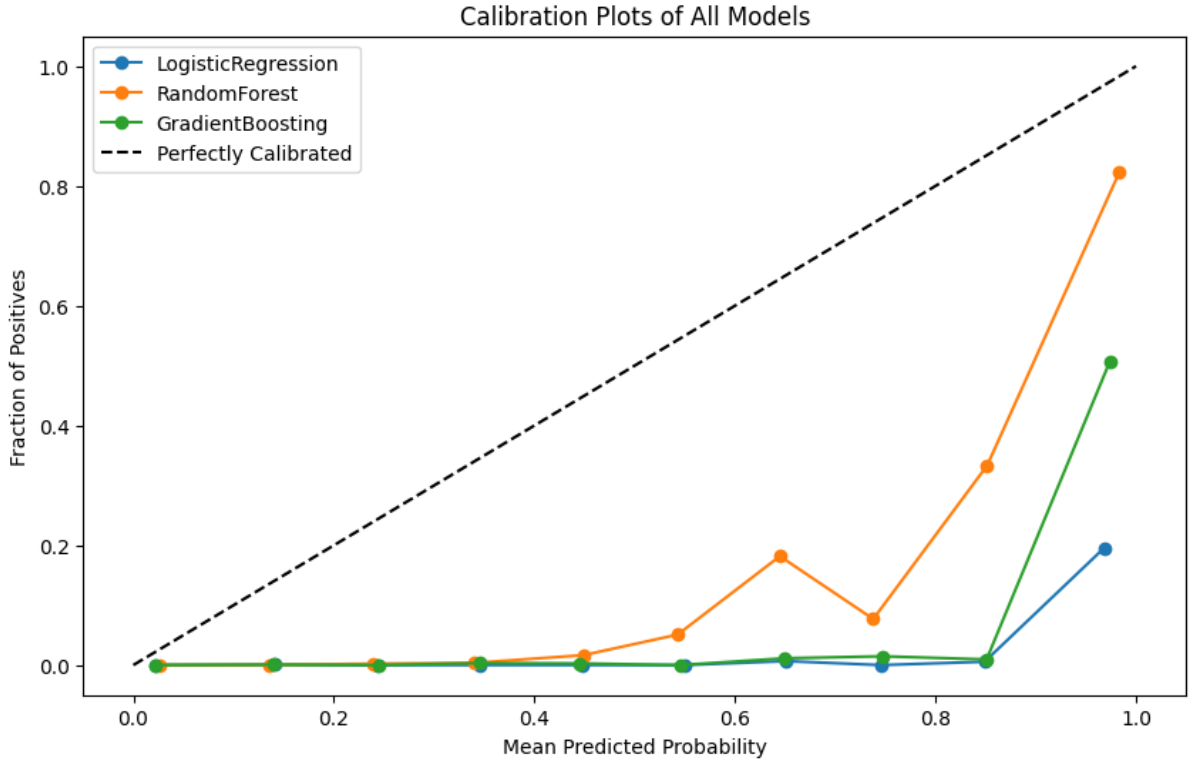

Figure 21: Calibration Plot for All Models

Discussions

In this research, the predictive models, including Logistic Regression, Random Forest, and Gradient Boosting, were evaluated for detecting fraudulent transactions in a large imbalanced financial data. Every model has advantages and disadvantages; however, for fraud detection in financially responsible institutions, the Random Forest algorithm has the most suitable hit rate recall, reliable and accurate probability estimates, and model interpretability.

The Logistic Regression model was used as a benchmark to illustrate the difficulties in applying linear models to fraud identification tasks. Although the model had high accuracy in identifying most fraud cases, with a recall score of 0.92, it was less precise, yielding a score of 0.06, causing many false positives. This high rate of false alerts shows that the Logistic Regression model must have the necessary complexity to differentiate between fraudulent and legitimate transactions adequately. Moreover, the calibration was also very low, providing inconsistent probability estimates crucial in practical applications where risk levels must be determined accurately.

Random Forest had a better accuracy of 0.44 for precision and 0.88 for recall, providing a better approach to reducing both accurate positive and true negative results. This balance is important in fraud detection because most fraud cases have to be detected while at the same time,

fallout or legitimate consumers have to be kept minimal. The AUC-ROC of the proposed model was 0.98, confirming its high ability to distinguish between classes with high specificity, while the proximity of the calibration plot to the ideal line shows that the model provides reliable probability estimates. These characteristics make the Random Forest especially applicable for practical implementation in financial institutions where predictive accuracy and model stability are critical. Also, Random Forest suggested V14, V10 and V4 as significant features in the fraud signs, which helped improve the model interpretability and aimed institutions at important fraud patterns.

Regarding the outcome metrics, Gradient Boosting had a high level of accuracy, as indicated by the AUC-ROC index, equal to 0.98, and the recall of the model equal to 0.90, which implies that this model is sensitive enough to explore fraud cases. Nevertheless, it had a low precision of 0.11 resulting in falser positive and hence could have been more helpful, especially when false alarms need to be avoided. Additionally, the calibration analysis showed that Gradient Boosting produced predictions in terms of probability estimates that were considerably less accurate than those of Random Forest, thus potentially limiting its usefulness in contexts that require accurate fraud risk measurement systems. This sequential learning arrangement makes Gradient Boosting a machine learning model capable of capturing intricate and nested patterns, and, from this perspective, it can be argued that this is where its sensitivity lies; however, it is observable that with this sensitivity comes a level of imprecision, which may put it at a disadvantage when used in fraud identification applications requiring higher levels of accuracy.

Random Forest, therefore, emerges as the least-gain most loss model proposed in the current study in the sense that it presents a balanced solution to the challenge of fraud identification coupled with probability estimation and most favoured unique features. The model's high precision, low recall, and good calibration ratios indicate it can protect organisations from both these losses and costs, making it a perfect solution for financial institutions searching for an efficient fraud detection solution.

Some limitations include using SMOTE in handling imbalanced data, a potential source of generalisation bias. Future studies can focus on different resampling approaches and look into semi-supervised or deep learning to learn new fraud structures. Future works can employ a broader range of models and utilise more data to extend the results and performance of fraud detection systems.

Conclusions

This research investigated how three algorithms namely Logistic Regression, Random Forest, and Gradient Boosting can perform in identifying fraudulent transactions from a large-scale imbalance of financial data. The evaluation was based on accuracy, precision, recall, F1-score, ROC-AUC, and calibration analysis, and the developed Random Forest model was found to be the most suitable for fraud detection in real-world practical contexts.

These problems led to the exclusion of logistic regression since it was simple to interpret, but its low precision and high FPR made it less suitable for application. While the algorithm provided a high recall rate, which means it identifies the vast majority of the fraudulent

transactions, the high false positive rate means it would not be efficient if it is used alone in managing fraud in financial operations. The comparatively low precision and calibration of the Gradient Boosting mean that while it was good at recall and sensitivity on complex patterns in the data, its utility was reduced by the fact that many of the results were ultimately false positives. This signal noise trade-off suggests that Gradient Boosting may need further tweaking to be helpful in high-consequence financial situations.

Out of all the models compared, it was identified that the Random Forest model yielded the best results, and met the desired level of fraud detection while maintaining low false positive results. Thus, the Random Forest is fast, accurate with AUC-ROC equals 0.98, has high precision and recall, and possesses good reliability of probability estimations, which makes it possible to use this model for practical application in fraud detection. The importance of features was computed in the next step to get more information about features that should be targeted to improve the detection rate. The glories of Random Forest make it favourable to financial institutions that need a stable and accurate interpretable fraud-detecting model.

This paper therefore connects the growing literature on the application of machine learning in fraud detection. The study provides valuable information on each model's strategic advantages and disadvantages to help institutions seeking to enhance transaction security. However, the remaining limitations, for example, synthetic data resampling (SMOTE), and the call for increased data set size and variance point to areas of future research. Future research should consider using other forms of resampling that might improve the effectiveness of fraud detection, such as semisupervised learning and experimenting with deep learning architectures.

Thus, Random Forest being a well-calibrated model with a satisfactory precision and recall score can be considered a valuable tool for fraud detection in financial institutions to minimise the financial and organisational impact of fraud as well as fraudulent transactions that are flagged accidently. New developments in model techniques, particularly in adaptive learning could add further readiness against fraud and other financial crimes as new patterns in transactions and fraud methodologies are identified.

References

Al-dahasi, E.M., Alsheikh, R.K., Khan, F.A. and Jeon, G., 2024. Optimising fraud detection in financial transactions with machine learning and imbalance mitigation. Expert Systems, p.e13682.

Bello, H.O., Ige, A.B. and Ameyaw, M.N., 2024. Adaptive machine learning models: concepts for real-time financial fraud prevention in dynamic environments. World Journal of Advanced Engineering Technology and Sciences, 12(02), pp.021-034.

Bello, O.A., Folorunso, A., Ejiofor, O.E., Budale, F.Z., Adebayo, K. and Babatunde, O.A., 2023. Machine Learning Approaches for Enhancing Fraud Prevention in Financial Transactions. International Journal of Management Technology, 10(1), pp.85-108.

Bodria, F., Giannotti, F., Guidotti, R., Naretto, F., Pedreschi, D. and Rinzivillo, S., 2023. Benchmarking and survey of explanation methods for black box models. Data Mining and Knowledge Discovery, 37(5), pp.1719-1778.

Endres, M., Mannarapotta Venugopal, A. and Tran, T.S., 2022, August. Synthetic data generation: A comparative study. In Proceedings of the 26th International Database Engineered Applications Symposium (pp. 94-102).

Ghaleb, F.A., Saeed, F., Al-Sarem, M., Qasem, S.N. and Al-Hadhrami, T., 2023. Ensemble Synthesised Minority Oversampling based Generative Adversarial Networks and Random Forest Algorithm for Credit Card Fraud Detection. IEEE Access.

Hilal, W., Gadsden, S.A. and Yawney, J., 2022. Financial fraud: a review of anomaly detection techniques and recent advances. Expert systems With applications, 193, p.116429.

Mohan, P.V., Dixit, S., Gyaneshwar, A., Chadha, U., Srinivasan, K. and Seo, J.T., 2022. Leveraging computational intelligence techniques for defensive deception: a review, recent advances, open problems and future directions. Sensors, 22(6), p.2194.

Nesvijevskaia, A., Ouillade, S., Guilmin, P. and Zucker, J.D., 2021. The accuracy versus interpretability trade-off in fraud detection model. Data & Policy, 3, p.e12.

Pazarbasioglu, C., Mora, A.G., Uttamchandani, M., Natarajan, H., Feyen, E. and Saal, M., 2020. Digital financial services. World Bank, 54, pp.1-54.

Rodríguez Vaquero, P., 2023. Literature Review of Credit Card Fraud Detection with Machine Learning.

Thétard, H.M., 2021. Automated payment fraud detection using logistic regression and support vector machines (Doctoral dissertation, Stellenbosch: Stellenbosch University).

Vanini, P., Rossi, S., Zvizdic, E. and Domenig, T., 2023. Online payment fraud: from anomaly detection to risk management. Financial Innovation, 9(1), p.66.

Vichare, S.S., 2024. PROBABILISTIC ENSEMBLE MACHINE LEARNING APPROACHES FOR UNSTRUCTURED TEXTUAL DATA CLASSIFICATION (Doctoral dissertation, Purdue University Graduate School).

Zhao, Y., Zhu, W., Wei, P., Fang, P., Zhang, X., Yan, N., Liu, W., Zhao, H. and Wu, Q., 2022. Classification of Zambian grasslands using random forest feature importance selection during the optimal phenological period. Ecological indicators, 135, p.108529.