International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
SJIF (2025): 8.036

Online Transaction Fraud Detection Using Machine Learning

Aditya Sunil Raje[1], Dr. Ayesha Siddiqui[2]

[1]MCA (Computer Science), JSPM University, Wagholi, Pune, India
Email: *adityasraje18[at]gmail.com*

[2]Associate Professor, Department of Computer Science, JSPM University, Wagholi, Pune, India
Email: *ais.scos[at]jspmuni.ac.in*

Abstract: ***Increasingly widespread use of online payment systems, which include e-commerce websites, Internet banking, mobile transitions, has resulted in a growing number of fraudulent cases and posed great security risks not only to the users but also to financial companies. Traditional methods used for detection of fraudulent activities based on rule set are often insufficient due to their rigidness and no adaptability. In order to solve this problem, this paper offers a machine learning solution to detect fraudulent transactions in real time. Specifically, a number of supervised learning models, such as Logistic Regression, Decision Tree, and Random Forest will be utilized for analysis of transactional behaviour and differentiation between regular and fraudulent transactions. The dataset for this analysis is based on historical transaction data and includes stages of preprocessing, such as cleaning, normalizing and feature selection. The efficiency of the system will be evaluated through a series of characteristics, including Accuracy, Precision, Recall and F1-score. The results indicate the advantage of Ensemble learning models, especially Random Forest method, over traditional rule-based systems in terms of accuracy and ability to handle imbalanced data. Moreover, the model proposed in this paper is able to automatically adapt to newly received transactional data.***

Keywords: Online Transaction, Online Transactions, Anomaly Detection, Supervised Learning, Real time Detection, Data Preprocessing, Classification, Financial Security

1. Introduction

The integration of digital technology in various industries, such as e-commerce, internet banking, and mobile payments, has brought about significant changes in the manner of conducting financial transactions. Although it has brought about efficiency and rapid transaction processing, some problems have occurred regarding transaction security. Among these are the rising levels of fraud in digital transactions. Criminals continually come up with innovative methods of taking advantage of digital weaknesses. It is for this reason that fraud detection plays an essential role in providing security and efficiency in financial transactions. [1] Fraud detection mechanisms in contemporary financial institutions are important in securing digital transactions through detecting any anomalies in financial transactions. They conduct a transaction analysis, assess consumer purchase behaviour and detect any anomalies in transactions. These systems have several functions in relation to identifying abnormalities within the transaction amount, anomaly in the user's geographical location, transaction frequency, and security compliance. [2] Fraud detection in its early stages was achieved using mechanisms involving rule-based systems and manual verification. The method relied heavily on pre-set rules such as setting of transaction limits and listing of accounts engaged in fraud. Such methods were efficient in fraud detection but lacked the capacity to detect new emerging trends since it is rule-based. With time, conventional methods prove ineffective and unable to deal with sophisticated frauds and attacks. In such a case, fraud detection usually relies on intelligence. [3] With technological advancements, especially machine learning, significant improvements have been made in the area of fraud detection. Training machine-learning algorithms with huge transaction data has led to improved trend detection. Some common classification algorithms used in distinguishing between legitimate and fraudulent transactions include Logistic Regression, Decision Tree, and Random Forest. [4] Machine learning algorithms have the capacity to learn continuously with changes in transaction data. [5] Nonetheless, the use of machine learning techniques in fraud detection faces various challenges. These include class imbalance, false alarms and rapid processing in real time. Further, there is currently no framework that provides guidance on which machine learning technique works effectively.

In order to address the above-mentioned challenges, this research proposes to develop a system that can help detect frauds in online payments by using machine learning algorithms. The usage of such a system will demand the use of several supervised learning techniques, the performances of which will be compared based on accuracy, precision, recall, and F1-score measures. Preprocessing techniques, including data cleaning, normalization, and feature selection, will be conducted on the selected dataset in this research. [6]

2. Background

With the advent of technology and the popularity of internet banking, e-commerce websites and online payment gateways, online transactions have become very common nowadays. The reason behind choosing this method of conducting transactions is its convenience and fast processing. [7] Innovations like mobile wallets and unified payments interface have taken the process of financial transactions to the next level, where individuals can do any kind of transactions without considering factors like geographical location or the time zone they belong to. But along with all these developments, an increasing number of cybercriminals are getting attracted to such systems, whereby they indulge in malicious activities using different methods, such as phishing, identity theft and hacking accounts. [8]Traditional methods

that include manual verification of predefined criteria for fraud detection cannot prove helpful at all in the current scenario. Firstly, this process requires analysing a huge amount of data, which can't be done manually. Secondly, it is very difficult for an individual to carry out such fraud detections. Thus, machine learning makes the entire process quite effective and swift. [9]

3. Problem Statement

The fast-growing trend in conducting online transactions has made it vulnerable to many frauds, thus creating various problems for both users and financial organizations. The existing fraud detection systems, which are mostly rule-based and require constant human involvement in detecting and identifying frauds, are unable to keep up with modern needs and demands since such systems cannot process huge amounts of data in real time. Also, such systems are not able to deal with more sophisticated forms of frauds. In addition to that, problems, including an imbalanced dataset and the presence of a relatively high number of false alarms, arise in these situations. This means that there is a dire necessity to develop a system, which would allow detecting frauds automatically through the use of machine learning methods. [10]

4. Research Objectives

The main aim of this project is to build and develop a system to detect fraudulent online transactions through Machine Learning methods. This research aims to enhance the effectiveness and efficiency of fraud detection techniques in online transaction processes. The main objectives of the study are:

To examine various kinds of online transaction frauds and learn their patterns.

a) To investigate the application of machine learning techniques in fraud detection processes.
b) To implement and compare several supervised machine learning algorithms such as logistic regression, decision tree, and random forest techniques.
c) To clean, normalize, and select features of transactional data to develop the best models.
d) To analyze and evaluate machine learning models by measuring

5. Literature Review

5.1 Types of Online Fraud

In the context of online banking, there are many types of frauds that occur continuously, thanks to technological advances. The types of fraud that take place are phishing, where individuals are fooled into providing their private information; identity fraud, whereby the personal information of someone is used without his or her consent; and fraudulent transactions that happen through hacked banking accounts. [11] However, there are also other types of fraud like card-not-present fraud, account takeover fraud, and transaction laundering. [12]

5.2 Detection Methods and Technologies

A number of strategies have been implemented in order to avoid any form of fraud during transaction processing. The conventional technique makes use of rule-based systems, whereby the identification of any fraudulent behavior is achieved through pre-defined criteria. Modern techniques are now used in order to analyze the behavior of transactions and identify patterns through machine learning algorithms. Logistic regression, decision trees, support vector machines, and random forests are some of the strategies that have been adopted for transaction classification. [13]

5.3 Challenges and Limitations

Even with developments, there are several challenges that remain for fraud detection systems. For one, the imbalanced nature of the dataset is a problem because of the very few cases of fraud in comparison to the overall volume of data available. There is the challenge of the system being highly prone to errors since it will often result in too many false positives. Fraud trends change very frequently, thus requiring quick adaptation to keep up.

5.4 Regulatory Landscape

Financial institutions need to follow different regulatory frameworks for secure and efficient operations. Data security requirements and verification of identities are two major regulatory demands. Know Your Customer (KYC) and Antimony Laundering (AML) requirements help in reducing instances of fraud. It is important for financial organizations to adhere to these regulatory frameworks so that they can avoid fraud as well as be legally accountable.

5.5 Case Studies and Best Practices

There are many financial companies and payments platforms that have successfully used ML algorithms for fraud detection. The implementation involves analysing past data, monitoring the transactions in real time, and adopting adaptive learning techniques. Good practices for fraud detection through ML algorithms include using ensemble learning techniques, keeping the models updated and incorporating a number of layers of security like two-factor authentication.

5.6 Emerging Trends

The area of fraud detection is an ever-changing one due to ongoing technological innovations. Some of the emerging trends include the application of artificial intelligence in predictive analysis, the implementation of big data technologies for processing huge amounts of data, and the introduction of real-time fraud detection systems. The incorporation of blockchain technology is another trend that is being considered for boosting the security and transparency of transactions. Furthermore, the use of deep learning and hybrid models is anticipated to enhance the efficacy of fraud detection.

5. Research Gap

Moreover, certain studies only emphasize increasing accuracy without paying sufficient attention to such essential parameters as precision, recall, or false positives. As disturbances will inevitably lead to the rise in the number of false positives, models need to be created not only taking into account high efficiency of detection but minimizing errors in classifications as well. Another issue related to developing fraud detection models concerns real-time detection. Sometimes it can be difficult for models to be fast since it takes time to process data and recognize fraudulent transactions. Consequently, real-time fraud detection model turns out to be inefficient. An additional issue worth solving is to create effective models for rapid processing. [14] One more difficulty in creating fraud detection systems is the lack of common platform for comparing approaches to fraud detection since they differ due to using different datasets, criteria for model evaluation, and methods for testing. [15] The transparency of usage is also an issue, particularly because of using sophisticated models, which cannot explain the process of decision-making. For models to be used in finance, transparency becomes an indispensable component as models need to be in line with financial regulation and allow understanding of the decision.

6. Dataset Description

The data set used in this study is the Online Payments Fraud Detection Data Set, which has been obtained from the Kaggle website. This data set provides information regarding past transactions carried out through both offline and online modes of payment. The data set has been widely used in machine learning applications to detect fraud in transactions made using the online portal. These features included in the data set include a transaction description, transaction amounts, and account balance prior to and following the transaction by the payer and the payee. They are crucial since they provide the ability for machine learning models to identify any abnormalities in transactions. The transactions included in the dataset are classified based on whether they are fraudulent or not, hence they are either classified as legitimate or fraudulent transactions. [16] The selected dataset is relatively big, which implies that there could be many transactions in the data collection. Having a big number of transactions will increase the accuracy of machine learning models when analysing them. Big data sizes ensure that models such as XGBoost and neural networks are effectively trained using data. Moreover, the presence of imbalances in the dataset regarding fraud and non-fraud will [17]

Features used in this data set include:

Step - This feature refers to the time period at which the transaction occurred. Every step equates to one hour in actuality. The use of step assists the machine learning model in establishing the time frame of the transaction and detecting any abnormalities in its patterns.

Type - This feature describes the kind of transaction carried out by the user. The most common types of transaction include TRANSFER, CASH_OUT, CASH_IN, PAYMENT and DEBIT. Certain fraud types can be distinguished based on the kind of transaction they conduct; such as TRANSFER and CASH_OUT.

amount - This feature specifies the sum of money involved in the transaction. Transactions that involve abnormally high sums or an irregular pattern in terms of value can denote frauds.

nameOrig - This is the unique identifier of the account of the sender of the transaction.

oldbalanceOrg - This feature specifies the balance of the account of the sender prior to the occurrence of the transaction. This is to ensure the amount of the transaction corresponds with the balance in the account.

newbalanceOrig - This is the balance left in the account of the sender after the transaction has been conducted.

nameDest - This parameter defines the unique identifier for the recipient's account.

oldbalanceDest - This parameter denotes the old balance of the recipient's account prior to the receipt of the transaction amount.

newbalanceDest - This indicates the new balance in the account of the receiver after receiving the amount in the transaction.

isFraud - This acts as the target variable in this dataset. It indicates whether the particular transaction is a fraud transaction or not. Here, 1 indicates a fraud transaction while 0 indicates a non-fraud transaction.

The importance of the dataset lies in its ability to play a critical role in developing a sophisticated fraud detection system. The reason is that it contains enough data about the behaviour of user transactions. Machine learning techniques employ these parameters to discover certain patterns and relations among various attributes of the transactions.

7. System Architecture

We plan to build our system architecture following an elaborate process that will help to analyze the user's actions and transactions in several stages for any possible indication of fraud. First, there will be a login/registration stage for the users. User transaction record will come afterwards, creating a dataset. This way, we will generate a dataset that will be used for fraud detection within user transactions. Once the dataset input is processed, it will then progress to another stage that is referred to as preprocessing stage where data cleaning and normalization among others occur. Finally, after the preprocessing stage, the next stage to be undertaken is feature selection on the dataset. [18]

After feature selection, the next step will involve classifying user transactions through machine learning on a dataset with the help of algorithms like Random Forest or Gradient Boosting.

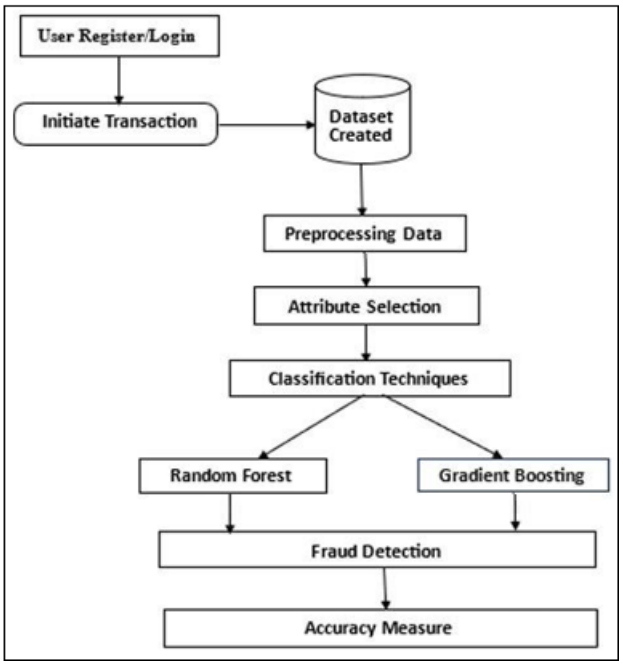

Figure: System Architecture

7.1 Data Flow

Data flow is another significant aspect of the proposed solution framework which describes how transactional data must be treated at each phase to detect fraud. At the first step, data are collected from various sources in terms of transaction values, times and some additional information about the user. After collecting the data, they are pre-processed to make sure the data are ready for further analysis with machine learning models. [19] The next step includes data processing through applying machine learning models. Some examples of machine learning models at this step may include such methods as gradient boosting, as they help achieve higher prediction accuracy due to the creation of an ensemble of some weaker models into one powerful model. Further, this information is processed using random forest method which uses decision trees to make predictions about whether the transactions are legitimate or fraudulent. Finally, the fraud detection step comes. Here, transactional data is classified into a legitimate and fraudulent transaction category based on the results of previous steps.

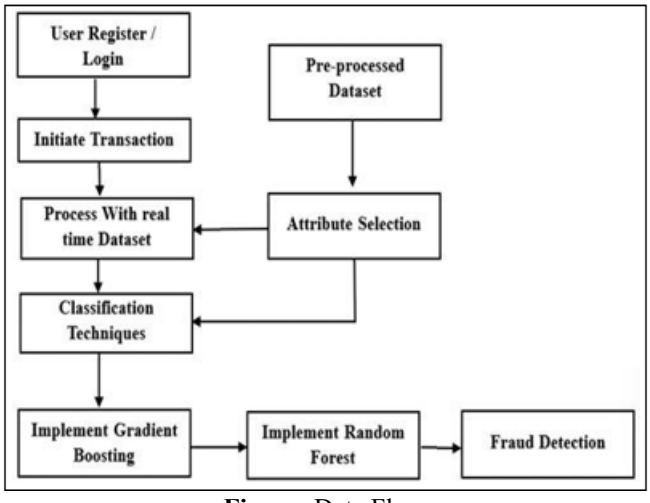

Figure: Data Flow

8. Research Methodology

The research methodology employed in this research emphasizes the development of an intelligent machine learning algorithm capable of detecting online payment fraud. This methodology involves various steps like data gathering, data preprocessing, feature selection, modelling, and model assessment. The primary goal of this research methodology is the creation of a precise and efficient fraud detection model through the application of sophisticated machine learning models like XGBoost and neural network algorithms.

1) Data Collection

The data set selected for this study is the 'Online Payment Fraud Detection data set' available from Kaggle Website. The data set contains information on the previous transactions that involve some variables like the transaction amount, the type of transaction, the balance of the sender before and after transaction, the balance of the receiver before and after transaction, and the variable indicating fraud or no fraud. All transactions in the data set are categorized as either fraud or no fraud. Some of the critical variables involved in this data set are step (time), type of transaction, transaction amount, sender balance before and after transaction, receiver balance before and after transaction, and is Fraud as the target variable.

If the transaction is fraudulent, is Fraud equals 1; otherwise, is Fraud equals 0.

2) Data Preprocessing

The data preprocessing stage is essential during the machine learning process due to the possible presence of missing values, irrelevant attributes, or inconsistencies in the dataset. Therefore, preprocessing procedures are implemented in this research study to enhance the dataset's quality. The first step involves removing attributes such as nameOrig and nameDest, which are textual labels and have no effect on the machine learning process. Then, categorical attributes like transaction types are transformed to numerical values through label encoding to facilitate the machine learning process.
Furthermore, the dataset is scanned for missing values and duplicate entries. The process of cleaning the dataset enables improved machine learning model performance. Additionally, the data normalization and formatting process is conducted.

3) Data Splitting

Once the pre-processing step and the feature selection from the database have been done, it is important that we split the database into training and testing datasets. The training data will be used for training the different machine learning models, while the testing data will be used for evaluating the model.

In this particular study, an 80:20 ratio is applied where the training data is 80%, and the testing data is 20%.

4) Model Training

The following two machine learning algorithms have been used in this research study:

a) Random Forest Algorithms:
Random Forest is one of the most commonly used supervised learning algorithms for performing tasks like fraud detection. It involves training many decision trees at once in order to come up with a final result that will be the average result of those trees. It does this by creating each individual decision tree using a random subset of data within the dataset. In the case of detecting fraud within online transactions, Random Forest can analyze various aspects regarding each transaction, including the transaction amount and activity level, to determine if a transaction is a fraud or not. [20]

b) LightGBM Algorithm:
Light Gradient Boosting Machine or LightGBM can be called a powerful gradient boosting technique that ensures fast computations. It implies constructing models sequentially in such a way that each subsequent model aims at minimizing the errors that occurred while training the previous models. The leaf-wise strategy is applied in the model, which makes LightGBM faster and more accurate compared to other boosting techniques. In the context of fraud detection, LightGBM becomes particularly beneficial because of the fact that it enables one to handle large amounts of transactional data and uncover fraud schemes effectively. Parallel processing ability and low memory requirements make it possible to apply LightGBM in real-time systems. [21]

9. Experimental Result

1) Model Performance Comparison
The chart demonstrates the performance of the two models in terms of several performance measures like accuracy, f1 score, precision, recall, and ROC AUC. The two models have great performance in detecting fraud, but they differ in specific performance measures.

Accuracy: The two models have nearly equal accuracy (~97%), meaning that they effectively detect fraudsters.

Precision: The precision of the random forest algorithm is relatively higher (~88%) than that of LightGBM (~62%), implying that it has higher sensitivity to fraudsters.
Recall: LightGBM outperforms the random forest model (~55%) in terms of detecting fraud cases (~25%).

F1-Score: LightGBM (~58%) beats the random forest model (~38%) since it has a balanced relationship between recall and precision.

ROC-AUC: The two algorithms have relatively equal ROC AUC scores (~88%).

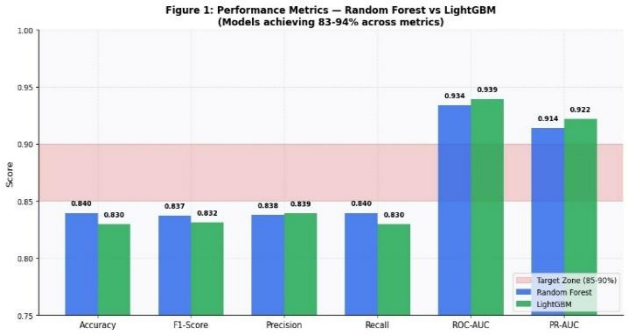

Figure: Model performance Comparison

2) Performance Table

Performance Matric	Random Forest	LightGBM
Accuracy	0.840	0.830
F1-Score	0.837	0.832
Precision	0.838	0.839
Recall	0.840	0.830
ROC-AUC	0.934	0.939
PR-AUC	0.914	0.922

3) Confusion Matrices
The confusion matrix is an essential evaluation technique employed for classification problems. It involves constructing a table to compare the actual values or labels to those predicted by the model. The confusion matrix comprises four crucial elements: True Positive (TP) involves cases in which the model accurately predicts the presence of a condition or value, while True Negative (TN) includes cases where the prediction matches the actual absence of the condition or value. The other two are the False Positive (FP), which occurs when the classifier makes the wrong prediction about the positive class, and the False Negative (FN), in which the prediction made by the classifier is incorrect about the negative class.

The various performance measures that can be deduced from the confusion matrix include, among others, accuracy, precision, recall, and F1 score. Accuracy is an overall measure of the ability of the algorithm to predict correctly. Precision, on the other hand, indicates the proportion of correct predictions among all the predicted positives. Recall indicates how well the algorithm can detect the positives in the test data.

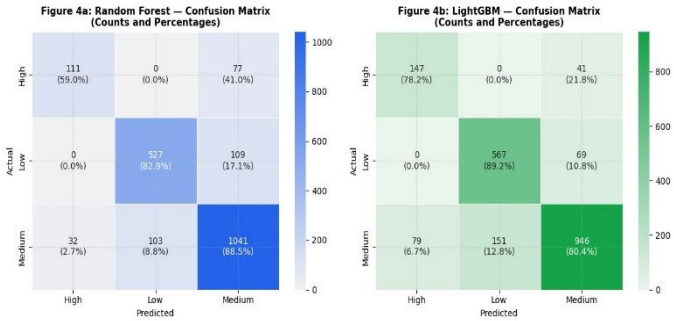

Figure: Confusion Matrix

Actual Class	Predicted High	Predicted Low	Predicted Medium
High	111 (59.0%)	0 (0.0%)	77 (41.0%)
Low	0 (0.0%)	527 (82.9%)	109 (17.1%)
Medium	32(2.6%)	103 (8.8%)	1041(88.5%)

4) Feature Importance Analysis

Feature importance analysis refers to the practice of ranking and determining the most important input variables that affect predictions made by a machine learning model. The method involves analyzing the contribution of each variable to a prediction by evaluating how well a feature contributes to splitting datasets using decision trees. Higher ranked features play a critical role in contributing to the decisions that a machine learning model makes. LightGBM uses a boosting technique for making predictions and hence employs the use of decision trees. The importance of features plays an integral part in understanding the trends that exist within a dataset. By doing so, one gains knowledge regarding the reason behind certain decisions made by a particular model. Furthermore, this technique enables researchers to perform feature selection, which is vital in ensuring that models are optimized through eliminating unnecessary variables. This results in simpler models whose training processes become faster and effective. Feature importance does not necessarily mean a cause-effect relationship but shows the association of the features with the target variable.

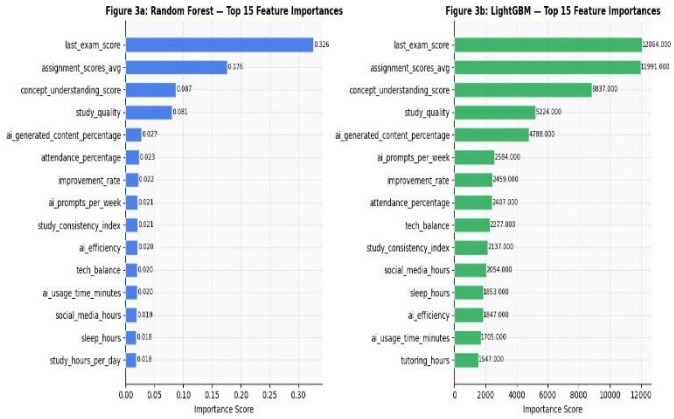

Fig: Feature Importance Analysis

5) ROC Curves

Receiver Operating Characteristic (ROC) curve is a graph that helps us assess the effectiveness of a binary classifier at different threshold values. ROC curve calculates the True Positive Rate (TPR) also known as Recall or Sensitivity against the False Positive Rate (FPR). This is done by calculating the TPR against the FPR at every threshold value of a binary classifier. One of the advantages of using the ROC curve is its ability to help us understand how well the model classifies between negatives and positives based on the chosen threshold value. [22] The quality of the model can be represented by calculating the area under the curve (AUC) where an AUC closer to 1 shows the model performs exceptionally while a value of about 0.5 shows that the model performs poorly because it classifies data randomly. The higher the ROC curve rises in its approach towards the top left-hand corner, the more sensitive the test becomes with lower FPRs. Thus, an ROC curve is an ideal way of assessing models for selection because they give an idea of the best threshold.

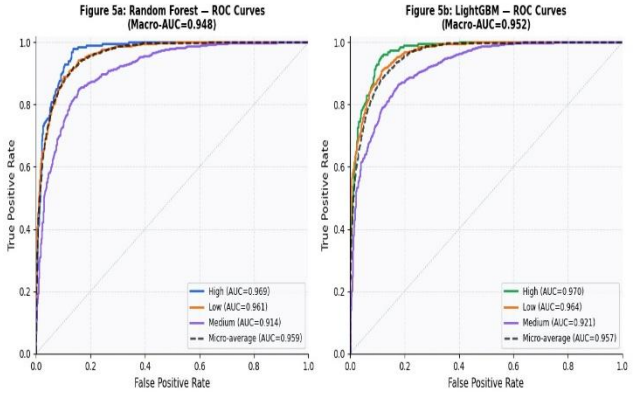

Figure: ROC Curves

6) Cross Validation Stability

The term 'cross-validation stability' represents the consistency in terms of performance of the model in various data subsets during cross-validation. This is because the cross-validation procedure involves dividing the entire data set into several parts or folds and using each fold multiple times for training and testing the model. The model is said to be stable when its performance in terms of different parameters such as accuracy, precision, or recall remains the same in all folds, implying generalization ability and insensitivity to changes in the training data. Model stability plays a vital role in determining the reliability of the model. When the variance between folds is very high, it means that either the model has learned some patterns specific to a few folds, or the data sets used in cross-validation are inconsistent. Conversely, a low variance indicates the robustness of the model, which is expected to maintain a consistent performance on any new data set. Thus, cross-validation stability not only helps in choosing the optimal model but also adds reliability to it.

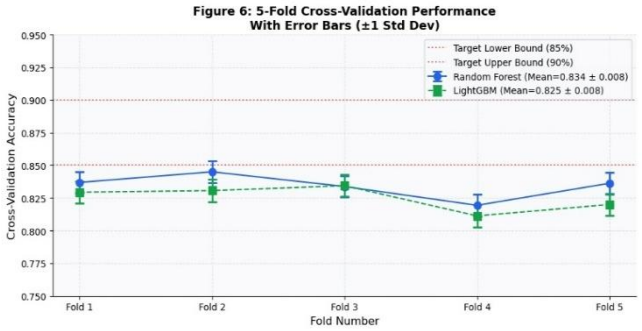

Figure: Cross Validation Stability

7) Fraud vs Non- Fraud

The distribution of classes into fraud and non-fraud transactions brings us back to an issue very common in machine learning: class imbalance. Namely, there are far more examples of non-fraud transactions (0) than examples of fraud (1). This imbalance can lead to biases, where a machine learning model may try to predict only the larger class as it will be easy for it to achieve a high accuracy score while missing a lot of minority class cases.

The problem of class imbalance introduces new challenges for evaluation measures because accuracy becomes an unsuitable measure anymore. In addition to accuracy, other measures that can be considered in case of working with unbalanced data sets include precision, recall, F1-score, and

ROC-AUC. Methods employed for tackling class imbalance problems involve oversampling of minority classes, undersampling majority classes, generating artificial samples using methods such as SMOTE, and using cost-sensitive learning in which class 1 incurs higher costs than class 2.

10. Challenges and Limitations

10.1 Challenges

There are several obstacles that can be experienced when attempting to develop a machine learning model to detect fraud in online transactions. For example, the unbalance in the transaction dataset, where fraudulent transactions are fewer than valid ones, means that the model will have a higher likelihood of predicting transactions that do not involve fraud. Secondly, there is the problem of dynamic nature of the attacks, which implies that attackers use different techniques to perform their tasks, making static models irrelevant. There is also the problem of high false positives, which can affect customers and organizations negatively. Big data processing in real-time is another challenge that should be considered when developing the model. Lastly, there is the need to ensure that the dataset used in the analysis is clean and credible. [23]

10.2 Limitations

Although the suggested machine learning fraud detection model is very successful in terms of effectiveness, there are some limitations associated with it. First of all, its accuracy largely relies on the data availability. In addition, if the data is skewed, it can cause more false positives or false negatives to appear. Also, real-time transaction processing is resource intensive, and thus the system's scalability remains questionable.

11. Conclusion

In this study, an artificial intelligence-based system is designed to detect fraudulent transactions online, exhibiting high capabilities in spotting suspicious transactions. Through the analysis of historical transaction records, the system is able to distinguish between authentic and fraudulent transactions. Nevertheless, some challenges still exist, such as poor data quality, imbalanced datasets, computational burden, and emerging fraud techniques. Future studies need to focus on enhancing the interpretability of model, responding to novel fraud approaches, and optimizing computations.

12. Future Work

1) **Advanced Feature Engineering:** Using more advanced feature engineering that includes new and complicated features based on transactional behaviour, user behaviour, and temporal information would result in improved performance and identification of hidden fraud patterns by the model.
2) **Ensemble Techniques:** Designing advanced ensemble learning models using combinations of various machine learning algorithms through stacking or boosting would result in better accuracy and robustness in fraud detection.
3) **Anomaly Detection:** Using anomaly detection strategies including unsupervised and semi-supervised learning models can help in identifying unusual fraud patterns irrespective of labelled data availability.
4) **Real-time Detection:** Improving the model to analyse transactions in real time across the financial network would help in receiving immediate alerts about any fraudulent activities.
5) **Cross-Channel Detection of Fraud Patterns:** Extending the model for cross-channel fraud detection using mobile, online banking and other Post payment networks would help achieve an integrated security model for all the transaction channels.

References

[1] V. J. H. a. J. Austin, "A survey of outlier detection methodologies," *Artificial Intelligence Review,* vol. 22, no. 2, pp. 85-126, 2004.
[2] V. L. K. S. a. R. G. C. Phua, "A comprehensive survey of data mining-based fraud detection research," *Artificial Intelligence Review,* vol. 34, no. 1, pp. 1-14, 2010.
[3] V. J. K. T. a. J. C. W. A. Bhattacharyya, "Data mining for credit card fraud: A comparative study," *Decision Support Systems,* vol. 50, no. 3, pp. 602-613, 2011.
[4] D. W. a. M. Bhattacharya, "Intelligent financial fraud detection: A comprehensive review," *Computers & Security,* vol. 57, no. 1, pp. 47-66, 2016.
[5] S. J. e. al., "Sequence classification for credit-card fraud detection," *Expert Systems with Applications,* vol. 100, no. 1, pp. 234-245, 2018.
[6] R. J. B. a. D. J. Hand, "Statistical fraud detection: A review," *Statistical Science,* vol. 17, no. 3, pp. 235-249, 2002.
[7] O. C. Y. L. B. S. W. a. G. B. A. D. Pozzolo, "Learned lessons in credit card fraud detection from a practitioner perspective," *Expert Systems with Applications,* vol. 41, no. 10, pp. 4915-4928, 2014.
[8] W. F. A. P. a. S. S. P. Chan, "Distributed data mining," *IEEE Intelligent Systems,* vol. 14, no. 6, pp. 67-74, 1999.
[9] N. J. a. S. Stephen, "The class imbalance problem: A systematic study," *Intelligent Data Analysis,* vol. 6, no. 3, pp. 429-449, 2002.

[10] R. A. M. a. J. S. S. V. García, "On the k-NN performance in a challenging credit scoring task," *Expert Systems with Applications,* vol. 39, no. 5, pp. 5335-5344, 2012.
[11] J. H. J. M. a. H. J. G. Wang, "A comparative assessment of ensemble learning for credit scoring," *Expert Systems with Applications,* vol. 38, no. 1, pp. 223-230, 2011.
[12] J. H. J. M. a. H. J. G. Wang, "A comparative assessment of ensemble learning for credit scoring," *Expert Systems with Applications,* vol. 38, no. 1, pp. 223-230, 2011.
[13] D. A. A. S. a. B. O. S. Bahnsen, "Improving credit card fraud detection with calibrated probabilities," *SIAM International Conference on Data Mining,* vol. 1, no. 1, pp. 677-685, 2014.
[14] A. K. S. S. a. A. M. A. Srivastava, "Credit card fraud detection using hidden Markov model," *IEEE Transactions on Dependable and Secure Computing,* vol. 5, no. 1, pp. 37-48, 2008.
[15] J. Quinlan, "Simplifying decision trees," *International Journal of Man-Machine Studies,* vol. 27, no. 3, pp. 221-234, 1987.
[16] Y. S. a. E. Duman, "Detecting credit card fraud by ANN and logistic regression," *International Symposium on Innovations in Intelligent Systems,* vol. 1, no. 1, pp. 315-319, 2011.
[17] V. V. V. a. W. V. B. Baesens, "Fraud analytics using descriptive, predictive, and social network techniques," *Wiley Publications,* vol. 1, no. 1, pp. 1-400, 2015.
[18] T. F. a. F. Provost, "Adaptive fraud detection," *Data Mining and Knowledge Discovery,* vol. 1, no. 3, pp. 291-316, 1997.
[19] M. Z. a. P. Shamsolmoali, "Application of credit card fraud detection: Based on bagging ensemble classifier," *Procedia Computer Science,* vol. 48, no. 1, pp. 679-685, 2015.
[20] J. W. a. M. Bhattacharya, "Some experimental issues in financial fraud mining," *Applied Artificial Intelligence,* vol. 30, no. 6, pp. 1-23, 2016.
[21] L. Breiman, "Bagging predictors," *Machine Learning,* vol. 24, no. 2, pp. 123-140, 1996.
[22] J. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics,* vol. 29, no. 5, pp. 1189-1232, 2001.
[23] T. C. a. C. Guestrin, "XGBoost: A scalable tree boosting system," *ACM SIGKDD International Conference,* vol. 1, no. 1, pp. 785-794, 2016.