International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/CSEIT25112825

Fraud Detection – A Hybrid Machine Learning Approach

Sahil Jena, Divyansh Rajput, Tatsat Pathak

Department of Computer Engineering, Sigma University, Vadodara, Gujarat, India

ARTICLEINFO

Article History:

Accepted : 05 April 2025

Published: 09 April 2025

Publication Issue

Volume 11, Issue 2

March-April-2025

Page Number

3463-3470

ABSTRACT

Online Fraud is an often a major crime in online money transactions between a sender and recipient. While the transaction occurs online, some incidents may occur where the transactions are made without sender's consent. There is a potential of many of us who becomes victim or may can be which is intangible and dynamic. This paper argues and help with the methods to prevent from happening the fraudulent. Once the fraud occurs there is less chances for someone to get their hard-earned money back, whereas some of them had also lost their lifetime savings. However, the decisions over it should be more transparent and accurate to win the trust of regulators, businesses and bank. Also, it is a critical issue in financial transactions, especially with the rise of digital payments. The data for this study is obtained with the help of real-time transactions which shows daily transactions scheme known as PaySim. It is a simulator for mobile money transactions, provides a realistic dataset to analyse fraud patterns. At the same time fraud detection methods will help to resolve the issue. Whereas the notorious fraud detection dataset is well known for its several legitimate transactions rather than fraudulent ones. Our paper aims to use the methods of machine learning which helps to predict fraud transactions occurs. Although the Online Fraudulent of money is crime in world but criminals often do this without a guilt which comes under IT ACT, 2000 and PREVENTION OF MONEY LAUNDERING ACT(PMLA) 2000.

KEYWORDS: Fraud detection, Ensemble models, Classification.

Introduction

There is no doubt online money transaction has been grown very faster in these recent years. With increasing amount of money being transacting from one account to another the risks of getting money got lost from a bank account of one without their consent. It is also a major factor in this area of online transaction. Money can be sent through different mode like USER PAYMENT INTERFACE(UPI), Bank Transfer, and Electronic Payments where the sender and receiver relation is based upon how safe and secure the transaction occurs. Mobile Money

Transaction allows a one to establish financial account from having no financial account. MMTs which was started in Kenya in year 2007 refers as M-Pesa, which was introduced by Safaricom [1]. It quickly spread to the worldwide technologies.

According to the **Ministry of Electronics and Information Technology (MEITY)**, Government of India the growth of MMT is substantially increased over the years. In INDIA, the volume of online money transaction has been increased from $645.15 million in F.Y. 2019-20 to $1.41 billion in F.Y. 2023-24. India's mobile money industry is grown at a compound annual growth rate (CAGR) of 45% in the year of 2017-2022. However, by the time new technology is introduces to prevent such vulnerabilities, crooks find new ways to fraudulent from someone bank [2]. One of the popular vulnerability in USSD is known as SMShing. It is a phishing attack performed through SMS. The work is however in its initial stages and many conclusions cannot be drawn about it [2]. However, over the years the many methods are been developed which helps to stop the occurrence of the fraudulent. It has been seen that the most fraudulent does occur on the mobile number linked with the particular bank account which is then, the fraudster does their work.

One of the methods which was delivered as to manual verification which will be held at the time of the person who is withdrawing the money, which will help to secure their money. As this method is also not so reliable due to the human work which cannot be done at a very large number of bank accounts compare to a Machine which can work unfailingly and continuously 24/7 [3].

Our approach result helps the society to dealt with any fraudulent or any unknown activity occurs while the transaction occurs. We have experimented with many supervised learning and ensemble models. Some of our strengths are:

(1) Robust prediction from the ensemble models and deep learning.
(2) Concise results with different method applied.

RELATED WORK

The methodologies of Machine learning are widely used in many fields of research, especially given the rise of digital payment systems and mobile money platforms. A common challenge across many of these studies is the **class imbalance problem**, it refers to the small part of fraudulent transaction present in any dataset. Traditional statistical models often struggle in such scenarios, which motivates to use of **Advanced machine learning algorithms**.

Bhattacharyya et al. (2011) [4] evaluated multiple classifiers which includes Neural Network and Decision Tree for credit card fraud detection, emphasizing the importance of precision and recall in heavily imbalanced datasets. In a seminal survey, *Ngai et al. (2011)* [5] categorized machine learning techniques applied in financial fraud into five types: classification, clustering, prediction, outlier detection, and visualization — with classification models showing the highest success rates.

The **PaySim simulator**, developed by *Lopez-Rojas and Axelsson (2016)* [6], replicates mobile money transactions and has been frequently used in recent works for fraud detection benchmarking. Its realistic transaction flow makes it a compelling alternative to traditional datasets. *Adewumi et al. (2019)* utilized PaySim to compare supervised models like SVM, Random Forest, and Neural Networks, noting that ensemble models outperformed single classifiers in terms of accuracy and robustness.

Random Forest and **AdaBoost** types of Ensemble models have gained significant attention due to their ability to handle noise and imbalance effectively (*Jha et al., 2012*) [7]. *Chen et al. (2004)* introduced **Gradient Boosting Machines**, another ensemble approach, which has since been adapted for fraud detection with notable success.

Meanwhile, **deep learning models**, particularly feedforward neural networks and LSTM-based architectures, are being increasingly explored for detecting complex temporal fraud patterns (*Roy et al., 2020*) [8]. However, these models often require more

data and tuning, and their black-box nature remains a concern for financial regulators.

Despite these advancements, many models are either overfitted to specific datasets or fail to generalize across platforms. This research builds on these prior works by performing a comprehensive evaluation of six diverse machine learning techniques — including deep learning and ensemble methods — on the PaySim dataset, and presenting a comparative performance analysis across multiple error metrics (accuracy, precision, recall, f-measure). This study thus contributes toward developing more practical, scalable, and legally-aware fraud detection frameworks.

METHOD

The dataset which is used for the model training is showcase below.

A. Dataset

We have used a dataset taken from Kaggle which is based on a real-world transaction from an internation provider. Dataset consists of 6354407 data points with 10 attributes.

Attribute	Description
type	Cash-deposit, CASH-withdraw, Cash-received, Cash-Payee, and Cash-sent
amount	Local currency transaction
nameOrig	Transaction done by customer
oldbalanceOrg	Initial balance before transaction
newbalanceOrig	Balance after transaction
nameDest	recipient ID of the transaction
oldbalanceDest	initial recipient balance before the transaction
newbalanceDest	New balance in recipient after transaction
isFraud	Fraudulent transactions as 1 and non-fraudulent as 0
isFlaggedFraud	This flagged the fraud transactions which occurs in an account

TABLE 1. DATASET DESCRIPTION

B. PREPROCESSING

The cleaning of this dataset was the most important part for which the models can be then performed. Further the data is divided using different metrics. It doesn't assume the data is normally distributed so that it can be processed easily.

C. LOGISTIC REGRESSION

It is based upon a statistical method which is solve binary classification which predicts the categorical outcome. It predicts that a given input can be of a class. It works likely Linear Regression but instead it gives either one or two response variables. It allows multiple explanatory variables being analysed simultaneously. The result we get is due to every variable with the odds ration [9].

D. ADA BOOST

Adaptive Boost is an ensemble method which combines multiple weak learners and vote out the best one among it to find out the best results. It trains all the samples and with equal weights, once a misclassified sample is found it boosts it weight and at the end it combines all the weak learner together and find the strongest one. As our PaySim Dataset is highly imbalanced data with groups of fraud and non-fraud it focuses more on the misclassified samples and give the best results. Ada boost allow itself to combined with different method which found the weak hypothesis without any need to get the knowledge of Weak Learn [10].

E. RANDOM FOREST

RF is a tree-based supervised machine learning algorithm that can be used for both classification and regression problems. It was used for classification in this paper [2]. RF use large number of decision tree inside it, which sums up the results of it to produces final output which shows the best vote in it. Its ability to handle multiple non-linear and imbalanced data made it to the highest used classification algorithms [11–12].

F. GAUSSIAN NAÏVE BAYES(GNB)

GNB is supervised learning that gives the best probabilities not predictions which helps our dataset

for this paper to get the best results for the high dimensions. This algorithm is used for the classification part. Looking on the computational efficiency performed by GNB made it to largely used. We evaluate the value of probability of P(x|y), where x shows the object and y shows the class. The given details in the dataset used by GNB makes an apparatus [13].

G. DECISION TREE

Decision tree refers to machine learning algorithm, which gives multiple output which we can further aggregates to get our final result. It mimics like human which first splits the data into parts/branches. Most node consists of single incoming edge. An outgoing edges node can be called as **Test** or **Internal node**. We can understand this with the help of Figure 3.1 which describe the response of direct mailing with spreading out of different nodes. Rest nodes present in the branch is known as Leaves. In this method, the present Test node gets splits in many to the input attribute values [14].

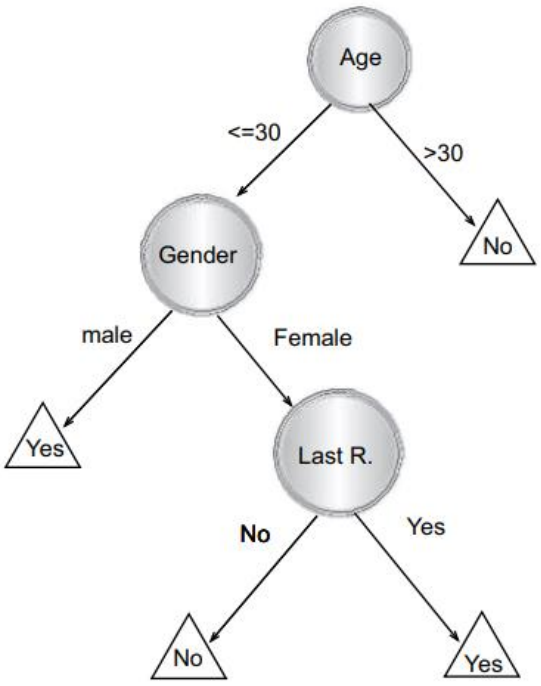

Figure 3.1. Response to Direct Mailing with the Customer.

H. NEURAL NETWORKS

The neural network is a machine learning model which works like a human mind, which is used to find patterns and make predictions on given dataset. Figure 3.2 shows how it processes.

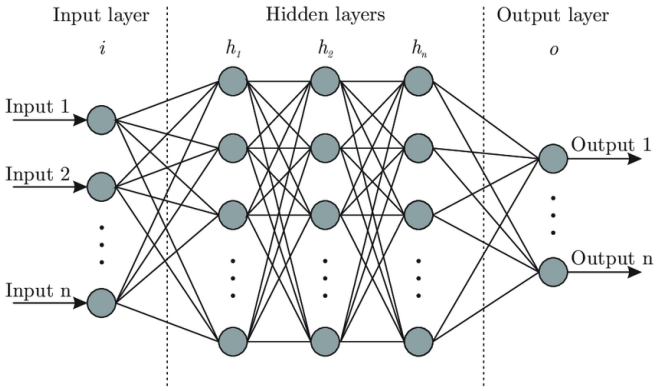

Figure 3.2 Process of Neural Network.

The input layer takes the input as a pixel value from an image or features of dataset given to it train. The hidden layer processes the data with the help of mathematical operations and the Output Layer or Final layer gives the predictions based on the input.

I. EVALUATION METRICS

This section of paper helps us to evaluates the metrics which will performs by different models and get the best results. The criteria consists of different factors such as True Positive, False Positive, True Negative, and False Negative , which will be used to evaluate the results.

Accuracy: Accuracy is mainly a ratio of TP+TN and TP+TN+FP+FN. It gives the results which describe which model is best to predict the result. It may show false sometimes due to many imbalanced data presents in any dataset. For evaluating classification models, accuracy is a major metric. It has the following definitions formally:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Precision: Precision is a ratio of True Positive and the total of True Positive and False Positive. The correctly predicted positive instances by any model we use is shown in the precision table and helps to find the optimal solution. The following defines it:

$Precision = \frac{TP}{TP+FP}$ (3)

Recall: The ratio of True Positive to the total of True Positive and False Negative which gives us the recall value of any model. It focuses on the classification algorithm which gives TP's. The following defines it:

$Recall = \frac{TP}{TP+FN}$ (4)

F-Measure: It is the ratio of double the product of Precision and Recall and the total of Precision and Recall. The F-Measure combines both the properties of Precision and Recall. We can get the best model by looking at Accuracy and F-Measure. The following defines it:

$F - Measure = \frac{2*Precision*Recall}{Precision+Recall}$ (4)

RESULTS AND DISCUSSION

The experiments result for overall **Accuracy** is present in Table 2. It ranged from 99.92% for Logistic Regression, 99.91% for Ada Boost, 99.97% for Random Forest (RF), 99.22% for Gaussian Naïve Bayes (GNB), 99.97% for Decision Tree and 99.94% for Neural Networks. Random Forest and Decision Tree have performed the best in the overall accuracy compared to other 4 methods performed and show the best outcome for this. However, RF and Decision Tree are the best classifier, following Neural Networks as second and GNB as third.

The result of **Precision** is present in Table 3. It ranged from 86.47% for Logistic Regression, 99.23% for Ada Boost, 98.09% for Random Forest (RF), 3.20% for Gaussian Naïve Bayes (GNB), 87.68% for Decision Tree and 95.73% for Neural Networks. Ada boost shows the best Precision Results of the dataset, along with GNB as the lowest. It indicates that all the rest classifiers had a high positive instance.

The result **Recall** is present in Table 4. It ranged from 45.37% for Logistic Regression, 31.35% for Ada Boost, 78.03% for Random Forest (RF), 18.44% for Gaussian Naïve Bayes (GNB), 86.67% for Decision Tree and 76.51% for Neural Networks. Decision Tree shows the best results among the other instance, whereas others performs well and above average.

The results for the **F-Measure** are in Table 5. It ranged from 45.37% for Logistic Regression, 47.64% for ADA BOOST, 86.92% for RF, 5.46% for GNB, 87.17% for Decision Tree and 85.05% for NN. We can predict the best model with the help of this report table.

Classifier	Accuracy
Logistic Regression	99.91%
Ada Boost	99.91%
Random Forest	99.97%
Gaussian Naïve Bayes	99.22%
Decision Tree	99.97%
Neural Networks	99.94%

TABLE 2: ACCURACY REPORT

Classifier	Precision
Logistic Regression	86.47%
Ada Boost	99.23%
Random Forest	98.09%
Gaussian Naïve Bayes	3.20%
Decision Tree	87.68%
Neural Networks	95.73%

TABLE 3: PRECISION REPORT

Classifier	RECALL
Logistic Regression	45.37%
Ada Boost	31.35%
Random Forest	78.03%
Gaussian Naïve Bayes	18.44%
Decision Tree	86.67%
Neural Networks	76.51%

TABLE 4: RECALL REPORT

Classifier	F-MEASURE
Logistic Regression	59.51%
Ada Boost	47.64%
Random Forest	86.92%
Gaussian Naïve Bayes	5.46%
Decision Tree	87.17%
Neural Networks	85.05%

TABLE 5: F-MEASURE REPORT

4.1 EVALUATION

In total of all the models performed on the dataset, **Decision Tree Classifier** demonstrated superior performance with an **accuracy of 99.97%**. However other all the models performed excellent in these criteria.

The **Precision** score also performs a vital role to measure the model 3 of the models performing above 98%. Rest models were above average too.

The **Recall** scores show us 3 of the models performing above 75% which is best for the models.

As above stated, **F-Measure** It is the ratio of double the product of Precision and Recall and the total of Precision and Recall. It gives us 3 models above the range of 85%, which decided our best model.

The work of Xuan et.al [15] on credit card fraud detection gives us the accuracy of 98.67%.

Lu et.al [11] work also performed some similar method on the same dataset we used for this paper and gives results like 99.2% for Random Forest, 99.2% for Decision Tree and other models like Logistic Regression, Support Vector Machine, and Shallow Neural Network.

CONCLUSION

In the growing landscape of digital transactions, the detection of online financial fraud has become both a technical and social necessity. This study explored multiple machine learning techniques to develop an effective fraud detection system using the PaySim dataset, a realistic simulation of mobile money transactions. Our main goal is to get the best performance of the models and vary it with different metrics and showcase the best suited for real-time fraud prevention.

A total of six machine learning models were analyzed: **Logistic Regression, AdaBoost Classifier, Random Forest Classifier, Gaussian Naive Bayes, Decision Tree Classifier,** and a **Neural Network (Sequential Model)**. Each model was assessed based on metrics like **Accuracy**, **Precision score**, **Recall Score** and **F-Measure**.

Among them, the **Decision Tree Classifier** demonstrated superior performance with an **accuracy of 99.97%**, a **F-Measure of 87.17%**, indicating high precision in distinguishing between legitimate and fraudulent transactions. The **Random Forest Classifier** closely followed with a similar accuracy and error rates. Meanwhile, the **Neural Network model** also performed robustly, achieving **99.94% accuracy**, showcasing the best results of deep learning in handling the imbalanced datasets.

On the other hand, **GaussianNB** lagged behind with a accuracy of **99.22%**, and significantly low F-Measure values, likely due to its assumptions not aligning well with the distribution of the dataset. The Logistic Regression and AdaBoost models, achieved **99.92%** and **99.91% accuracy**, showed slightly higher error rates compared to the ensemble and tree-based models.

These results affirm that **ensemble methods (Random Forest, AdaBoost)** and **deep learning approaches** are particularly effective in fraud detection tasks involving large-scale, imbalanced financial datasets. Additionally, metrics like Accuracy, Precision, Recall and F-Measure provided deeper insight into model precision beyond basic accuracy.

The study also emphasizes the broader societal and legal context of fraud, referencing the **IT Act (2000)** and **Prevention of Money Laundering Act (2002)**, reinforcing the need for technologically sound and legally aware fraud detection systems. Future work should focus on deploying these models in real-time environments, refining model interpretability, and enhancing adaptive learning to counteract evolving fraud tactics.

DATASET

https://www.kaggle.com/datasets/ealaxi/paysim1

References

[1]. Buku, M.W.; Meredith, M.W. Safaricom and M-Pesa in Kenya: financial inclusion and

financial integrity. Wash. Jl tech. & arts 2012, 8, 375.

[2]. Botchey, F. E., Qin, Z., & Hughes-Lartey, K. (2022). An Evaluation of Machine Learning Methods to Predict Fraud in Mobile Money Transactions. International Journal of Engineering Research & Technology (IJERT), 11

[3]. Shaukat, K.; Iqbal, F.; Alam, T.M.; Aujla, G.K.; Devnath, L.; Khan, A.G.; Iqbal, R.; Shahzadi, I.; Rubab, A. The Impact of Artificial intelligence and Robotics on the Future Employment Opportunities. Trends in Computer Science and Information Technology 2020, 5, 050–054.

[4]. Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems, 50(3), 602–613.

[5]. Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559–569.

[6]. Lopez-Rojas, E., & Axelsson, S. (2016). Money laundering detection using synthetic data. In The 27th European Modeling and Simulation Symposium.

[7]. Jha, S., Guillen, M., & Westland, J. C. (2012). Employing transaction aggregation strategy to detect credit card fraud. Expert Systems with Applications, 39(16), 12650–12657.

[8]. Roy, A., Sunitha, C., & Saha, S. (2020). LSTM-based deep learning model for detecting financial fraud. In International Journal of Engineering Research & Technology (IJERT), 9(9), 45–49.

[9]. Sperandei S. Understanding logistic regression analysis. Biochem Med (Zagreb). 2014 Feb 15;24(1):12-8. doi: 10.11613/BM.2014.003. PMID: 24627710; PMCID: PMC3936971

[10]. Chengsheng, Tu & Huacheng, Liu & Bing, Xu. (2017). AdaBoost typical Algorithm and its application research. MATEC Web of Conferences. 139. 00222. 10.1051/matecconf/201713900222.

[11]. Lu, C.; Lee, C.T.; Qiu, H.; Liu, M. Compare Shallow Neural Network and Conventional Machine Learning in Predicting Money Laundering Crimes

[12]. Liu, X.Y.; Wu, J.; Zhou, Z.H. Exploratory undersampling for class-imbalance learning. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 2008, 39, 539–550.

[13]. Anand, M. V., KiranBala, B., Srividhya, S. R., Kavitha, C., Younus, M., & Rahman, M. H. (n.d.). Gaussian Naïve Bayes Algorithm: A Reliable Technique Involved in the Assortment of the Segregation in Cancer.

[14]. Rokach, Lior & Maimon, Oded. (2005). Decision Trees. 10.1007/0-387-25465-X_9.

[15]. Xuan, S.; Liu, G.; Li, Z.; Zheng, L.; Wang, S.; Jiang, C. Random Forest for credit card fraud detection. 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC). IEEE, 2018, pp. 1–6.

[16]. Lopez-Rojas, E.; Elmir, A.; Axelsson, S. PaySim: A financial mobile money simulator for fraud detection. 28th European Modeling and Simulation Symposium, EMSS, Larnaca. Dime University of Genoa, 2016, pp. 249–255.

[17]. Adewumi, A. O., Akinyelu, A. A., & Dada, E. G. (2019). Comparative analysis of machine learning techniques for financial fraud detection using PaySim dataset. Procedia Computer Science, 163, 247–254.

[18]. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[19]. Paul, A.; Mukherjee, D.P.; Das, P.; Gangopadhyay, A.; Chintha, A.R.; Kundu, S. Improved random forest for classification. IEEE Transactions on Image Processing 2018, 27, 4012–4024.

[20]. Evgeniou, Theodoros & Pontil, Massimiliano. (2001). Support Vector Machines: Theory and Applications. 2049. 249-257. 10.1007/3-540-44673-7_12.