Enhancing Mobile Money Transaction Fraud Detection: A Comparative Analysis of Machine Learning Models

Humaira Noor
Deptartment of Computer Science and Engineering
United International University, Dhaka, Bangladesh
hnoor222007@mscse.uiu.ac.bd

Kangkhita Hassin
Deptartment of Computer Science and Engineering
United International University, Dhaka, Bangladesh
khassin221015@mscse.uiu.ac.bd

Khondokar Riaz Mahmud
Deptartment of Computer Science and Engineering
United International University, Dhaka, Bangladesh
kmahmud2310008@mscse.uiu.ac.bd

Dr. SWAKKHAR SHATABDA, Ph.D.
Department of Computer Science and Engineering
United International University, Dhaka, Bangladesh
swakkhar@cse.uiu.ac.bd

***Abstract*—Financial transactions have been changed by the rise of mobile money services, which are now accessible and convenient to millions of people globally. However, this ease has also led to an increase in fraudulent mobile money transactions. Maintaining the integrity of these services and protecting users' financial assets in this changing environment depends heavily on identifying and combating fraud. The complete method used in this work to predict mobile money transaction fraud makes use of cutting-edge machine learning methods. We investigate the use of logistic regression, KNN, decision trees, random forests, and neural networks to detect fraud using real-time transaction data. Our research shows that these algorithms can significantly improve fraud detection, strengthening the safety of mobile money transactions.**

***Index Terms*—mobile money transactions ;fraud; Machine Learning**

I. Introduction

Mobile money services have revolutionized how people and organizations conduct financial transactions, making it possible to transfer money, pay bills, and access financial services without a traditional bank account in many cases. But this shift hasn't been without difficulties. The trust and dependability that support mobile money services (MMSs) are in danger due to a rise in fraud incidents that have coincided with MMSs' rising popularity. The need to properly combat fraud is one of the most serious concerns that mobile money providers confront today. Fraudsters constantly improve their strategies, taking advantage of flaws in the infrastructure and using the simplicity and quickness of mobile money transfers to their advantage.

To address this growing concern, financial institutions, and law enforcement agencies must deploy sophisticated tools and techniques capable of detecting fraudulent activities in real time. This paper, reflecting the urgency of the matter, goes beyond traditional approaches by exploring a comprehensive set of machine learning models, including logistic regression, K-nearest neighbors (KNN), decision trees, random forests, and neural networks, in the quest to predict mobile money transaction fraud effectively. These models are well-suited for sifting through vast datasets and identifying patterns and anomalies that may be indicative of fraudulent behavior. In this paper, we present our research findings, which contribute significantly to the ongoing efforts to combat mobile money transaction fraud. By employing a diverse set of machine learning models, we explore their strengths and weaknesses and emphasize their potential to revolutionize fraud detection in mobile money services. Ultimately, we aim to equip financial institutions and authorities with a powerful and versatile toolkit to maintain the security and integrity of mobile money transactions in an evolving digital financial landscape.

Identify applicable funding agency here. If none, delete this.

II. Dataset

[1]In this dataset, mobile money transactions were simulated using PaySim. The simulations were built using data from financial logs produced by an MMS installed in an African country, which included a sample of real mobile money transactions. A multinational company that delivers mobile financial services, which is currently active in over 14 countries, gave the initial logs. The initial logs, which served as the dataset's foundation, were shared by the corporation. Nine distinct features were gathered in a total of 6362620 rows of data. Except for "Step" and "Type," all characteristics are continuous variables with a numeric value. When categorizing transfers, the "Type" feature divides them into five groups: cash-ins, cash-outs, debit, payment, and transfer. "Step" stands for the unit of time. Fraud is the dependent variable in this investigation. The transactions made by dishonest actors in the simulation are referred to here as "fraud" in this context. These dishonest agents carry out actions intended to seize control of client accounts and launder money by moving it to another

system. The money is then taken out of the system by cashing it out.

III. Related works

Numerous studies on automated fraud detection have been carried out.

[6]The Authors of this review paper stated that due diligence, continuous monitoring, and robust mechanisms like credit scoring can reduce NPAs and frauds. They also elaborated Machine learning (ML) techniques are being explored to improve credit risk evaluation accuracy, but a lack of comprehensive public datasets remains a concern. [9]Han et al. examined the drawbacks of conventional AML techniques in more recent research, as well as the potential advantages of AI in identifying and forecasting suspicious transactions. The authors examined various anomaly detection, clustering, classification, and predictive modeling AI-based techniques used in AML compliance. They discovered that these methods improve the efficacy and efficiency of AML initiatives. To ensure AI's effective and moral application in stopping money laundering, they also emphasized the significance of careful implementation and monitoring.

Other studies have looked at using ML techniques to identify and stop the financing of terrorism and money laundering. [8] [5]

Other researchers have used ML specifically to look at mobile money transfer fraud. [4]

The use of case-based studies to identify fraud in mobile money transfer systems has also been investigated, with encouraging outcomes. Adedoyin et al. [3] suggest a novel method for predicting fraud in mobile money transfer systems in a recent study. The authors investigate the use of case-based reasoning using ML techniques to identify suspicious transactions in light of the various challenges providers face in preventing fraud. According to the study, the case-based approach performs better than conventional techniques at accurately detecting fraud. These results underline the need for additional research in this field and highlight the potential of case-based reasoning to improve the security of mobile money transfer systems. The performance of ML algorithms is comparable in experiments using real-world mobile money transaction data in terms of overall prediction accuracy. When choosing ML algorithms for fraud prediction, Botchey et al. [7] emphasize the trade-offs between performance accuracy and computational efficiency and discuss the implications of these findings.

IV. Methodology

A. Selection of ML Algorithms:

The algorithms chosen for this project to anticipate financial crimes was based on the characteristics of the data and their use in financial fraud detection in various papers. These algorithms have successfully been used in financial crime research and as the dependent variable is a categorical one, it's a classification problem. The models being trained include those for logistic regression, KNN, decision trees, random forests, and neural networks.

B. EDA:

The "Type" feature divides them into five groups: cash-ins, cash-outs, debit, payment, and transfer.

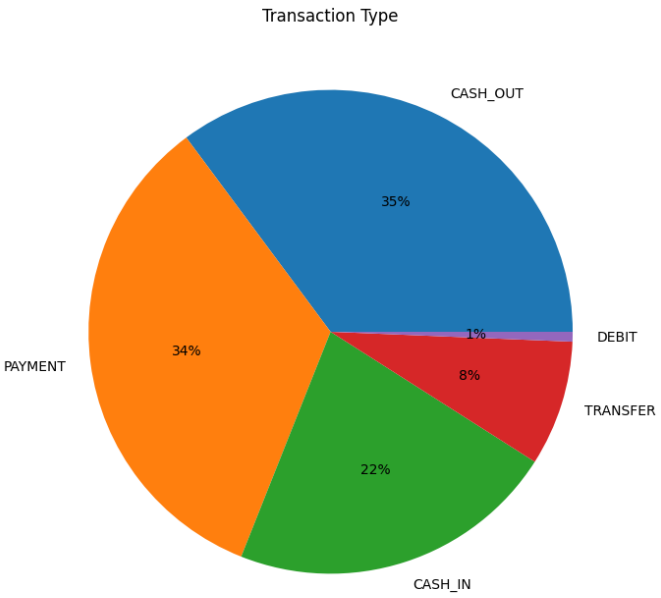

Fig. 1.

There are two flags: isFraud: indicates the actual fraud transactions isFlaggedFraud : The system prevents the transaction when an attempt to transfer more than 200.000 in a single transaction made Let's check what kinds of transactions are being flagged and are fraud

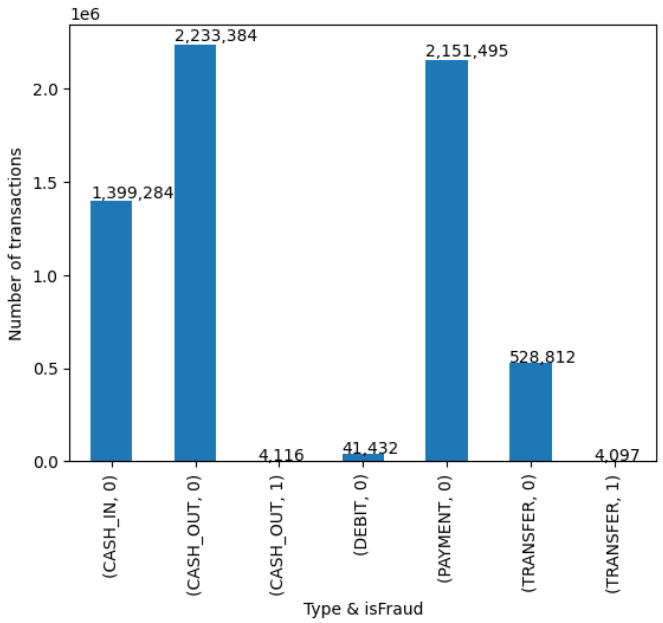

Fig. 2.

Separating the dataset where isFraud==1, which signifies the transactions that were frauds. If we observe the type of transaction, we can see that 'TRANSFER', and 'CASH_OUT' are the only types of transactions that had fraudulent activity.

Fig. 3.

Also, from checking the "nameOrig" and "nameDest" we can see that fraudulent activity didn't happen from the merchant side.

C. Data cleaning and preprocessing:

Before the building of the model, several aspects were considered unnecessary and dropped. As a result, the features "nameOrig" and "nameDest" have been deleted because they are no longer relevant. The longitude and latitude data required to identify the destinations are missing from these features. The feature "isFlaggedFraud" also showed no change and was therefore removed from the model. If there was any missing value, that was checked and handled.

D. SMOTE for imbalanced data:

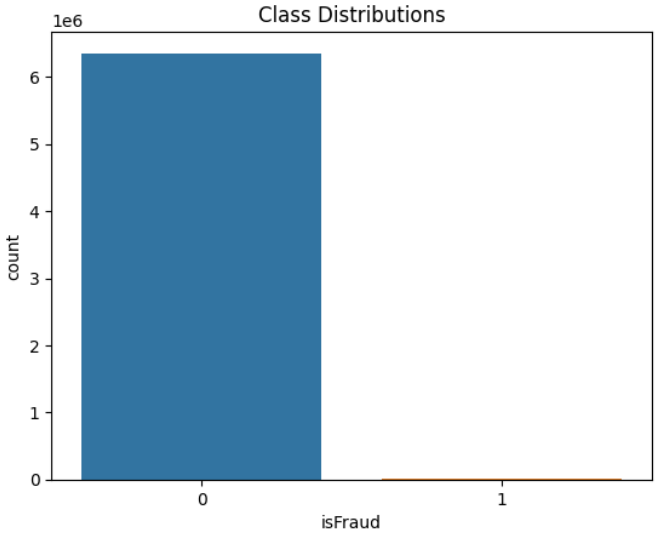

Fig. 4. Before smote

The dataset that was used for this investigation showed a significant level of class imbalance. The ratio of fraud to no fraud was 1% for fraud and 99% for no fraud. Upsampling the minority class is a frequent strategy for dealing with unbalanced datasets. There are several techniques to upsample data, but one technique that is frequently used is the synthetic minority oversampling technique (SMOTe).

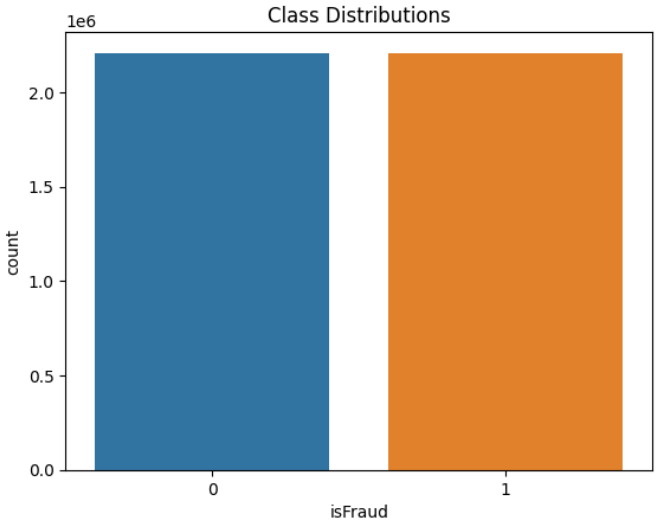

Fig. 5. After smote

E. Training and results:

After Smote, The set was split into train and test datasets. 80% of the data was for training and the rest was for testing. The test data accuracy is the primary focus.

F. KNN

Accuracy of KNN classifier with k=3: 99.35%

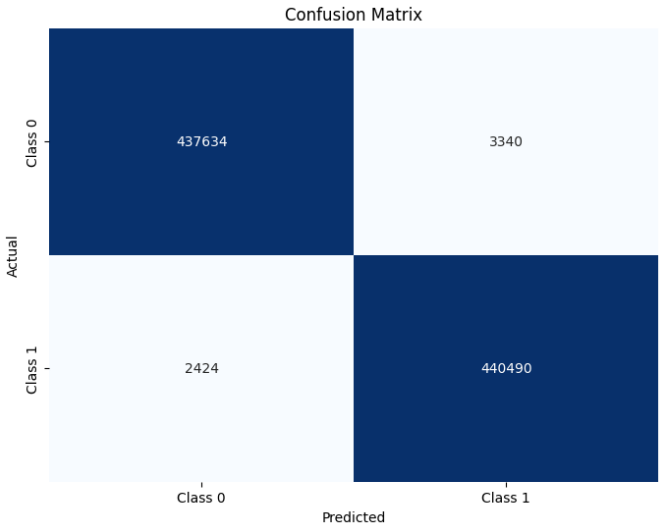

Fig. 6. Confusion Matrix:KNN

G. Logistic Regression

Accuracy of Logistic Regression classifier: 93.51%

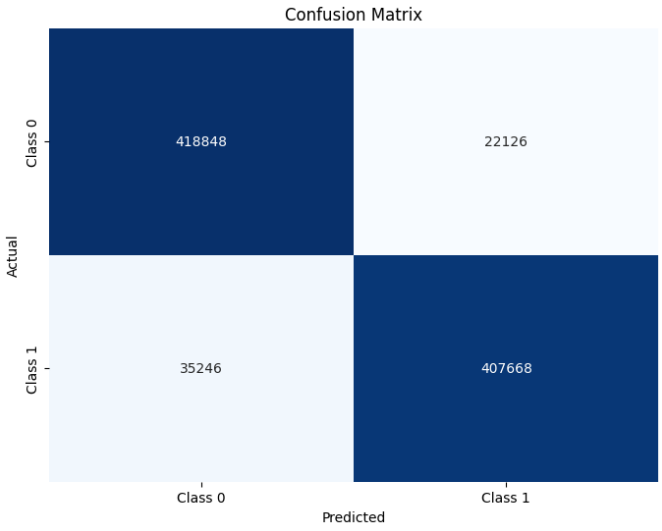

Fig. 7. Confusion Matrix:Logistic Regression

H. Decision Tree classifier

Accuracy of Decision Tree classifier: 91.11%

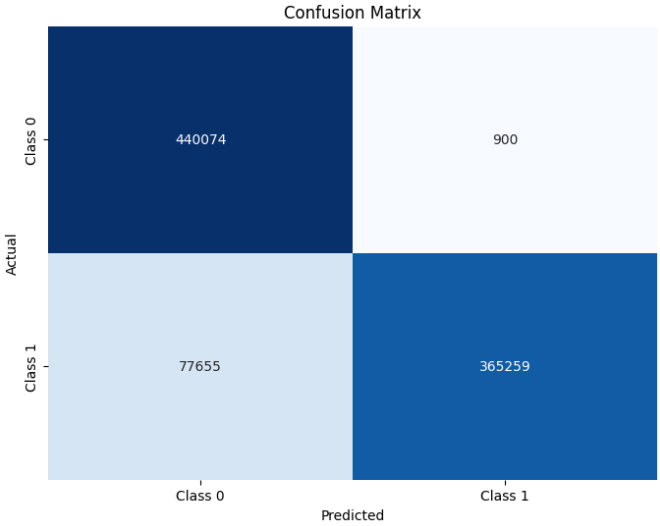

Fig. 8. Confusion Matrix:Decision Tree classifier

I. Random Forest classifier:

Accuracy of Random Forest classifier: 81.08%

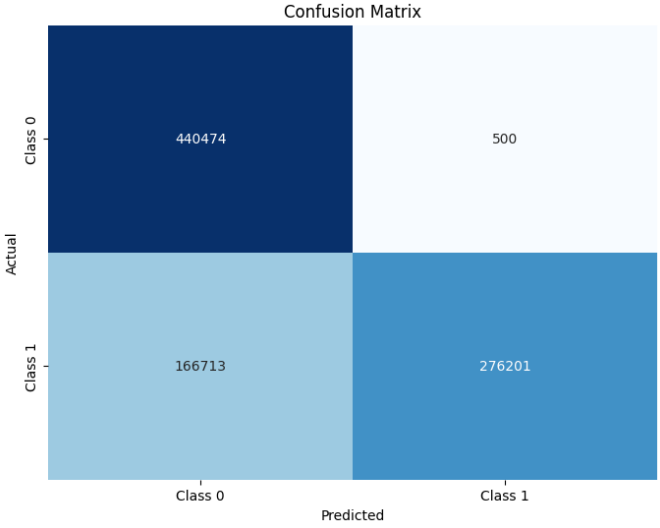

Fig. 9. Confusion Matrix:Random Forest classifier

J. Neural Network classifier:

Accuracy of Neural Network classifier: 99.08%

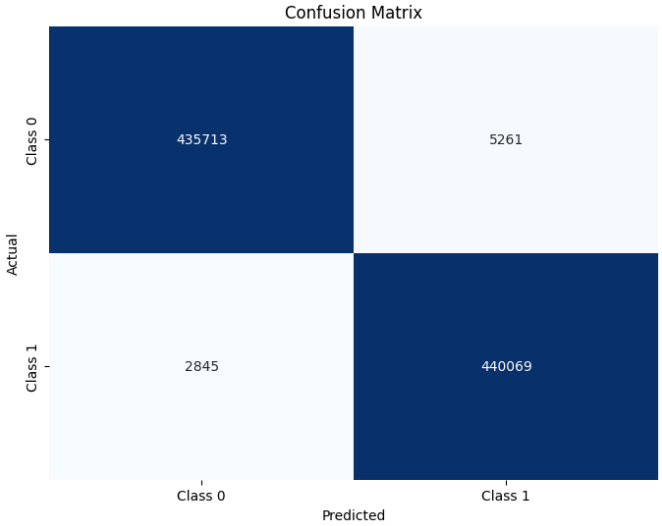

Fig. 10. Confusion Matrix:Neural Network classifier

V. Conclusion

This study is unique primarily because it focuses on mobile money fraud detection. From all the model results and confusion matrix, we can observe that KNN with 99.38% and Neural network Classifier with 99.08% did the best among all the models in the test data. Although machine learning (ML) is a useful tool for detecting fraud, it has some drawbacks, including the need for human intelligence and expert analysis, bias from flawed methodologies, and inaccurate predictions caused by non-representative datasets. Despite these, ML has a significant impact on fraud research, and with further development, it can be combined with human intelligence to detect financial crime more effectively and identify patterns of fraudulent behavior.
code: [2]

References

[1] https://www.kaggle.com/datasets/ealaxi/paysim1.
[2] https://github.com/riazzzz/Financial-informatics-project.
[3] Adeyinka Adedoyin, Stelios Kapetanakis, Georgios Samakovitis, and Miltos Petridis. Predicting fraud in mobile money transfer using case-based reasoning. pages 325–337, 11 2017.
[4] Rizik Al-Sayyed, Esra' Alhenawi, Hadeel Alazzam, Ala' Wrikat, and Dima Suleiman. Mobile money fraud detection using data analysis and visualization techniques. *Multimedia Tools and Applications*, pages 1–16, 2023.
[5] Noura Ahmed Al-Suwaidi and Haitham Nobanee. Anti-money laundering and anti-terrorism financing: a survey of the existing literature and a future research agenda. *Journal of Money Laundering Control*, 24(2):396–426, 2021.
[6] Siddharth Bhatore, Lalit Mohan, and Y Raghu Reddy. Machine learning techniques for credit risk evaluation: a systematic literature review. *Journal of Banking and Financial Technology*, 4:111–138, 2020.
[7] Francis Effirim Botchey, Zhen Qin, and Kwesi Hughes-Lartey. Mobile money fraud prediction—a cross-case analysis on the efficiency of support vector machines, gradient boosted decision trees, and naïve bayes algorithms. *Information*, 11(8):383, 2020.
[8] Ana Isabel Canhoto. Leveraging machine learning in the global fight against money laundering and terrorism financing: An affordances perspective. *Journal of business research*, 131:441–452, 2021.
[9] Jingguang Han, Yuyun Huang, Sha Liu, and Kieran Towey. Artificial intelligence for anti-money laundering: a review and extension. *Digital Finance*, 2(3-4):211–239, 2020.