

Online Transactions Fraud Detection using Machine Learning

Ms. Kishori Dhanaji Kadam, Ms. Mrunal Rajesh Omana,
Ms. Sakshi Sunil Neje, Ms. Shraddha Suresh Nandai.

^{1,2,3,4} Student, Sharad Institute of Technology College of Engineering, Ichalkaranji.

Date of Submission: 10-06-2023

Date of Acceptance: 20-06-2023

ABSTRACT: Now a days Digital transactions are rapidly increasing as it results in increasing online payment frauds too. In fact, according to the Reserve Bank of India, comparing March 2022 to March 2019, digital payments have risen in volume and value by 216% and 10%, respectively. People are starting to go all-in with digital transactions, but one can't deny the security issues that loom, and know-how when it comes to online payments. Few years ago, we could have barely seen the online payment, but today UPI payment QR code installed at doorstep. This invited the hoaxers and attackers to develop fraudulent transactions and fool people for some amount of money. Fortunately, the online transactions are monitored and hence could be analysed using the latest tools. In this system, an attempt is made to develop a machine learning model to identify fraudulent transactions in a transaction's dataset.

KEYWORDS: Transaction, Payment, UPI, Attackers, Fraudulent, Hoaxers, Money, Dataset.

I. INTRODUCTION

Mobile payment has gained significant popularity as a mainstream payment method, leading to a high volume of transactions on online trading platforms. Unfortunately, this popularity also attracts criminals who exploit the complex network environment to commit fraud. Such fraudulent activities not only harm consumers but also impede the healthy growth of the online economy. Consequently, effective transaction fraud detection becomes a vital tool in combating network transaction fraud.

Traditional fraud detection approaches primarily rely on statistical and multi-dimensional analysis techniques. However, these verification-based methods struggle to uncover the underlying patterns in transaction data, limiting their

effectiveness. On the other hand, big data technology and machine learning algorithms offer efficient solutions for detecting transaction fraud. Machine learning, particularly when applied to large datasets, can capture important features that traditional statistical methods fail to describe. By utilizing suitable machine learning techniques, we can build models based on existing transaction data to detect network transaction fraud, thereby mitigating associated losses.

In 2018, Zhaohui Zhang proposed a reconstructed feature convolutional neural network prediction model specifically tailored for transaction fraud detection. This model demonstrated improved stability and classification effectiveness compared to other convolutional neural network models. However, a challenge remains in achieving high detection accuracy due to imbalanced sample labels. To address this, the paper introduces two fraud detection algorithms: one based on a Fully Connected Neural Network and another utilizing XGBoost. The former algorithm integrates two neural network models with different cross-entropy loss functions, enabling a quick and convenient design process for the combined model. The latter algorithm leverages Hyperopt to optimize the XGBoost classifier, resulting in a fraud detection model with superior performance by selecting the best parameters. These two algorithms serve different application scenarios.

II. LITERATURE SURVEY

Different Supervised machine learning algorithms like Decision Trees, Naive Bayes Classification, Least Squares Regression, Logistic Regression and SVM are used to detect. Although real-time datasets help identify fraudulent transactions, there are remaining challenges when dealing with imbalanced data. The future efforts

will concentrate on addressing the problem mentioned earlier. The algorithm of the random forest itself should be improved. Multiple Supervised and Semi-Supervised machine learning techniques are used for fraud detection, but we aim is to overcome three main challenges with card frauds related dataset. i.e.

1. Strong class imbalance
2. The inclusion of labelled and unlabelled samples, and
3. To increase the ability to process many transactions.

III. METHODOLOGY

Initially, we apply a clustering technique to categorize cardholders into distinct groups (high, medium, low) based on their transaction amounts. This is accomplished by utilizing range partitioning. Subsequently, using the Sliding-Window method, we combine the transactions within each group and extract specific features from the window to identify patterns in cardholders' behaviour. These features include the maximum and minimum transaction amounts, the average amount within the window, and the duration of time elapsed.

Algorithm 1: Sliding Window-Based Aggregated Transaction Details and Cardholder Feature Extraction

Input:

- Customer's cardholder ID
- Sequence of transactions, denoted as t
- Window size, denoted as w

Output:

- Aggregated transaction details
- Cardholder features indicating genuine or fraud

Steps:

1. Initialize an empty list to store aggregated transaction details.
2. Initialize an empty list to store cardholder features.

3. Set the start index, denoted as $start_idx$, to 0.
4. Set the end index, denoted as end_idx , to w .
5. While end_idx is less than or equal to the length of transaction sequence t , do:
 - a) Retrieve the transactions within the current window, from $start_idx$ to end_idx .
 - b) Calculate and store the aggregated details, such as the maximum transaction amount, minimum transaction amount, average amount, and time elapsed, based on the transactions within the window.
 - c) Extract cardholder features from the window, considering the aggregated details and any other relevant information.
 - d) Append the aggregated transaction details and cardholder features to their respective lists.
 - e) Move the window by incrementing both $start_idx$ and end_idx by 1.
6. Return the aggregated transaction details and cardholder features.

Algorithm 2: Classifier Rating Score Update for Model Accuracy Assessment.

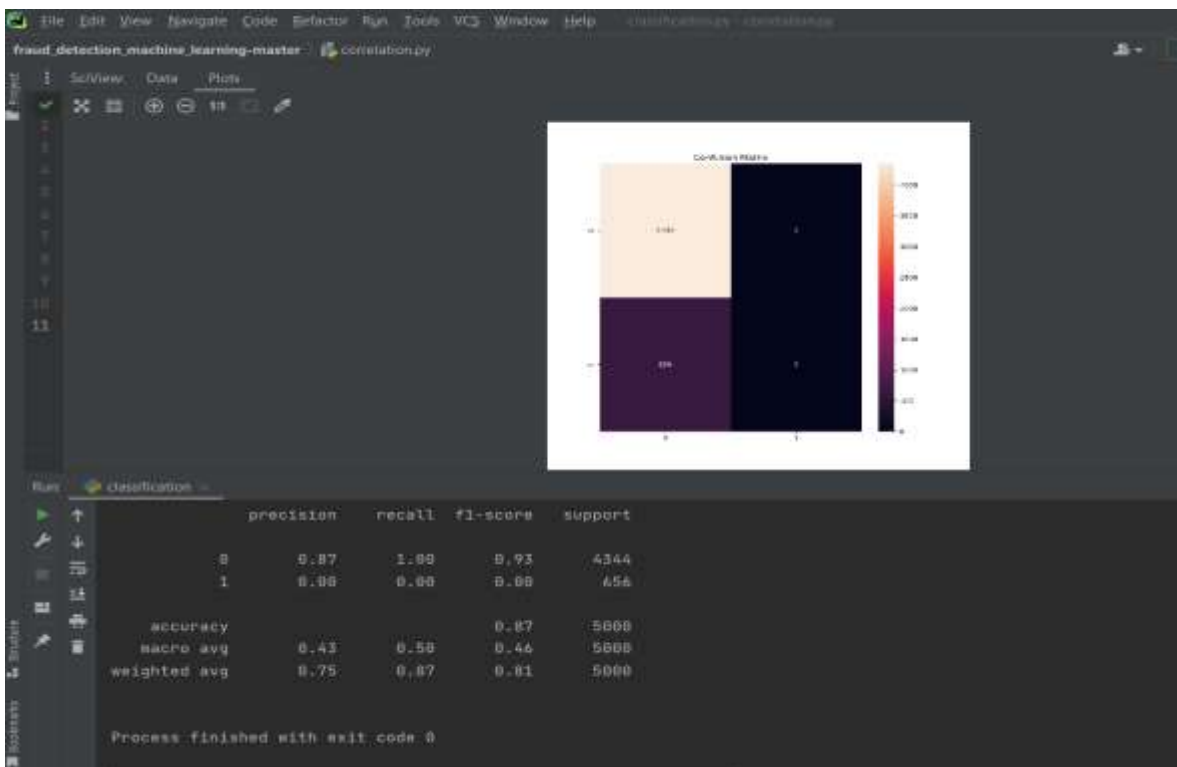
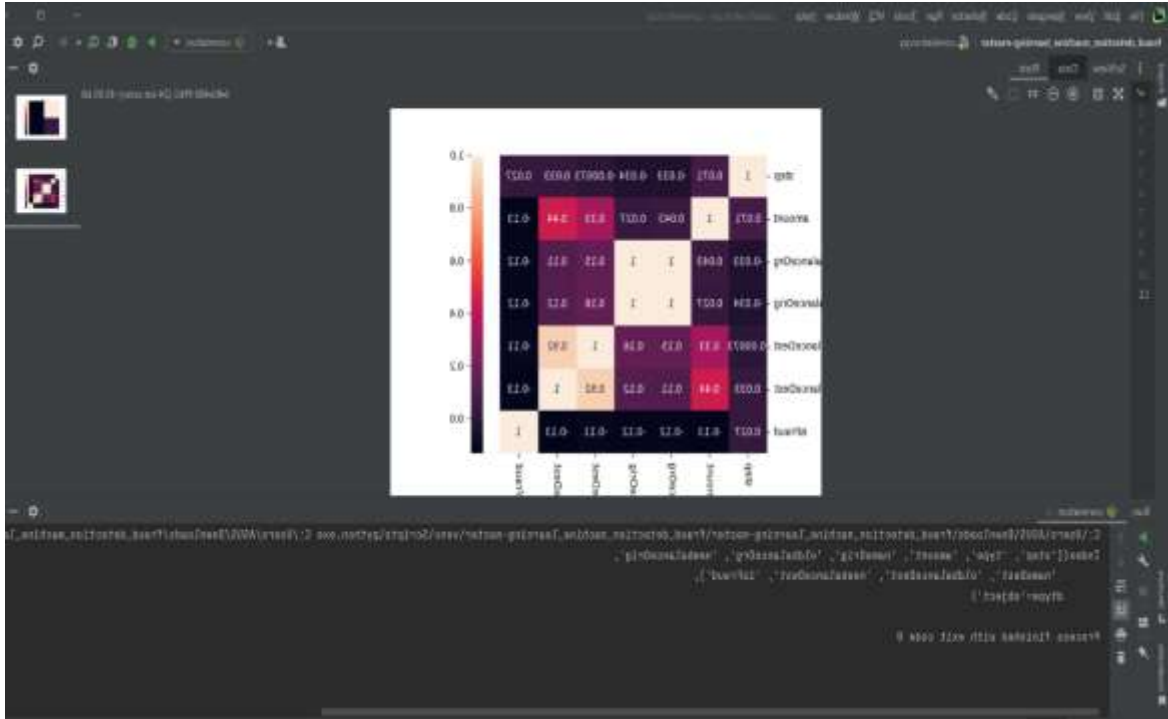
Input:

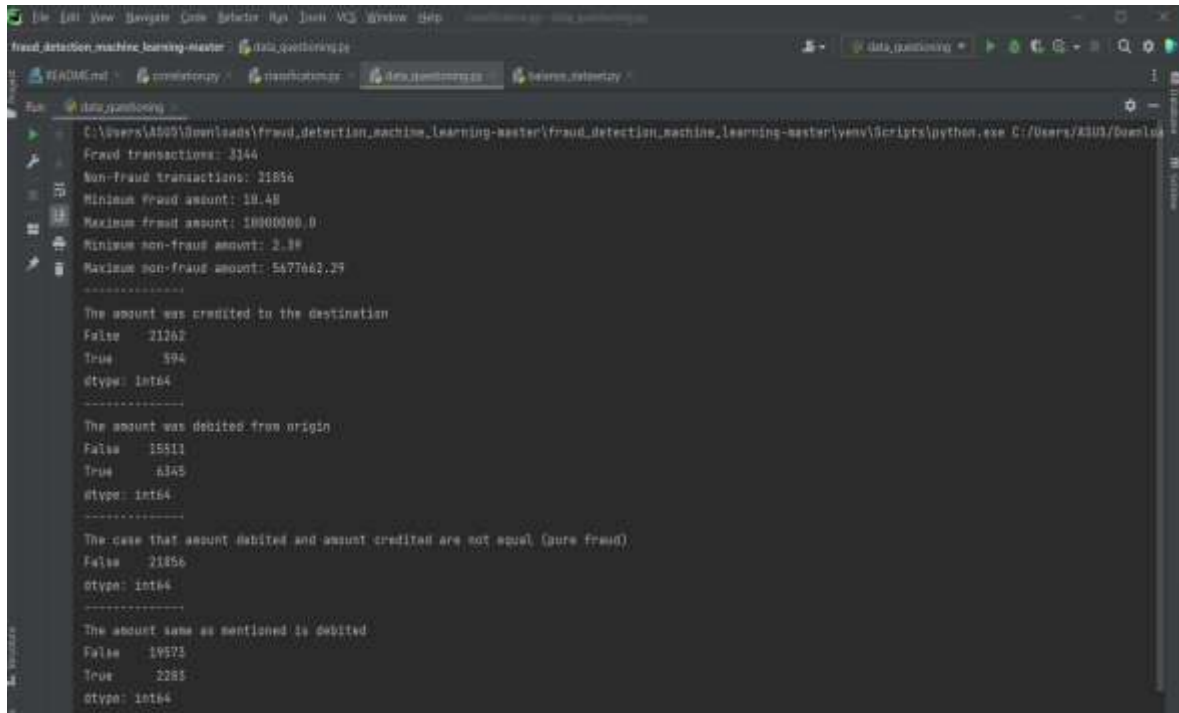
- Cardholder ID
- Previous transaction and current transaction

Output:

- Rating score of the model after each transaction
1. Initialize the rating score of the classifier to 0.
 2. For each new transaction received: a. Update the classifier with the previous transaction and the current transaction. b. Evaluate the performance of the classifier by comparing its prediction for the current transaction with the actual label. c. Adjust the rating score based on the accuracy of the classifier's prediction. Increase the score if the prediction is correct or decrease it otherwise. d. Update the previous transaction to be the current transaction for the next iteration.
 3. Return the final rating score of the classifier.

IV. RESULT





```
fraud_detection_machine_learning-master | data_quantization | data_quantization | release_release |  
C:\Users\ASU3\Downloads\fraud_detection_machine_learning-master\fraud_detection_machine_learning-master\env\Scripts\python.exe C:/Users/ASU3/Downloa  
Fraud transactions: 3144  
Non-Fraud transactions: 21856  
Minimum fraud amount: 19.48  
Maximum fraud amount: 1000000.0  
Minimum non-fraud amount: 2.38  
Maximum non-fraud amount: 5677662.79  
-----  
The amount was credited to the destination  
False    21262  
True     894  
dtype: int64  
-----  
The amount was debited from origin  
False    19511  
True     6345  
dtype: int64  
-----  
The case that amount debited and amount credited are not equal. (pure fraud)  
False    21856  
dtype: int64  
-----  
The amount same as mentioned is debited  
False    19573  
True     2281  
dtype: int64
```

V. CONCLUSION

In this system we developed a novel method for fraud detection, where customers are grouped based on their transactions. We finally observed that Logistic regression, decision tree and random forest are the algorithms that gave better results.

REFERENCES

- [1]. Mohammed, Emad, and Behrouz Far. "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study." *IEEE Annals of the History of Computing*, IEEE, 1 July 2018, doi.ieeecomputersociety.org/10.1109/IRI.2018.00025.
- [2]. Randhawa, Kuldeep, et al. "Credit Card Fraud Detection Using AdaBoost and Majority Voting." *IEEE Access*, vol. 6, 2018, pp. 14277–14284., doi:10.1109/access.2018.2806420. doi:10.1109/colcomcon.2017.8088206.
- [3]. <http://www.rbi.org.in/Circular/CreditCard>
- [4]. <https://www.ftc.gov/news-events/press-releases/2019/02/imposter-scams-top-complaints-made-ftc-2018>