IJISEM
INTERNATIONAL JOURNAL OF INNOVATIONS IN SCIENCE ENGINEERING AND MANAGEMENT

OPEN ACCESS

Volume: 4

Issue: Special Issue 1

Month: April

Year: 2025

ISSN: 2583-7117

Citation:

Dr. Rakesh Kumar Pathak, Dr. Prakash Upadhyay "An Anomaly and Fraud Detection Model in Online Financial Transactions: A Machine Learning Approach" International Journal of Innovations in Science Engineering and Management, vol. 4, no. Special Issue 1, 2025, pp. 103–108.

DOI:

10.69968/ijisem.2025v4si1103-108

An Anomaly and Fraud Detection Model in Online Financial Transactions: A Machine Learning Approach

Dr. Rakesh Kumar Pathak[1], Dr. Prakash Upadhyay[2]

[1]*Assistant Professor, Department of Computer Science, St. Xavier's College of Management & Technology, Patna.*
[2]*Assistant Professor, Department of Computer Science, St. Xavier's College of Management & Technology, Patna.*

Abstract

Indian economy has gone through a massive transformation in an unprecedented pace in the last decade. Nobody has ever imagined that a developing and growing economy like India can beat the western world in the digitalization of economy. A larger chunk of monetary transactions are being carried out digitally in India. People and institutions are using digital technology such as UPI, Net Banking, IMPS and RTGS. The use of digital technology in financial transactions has become a new normal in Indian economy. The worrying part of financial transactions carried over the internet are the events of frauds and cheatings. Both the people and businesses are equally concerned about frauds in financial transactions as both the parties are victims of fraudulent activities.

The perpetrators of financial frauds are not untrained criminals, rather they are highly skilled and they make use of sophisticated techniques. As the fraudster are getting smarter day by day, conventional methods of catching them often fall short. This paper explores various machine learning algorithms to detect anomalies and frauds in financial transactions and suggest some means to deal with them.

Keywords; Financial Transaction, Anomaly, Fraud, Artificial Intelligence, Machine Learning, Deep Learning.

INTRODUCTION

Financial transactions have always been an integral part of every society ever since business or trade has existed on this planet. However the methods of financial transactions have evolved over the years. From exchange of articles to precious coins, coins to paper money, from paper money to cards or plastic money and from cards to virtual money or e cash. Use of virtual money or online transactions have changed the entire eco system of trade and business. The use of online transactions have become a new normal of modern society. Our own country India has developed UPI or unified payment interface which has been accepted by the masses here. The entire world is mesmerised on its success. Although India is a developing country but it has been a pioneer in this regard. Even the Europeans and Americans are amazed with its ease of use and seamless integration with all sorts of payment platforms and across businesses, no matter how big or small their sizes are. One can pay online using debit or credit card to a merchant be it a tea seller or a 5 star hotel or a giant hospital. This user friendliness of online transaction platform is almost making India a cashless economy. Every day a whopping 80 thousand crore transaction are taking place in India through UPI platforms without any hustle and almost cent percent success rate. The transaction gets settled within a blink of an eye and both the payer as well as the payee getting intimation of the same almost instantly.

Every technology has two sides. Technology in itself is not bad, however if it is used with a negative intentions, it can cause damage. The same is true for online transactions too.

Frauds and cheatings are the ill effects or bad uses of online payment technology. Fraudulent activities in the trade and commerce arises from countless sources. Some of the most common avenues of frauds in financial transactions are ecommerce, credit cards, UPI payments, clone cheques, phone banking and online banking are the few examples. According to [1], various stakeholders have lost almost a whopping figure of $ 200 billion from 2021 to 2025 in the fraudulent activities while performing transactions such as ecommerce, ticket booking, money transfer and similar other activities. The people involved in online frauds are usually not lay criminals, rather they are highly skilled and technically qualified people who know how to misuse sophisticated technology. They rarely leave any clue or footprints behind them and hence it is very difficult to catch them. The authors of [2], in their paper have discussed the techniques used by the fraudsters to steal user's personnel credentials to carry out fraudulent activities. In this the fraudsters try to steel customer's personnel credentials by hacking or any other means and use these stolen credentials to carry forward the fraudulent activities. This act of impersonation is really very challenging to combat. However the activities of the fraudsters shows certain anomalous feature that can be detected by the fraud detection processes. Some of the signs of suspicious behaviour includes an abnormally bigger amount of the transaction, a payee of the transaction does not matches with the jurisdiction of the account holder life context. The assumption is that the through machine learning algorithms such anomalies in the financial transaction can be detected.

FRAUDS IN ONLINE TRANSACTIONS: AN OVERVIEW

Gone are days when online payment services were used only by high class people or so called technologically advanced people. In the present era, it has become a technology which is being used by the masses. The sear volume and velocity of online transactions are simply unprecedented. With this increasing volume and value of online transactions, threats of fraud on this mode of transactions is also increasing every moment. Fraud prevention in online transaction is only possible if we are able to track the discrepancies that the fraudulent transactions show. However the volume and velocity of transactions over the internet makes it very difficult for the lay people to detect and prevent such transactions within very little response time. However of if machine learning or other similar techniques are used then not only detection of fraudulent activities becomes possible but also their prevention too becomes achievable. To identify online fraudulent transactions, researches in future may combine ML algorithms with various I/O consideration configurations[3]. The most popular delivery strategies are balanced ones. Three categories comprise the most popular suggested solutions: synthesis, algorithm, and data layer solutions. Class imbalance may suffer if pre-processing is shortened as a resample to apply data-level solutions. The algorithmic level solution is to modify the learning biases of the current algorithms or develop new ones for the minority class. [4]. One of the biggest challenge in the development of fraud identification model is the lack of real wold data set. As nobody want to reveal or even share with anybody their credentials, therefore a true representative data set is missing. ML algorithms are being trained on simulated data set. These simulated data set cannot be used to train a real world model of a machine learning system that can detect fraud in a real world scenario having tremendous volume of transactions with an unbelievable high speed. One of the most common mistakes in the ML models trained on simulated data is in their wrong predictions both with positive and negative outcomes. In simple word the transaction that the system labels as True Positive (the transaction is a fraud) or false negative (those cases where the system detects that is fraud, which in reality is not (false positive) or false negative which means that the tested condition as per the system is not a fraud, which in reality it is. The challenge that the previous studies in the similar domain has faced was the lack of availability of genuine data set. Because of this most of the past studies have utilised artificially created datasets. The introduction of the PaySim financial simulator, which imitates usual transactions done using handheld instruments and introduces dubious activities to intensify the occurrence of frauds, is a notable advancement [5].

Types of fraudulent activities

By every passing day, perpetrators of online frauds are enhancing themselves and are using more and more improvised means and they are coming up with novel techniques of carrying out fraudulent activities, however most of these activities include use of plastic money, i.e. some type of caard. Following are some of the most common methods of carrying out fraudulent activities

- **Lost or missing card**

Fraudsters might use stolen or lost credit cards to make illegal purchases. Before the cardholder notices the card is gone and reports it, they might use it at actual establishments like restaurants, shops, or other establishments. [9]

- **Counterfeit Cards**

By copying the data from real credit cards, scammers produce fake credit cards. Then, fraudulent cards are used for in-person transactions, frequently at ATMs or retail establishments. [9]

- **Card Impersonation**

In card impersonation fraud, a fraudster presents a real card with the cardholder's information on it in order to pose as the actual cardholder. This may occur when a thief obtains a real card or when they mix information from a stolen card with a forged identity document. [9]

- **Touchless Transaction**

These days various cards are coming with wi-fi enabled feature which allows transaction to be carried out without having to physically swiping the card. Fraudsters are using handheld POS devices and if the cardholder is not alert, their card can be used to carry out the transaction. [9]

These examples are to name a few. It is quite clear that fraudsters are carrying out their activities both in online as well as offline mode.

ANOMALY BASED FRAUD DETECTION

Online transactions are essential in today's world. Online transactions encompass any transaction that uses an internet-based payment system, not just those conducted on the webpages. Any POS, UPI, or swipe machine can be used. Whenever we use any such payment platform, we are exposing our vital accounting information or card details or UPI credentials over the internet. There is no way out of this situation. Nonetheless, the payment gateways are making every effort to shield their subscribers from fraudulent activity of any kind. Here, the issue is how to determine whether a transaction is being carried out legitimately or fraudulently. The explanation is that there is a difference between a fraudulent transaction and a genuine one. The fraudulent transaction exhibits some unusual characteristics.

An anomaly can be defined in simple words as some irregularity or abnormality in the properties of an activity such as a financial transaction. To detect fraud in any financial transaction algorithms can be developed that can identify any slight deviation or anomaly in the properties of transaction and can alert the user to take protective measure or it can itself initiate the protective measures.

An anomaly based fraud detection model is an approach of detection of fraud in any financial transaction that tries to find out unusual pattern in the transaction by comparing the properties of an ongoing transaction with the properties of a transaction which is labelled as normal or safe transaction. To do this the machine has to be trained with a substantial number of transaction's data. The size of data set used to train the model plays a significant role in the accuracy of the model [8]. The system keeps learning and refining itself gradually. The longer the learning period the better the accuracy of the system will be.

METHODOLOGY

Fraud detection in online transaction is a difficult and challenging task, not because the fraudsters are very clever or technically well-equipped but because the whole scenario is very dynamic. The sheer volume and velocity of the transactions is enormous and besides this the methodology adopted by the fraudsters is getting upgraded every moment. The biggest challenge in coming up this kind of model is that any real dataset was not at all available. None of the financial or banking house will ever be agree to share their customer's sensitive data such as the card number, IDs or passwords with anyone because they are the ones who have to protect and secure their customer's privacy. To other extreme any model to detect attempts of frauds is not possible unles the model is trained on reliable data set. The situation is really tricky and paradoxical. Simulation of financial transaction in an online environment came as a ground breaking solution. Public datasets like PaySim, Kaggle Credit Card Fraud Detection, and IEEE-CIS Fraud Detection data set are there to help researchers and developers to build machine learning model and train and test the model using these datasets.

Data Collection

In the proposed system we have used PaySim data set. The data set contains around 1050000 records with 11 attributes or features. The dataset has the following features

- Transaction amount
- Transaction time
- Merchant category
- User behaviour (e.g., location, device, IP)
- Historical fraud records

Data Pre-processing

Before using the collected data for the training of ML model, pre-processing on the data needed to be performed. In a broad sense, following pre-processing was performed

- **Handle Missing Values:** imputation techniques is used to tackle the missing data.

- **Feature Engineering** has been used to convert categorical data.
- **Scaling & Normalization** is used to standardize transaction amounts and times.
- **Handling Imbalanced Data:** for this purpose SMOTE (Synthetic Minority Over-sampling Technique) or undersampling is used.

The Anomaly based Fraud Detection in online transaction using machine learning model has been implemented using python programming language. We have used Google colab platform to implement this model. The model has been implemented and tested in the following steps

a. First our google drive was imported and mounted on the Google colab platform

b. Necessary libraries were imported to help implement our model

c. After importing google drive and importing the libraries, we connected our data set and performed data pre processing

d. After finishing data pre-processing, we performed splitting of our data set into training and testing sub divisions.

e. Once the model got trained we have used various visualization techniques to showcase finding of this model and also used various parameters to test the accuracy of model's findings.

FINDINGS AND ANALYSIS

To begin with the proposed model performed the classification of data in the data set into two categories or classes, i.e. Normal or legitimate transactions and fraudulent transactions

Class distribution:

	Class
0	284315
1	492

A confusion matrix was produced by the model along with a classification report

Confusion Matrix:

$$\begin{bmatrix}[56862 & 2 \\ 23 & 75]\end{bmatrix}$$

Table 1 Classification Report:

	Precision	recall	f1-score	support
0	1.00	1.00	1.00	56864
1	0.97	0.77	0.86	98
Accuracy			1.00	56962
Macro avg	0.99	0.88	0.93	56962
Weighted avg	1.00	1.00	1.00	56962

A confusion matrix is utilised to summarize the outcomes of a classification algorithms. For binary classification, as in our model's case, the dimension of the confusion matrix is 2X2. For binary classification, the model predicts that each instance of the dataset belongs to one class or the other[6]. The performance matrix evaluated the similarity or differences between the real verses the predictions. For this it uses following evaluation methodologies [7].

True Positive

True Negative

False Positive

False Negative

The performance matrix or the confusion matrix uses following four parameters to evaluate the prediction models

- **Accuracy:** It is the ratio of correct prediction with that of total prediction. It is very simple technique of measuring a models accuracy. Formula to calculate accuracy is
  Accuracy= correct prediction / (Total correct + Total incorrect) predictions
  Accuracy = (TP + TN) / (TP + FP + TN + FN)
- **Precision Score:** It is the measure of proportion of positively predicted labels that are in reality true. It is a useful measure of success of prediction when the classes have imbalance.
  Precision Score=Total Positive / (TP + FP)
- **Recall Score:** Recall score is the ratio of predicted positive value and the actual total positive values. It calculates the ratio of predicted positive values to that of actual positive values plus those positive values in the dataset that are wrongly predicted as negative values. It formula is
  Recall= True Positive / (True Positive + False Negative)
  Recall= TP / (TP + FN)

- **F1 Score:** F1 score is actually the harmonic mean of the Precision and the recall scores. It is calculated using the following formula
  Recall= 2 * (Recall * Precision) / (Recall + Precision)
  F1 Score is also called F-beta score and the F-beta score reaches its best value when the beta is 1, and worst score when beta value=0.
- **Support:** it indicates total number of true occurrences of each class. is the sum of total true instance of a label.

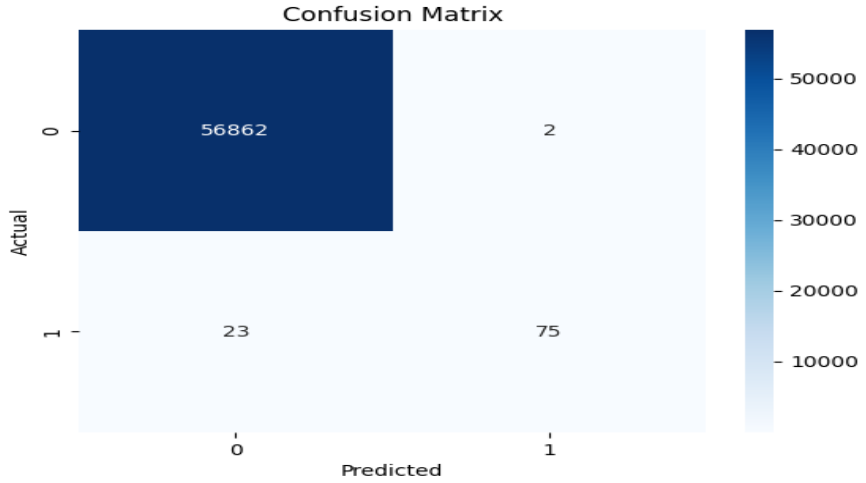

Figure 1 Confusion Metrix

After producing confusion matrix, our model produced the box plot to further elaborate the classification of data

Axes: xlabel='Class', ylabel='Amount'

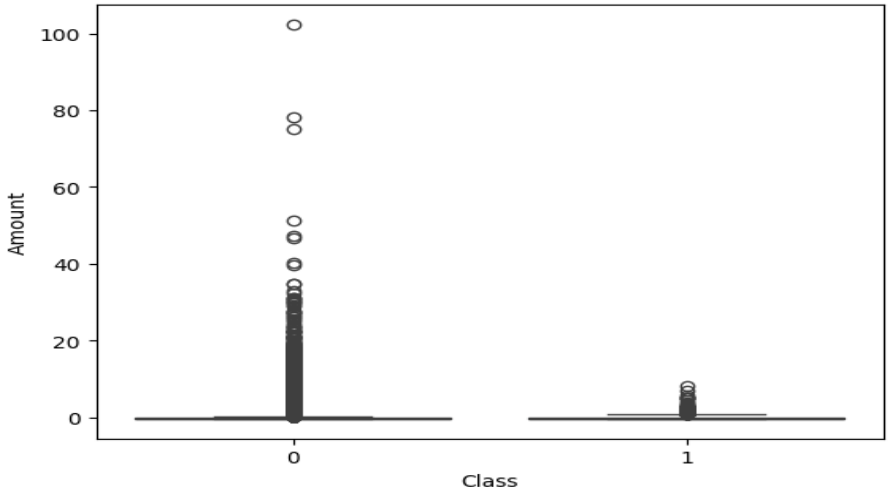

Figure 2 Boxplot

CONCLUSIONS

The task of fraud detection in online financial transaction is very challenging. The perpetrators of these types of crime are highly skilled and technically qualified. Besides these they can be living in any part of the worlds and can be carrying out their activities anywhere they want. These people are looking for weak or vulnerable networks where they can easily penetrate and keep hunting pray. The problem of jurisdiction of security agencies is the biggest hurdle in dealing with such offences. Our model is an attempt to help curb these offences by using sophisticated technology. Such crimes are committed online therefore they should be best dealt with online tools. The model

proposed in this paper tries to identify fraudulent activities on the basis of its feature or by finding anomalies in the transactions. However the algorithm needs to continuously upgrade itself as the perpetrators are not using stagnant technologies, rather they are upgrading or enhancing themselves much faster.

REFERENCES

[1] *Juniper Research* (2020) Online payment fraud: Emerging threats, segment analysis and market forecasts 2021-2025. www. Juniperresearch. Com

[2] Amiri M, Hekmat S (2021) Banking fraud: a customer-side overview of categories and frameworks of detection and prevention. *J Appl Intell Syst Inf Sci* 2(2):58–68

[3] U.Siddaiah, P. Anjaneyulu, Y. Haritha, M Ramesh. 2023 7th *International Conference on Intelligent Computing and Control Systems* (ICICCS) 2023 IEEE|DOI:10.1109/ICICCS56967.2023.10142404

[4] Yeming Chen, Xinyuan Han Clarity AI Beijing, China. 2021 IEEE *International Conference on Consumer Electronics and Computer Engineering* (ICCECE) 2021 IEEE | DOI: 10.1109/ICCECE51280.2021.9342475.

[5] Petr Hajek, Mohammad Zoynul Abedin, Uthayasankar Sivarajah.

[6] Kumar, A. (2023, March 17). *Data Analytics - AI, Data, Data Science, Machine Learning, Blockchain, Digital.* Retrieved from vitalflux.com: https://vitalflux.com/accuracy-precision-recall-f1-score-python-example/

[7] sklearn.matrics.*precision_recall_fscore_support*. (2023). Retrieved from scikit-learn.org:https://scikitlearn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html

[8] Durgesh Kumar Mishra, Nilanjan Dey, Bharat Singh Deora, Amit Joshi. "*ICT for Competitive Strategies"*, CRC Press, 2020

[9] Sumedh N. Pundkar, Mohd Zubei. "*Credit Card Fraud Detection Methods: A Review*", E3S Web of Conferences, 2023

[10] Alok Baran 2024. Concept of E-Currency: A Broader View on A Wider Canvas. *International Journal of Innovations in Science, Engineering And Management*. 2, (May 2024), 108–114.