IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Upi Fraud Detection Using Machine Learning

Amrapali Santosh Gayakwad
Post Graduate Student, Department of Data Science, Zeal College of Engineering and Research, Pune, Maharashtra, India

ABSTACRT: forest algorithm and the Adaboost algorithm. The results of the two algorithms are based on accuracy, precision, recall, and F1-score. The ROC curve is plotted based on the confusion matrix. The Random Forest and the Adaboost algorithms are compared and the algorithm that has the greatest accuracy, precision, recall, and F1-score is considered as the best algorithm that is used to detect the fraud. UPI fraud detection is presently the most frequently occurring problem in the present world. This is due to the rise in both online transactions and e-commerce platforms. UPIfraud generally happens when the card was stolen for any of the unauthorized purposes or even when the fraudster uses the UPI information for his use. In the present world, we are facing a lot of UPI problems. To detect the fraudulent activities the UPIfraud detection system was introduced. This project aims to focus mainly on machine learning algorithms. The algorithms used are random.

Keywords
UPI, Fraud Detection, Machine Learning, Anomaly Detection, Financial Security, Real-time Analysis

I. INTRODUCTION

1.1 OVERVIEW

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 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 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 Hyper opt 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.

1.2. Fraud Detection

We are living in a world which is rapidly adopting digital payments systems. UPI and payments companies are experiencing a very rapid growth in their transaction volume. In third quarter of 2018, PayPal Inc (a San Jose based payments company) processed 143 billion USD in total payment volume [4]. Along with this transformation, there is also a rapid increase in financial fraud that happens in these payment systems. An effective fraud detection system should be able to detect fraudulent transactions with high accuracy and efficiency. While it is necessary to prevent bad actors from executing fraudulent transactions, it is also very critical to ensure genuine users are not prevented from accessing the payments system. A large number of false positives may translate into bad customer experience and may lead customers to take their business elsewhere. A major challenge in applying ML to fraud detection is presence of highly imbalanced data sets. In many available datasets, majority of transactions are genuine with an extremely small percentage of fraudulent ones.

1.3 Problem Statement

Machine Learning is revolutionizing the way organizations detect and prevent fraud. From utilizing advanced fraud detection algorithms to detect suspicious activity to preventing hackers from accessing sensitive data, Machine Learning is making a huge impact on organizations' fraud prevention efforts.

In this blog, we will explore the benefits, limitations, and use cases of machine learning for fraud detection. We will delve into how machine learning can empower businesses to detect and prevent fraudulent activities, highlighting the role of unsupervised learning algorithms and the expertise of ML engineers. Additionally, we will discuss the impact of machine learning on fraud prevention strategies and the advantages of adopting this technology.

II. LITERATURE SURVEY

1. ONLINE TRANSACTIONS FRAUD DETECTION USING MACHINE LEARNING

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.

2. Fraud Detection using Machine Learning

Recent research has shown that machine learning techniques have been applied very effectively to the problem of payments related fraud detection. Such ML based techniques have the potential to evolve and detect previously unseen patterns of fraud. In this paper, we apply multiple ML techniques based on Logistic regression and Support Vector Machine to the problem of payments fraud detection using a labeled dataset containing payment transactions. We show that our proposed approaches are able to detect fraud transactions with high accuracy and reasonably low number of false positives.

3. Predictive-Analysis-based Machine Learning Model for Fraud Detection with Boosting Classifiers

Fraud detection for credit/debit card, loan defaulters and similar types is achievable with the assistance of Machine Learning (ML) algorithms as they are well capable of learning from previous fraud trends or historical data and spot them in current or future transactions. Fraudulent cases are scant in the comparison of non-fraudulent observations, almost in all the datasets. In such cases detecting fraudulent transaction are quite difficult. The most effective way to prevent loan default is to identify non-performing loans as soon as possible. Machine learning algorithms are coming into sight as adept at handling such data with enough computing influence. In this paper, the rendering of different machine learning algorithms such as Decision Tree, Random Forest, linear regression, and Gradient Boosting method are compared for detection and prediction of fraud cases using loan fraudulent manifestations. Further model accuracy metric have been performed with confusion matrix and calculation of accuracy, precision, recall and F-1 score along with Receiver Operating Characteristic (ROC)curves.

4. Fraud Detection Using Machine Learning and Deep Learning

Fraud detection is a critical task in various industries, aiming to identify and prevent fraudulent activities. In this report, we explore different machine learning models and the impact of applying Borderline SMOTE, a data augmentation technique, on their performance. We evaluate the precision, recall, and area under the precision-recall curve (AUPRC) metrics before and after applying Borderline SMOTE. Our findings indicate that Borderline SMOTE does not improve the results significantly for fraud detection in this dataset. Despite the scarcity of fraud instances, the generated synthetic data introduces noise and adversely affects most models' performance. Decision Trees and Random Forest, leveraging the inherent nature of fraud occurrences as closely related and rare events, outperform other models in this scenario. Logistic Regression, Support Vector Machines (SVM), Adaboost, and Neural Networks show limitations in effectively capturing the intricacies of fraud patterns in this dataset.

III. PROJECT DECRIPTION

EXISTING SYSTEM

- To detect counterfeit transactions, three machine-learning algorithms were presented and implemented.
- There are many measures used to evaluate the performance of classifiers or predictors, such as the Gradient Boost Classifier, Vector Machine, Random Forest, and Decision Tree.
- These metrics are either prevalence dependent or prevalence-independent.
- Furthermore, these techniques are used in UPI fraud detection mechanisms, and the results of these algorithms have been compared.

Disadvantages

- 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.

IV. PROPOSED SYSTEM

- Various modern techniques like artificial neural network
- Different machine learning algorithms are compared, including Auto Encoder, Local Outlier Factor, Kmeans Clustering.
- This project uses various algorithm, and neural network which comprises of techniques for finding optimal solution for the problem and implicitly generating the result of the fraudulent transaction.
- The main aim is to detect the fraudulent transaction and to develop a method of generating test data.
- This algorithm is a heuristic approach used to solve high complexity computational problems.
- The implementation of an efficient fraud detection system is imperative for all UPI issuing companies and their clients to minimize their losses.

ADVANTAGES

- A typical organization loses an estimated 5% of its yearly revenue to fraud. In this course, learn to fight fraud by using data. Apply supervised learning algorithms to detect fraudulent behavior based upon past fraud, and use unsupervised learning methods to discover new types of fraud activities.
- Fraudulent transactions are rare compared to the norm. As such, learn to properly classify imbalanced datasets.
- The course provides technical and theoretical insights and demonstrates how to implement fraud detection models. Finally, get tips and advice from real-life experience to help prevent common mistakes in fraud analytics.

V. SYSTEM CONFIGURATION

H/W SYSTEM CONFIGURATION:-

- Processor – Intel core2 Duo
- Speed - 2.93 Ghz
- RAM – 2GB RAM
- Hard Disk - 500 GB
- Key Board - Standard Windows Keyboard
- Mouse - Two or Three Button Mouse
- Monitor – LED

S/W SYSTEM CONFIGURATION:-

- Operating System: XP and windows 7
- Python coding
- s/w –googlecolabs,or spider

VI. MODULES DESCRIPTION

DATASET AND ANALYSIS

In this project, we have used a Kaggle provided dataset of simulated mobile based payment transactions. We analyze this data by categorizing it with respect to different types of transactions it contains. We also perform PCA - Principal Component Analysis - to visualize the variability of data in two dimensional space. The dataset contains five categories of transactions labeled as ’CASH IN’, ’CASH OUT’, ’DEBIT’, ’TRANSFER’ and ’PAYMENT’ - details are provided

UPIFRAUD

Credits are typically used to refer to electronic financial transactions made without the use of physical cash. A UPI that is extensively used for online transactions is a small piece made up of thin plastic material with credit services and customer details. Fraudsters use credit cards to make unlawful transactions that result in massive losses to banks and card holders. Moreover, the invention of counterfeit cards has aided fraudsters in performing illicit transactions more easily. In general, it is regarded as illegitimate to use the card without the proper owners’ authorization. By obtaining access to a certain account illegitimately, any transaction that is carried out is considered as fraudulent. UPI fraudulent activities can be divided into two aspects, namely, offline and online fraud. In offline fraudulent activity, the fraudsters conduct their illicit transactions with stolen credit cards such as genuine card holders, while online fraudsters conduct their activities in online transactions through Internet Online fraud.

FINANCIAL STATEMENT FRAUD

Fraud in financial statements involves forging financial reports to claim that a company is more profitable than usual, avoid the payment of taxes, increasing stock prices, or obtaining a bank loan. It can also be regarded as the confidential records generated by organizations that contain their financial records that comprise their expenses, profits made, income loans, etc. These statements also comprise some write-ups made by management for discussing business performances and predicted future tendencies.

INSURANCE FRAUD

Insurance fraud can be defined as the act of misusing an insurance policy for gaining illegitimate benefits from an insurance business. Usually, insurance is made to protect the organization’s transactions or individual’s transactions against any financial risks. The main sectors of target by fraudulent insurance claims include healthcare and automobile insurance companies although home and crop insurance fraudulent also occur, however, there is a paucity of the literature on both. It has been estimated recently that the total cost of insurance fraud in the United States is over a billion USD yearly and it is finally passed on to consumers in the form of higher insurance premiums.

In order to cover the relevant costs of theft or accidental damages to a car, an agreement between the insurance provider and the insured person or organization is typically involved in automobile insurance claims. Individual fraudsters are capable of committing fraudulent claims, and one method of committing fraud is through deception during the claims process. Evidence of organized groups working together to conduct insurance fraud also exists.

VII. CONCLUSION

Online transaction are increasing with UPI, cards, net banking as it is easier and saves time of the customer purchase. But as the online transaction is increasing so is the fraud. Hence using machine learning algorithms and ANN model, an attempt has been made to gain the knowledge about the fraud and genuine transaction. With the increasing usage of ecommerce and online shopping, customer's vital information like card's CVV, UPI pin, OTP, passwords etc. are always vulnerable. The reason for rise in fraud cases is that fraudsters easily barge into the customers private information. Even the banking systems are vulnerable to frauds, also the fraudsters are finding new ways of fraud like identity theft from social media platform so the fraud prevention and detection is an ever evolving process. As the online frauds are ever increasing, the traditional way of detecting and preventing frauds are being replaced by various machine learning algorithms.

REFERENCES

1. Elisa Indriasari; Ford LumbanGaol; Tokuro Matsuo, "Digital Banking Transformation: Application of Artificial Intelligence and Big Data Analytics forLeveraging Customer Experience in the Indonesia Banking Sector",presented at 2019 8th International Congress on Advanced Applied Informatics(IIAI-AAI) ,Toyanama, Japan,7-11 Year: 2019, pp: 863-868.

2. Dr. Navleen Kaur, Ms. SupriyaLambaSahdev, "Banking 4.0:the influence of artificial intelligence on the banking industry & how AIis changing the face of modern day banks",International Journal of Management (IJM) Volume 11, Issue 6, June 2020,pp 39-45.
3. Tushar Gupta; Naman Gupta; Ankit Agrawal; Aksh Agrawal; KartikKansal, "Role of Big Data Analytics In Banking", 2019 International Conference on contemporary Computing and Informatics (IC3I).
4. KaithekuzhicalLeenaKurien*1 & Dr. AjeetChikkamannur, "Detection and Prediction of credit card fraud transactions using machine learning",International Journal of Engineering Sciences & Research Technology(IJESRT),ISSN:2277-9655,pp. 205-207
5. CorreaBahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for UPIfraud detection," Expert Systems with Applications, vol. 51, pp. 134–142, 2016.
6. Varun Kumar K S, Vijaya Kumar V G, Vijay Shankar A, Pratibha K, "UPIFraud Detection using Machine Learning Algorithms",International Journal of Engineering Research & Technology (IJERT) ,Volume 09, Issue 07 (July 2020),pp 5-8.
7. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi and G. Bontempi, "UPIFraud Detection: A Realistic Modeling and a Novel LearningStrategy," in IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 8, pp. 3784- 3797, Aug. 2018.
8. Pratyush Sharma, Souradeep Banerjee, Devyanshi Tiwari, and Jagdish Chandra Patni, "Machine Learning Model for UPIFraud Detection-A Comparative Analysis",The International Arab Journal of Information Technology, Vol. 18, No. 6, November 2021 , pp 789-790.
9. Ruttala Sailusha ; V. Gnaneswar ; R. Ramesh ; G. Ramakoteswara Rao ,"UPIFraud Detection Using Machine Learning ",2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS) ,19 June 2020
10. Kha Shing Lim, Lam Hong Lee and Yee-WaiSim,"A Review of Machine Learning Algorithms for Fraud Detection in Credit Card Transaction",IJCSNS International Journal of Computer Science and Network Security, Vol.21 No.9, September 2021,pp 32-37