A Review on Credit Card Fraud Detection Systems (CCFDS) using Machine Learning (Apache Spark)

Vinaya D S [1], Satish B Basapur[2], Vanishree Abhay[3], Neetha Natesh[4]

[1]M.Tech student Information science & Dr.Ambedkar Institute of Technology, Bangalore, Karnataka, India
[2,3, 4] Asst Professor Information science & Dr.Ambedkar Institute of Technology, Bangalore, Karnataka, India

---***---

Abstract - *As the payment method is simplified by the combination of the financial industry and IT technology, the payment method of consumers is changing from cash payment to electronic payment using credit card, mobile micropayment, and app card. As a result, the number of cases in which anomalous transactions are attempted by abusing e-banking has increased and financial companies started establishing a Fraud Detection System (FDS) to protect consumers from abnormal transactions. The abnormal transaction detection system aims to identify abnormal transactions with high accuracy by analysing user information and payment information in real time. Although FDS has shown good results in reducing fraud, but the majority of cases being flagged by this system are False Positives that resulting in substantial investigation costs and cardholder inconvenience. The possibilities of enhancing the current operation constitute the objective of this research. Based on variations and combinations of testing and training class distributions, experiments were performed to explore the influence of these parameters. In this study, we investigated the trend of abnormal transaction detection using payment log analysis and data mining, and summarized the data mining algorithm used for abnormal credit card transaction detection. We used python programming with apache spark for advanced processing of data and high accuracy.*

Keywords: *credit card, Fraud detection, Outlier detection, GBT classifier*

I-INTRPDUCTION

Million and billions of people use the credit card for payment in both online and offline transaction, due to existence of widespread point of sale (POS).countless transaction occurred per minute everywhere in the planet. The reason behind fraud is negligence of user .when third person steal the most important information about credit card and user details easily fraud can be achieved. To detect what type of fraud occur during transaction, we need to face Several challenges. Fetching that among all the transactions is occurred and which one is real could be a task.

Amongst the standard and very common ways of making payment globally and especially in North America, because of the presence of a far reaching point of sale. A huge number of individuals around the globe use charge cards to buy products and services by getting credit for a time of half a month. Any helpful framework could be mishandled and charge card is no exemption to this. Alongside the ascent of charge card use, extortion is on the ascent. Monetary Institutions (FIs) endure refined fake exercises and bear a large number of dollar misfortunes every year. In light of statistics [2] frauds account to more than $1 billion every year for Visa and MasterCard around the world.

Credit card companies and their part banks attempt to discover better approaches to forestall scams. A portion of the precautionary measures on the cards are magnetic stripes, 3D monograms, and CVC. Credit card companies are likewise taking steps to have alternate for credit cards as Smart Cards, be that as it may, in view of assessments this substitution will be over the top expensive because of the broad POS network in USA and the gigantic no. of cards available for use in those places. FIS additionally utilizing an assortment of computing mechanism, such as Neural Networks (NNs), to follow and distinguish dubious exchanges and ban them for additional examination.

II. LITERATURE REVIEW

In few years, the significant increase in the number of credit card issuances has also brought about an increasing number of fraud transaction cases, fraud methods are constantly updated, and crime-making skills are becoming increasingly sophisticated. In the past, credit card fraud recognition mechanisms based on human recognition and machine recognition are no longer enough to meet the needs of today's credit card fraud recognition. How to effectively, quickly and accurately identify credit card fraud transactions has become an urgent problem in the banking industry. The article reviews recent literature on credit card fraud identification. Existing literatures apply neural network, decision tree, random forest and other machine learning methods to credit card fraud problems, proving the effectiveness of machine learning models to solve credit card fraud problems. Credit card fraud identification is an important issue that needs to be resolved urgently. The number of card holders per capita in the United States has reached 2.9 in 2017. The problem of credit card fraud not only inflicts huge economic losses on credit card holders and banks, but also negatively affects the bank's safety and reputation. Therefore, it is urgent to establish a reasonable and effective credit card fraud recognition mechanism. At this stage, credit card fraud recognition mainly adopts the machine learning methods.

In last few years, the quantity of credit card issuance has also increased significantly It has brought about an increasing number of credit card fraud transactions worldwide, fraud methods are constantly updated, and crime-making skills are becoming increasingly sophisticated. In the past, credit card fraud recognition mechanisms based on human recognition and machine recognition are no longer sufficient to meet the needs of today's credit card fraud recognition. How to effectively, quickly and accurately identify credit card fraud transactions has become an urgent problem in the banking industry. The section reviews recent literature on credit card fraud identification. Existing literatures apply neural network, decision tree, random forest and other machine learning methods to credit card fraud problems, proving the effectiveness of machine learning models to solve credit card fraud problems. In recent years, due to the surge in credit card transaction volume, the huge transaction volume has high requirements for data processing methods, and traditional credit card fraud identification methods have been unable to meet actual needs. In addition, there are serious data imbalances in the data set of credit card transaction records, that is, the volume of legal transactions far exceeds the volume of fraudulent transactions, which brings great obstacles to behavior recognition and puts forward higher requirements for credit card fraud recognition mechanisms.

2.1 Characteristics:

(1) Objects of credit card fraud
From a subjective point of view, there are four kinds of objects directly affected by credit card fraud, namely cash currency, commodity currency, broad financial services and credit card itself.

Money cash and product money are transporters of social riches and have a place with substantial property. The culprit of credit card extortion wrongfully has non-individual property as his own, which is an infringement of social riches. Clearly, substantial property is the immediate objective of credit card misrepresentation. Dealer administration alludes to the administration gave via cardholders by administration staff of extraordinary credit card vendors, for example, lodgings and diversion focuses. This is a sort of emerged social work that exists in the administration procedure and has a place with a sort of social ware. Correspondingly, the charge card itself has a physical structure, that is, a plastic card, which is utilized as a voucher for product exchanges in the course field, and mirrors a specific money related capacity, so it is a unique budgetary item. From the point of view of social products, both dealer administrations and charge cards themselves might be straightforwardly encroached with Visa extortion and become the object of Visa misrepresentation.

However, only a visual investigation is not sufficient to detect the full meaning of the object of credit fraud. Because, fraud is first and foremost an infringement of the bank credit represented by the credit card. Illegal use of credit cards means illegally using bank credit. No matter what kind of fraudulent behavior, it directly acts on bank credit to achieve the illegal results sought by the perpetrator. It can be seen that the infringement of bank credit is the essence of credit card fraud.

(2) Objects of Credit card fraud mainly based on

- It includes public and private property ownership, bank financial management order and merchant management order
- It has essential and phenomenal relationship
- External manifestation
- Internal nature
- Complex object.

(3) The intermediary of credit card fraud is the credit card itself
The major characteristics of credit card fraud is that its fraud intermediary is the "credit card" itself, example; credit card fraud is done by using the "credit card" medium. The "credit card" here includes duplicate cards of original (such as after the original card is lost or stolen, the third person acts as a cardholder to commit fraudulent property, which is the scope of fraudulent use of real cards), as well as fake cards, "Black card" and so on. Whether it is a real card, a fake card or a "black card", it has become one of the tools used by criminals to commit frauds in credit card fraud and criminal activities.

Forms of credit card fraud

(1) Lost card fraud
There are generally three cases of lost cards. One is that the card-issuing bank lost the card when it was sent to the cardholder, that is, the card was not reached; the second is that the cardholder lost it by improper custody; the third is the theft by a criminal.

(2) Counterfeit application
Generally, they use other people's information to apply for credit cards, or intentionally fill in false information. The most common is to forge ID cards and fill in false work or home addresses.

(3) Forged credit card
Over 60% of international credit card fraud cases are fraudulent card frauds. It is characterized by the nature of gangs, from stealing card information, manufacturing fake payment cards.

Impersonators frequently use some of the latest technological methods to steal original credit card information, some use micro-recorders to steal credit card information, and some are opportunistic to steal authorization card terminal functions to steal credit card information. After stealing real credit card information, they will provide Manufacture fake cards in batches, and then commit crimes by selling fake cards to make huge profits.

Countermeasures against credit card fraud [1]

(1) Improve rules and regulations
First, we must start with a sound system, and we must regularly check the implementation of rules and regulations, find problems in a timely manner, and constantly improve. We must strictly control the issuance of blank cards, vouchers, and documents, and carefully implement the registration and storage system. The registration and sending of "blacklists" should be timely and accurate. Authorization must be done in strict accordance with the operating procedures. In order to adapt to the continuous development of the credit card business, not only must the credit card staff be repeatedly trained, but also all banking business personnel must have a certain degree of understanding and understanding of the functions, business requirements and regulations of the credit card, and give full play to the enthusiasm of credit card staff in practice The problem is constantly found in China, and suggestions are made to improve the system.

(2) Improve the ability of credit card practitioners and special merchants to argue against fake credit cards
A large part of credit card fraud crimes is due to the use of fake credit cards by criminals, and due to the lack of recognition or negligence of staff of credit card business staff and special merchants, the credit card business staff and staff of special merchants must be greatly improved. The ability to identify serious and fake credit cards, while improving their sense of responsibility for their work.

(3) Vigorously improve the anti-counterfeiting technology of credit cards
Improving the anti-counterfeiting technology of credit cards will undoubtedly set up a barrier for those criminals who are profit-making and eager to move around, and will definitely reduce the occurrence of fraudulent credit card fraud.
(4) Strict authorized management

Authorization management refers to a method to control the consultation of credit doubt and consumption exceeding the limit. If the authorization management is done well, it can reduce the occurrence of fraud, reduce the risk loss, and give criminals no opportunity. It is necessary to formulate detailed hierarchical authorization principles and starting points of hierarchical authorization amounts, insist on authorization within the scope of authority, adhere to the 24-hour authorization system and implement a strict authorization record and handover registration system to ensure the accuracy of every transaction.

(5) Bank credit card business practitioners and staff of special merchants shall strictly follow the operation steps when accepting credit cards.

(6) Establish a blank voucher and punch card management system
After the blank card is easy to be stolen by criminals, the self-scaling fake card number and other information are forged into a "credit card". Therefore, the management of blank cards must be strengthened to prevent undue losses. The receipt, issuance and storage of all kinds of blank certificates (cards) must be included in the "Blank Important Witness" table. Business executives should conduct regular checks from time to time to ensure that the accounts are consistent.

(7) Pay close attention to the construction of electronic engineering
Credit card business needs to use advanced scientific and technological equipment to transfer information and provide services. Because the delivery and prevention of stop lists and malicious overdrafts require a developed electronic network, if the stop list is not delivered in time, it will cause greater economic losses. Each payment stop list goes through internal sorting, printing, mailing and other procedures. It usually takes about 10 days before it can be sent to the special merchants and cash withdrawal points. If it is used by criminals during this period of time, it may cause certain economic losses. Therefore, an POS automatic authorization system should be established as soon as possible; the stop payment list inquiry machine realizes the electronic exchange of data files across the country and quickly transmits relevant information. American Express uses a CAT90 credit card verification terminal. In addition, the computer knowledge of credit card practitioners should be strengthened. Only in this way can advanced technology be used to combat criminal activities and stop credit card fraud. To this end, we must invest a lot of manpower and material resources to do a good job in electronic engineering construction.

(8) Strict law enforcement and strengthen efforts to combat credit card fraud
For the use of credit card crimes to forge credit cards or malicious overdrafts, how to conviction and punishment is not clearly stipulated in China's old criminal code, because China's criminal code was born in 1974, and China began to use and issue credit cards in 1986. The relevant legislation has not kept up, which has created an unreliable situation for our judicial practice. To a certain extent, this has affected the strength of our judicial department in cracking down on credit card fraud. Because for serious crimes of credit card fraud, only criminals are liable to bear civil or administrative responsibility, which is obviously not conducive to combating such illegal criminal activities. To some extent, it also constitutes an indulgence for credit card fraud. According to a survey conducted by the Bank of Korea in 2014 on the use of payment methods, consumers recognized convenience as the most important factor when choosing payment methods, and responded that the most convenient payment method was a credit card [4]. Fig. 1. Shows the amount of transaction amount by means of payment in various countries. Among the total transaction amount, credit card payments accounted for the highest proportion of credit card transactions such as Korea 71%, Canada 71%, Netherlands 64%, USA 55%, Australia 50%, and France 46%.

As non-face-to-face transactions are revitalized worldwide, the need for research on abnormal transaction detection is emerging. The rule-based detection algorithm that can be installed in multiple systems has the advantage of speed, but it cannot detect unknown anomalies and the policy or payment pattern of each country is different, so it is necessary to determine the ideal rule for each system. The disadvantage lies in that scenario. On the other hand, anomalous transaction detection using data mining can find a meaningful rule from it with a large amount of payment, and as the system becomes highly efficient, it can overcome the disadvantages that require a lot of computation, so many studies are currently underway.

Nagi et al. presented an intrusion detection frauds using data mining techniques for financial fraud detection. The papers from 49 journals published in 2018. This paper allows the analyzed and classified into four fraud categories and six data mining techniques [5].

Sanjeev et al. Is a paper that analyzes and classifies fraud type classification and fraudulent transaction frequency and amount by country through actual credit card fraud transaction data, and visually expresses the distribution using a box plot [6].

Michael et al. presented a signature based research on fraud detection and a fraud detection model [7] to provide a comprehensive survey of existing research related to fraud detection and to conduct fraudulent transactions in real time [7].

In the case of real-time detection, it is necessary to make accurate judgments in an instant and consider the characteristics of the data mining algorithm. TS Quah et al. Implemented real-time detection by separating detection mechanisms into the initial authentication layer, inspection layer, core layer for risk score evaluation and behavior analysis, and additional review layer using the SOM algorithm [17].

III. COMPARISONS

Performance of all learning algorithms used for fraud detection in credit card transactions are compared in table 1 The comparison is based on their accuracy, precision and specificity.

Table 1: comparison of machine learning techniques

Classifiers	Accuracy	Precision	Specificity
Outlier Detection and GBT classifier	0.989	0.997	0.987
Random forest	0.949	0.996	0.979
Logistic egression	0.941	0.410	0.975
KNN	0.930	0.410	0.975
SVM	0.928	0.782	0.912
Navie bayes	0.917	0.505	0.954

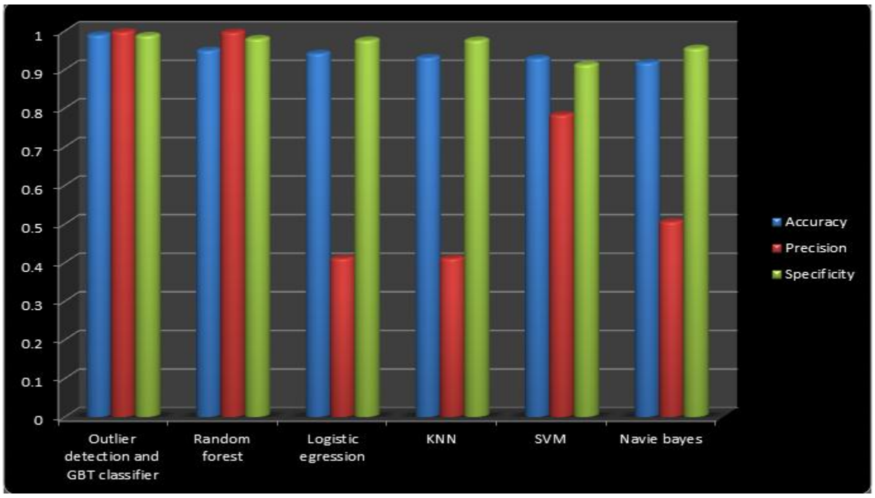

Fig.1. Accuracy, precision and specificity performance of all classifier

From table 1 we can see that accuracy of GBT Classifier and outlier detection is far better than the other learning algorithms. From fig. 1 we can see that precision, accuracy and specificity of GBT Classifier highest followed by Logistic regression, SVM, Decision Tree, Naive Byes and KNN. Hence the proposed system using Outlier detection will show better accuracy for larger number of training data.

3.1 PROBLEM STATEMENT

Not al1 the doubtful transactions consider as fraudulent. It is commonly called as false positive (FP) which means that the case was not fraud although it was flagged as being potentially scam. This process of affirming each transaction those outliers from the cardholder's normal routine brings doubt about possible client disappointment. Additionally, the expenses related with exploring an enormous no. of false positives are high.

3.2 Motivation

As of now, a considerable amount of time is given for examining countless genuine cases (FPs). On the off chance that the quantity of examination on FPs could be dropped down, scam analysts can invest more energy and time in genuine fraud transactions that restricts the losses to the FIs.

IV. SYSTEM ANALYSIS

4.1 Proposed System

The key objective of current research is improvising the procedure of personal follow up on a large number of suspicious transactions and to discover a path to preprocess the flagged records to recognize the probable genuine entries from the list of genuine/falsified entries. Here, the volume of needless analysis is decreased leading to significant savings for the financial institutions. Moreover, the current FDS threshold can also be lowered and a number of fraudulent cases, being missed under this level, can be detected. As a result, the fraud is discovered earlier and the overall losses may be reduced. For addressing these challenges, outlier detection and GBT Classifier is used, i.e. among the very common used applications of Machine Learning for addressing the pattern recognition and classification problems. The results indicate that the used method has a very good possibility to improvise the present system.

Advantages

- The proposed method overcomes the low accuracy forecast problem.
- Utilizing latest AI methods, the fraudulent transactions are recognized and the false alerts are reduced.
- Fast and reliable solution is attained.

4.2 System Architecture:

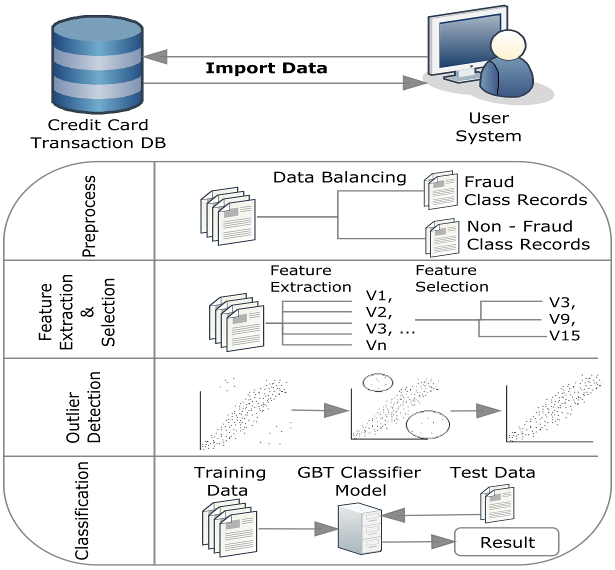

Figure 2: System Architecture

4.3 Module Description

1. **Data Collection:** It contains 2,84,807 records of credit card transactions that happened in duration of just 2 days. This dataset is very much unbalanced as it has a total of 492 fraud entries 2,84,515 genuine entries i.e. is just 0.17% of total records. The original features are masked with V1, V2, V3, ...V28. The last column here represents fraud or non-fraud class i.e. represented by 0 and 1 respectively.

2. **Data Balancing:** Imbalanced classes are a general issue in ML based classification where there is an abnormal count of each class. It occurs due to the fact that ML Algorithms are typically intended to improve precision by diminishing the errors. In this manner, they don't consider the class or balancing the ratio of classes. As out of 2.84807 transactions just 492 fraud transactions exist, which makes it quite difficult to build a standard model with this much less number of fraud transactions. Thus, we use pandas in python to make it 50-50 i.e. we decrease the no. of legitimate transactions to balance it with the number of fraud transactions in equal proportion.

3. **Feature Extraction:** We use heatmap technique to find the significant feature that can distinguish the classes properly and ultimately that affects the accuracy of detection algorithm. Heatmap provides a good visualization of the major and minor values in the matrix as different colored cells that define the values. Here, rows/columns of the matrix are clustered in sets. Thus, the features which look most significant are recognized and used further for model training.

4. **Outlier Detection**: The outlier detection technique measures the distance of each data similar to the clustering technique, but is used to find specific data and rules that are separated from the total data. The values which are not in flow of the linear graph are considered as outliers. Here our aim is to reduce the outliers to have a better trained model. We use numpy library in python for this.

5. **Classification:** The task of classification occurs in a wide range of applications. In a broad sense, the term could relate to any context in which some decision or forecast is made on the basis of currently available information. It works on a set of pre-defined classes on the basis of observed attributes or features. Here the aim is to establish a rule whereby one can classify a new observation into one of the existing classes. The construction of a classification procedure from a set of data for which the true classes are known has also been variously termed as pattern recognition, discrimination, or supervised learning. We use PySpark and GBT Classifier for data streaming and classification purpose. PySpark library is applied as a SQL-like analysis to a large amount of structured or semi-structured data. GBT Classifier does the classification of data coming through the stream.

V.CONCLUSIONS

With the development of electronic financial transaction technology and the emergence of simple payment, the risk of fraudulent payment and fraudulent payment increases as the authentication process is simplified. The types of fraudulent use of credit cards include theft and loss, identity theft, new card not received, card forgery, and card information theft. In particular, as phishing, pharming as well as card information leakage due to card information leakage, card information theft accidents are occurring. In response, the government tried to deal with electronic financial fraud by implementing the 'e-financial fraud prevention service'. It is difficult to cope with financial fraud by simply setting the existing keyboard security, public certificate, and additional password. The abnormal transaction detection system is used to analyze the user's data and payment data in real time to inform the financial institution and the user of the detection if it is different from the usual pattern, and further to arbitrarily stop the transaction. Therefore, an abnormal transaction detection system is important for fast and accurate detection, and research is needed to improve the algorithm. In this study, the method of detecting anomalous transactions using the electronic payment log analysis and machine learning technique was investigated. Results show the significance of algorithms used over the dataset and efficient classification is performed.

In future deep learning concepts can be applied using convolution networks for improved accuracy. Also some other datasets can be used for further testing of proposed mechanisms.

REFERENCES

[1] Donald V. Macdougall, Richard G. Mosley, Garioch J. l. Saunders; Credit card crime in Canada: Investigation - Prosecution; The Canadian Association of Crown Counsel; page 1-56; January 1985.

[2] Isabelle Sender; Detecting and combating fraud; Chain Store Age; New York; Vol. 74; Issue 7; Page 162; July 1998.

[3] Elford Dean, Raj Thomas, Lorry; Visa security center; Personal meetings; January 7 and February 11,1999.

[4] Gyusoo Kim and Seulgi Lee, “2014 Payment Research”, Bank of Korea, Vol. 2015, No. 1, Jan. 2015.

[5] EWT Nagi, Yong Hu, HY Wong, Yijun Chen, Xin Sun, “The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of Literature,” Decision Support Systems, Vol. 50, No. 3, Feb. 2011.

[6] Jha, Sanjeev, J. Christopher Westland, “A Descriptive Study of Credit Card Fraud Pattern,” Global Business Review, Vol. 14, No. 3, pp. 373-384, 2015.

[7] Edge, Michael Edward, Pedro R. Falcone Sampaio, “A Survey of Signature based Methods for Financial Fraud Detection,” Computers & Security, Vol. 28 No. 6, pp. 381-394. 2009.

[8] Aihua Shen, Rencheng Tong, Yaochen Deng, "Application of Classification Models on Credit Card Fraud Detection," Service Systems and Service Management of the 2007 IEEE International Conference, pp. 1-4, Jun.2007.

[9] Chengwei Liu, Yixiang Chan, Syed Hasnain Alam Kazmi, Hao Fu, “Financial Fraud Detection Model: Based on Random Forest,” International Journal of Economics and Finance, Vol. 7, No. 7, pp. 178-188, 2015.

[10] Ganesh Kumar.Nune and P.Vasanth Sena, “Novel Artificial Neural Networks and Logistic Approach for Detecting Credit Card Deceit,” International Journal of Computer Science and Network Security, Vol. 15, No. 9, Sep. 2015..