REAL-TIME CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING

Dr.E.Punarselvam[1,] G.Nivedhitha[2], B.Ilavarasan[3], R.Naveen Kishore[3],C.S.PranavAdhithya[3],P.Prithivi Raj[3]

E-Mail:1.punarselvam83@gmail.com, 2.nivedhitha.g.it@mec.edu.in

1. Head of the Department/IT, 2.Assistant professor/IT 3.Final Year Student

Department of Information Technology, Muthayammal Engineering College, Rasipuram, Tamilnadu.

ABSTRACT

Credit card fraud events take place frequently and then result in huge financial losses. The number of online transactions has grown in large quantities and online credit card transactions hold a huge share of these transactions. Online transactions have become an important and necessary part of our lives. As frequency of transactions is increasing, number of fraudulent transactions is also increasing rapidly. As credit card becomes the most popular mode of payment for both online as well as regular purchase, cases of fraud associated with it are also rising. Implementation of efficient fraud detection systems has thus become imperative for all credit card issuing banks to minimize their losses. Many modern techniques based on Artificial Intelligence, Data mining, Fuzzy logic, Machine learning, Sequence Alignment, Genetic Programming etc., has evolved in detecting various credit card fraudulent transactions. A clear understanding on all these approaches will certainly lead to an efficient credit card fraud detection system. The most commonly used fraud detection methods are Neural Network (NN), rule-induction techniques, fuzzy system, decision trees, Support Vector Machines (SVM), Artificial Immune System (AIS), genetic algorithms, K-Nearest Neighbor algorithms. These techniques can be used alone or in collaboration using ensemble or meta-learning techniques to build classifiers. This thesis presents a survey of various techniques used in credit card fraud detection and evaluates each methodology based on certain design criteria.

Key Words: Data mining, Fuzzy logic, Machine learning, NN, SVM, AIS, K-Nearest Neighbor Algorithm.

1 INTRODUCTION

Credit cards are the plastic cards issued by financial companies and institutions such as banks. The payment card allows the cardholder to repeatedly borrow funds to buy products, food, merchandises or services. Here the card user agrees to pay the amount back within the due date. It is a lot easier to carry payment cards over cash while on the move. We can even perform online payments using credit cards. In such type of transaction, the card holder has to enter few details (card number and expiry date) and make his/her purchase online. This ease of use has made credit cards more popular nowadays.

The growing popularity of e-commerce in a way has led to the increase in credit card users and cashless payments. If we consider the growth of credit cards in India in May 2015 , 21480389 were the number of credit cards issued and 570813794 debit cards were issued. In the year 2016, online spending value increased rapidly than that of offline payments[2] i.e., increase in percentage was 26%(online) and 17%(offline).

THE VARIOUS ADVANTAGES OF USING CREDIT CARD INCLUDE

(1) Easy to carry

(2) Helps to keep track of expenses

(3) Instant cash

(4) Flexibility and convenience.

Fraud has been increasing drastically with the progression of state-of-art technology and worldwide communication. Fraud can be avoided in two main ways: prevention and detection. Prevention avoids any attacks from fraudsters by acting as a layer of protection. Detection happens once the prevention has already failed. Therefore, detection helps in identifying and alerting as soon as a fraudulent transaction is being triggered. Recently, card not- present transactions in credit card operations have become popular among web payment gateways. According to the Nilsson Report in October 2016, more than $31 trillion were generated worldwide by online payment systems in 2015, increasing 7.3% than 2014. Worldwide losses from credit card fraud have been rising to $21 billion in 2015, and will possibly reach $31 billion by 2020. However, there has been an extreme increase in fraudulent transactions that affect the economy dramatically. Credit card fraud can be classified into several categories.

The two types of frauds that can be mainly identified in a set of transactions are **Card-not-present** (CNP) **frauds** and **Card-present** (CP) **frauds**. Those two types can be described further by bankruptcy fraud, theft/counterfeit fraud, application fraud, and behavioral fraud. The machine learning algorithms are used in various sectors of industry especially when it comes to computing field.

These algorithms are developed by machine learners who are specialized in machine learning field with proper study of all the tools handled in the respective field. Using machine learning tools, we can perform various instructions to execute the system works. Machine Learning is now playing important aspects in mobile communication field too. Since several big and popular companies are totally depended on these experts and in future, they need more and more of these peoples in order to match the rivalry with their opponents. These machine learning algorithms now used to find aspect such as emotion detection, fraud detection, path finding etc. such projects can be achieved by using machine learning algorithms.

However there is lack of Data scientist in coming future .Generally Data scientist are group of people who study about machine learning, data processing, data analyst etc. Though machine learning are now further improvised and now we have deep learning concepts which is more accurate and uses tools like Tensor Flow etc. and advanced algorithms which makes it more complicated than our general machine learning algorithms and also it takes much more time in writing code but it is more time efficient than machine learning algorithms with much more higher level of accuracy.

2 CREDIT CARD FRAUDS CAN BE DIVIDED INTO 2 TYPES:

Inner card fraud

External card fraud.

Inner card fraud

Inner card fraud intends to defraud the cash. Usually it is the collusion between merchants and cardholders, using false transactions to defraud banks cash.

External card fraud.

External card fraud is mainly embodied at using the stolen, fake or counterfeit credit card to consume, or using cards to get cash in disguised forms, such as buying the expensive, small volume commodities or the commodities that can easily be changed into cash.

Fraud detection involves monitoring and analyzing the behavior of various users in order to estimate detects or avoid undesirable behavior. In order to identify credit card fraud detection effectively, we need to understand the various technologies, algorithms and types involved in detecting credit card frauds. Algorithm can differentiate transactions which are fraudulent or not. Find fraud, they need to passed dataset and knowledge of fraudulent transaction. They analyze the dataset and classify all transactions.

The credit card is a small plastic card issued to users as a system of payment. it allows its cardholder to buy goods and services based on the cardholder's promise to pay for these goods and services. Credit card security relies on the physical security of the plastic card as well as the privacy of the credit card number. Globalization and increased use of the internet for online shopping has resulted in a considerable proliferation of credit card transactions throughout the world. Thus a rapid growth in the number of credit card transactions has led to a substantial rise in fraudulent activities. Credit card fraud is a wide-ranging term for theft and fraud committed using a credit card as a fraudulent source of funds in a given transaction. Credit card fraudsters employ a large number of techniques to commit fraud.

To commit fraud in these types of purchases, a fraudster simply needs to know the card details. Most of the time, the genuine cardholder is not aware that someone else has seen or stolen his card information. In real life, fraudulent transaction are scattered with genuine transactions and simple pattern matching Techniques are not often sufficient to detect those frauds accurately.

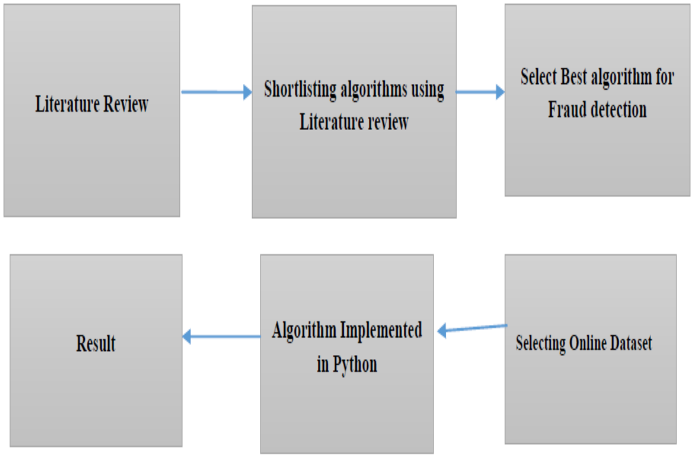
Literature Review
Shortlisting algorithms using Literature review
Select Best algorithm for Fraud detection
Selecting Online Dataset
Algorithm Implemented in Python
Result

Fig. 1 Flow of finding fraud detection

3 CREDIT CARD FRAUD VARIANTS

1) ID theft: When an attacker obtains the personal information of a victim such as date of birth, gender, email id, he can easily get access to a new account using victim's details or even a step further by taking hold of the existing account. Identity theft constitutes 71% of the most common type of fraud.

2) Fake cards: Card which is not authorized or not issued by financial institutions is termed as fake cards. Fake cards are developed by skimming the actual data of genuine card which was swiped over an EDC machine. This data is encoded from the magnetic strips and later used to create fake cards.

3) Stolen/lost cards*:* A scenario where a card holder accidentally loses his card or his card has been stolen, if the cardholder fails to report it to the concerned bank there might be chances that the card can be misused by a criminal.

4) CNP fraud: Card, not present fraud is a type of fraud where the criminal requires minimal information such as card number and expiry date. In such situation, the card need not be present while making the purchases online.

5) Clean frauds: These frauds are not as clean as they sound. The purchases are made with stolen cards and later transactions are modified thus making it find a way around the FDS.

6) Friendly fraud: In friendly fraud the actual cardholder himself makes the purchases and pays for the services using "pull" mode of payment with his credit/debit card. Later reports a complaint stating loss of card and claims for reimbursement.

7) Affiliate fraud: It is the most widely distributed fraud where either an individual logs into a website and makes purchases using a false account or a program is designed to carry out fraud activities.

8) Triangle fraud: Such fraud mainly involves 3 steps: (a) Creating a fake website (b) Providing offers such as immediate delivery upon credit card payment mode (c) Stolen or fake cards are used to make the payments and the name obtained at the real store is misused by the criminal to later ship the product to the customer.

Fraud detection systems are trained using older transactions to decide about future ones. The faster a fraud detection system performs the better. In fraud detection, the count of normal cases is much more than the unauthorized cases. This causes a condition called "imbalanced data" where one class of data has a very high number of instances as compared to the other class of data. This leads us to the "class imbalance problem".

4 EXISTING SYSTEM

Due to rise and acceleration of E-Commerce, there has been a tremendous use of credit cards for online shopping which led to High amount of frauds related to credit cards. In order to identify credit card fraud detection effectively, we need to understand the various technologies, algorithms and types involved in detecting credit card frauds. Financial fraud is increasing significantly with the development of modern technology and the global superhighways of communication, resulting in the loss of billions of dollars worldwide each year. The companies and financial institution loose huge amounts due to fraud and fraudsters continuously try to find new rules and tactics to commit illegal actions.

Disadvantages

- In real life, fraudulent transactions are scattered with genuine transactions and simple pattern matching techniques are not often sufficient to detect those frauds accurately.
- Algorithm can differentiate transactions which are fraudulent or not.
- Thus, fraud detection systems have become essential for all credit card issuing banks to minimize their losses.
- It should not classify a genuine transaction as fraud

5 PROPOSED SYSTEM

Fraud can be avoided in two main ways: prevention and detection. Prevention avoids any attacks from fraudsters by acting as a layer of protection. Detection happens once the prevention has already failed. The most commonly used fraud detection methods are Neural Network (NN), rule-induction techniques, fuzzy system, decision trees, Support Vector Machines (SVM), Artificial Immune System (AIS), genetic algorithms, K-Nearest Neighbor algorithms. These techniques can be used alone or in collaboration using ensemble or meta-learning techniques to build classifiers. The detection process, the payment transactions are ingested to the analytical engine using Apache Spark. Genuine transactions are allowed to complete whereas the fraudulent transaction is aborted. Though there are various benefits of using credit cards such as convenience, instant cash, but when it comes to security credit card holders, banks, and the merchants are affected when the card is being stolen, lost or misused without the knowledge of the cardholder (Fraud activity).

Advantages

- It is a lot easier to carry payment cards over cash while on the move.
- The growing popularity of e-commerce
- A clear understanding on all these approaches will certainly lead to an efficient credit card fraud detection system**.**
- It should identify the frauds accurately
- It should detecting the frauds quickly

6 METHODOLOGIES

Data description

The dataset was created combining two data sources; the fraud transactions log file and all transactions log file. The fraud transactions log file holds all the online credit card fraud occurrences while all transactions log file holds all transactions stored by the corresponding bank within a specified time period. Due to the confidential disclosure agreement made between the bank and the authors of the paper, some of the sensitive attributes such as card number were hashed.

Credit Card Fraud Detection Using Bayesian and Neural Networks

The credit card fraud detection using Bayesian and Neural Networks are automatic credit card fraud detection system by means of machine learning approach. These two machine learning approaches are appropriate for reasoning under uncertainty. An artificial neural network consists of an interconnected group of artificial neurons and the commonly used neural networks for pattern classification is the feed- forward network. It consist of three layers namely input, hidden and output layers. The incoming sequence of transactions passes from input layer through hidden layer to the output layer. This is known as forward propagation.

Credit card fraud Detection

Illegal use of credit card or its information without the knowledge of the owner is referred to as credit card fraud. Different credit card fraud tricks belong mainly to two groups of application and behavioral fraud [3]. Application fraud takes place when, fraudsters apply new cards from bank or issuing companies using false or other's information. Multiple applications may be submitted by one user with one set of user details (called duplication fraud) or different user with identical details (called identity fraud). Fraud detection systems are prune to several difficulties and challenges enumerated bellow. The system should take into account both the cost of fraudulent behavior that is detected and the cost of preventing it.

Modeling and testing

They are Support Vector Machine, Naive Bayes, K-Nearest Neighbor and Logistic Regression. We applied those selected supervised learning classifiers to our resample data. When selecting machine learning models which can capture each fraud, the accuracy and performance of each model were taken into consideration. Optimal models were selected by filtering them out comparatively against an appropriate performance matrix.

Fraud Detection Systems

Real-time detection of credit card fraud can be stated as one of the main contributions of this project. The real-time fraud detection system consists of three main units; API MODULE, FRAUD DETECTION MODELS. All the components are involved in fraud detection simultaneously. The predicted results and other important data of the machine learning models. The user can interact with the fraud detection system with GUIs where it shows the real time transactions, alerts regarding frauds and historical data regarding frauds in a graphical representation.

7 SYSTEM ARCHITECTURE

The sender (source) transmits the data to receiver (destination) through multiple paths. The data are divided into multiple segments and transmitted over a multiple nodes. The segmented data's are then send to base station through relay or direct transmission. The direct transmission is sending the data without any intermediate nodes whereas relay transmission is sending the data with intermediate nodes. The segmented data's are then reaches the final base station where the data's are rearranged to their original order and send to the destination node.

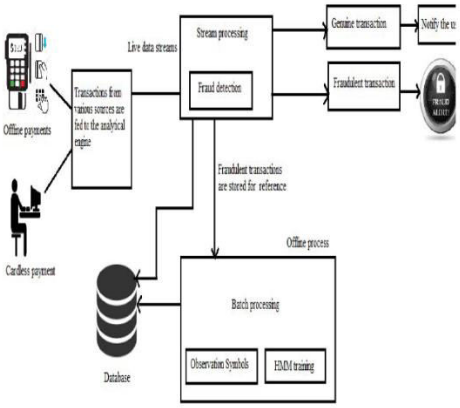

Fig 2 System Architecture

8 RESULTS AND DISCUSSION

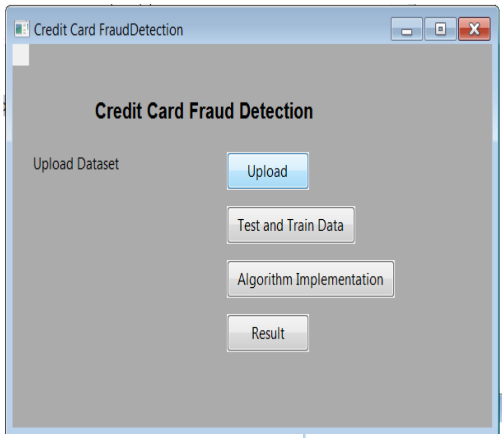

Fig 3 Home Page

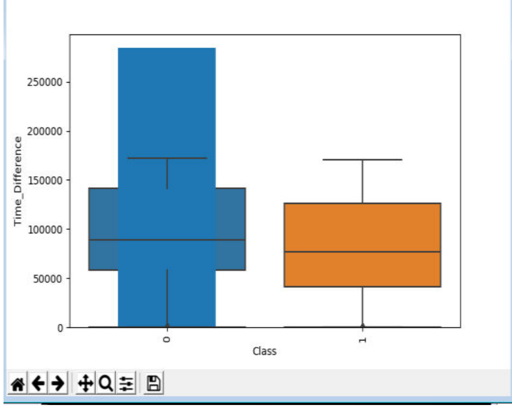

Fig 4 Time Difference

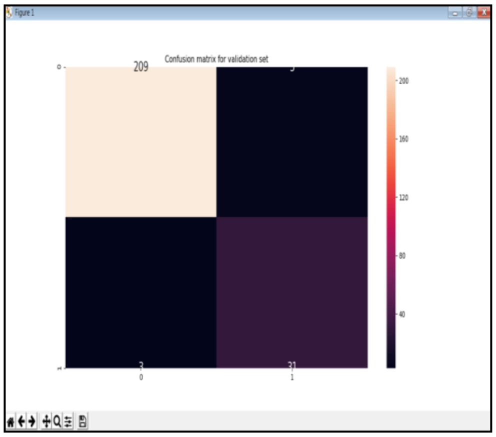

Fig 5 Performance Result

9 CONCLUSIONS

Credit card fraud detection has been a keen area of research for the researchers for years and will be an intriguing area of research in the coming future. This happens majorly due to continuous change of patterns in frauds. In this paper, we propose a novel credit-card fraud detection system by detecting four different patterns of fraudulent transactions using best suiting algorithms and by addressing the related problems identified by past researchers in credit card fraud detection. By addressing real time credit-card fraud detection by using predictive analytics and an API module the end user is notified over the GUI the second a fraudulent transaction is taken place. This part of our system can allow the fraud investigation team to make their decision to move to the next step as soon as a suspicious transaction is detected.

10 FUTURE ENHANCEMENTS

Optimal algorithms that address four main types of frauds were selected through literature, experimenting and parameter tuning as shown in the methodology. As the developed machine learning models present an average level of accuracy, we hope to focus on improving the prediction levels to acquire a better prediction. Also, the future extensions aim to focus on location-based frauds.

11 REFERENCES

[1] V. Bhusari and S. Patil.(2011). Use of concealed markov show in Visa misrepresentation discovery.Worldwide Journal of Distributed and Parallel Systems (IJDPS) Vol.2, No.6.

[2] E.Punarselvam,“Robust Facial Expression Recognition using Local Directional Number Version”, International Journal of Innovative Research in Science, Engineering and Technology, ISSN(Online) : 2319 – 8753,ISSN (Print) : 2347-6710 Vol. 4, Special Issue 6,May 2015,pp 182-186

[3] Sen, Sanjay Kumar., and Dash, Sujatha. (2013). Meta learning calculations for charge card extortion location. Universal Journal of Engineering Research and Development Volume 6, Issue 6, pp. 16-20.

[4] Dr.E.Punarselvam,"Supervised and Semi Supervised Machine Learning Clustering Algorithm based on feature selection", International Journal on Applications in Information and Communication Engineering, Volume 5 : Issue 2: November 2019, PP 19 –24, ISSN (Online) : 2394 – 6237

[5] N.Malini and Dr.M.Pushpa , "Analysis on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection" , 3rd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEEICB17) , 2017

[6] Dr.E.Punarselvam,"Effective and Efficient Traffic Scrutiny in Sweet Server with Data Privacy", International Journal on Applications in Information and Communication Engineering Volume 5 : Issue 2: November 2019, PP 1 – 5.

[7] John Richard D. Kho and Larry A. Vea, "Credit card Fraud detection based on transaction Behavior", IEEE Region 10 Conference (TENCON), Malaysia, pp 1880 – 1884 , November 2017

[8] Fahimeh Ghobadi and Mohsen Rohani, "Cost Sensitive Modeling of Credit CardFraud Using Neural Network Strategy" , IEEE ICSPIS 2016, Dec 2016

[9] Dr.E.Punarselvam,"Effective and Efficient Traffic Scrutiny in Sweet Server with Data Privacy", International Journal on Applications in Information and Communication Engineering Volume 5 : Issue 2: November 2019, PP 1 – 5 [10] Sarween Zaza and Mostafa Al-Emran, "Mining and Exploration of Credit Cards Data in UAE", Fifth International Conference on e-Learning , pp 275-79 , 2015

[11] E.Punarselvam, "Big Data using Hadoop Database using python Language to implement Real Time Applications", International Journal of Engineering Research and development,Vol.8 Issue No.12 Oct 2013 PP(19-22) e-ISSN:2278-067X,p-ISSN:2278-800X.

[12] Rajeshwari U and Dr B Sathish Babu, "Real-time credit card fraud detection using Streaming Analytics", 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), pp 439 – 444, 2016