International Journal of Knowledge Based Computer Systems
10 (1) June 2022, 22-26
http://www.publishingindia.com/ijkbcs/

Various Credit Card Fraud Detection Techniques based on Machine Learning

Chinnu Maria Varghese[1*], Deepika M. P.[2] and Abraham Varghese[3]
[1]PG Student, Computer Science and Engineering, Adi Shankara Institute of Engineering & Technology, Kerala, India. Email: chinnumariav@gmail.com
[2]Associate Professor, Computer Science and Engineering, Adi Shankara Institute of Engineering & Technology, Kerala, India. Email: deepika.it@adishankara.ac.in
[3]Faculty, University of Technology & Applied Science, Muscut, Oman. Email: abraham.v@hct.edu.com
*Corresponding Author

Abstract: Use of mobile devices facilitates more people to do online shopping using credit cards. As a result, Internet shopping has become a popular method of making daily purchases. Online shopping offers benefits like efficiency, convenience and greater selection as well as better pricing. With the number of transactions by credit cards are increasing rapidly transaction fraud are also increasing. A fraud activity results in financial loss to the individuals. Therefore financial institutions provide more value and demand for fraud detection applications. We need to also make our systems learn from the past submitted frauds and make them fit for adapting to future new methods of frauds. This paper shares the concept of frauds related to credit cards and their various types. We have considered various techniques available for a fraud detection system such as, Hidden Markov Model (HMM), Bayesian Network, Hybrid Support Vector Machine (HSVM), K-Nearest Neighbor (KNN), Naïve Bayes, Logic regression, Decision Tree and a feedback mechanism. This paper covers existing and proposed models for credit card fraud detection and focus to find the best method by comparing different techniques on the basis of quantitative estimations like accuracy, detection rate and false alarm rate.

Keywords: Bayesian network, Hybrid support vector machine, K-Nearest neighbor, Logic regression, Naïve Bayes.

I. Introduction

In the course of the most recent couple of years, e-commerce has become an essential part of the global retail framework. Like many other industries, the retail landscape has gone through a considerable change following the advent of the internet, and thanks to the on-going digitalization of modern life, consumers from virtually every country now profit from the perks of online transactions. As internet access and adoption are quickly expanding around the world, the number of online buyers keeps climbing every year. According to Statista study, around two billion people purchased goods through online, and e-retail sales surpassed 4.2 trillion U.S dollars worldwide during the year 2020 [1].

Today use of credit card in developing countries has become a common scenario. Credit cards can be used for shopping, paying bills, loan payments and for online transactions. However, as the number of credit card users grows, so does the number of credit card fraud instances. Credit card related cheats cause worldwide a loss of billions of dollars. Fraud can be defined as any activity with the intent of deception to obtain financial gain by any manner without the knowledge of the cardholder and the issuer bank. With the advancement of mobile phones, online shopping becomes a popular mode of everyday buys. However, the Internet environment is open, online shopping systems have problems, and criminals can use some bad techniques suchas Trojan and pseudo base-station. All these ends up in a serious increase of credit card fraud events. When a criminal steals or cheats the information of the cardholder, the criminal can utilize the credit card to consume. Credit Card fraud can be done in numerous ways such as by phishing, by producing fake cards, by stolen cards, by cloning the original site, by changing the magnetic strip present at the card which contains the user's information, by skimming or by stealing data from a merchant's side. With continued advancement in fraudulent strategies it is important to develop effective models to combat these frauds in their initial stage only, before they can take to completion. However, the main issue in constructing such a model is that the number of fraudulent transactions is quite tiny in comparison to the total number of transactions, making the task of discovering a fraudulent transaction in aneffective and efficient manner rather difficult.

Detecting a fraud is a complicated computational task. The number of parameters to select, cluster and classify are

tremendous and classification of parameters will determine the success of any fraud detection technique. Moreover a transaction can't be exactly classified as a fraud or a genuine one by the existing systems; they just find the chances of a transaction being fraud based upon the elaborated study of customer's behavior, their spending habits and also analyzing the previously committed frauds and identifying their patterns. Limited time period to determine the transaction is genuine or fraud and processing of large number of parameters during training are the two main challenges in making a decision.

A good fraud detection system should have the following properties:

- The number of wrong classifications should be minimum that is identifying the frauds accurately.
- It should be able to detect the fraud while it is in transit.
- It ought not term any genuine transaction as fake.

In our paper we have tried to study some techniques that can be used in a fraud detection system and have tried to do a comparative study to show which technique performs better under what scenarios.

II. Related Work

Credit-card-based purchases can be categorized into physical card and virtual card. In a physical-card based purchase, the cardholder submits his card physically to a dealer for making a payment. In this kind of purchase, an attacker has to steal the credit card for fraudulent transactions. If the cardholder doesn't understand the loss of card, it can lead to a considerable financial loss to the credit card organization. In the virtual card based purchase, only some important information about a card is required to make the payment. Such buys are normally done on the internet or via telephone. A scammer only has to know the card details to conduct fraud in these types of transactions. Most of the time, the certified cardholder is not aware that another person has seen or stolen his card information. The best way to find this kind of fraud is to analyze the spending patterns on each card and to sort out any irregularity with respect to the "usual" spending patterns. Fraud detection based on the analysis of existing purchase data of cardholder is a promising method to lessen the rate of successful credit card frauds. Since humans tend to exhibit specific behavioristic profiles, each cardholder can be addressed by a set of patterns containing information about the typical purchase category, the time since the last purchase, the amount of money spent, etc. Deviation from these patterns poses a risk to the system.

In 2008, Abhinav Srivastava and Amlan Kundu [2] in their paper proposed a Hidden Markov Model (HMM). Intheir model they model sequence of operations in credit card transaction processing using a Hidden Markov Model (HMM) and show how it can be used for the detection of frauds. An HMM is first trained with the normal pattern of acardholder. If an incoming credit card transaction is notaccepted by the trained HMM with sufficiently high probability, it is viewed as fake. At the same time, try to ensure that genuine transactions are not rejected. The different steps in credit card transaction processing are addressed as the underlying stochastic process of an HMM. For the observation symbols ranges of transaction amount is used and types of item is considered as HMM states .The paper proposed a method for discovering the spending profile of cardholders, as well as utilization of this knowledge in deciding the value of observation symbols and initial estimate of the model parameters. The paper also explain how the HMM can recognize whether an incoming transaction is fake or not. Test results show the performance and effectiveness of system and demonstrate the usefulness of learning the spending profile of the cardholders. Accuracy of the system is nearly 80 percentages over a wide variation in the input data. The system is also applicable for large volumes of transaction handlings.

V. Filippov and L. Mukhanov [3] have proposed a Fraud Detection System in their paper. The use of credit cards is very common in modern day society. However, the number of credit card fraud cases is continually increasing despite the chip cards worldwide integration and existing protection systems. This is why the issue of fraud detection is vital at this point. A general description of the developed fraud detection system and examinations between models based on using of artificial intelligence are given. In the fraud detection system two modules FDS ONLINEP and FDS OFFLINEP are used for fraud detection. The FDS ONLINEP module is used for online fraud detection. FDS OFFLINE module permits the system to detect fraud among transactions that have been already authorized and grouped in the FDS ONLINEP module. For the storage of incoming transactions, statistical data for corresponding models, results of classification and generic parameters a FDS Data Warehouse is used. Module FDS ALERT is used for alerting credit card holders in case of fraud detection by the FDS ONLINEP module using SMS or email messages. For building statistical data the FDS BUILDSTATP module is used. A clustering model is developed based on the use of the parameters' data clusterization regions. In this system 24 real parameters of transactions are used for classification. All of them are discrete. It is impossible to use the Naive Bayesian Classifier based on the discrete distribution, the normal distribution and the kernel density estimation for this type of fraud detection. To solve this problem they developed clusterization algorithm which gives an input attribute value 1 when a real value of this parameter hits into some region of data clusterization, or else it has value 0. Finally the evaluative testing of transactions have been generated for training process and the testing process

John O. Awoyemi, Adebayo O. Adetunmbi and Samuel A. Oluwadare [4] in their paper in 2017 did a detailed comparative study of Credit card fraud detection using Machine Learning Techniques. Data mining technique is one prominent techniques used in solving credit fraud detection problem. Credit card fraud detection is the way toward recognizing those transactions that

are fraudulent into two classes of legitimate and fraudulent transactions. The paper examines the performance of Naïve Bayes, K-Nearest Neighbour and Logistic regression on highly skewed credit card fraud data. A hybrid technique of under- sampling and oversampling is carried out on the skewed data. The performance of fraud detection in credit card exchanges is greatly influenced by the sampling approach on dataset, selection of variables and detection techniques used. A hybrid technique of under-sampling and oversampling is completed on the skewed data. The three methods are applied on the raw and pre-processed data. The work is carried out in Python. The performance of the techniques is assessed based on accuracy, sensitivity, specificity, precision, Matthews correlation coefficient and adjusted classification rate. The outcomes shows of ideal accuracy for Naïve Bayes, K-Nearest Neighbour and Logistic regression classifiers are 97.92%, 97.69% and 54.86% respectively. The similar results show that K-Nearest Neighbour performs better compared to Naïve Bayes and Logistic regression techniques. They concluded through experiments that Random Forest technique shows most accuracy followed by Logistic Regression and Support Vector.

In 2017 J. Vimala Devi and K. S. Kavitha [5] did a detailed study of different Fraud Detection methods in Credit Card Transactions by using Classification Algorithms. This paper mainly focused on three algorithms Support Vector Machine, Random Forest and Decision Tree and has compared the results of these algorithms. These three classification techniques are chosen since the dataset is huge, an imbalanced one. Each of these methods has its own merits and demerits depending on the application. Therefore based on the implementation, the decision tree classification algorithm is displaying the best accuracy for the given dataset compared to other classification algorithms. The actual dataset is considered since the data is highly imbalanced one, classification may be done by oversampling the data.

Specific crime within the banking industry is credit card fraud. Credit card use has been expanded because of the fast development of E-business strategies. Credit card fraud also increased simultaneously. Avoidance is better than detection. So the current system prevented the credit card fraud by recognizing fraud in the application of the Credit card. To overcome the limitations like scalability issues, extreme imbalanced class and time constraints of the existing system, V. Mareeswari and G. Gunasekaran [6] proposed new algorithm along with the existing algorithm. Those limitations are overwhelmed by hybrid support vector machine (HSVM) along with communal and spike detection for credit card application fraud detection. Credit card fraud detection is done at the initial stage of credit card application and general fraud and crime activities are predicted by this system. In order to perform the identification of frauds, system uses the hybrid support vector machine (HSVM) for computing the weight of the each attribute for communal and spike detection for credit card application fraud detection. The system is working with existing two layers Communal Detection (CD) and Spike Detection (SD) along with HSVM for computing weight of each attribute of applicant's application. It uses to identify for legal behavior and data errors of the applicant. The CD algorithm performs with communal data of the applicant andworks in real time by exact or similar matches between categorical data, giving scores. Blacklist or attribute- oriented approach on the variable - size set of attributes is done by SD. After CD evaluation, SD takes care of the further process.

Support vector machine is the most used method for the pattern recognition and classification. In their approach it performs prediction and classification on the credit card dataset and classifies into two classes; fraud and genuine transaction. Credit card application is an online application process under which users enter their personal info for registration. According to proposed model when the user feeds the data to the online application form, this data is matched with the Whitelist (WL) database by CD algorithm, if it is a valid then pass it to SD for verifying civil score according to the Blacklist (BL) Database. Prior to this data matching process the attributes are learned by proposed algorithm and then the dual optimization process is performed during this the attribute are divided based on their priorities. CD and SD algorithm are computing weight of the attribute by using hybrid support vector machine prediction method. The application for a credit card is granted or else it is rejected based on the verifying of the both algorithm's outcomes. The main target that focused on this system is to preserve the credit fraud in the initial stage of the credit life cycle. The implementation of this algorithm in order to perform the identification of frauds, this system uses the Hybrid Support Vector Machine (HSVM) for computing the weight of the each attribute for communal and spike detection for credit card application fraud detection.

Credit card fraud detection is a significant method to prevent fraud events, which is usually categorized into anomaly detection and classifier-based detection techniques. Anomaly detection means calculating the distance between the data points in space. By calculating the distance between the incoming transaction and the cardholder's profile, an anomaly detection method can filter any incoming transaction which is inconsistent with the cardholder's profile. The classifier-based detection technique utilizes some supervised learning methods to train a classifier on thebasis of the given normal transactions and fraud ones. The supervised learning extracts fraud features from fraud transactions. However, both of these techniques have limitations. For the anomaly detection, it has no capacity to portray fraud features although it can portray cardholders' transaction behaviors. For the classifier-based detection, it fails to recognize different normal behaviors from different cardholders although it can catch fraudsters' behaviors. Transaction habits of an individual vary regularly since they are easily influenced by their incomes, resources, ages and characters. Thus their distribution evolves over time because of seasonality and new attack methodologies. This is known

as the problem of concept drift that is difficult to be resolved by the above detection methods. On the other hand, both of them don't know about the adaptive capacity of the model. For instance, a person may involve some new transaction behaviours in a particular period which has never occurred in his/her history. The vast majority of the proposed methods just keep the recent instances for model training, but don't consider the adaptiveness of the model.

Facing the existing challenges J. Changjun Jiang, Jiahui Song and Guanjun Liu [7] proposed a feedback mechanism which can adapt to the transaction of cardholder's behaviours seasonally. This paper proposes a novel fraud detection method that consists of four stages. To enrich a cardholder's behavioural patterns, first use clustering method, all cardholders are isolated into three groups depending up on the transaction amount, that is high, medium and low. Then introduces a sliding-window-based technique to aggregate the transactions in each group, i.e., derive a set of extra features from windows to characterize a cardholder's behavioural patterns. After pre-processing features, train a set of classifiers for each group using the data comprising of each specific behavioural pattern and extracted fraud features. Finally, use a classifier set prepared for a group is assigned to each cardholder in the group as his/her own behavioural patterns, and the classifier with the most rating score is viewed as his/her recent behavioural pattern. Based on the three classifier sets, propose a fraud detection method in which a feedback mechanism is taken to take care of the concept drift problem.

III. Comparison

Table I: Comparison Table

Paper Name	Comparison		
	Method	Advantages	Disadvantages
Credit Card Fraud Detection using Hidden Markov Model [2]	Detection of frauds using a Hidden Markov Model (HMM)	Scalable for handling large volumes of transactions	Suffer performance degradation
Credit Card Fraud Detection System [3]	Developed a fraud detection system and clustering model	Clustering model- allows fast monitoring of incoming transactions	The model is less accurate
Credit Card Fraud Detection using Machine Learning Techniques [4]	Three classifier models based on Naive Bayes, K-Nearest Neighbour and Logistic regression are used	Hybrid sampling greatly improves the performance	Logistic regression did not show better improvement
Fraud Detection in Credit Card Transactions by using Classification Algorithms [5]	Decision tree, Random forest and Support Vector Machines (SVM) are used to classify the data	Decision tree classification algorithm is displaying the best accuracy	Sometimes require long raining times and be deficient in available memory, if dealt with large databases
Prevention of Credit Card Fraud Detection based on HSVM [6]	Uses hybrid support vector machine (HSVM) for computing the weight of the each attribute	Improved the efficiency as well as performance less training time	Not suitable for large datasets
Credit Card Fraud Detection: A Novel Approach using Aggregation Strategy and Feedback Mechanism [7]	A feedback mechanism is introduced	Solved the concept drift problem. Increase the average recall and accuracy	Low performance

IV. Conclusion

Although there are several fraud detection techniques available today but none is able to detect all frauds completely when they are actually happening, they usually detect it after the fraud has been committed. This happens because a very minuscule number of transactions from the total transactions are actually fraudulent in nature. So we need a technology that can detect the fraudulent transaction when it is taking place so that it can be stopped then and there and that too in a minimum cost. So the major task of today is to build an accurate, precise and fast detecting fraud detection system for credit card frauds that can detect not only frauds happening over the internet like phishing and site cloning but also tampering with the credit card itself i.e. it signals an alarm when the tampered credit card is being used. The biggest disadvantage is that the existing procedures

will not produce the same outcomes in various contexts. They produce superior outcomes with one type of dataset while producing poor or unacceptable results with others.

The HMM and Logic regression suffer some sort of performance degradation. But HMM is scalable for handling large volumes of transactions. Some techniques like clustering model and Naive Bayesian Network though have high detection rates and gives high accuracy they are very expensive to train. Some like KNN and HSVM gives excellent results with small datasets but are not scalable to large datasets. The problem of concept drift can be solved by using a Feedback mechanism. Some techniques like decision tree and support vector displaying the best accuracy and gives better results on sampled and pre-processed data. A solution to these gaps by creating a hybrid of various techniques that are already used in fraud detection to cancel out their limitations and get enhanced performance.

References

[1] D. Coppola, "E-commerce worldwide - statistics & facts," 2021. [Online]. Available: https://www.statista.com/topics/871/online-shopping/

[2] A. Srivastava, A. Kundu, S. Sural, and A. Majumdar, "Credit card fraud detection using Hidden Markov Model," *IEEE Transactions on Dependable & Secure Computing*, vol. 5, no. 1, pp. 37-48, 2008.

[3] V. Filippov, L. Mukhanov, and B. Shchukin, "Credit card fraud detection system," *IEEE International Conference on Cybernetic Intelligent Systems,* 9-10 Sep. 2008, pp. 1-6.

[4] J. O. Awoyemi, O. A. Adetunmbi, and A. Samuel, "Credit card fraud detection using machine learning techniques," *International Conference Computer Networking and Informatics (ICCNI)*, 2017, pp. 1-9.

[5] J. V. Devi, and K. S. Kavitha, "Fraud detection in credit card transactions by using classification algorithms," *International Conference on Current Trends in Computer, Electrical, Electronics and Communication (ICCTCEEC)*, 2017.

[6] V. Mareeswari, and G. Gunasekaran, "Prevention of credit card fraud detection based on HSVM," *International Conference on Information Communication and Embedded Systems (ICICES)*, 2016.

[7] C. Jiang, J. Song, and G. Liu, "Credit card fraud detection: A novel approach using aggregation strategy and feedback mechanism," *IEEE Internet of Things Journal*, 2018.