ISSN XXXX XXXX © 2017 IJESC

Research Article Volume 7 Issue No.6

Analyzing the Credit Card Fraud Detection using Data Mining Techniques

Deepika .N[1], Prof. Roopa .H[2]
M.Tech Student[1], Assistant Professor[2]
Department of ISE
Bangalore Institute of Technology, Bangalore, India

`

Abstract:
The credit card has become the most popular mode of payment for both online as well as regular purchase and therefore fraud associated with it are rising rapidly. Credit card frauds are increasing day by day regardless of the various techniques developed for its detection. As a result, fraud detection is the essential method and it is the best way to stop various fraud types. The classification models based on support vector machine (SVM) which is used to achieve neural network are developed and applied on credit card fraud detection problem and compares the performance of credit card details with a data set. Data is developing haphazardly for credit card and the k-means algorithm is used for discovering transaction weather it is fraud transaction or legitimate transaction.

Keywords: credit card fraud detection, support vector machine, ANN, association rules

I. INTRODUCTION

Credit Card Fraud is defined as, when an individual uses another individual credit card for personal use while the owner of the card as well as the card issuer are not aware of the thing that the card is being used. Fraud can be defined as wrongful or criminal deception intended to result in financial or personal gain, or to harm another person without essentially prompting direct lawful outcomes. The two main mechanisms to avoid frauds and losses due to fraudulent activities are fraud prevention and fraud detection systems. Fraud prevention is the proactive mechanism with the goal of disabling the occurrence of fraud. Fraud detection frameworks become possibly the most important factor when the fraudsters outperform the misrepresentation anticipation frameworks and begin a false exchange. No one can comprehend whether a fake exchange has passed the counteractive action components. Accordingly, the goal of the fraud detection systems is to check every transaction for the possibility of being fraudulent regardless of the prevention mechanisms, and to recognize fake ones as fast as conceivable after the fraudster has started to execute a fake exchange. The most well-known fraud types are fraudulent transactions in credit card systems and e-commerce systems, money laundering in financial systems, intrusions to computer systems, fraudulent calls or service usages in telecommunication systems, and fraudulent claims in health and auto insurance systems. Besides, with the upgrades in the data transfer capacity of internetworking channels, fraudsters have the opportunity to frame fraud systems among themselves through data change and joint effort everywhere throughout the world. Thus, fakes submitted over Internet, for example, online Master card fakes turn into the most mainstream ones in light of their temperament.. Credit card cheats can be made from multiple points of view, for example, basic robbery, application misrepresentation, fake cards, never received issue (NRI) and online fraud (where the card holder is absent). In online fraud, the exchange is made remotely and just the card's points of interest are required. A manual mark, a PIN or a card engraving are not required at the season of procurement. Though prevention mechanisms like CHIP&PIN decrease the fraudulent activities through simple theft, counterfeit cards and NRI; online frauds (Internet and mail order frauds) are still increasing in both amount and number of transactions. There has been a growing amount of financial losses due to credit card frauds as the usage of the credit cards become more and more common. Techniques used in fraud detection can be divided into two: Supervised techniques where past known legitimate/fraud cases are used to build a model which will produce a suspicion score for the new transactions; and unsupervised ones where there are no prior sets in which the state of the transactions are known to be fraud or legitimate. In this study, a credit card fraud detection system based SVM methods is developed. In this system, each account is monitored separately using suitable descriptors, and the transactions are attempt to be identified and flagged as legitimate or normal. The identification will be based on the suspicion score produced by the classifier models developed. When a new transaction is going, the classifier can predict whether the transaction is normal or fraud. Fraud detection discovery frameworks assess the exchanges and deliver a doubt score (for the most part a likelihood in the vicinity of 0 and 1) which demonstrates the likelihood of that exchange to be deceitful. Computational methodology of these scores are pertinent to the procedures used to assemble the model/models in the misrepresentation location frameworks. These scores are utilized with a predefined edge an incentive to separate the deceitful exchanges from the genuine ones. In any case, more often than not, these scores are not specifically utilized; but rather help the eyewitness staff with space aptitude who inspect and attempt to distinguish the fakes. Since the associations have restricted staff for this procedure, the capacity of the location frameworks to deliver precise doubt scores helps these staff from multiple points of view. By the by, the accomplishment of the identification frameworks lies in recognizing the false exchanges from honest to goodness ones through delivering doubt scores with high precisions. In this paper we propose a fraud detection method which is helpful to detect the fraud based on the customer behavior. For this data mining techniques attempted to detect the fraud. The rest of the paper is organized as follows – Section 2 describes the related work, Section 3 gives the methodology. Section 4 contains the experimental results and observations. The paper is concluded with conclusions and references.

II. RELATED WORKS

Duman.Ekrem et, al [1], a technique was evaluated such as genetic algorithm and scatter search to score each transaction which is based on these scores the transaction can be classified as fraudulent or genuine transaction and these approaches are based on the classification problem. Sahin.Yusuf et, al [2], the security mechanism such as CHIP and PIN are developed for credit card system that does not prevent from fraudulent credit card usages on online fraud and the author have developed and which is implemented a cost sensitive decision tree approach to detect fraudulent transactions and this approach is compared with the traditional classification models on a real world credit card data set. Dharwa.N.Jyotindra et, al [3], Transaction Risk Score Generation Method was used to calculate certainty factor to identify whether the transaction is fraudulent or genuine and Risk score is analyzed based on identification of spending profile of customer of a bank by implementing DBSCAN algorithm and address mismatch in which it will identify whether the customer billing and shipping is same. Farvaresh.Hamid et, al [4] , a framework was proposed to detect fraud telecommunication subscribers by using various techniques such as data cleaning, dimension reduction, clustering and classification and the main problem in this framework to determine customer is fraudster or genuine requires the historic data. Sanchez‟s. at, al [5], Association rules (Fuzzy Rules) are used to detect new, undesired behavior of bank customer in the online verification process and Association Rules (Fuzzy Rules) which are applied in the area of Business Management and from a large database planning to extract data of fraudulent transaction.

III. METHODOLOGY

Fraud detection is a binary classification task in which any transaction will be predicted and marked as a fraud or legit. In this paper the classification techniques are used for this task and their performances were compared.

1. Support vector machine: Support vector machines (SVMs) are statistical learning techniques which can be used in a classification tasks. This technique is based on the supervised learning algorithm. An SVM model is as focuses in space, and diverse focuses are mapped so that the different classifications are partitioned by a reasonable hole that is as wide as could be allowed. SVM is used to achieve neural network and used for comparing the data set based on the attributes.

2. Neural network[1]: Neural networks[NN] have been widely used in fraud detection. Neural network is a set of connected input/output units and each connection has a weight present with it. During the learning phase, network learns by adjusting weights to predict the correct class labels. Fraud detection methods based on neural network are the most popular ones. An artificial neural network consists of an interconnected group of artificial neurons .The pattern recognition and associative memory is the principal of neural network. The neural network recognizes similar patterns and predicts future values based on the associative memory of the patterns it was learned. It is mostly applied in classification and clustering. The advantages of neural networks over other techniques are that these models are able to learn from the past and thus, improve results as time passes. They can also extract rules and predict future activity based on the current situation. By employing neural networks, effectively, banks can detect fraudulent use of a card, faster and more efficiently. Among the reported credit card fraud studies most have focused on using neural networks. In more practical terms neural networks are nonlinear statistical data modelling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data. There are two phases in neural network training and recognition. Learning in a neural network is called training. There are two types of NN training methods supervised and unsupervised. In supervised training, samples of both fraudulent and non fraudulent records are used to create models. In contrast, unsupervised training simply seeks those transactions, which are most dissimilar from the norm. On other hand, the unsupervised techniques they doesn’t require the previous knowledge of fraudulent and non fraudulent transactions in database. NNs can produce best result for only large transaction dataset. And they want a long training dataset.

IV. FLOW CHART

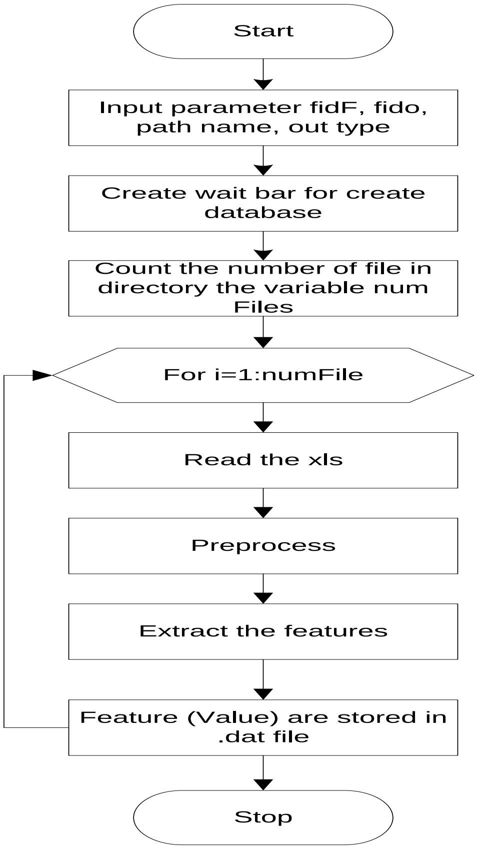
Start
Input parameter fidF, fido, path name, out type
Create wait bar for create database
Count the number of file in directory the variable num Files
For i=1:numFile
Read the xls
Preprocess
Extract the features
Feature (Value) are stored in .dat file
Stop

Figure.1. Feature extraction

The workflow is described briefly:

i) Input: Data collection has given as input. First , should collect the data and store complete data in database. Now, use stop words which removes the unwanted words such as, is, an, a, the, on, also and etc.

ii) Pre-processing: The dataset was pre-processed for the purpose of improving the performance of the classifiers and reducing their training and operating time. The pre-processing includes investigating the dataset feature space and handling the imbalanced nature of the dataset. It removes the natural language common text.

iii) Data transformation: Feature collection and feature extraction methods are used. SVM is used to achieve neural network and also used for comparing the data set based on attributes property, where we have trainer set and that set is compared with testing data set and gives good performance results. Figure 1 shows the feature extraction.

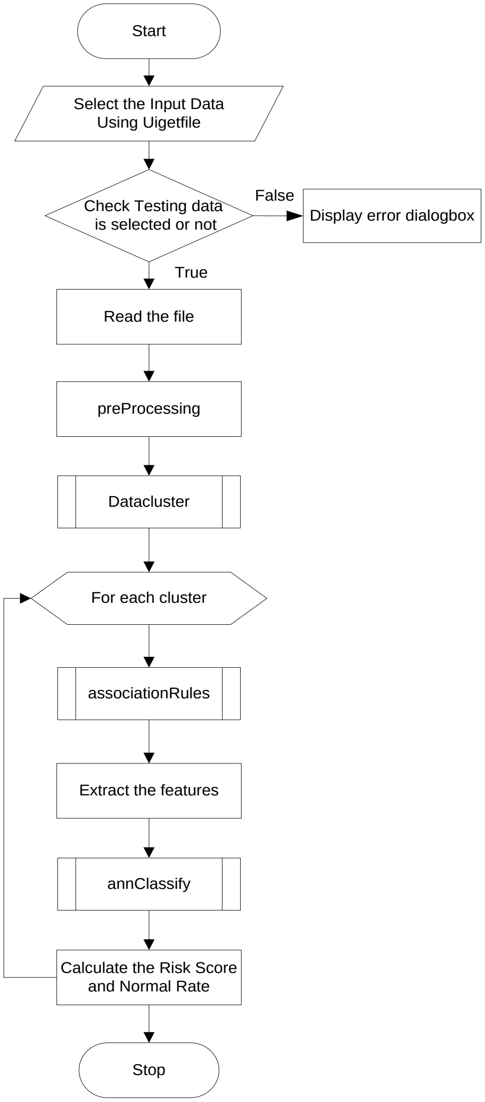

Figure.2. Select Query

iv) Mining: To mine the useful information association rules and clustering methods are used. Association rules are created by analyzing data for frequent pattern and identify relationships between data set. Clustering technique is used to cluster the data into different groups. The work is divided into n clusters to reduce the time complexity. Figure 2 shows the Select Query

v) Performance: Calculate mean score and normal rate for each cluster using association rules and an classify and results are shown in SVM.

The next section describes the experimental results.

V. EXPERIMENTAL RESULTS

The fraud detection models were trained and tested using data mining techniques and data pre-processing techniques. The data set is divided into subsets; one of them is used as the testing set and the others are used as the training set. This process is repeated taking a different subset as the testing set. The average performance results are then recorded. This methodological approach ensures that all data were represented once as a test data and several times as a training data producing accurate results SVM helps to achieve neural networks and used for comparing the data set based on attributes property, where we have trainer set and that set is compared with testing data set and gives good performance results.

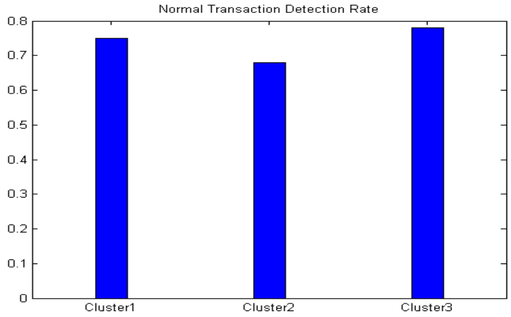

Figure.3. Normal transaction detection rate

Figure 3 shows normal transaction detection rate where x-axis is attributes and y-axis is time.

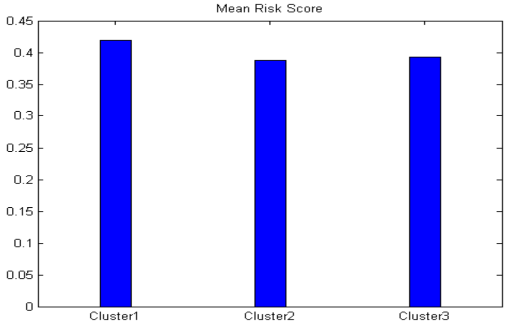

Figure. 4. Mean risk score

The fraud has been detected in mean risk score which shown in Figure 4.

VI. CONCLUSION

As more and more the financial and other data are digitized, the opportunities for payment card fraud increases. SVM is used to achieve neural network and also used for comparing the data set based on attributes property, where we have trainer set and that set is compared with testing data set which gives good performance results. The Data Mining algorithm KMEAN Clustering algorithm is actualized to distinguish fake exchanges in light of the spending conduct of a client. The performance results in ratios.

VII. REFERENCES

[1].Duman.E. and Ozcelik.H.M , “Detecting credit card fraud by genetic algorithm and scatter search”, Science Direct, Expert System with Applications 38 , pp 1305713063,2011.

[2]. Sahin.Y, Bulkan.S and Duman.E, “A cost-sensitive decision tree approach for fraud detection”, Science Direct, Expert System with Applications 40 pp-59165923, 2013

[3].Dharwa.J.N and Patel.A.R, “A Data Mining With Hybrid Approach Based Transaction Risk Score Generation Method for Fraud Detection of Online Transaction”, International Journal of Computer Applications Volume 16 No.1, pp 18-25, 2011.

[4].Farvaresh.H and Sepehri.M.M, “A data mining framework for detecting subscription fraud in telecommunication”, Science Direct, Engineering Applications of Artificial Intelligence 24, pp 182-194, 2010.

[5].Sanchez.D, Cerda.L, Serrano.J.M and Vila-.M.A, “Association Rules applied to Credit Card Fraud Detection”, Science Direct Expert System with applications 36 pp 3630-3640, 2009.

[6].Neha Bharill, Aruna Tiwari, Aayushi Malviya, "Fuzzy Based Scalable Clustering Algorithms for Handling Big Data Using Apache Spark", IEEE Transactions on Big Data, vol. 2, no. , pp. 339-352, Dec. 2016.