International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 5 | ISSN : 2456-3307

A Survey on Different Data Mining & Machine Learning Methods for Credit Card Fraud Detection

Vipul Patil[1], Dr. Umesh Kumar Lilhore[2]
[1]M. Tech Scholar Department of CSE NIIST Bhopal, Madhya Pradesh, India
[2]Head Department of CSE NIIST Bhopal, Madhya Pradesh, India

ABSTRACT

Due to rapid growth in field of cashless or digital transactions, credit cards are widely used in all around the world. Credit cards providers are issuing thousands of cards to their customers. Providers have to ensure all the credit card users should be genuine and real. Any mistake in issuing a card can be reason of financial crises. Due to rapid growth in cashless transaction, the chances of number of fraudulent transactions can also increasing. A Fraud transaction can be identified by analyzing various behaviors of credit card customers from previous transaction history datasets. If any deviation is noticed in spending behavior from available patterns, it is possibly of fraudulent transaction. Data mining and machine learning techniques are widely used in credit card fraud detection. In this survey paper we are presenting review of various data mining and machine learning methods which are widely used for credit card fraud detections.

Keywords: Data Mining, Machine Learning, Credit Card Fraud, Cashless Transactions.

I. INTRODUCTION

Due to a rapid advancement in the electronic commerce technology, use of credit cards has dramatically increased. As credit card becomes the most popular mode of payment, credit card frauds are becoming increasingly rampant in recent years [1]. In present scenario when the term fraud comes into a discussion, credit card fraud clicks to mind so far. With the great increase in credit card transactions, credit card fraud has increasing excessively in recent years.

Fraud detection includes monitoring of the spending behavior of users in order to determination, detection, or avoidance of undesirable behavior [4]. As credit card becomes the most prevailing mode of payment for both online as well as regular purchase, fraud relate with it are also accelerating. Fraud detection is concerned with not only capturing the fraudulent events, but also capturing of such activities as quickly as possible. The use of credit cards is common in modern day society [7]. Fraud is a millions dollar business and it is rising every year. Fraud presents significant cost to our economy worldwide. Modern techniques based on Data mining, Machine learning, Sequence Alignment, Fuzzy Logic, Genetic Programming, Artificial Intelligence etc., has been introduced for detecting credit card fraudulent transactions. In this review paper we are presenting analysis of various CC fraud detection methods based on various machine learning and data mining techniques [1, 4]. This complete paper is organized in various chapters includes introduction, existing work, existing fraud detection methods, challenges in existing methods and finally covers conclusions and future works.

II. EXISTING WORK

Credit card fraud detection requires data analysis and behaviors monitoring. In this review paper following existing research works used for analysis.

Kuldeep Randhawa et. al. 2018 worked on "Credit Card Fraud Detection Using AdaBoost and Majority Voting" [1]. As per author Credit card fraud is a serious problem in financial services. Billions of dollars are lost due to credit card fraud every year. There is a lack of research studies on analyzing real-world credit card data owing to confidentiality issues. In this paper, machine learning algorithms are used to detect credit card fraud. Standard models are first used. Then, hybrid methods which use AdaBoost and majority voting methods are applied. To evaluate the model efficacy, a publicly available credit card data set is used. Then, a real-world credit card data set from a financial institution is analyzed. In addition, noise is added to the data samples to further assess the robustness of the algorithms. The experimental results positively indicate that the majority voting method achieves good accuracy rates in detecting fraud cases in credit cards.

N Malini et. al. 2017 worked on "Analysis on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection". AS per authors [3] an efficient method of fraud detection has become a need for all banks in order to minimize chaos and bring order in place. There are several techniques like Machine learning, Genetic Programming, fuzzy logic, sequence alignment, etc are used for detecting credit card fraudulent transactions [2]. Along with these techniques, KNN algorithm and outlier detection methods are implemented to optimize the best solution for the fraud detection problem. These approaches are proved to minimize the false alarm rates and increase the fraud detection rate. Any of these methods can be implemented on bank credit card fraud detection system, to detect and prevent the fraudulent transaction.

John O. Awoyemi et. al. 2017 worked on, "Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis". As per authors [3] A hybrid technique of under-sampling and oversampling is carried out on the skewed data. The three techniques are applied on the raw and preprocessed data. The work implemented in Python. The performance of the techniques is evaluated based on accuracy, sensitivity, specificity, precision, Matthews's correlation coefficient and balanced classification rate. The comparative results show that k-nearest neighbour performs better than naïve bayes and logistic regression techniques.

B.Pushpalatha et. al. 2017, worked on" Credit Card Fraud Detection Based on the Transaction by Using Data mining Techniques". As per authors [4] the most common techniques used to make the fraud detection model. Incidentally, detection and prevention of credit card frauds are one of the vital problems in the digital world that need exact transactions analysis. One method for detecting fraud is to check for suspicious changes in user behavior. The purpose of this paper is to investigate Data mining techniques like Bayesian networks, Bayes Minimum Risk, Genetic algorithm, Hidden markov model (HMM) and Ontology for improve fraud detection in credit cards. This work primarily aims to improve current fraud detection processes by improving the prediction of fraudulent accounts.

You Dai et. al. 2016 worked on,"Online Credit Card Fraud Detection: A Hybrid Framework with Big Data Technologies". In this work [5] researchers focused on designing an online credit card fraud detection framework with big data technologies, by which achieved three major goals: 1) the ability to fuse multiple detection models to improve accuracy; 2) the ability to process large amount of data and 3) the ability to do the detection in real time. To accomplish that, we propose a general workflow, which satisfies most design ideas of current credit card fraud detection systems.

Nuno et. al. 2017, worked on" A data mining based system for credit-card fraud detection in e-tail". As per authors [6] a credit-card fraud leads to billions of

dollars in losses for online merchants. With the development of machine learning algorithms, researchers have been finding increasingly sophisticated ways to detect fraud, but practical implementations are rarely reported. The paper can thus help researchers and practitioners to design and implement data mining based systems for fraud detection or similar problems. This project has contributed not only with an automatic system, but also with insights to the fraud analysts for improving their manual revision process, which resulted in an overall superior performance.

III. CC FRAUD DETECTION METHODS

Following methods are widely used for CC fraud detection.

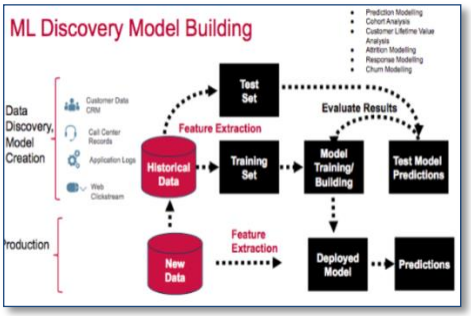

Figure 1. ML process in CC Fraud detection

3.1 Genetic algorithms-Algorithms are often recommended as predictive methods as a means of detecting fraud. One algorithm that has been suggested by Bentley et al. 2017 is based on genetic programming in order to establish logic rules capable of classifying credit card transactions into suspicious and non-suspicious classes. Basically, this method follows the scoring process [15]. In the experiment described in their study, the database was made of 4,000 transactions with 62 fields. As for the similarity tree, training and testing samples were employed. Different types of rules were tested with the different fields. The best rule is the one with the highest predictability. Their method has proven results for real home insurance data and could be one efficient method against credit card fraud [4].

3.2 Decision Tree- A decision tree is a graphical representation of possible solutions to a choice based on certain situations. Decision tree starts with root node, divides into separate branches, and these branches are connected with other nodes and so on. Decision tree end up in node called leaf node. Each node in Decision tree represents a test, a branch associated with it represents its possible results and a leaf node has a label of class. With this tactical approach of separating and deciding, decision tree usually isolate the complex problem into simple ones [6].

3.3 Artificial Neural Network-Artificial Neural Network (ANN) is one of the powerful classifiers to find out hidden pattern among different attributes. ANN works same as human's brain. ANN consists of different layers in which first layer is input layer and last layer is output layer. It may have number of hidden layer or no hidden layer. If neural network consist of more than one hidden layer, then it is deep learning. Each layer has different neurons, and each neuron is connected with weighted edges. Output of each neuron is a function of its unit. This function is called activation function. Example of different activation functions used is sigmoid function, step function, threshold function, linear function etc. Most used function is sigmoid function among all [10].

3.4 Convolution Neural Network (CNN)- Convolution Neural Network (CNN) is a part of deep learning. Mapping of input into hidden layer represents one feature map. Each feature map represents one characteristic. Process of compressing neurons into feature map is called convolution. Sub sampling reduces parameters of feature map. Fully connected layer is same as neural network [8].

3.5 Outlier Detection-Outliers are a basic form of non-standard attention that can be used for fraud detection. An observation that deviates much from other observations that arises suspicion that it was generated by a different mechanism is known as

outlier. Unsupervised learning approach is employed by this model. Generally, the result of unsupervised learning is a new explanation or representation of the observed data, which will then lead to improved future decisions. Unsupervised methods do not need the prior knowledge of fraudulent and non-fraudulent transactions in historical database, but instead unsupervised learning detect changes in behavior and unusual transactions [9,13].

3.6 Clustering techniques- Two clustering techniques have been suggested for behavioral fraud by Bolton & Hand (2002). Peer group analysis is a system that allows identifying accounts which are behaving differently from others at one moment in time whereas previously, they were behaving the same. These certain accounts are then flagged as suspicious. Then fraud analysts have been used to uncover those cases [11]. Hypothesis behind peer group analysis is that if accounts that were behaving the same for a certain period of time and then one account, still behaving significantly differently, then this account has to be notified. Another approach, Breakpoint analysis uses a different hypothesis which states that if a change of card usage is notified on an individual basis, the account must be investigated.

3.7 Logistic Regression- Data mining tasks has more and more statistical model that involves Discriminant analysis, regression analysis, multiple- logistic regression, etc. Logistic regression (LR) is useful for situations in which we want to be able to predict the presence or absence of a characteristic or outcome based on values of a set of predictor variables. It is similar to a linear regression model but is suited to models where the dependent variable is dichotomous [11].

3.8 Deep learning -Deep learning is the state of the art technology that recently attracted the IT circle's considerable attention. The deep learning principle is an ANN that has many hidden layers. Conversely, non-deep learning feed forward neural networks have only a single hidden layer.

3.9 Rule based method-Association Rules are generated to detect fraudulent transactions and normal transactions. In fraud detection, generated rules will be used to classify fraudulent and legitimate transactions. There for rules are generated as per behavior. This method is similar to decision tree.

3.10 Hidden Markov Model- Hidden Markov Model is a hybrid and embedded stochastic process for modeling. Complexity of this stochastic process is higher than common Markov model. If bank transaction is not accepted by learner [2, 8] Hidden Markov Model with high probability, it is considered as unsafe and fake transaction. Baum Welch algorithm is used for model learning and K-Means algorithm is used for data classification. The model classifies transactions in terms of high, average and low level.

IV. CHALLENGES

Based on survey of various CC fraud detection methods following challenges are identified which needs to resolve.

Some of the identified issues and challenges are as follows:

- **Typical classification problems:** CI and data mining-based financial fraud detection is subject to the same issues as other classification problems, such as feature selection, parameter tuning, and analysis of the problem domain [3].
- **Fraud types and detection methods:** Financial fraud is a diverse field and there has been a large imbalance in both fraud types and detection methods studied: some have been studied extensively while others, such as hybrid methods, have only been looked at superficially [8].
- **Privacy considerations:** Financial fraud is a sensitive topic and stakeholders are reluctant to

share information on the subject. This has led to experimental issues such as under sampling [5].

- **Computational performance:** As a high-cost problem it is desirable for financial fraud to be detected immediately. Very little research has been conducted on the computational performance of fraud detection methods for use in real-time situations [11].
- **Evolving problem:** Fraudsters are continually modifying their techniques to remain undetected. As such detection methods are required to be able to constantly adapt to new fraud techniques.
- **Disproportionate misclassification costs:** Fraud detection is primarily a classification problem with a vast difference in misclassification costs. Research on the performance of detection methods with respect to this factor is an area which needs further attention.
- **Generic framework:** Given that there are many varieties of fraud, a generic framework which can be applied to multiple fraud categories would be valuable.

V. CONCLUSIONS & FUTURE WORK

In the current paper, different methods of fraud detection in credit cards were investigated. Firstly, significance of the subject was stated and existing deficiencies in traditional systems were mentioned. Fake transactions have also varying degrees of risks and ways should be found for finding transactions with highest risk in quicker and more accurate manner. For identification of these transactions, common data mining methods alone do not suffice. Innovative algorithms should be used for finding the best answer.

In future work we will develop an efficient CC fraud detection method based on existing data mining and machine learning methods.

VI. REFERENCES

[1]. Kuldeep Randhawa, Chu kiong loo, manjeevan seera, chee peng lim, and asoke k. nandi," Credit Card Fraud Detection Using AdaBoost and Majority Voting", IEEE Access Volume 6, 2018, PP 14277-14285.

[2]. N. Malini, 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), IEEE, PP 978-985.

[3]. John O. Awoyemi, Adebayo O. Adetunmbi, Samuel A. Oluwadare,"Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis", 2nd International conference on New IT trends 2017, IEEE, PP 978-988.

[4]. B.Pushpalatha, C.Willson Joseph," Credit Card Fraud Detection Based on the Transaction by Using Data mining Techniques", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 2, February 2017, PP 1785-1794.

[5]. You Dai, Jin Yan, Xiaoxin Tang, Han Zhao and Minyi Guo," Online Credit Card Fraud Detection: A Hybrid Framework with Big Data Technologies", 2016 IEEE TrustCom/BigDataSE/ISPA, PP 1644-1653.

[6]. T Nuno Carneiroa, Gonc¸alo Figueiraa,∗, Miguel Costab," A data mining based system for credit-card fraud detection in e-tail", Journal of DSS, Sep 2016, PP 1-11

[7]. D. Viji, S. Kothbul Zeenath Banu,"An Improved Credit Card Fraud Detection Using K-Means Clustering Algorithm", International Journal of Engineering Science Invention (IJESI), One Day National Conference On"Internet Of Things The Current Trend In Connected World" NCIOT-2018, PP 59-64.

[8]. Shivangi Sharma,Puneet Mitta,Geetika,"An Approach to Detect Credit Card Frauds using Attribute Selection and Ensemble Techniques", international Journal of Computer Applications, Volume 180, No. 21, February 2018, PP 1-6.

[9]. S. K. Saravanan1, Dr. G. N. K. Suresh Babu," Literature Study Data Mining Techniques on Detecting Fradulent Activities in Credit Card", International Journal of Emerging Research in Management &Technology, Volume-6, Issue-10, PP 60-70.

[10]. R. Mankame, S. Nikam and A. Gurav,"Using Data Mining Detection of Fraud in Transaction", International Journal of Engineering research Online, Vol. 5, No.2, pp. 152-156, 2017.

[11]. S. Kumari and A. Choubey,"A Review on Various Techniques and Approaches for Credit Card Treachery Detection", International Journal of Scientific Research Engineering & Technology (IJSRET),Vol. 6, No. 5, pp. 485-489, 2017.

[12]. B. Pushpalatha and C.W. Joseph,"Credit Card Treachery Detection Based on the Transaction by Using Data mining Techniques", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, No. 2, pp. 1785-1793, 2017.

[13]. Deepika .N and Roopa .H"Analyzing the CC (credit card) Treachery Detection using Data Mining Techniques", IJESE, Vol. 7, No. 6, pp. 12851-12854, 2017.

[14]. M. Divya"Credit Card Treachery Detection Using HMM in Proposed Distributed Data Mining", International Journal of Advanced Research in Computer Science & Technology, Vol. 5, No. 1, pp. 49- 51, 2017. Ruchi Oberoi,"Credit – Card Fraud finding System: Using Genetic Algorithm", IJCMS, Vol. 6, No. 6, pp. 59-63, 2017.