A Survey on Credit Card Fraud Detection

Sampurna J[1], Rashmi R[2], K.S.Shwetha[3], Jitendranath Mungara[4]

B.E Students, Department of ISE, New Horizon College of Engineering, Bangalore, Karnataka, India

Asst. Professor, Department of ISE, New Horizon College of Engineering, Bangalore, Karnataka, India

HOD, Department of ISE, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT: Credit card fraud detection is a wide-ranging term for theft and fraud committed using or involving a payment card, such as a credit card or debit card, as a fraudulent source of funds in a transaction. Due to fast growth of E-Commerce, use of credit card for online purchases has dramatically increased and it caused and increase in the credit card fraud.
As credit card has become the most popular mode of payment for online and regular purchase, frauds associated with it are rising. In real life, fraudulent transactions are scattered with real transactions and simple pattern matching techniques are not often sufficient to detect those frauds accurately. Many techinques based on data mining, Random Forest, Machine Learning has involved in detecting various credit card fraudulent transactions.

KEYWORDS: Fraud Detection, Machine Learning, Random Forest, Spark, Semi-Supervised Learning Techniques.

I. INTRODUCTION

Credit card fraud detection is a relevant problem that draws the attention of machine-learning and computational intelligence communities, where a large number of automatic solutions have been proposed. In day to day life credit cards are used for purchasing goods and services with the help of virtual card for online transaction or physical card for offline transactions. In a physical-card based purchase, the card holder presents his card physically to a merchant for making the payment. The only way to detect this kind of fraud is to analyse the spending patterns on every card and to figure out any inconsistency with respect to the "usual" spending patterns. Fraud detection based on the analysis of existing purchase data of card holder is a promising way to reduce the rate of successful credit card frauds.

Data mining is the extraction of hidden predictive information from large databases and it is an important method for each activity of the credit card process. For example, it can be used for classifying good or bad customers based on their application information and also detecting a misuse of a credit card based on the purchase information of a customer. The capability of predicting goodness or badness of an applicant can reduce the credit card risk of a credit card issuer. However, if the issuer makes a wrong decision by issuing credit card to bad customers, it will result in the loss of revenue and liquidity. Data mining methodology is important for fraud detection models. This is due to its computation performance on large data.

Random Forest is the best performing model to find fraud or anomalies. To perform well it requires deep trees. A random forest use cleaver averaging tree to improve the predictive accuracy and also controls over-fitting. It is an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of classes or mean prediction of individual trees. It is also a statistical algorithm which is used to cluster points of data in functional groups.
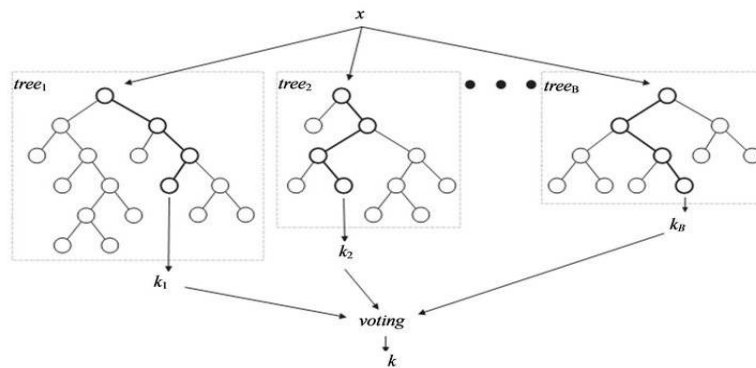
Fig 1: Random Forest.

Random Forest Pseudocode:

Step 1: Randomly select "k" features from total "m" features.
         Where k<<m
Step 2: Among the "k" features, calculate the node "d" using the best split point.
Step 3: Split the node into daughter nodes using the best split.
Step 4: Repeat 1 to 3 steps until "l" number of nodes has been reached.
Step 5: Build forest by repeating steps 1 to 4 for "n" number times to create "n" number of trees.

Machine Learning techniques have emerged as an alternative for the development of automatic fraud detection systems. It makes possible to take into account the wide range of ways in which fraud can be performed. It explores a study in construction of algorithms that can learn from and make predictions on data. It has 2 techniques:

1. Supervised Learning: The computer is presented with example inputs and their desired outputs given by a teacher and the goal is to learn a general rule that maps inputs to outputs.
2. Semi-Supervised Learning: The computer is given only an incomplete training signal. A training set with some of the target outputs missing.

Apache spark is an open source big data processing framework built around speed, ease of use and sophisticated analytics. Spark gives us a comprehensive, unified framework to manage big data processing requirements with a variety of data sets that are diverse in nature. It holds intermediate results in memory rather than writing them to disk which is very useful especially when you need to work on the same data set multiple times. It can also be used for processing the data sets that larger than aggregate memory in a cluster.

II. LITERATURE SURVEY

M. Behdad, L. Barone proposed, Electronic fraud, is highly lucrative, with estimates suggesting these crimes to be worth millions of dollars annually. Because of its complex nature, electronic fraud detection is typically impractical to solve without automation. However, the creation of automated systems to detect fraud is very difficult as adversaries readily adapt and change their fraudulent activities which are often lost in the magnitude of legitimate transactions. This study reviews the most popular types of electronic fraud and the existing nature inspired detection methods that are used for them. The common characteristics of electronic fraud are examined in detail along with the difficulties and challenges that these present to computational intelligence systems.

Ayushi Agrawal, Shiv Kumar, Amit Kumar Mishra, implemented, Credit card, is foremost well liked mode of payment. And fraudsters are increasing day by day. A good fraud detection system should be able to identify the fraudulent activities accurately and also as quickly as possible. Credit card fraud is security weakness of credit card companies. There are multiple approaches for fraud detection. To improve merchants' risk management level in an automatic and effective way, building an accurate and easy handling credit card risk monitoring system is one of the key tasks for the merchant banks. One aim of this study is to identify the user model that best identifies fraud cases. The models are compared in terms of their performances.

John O. Awoyemi, Adebayo O. Adetunmbi, Samuel A. Oluwadare, proposed This paper investigates the comparative performance of Naïve Bayes, K-nearest neighbour and Logistic regression `models in binary classification of imbalanced credit card fraud data. The rationale for investigating these three techniques is due to less comparison they have attracted in past literature. However, a subsequent study to compare other single and ensemble techniques using our approach is underway.

N.Malini, M.Phil student, Dr.M.Pushpa, proposed Credit card scam has become much more extensive. To progress safety measures of the monetary transaction systems in a habitual and effectual way, structure a precise and well organized credit card scam detection system is one of the essential functions for money transactions. By performing over sampling and extracting the principal direction of the data we can use our KNN method to determine the anomaly of the target instance. Hence the KNN method can suit for detecting fraud with the limitation of memory. By the mean time outlier detection mechanism helps to detect the credit card fraud using less memory and computation requirements. Especially outlier detection works fast and well on online large datasets. But compared with power methods and other known anomaly detection methods, experimental results prove that the KNN method is accurate and efficient.

Dr.R.Jayabrabu, Dr.J.Jebamalar Tamilselvi, Frauds are happened based on instance or incidents, but they are repeated offences using some methods (old and new),instances are more similar in content and appearance but they are non - identical while comparing. Fraud deduction is one of difficult process not only technology, but also in crime investigations. The traditional method of fraud detection is based on simple comparisons, but also based on association, good based on time complexity and heterogeneous databases, different syntax and needs domain experts. The proposed system addresses these issues with various parameters by considering a expert system names as automated fraud detection frame work to identify the fraud using intelligent agents, data fusion techniques and various data mining techniques. The agent properties are incorporated and implemented in this proposed system is one of the most added advantage and implementation of intelligent agents in data mining process is another advantage of this proposed system. The most important thing is finding the better result after mining process even though better techniques were considered for mining. These issues were also addressed in this paper using cluster quality index parameters. The proposed framework produced better result while compared to traditional fraud detection methods. Further, the proposed framework is going to consider some more techniques for fraud detection such as link analysis, Bayesian networks, decision theory, land sequence matching, etc. clustering, perdition and outlier detections is also addressed in this paper.

III. PROPOSED SYSTEM

A strategy based on a meta-classification approach that combines BRF and Co-Trained BRFs achieves the best performance.
A BRF based on the Spark RF model was implemented, in order to compensate the class imbalance of datasets and unlabeled samples. All strategies were implemented on apache Spark.
Both supervised and unsupervised methods have been proposed for credit card fraud detection purposes. Unsupervised methods consist of outlier/anomaly detection techniques that consider as a fraud any transaction that does not conform with the majority.

Supervised methods are by far the most popular in fraud detection, and exploit labelled transactions for training a classifier. Several studies have reported random forest (RF) to achieve the best performance.

In FDS we have described the alert–feedback interaction, which is the mechanism providing recent supervised samples to train/update the classifier. Feedbacks play a central role in the proposed learning strategy, which consists in separately training a classifier on feedbacks and a classifier on delayed supervised samples, and then aggregating their posteriors to identify alerts.

IV. CONCLUSION AND FUTURE WORK

A good fraud detection system should be able to identify the fraudulent activities accurately and also as quickly as possible. Credit card fraud is security weakness of credit card companies. There are multiple approaches for fraud detection. To improve merchants risk management level in an automatic and effective way, building an accurate and easy handling credit card risk monitoring system is one of the key tasks for the merchant banks. One aim of this study is to identify the user model that best identifies fraud cases. The models are compared in terms of their performances.

Several surrogate methods based on signal processing on graphs has been proposed for improving automatic credit card fraud detection. The proposed methods were applied to different scenarios for training of the detectors, considering several ratios of fraud operation number to legitimate operation number and surrogate proportions. The capabilities of proposed methods to improve detection performance were demonstrated using real data and measured by ROC curves and KPIs commonly used in financial business.

By performing over sampling and extracting the principle direction of data we can use our KNN method to determine the anomaly of the target instance. Hence the KNN method can suit for detecting fraud with the limitation of memory. But compared with power methods and other known anomaly detection methods, experimental methods prove that the KNN method is accurate and efficient.

REFERENCES

1. Mohammad Behdad, Luigi Barone, Mohammed Bennamoun, and Tim French. Nature-Inspired Techniques in the Context of Fraud Detection, IEEE TRANSACTIONS ON SYSTEMS,MAN,AND CYBERNETICS-PART C:APPLICATIONS AND REVIEWS,VOL.42,NO.6,NOVEMBER-2012.
2. Ayushi Agrawal,ShivKumar,Amit Kumar Mishra,M-Tech-Scholar, Department of Information Technology, Technocrats Institute of Technology, Bhopal, M.P.,2015 IEEE.Implementation of Novel Approach for Credit card fraud detection.978-9-3805-4416-8/15/$31.00_c 2015 IEEE.
3. John O. Awoyemi, Department of Computer Science, Federal University of TechnologyAkure, Nigeria, johntoba online Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis, **2017 IEEE**
4. N.Malini, M.Phil student, PG & Research department of Computer Science, Dr.M.Pushpa, Assistant Professor PG & Research department of Computer Science, Quaid-E-Millath Government College For Women(A), Annasalai, Chennai, India."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).978-1-5090-5434-3c 2017 IEEE.
5. Dr.R.Jayabrabu Associate Professor,Dr.N.G.P.Institute of Technology,Coimbatore,India,Dr.V.Saravanan Professor and Dean, Dept of Computer Application, Dr.J.Jebamalar Tamil selvi, Associate Professor, Dept of Computer Application, Chennai, India. A Framework for Fraud Detection System using Intelligent Agent for better decision making process.
6. Aihua Shen: Rencheng Tong: Yaochen Deng."Application of Classification Models On Credit Card Fraud Detection" Service Systems and Service Management, International Conference on Volume,Issue,9-11 June page(s):1-4,2007.