RESEARCH ARTICLE OPEN ACCESS

Credit Card Nearest Neighbor Based Outlier Detection Techniques

Mrs.C.Navamani MCA., M.Phil., M.E.,
(Assistant Professor Department of Computer Application, Nandha Engineering College/Anna University, Erode-52)
S.KRISHNAN
(Department of Computer Application, Nandha Engineering College/Anna University, Erode-52)

----------**********----------

Abstract:

Popular payment mode accepted both offline and online is credit card that provides cashless transaction. It is easy, convenient and trendy to make payments and other transactions. Credit card fraud is also growing along with the development in technology. It can also be said that economic fraud is drastically increasing in the global communication improvement. It is being recorded every year that the loss due to these fraudulent acts is billions of dollars. These activities are carried out so elegantly so it is similar to genuine transactions. Hence simple pattern related techniques and other less complex methods are really not going to work. Having 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. 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.

Keywords **— Classification, Fraud Detection, K-Nearest Neighbor Algorithm, Outlier Detection.**

----------**********----------

I. INTRODUCTION

Last few decades, technology has rapid growth, the e-commerce and an online expenditure have grown to such a large extent and people rely on it for most of their requirements. It has become a great boon to the modern world to carry out an effortless way of life. As the credit card gives a lot of expediency to the users, Frauds caused due to these are potentially hazardous and are even more. As our lives become increasingly digital, a growing amount of financial transactions are conducted online. Fraudsters have been quick to adapt to this trend, and to devise clever ways to defraud online payment systems. While this type of action involves illegal rings, a well-educated fraudster can create a very large number of artificial identities on his own, and use these to carry on sizeable schemes. New types of frauds are getting devised and hence the detection of frauds is flattering difficult. While taking ecommerce transactions the main problem that has been faced is that, the fraudulent transactions appears in a most cunning way as it looks similar as the legal one's. This puts many financial institutions and enterprises in problem. Hence the efficient way of fraud detection mechanism is very much compulsory rather than using simple classification techniques and the pattern matching techniques. The challenging part is to detect frauds in the highly unwarranted datasets were the legal transactions are on the maximum and the fraudulent transactions are about very less amount. The research papers about credit card fraud detection are very few and that is mainly because of the lack of publicly available datasets. This makes the researchers to have great difficulty in performing experiments. Since the credit card information is confidential, the bank owners and the other financial enterprise owners are not ready to disclose the credit card information's about their customers

because of the privacy concerns. Due to this, only fewer papers seems to be implemented and still there are some of the successful applications have been developed and evolved from numerous research communities, which are driven by KNN. Since the emergence of advanced computing and classification systems KNN shows a greater fluctuation in the **Web**
different technologies due to the accuracy and efficiency it produces. This project is to perform credit card fraud using a hybrid approach of KNN algorithm. KNN proves accurate in deducting fraudulent transaction and minimizing the number of false alert.

TYPES OF FRAUDS

Various types of frauds in this paper include credit card frauds, telecommunication frauds, and computer intrusions, Bankruptcy fraud, Theft fraud/counterfeit fraud, Application fraud, Behavioral fraud.

1) Credit Card Fraud: Credit card fraud has been divided into two types: Offline fraud and On-line fraud. Offline fraud is committed by using a stolen physical card at call center or any other place. On-line fraud is committed via internet, phone, shopping, web, or in absence of card holder..

2) Telecommunication Fraud: The use of telecommunication services to commit other forms of fraud. Consumers, businesses and communication service provider are the victims.

3) Computer Intrusion: Intrusion is defined as the act of entering without warrant or invitation; That means "potential possibility of unauthorized attempt to access Information, Manipulate Information Purposefully. Intruders may be from any environment, an outsider (Or Hacker) and an insider who knows the layout of the system.

4) Bankruptcy Fraud: This column focuses on bankruptcy fraud. Bankruptcy fraud means using a credit card while being absent. Bankruptcy fraud is one of the most complicated types of fraud to predict.

5) Theft Fraud/ Counterfeit Fraud: In this section, we focus on theft and counterfeit fraud, which are related to one other. Theft fraud refers using a card that is not yours. As soon as the owner give some feedback and contact the bank, the bank will take measures to check the thief as early as possible. Likewise, counterfeit fraud occurs when the credit card is used remotely; where only the credit card details are needed.

6)Application Fraud: When someone applies for a credit card with false information that is termed as application fraud. For detecting application fraud, two different situations have to be classified. When applications come from a same user with the same details, that is called duplicates, and when applications come from different individuals with similar details, that is termed as identity fraudsters.

II. RELATED WORK AND MOTIVATION

A. A.J. Graaff et al [1]

gave an overview of the natural immune system and presented the artificial immune system as a classifier between positive and negative patterns (self and non-self). Applications of the existing AIS models to detect network intrusion, virally infected files etc. were presented and the AIS as a fraudulent call detector was proposed. The results obtained from the AIS model to classify the Iris plant dataset was also presented. In theory, the AIS model seems to be a good detector for illegitimate patterns. For future work, the AIS model needs to be trained on legitimate calls and then tested with a set of calling patterns consisting out of both legitimate and illegitimate calling patterns.

B. Abhinav Srivastava et al [2]

Have used the ranges of transaction amount as the observation symbols, whereas the types of item have been considered to be states of the

HMM. We have suggested a method for finding the spending profile of cardholders, as well as application of this knowledge in deciding the value of observation symbols and initial estimate of the model parameters. It has also been explained how the HMM can detect whether an incoming transaction is fraudulent or not. Experimental results show the performance and effectiveness of our system and demonstrate the usefulness of learning the spending profile of the cardholders. Comparative studies reveal that the Accuracy of the system is close to 80 percent over a wide variation in the input data. The system is also scalable for handling large volumes of transactions.

C. *Divya.Iyer et al [3]*

Present the necessary theory to detect fraud in credit card transaction processing using a Hidden Markov Model (HMM). An HMM is initially trained with the normal behavior of a cardholder. If an incoming credit card transaction is not accepted by the trained HMM with sufficiently high probability, it is considered to be fraudulent. At the same time, we try to ensure that genuine transactions are not rejected by using an enhancement to it (Hybrid model). In further sections, we compare different methods for fraud detection and prove that why HMM is more preferred method than other methods.

D. *K. Ram Kalyani et al [4]*

Develop a method of generating test data and to detect fraudulent transaction with this algorithm. This algorithm is an optimization technique and evolutionary search based on the principles of genetic and natural selection, heuristic used to solve high complexity computational problems. This paper presents to find the detection of credit card fraud mechanism and examines the result based on the principles of this algorithm. The benefit of detecting fraud is to clear for both credit card companies and their clients. The fraudulent transactions are not prevented from being cleared; the company must accept the financial cost of that transaction. This reduces the cost associated with higher interest rates, and its charges.

E. *Renu et al [5]*

Fraud detection involves monitoring the activities of populations of users in order to estimate, perceive or avoid undesirable behavior. Undesirable behavior is a broad term including delinquency, fraud, intrusion, and account defaulting. This paper presents a survey of current techniques used in credit card fraud detection and telecommunication fraud. The goal of this paper is to provide a comprehensive review of different techniques to detect fraud

F. *Problem statement*

Classification scheme usually uses one of the following approaches: statistical and syntactic. Statistical pattern classification is based on statistical characterizations of patterns, assuming that the patterns are generated by a probabilistic system. Structural pattern recognition is based on the structural interrelationships of features. A wide range of algorithms can be applied for pattern recognition, from very simple Bayesian classifiers to much more powerful neural networks.

Supervised classification is a technique in which we identify examples of information classes of interest within the dataset known as training data. Statistical characterization of each of the information class is obtained. Once a statistical characterization has been achieved for each information class, the test data is then classified by examining the best possible match against the training data and making an informed decision about the class it resembles most.

Unsupervised classification is a method which examines a large number of datasets and divides into a number of classes based on natural groupings present within the dataset. Unlike supervised classification, unsupervised classification does not require a priori labeling of training data. The basic phenomenon is that data within a given class should be as close as possible in the measurement space, whereas data in different classes should be comparatively well separated.

Although classification depends on the application and the information available from that problem, still research is going on to generalize these techniques and to conclude which technique is suitable for what kind of problems. The main objective of this paper is to review, implement K- Nearest Neighbor (KNN) and investigating the performance of K- Nearest Neighbor (KNN) classifier based on an application "Credit card approval", and also to show how classification rate improves in KNN by varying the value of k.

III. OUR CONTRIBUTION

Due to a rapid improvement in the electronic commerce technology, the utilize of credit cards has augmented. As credit card becomes the trendiest style of payment for individually online as well as habitual acquisition, luggage of credit card fraud also growing. Economic fraud is increasing radically with the development of modern technology and the global super highways of communication, consequential in the loss of billions of dollars worldwide each year. The falsified transactions are sprinkled with genuine transactions and simple pattern corresponding techniques are not often sufficient to detect those frauds accurately. 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 K-Nearest Neighbor (KNN) algorithm is an evolutionary search and optimization technique that Mimics natural evolution to find the best solution to a problem. Here the characteristics of credit card transactions undergo evolution to allow a modeled credit card fraud detection system to be tested. This method proves accurate in deducting fraudulent transaction and minimizing the number of false alert. If this algorithm is applied into bank credit card fraud detection system, the probability of fraud transactions can be predicted soon after credit card transactions.

A. Spending history databases

It comprises of genuine Transaction Record (for individual customers from their past behavior) and Fraud Transaction Record (from different types of past fraud data). We represent each history transaction by set of attributes containing information like card number, transaction amount and time since last purchase. to extract characteristic information about genuine and fraud transactions.

B. Behavior based Model

The key concept in fraud detection is to analyze the spending behaviors of the user. If any discrepancies occur with the respect to the usual spending behavior, then it is considered as suspicious behavior. And it is taken for further consideration. The behavior of spending varies from person to person. Fraud detection based on the analysis of existing spending behavior of cardholder is a promising way to find the credit card frauds. Behavior based fraud detection model means that the data use in the model are from the transactional behavior of cardholder directly or derived from them. Each person may have a different spending behavior pattern. Most of the existing fraud detection methods use the behavior pattern as measure to find the destruction in the transactions. Based on the spending pattern the customer's usual activities such as transaction amount, billing address etc. are learned. Some of the count measures to suspect the behaviors are the variation of billing address and shipping address, maximum amount of purchase, large transaction done far away from the living place etc. Like that the behaviors deviate from the normal ones are suspected and taken for further consideration

C. Address matching

In this method, an incoming transaction is checked for the address mismatch. If shipping address and billing address is found same, then the transaction is considered to be genuine and is approved else the transaction is fraudulent. But

there may be a case that the real customer has ordered this product for different address and may be the customer is giving a gift to his friend.

- ✓ These cases show that the customer behavior changes constantly. So, it is difficult to identify customer behavior.
- ✓ Real World Data consists of a large number of errors in a data. In order to identify whether the transaction is fraudulent or genuine the data must be free from errors.

So, the main objective is to identify fraudster behavior and to propose a security mechanism that should be able to adapt themselves to detect new kinds of fraud and to improve the quality of data by removing the errors or noise in the data by using data mining technique.

D. P – Location

Here the characteristics of credit card transactions undergo evolution to allow a modeled credit card fraud detection system to be tested. This method proves accurate in deducting fraudulent transaction and minimizing the number of false alert. If this algorithm is applied into bank credit card fraud detection system; the probability of fraud transactions can be predicted soon after credit card transactions. An outlier is an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism. Unsupervised learning approach is employed to this model. Usually, the result of unsupervised learning is a new explanation or representation of the observation data, which will then lead to improved future responses or decisions. Unsupervised methods do not need the prior knowledge of fraudulent and non-fraudulent transactions in historical database, but instead detect changes in behavior or unusual transactions. These methods model a baseline distribution that represents normal behavior and then detect observations that show greatest departure from this norm. Outliers are a basic form of non-standard observation that can be used for fraud detection. In supervised methods, models are trained to discriminate between fraudulent and non-fraudulent behavior so that new observations can be assigned to classes. Supervised methods require accurate identification of fraudulent transactions in historical databases and can only be used to detect frauds of a type that have previously occurred. An advantage of using unsupervised methods over supervised methods is that previously undiscovered types of fraud may be detected.

E. K-Nearest Neighbor Algorithm

The concept of nearest neighbor analysis has been used in several anomaly detection techniques. One of the best classifier algorithms that have been used in the credit card fraud detection is k-nearest neighbor algorithm that is a supervised learning algorithm where the result of new instance query is classified based on majority of K-Nearest Neighbor category. It was first introduced by Aha, Kibler, and Albert (1991) The performance of KNN algorithm is influenced by three main factors:

- The distance metric used to locate the nearest neighbors.
- The distance rule used to derive a classification from k-nearest neighbor.
- The number of neighbors used to classify the new sample.

Among the various credit card fraud detection methods of supervised statistical pattern recognition, the K Nearest Neighbor rule achieves consistently high performance, without a priori assumptions about the distributions from which the training examples are drawn. K-Nearest neighbor based credit card fraud detection techniques require a distance or similar the measure defined between two data instances. In process of KNN, we classify any incoming transaction by calculating of nearest point to new incoming transaction. Then if the nearest neighbor be fraudulent, then the transaction indicates as a fraud. The value of K is used as, a small and odd to break the ties (typically 1, 3 or Larger K values can help to reduce the effect of

noisy data set. Inthis algorithm, distance between two data instances can be calculated in different ways. For continuous attributes, Euclidean distance is a good choice. For categorical attributes, a simple matching coefficient is often used. For multivariate data, distances usually calculated for each attribute and then combined. The performance of KNN algorithm can be improved by optimizing the distance metric. This technique required legitimate as well as fraudulent samples of data for training. It is fast technique along with high false alert.

ALGORITHM: KNN

INPUT: DS, current window size N, integer k, query time U query, number of outlier m
OUTPUT: m outliers
METHOD:
BEGIN
SM (DS, N, k);
when (U query) QM(m);
END

Stream Manager Procedure

PROCEDURE SM (DS, N, k)
BEGIN
Step 1: FOR each data stream object obj with arrival time t DO
Step 2: IF the oldest object q of current window eExpires S 3: FOR all objects 0 III q. knn list DO o. rknn list delete(q);
Step 4: FOR all objects 0 in q. rknn list DO
o. knn list delete(q);
Step 5: ENDIF
Step 6 : remove object q from current window
Step 7 : object p (obj, t, cD, cD);
Step 8 : FOR all objects 0 in current window DO
Step 9 : dis t= distance(p);
Step10 : p.knn list insert (0); list k nearest neighbor of p
Step 11: o. rknn list insert(p);
Step 12: IF dis<=o. distance
Step 13: o. knn list insert(p);
Step 14: p. rknn list insert (0);
Step 15: ENDIF
Step 16: END FOR
Step 17: Insert object p into current window.
Step 18: END FOR
END

Outlier Query Management Procedure

PROCEDURE QM(m)
BEGIN
Step 1: Perform a single scan of current window;
Step 2: Return m objects with minimal I

RNNk(p) I as Outliers
END

CONCLUSION & FUTURE WORK

A novel data stream outlier detection algorithm KNN is presented. This algorithm reduces the number of scans to only one. Experiments conducted on both synthetic and real data sets show that the proposed method is efficient and effective. At the end of the transaction process, it also notifies cardholder and the merchant about the transaction status. Finally, it will alert a mail to the user, fraudulent transaction as try to make. Our aim is to detecting the presence of outliers from a large amount of data via an online updating technique. Our proposed framework is favored for online applications which have computation or memory limitations. Compared with the well-known power method and other popular anomaly detection algorithms, our experimental results verify the feasibility of our proposed method in terms of both accuracy and efficiency.
A Fraud Detection System that combines both these methods will be more robust and efficient in identifying various types of credit card frauds. Mobile OTP is use to generate 8-digit unique security code. In proposed model, we have use Mobile OTP to provide more security and to reduce the fraud.

REFERENCES

[1]. A.J. Graaff A.P. Engelbrecht agraaff "The Artificial Immune System for Fraud Detection in the Telecommunications Environment" Telkom SA Ltd., Pretoria 0001, South Africa, 20 November 2014.

[2] Abhinav Srivastava, Amlan Kundu, Shamik Sural, and Arun K. Majumdar" Credit Card Fraud Detection Using Hidden Markov Model" IEEE Transactions on Dependable and Secure Computing Vol. 5, No. 37, January-March 2008

[3] Divya.Iyer, Arti Mohanpurkar, Sneha Janardhan, Dhanashree Rathod, Amruta Sar Deshmukh" Credit Card Fraud Detection Using Hidden Markov Model" IEEE, 978-1-4673-0126-8/11/$26.00_c 2011

[4] K. Ram Kalyani, D. Uma Devi" Fraud Detection of Credit Card Payment System by Genetic Algorithm" International Journal of Scientific & Engineering Research Volume 3, Issue 7, July-2012 1 ISSN 2229-5518

[5] Renu, Suman" Analysis on Credit Card Fraud Detection Methods"International Journal of Computer Trends and Technology (IJCTT) – volume 8 number 1– Feb 2014

[6] EkremDuman, M. Hamdi Ozcelik "Detecting credit card fraud by genetic algorithm and scatter search". Elsevier, Expert Systems with Applications, (2011). 38; (13057–13063).

[7] S. Benson Edwin Raj, A. Annie Portia, "Analysis on Credit Card Fraud Detection Methods", International Conference on Computer, Communication and Electrical Technology – ICCCET 2011, 18th & 19th March, 2011. (152-156).

[8] Y. Sahin and E. Duman, "Detecting Credit Card Fraud by Decision Trees and Support Vector Machines", International Multiconference of Engineers and computer scientists' volume 1, March,

[9] FahimehGhobadi, Mohsen Rohani, "Cost sensitive Modeling of Credit Card Fraud Using Neural Network strategy", ICSPIS 2016, 14-15 Dec 2016, Amirkabir University of Technology Tehran, Iran.

[10] NabhaKshirsagar, Neha Pandey, Shraddha kotkar," Credit card Fraud Detection System using Hidden Markov Model and Adaptive Communal Detection", International Journal of Computer Science and Information Technologies, vol 6 (2), 2015.

[11] A.A. El Masri and J.P. Sousa, "Limiting Private Data Exposure in Online transactions: User-Based Online Privacy Assurance Model", in International Conference on Computational Science and Engineering, CSE '09, vol. 3, pp. 438 – 443, 2009.