FRAUD DETECTION FOR CREDIT CARD TRANSACTIONS USING RANDOM FOREST ALGORITHM

Dr.P.Indira Priyadarshini[1]

Faculty, AVN Institute of Engineering and Technology, Hyderabad

S.Ramesh Babu[2]

UG Scolor AVN Institute of Engineering and Technology, Hyderabad

Govind.D[2]

UG Scolor AVN Institute of Engineering and Technology, Hyderabad

Sandeep.A[2]

UG Scolor AVN Institute of Engineering and Technology, Hyderabad

Manish.K[2]

UG Scolor AVN Institute of Engineering and Technology, Hyderabad

Siddartha.G[2]

UG Scolor AVN Institute of Engineering and Technology, Hyderabad

Abstract

In these days, credit card fraud detection is a major concern in the society. The use of credit cards in e-commerce sites and various banking sites has been increased rapidly in recent times. As modernization will have both positive and negative impacts, the use of credit cards in online transactions has made it simple; likewise, it also led to the increase of the number of fraud transactions. As part of the activities happening, it is always advised for the e-commerce sites and the banks to have automatic fraud detection system. Credit card fraud might result in huge financial losses. While look for the solutions for credit card frauds that are happening, machine learning techniques provide favorable solutions. The proposed system uses a random forest application in solving the problem and to attain more accuracy when compared to the other algorithms used till now. All the basic classifiers have the same weight but random forest algorithm has relatively high and others have relatively low weights because of the randomization of bootstrap sampling of a making decision and selection of attributes cannot guarantee that all of them have the same stability in decision making.

KEYWORDS: Decision tree · Fraud detection · Random forest

1. INTRODUCTION

In our everyday life, various transactions are done through credit card payments, cardless transactions like Google Pay, PhonePe, Samsung Pay, and PayPal. There is an ongoing concern in recent days which is fraud detection, and it leading to the great loss of money every year. If the fraud continues this way, it is said that by the year 2020, it will reach double digits. Nowadays, the presence of the card isn't physically required to finish the exchange which is prompting increasingly more extortion exchanges [1]. Fraud detection has an emotional impact on the economy. In this way, fraud detection is fundamental and vital. Financial institutions have to employ various fraud detection techniques for tackling this problem [2, 3]. But when given time the fraudsters find ways to overcome the techniques established by the company holders. Despite all the preventive methods taken by the financial institutions and strengthening of law and government putting their best efforts to eradicate fraud detection, fraud detection continues to rise and it remains as a major concern in

the society [4, 5]. Credit cards are generally utilized in the improvement of the Internet business and furthermore portable applications and primarily in the online-based exchanges. With the help of the credit card, the online transactions and online payment are easier and convenient for usage [6, 7]. Fraud transactions have a great influence on enterprises [8]. Machine learning techniques have been widely used, and it has become very important in many areas where spam classifiers protect our mail id. The fraud detection systems learn the features of extraction and helps in controlling the fraud detection.

2. LITERATURE REVIEW

[1] The Use of Predictive Analytics Technology to Detect Credit Card Fraud in Canada. "Kosemani Temitayo Hafiz, Dr. Shaun Aghili, Dr. Pavol Zavarsky." This research paper focuses on the creation of a scorecard from relevant evaluation criteria, features, and capabilities of predictive analytics vendor solutions currently being used to detect credit card fraud. The scorecard provides a side-byside comparison of five credit card predictive analytics vendor solutions adopted in Canada. From the ensuing research findings, a list of credit card fraud PAT vendor solution challenges, risks, and limitations was outlined. [2] BLAST-SSAHA Hybridization for Credit Card Fraud Detection. "Amlan Kundu, Suvasini Panigrahi, Shamik Sural, Senior Member, IEEE, and Arun K. Majumdar" This paper propose to use two-stage sequence alignment in which a profile Analyser (PA) first determines the similarity of an incoming sequence of transactions on a given credit card with the genuine cardholder's past spending sequences. The unusual transactions traced by the profile analyser are next passed on to a deviation analyser (DA) for possible alignment with past fraudulent behaviour. The final decision about the nature of a transaction is taken on the basis of the observations by these two analysers. In order to achieve online response time for both PA and DA, we suggest a new approach for combining two sequence alignment algorithms BLAST and SSAHA. [3] Research on Credit Card Fraud Detection Model Based on Distance Sum. "Wen-Fang YU, Na Wang". Along with increasing credit cards and growing trade volume in China, credit card fraud rises sharply. How to enhance the detection and prevention of credit card fraud becomes the focus of risk control of banks. It proposes a credit card fraud detection model using outlier detection based on distance sum according to the infrequency and unconventionality of fraud in credit card transaction data, applying outlier mining into credit card fraud detection. Experiments show that this model is feasible and accurate in detecting credit card fraud. [4] Fraudulent Detection in Credit Card System Using SVM & Decision Tree. "Vijayshree B. Nipane, Poonam S. Kalinge, Dipali Vidhate, Kunal War, Bhagyashree P. Deshpande". With growing advancement in the electronic commerce field, fraud is spreading all over the world, causing major financial losses. In current scenario, Major cause of financial losses is credit card fraud; it not only affects trades person but also individual clients. Decision tree, Genetic algorithm, Meta learning strategy, neural network, HMM are the presented methods used to detect credit card frauds. In contemplate system for fraudulent detection, artificial intelligence concept of Support Vector Machine (SVM) & decision tree is being used to solve the problem. Thus by implementation of this hybrid approach, financial losses can be reduced to greater extend. 5] Supervised Machine (SVM) Learning for Credit Card Fraud Detection. "Sitaram patel, Sunita Gond". This thesis propose the SVM (Support Vector Machine) based method with multiple kernel involvement which also includes several fields of user profile instead of only spending profile. The simulation result shows improvement in TP (true positive), TN (true negative) rate, & also decreases the FP (false positive) & FN (false negative) rate. [6] Detecting Credit Card Fraud by Decision Trees and Support Vector Machines. "Y. Sahin and E. Duman" In this study, classification models based on decision trees and support vector machines (SVM) are developed and applied on credit card fraud detection problem. This study is one of the firsts to compare the performance of SVM and decision tree methods in credit card fraud detection with a real data set

3. EXISTING SYSTEM

In existing System, a research about a case study involving credit card fraud detection, where data normalization is applied before Cluster Analysis and with results obtained from the use of Cluster Analysis and Artificial Neural Networks on fraud detection has shown that by clustering attributes neuronal inputs can be minimized. And promising results can be obtained by using normalized data and data should be MLP trained. This research was based on unsupervised learning. Significance of this paper was to find new methods for fraud detection and to increase the accuracy of results. The data set for this paper is based on real life transactional data by a large European company and personal details in data is kept confidential. Accuracy of

an algorithm is around 50%. Significance of this paper was to find an algorithm and to reduce the cost measure. The result obtained was by 23% and the algorithm they find was Bayes minimum risk.

DISADVANTAGES

1. In this paper a new collative comparison measure that reasonably represents the gains and losses due to fraud detection is proposed.

2. A cost sensitive method which is based on Bayes minimum risk is presented using the proposed cost measure.

4. PROPOSED SCHEME

In proposed System, we are applying random forest algorithm for classification of the credit card dataset. Random Forest is an algorithm for classification and regression. Summarily, it is a collection of decision tree classifiers. Random forest has advantage over decision tree as it corrects the habit of over fitting to their training set. A subset of the training set is sampled randomly so that to train each individual tree and then a decision tree is built; each node then splits on a feature selected from a random subset of the full feature set. Even for large data sets with many features and data instances training is extremely fast in random forest and because each tree is trained independently of the others. The Random Forest algorithm has been found to provide a good estimate of the generalization error and to be resistant to over fitting.

ADVANTAGES OF PROPOSED SYSTEM

- Random forest ranks the importance of variables in a regression or classification problem in a natural way can be done by Random Forest.
- The 'amount' feature is the transaction amount. Feature 'class' is the target class for the binary classification and it takes value 1 for positive case (fraud) and 0 for negative case (not fraud).

IMPLEMENTATION

SERVICE PROVIDER

In this module, the sp has to login by using valid user name and password. After login successful he can do some operations such as Upload Products, View all product details, view rating results, view dislikes, view rank results, view all remote users, View all Products reviews, View all trending products, View products recommended, View Products Purchased, View purchased status, View Credit Card Fraud Detections.

USER

In this module, there are n numbers of users are present. User should register before doing some operations. After registration successful he has to login by using authorized user name and password. Login successful he will do some operations like Search Products, Give rate, rank, review, dislikes ,View all Products reviews ,View Trending products, View Your Profile, View Recommended Products, View Collusion Sellers, View Account details View Fraud Sellers

ARCHITECTURE

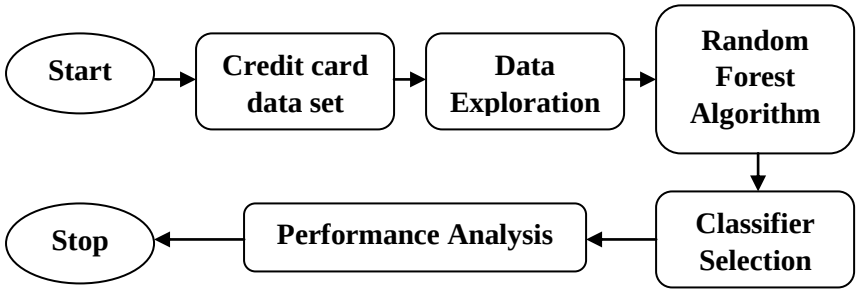

Figure: 1 Architecture diagram of the proposed system

THE PROPOSED RANDOM FOREST TECHNIQUE

Random forest algorithm first delivers a forest, and it makes them randomly; the forest is built as a collaborative decision tree which is otherwise known as the bagging method. The algorithm is dependent upon the quality of the discrete trees and furthermore the correlation between various trees. It shows the distinction of various information factors that simultaneously permit a high number of considerations to contribute to prediction. The algorithm works well with an insignificant amount of data. Aggregating un-correlated trees are the main concept that makes the random forest algorithm better than the decision trees. The main idea is to create several model trees and make an average of these trees to create a better random forest.

IMPLEMENTATION OF RANDOM FOREST ALGORITHM

The procedure of the random forest algorithm execution is done in the below diagram. As shown in Fig. 1, there are several steps involved; first, there is a requirement to gather the information and to store the information. The gathered information are in the form of data set in an excel sheet. In data exploration, the entire data set is checked and removed the unnecessary data that is present. However, the data which is preprocessed is further treated using a random forest algorithm in two ways by using the train data set and then using the test data set. The attained results are verified as legal and fraudulent transaction process.

COLLECTION OF DATA SETS

Here, the first step is to collect the data sets. The data sets can be collected from various methods like crawling or application program interface. The data sets must contain attributes like name of the customer, customer email address, card number, payment method, customer mobile number, bank account number, and pin number. After collecting the data sets from the above attributes, the data set is used for performing analysis. The primary distinction between the training data set and the test data set is that the training data set is labeled; however, the test set is unlabeled. So at first, the data set is trained by the regression analysis and then it is been tested by the random forest algorithm.

ANALYSIS OF DATA

After preparing the data, analysis is performed using various algorithms. This data has a set of functions for training the data and creating classification predictive models. Random forest algorithm is used for grouping the data sets, and it is divided into training set and rest as test sets. It is composed of various tools such as splitting of data and data preprocessing; which is done by using the resampling method.

REPORTING RESULTS

After the above stages, the complete analysis of data is done and the results are produced by using random forest algorithm. Hence, the random forest algorithm is performed by the classification to obtain the results. The result is to gain more accuracy in fraud detection. Random forest algorithm has nearly the same hyper parameters as a decision tree or a bagging classifier. Fortunately, it is not required to combine a decision tree with a bagging classifier. Random forest algorithm also deals with regression tasks. The algorithm adds additional randomness to the model while growing the trees. Therefore, only a random subset of the features is taken into consideration by the algorithm for splitting a node. Instead of searching for the most important feature node, it searches for the best feature among the random subset of features by using random thresholds. This results in a wide diversity and yield a better model. The structure of the random forest tree is portrayed in Fig. 2. The acquired information is classified from the data set and represented as in the form of attributes such as card number, card limit, and personal information. The information is presented in the form of a matrix to decide the sample belongs to which of the decision tree. This process requires a large number of trees to create similar trees so as to provide various decision trees. By analyzing all the trees present in the graph, which tree gets the most number of relevant solutions, are identified. Based on that conclusion, the decision tree is chosen.

5. CONCLUSION

The proposed unsupervised random forest algorithm reduces the number of fraud transactions. There are several experiments performed using random forest algorithm. The obtained results ensured that the number

of fraud transactions is greatly reduced. This improves more secure transactions through online and makes the system more accurate

REFERENCES

[1] U. Fiore, A. De Santis, F. Perla, P. Zanetti, F. Palmieri, Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. Inf. Sci. 479, 448–455 (2019)
[2] N. Carneiro, G. Figueira, M. Costa, A data mining based system for credit-card fraud detection in e-tail. Decis. Supp. Syst. 95, 91–101 (2017)
[3] A.C. Bahnsen, D. Aouada, A. Stojanovic, B. Ottersten, Feature engineering strategies for credit card fraud detection. Exp. Syst. Appl. 51, 134–142 (2016)
[4] M. Zareapoor, P. Shamsolmoali, Application of credit card fraud detection: based on bagging and ensemble classifier. Proc. Comput. Sci. 48, 679–685 (2015)
[5] K. Randhawa, C.K. Loo, M. Seera, C.P. Lim, A.K. Nandi, Credit card fraud detection using AdaBoost and majority voting. IEEE Access 6, 14277–14284 (2017)
[6] P. Save, P. Tiwarekar, K.N. Jain, N. Mahyavanshi, A novel idea for credit card fraud detection using decision tree. Int. J. Comput. Appl. 161(13), 0975–8887 (2017)
[7] S. Sorournejad, Z. Zojaji, R.E. Atani, A.H. Monadjemi, A survey of credit card fraud detection techniques: data and technique oriented perspective. ArXiv (2016)
[8] A. Singh, A. Jain, Adaptive credit card fraud detection techniques based on feature selection method. Adv. Comput. Commun. Comput. Sci., 167–178 (2019)
[9] Z. Li, G. Liu, S.Wang, S. Xuan, C. Jiang, Credit card fraud detection via kernel-based supervised hashing, in 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (2018)
[10] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, C. Jiang, Random forest for credit card fraud detection, in 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC) (2018)

