International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 05 | May-2018 www.irjet.net p-ISSN: 2395-0072

Hypocrisy Detection in E-Transaction using Distributed Data Mining Technique

Rithesh Pakkala P[1], Ravi Yadav[2], Pooja B[3], Rahin Shama[4], Vaishak Chandra K S[5]

[1]*Assistant Professor, Department of Information Science & Engineering, Sahyadri College of Engineering & Management, Mangaluru, Karnataka, India,*
[2,3,4,5] *Final Year BE Students, Department of Information Science & Engineering, Sahyadri College of Engineering & Management, Mangaluru, Karnataka, India.*

Abstract - *These days there has been a massive increase in the usage of e-transactions in the past few years. There is a larger increase in e-transactions day by day due to the ever increasing online traders such as EBay, Wal-Mart and Amazon etc. The customers will usually make use of online payment services in order to fulfill their purchasing requirements. Nowadays credit and debit card is more frequently used for both your online as well as the regular purchases. Since there are large numbers of e-transactions, the dataset associated with these transactions is also larger. In order to find these hypocritic transactions we need to make use of efficient algorithms. We can thus notice a large number of hypocritic cases that are coming up these days, due to billions of transactions taking place each day. Most of the hypocrites are identified by making use of rule-based patterns so that when a particular pattern is identified the hypocrites are been detected automatically. Here the system needs to be trained based upon the certain patterns so that they will be able to identify them. The hypocritic case is usually examined by the difference in the hypocrisy and non-hypocrisy classes. Here the hypocrisy detection models are made up of feature selection and the data classification techniques.*

Key Words: E-Transactions, Hypocrisy, Feature Selection, Classification, Rule-Based Patterns

1. INTRODUCTION

Hypocrisy is synonym for fraud which is been seen in many of the industries such as banking and financial sectors. These fraudulent behaviors can be observed or detected by thus making use of data mining tools. Since millions of transactions will take place in every minute there will always be a possibility that such hypocritic cases may be observed. Hence Hypocrisy is one the major factors used for detecting the fraud related activities taking place in e-transactions.

The Hypocrisy will be detected in one of the following two ways namely: offline hypocrisy and online hypocrisy.

- **Offline Hypocrisy**: It usually involves human related activities such as stealing wallet which consists of the crucial information such as ID Proofs , credit card and so on.
- **Online Hypocrisy**: Here the scammer will create a fake website and try to hack the personnel details related to that of a customer.

Some of the ways in which they can hack the personnel details would be hacking, spoofing and so on. The number of transactions taking place per day is very huge, it is more likely for the threat to appear, thus detecting the new types of threat and creating new rules for them would be time consuming. Thus when Hypocrisy is detected, there is a higher probability of affecting the crucial information present in your system. In order to turn down the hypocrisy we need to spot the factors that will lead to hypocrisy. The Hypocrisy detected in the system will usually be larger than the cost of the goods which is been sold.

The Bank and the credit card companies will gather a large amount of transaction data in order to keep a track of customers spending habit, so as to detect the hypocrisy which is one of the major aspect to provide a better service. The online transaction associated with that of a customer will take place in a very small amount of time, thus providing the results in faster and a better way to that of a customer. The credit card is a small piece of plastic card which is been given to that of the card holder in order to make purchases. The customer thus can purchase any kind of commodity through that of online transactions, thus by making the payment with the help of credit/debit cards. When multiple transactions are made by the customers there is always a chance that threat or hypocrisy make occur. Hence we must make sure that these transactions needs to be safe, no problem should occur. One of the best solution is to safeguard our online transactions is by making use of firewalls.

It is our responsibility to safeguard our transactions and make sure that each of the transaction which is been made is safe and secure. It is best to make use of the data mining techniques in order deal with large amount of data. We can also make use computational intelligence techniques in order to identify the fraud which makes occur in the transactions. We can thus make use of electronic payment services such as Brazilian payment services. The results obtained from the system can thus be used for analyzing the hypocrisy present in the system which is mainly detected by that of hypocrisy analysts.

1.1 RELATED WORKS

Amitha Raghava-Raju [2], discussed few of the features in basically finding the fraudulent cases from the dataset that they had and those were the throttle through which the differences varied from logging into the system to purchase of a particular item and the number of particular ids that would be attached. From the inference that they obtained from their dataset they flagged all the hypo critic users. If we look at the general pattern there is certain time that user spends in working with the system while making his transaction and if this tie period is close to his average time taken normally he might be a genuine user, On the counter part if the time used in the process is less than this time he is ought to be a fraudulent user.

In [3], in order to detect hypocritic transactions, the authors gave prominence to Hidden Markov model which detects the fraud transaction and simultaneously report the timestamp and IP address of the intruder's machine. Every time a new transaction made is recorded in the system. If any intruder tries to make transaction with any registered credit card, then its spending habit will be different from that of authorized user and can be easily captured. Through this system we make sure that no genuine transaction is rejected. The system is capable of recording the timestamp and IP address of the attacking machine so that the geographic location of intruder can be traced.

In [4], In this paper they have discussed about the HSVM technique basically they have tried to exploit the conditions and find the matching patterns and find the IP address from where these frauds might take place from and to rectify the same the future enhancement block the card temporarily down and get hold of the fraud. HSVM as seen in the recent history has been widely recommended for the classification process.

The identity crime in the recent history has taken the toll as we see the hackers have easy access to the real deal and can make use of the same for theft or frauds.

In [5], come across techniques like feature selection that is taking certain important features into account and checking for the accuracy and precision by avoiding humongous number of attributes. Feature selection, predominantly also called variable selection as we select the number of variables from the available lot. Here we select relevant features that they use in their model construction. It's been seen that the selection process was carried out in two ways forward and backward. The forward approach is where they had begun with no variables and kept adding until further adding does not lead to any error and in the other way around we start with all the variables and go on reducing. The disadvantage is that the efficiency is low in most cases.

In [6], introduced a method where there are nodes that form a network and are directed in nature technically called Bayesian network. The different data mining as well as machine technique are evaluated and applied. Here the network is formed to catch or highlight or flag fraud in e-transactions. These nodes are used to form directed graphs that are acyclic in nature. The nodes in the graph notify the different random variables and the relationship or the similarity is shown by arcs.

In [7], introduced a fusion approach using Bayesian learning and also Dempster -shafer. In this approach the combination of the current and the past history of the user is taken into account. Whenever a customer does regular transaction he gets into a certain type of behaviour or pattern, thus forming a profile for himself. In this suggested once it is suspected to be fraud the belief might be strengthened or not in regard to the history of the profile.

Wen-Fang Yu and Na Wang [8], implements the outlier mining algorithm based upon the sum that is been obtained from source and destination. It usually makes use of Euclidean distance in this approach. The distance between the two clusters is found using either k-nearest neighbour using classification or k-means clustering now this calculating distance is taken as the Euclidean distance. The k-means finds the k nearest points and outlier mining algorithm uses this distance and predict the transaction to be genuine or not. The outliers sets are decided by setting the threshold of outliers but it is noticed that threshold varies in different situations of applications.

V. Filippov et al. [9], describes the credit card hypocrisy detection system which includes various fraud detection models in case of clustering model if the attributes of instances are legal then the transaction is legal. In case of Bayesian network it gives an estimation with the genuine transaction and the fraud instances. It also takes care of geographic locations when a transaction takes place if there is difference in time between the current and previous transaction then it is labelled as fraud.

Credit card hypocrisy detection is precisely confidential, Yufeng Kou et al. [10], used neural network as one of the techniques to detect the hypocrisy in the credit cards. A neural network functions like a brain when the nodes imitate the functioning of it, these networks can be constructed for two kinds of learning, supervised and unsupervised. In case of unsupervised methods we do not give any information regarding the genuine or fraud instances

from the historical database instead it tries to detect and learn abnormal transactions. In case of supervised method, model or system is trained on genuine or fraud behavior hence using this the new instance can be assigned to a particular class.

Tao Guo and Gui-Yang Li [11], introduced a new approach which is based on Neural networks. The topologies are formed when the nodes which act as local attributes that are stacked into layers, these layers form an interconnection. Here Back propagation algorithm is being used it compares the prediction to each instance with known target, however

the disadvantage is that there is no good way to find hidden layers of networks.

1.2 Problem Definition

To design a system to detect hypocrisy cases in e-transactions using distributed data mining.

2. Methodology

The implementation phase was divided into mainly three modular components. In the very first module we collected the details of a customer and try to match a pattern with the database and find a co-relation to check if the case is genuine. In the second module we went with the main component of the system which was again divided into the three phase's genuine customers, update databases and check for hypocrisy.

There are 3 modules in our system namely,

- Customer
- Detection System
- Class

From the dataset that we had acquired we separated the genuine and fraud cases and formed 2 separate databases, now these 2 databases are used to form the rule base and patterns that would be used to get the inference out regarding the genuine and fraud transactions each time the new user enters his credentials and further again that data would be updated into the respective database.

2.1 Dataset

The dataset that we have taken for online transaction is not the original data that might have been generated at the time of transaction. The original data is highly confidential and therefore difficult to access. Hence we have used a data set which we found online it contains 20 attributes and also 1000 instances, the data has mixture of both categorical and numerical values which have been cleaned for further processing.

The data set we have chosen for our project was fairly used by many others hence we have gone with the same for our study. In the dataset we found many attributes, all the attributes do not contribute to the classification and the last attribute gives the class whether it is good or bad that is if its fraud or genuine case.

2.2 Working Principle

The system will be trained priorly based on the training data set. Based on former occurrence of fraud cases, the rule sets will be designed. The input that outlines each transaction will be the input to the system.

Every time a new transaction is made by an individual user in the system, his essentials, spending habits, purchase time, credit card number and history or state of former purchase, location etc., are the few attributes to be acknowledged.

The input will be preprocessed, cleaned and feature selected values are transferred to the inference engine. The rules which are present in the Rule base are sent to the systems inference engine. The inference engine will perceive the inputs based on the rules in the Rule base. The system will adapt and learn by itself. Every time the fraud case is detected the system will raise an alarm and will block the transaction, concurrently the database will be updated. Similarly when the transaction is found to be genuine the system will grant access and the database will be updated.

Fig -1: Instances of Fraud

Every time a new transaction is made by an individual user in an e-commerce system, his credentials, spending habits, purchase time, purchase history, credit card number and history or state of former purchase, location etc., are the few attributes to be acknowledged. The input will be preprocessed and it will be fed to the inference engine. The rules which are present in the Rule base are fed to the inference engine. The inference engine will perceive the inputs based on the rules in the Rule base. The system will adapt and learn by itself. Every time the fraud case is detected the system will raise an alarm and will block the transaction, concurrently the database will be updated. Similarly when the transaction is found to be genuine the system will grant access and the database will be updated.

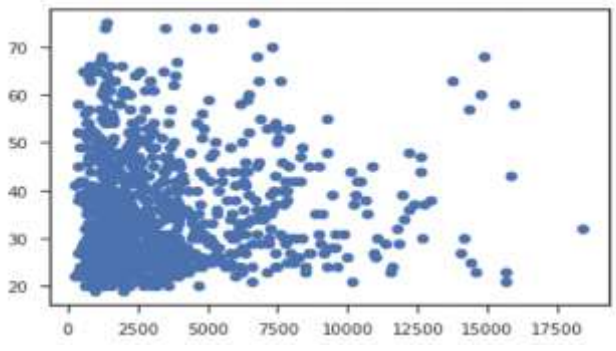

Fig -2: Amount Spent v/s Age

During the testing phase we saw various relations among the attributes and one interesting plot we come across is that of a relation between amount spent v/s age. Here we see that people between the age group of 20-40 spent maximum in the range of 2500 to 5000 rupees and as the age increases we see the decline in it, we also see a decline in expense after 10000 rupees.

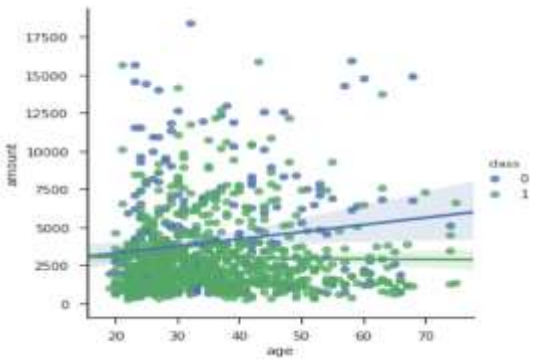

Fig -3: Classification with Amount Spent and Age

When we compare two graphs we see that the transactions which were seen above 6000 were mainly fraudulent cases were observed. There are many other techniques that can be considered such as Bayesian networks instead of ANFIS but the problem again that we would face is each time a new instance or a new type of fraud is encountered the system would face difficulty.

3. CONCLUSION

In this work we presented an analysis of different classification techniques and have found that the random forest gives the best accuracy. We have also seen that when we use adaptive Neuro-fuzzy approach the system need not be retrained as it learns from the behavior of a particular instance if found to be fraud.

REFERENCES

[1] Shaji Jisha and Dakshata Panchal, "Improved Fraud Detection in E-Commerce Transactions" IEEE 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA), pp. 121-126, 2017.

[2] Amitha Raghava-Raju, "Predicting Fraud in Electronic Commerce: Fraud Detection Techniques in E-Commerce ", International Journal of Computer Applications, Volume 171 – No. 2, August 2017.

[3] Aayushi Gupta, "Hidden Markov Model based Credit Card Fraud Detection System with Time Stamp and IP Address", International Journal of Computer Applications, Volume 166 – No. 5, May 2017.

[4] Mareeswari, V., and G. Gunasekaran, "Prevention of Credit Card Fraud Detection based on HSVM" International Conference on Information Communication and Embedded Systems (ICICES), IEEE, 2016.

[5] Lima, Rafael Franca, and Adriano Cesar Machado Pereira, "A Fraud Detection Model based on Feature Selection and Under Sampling applied to Web Payment Systems. "International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), IEEE/WIC/ACM. Vol. 3, 2015.

[6] Caldeira, Evandro, Gabriel Brandao, and Adriano CM Pereira, "Fraud Analysis and Prevention in E-Commerce Transactions." Web Congress (LA-WEB), 9th Latin American. IEEE, 2014.

[7] Yu, Wen-Fang, and Na Wang, "Research on Credit Card Fraud Detection Model based on Distance Sum.". International Joint Conference on Artificial Intelligence, IEEE, 2009.

[8] Filippov, V., L. Mukhanov, and B. Shchukin, "Credit Card Fraud Detection System." 7th International Conference on Cybernetic Intelligent Systems, IEEE, 2008.

[9] Kou, Yufeng, "Survey of Fraud Detection Techniques." , international conference on Networking, sensing and control. Vol. 2. IEEE, 2004.

[10] Brause, R., T. Langsdorf, and Michael Hepp, "Neural Data Mining for Credit Card Fraud Detection." Proceedings of 11th International Conference on Tools with Artificial Intelligence, IEEE, 1999.

BIOGRAPHIES

Rithesh Pakkala P.
B.E (CSE), M.Tech. (CSE)
Asst. Professor

Mr. Rithesh Pakkala P. obtained his B.E and M.Tech degree in Computer Science & Engineering from Sahyadri College of Engineering and Management, Mangaluru under VTU. He is having 3.5 years of teaching experience and 1 Year 5 Months of experience in the IT field. Mr. Pakkala worked as a NAAC Coordinator of department of ISE at Sahyadri College of Engineering and Management, Mangaluru and also Software Engineer for Karnataka Examinations Authority (KEA), CET Help-Line Centre, Mangaluru in 2013. He is having 6 publications in International Journals and 1 publication in the proceedings of International Conference. He has also participated in more than 5 workshops. His specialized subjects are Formal Languages & Automata Theory, System Modeling & Simulation, Discrete Mathematical Structures and Operations Research. And research areas are data mining, theory of computation and cryptography.

Ravi Yadav currently pursuing Bachelor's in Information Science. Who has also served as a club captain of Mozilla campus clubs is keen and passionate regarding programming and also is interested in machine learning concepts.

Pooja B currently pursuing Bachelor's in Information Science from Sahyadri college of Engineering and Management is keen and extremely passionate about data mining and is been learning in same field.

Rahin Shama currently pursuing Bachelor's in Information Science from Sahyadri college of Engineering and Management is keen and extremely passionate about Big Data and is been learning in same field.

Vaishak Chandra K. S. currently pursuing Bachelor's in Information Science from Sahyadri college of Engineering and Management is keen and extremely passionate about Big Data and is been learning in same field.