International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 2 (2019) pp. 393-396

A Cluster Based Approach for Credit Card Fraud Detection System using Hmm with the Implementation of Big Data Technology

M. Sathyapriya[#1], Dr. V. Thiagarasu*[2]

[#]*Assistant Professor of Computer Science, Gobi Arts and Science College (Autonomous), Erode, Tamilnadu, India.*

**Associate Professor of Computer Science, Gobi Arts and Science College (Autonomous), Erode, Tamilnadu, India.*

Abstract

In today's framework, Credit card is used for online transaction and security of the usage of credit card is also a big issue. Here, the fraud detection is done in two Steps. First is training of transactions and second is detection of fraud in the incoming transactions. In the first Step, sequences of transactions are added in to the system. In the second step, the fraud is detected in the credit card using the cluster comparison of the transactions. In first stage, HMM (Hidden Markov Model) is used with the expenditure behavior of card holders and categorized as low, medium and high expenditure behavior. The probability for every transaction sequence is calculated and checked with the threshold value. If any deviation is encountered then it is declared as fraud. To analyse the transactions into clusters clustering is used with the implementation of Big Data technology.

Keywords: Cluster, Hidden Markov Model, Expenditure Behavior, Big Data, Spark.

I. INTRODUCTION

This paper is concerned with the study and analysis of detecting credit card frauds and proposes a model for improving the efficiency in processing huge amount of data and detecting frauds among the transactions with a minimal duration of time. Globalization has amplified the use of the internet for online shopping that result in an extensive escalation of credit card transactions throughout the world. Thus a rapid growth in the number of credit card transactions has led to a considerable rise in fraudulent activities. Therefore, it is mandatory to employ mechanisms that are able to assist in fraud detection system [12].

The study addresses the fraud detection system to analyze the customer transactions in order to identify the patterns that lead to frauds. In order to facilitate this pattern recognition work, the k-means clustering algorithm is used which is an unsupervised learning algorithm and applied to find out the normal usage patterns of credit card users based on their past activity [5]. The past transactions are trained and stored in the database for analyzing the incoming transactions and used for detection of fraud patterns.

The focus of the research is to develop a prototype for fraud detection system that would attempt maximally to detect credit card fraud by generating clusters and analyzing the clusters generated by the dataset for anomalies.

II. LITERATURE SURVEY

In fraud detection, using Bayes minimum risk classifier is used that gives rise in much better fraud detection results. In this work, Hadoop Distributed File System is used for storing and fast accessing of the user logs. Bayes rule depends upon Prior Probability and stores all the data and transactions details of the customer. If any unusual thing is encountered then it is considered as fraud [Anushree Naik, Kalyani Phumamdikar et. al, 2016]. In consideration of Frequent Item set Mining, a matching algorithm is proposed and according to the transactions closer to the patterns is identified and decisions are made whether it is legal or illegal. The matching algorithm detects to which pattern the incoming transaction matches more. If the incoming transaction matches more with legal pattern of the particular customer, then the algorithm returns 0 which is a legal transaction else the algorithm returns 1 which is a fraudulent transaction [K. R. Seeja and Masoumeh Zareapoor, 2014].

In security level based system there are three levels of security as login and password, SVM (Support Vector Machine) for detecting user behavior and decision tree for determining whether user behavior is normal or abnormal and questionnaire for additional verification. Initially clustering technique is used to find the data pattern that does not belong to current data pattern. Then SVM classification is used for next level of security which is a binary classification so the transactions are labeled as fraudulent or legitimate [Vijayshree B. Nipane, Poonam S. Kalinge et.al, 2016].

For efficient detection of credit card fraud, clusters for training set are generated and spending profile of card holder is identified. It does not consider the number, type of items purchased but it only concentrates on the amount of item purchased. It stores data of different amount of transactions in the form of clusters depending on transaction amount. If the transactions deviate from stored ranges, then it is considered as fraud [Ayushi Agarwal, Shiv Kumar, 2015]. Artificial Immune System called AFDM was developed where normal detectors are generated for each user and the user's previous transactions are considered and processed. Fraud detectors encounter the fraud patterns in a dataset. Also suggested to update the user model based on user's latest transactions because user's behavior need not be same all the times which may lead to Higher False alarm Rate [Neda Soltani, Mohammad Kazem, 2012].

III. MOTIVATION

Rule-based systems have been the common fraud detection tools for current financial systems where fraud experts define the rules according to past cases and obtained results. If a new transaction matches one or more of the previously defined rules, an alarm is raised to indicate that the new transaction is

potentially fraudulent. The rule-based approach is successful for previously observed fraud patterns [11]. Before adding a new rule to the existing rule-set, a considerable number of fraudulent transactions matching the rule must have occurred. In this period, the fraud strategies may change, causing the induced rule to expire. Therefore, the focus should be on using the past transactions that follow rule-based approach together with unsupervised approach which also detect the previously undiscovered fraud activities.

There is a need to employ fraud detection systems that cope up with updated expenditure behaviour of the card holder. The strategy of the detection process is to detect as much fraud as possible by minimizing the false positive rate which gives a negative impact on card holder satisfaction if there is an increase in expense of giving more false alarms [2]. To achieve this strategy, the threshold value is calculated at card holder account level by analyzing the probability sequence of past and new incoming transactions. Moreover, the detected fraud transactions are labelled in the database for future analytics in the case of extra evaluation, if needed.

IV. OBJECTIVES

The expenditure profile of the card holder is used to find the fraudulent activities based on the probability outcome of the system. Although it is avail with various existing methodologies, but it still need to be refined for detecting the fraud transactions by using real-time data and scalability of huge amount of transactions that occur at once [15].

The objectives of this research:

1. Study and analyze the existing fraud detection algorithms for clustering, and detecting the fraud transactions with improved performance.
2. Design and develop an improved algorithm for detecting fraud transactions that improves the system in performance, accuracy and reduced false positive rate.
3. Analyzing the incoming new transactions based on past expenditure stored in the database repository.
4. Make an experimental testing of the proposed system to prove the validity of the algorithm.
5. Performance Analysis has to be performed over the developed system to reveal the detecting efficiency of fraud and genuine transactions.

V. DESIGN OF PROPOSED SYSTEM

The credit card holder expenditure system is created as a prototype implementation which combines incoming credit card transactions with most recent transaction history and makes real time prediction based on the likelihood of the expenditure pattern. The importance is on analyzing card holder expenditure behavior and states a new behavior model for detecting credit card fraud. The system has been proposed to detect credit card frauds along by reducing the rate of false alarm which poses two distinct phases [8].

In the first phase, training phase, the credit card holder's past expenditure behavior is stored in the database and k-means clustering algorithm is applied to cluster the transaction amount based on the expenditure pattern. After clustering, using Hidden Markov Model, the transactions are generated as observation symbols such as {low, medium, high} where sequence of states are recorded by considering the transaction amount of each card holder [1].

The second phase is detection phase where generation of observation symbols is made for new incoming transaction and the probability of the new transaction is compared with the stored expenditure value of the card holder. If the result value deviates from the threshold value, then it is assumed as fraudulent transaction else it is marked as a genuine transaction [10]. Thus proposed system is designed with the developed algorithm given as Figure 1 to attain the accuracy over the credit card transactions.

Step 1: Read the input transaction
Step 2: Read the expenditure amount of the card holder from the database
Step 3: Find to which cluster the transaction amount data fall into.
Step 4: Calculate the probability difference and test the result with trained data.
Step 5: Perform $\Delta\alpha=|\alpha_1- \alpha_2|$
Step 6: To calculate the acceptance the formula used is, $\Delta\alpha / \alpha_1$
Step 7: The resultant value is compared with threshold value
Step 8: The transaction is accepted as a genuine one if, $\Delta\alpha/ \alpha_1 \leq$ threshold value (θ)
Step 9: With the obtained result, Sensitivity, Specificity, False Positive Rate, Precision and Accuracy are calculated as,

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TP}{TP+FN}$$

$$\text{False Positive Rate} = 1 - \text{Specificity}$$

$$\text{Precision} = \frac{TP}{TP+FN}$$

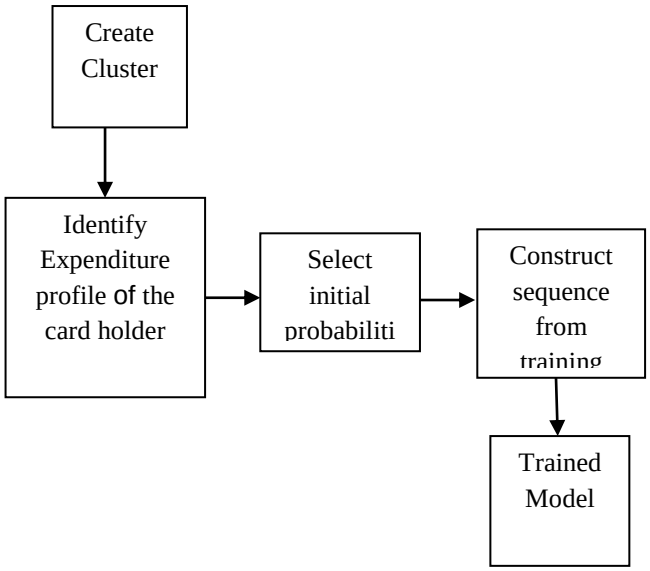
$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN}$$

TP = No. of fraud transactions (True Positive)
FN = No. of missed fraud transactions (False Negative)
FP = No. of legitimate transactions (False Positive)
TN = No. of missed legitimate transaction (True Negative)

Figure 1. Algorithm for Cluster Based Fraud Detection

The schematic representation of the proposed system is as specified below in Figure 2 stating the training phase of the historical transactions, the detection phase process and evaluates the new incoming transactions.

Training Phase



Detection Phase

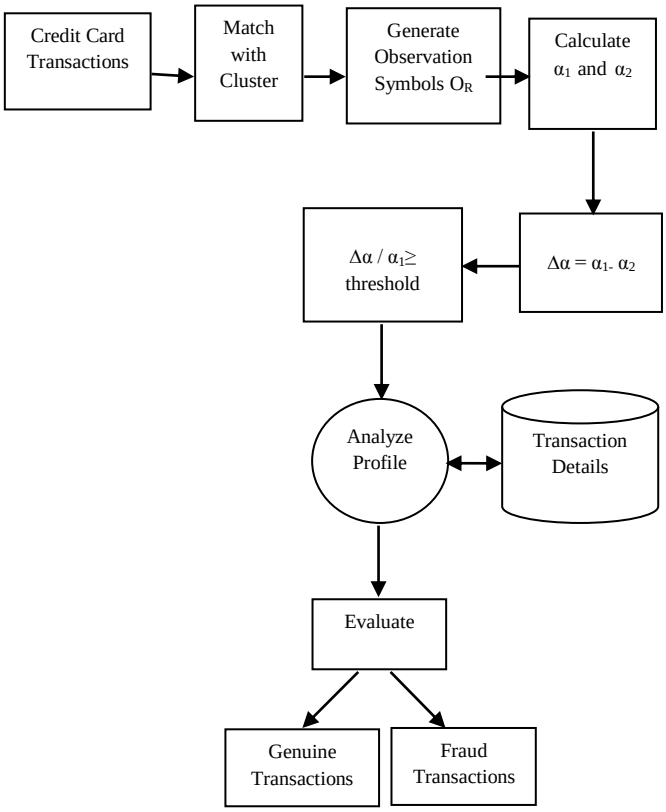


Figure 2. Schematic Representation of Credit Card Fraud Detection Model

VI. EXPERIMENTAL RESULTS

The proposed system has been developed for detection of credit card frauds in accordance with improvement over the performance metrics such as sensitivity, specificity, false positive rate, precision and accuracy with the dataset created using Spark as a software framework [7]. The system stores the historical transactions in the database after featuring the data parameters as the observation symbols. The data are clustered as three groups as low, medium and high for each card holder according to the transaction amount using k-means clustering [13]. The screen shots of the proposed system showing Figure 3. Table representation of dataset, Figure 4. Clustering the transaction amount in to three different clusters and Figure 5. Representing the performance analysis of the proposed system.

```
user1@gasc-desktop:~$ python CCarddetails.py
      ------------------------------------
      | Credit Card Number   |      Amount |
      ---------------------|-------------|
      |1234 5678 9012 3456   |      2000   |
      |1234 5678 9012 3456   |      2500   |
      |1234 5678 9012 3456   |      3000   |
      |1234 5678 9012 3456   |      3600   |
      |1234 5678 9012 3456   |      1800   |
      |1234 5678 9012 3456   |      2280   |
      |1234 5678 9012 3456   |      2090   |
      |1234 5678 9012 3456   |      4850   |
      |1234 5678 9012 3456   |      3080   |
      |1234 5678 9012 3456   |      4500   |
      ------------------------------------
user1@gasc-desktop:~$ _
```

Figure 3. Table representation of data set

```
user1@gasc-desktop:~$ python cluster.py
       ------------------------------------
      |Amount | ClusterId | ClusterCentroid |
      |------------------------------------|
      |2000   |     1     |    4.75         |
      |2500   |     1     |    4.75         |
      |3000   |     2     |     5           |
      |3600   |     2     |     5           |
      |1800   |     1     |    4.75         |
      |2280   |     1     |    4.75         |
      |2090   |     1     |    4.75         |
      |4850   |     3     |     8           |
      |3080   |     2     |     5           |
      |4500   |     3     |     5           |
      ------------------------------------
user1@gasc-desktop:~$ _
```

Figure 4. Clustering of transaction amount

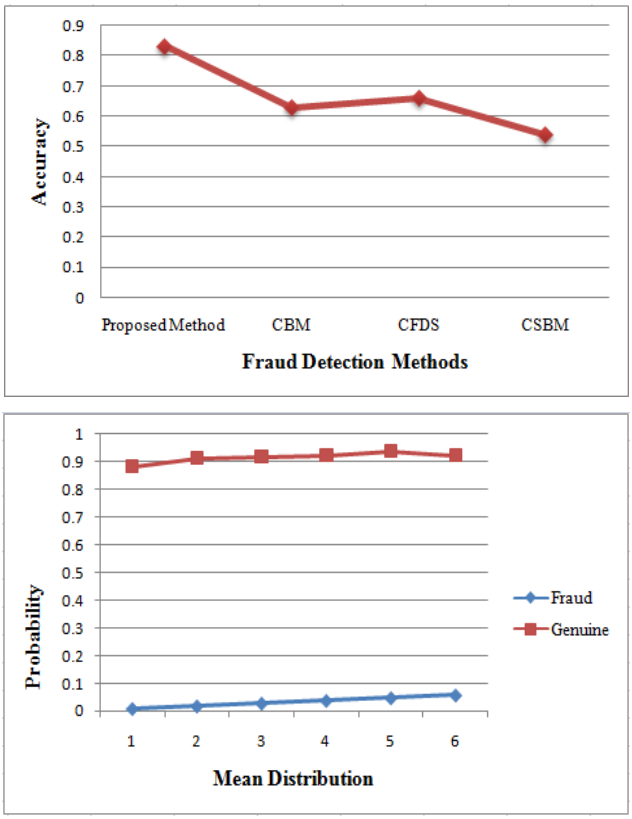


Figure 5. Performance Analysis of Fraud Detection Methods

VII. CONCLUSION

The credit card expenditure behavioural model has been proposed for improving the performance, accuracy and scalable storage of data that cope with the recent issues in the research. The focus is on analyzing card holder expenditure behavior and proposed a system that can support large volumes of data. The proposed system is better in reducing the rate of false alarms that can be achieved by examining the relationship between the transactions that were marked as actual frauds and the transactions that were guessed as fraud. The reduced rate in false alarm improves the accuracy of the system in detecting exact genuine and fraud transactions.

REFERENCES

[1] A. Prakash, C. Chandrasekar , "A Novel Hidden Markov Model for Credit Card Fraud Detection", International Journal of Computer Applications, 2012.

[2] Abhinav Srivastava, Amlan Kundu, Shamik Sural and Arun Majumdar, "Credit Card Fraud Detection Using Hidden Markov Model", IEEE Transactions, 2008.

[3] Anushree Naik, Kalyani Phulmamdikar, Shreya Pradhan, Sayali Thorat, Prof.Sachin V. Dhande, "Real time Credit card transaction analysis", International Engineering Research Journal (IERJ), Vol 1, Issue 11, 2016.

[4] Ayushi Agarval, Shiv Kumar, Amit Kumar Mishra, "Implementation of Novel Approach for Credit Card Fraud Detection", IEEE Transactions, 2015.

[5] Florin Dragan, Ioan-Danial Borlea and Rad-Emil Precup, "On the architecture of a clustering platform for the analysis of big volumes of data", IEEE Transactions, 2016.

[6] K. R. Seeja and Masoumeh Zareapoor, "Fraud Miner: A Novel Credit card fraud detection model based on Frequent Item set Mining", The Scientific World Journal, 2014.

[7] M. Sathyapriya, Dr. V. Thiagarasu, "Big Data Analytics Techniques for Credit Card Fraud Detection: A Review", International Journal of Science and Research, 2017.

[8] M. Sathyapriya, Dr. V. Thiagarasu, "Implementation of Big Data Technology for Credit Card Fraud Detection System Using Hidden Markov Model", International Journal of Scientific Research in Computer Science Applications and Management Studies, Volume 7, Issue 3, 2018.

[9] Neda Soltani, Kazem Akbari, "A new user-based model for credit card fraud detection based on artificial immune system", Research Gate, 2012.

[10] Rajeshwari. U, Dr. B. Sathish Babu, "Real Time Credit card Fraud Detection using Streaming Analytics", IEEE Transactions, 2016.

[11] S. Athmaja, M. Hanumanthappa and Vasantha Kavitha, "A Survey of machine learning algorithms for Big Data Analytics", IEEE Transactions, 2017.

[12] Sunil S Mhamane, L.M.R.J Lobo, "Internet Banking Fraud Detection Using HMM", IEEE Transactions, 2018.

[13] Vaishali, "Fraud Detection in Credit Card by Clustering Approach", International Journal of Computer Applications, 2014.

[14] Vijayshree Bhaskar Nipane, Poonam S. Kalinge, Dipali Vidhate, Kunal War, Bhagayashree P.Deshpande, "Fraudulent Detection in Credit Card System Using SVM & Decision Tree", IJSDR, 2016

[15] You Dai, Jin Yan, Xiaoxin Tang, Han Zhao and Minyi Guo, "Online Credit Card Fraud Detection: A Hybrid Framework with Big Data Technologies", IEEE Transactions, 2016.