

# Survey Paper for Credit Card Fraud Detection Using Data Mining Techniques

S.Vimala<sup>1</sup>, K.C.Sharmili<sup>2</sup>

PG Scholar, Department of Computer Science, Stella Maris College, Chennai, India<sup>1</sup>

Assistant Professor, Department of Computer Science, Stella Maris College, Chennai, India<sup>2</sup>

**ABSTRACT:** Nowadays, the Credit card usage is increasing day by day for both regular purchases as well as online. In every transaction of credit card, the bank should check the fraud detection. Everywhere, the fraudulent transactions are generated in new ways by the fraudsters. Banking industry, mostly use Data Mining techniques for credit card fraud detection. The detection techniques is mostly based on the methods like Decision Tree, Clustering techniques, Neural Networks and Hidden Markov Model, these are evolved in detecting the various credit card fraudulent transactions. This paper presents the survey of those techniques and identifies the best fraud cases.

**KEYWORDS:** credit card, data mining, fraud detection, Decision Tree, Hidden Markov Model.

## I. INTRODUCTION

Nowadays the modes of payment methods are changed into online transactions. Banking system provides different type of payments like e-cash, card payments, internet banking, and e-services for improving online transaction. Credit card is one of the most custom ways of online transaction. Credit Card purchasing is categorized into two types:

- > Physical Card
- > Virtual Card

A physical card is used for making a payment at a physical payment terminal in your shop or via an online shop. A virtual card is not available as a card; it is a digital payment instrument that can be used for internet banking.

### **Credit Card:**

A Credit card is used for a payment purpose to enable the user to pay an amount for goods and services based on the cardholder's promise to the card and other agreed charges. Figure 1.1.1 will show the process of credit card transaction. Generally, the issuer creates a temporary account to the card holder along with a grant statement in order to draw money for payment of purchase. The bank charges a small interest in this regard.

### **Credit Card Fraud:**

Credit card is the most frequently used payment method in online transactions and gradually there is an increase in fraud transactions as well. The intention behind these kinds of fraudulence may be obtaining goods without paying, or unauthorized funds from an account. Most people associate crime of credit card fraud with ID theft. Now the ID theft decreased into a percentage. A survey proved that about 0.1% fraudulence is by credit card. This has resulted in huge financial losses.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

## Fraud Detection:

Credit card fraud detection stores the previous transaction patterns of each cardholder which helps in decrease the rate of credit card frauds. The data is based on the user spending ability and the country they live. It calculates user's characteristics. Deviation of 20-30, user transaction history and operating is considered as an invalid attempt and system takes action. The system will block the user for more than three invalid attempts.

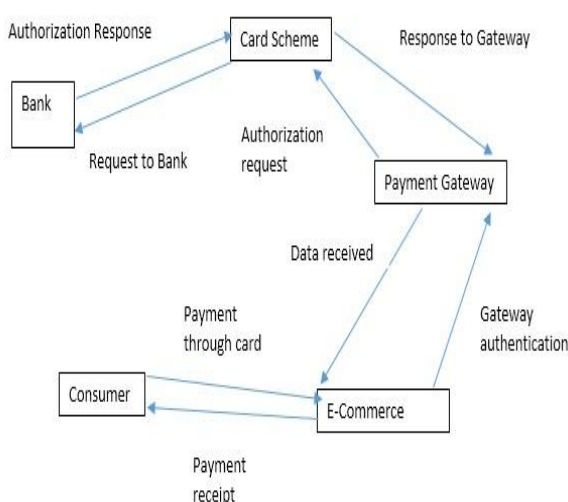


Figure 1.1.1: The process of credit card transaction.

## II. LITERATURE SURVEY IN FRAUD DETECTION

In [6] SnehalPatil et al., describes the “Decision Tree Induction Algorithm” which is used for Credit Card Fraud Detection. In this paper it discusses about the method, decision tree approach is a new cost sensitive technique compared with well-known traditional classification models on a real world credit card fraud data set, which reduces the sum of misclassification cost while selecting the splitting attribute at each non-terminal node is advanced. Credit card fraud detection is to reduce the bank risks, also used to equalize the transaction information with credit card fraud transaction of historical profile pattern to predict the probability of being fraudulent for a new transaction. In [3] Dr.R.Dhanapal, Gayathiri.P, describes the “Credit Card Fraud Detection Using Decision Tree for tracing Email and IP Address. By using this technique, we can able to find out the fraudulent customer/merchant through tracing the fake mail and IP address. If the mail is fake, the customer/merchant is suspicious and information about the owner/sender is traced through IP address.

In [1] RaghavendraPatidar, Lokesh Sharma, describes the “Neural Network” is used for the comparison on one way solution by using neural network, for matching the previous stored patterns and currently used patterns from which we can detect such patterns for Credit Card Fraud Detection.

In [8] Yogesh Bharat Sonawane, Akshay Suresh Gadgil, AniketEknath, NiranjanaKamalakarJathar, describes the “Clustering Based Approach” with K-means cluster analysis is a technique for smashing dataset down into narrated constituents in such a way that patterns and order becomes perceivable. Searching outliers is a main duty in k-means clustering for Credit Card Fraud Detection.

In [5] Mr. P. Matheswaran, Mrs. E.SivaSankariME, Mr.R.Rajesh ,describes the “Data Mining Techniques” for Fraud Detection in Credit Card. In this paper, we can detect the hidden information like whether an incoming transaction is fraudulent or not. It also dividing the transaction amount in three categories used on different ranges of transaction amount each group show the aberration symbols. The different steps in credit card transaction processing are represented as the underlying stochastic process of an Hidden Markov Model. HMM is used to model the sequence of operation in credit card transactions with the behavior of cardholder. In [4] Dinesh L. Talekar , K. P. Adhiya, describes

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

the “Credit Card Fraud Detection System: A Survey”. This survey paper they tell the steps to detect the fraud transaction using Hidden Markov Model.

In [2] K.RamaKalyani, D.UmaDevi, describes the “Genetic Algorithm is the technique of finding optimal solution for the problem and implicitly generate the results using genetic algorithm, In this paper we use Genetic Algorithm for Detecting the fraud of Credit Card Payment System by the probability of fraud transactions that can be predicted soon after credit card transactions by the banks. In [7] PoojaChougule, A.D. Thakare ,Prajakta Kale , MadhuraGole, PriyankaNanekar, describes the “Fraud Detection of Credit Card Payment System by Genetic Algorithm” used genetic algorithm to optimize the output of the transaction for the Credit Card Fraud Detection”.

### III. MOTIVATION

People are using credit card payment for shopping, bill payment, online transaction has considerably increased in number. We should avoid the fraud transaction which would rise by a fraudster. Data mining has various techniques to avoid that risk from the various types of attacks. In this paper we discuss about some of the techniques which is used to detect fraud operations. Out of these, Decision Tree and Hidden Markov Model are considered to be the best for fraud detection.

### IV. DATA MINING

The Methods that involves in identifying patterns from large data sets at the intersection of machine learning, statistics and database systems is called Data Mining. Banking Industry mostly implement data mining techniques into their dataset, because bank has large set of data to take important decision, focus on important pattern of data from the database and secure these data.

By fetching the information and converting them into some patterns and to understand the meaning of data, Data mining is the most wanted tool to take important decisions. It is well known as Knowledge Discovery in Database. The fraud is also done using data mining techniques. Figure 4.1 will show the fraud detection transaction.

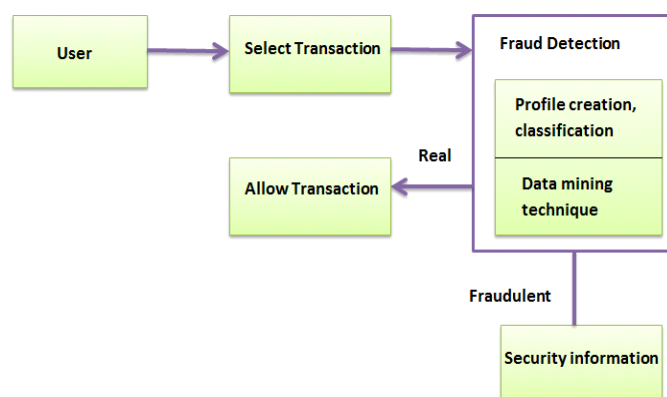


Figure 4.1 Credit Card Fraud Detection using data mining technique.

The steps for detecting the frauds using data mining techniques are first the user will select the transaction for the one they want to do the payment. Here the Fraud detection section has two steps,

- First the bank that offers the credit card to the customer, bank would create one profile and classify the cardholder data transaction details.
- In later they apply data mining technique to the transaction data; it will pass the transaction for verification.

If there is any fraudulent operation done, for it will get that transaction into security information section. In security information section current transaction data would check with the previous transaction information and the agreement. Otherwise the transaction would just be in process.

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

### V. VARIOUS TECHNIQUES FOR CREDIT CARD FRAUD DETECTION

In credit card fraud detection there are many methods in data mining. In this Survey paper we discuss some most useful methods.

- Decision Tree
- Neural Network
- K-Means Clustering
- Hidden Markov Model
- Genetic Algorithm

#### Decision Tree:

Decision Tree Algorithm is a Data mining induction technique that recursively shares a set of records. This algorithm used for solving regression and classification problems using tree representation. Decision tree contains one root node, internal nodes and leaf nodes. The nodes are labeled with the use of attribute names, edges are labeled with the values of attributes. For predicting a class label for a record we start from the root of the tree. Compare the values of the root attribute with record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node, comparing our record's attribute values continuously with other internal nodes of the tree until we reach a leaf node with predicted class value [6]. Decision Tree is easy to implement, understand and display comparing to other classification algorithm. Using decision tree we can also tracing the mail and IP address through the credit card fraud detection [3]. This fraud detection depends on the location where the cardholder use the previous credit card transaction compares with the location of current places transaction.

#### Neural Network:

Neural network also recommended for credit card fraud detection it is effective result in myriad complications. Using neural network on fraud detection is like a human brain working principal. Human brain stores the experience which they crossed used that experience in current life, which same process is did in neural network. Here in credit card fraud detection, neural network splits the information into various categories; first one is based on cardholder income, occupation. Another one category will store the payment details like, number of larger purchasing, frequencies of largest purchases, location. This wholes the detail is going to judge the future transaction whether the transaction is fraud or genuine [1]. Figure 5.2.1 show how the neural network would process. In this architecture, there are three different types of layer.

- **Input Layer:** This layer has input nodes, it will identify the cardholder use the information it will check the characteristics of the transaction.
- **Hidden Layer:** Hidden nodes are do neural network operation to identify the transaction fraud or not.
- **Output Layer:** After analysis the transaction output nodes give the output value, it between 0 and 1.

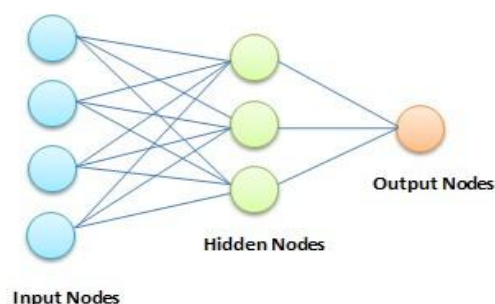


Figure 5.2.1: Neural Network Architecture.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

## **K-means Clustering:**

Clustering refers to making a group of data from the provided database. When it is required means if we want some group of information or it will recognize the same pattern for analysis. Clustering helps in bank database for finding priority of customers and to draw similar kind of fraud techniques used in fraud detection. To know the transaction is whether fraud or legal clustering uses one of the most used K-means algorithm [7].

In transaction we declare some variable like, transaction amount, credit card number, transaction date and id, country, merchant category id. Here credit card number is must if we forget we can't do the transaction this process done in transaction validation section. The information which we got as an input those will going to store in transaction dataset. Next we assign of the cluster name in which type of transaction that is and label it as, low cluster or high cluster or risky cluster [8]. The transaction detail will take over to k-means algorithm. If the transaction is fraud or legal it displays a message.

## **Hidden Markov Model:**

Hidden Markov Model is a set of states associated with the probability distribution. Each state generates outcome or observation according to the probability distribution which is associated with the particular state. The state is visible but the only observation or outcome are can be visible that is called Hidden Markov Model [5].

In credit card detection, HMM is used to find the fraud transactions by the spending behavior of cardholder. Spending behavior of the cardholder is calculated by the past history of transaction, that history has the attributes like transaction amount, IP address, shipping address & location of last transaction, etc. Here we categorize the cardholder behavior into three types like,

1. Low spending behavior
2. Medium spending behavior
3. High spending behavior

If the cardholder spends low amount, they under in low spending behavior, if the cardholder spending Medium level of amount they under in medium spending otherwise the cardholder who spends high amount they came in the category of high spending behavior.

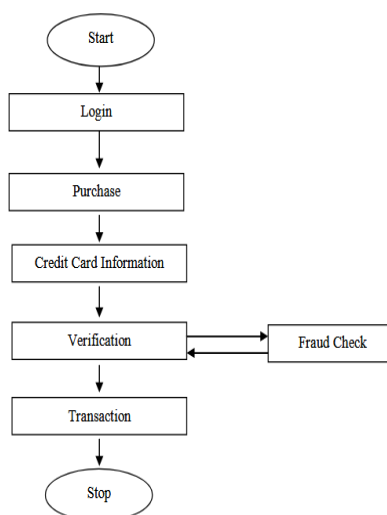
Figure 5.4.1 show the credit card fraud detection using HMM. In first level is profile identification, here we analysis the profile in which category the cardholder came it depend on the purchasing detail [4]. Hidden Markov Model uses two steps of training sequence to find the fraud detection.

- 1) From the past transaction history, Hidden Markov Model is been trained.
- 2) Take the input and check whether the transactions details are accepted by training sequence are not.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017



**Figure 5.4.1: Fraud detection Using Hidden Markov Model.**

## Genetic Algorithm:

The main aim of genetic algorithm is to obtain the better optimal solution, and also fraud transactions are generated with the given ample data set. This algorithm work in the manner of efficient and secure electronic payment detect whether a transaction is fraud or not [7].

In credit card transaction it has n number of attributes. In first step it would select the data set which one going to the process from that we select the standardize the data, which will include the full detail about the cardholder. Figure 5.5.1 shows the detection using genetic algorithm [2]. Calculate the critical values, from the frequency use of credit card count, current bank balance, credit card overdraft, location where they use credit card on that particular transaction, average daily spending. Then we compare the data critical, finally we know the transaction is fraud or that one is legal.

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 11, September 2017

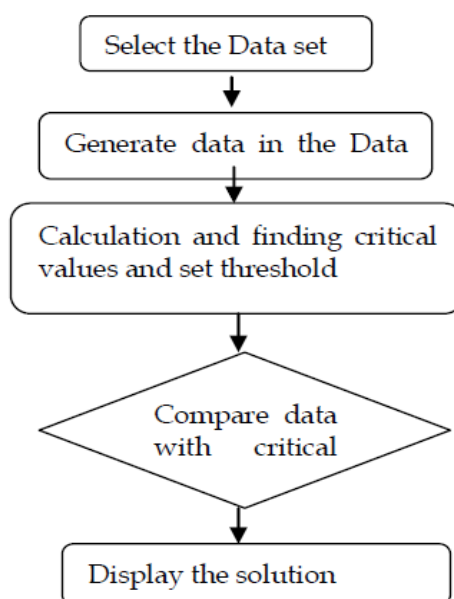


Figure 5.5.1: Credit Card Fraud Detection using Genetic algorithm.

### VI. CONCLUSION

In recent years, credit card usage has increased significantly. Fraud operations are also newly arriving in another way; there are more techniques introduced to detect the frauds. Main aim of this study is to know about the best technique that identifies fraud cases. One of the data mining techniques or combination of these techniques can be applied for credit card fraud detection system. The best way of credit card fraud is finding from the history of transactions; it predicts if the transaction is legal or fraudulent. In this paper we survey about four best data mining techniques for credit card fraudulence detection. Comparing their results we conclude the techniques decision tree and hidden Markov Model are the best way to find the fraud detection. Decision tree mostly detect the fraud using location. Hidden Markov Model detect the fraud depend upon the cardholders behavior and history of transactions, but we want further improvement in both the techniques to avoid frauds more in future.

### REFERENCES

- [1] RaghavendraPatidar, Lokesh Sharma, "Credit Card Fraud Detection Using Neural Network", International Journal of Soft Computing and Engineering (IJSCE)ISSN: 2231-2307, Volume-1, Issue-NCAI2011, June2011.
- [2] K.RamaKalyani, D.UmaDevi, "Fraud Detection of Credit Card Payment System by Genetic Algorithm", International Journal of Scientific & Engineering Research Volume 3, Issue 7, July-2012.
- [3] Dr R.Dhanapall, Gayathiri.P2, "Credit Card Fraud Detection Using Decision Tree For Tracing Email And Ip", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 2, September 2012.
- [4] Dinesh L. Talekar, K. P. Adhiya, "Credit Card Fraud Detection System: A Survey", International Journal Of Modern Engineering Research(IJMER),Vol. 4, Iss.9, Sept. 2014.
- [5] Mr.P.Matheswaran, Mrs.E.SivaSankari ME, Mr.R.Rajesh, "Fraud Detection in Credit Card Using DataMining Techniques, Volume II, Issue I", IJRSET-February-2015.
- [6] SnehalPatilet al, "Credit Card Fraud Detection Using Decision Tree Induction Algorithm",International Journal of Computer Science and Mobile Computing, Vol.4 Issue.4, April-2015.



ISSN(Online): 2319-8753  
ISSN (Print): 2347-6710

## International Journal of Innovative Research in Science, Engineering and Technology

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 6, Special Issue 11, September 2017**

- [7] Pooja Chougule#1, A.D. Thakare , Prajakta Kale, MadhuraGole, PriyankaNanekar, “Genetic K-means Algorithm for Credit Card Fraud Detection”, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (2) , 2015.
- [8] Yogesh Bharat Sonawane , Akshay Suresh Gadgil , AniketEknath More4 , NiranjanaKamalakarJathar , “Credit Card Fraud Detection Using Clustering Based Approach”, IJARIE-ISSN(O)-2395-4396 Vol-2 Issue-6 2016.