iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** V **Month of publication:** May 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Credit Card Fraud Detection Using Hidden Markov Model, Dempster–Shafer Theory and Bayesian Learning a Better Approach to Credit Fraud Detection

Montek Singh[1], Ashraf Zakee[2]
[1]VIT University, Vellore, India VIT University, Vellore, India
[2]Bhavya Joseph VIT University, Vellore, India

Abstract: ***Internet banking is a very common means of payments these days. It allows the users to instantly pay their dues but at the same time it also has some disadvantages; like online payment uses the credentials of the credit card holder to make the payment but if these details get leaked, then the funds can be misused by the hacker. Therefore we propose a credit card fraud detection system using a Dempster-Shafer adder and Bayesian learning and Hidden Markov Model. If an Incoming credit card transaction is not accepted by the trained DBM with sufficiently high probability, it is considered to be fraudulent. At the same time, we try to ensure that genuine transactions are not rejected. The developed system is simpler to implement and requires no training, when compared to the complex techniques like SVM and neural networks. The proposed algorithm is capable of updating the user's withdrawal behavior in real time. The development of such a system will not increase the faith of people in online transactions but at the same time will also encourage several small businesses and merchants to run their businesses online to earn better profits'.***

I. INTRODUCTION

The present day techniques are not efficient enough to track the sequence of credit card transactions and detect fraudulent transactions, as they were based on the old methods of banking. The main reason of developing a new algorithm is that the number of people using online banking services has significantly increased in last 3 yrs due to the launch of online shopping websites like Flipkart, Myntra, etc; where most of the users are not fully aware of the online transaction rules.

Such a situation creates an environment where hackers can exploit the loosely guarded resources to misuse the funds.

The main objective of this research is to model a sequence of operations in credit card transaction processing using a Dynamic Bayesian Network and Bayesian learning and show how it can be used for the detection of frauds Any violation in the system will lead into a trap where the suspected will have to enter the personal information related to the person making a transaction. If these details match the given information then the transaction will be accepted else it will be rejected.

The developed algorithm can very easily be integrated with the current banking techniques as the old algorithms can be replaced with the proposed algorithm with small changes to the Current system.

The developed system will be very useful for the Customers as well as the Merchants who are highly dependent on the online Banking for their Businesses. When this system gets operational over the internet the system can be further be used to track the suspected person using its IP address and inform the bank authorities about it.

II. RELATED WORK AND MOTIVATION

The Credit Card Fraud can be detected using numerous techniques which may include Hidden Markov Model, Dempster-Shafer adder**,** Bayesian learning neural network, support vector machine and neural networks.

Credit Card Fraud Detection by artificial neural networks in Ghosh and Reilly [1] and Aleskerov et al. [2] helped a particular bank to detect credit card fraud using artificial neural networks. They deduced the number of fraud transactions by 20-40%, which could have been detected. [2] Presents a technique, to detect credit card fraud by neural networks using Data Mining. The system includes the previous data of a customer and uses neural networks to predict the expenditure of the customer. The model has three techniques to represent customer's expenditure pattern– purchase category, transaction amount of item bought and last item bought time. The

model underwent tests. All the methods discussed above need loads of data to train the system first and then the system starts functioning'.

Syeda [3] suggested the use of granular neural networks to speed up data mining and knowledge discovery process. Mayes [4] has outlined an automated credit card fraud detection system by Bayesian belief networks (BBN) as well as Artificial Neural Networks (ANN). The research showed that BNN will give better results related to detect frauds and the training time is fast with ANN. Neural networking is designed to imitate the human brain. Therefore it is complex to implement is at first place and at the same time it requires a large size data set to initialize the system which will help in the machine learning process. The other disadvantages include its poor explanation capability, difficult to setup and operate and the need to convert every data to numerical vault between 0 and 1. The neural network based methods not fast but they can be accurate. Re-training neural networks is the main problem as the training time is quite high. Chen [5] devised a model which collected online questionnaire transaction data. An SMV was used to figure out new withdrawal. Chen has proposed a system which uses both ANN and SMV. It wishes to detect fraud without any data. However, it very complicated to develop a SVM but it will try. The main issue with SMV is that it requires huge amount of memory and has a high algorithmic complexity.

'Other researchers applied data mining algorithms to detect fraud. Chan [6] will divide the user information to predict his behavior and to predict his transactions. Data mining algorithms also require a large data set which will be unavailable if the customer has opened a new bank account few weeks ago. While data mining techniques are relatively accurate, they are inherently slow. One of the biggest disadvantages about Data Mining is that the database is extremely big and it has a complete history of every customer. This means that if there is any breach in the security the complete data can accessed by the hacker and he can misuse it which will have a seriously affect the customers. Therefore during its implementation the servers must be highly secure and at the same time redundant data must be stored to recover from any emergency situation. A fraudster might use numerous other ways to try and attack the system and find loopholes to exploit it. This will allow him to understand the actual functioning of the FDS.

Therefore if the system is meant to check only the variations then the above case gets neglected.

We have devised a new method which will use Hidden Markov Model with Bayesian Learning and Dempster-Shaffer method to detect such frauds. The system will evaluate multiple factors to reach the final conclusion. These factors will include the variation of the new transaction amount from the mean transaction amount, the variation of the address to which the product must be delivered and the variation of the amount of money spend by the user on different websites. E.g. consider a situation where a person spends more money on only one particular shopping website which is much more as compared to the other shopping web sites. Therefore if a large transaction is made from other website, the system will suspect danger and will ask the suspected person to enter more details. If the details are correct the transaction continues else it is cancelled'.'

III. PROPOSED FRAUD DETECTION SYSTEM

The functionality of the proposed system is divided into six hierarchical steps.

'Address mismatch- this part of the system needs an initialization database of the Bank Customer. Using this database the system generates a probability value for the number of times the customer's actual address in the database matches the shipping address. For example- if such a probability value is very high and then a new transaction is made with a different address then there is strong evidence that the new transaction might be a fraudulent transaction. When the new transaction gets verified the probability value of the database also gets updated. ShipProb- means the probability that the transaction is fraudulent on the basis of address mismatch'.

Transaction Amount Outlier- this part of the system also needs an initialization database of the Bank Customer. Using this information the system generates an average value for the amount of transactions made by the user. After this the Banking Administration will have to set a parameter. This parameter will decide the maximum allowed variation of the new transaction amount from the mean. This means that if the mean of the transaction amount is 100 and the parameter value is 0.2 (20%), then the maximum allowed amount is 120. Allowed here means that the system will not suspect such a transaction. If this variation value is more, then the chances that such a transaction is fraudulent are also more. Per- means the percentage of variation.

Dempster-Shafer adder (DSA) - The role of DSA is to combine the results obtained from the above two tests to obtain an overall believe value for each transaction.

A. Formula- Let (F, m_1) and (F, m_2) be two body of evidences given by two Experts.

Suppose A_1, A_2, ... , A_n be the fal elements. The combined body of evidence is constructed by assigning new BPA to each focal elements A_1, A_2, ... , A_n. This is done as follows:

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

For each A_i we associate a value m'(A_i) as

$$m'(A_i) = \frac{1-(1-m_1(A_i)\times(1-m_2(A_i)}{1+(1-m_1(A_i)\times(1-m_2(A_i)}$$

BPA assigned to A_i is-

$$m(A_i) = \frac{m'(A_i)}{\sum_n m'(A_i)}$$ ------(1This value can be interpreted as the combined evidence in support of the focal elements *Ai* .

B. Application

This theory is applied on Per and ShipProb values to obtain P(h) value. Based on this resultant value P(h), the transaction on a particular Card can be initially classified as normal, abnormal or suspicious. Therefore P(h) value calculates the degree of fraud.

Transaction history database- This database will store the information about good as well as the fraud transactions. The transaction gap is divided into 4 intervals- D1, D2, D3 and D4 depending on the time gap from last purchase.

D1 means that the time interval between the last and the new transaction is 8 hrs.

Similarly D2 means that this gap is between 8 to 16hrs.

D3 means that this gap is between 16 to 24 hrs.

And finally D4 means that this gap exceeds 24hrs. Di marks a new transaction.

Next P(Di|h) and P(Di|(bar)h) values are computed from the Fraud Transaction Database and the Good Transaction Database, respectively. P(Di|h) measures the probability of occurrence of Di given that a transaction is originating from a fraudster and P(Di|(bar)h) measures the probability of occurrence of Di given that it is genuine. The values of P(Di|h) and P(Di|(bar)h) are given by the following equations:

$$P(D_i|h) = \frac{\#(\text{Occurrences of } D_i \text{ in FTH})}{\#(\text{Transactions in FTH})}$$

$$P(D_i|\bar{h}) = \frac{\#(\text{Occurrences of } D_i \text{ on } C_k \text{ in GTH})}{\#(\text{Transactions on } C_k \text{ in GTH})}$$

The probability of the occurrence of the transaction Di can be calculated from the formula-

$$P(D_i) = P(D_i|h) * P(h) + P(D_i|\bar{h}) * P(\bar{h})$$

To calculate the probability of occurring a fraud transaction given a transaction will be calculated by combing the probability of the occurrence of the transaction Di and P(h) using Bayesian Learning.

Bayesian Learning- It is a tool used to update pervious believes when new information is available.

Therefore P(h|Di) calculates the probability of occurring a fraud transaction given a transaction i.e. it calculates the suspicion score.

$$P(h|D_i) = \frac{P(D_i|h) * P(h)}{P(D_i)}$$

By substituting we get:

$$P(h|D_i) = \frac{P(D_i|h) * P(h)}{P(D_i|h) * P(h) + P(D_i|\bar{h}) * P(\bar{h})}$$

Same as-

$$P(fraud|D_i) = \frac{P(D_i|fraud) * P(fraud)}{P(D_i|fraud) * P(fraud) + P(D_i|\neg fraud) * P(\neg fraud)}$$ ----- (2)

$$P(\neg fraud|D_i) = \frac{P(D_i|\neg fraud) * P(\neg fraud)}{P(D_i|\neg fraud) * P(\neg fraud) + P(D_i|fraud) * P(fraud)}$$ ----- (3)

Depending on which of the two values is greater (2 or 3), future actions are decided by the system.

'Final Value- If the value of (2) > (3) then Dempster-Shafer adder is used to combine (2) and P(h) value to give the final result. This final result value is used to declare that whether a transaction is fraudulent or not based on the parameter value. Example- if the parameter value is 0.7 and the value of the final result exceeds 0.7 then FRAUD DETECED, else NO FRAUD'.

IV. PROPOSED METHODOLOGY

The working of the system is based on the evidences it obtains from the multiple factors which can cause a fraudulent transaction. Given below is a list of factors which can help to predict a fraudulent transaction-

A. 'Address Mismatch

This factor if measured by shipProb. If the probability that the shipping address is same as actual banking address is very high then if a new transaction is made with an address different from the actual banking address then the probability that this transaction is fraudulent is very high. This quantity is labelled as shipProb. But unfortunately only this factor cannot be considered to declare a transaction is fraudulent or not, because for some users this data can have a lot of variation.

B. Transaction Amount Outlier

This factor is measured by per. To initialize this system the customer's database is needed. This database will have all the information about the transactions made by the user like the transaction amount. This information will be used to generate the mean of transaction amounts. When a new transaction is made the deviation from this mean value will be measured. This is based on the concept of Hidden Markov Model. If this value is very high, then the chances that the transaction is fraudulent are high. But this information is not practical in cases when the transactions made by the user have a lot of variation. To overcome this difficulty the Dempster-Shafer adder (DSA) is used'.

C. Dempster-Shafer adder (DSA)

It is used to combine the per and shipProb probability values to get combined believe value. This value is labelled as P(h) value. If this value is very high it means the probability that the transaction is fraudulent is also very high.

D. Transaction history database

If the fraudster is aware of the spending pattern of the user then the above value becomes useless. Therefore to avail maximum benefits the fraudster will have 2 options-

1) Make transaction with large amount where the gap between the transactions is more.

2) Make transactions with small amount where the gap between the transactions is less.

These factors can be measured by the P(h) value but the frequency of the transactions cannot be measured directly. For this the transactions are divided into 4 intervals D1, D2, D3 and D4. When a new transaction is made its time elapsed after the Last Good Transaction and Last Bad Transaction is measured. Using this data P(Di| h) and P(Di| hbar) values are calculated.

P(Di|h) is the probability that Di(new transaction) occurs given that a transaction is originating from a fraudster. Finally the P(h) and P(Di|h) values are combined using Bayesian Learning.

Bayesian Learning- As explained previously Bayesian learning is a method used to update pervious believes when new information is available. Therefore this tool is used to combine the P(h) and P(Di|h) values to obtain P(h|Di). P(h|Di) calculates the probability of occurring a fraud transaction given a transaction and P(hbar|Di) calculates the probability of occurring a good transaction given a transaction.

As these to values convey opposite meanings, one tells about the degree of fraud and the other value tells about the degree of genuine transaction.

E. Therefore if P(h|Di)> P(hbar |Di) then the chances of fraud transaction are mode.

Final Assessment- For cases when if P(h|Di)> P(hbar |Di) the P(h|Di) will be combined with P(h) value to get the final Suspicious score. This suspicious score is calculated used Dempster Shafer Adder.

This Suspicious Score will tell probability that whether this transaction is fraudulent or not. If its value is greater than the value of the parameter, then user will be asked to fill a verification form. If the data written in this form is correct then the transaction is made else the transaction will be cancelled.

F. Algorithm

supply Initializing Database

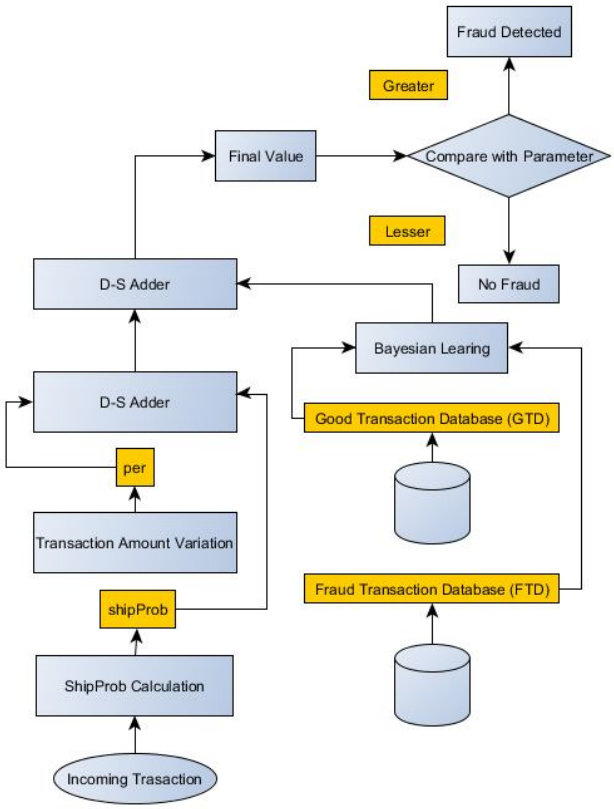
input: Transaction Amount

```
input: Shipping address
calculate shipProb
shipProb=address fraud detection probability
Trans=number of transaction made till date
addressEqual= true (if shipping address == billing address)
if((shipprob/trans)>=0.5 && addressEqual==true) then
        shipProb=(1-(shipprob/trans));
//address fraud probability
endif
else if((double)(shipprob/trans)<0.5 && addressEqual==true) then
        shipProb=(1-(shipprob/trans));
//address fraud probability
endif
else if((double)(shipprob/trans)>=0.5 && addressEqual==false) then
       shipProb=(shipprob/trans);
//address fraud probability
endif
else if((double)(shipprob/trans)<0.5 && addressEqual==false) then
        shipProb=(shipprob/trans);
//address fraud probability
endif
calculate variation amount (per)
sd=deviation from the mean
per=percentage of variation from the mean
Double sd=Math.sqrt(Math.pow((paymentAmount-mean),2));
Double per=(sd/mean);
// amount deviation fraud probability
P(h)=Dempster Shafer Adder ( shipProb, per)
//Using Eq.1
Calculate (P(h| Di) and P(hbar| Di)
// Using Eq. 2 and 3
if ( P (h| Di) > P(hbar| Di) ) then
            Final Value=Dempster Shafer Adder((P(h| Di), P(h) )     // Using Eq.1
            if (Final Value>0.7) then
                        OUTPUT (“Fraud Detected”)
             endif
            else
                        OUTPUT (“No Fraud Detected”)
endif
else
            OUTPUT (“No Fraud”)
```

V. SIMULATION AND RESULTS

As explained above the evaluation of the transaction is not based on a single value. The evaluation depends on the probability values obtain from the factors like – Address Mismatch, Transaction amount and the history of good and bad transactions. If the resultant value obtained is greater than parameter value, the user will have to enter the security check information. For this project the Parameter values was set as 0.7. This value can be set by the Banking Authorities. This information is stored into the Bank database. If the entered information is same as the bank database then the transaction is successful else the transaction will be cancelled.

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)



A. Database Initialization

The database for this project was created using ORACLE XE. Bank Customers details are set as follows:

FIRSTNAME set as “montek”; LASTNAME set as “singh”; USERNAME set as “montek”; PASSWORD set as “1”; AGE set as “19”; GENDER set as “M”; ADDRESS set as “vit”; ACCOUNTNUMBER set as “1”; BALANCE set as “38000”; TRANSCOUNT set as “20”; MEAN set as “1554”; SHIPPROB set as “18”.

Customer Security Question details are set as follows:

ACCOUNTNO set as “1”; QUESTIONNO set as “0”; ANSWER set as “bruno”.

Good Transaction History details are set as follows:

ACCOUNTNO set as “1”; D1 set as “4”; D2 set as “3”; D3 set as “4”; D4 set as “9”;

LASTTRANS set as “01:14:18”; DATETRANS set as “13-04-2017”.

Fraud Transaction History details are set as follows:

ACCOUNTNO set as “1”; D1 set as “8”; D2 set as “4”; D3 set as “2”; D4 set as “2”;

LASTTRANS set as “19:00:00”; DATETRANS set as “12-04-2017”.

Shopping Website Database details are set as follows:

NAME set as “montek”; USERNAME set as “montek”; PASSWORD set as “1”; EMAIL set as “montek01singh@gmail.com”.

1) Test Case 1

2) Initial status

At initialization of the test cases the values of FIRSTNAME, LASTNAME, USERNAME, PASSWORD, AGE, GENDER, ADDRESS and ACCOUNTNUMBER will be set as follows:

FIRSTNAME set as “montek”; LASTNAME set as “singh”; USERNAME set as “montek”; PASSWORD set as “1”; AGE set as “19”; GENDER set as “M”; ADDRESS set as “vit”; ACCOUNTNUMBER set as “1”.

Parameters that change for each test case will be BALANCE, TRANSCOUNT, MEAN and SHIPPROB.

For the given test case these values are:

BALANCE set as “38000”; TRANSCOUNT set as “20”; MEAN set as “1554”; SHIPPROB set as “18”.

A payment is made of Rs. 1800 with different address.

a) Result 1

When a payment is made of Rs-5000 with same address, fraud gets detected with per = 4.0; shipProb = 0.25; P (h) = 1.0; P (h|d) = 1.0; P (hbar|d) = 0.0.

As the 'per' value is extremely high therefore the chances of fraud are extremely high. Therefore, P(h) =1 and the fraud is detected.

In security form on entering the correct data the changes are made as follows:

In Good Transaction database, ACCOUNTNO set as "1"; D1 set as "2"; D2 set as "3"; D3 set as "4"; D4 set as "8"; LASTTRANS set as "10:42:24"; DATETRANS set as "13-04-2017".

In Bank Customer database, FIRSTNAME set as "montek"; LASTNAME set as "singh"; USERNAME set as "montek"; PASSWORD set as "1"; AGE set as "19"; GENDER set as "M"; ADDRESS set as "vit"; ACCOUNTNUMBER set as "1"; BALANCE set as "45000"; TRANSCOUNT set as "17"; MEAN set as "1235"; SHIPPROB set as "12".

The D1 value in Good transaction database gets incremented and the mean value in Bank Customer table gets updated.

2) Test Case 2

Initial status

At initialization of the test cases the values of FIRSTNAME, LASTNAME, USERNAME, PASSWORD, AGE, GENDER, ADDRESS and ACCOUNTNUMBER will be set as follows:

FIRSTNAME set as "montek"; LASTNAME set as "singh"; USERNAME set as "montek"; PASSWORD set as "1"; AGE set as "19"; GENDER set as "M"; ADDRESS set as "vit"; ACCOUNTNUMBER set as "1".

Parameters that change for each test case will be BALANCE, TRANSCOUNT, MEAN and SHIPPROB.

For the given test case these values are:

BALANCE set as "45000"; TRANSCOUNT set as "17"; MEAN set as "1235"; SHIPPROB set as "12".

A payment is made of Rs. 2000 with different address.

a). Result 2:

When a payment is made of Rs-2000 with same address, transaction is successful with per = 0.619; shipProb = 0.176; P (h) = 0.358; P (h|d) = 0.717; P (hbar|d) = 0.289; finalValue = 0.521.

As the 'finalValue' is less than 0.7 therefore allow transaction.

The changes are made as follows:

In Good Transaction database, ACCOUNTNO set as "1"; D1 set as "3"; D2 set as "3"; D3 set as "4"; D4 set as "8"; LASTTRANS set as "11:25:43"; DATETRANS set as "13-04-2017".

In Bank Customer database, FIRSTNAME set as "montek"; LASTNAME set as "singh"; USERNAME set as "montek"; PASSWORD set as "1"; AGE set as "19"; GENDER set as "M"; ADDRESS set as "vit"; ACCOUNTNUMBER set as "1"; BALANCE set as "43000"; TRANSCOUNT set as "18"; MEAN set as "1346"; SHIPPROB set as "14".

The D1 value in Good transaction database gets incremented and the mean value in Bank Customer table gets updated.

3) Test Case 3

Initial status:

At initialization of the test cases the values of FIRSTNAME, LASTNAME, USERNAME, PASSWORD, AGE, GENDER, ADDRESS and ACCOUNTNUMBER will be set as follows:

FIRSTNAME set as "montek"; LASTNAME set as "singh"; USERNAME set as "montek"; PASSWORD set as "1"; AGE set as "19"; GENDER set as "M"; ADDRESS set as "vit"; ACCOUNTNUMBER set as "1".

Parameters that change for each test case will be BALANCE, TRANSCOUNT, MEAN and SHIPPROB.

For the given test case these values are:

BALANCE set as "43000"; TRANSCOUNT set as "18"; MEAN set as "1346"; SHIPPROB set as "14".

A payment is made of Rs. 2100 with different address.

a) Result 3

When a payment is made of Rs-2100 with different address, fraud gets detected with per = 0.560; shipProb = 0.777; P (h) = 0.668; P (h|d) = 0.866; P (hbar|d) = 0.133; finalValue = 0.772.

As the 'finalValue' is greater than 0.7 therefore suspect fraud.

In security form on entering the wrong data the changes are made as follows:

In Good Transaction database, accountno set as "1"; D1 set as "3"; D2 set as "3"; D3 set as "4"; D4 set as "8"; LASTTRANS set as

"11:25:43"; DATETRANS set as "13-04-2017".
In Fraud Transaction database, ACCOUNTNO set as "1"; D1 set as "9"; D2 set as "4"; D3 set as "2"; D4 set as "1"; LASTTRANS set as "11:48:53"; DATETRANS set as "13-04-2017".
In Bank Customer database, FIRSTNAME set as "montek"; LASTNAME set as "singh"; USERNAME set as "montek"; PASSWORD set as "1"; AGE set as "19"; GENDER set as "M"; ADDRESS set as "vit"; ACCOUNTNUMBER set as "1"; BALANCE set as "43000"; TRANSCOUNT set as "18"; MEAN set as "1346"; SHIPPROB set as "14".
The Good transaction database and the Bank Customer table remain as it is but the Fraud transaction database gets updated.

4) Test Case 4
Initial status:
At initialization of the test cases the values of FIRSTNAME, LASTNAME, USERNAME, PASSWORD, AGE, GENDER, ADDRESS and ACCOUNTNUMBER will be set as follows:
FIRSTNAME set as "montek"; LASTNAME set as "singh"; USERNAME set as "montek"; PASSWORD set as "1"; AGE set as "19"; GENDER set as "M"; ADDRESS set as "vit"; ACCOUNTNUMBER set as "1".
Parameters that change for each test case will be BALANCE, TRANSCOUNT, MEAN and SHIPPROB.
For the given test case these values are:
BALANCE set as "43000"; TRANSCOUNT set as "18"; MEAN set as "1346"; SHIPPROB set as "14".
A payment is made of Rs. 1800 with different address.
a) Result 4
When a payment is made of Rs-1800 with different address, transaction is successful with per = 0.337; shipProb = 0.777; P (h) = 0.544; P (h|d) = 0.801; P (hbar|d) = 0.198; finalValue = 0.673.
As the 'finalValue' is less than 0.7 therefore allow transaction.
The changes are made as follows:
in good transaction database, accountno set as "1"; d1 set as "4"; d2 set as "3"; d3 set as "4"; d4 set as "8"; lasttrans set as "12:16:06"; datetrans set as "13-04-2017".
In fraud transaction database, accountno set as "1"; d1 set as "9"; d2 set as "4"; d3 set as "2"; d4 set as "1"; lasttrans set as "11:48:53"; datetrans set as "13-04-2017".
In bank customer database, firstname set as "montek"; lastname set as "singh"; username set as "montek"; password set as "1"; age set as "19"; gender set as "m"; address set as "vit"; accountnumber set as "1"; balance set as "41200"; transcount set as "19"; MEAN set as "1441"; SHIPPROB set as "14".
The Fraud transaction database remains as it is but the Good transaction database and the Bank Customer table get updated.

VI. DISCUSSION OF RESULTS

From Result 1- when a transaction amount (Rs.5000) is much greater than the mean of transaction amount (1000) then per value becomes extremely high i.e. greater than 1, because of which P(h) value becomes 1 and P(h| d) value also becomes 1. Therefore the fraud is detected. Such a type of situation is created when a fraudster tries to make a transaction with a very large amount not knowing thespending pattern of the user.
From Result 2- when a transaction of Rs. 2000 is made, when the mean is Rs. 1235 where the shipping address is same as the customer's bank information (shipProb=0.176) then such a transaction is allowed. Because the value of shipProb is very less, which means the probability of fraud due to address mismatch is very less. Although the per value is high (0.619) but the final value obtained is 0.512, which is less than 0.7 (parameter value). If this parameter value was less, then the system would have treated this transaction as a fraudulent transaction'.
'From Result 3- when the transaction of Rs.2100 is made, when the mean is Rs. 1346 where the shipping address is different from the customer's bank information (shipProb=0.777) then such a transaction is not allowed. Because the value of shipProb is very high, which means the probability of fraud due to address mismatch is very high. Although the per value is 0.560 but the final value obtained is 0.772, which is greater than 0.7 (parameter value). Therefore such a transaction is treated as a fraudulent transaction.
From Result 4- when the transaction of Rs.1800 is made, when the mean is Rs. 1346 where the shipping address is different from the customer's bank information (shipProb=0.777) then such a transaction is allowed. Although the value of shipProb is very high but the per value is 0.337. Therefore the final value obtained is 0.673, which is less than 0.7 (parameter value). If this parameter value

was less then, the system would have treated this transaction as a fraudulent transaction.

VII. COMPARATIVE ANALYSIS

'The base paper for this research was "Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning by Suvasini Panigrahi, Amlan Kundu, Shamik Sural a, A.K. Majumdar b". The hierarchical order of the factors presented in this research is very similar to factors considered by the base paper. But there is a significant amount of change when we talk about the ways these methods have been implemented in this research.

Let us consider: $\Theta_{LT} = 0.6$ and $\Theta_{UT} = 0.9$.

Let initially P (h) = 0.544 which is obtained by combining the evidences from the rules R1 and R2 stated in the paper. Since $0.544 < \Theta_{LT}$, the transaction is marked as good transaction. According to proposed method

A. Result

When a payment is made of rs-2100 with different address, fraud gets detected with per = 0.560; shipprob = 0.777; p (h) = 0.668; p (h|d) = 0.866; p (hbar|d) = 0.133; finalvalue = 0.772.

As the 'finalvalue' is greater than 0.7 therefore suspect fraud.

In security form on entering the wrong data the changes are made as follows:

in good transaction database, accountno set as "1"; d1 set as "3"; d2 set as "3"; d3 set as "4"; d4 set as "8"; lasttrans set as "11:25:43"; datetrans set as "13-04-2017".

In fraud transaction database, accountno set as "1"; d1 set as "9"; d2 set as "4"; d3 set as "2"; d4 set as "1"; lasttrans set as "11:48:53"; datetrans set as "13-04-2017".

In bank customer database, firstname set as "montek"; lastname set as "singh"; username set as "montek"; password set as "1"; age set as "19"; gender set as "m"; address set as "vit"; accountnumber set as "1"; balance set as "43000"; transcount set as "18"; mean set as "1346"; shipprob set as "14".

The good transaction database and the bank customer table remain as it is but the fraud transaction database gets updated. The methods used to calculate the probability values for each factor have been compared below'-

Transaction amount variation- the base paper uses a technique called as DBSCAN (density based spatial clustering of applications with noise) to measure the outlierness of a transaction. In this technique clusters are formed using the values from some attributes like transaction amount, billing address, shipping address and inter-transaction time gap. But the base paper uses this technique to define clusters only on the basis of Transaction Amount. To follow such an approach huge sets of data will be required to define all transaction clusters; otherwise for example if a transaction is made, but the amount does not fit in any of the cluster then such a transaction will be taken as a fraud transaction. For most of the cases the transaction amount do not vary so much that there is a need to define clusters. The similar results can be achieved by using variation in the mean of transaction amounts. This technique is much simplr to understand as well as apply in current systems.

Dempster–Shafer adder (DSA)- the formula used in the base paper works well on for some situations. If there are two doctors who agree after examining a patient that he is suffering from either of the three: meningitis (M), contusion (C) or brain tumour (T). Thus frame of discernment is {M, C, T}. Assuming that these two doctors agree with a low expectation of a tumor, but disagree in its cause and come up with the following diagnosis:

$$m_1(M) = 0.99,\ m_1(T) = 0.01$$
$$\text{and } m_2(C) = 0.99, m_2(T) = 0.01.$$

Based on Dempster rule of combination, we get the conclusion m(T)=1 which was not expected. This means that the patient suffers with certainty from brain tumour. This problem can be solved by using the formula explained in the methodology.

After finding the address mismatch and transaction variation, the Dempster Shafer adder is used to combine the results. In the base paper if the value of this combined result is less than a threshold value then the system declares this transaction as a genuine transaction, without calculation the transaction history of the customer. But the proposed system eliminates the use of this threshold value and goes for another test i.e. Bayesian Learning to measure the suspicious value. This makes the system more strong.

The base paper needs the following parameters Epsilon, MinPts, Theta-LT, Theta-UT, rou. But the proposed system needs only one parameter i.e. used to check the finalValue. Therefore the proposed system is more automated and less complex.

After a fraud in detected by the system, the user is asked to answer the security question. If this answer is correct then the

transaction is successful else the transaction is cancelled. This functionality is not provided by the base paper.

B. Comparison with other techniques

The most popular techniques used in detecting credit card fraud include neural network, data mining and support vector machine. Dempster–Shafer theory and Bayesian learning are old but simple techniques which require very less data to initialise the system. Results can be obtained by them very quickly when compared to other techniques as it involves very less mathematical calculations and at the same time they can be implemented to the current banking system with very less changes.

Neural Networking- neural networking is designed to imitate the human brain. Therefore it is complex to implement is at first place and at the same time it requires a large size data set to initialize the system which will help in the machine learning process. The other disadvantages include its poor explanation capability, difficult to setup and operate and the need to convert every data to numerical vault between 0 and 1.

Methods based on neural networking are ont efficient because although they are fast they are not accurate. To eliminate this problem the neural networks have to be retrained which is again a tiresome task.The proposed system has no training time. All the information required to calculate the probabilities is provided in the database itself.

Data Mining- Data mining algorithms also require a large data set which will be unavailable if the customer has opened a new bank account few weeks ago. This problem is faced by the proposed system also. Although data mining methods are more accurate, they are inherently slow. One of the biggest disadvantages about Data Mining is that the database is extremely big and it has a complete history of every customer. This means that if there is any breach in the security the complete data can accessed by the hacker and he can misuse it which will have a seriously affect the customers. Therefore during its implementation the servers must be highly secure and at the same time redundant data must be stored to recover from any emergency situation. All these problems are present in the proposed system, but it is much simpler when compared to the data mining.

Support Vector Machines- First of all it very complicated to develop a SVM but it will try to prevent fraud for users even without any transaction data but these systems cannot be made fully automated so they depend on the user's expertise level. SVMs are problematic because the quadratic programming so done reuires high level of algorithmic complexity and memory requirements that are extensive.

Therefore after understanding the problems of the alternative approaches, the Dempster–Shafer theory and Bayesian learning appear to be the best methods to develop a simple but efficient algorithm to detect the fraudulent credit card transaction.

VIII. THREATS TO VALIDITY

Our proposed methodology does not work efficiently until it has huge amount of data fed into it. The more the system is used for transaction the more efficient it will be in detecting fraudulent transactions.

There have been no methods evolved to detect fraudulent transactions even if the customer is new to the website as well as bank payment portal.

The databases involved in the current system store a lot of sensitive information. If any security breach occurs then serious problems can occur.

If the user makes highly varied transactions then the variation concept will not be much useful. Clusters based on appropriate attributes will be needed in such cases.

After the transaction has been labelled as a fraud transaction then the complete system depends on the answer of one security question. This is a bottleneck situation. If the fraudster knows the answer the complete system fails. A solution to this problem involves the use of OTP code generation.

The factors considered in the research are common. In today's world people use the different shopping websites for buying specific things, therefore a system much learn that what user buys from which website. This will create possibilities for defining new factors.

The important part for the functioning of the system is the parameter value. If this value is high the system will allow fraud transactions but if its value is low the system will generate unnecessary alerts.

IX. CONCLUSIONS AND FUTURE WORK

The main objective was to develop a fraud detection system which is able to understand the pattern of frauds but at the same time does not generate false alarms. The proposed method is simpler to understand and apply in the real world scenario without making many changes to the current databases. Therefore such a system can minimize the losses of the credit card company.

he fraud detection system is based on two concepts- Dempster- Shafer theory and Bayesian learning. Both these methods are used to

Volume 5 Issue V, May 2017
ISSN: 2321-9653

combine any two believes to get a single value. The proposed system is divided into 6 hierarchical steps which are interrelated to each other. The main advantage of such a system is that any new factor can be incorporated to this system easily. Therefore the developed system is flexible and can be updated according to the future requirements. A theoretical comparison of this system with pre existing systems reveal that significant accuracy and performance was obtained mainly due to the use of Dempster- Shafer theory and Bayesian learning. With the results we have received we come to the conclusion that the idea of comparing various cases and believes are the best for addressing such a modern problem as it covers all the complex possibilities in the system.

The developed system is highly flexible. Multiple factors can be added to it according to the future requirements. These days people use the different shopping websites for buying specific things, therefore a system much learn that what user buys from which website. Using this information a stronger system can be made. Other methods for compiling evidences like Bayesian combination network can be used too.

Future scope of the work includes its application in real banking systems with improvements to include and process large amounts of data that will come along with linking of various online shopping websites and banks offering net banking

REFERENCES

[1] S. Ghosh, D.L. Reilly, Credit card fraud detection with a neural-network, in: Proceedings of the Annual International Conference on System Science, 1994, pp. 621–630.

[2] E. Aleskerov, B. Freisleben, B. Rao, CARDWATCH: a neural network based database mining system for credit card fraud detection, in: Proceedings of the Computational Intelligence for Financial Engineering, 1997, pp. 220–226.

[3] M. Syeda, Y.Q. Zhang, Y. Pan, Parallel granular neural networks for fast credit card fraud detection, in: Proceedings of the IEEE International Conference on Fuzzy Systems, 2002, pp. 572–577.

[4] S. Maes, K. Tuyls, B. Vanschoenwinkel, B. Manderick, Credit card fraud detection using Bayesian and neural networks, in: Proceedings of the First International NAISO Congress on Neuro Fuzzy Technologies, 2002.

[5] R.C. Chen, M.L. Chiu, Y.L. Huang, L.T. Chen, Detecting credit card fraud by using questionnaire-responded transaction model based on support vector machines, in: Proceedings of the Fifth International Conference on Intelligent Data Engineering and Automated Learning, vol. 3177, October 2004, pp. 800–806.

[6] P.K. Chan, W. Fan, A.L. Prodromidis, S.J. Stolfo, Distributed data mining in credit card fraud detection, in: Proceedings of the IEEE Intelligent Systems, 1999, pp. 67–74.

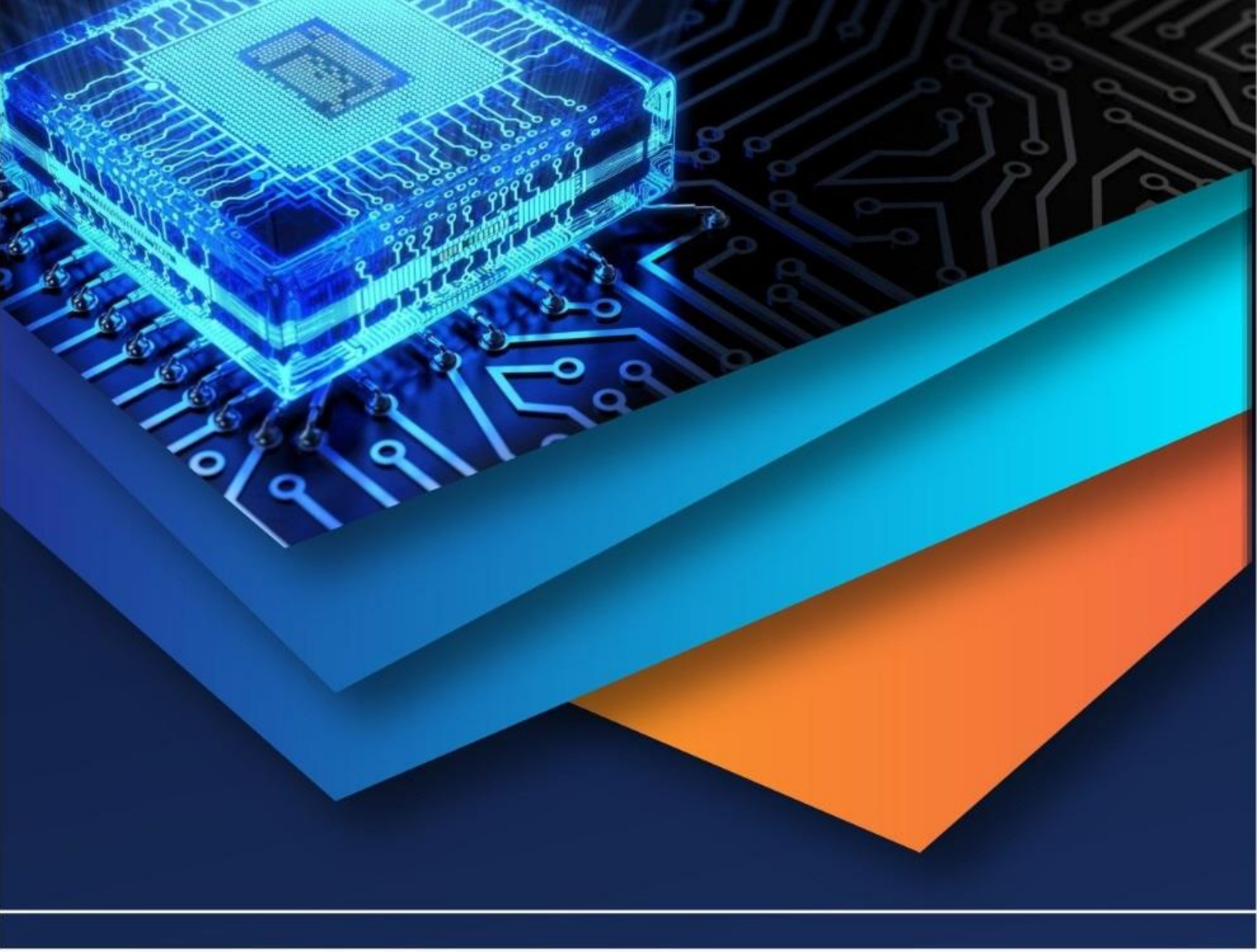
[7] Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning by Suvasini Panigrahi, Amlan Kundu, Shamik Sural a, A.K. Majumdar b

[8] Online fraud is 12 times higher than offline fraud, 20 June,2007. <http://sellitontheweb.com/ezine/news0434.shtml>.

[9] J.R. Dorronsoro, F. Ginel, C. Sanchez, C.S. Cruz, Neural fraud detection in credit card operations, IEEE Transactions on Neural Networks 8 (July) (1997)

[10] R.C. Chen, S.T. Luo, X. Liang, V.C.S. Lee, Personalized approach based on SVM and ANN for detecting credit card fraud, in: Proceedings of the IEEE International Conference on Neural Networks and Brain, October 2005, pp.810–815.

[11] T.M.Chen, V. Venkataramanan, Dempster–Shafer theory for intrusion detection in ad hoc networks, in: Proceedings of the IEEE Internet Computing, November–December 2005, pp. 35–41.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY