Fraud Detection by Monitoring Customer Behavior and Activities

Parvinder Singh
Assistant Professor
MMICT and BM
MM University, Mullana, Ambala, India

Mandeep Singh
Associate Professor
Department of Computer Science and Engg.
Chandigarh University, Mohali, Punjab, India

ABSTRACT
With the enhancement in technology e-banking like credit Card, Debit Card, Mobile Banking and Internet Banking is the popular medium to transfer the money from one account to another. E-Banking is gaining popularity day by day, which increases the online transaction with the increase in online shopping, online bill payment like electricity, Insurance Premium and other charges, online recharges and online reservation of railways, bus etc., so the fraud cases related to this are also increasing and it puts a great burden on the economy, affecting both customers and financial bodies. It not only costs money, but also a great amount of time to restore the harm done. The purpose is to prevent the customer from online transaction by using specific technique i.e. based on Data Mining and Artificial Intelligence technique. The risk score is calculated by Bayesian Learning Approach to analyze whether the transaction is genuine or fraudulent based on the two parameters: Customer Spending Behaviour and Geographical Locations. The customer than spending behaviour that can be identified by KMEAN clustering algorithm and in geographical location the current geographical location is compared with the previous location. If risk score is greater 0.5 then transaction is considered to be fraudulent transactions and then the security mechanism authenticates the user by entering the 4 digit random number that appears on the screen and the genuine user enters the code in a correct manner.

Keywords
Unique, Hybrid, Logical, Data Mining, Artificial Intelligence, Bayesian Learning Approach, K Mean algorithm, Authenticate.

1. INTRODUCTION
When any crime is committed on the Internet it is termed as cyber crime. Cybercrime generally are of two types:

1. Crimes that target computers directly.
2. Crimes facilitated by computer networks or devices.

When the profit on investment is high and less probability of risk, the people usually want to take advantage of such solution. This is exactly that happen in cyber crime. Obtaining sensitive information and data and using it to perform transactions and to catch such criminals is difficult. Therefore, this has resulted to increase in cyber crime across the world.

Fraud can be defined as the criminal activity i.e. committed by the criminal in order to obtain financial/personal gain. Fraud refers to the act of deceiving the people to obtain the advantage/benefit by using the name of another person. Fraud can be conducted by the use of digital computer i.e. termed as Cybercrime. In cybercrime the computer is used as a medium to perform illegal activities like committing fraud, trafficking in child pornography etc.

Due to the growth of modern technology, the mode of payment of individual has changed significantly. The use of Online Payment mode such as Online Banking, Debit Card, Credit Card etc. has become popular and is becoming important in day to day activities because it allows bank customers to purchase goods and services from the shopping websites or from the market.

Fraud deals with cases that happen due to criminal purpose which are difficult to identify. Fraud can be mainly divided into two types:

- **Offline Fraud:** Most of the offline fraud incidents occur due to the steal of purse/wallet that contains important documents. Documents such as Driving License, ID card etc. contains crucial information such as name, date of birth, transaction slips etc.
- **Online Fraud:** Online fraud occurs when fraudster present their website as a genuine website in order to obtain crucial personal data of a customer and perform illegal transactions on such customer account.

Credit card is also one of the most illegal types of fraud. Credit Card is a plastic card i.e. issued to customers of a bank as one of the mode of payment. It allows cardholders to purchase goods and services from the shopping websites or from the market. Credit Card Fraud is defined as, when an individual uses another individual credit card for personal use while the owner of the card as well as the card issuer are not aware of the thing that the card is being used.

Fig1.1(a) shows, If credential information of a customer has stolen and uses for online shopping, the card holder acknowledges transaction details after the fraud has been committed and then customer inquire the bank for transaction. There is no such process that can prevent fraudulent transaction at the time of happening. So there is a need of such interface that prevents from online transactions.

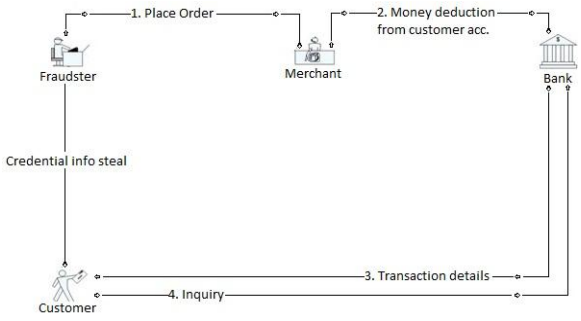

Fig 1.1: Example of Credit Card Fraud

1.1 Methods to Steal Personal Information:

There are various methods or techniques that cyber criminals used to commit the crime:

- **Hacking**: Hacker is a person who seeks and exploits weakness in computer system. Attacker breaks into industry or personal databases.
- **Phishing:** Phishing is a fraudulent attempt, usually made through email, to steal your personal information.
- **Spoofing:** The word "spoof" means to hoax, trick, or deceive. Spoofing refers to tricking or deceiving the computer users. This is typically done by hiding one's identity or faking the identity of another user on the Internet.
- **Spyware:** The computer user unknowingly downloads software from the Internet that contains spyware. Spyware collects personal information from your computer and transmits it to fraudster or attacker.
- **Shoulder Surfing:** An attacker watches a bank customer from a nearby location as the customer punches in his personal information. If the customer is giving his personal information over the phone (e.g., to a hotel or car rental company), the attacker may listen to the conversation so as to obtain personal information of bank customer
- **Dumpster Diving:** An attacker goes through a customer's garbage cans or trash bins to obtain personal information of bank customer such as bank statement, payment receipt etc.

1.2 Security Mechanism to Prevent from Fraud

These are the existing security mechanism that helps to prevent from fraudulent transactions:

- **Address Verification Service (AVS):** In this technique it matches the cardholder billing address and shipping address and identifies whether the cardholder has purchased product on this address. However, this technique contains some weaknesses i.e. the address information is available online; the banker feels boring to check record of every customer to prevent from fraudulent transaction; it cannot check the entire informational card.
- **Fraud Rates:** This technology checks for recognized patterns i.e. used by the fraudster to commit the fraud. The advantage that it is easy to configure and understand, but the disadvantage

  fraudster changes their pattern changes at regular interval.
- **Relocation:** This technology identifies the customer geographic location by identifying its IP addresses.
- **Chip & Pin:** A PIN is a 4 digit unique and secret number that customer has to enter before doing transaction by ATM/Debit Card/Credit Card. The 4 digit pin is used to identify whether the customer is genuine or not.
- **3D-Secure:** This technology works on the principle of authenticating the user password with the password i.e. stored in the database. The main advantage of this system is that fraudster needs a user's password to perform the transaction.
- **Biotechnology:** The unique characteristic of each customer such as fingerprints, voice, signature, iris, and other similar biological components is stored in a computer so that a computer can read it. Then the computer compares the stored patterns to the person who is performing the transaction to identify whether the customer is genuine. The main disadvantage of this technology is that is that it requires additional hardware cost.
- **One Time Password:** The random number is generated at server side and is send to the customer's mobile phone through the help of the web services to ensure that the correct user is performing the transaction at that instant of time. The user has to enter the same password for getting the authorization from the bank side..

1.3 Data Mining:

Data Mining is a technique that uses statistical, artificial intelligence, and neural network technique to extract and fetch useful information from a large database [30]. Data Mining is a technique to study data from different views and summarizes it into crucial information. Data Mining is a technique that is used to detect financial fraud detection because it can identify new attacks before financial fraud can be detected by human experts [31].

1.4 Challenges in Data Mining to detect Fra-ud:

- There are millions of transactions each day. To extract large amount of data from a database requires highly efficient techniques.
- The data or information is noisy.
- Data labels are not immediately available. Frauds or intrusions usually aware after they have already happened.
- It is hard to track user's behaviours. All types of users (good users, business, and fraudsters) change their behaviours frequently

2. RELATED WORK

Quah.John.TS [26], author evaluates a neural network based approach i.e. called SOM (Self Organizing Map) to detect spending pattern of the customer in credit card database and SOM is a multilayer approach that consists of: The initial Authentication, Screening layer, Risk Scoring layer, Behavior Analysis layer (Core Layer) and Decision making layer and The main purpose of SOM approach is to classify and cluster input data, to detect and derive hidden patterns in input data.

Srivastva.Abhinav [29], HMM (Hidden Markov Model) is proposed to detect fraudulent transactions which is initially trained with the normal behavior of a cardholder therefore if an incoming credit card transaction is not accepted by the HMM with sufficiently high probability, it is considered to be fraudulent and K Mean Clustering algorithm is used to identify spending behavior of a customer. HMM-based applications are used in various areas such as speech recognition, bioinformatics, and genomics so this model is able to detect fraud in large volumes of transactions.

Panigrahi.Suvasin [24], Fraud Detection System is proposed i.e. based on the combination of three approaches: Rule-based filtering, Dempster–Shafer theory and Bayesian learning in which Dempster rule is used to match customer

current behavior compared with the previous behavior ,rule based filtering approach is used to determine the suspicious level of each incoming transaction and Bayesian learning approach is used to update the suspicious score of transaction using history database of both genuine cardholder as well as fraudster.

Sanchez's. [28], Association rules (Fuzzy Rules) are used to detect new, undesired behavior of bank customer in the online verification process and Association Rules (Fuzzy Rules) are applied in the area of Business Management and planning to extract data of fraudulent transaction from a large database.

Farvaresh.Hamid[11] , a framework was proposed to detect fraud telecommunication subscribers by using various techniques such as data cleaning, dimension reduction, clustering and classification and the main problem in this framework is that it requires the historic data to identify whether the customer is fraudster or genuine.

Bhattacharyya.Siddhartha[1],author evaluate two advanced data mining approaches, decision tree approach, support vector machines and random forests together with logistic regression to detect credit card fraud and examines the performance of these techniques with the varying level of data under sampling and these techniques can detect only few fraudulent transaction when it is applied to a real world data set.

Duman.Ekrem [9], a technique was evaluated such as genetic algorithm and scatter search to score each transaction and based on these scores the transaction can be classified as fraudulent or genuine transaction and these approaches are based on the classification problem.

Dharwa.N.Jyotindra [7], Transaction Risk Score Generation Method was used to calculate certainty factor to identify whether the transaction is fraudulent or genuine and Risk score is analyzed by identify spending profile of customer of a bank by implementing DBSCAN algorithm and address mismatch in which it will identify whether the customer billing and shipping is same .

Cao.Longbing.Zhang. [2], combined mining as a general approach to extract informative knowledge in complex data to solve problems the problems of enterprise applications, such as telecom fraud detection and cross-market surveillance in stock markets and a framework that is flexible and customizable for handling a large amount of complex data.

Joris.Maervoet [18], a tool was proposed extract data for relational regularities and corresponding outliers in relational data based on WARMR algorithm whose input consist of interpretations, a background knowledge and a language bias, generated from the end user's data selection and mining preferences, so this tools assists a geographic content providing company in reasoning about the structure of the data and about the data itself and this tool is able to extract previously unknown knowledge in an automated way, which can be integrated in the quality maintenance process directly.

Li.Jinjiu.Wei.Wei. [22], online banking fraud detection framework recommended i.e. based on utilizing resources and advanced data mining techniques and algorithms such as contrast pattern mining, neural network and decision forest are implemented and their outcomes are integrated with an overall score measuring the risk of an online transaction being fraudulent or genuine.

Phua.Clifton [6], the two new layers such as Communal Detection (CD) and Spike Detection were used where CD is a white list oriented approach on a fixed set of attribute and finds real social relationships to reduce the suspicion score and SD is attribute oriented approach on a variable size set of attribute and find spikes in duplicates to increase the suspicion score to detect more types of attack on and remove the redundant attribute.

Edge.E.Michael [10], FFML a design was presented i.e. a rule-based policy modeling language where FFML is an architecture for facilitating the conceptual level expression implementation of proactive fraud controls within multi-channel financial service platforms by using approaches such as data mining, neural network, and machine learning techniques to identify suspicious transaction and these techniques supports real time monitoring component that finds financial fraud in areas of stock market ,money laundering etc.

R.Huang [15], a hybrid model recommended for online fraud detection of Video-on-demand system to improve the current Risk Management Pipeline (RMP) by adding Artificial Immune System (AIS) based fraud detection for logging data in which AIS based model combines two artificial immune system algorithms with behavior based intrusion detection using Classification and Regression trees (CART),so the proposed approach can help e-commerce better understand the issues and plan the activities involved in a systemic approach to E-fraud.

Drezewski.Rafal [8], a system supporting money laundering in which the imported data are used by the clustering algorithm that is the basis for money transfer and six algorithms that extract sets and sequences were implemented; these algorithms is able to provide all the frequent patterns ,closed or even the maximal ones, so this system is able to detect criminal i.e. is a complex process by gathering information from different sources may be of quantative character such as billing or bank account transactions but also of qualitative character such as eyewitnesses testimonies and the results obtained by using these algorithms can be visualized so that they can be easily explored by the police analyst and trace the criminals.

Chang.W.H. [4], author recommended a early fraud detection method i.e. based on the accuracy and timeliness simultaneously; this method prevents from online fraud in which fraudster deceive the individual, business or the organization by performing fraudulent transaction and early fraud detection method can be implemented by constructing decision trees or by instance-based learning.

Wu.Shu [32], two practical was proposed for outlier detection methods named ITB-SS and ITB-SP to solve real world problem such as Intrusion Detection, Criminal activity detection in E-Commerce etc. and these methods does not require user defined parameter to decide whether an object is outlier and author also proposed a new concept called 'Holoentropy' that takes both data and total correlation into consideration.

Sahin.Yusuf [27], the security mechanism such as CHIP and PIN are developed for credit card system that does not prevent from fraudulent credit card usages over online fraud and the author have developed and implemented a cost sensitive decision tree approach to detect fraudulent transactions and this approach is compared with the traditional classification models on a real world credit card data set.

Kim.Chan.Ae[21],digital forensics techniq-ues has been recommended to analyze system intrusion incidents traditionally is used to detect anomaly transactions that may occur in the user environment during electronic financial transactions and the risk point calculation model is proposed by scoring anomaly transaction cases in the detection step by items.

Kim.Kwanho [19], two methods was proposed i.e. based on the finite fixture model to detect fraudulent items automatically without the requirement of labeled items and modeled the dependency between the description and the price of an item by considering the possible combinations of the item description and price clusters according to item clusters and a real-world dataset to evaluate the effectiveness of the proposed models and compared them to existing outlier detection methods so the proposed model significantly identifies the fraudulent transactions. The proposed model requires further enhancement by utilizing historical logs to more accurately analyze the customer behavior.

Hajian.Sara [16], a new pre-processing discrimination prevention methodology consisting of different data transformation methods that can prevent data discrimination, indirect discrimination or both of them at the same time; in order to attain this objective the first step is measure discrimination, and identify categories and groups of individuals that have been directly and/or indirectly discriminated in the decision-making processes; the second step is to transform data in the proper way to remove all those discriminatory biases and finally discrimination free data model can be produced from the transformed data set without damaging the actual data. The proposed technique is successful in removing discrimination and preserving data accuracy.

Lee.YJ[33], In this paper, online oversampling principal component analysis algorithm (OSPCA) was recommended to solve real world applications problem such as intrusion detection or credit card fraud detection and the aim of the author is to detect the presence of outliers from a large amount of data via an online updating technique. The author proposed framework is favored for online applications which have computation and memory limitations. The author compared OSPCA algorithm with other anomaly detection algorithm and OSPCA algorithm provides better accuracy and efficiency as compared to other anomaly detection algorithms The issues in proposed anomaly detection algorithm are: normal data with multiclusterin structure, and data in an extremely high dimensional space.

Pozzolo.A.D [25], AP, AUC and Precision Rank as performance measure for a fraud detection task method was recommended.The proposed algorithm reduces the risk i.e. faced by the customer of a bank due to fraudulent transaction of credit card fraud and to reduce the losses the algorithm depends on advanced machine techniques to assist fraud investigator. The author proposed the algorithm to solve the problem of non stationary distribution of data, highly imbalanced classes distributions and the continuous stream of transactions. There are three main issues: unbalancedness, non stationarity and assessments and the advanced machine learning technique depends on three main factors: data distribution, classifier used and assessment. The proposed framework addresses the problem of non-stationary in data streams by creating a new model every time a new chunk is available.

3. DATA MINING TECHNIQUES

Fraud Detection has been usually seen as a data mining problem where the objective is to correctly classify the transactions as legitimate or fraudulent.

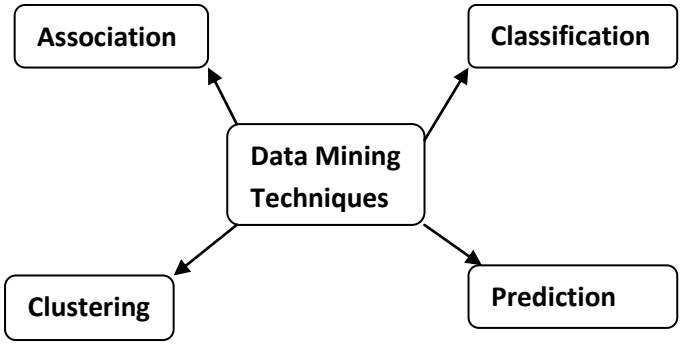

Fig 1.2: Data Mining Techniques

- **Association:** In association, patterns are discovered based on a relationship of a particular item on other items in the same transaction [27].For example, the association technique is used in market basket analysis to identify what products that customers frequently purchase together [10].
- **Classification:** Classification is a data mining technique i.e. based on machine learning. Classification is used to classify each item in a set of data into one of predefined set of classes or groups [7]. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics [27]. For example, Classification in application that "given all past records of employees who left the company, predict which current employees are probably to leave in the future." In this case, classification divides the employee's records into two groups that are "leave" and "stay" [10].
- **Clustering:** A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. For example, in a library, books have a wide range of topics available. The challenge is how to keep those books in a way that readers can take several books in a specific topic without hassle. In clustering technique, the books that have some kind of similarities in one cluster or one shelf and the books that are different in other cluster or in another shelf.
- **Prediction:** The prediction is one of a data mining techniques that discover relationship between independent variables and relationship between dependent and independent variables. In this technique, prediction analysis technique can be used in sale to predict profit for the future if we consider sale is an independent variable, profit could be a dependent variable [10].

Fraud detection is based on three approaches:

- **Supervised Approach:** In supervised approach, the fraud can be detected by analyzing the previous records of customer and compare these records with the current transaction. It uses Classification technique to detect fraud.
- **Semi Supervised Approach:** Semi supervised can be used in interactive learning system. It uses Prediction technique to predict all the outliers (unusual behavior

from a normal data set) from the input data set based on an outlier sample indicated by the end user.

- **Unsupervised Approach:** This Approach is used to detect unusual observation such as transactions, customers or accounts whose behavior may be different from the normal.

The fraud detection techniques are based on Supervised, Semi Supervised and Unsupervised approach and these techniques are:

Genetic Algorithm: Genetic Algorithm is based on the sexual reproduction in which the genes of two parents combined to form to those of their children. When this algorithm is applied to any problem the basic premise is that we can create an initial population of individual representing possible solutions to a problem. Each of those individual has certain characteristics that make them more or less fit as member of population. The fit member will have a higher probability of producing efficient solution. This method is very effective at finding optimal or near optimal solution.

Neural Network: Neural Network is an Artificial Intelligence technique that is used in biological learning system. Artificial Neural Network is composed of large number of highly interconnected nodes to solve the specific problem. Example: Neural Network can be used to differentiate legal transactions from fraudulent transaction, detect fraudulent behavior on an e-commerce site [34].

- **Self Organising Map (SOM):** SOM is a based on neural network technique i.e. based on the unsupervised learning approach. In SOM historical data present in the database gets classified into genuine and fraudulent sets through the process of self-organization and any new incoming transaction is pre processed and fed to the SOM [26]. Based on the threshold value, fraudulent or genuine transaction is identified.SOM is a multilayered approach consisting of [26]:
- The initial authentication and screening layer.
- The risk scoring and behavior analysis layer.
- Decision-making layer.

Bayesian belief network (BBN): BBN are graphical representations of probability distributions, derived from co-occurrence counts in the set of data items [30]. BBN is a directed, acyclic graph, where the nodes represent attribute variables and the edges represent probabilistic dependencies between the attribute variables. Bayes theorem gives the mathematical formula to detect fraud in the form of posterior probability. [7]

$$P(E|fraud) = P(E|fraud)*P(fraud)/P(E)$$

Support Vector Machines: Support vector machines (SVMs) are statistical learning techniques that can be used in a classification tasks. This technique is based on the supervised learning algorithm. An SVM model is a representation as points in space, and different points are mapped so that the separate categories are divided by a clear gap that is as wide as possible.

Logistic Regression: Logistic regression is a probabilistic statistical classification model. Logistic Regression is used to measure the relationship between the dependent & independent variable.

Card Based Reasoning: Case-based reasoning (CBR) is a technique that solves a given problem by using past experience and solutions of particular problem.

Decision Tree: Decision tree are also called Classification tree or Binary tree. It works on a binary choice that means 'yes' or 'no'[27]. A Decision tree is built by a recursive process from top to bottom. A decision tree can be used in a classification problem and is used to detect credit card fraud. For example, spending behavior of a customer can be represented by decision tree approach such as spending profile represents the root node and high, medium, & low spending profile represents the leaves of a tree. Fraudulent or genuine transaction may be identified by the Decision Tree approach.

- **Hidden Markov Model (HMM):** A HMM consists of a finite number of states; each state is associated with a probability distribution. Transaction among these states is identified by a set of probabilities called transition probabilities [29]. HMM model is able to detect credit card fraud.HMM model does not need biometric recognition to identify fraud. HMM model detects fraudulent transaction just by remembering customer spending behavior.

  In this process, the spending profile of customer can be categorized into three parts:

  1) Low Spending Profile
  2) Medium Spending Profile
  3) High Spending Profile

4. PROPOSED WORK

There are some cases that can analyze whether the transaction is suspicious or genuine:

- If the user is taking long time to perform the transaction than the routine time (number of times that the user can take time) then there may be a chance that user is suspicious means user is fraudster**.** Example: Suppose the bank customer normally takes 60 seconds to perform the transaction and suddenly the user is taking 10 minutes to perform the transaction then the user is found to be suspicious.
- Suppose the user performs the transaction in India at 11:00 AM by providing his details and same user details are used in England at 12:00 AM then there may be possibility that transaction is fraudulent.
- A customer performs similar type of transaction in terms of amount and suddenly the customer is purchases an item from shopping website which is costly then there may be possibility that transaction is fraudulent.

Existing system is based on two parameters i.e. Address Mismatch and Spending Behavior to prevent from fraudulent transactions. In case of address mismatch, the user shipping address is compared with the delivery address and risk score is calculated by Bayesian learning approach. In case of Spending Behavior, it will identify the customer previous transaction and compares the current transaction amount with the previous transactions amount. Therefore in existing system of fraud detection the fraudulent transaction can be identified once the fraud has happened. It is difficult to find out fraudulent transaction once it has done.

- It is difficult to track user behavior because the user behavior changes constantly.

- The main problem is that the existing fraud detection will give the false alarm that means transaction is fraudulent even if the transaction is genuine

The aim is to propose a security system that can prevent from fraudulent transactions. The security mechanism must be able to identify whether the transaction is genuine or fraudulent. In case of fraudulent transaction the security system must be able to prevent from online transactions. The main objective is to identify the customer behavior in case of spending behavior and their location. The customer spending behavior can be identified with the help of KMEAN clustering algorithm. In case of spending behavior the customer usually performs similar type of transactions in terms of amount which can be visualized as a part of cluster and suddenly the customer performs transaction of huge amount which can be seen as outlier. In case of geographical location, suppose the customer performs transaction in India at 10AM and at 10:30AM the transaction is performed from Poland by using the same details. It is impossible that customer can reach that location in a short time period. The advantages of proposed system are:

1. The detection of the fraud is found much faster than the existing system.
2. We can find the most type of fraud by using TRSGM.

There are some rules that help to identify fraudulent transaction:

1. If the transaction has performed during the late night and no past transaction exist in late night, and then transaction is considered as suspicious transaction.
2. If the user is active and performs the transactions continuously, but suddenly stops performing the transaction and after some time becomes again active to perform transaction, then transaction is considered as suspicious.
3. Generally customer doesn't like to purchase the costly and luxury product but suddenly customer purchases a costly and luxury product, then transaction is considered to be suspicious.
4. Overseas transaction is considered as sensitive transaction if in the past no overseas transaction has done on the same account.

The spending profile of a cardholder identifies the spending behavior of a cardholder. Cardholders can be broadly categorized into three groups based on their spending habits, namely, high-spending group, medium-spending group, and low spending group [18]. Cardholder who spent high amount on purchasing the goods/ services belongs to the high spending group and cardholder who belong to the low spending group purchase low priced item. The spending profile of a customer is analyzed by K-Mean Clustering algorithm. K-Mean is an unsupervised learning algorithm. K Mean Clustering algorithm groups the data based on the similarity of their attribute values. The groups formed by Mean Clustering algorithm is referred to as cluster. The grouping is formed based on the square of distance and centroid of their data values**.**

Step1: Compute the centroid of the cluster .

Step2. Compute the distance between the object to the centroid

Step 3.Grouping is done on the basis of minimum distance between each point.

An incoming transaction is first checked for the address mismatch. If current geographical location and previous geographical location is found same, then the transaction is considered to be genuine and is approved and no other check is performed. If current geographical address and previous geographical address is different then it will analyze that whether the past transactions are successfully performed on the same location. If products are successfully shipped on the current location, then it considers the transaction highly genuine and generate risk score 0. If this is first transaction on the given location then the incoming transaction amount is checked with the clusters formed by KMEAN clustering algorithm for its coverage. If the risk score < 0.5, the transaction is considered to be genuine and is approved. On the other hand, if risk score > 0.8 then the transaction is declared to be fraudulent.

The risk score is determined by following equation:

Risk Score= (1-threshold)$\sum_{n}^{i=1} Pi * Wi$ [7], where threshold=0, Pi=Parameter, W*i*=Weightage of the parameter which is given as input.

Table 1.1: Parameters and Weightage to calculate Risk Score

Sr. No.	Parameter	Weightage
1.	Amount of transaction	W1%
2.	Number of transaction	W2%
3.	Location from which product is ordered	W3%
4.	Time pass since last transaction	W4%

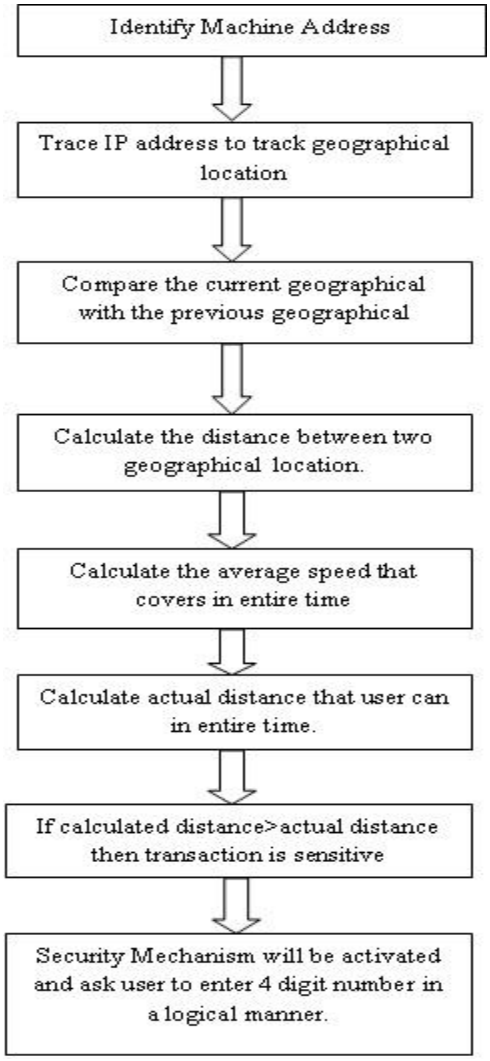

Fig 1.3 Flow Chart of Proposed Mechanism

Algorithm

Input:T_{amount} ,$C_{address}$,$P_{address}$, location, no_ of_ products (Number of the products customer has purchased online)

$\Psi = 0$

trans_amount=0;

i=1;

while(i<=number of products)

loop

trans_amount=tranas_amount+T_{amount} (i)

i=i+1;

end loop

if $C_{address} \neq P_{address}$

Then calculate distance between two location (cd)

Calculate actual distance that cover(ad)

If(cd>ad)

Then output("Transaction is suspicious")

Generate and update risk score

Cluster c=KMeanAlgorithm(trans_amount)

Count_percentage p=(Cluster c,trans_amount) If p<10

Output("High Spending Profile") ; Generate and update risk score

5. RESULTS AND DISCUSSION

Customer behavior is analyzed by spending behavior i.e. categorized into three broad categories: Low Spending Profile, Medium Spending Profile & High Spending Profile. Generally bank customer do not purchase product beyond their income. The customer of a bank generally purchases items of some share of their income. They do not spend the whole income to perform the transaction. The fraud can be identified by using K Mean Clustering algorithm. Table 6.1 contains the transaction i.e. done by the customer. The amount i.e. spent by the customer based on which the transaction can be considered as genuine or fraudulent. The most recent transaction is placed at the first position and correspondingly first transaction is placed at the last position in the table and so-on.

Table 1.2: List of all Transaction

Number of Transactions	Amount in Rs
1	2000
2	2500
3	3000
4	3600
5	1800
6	2280
7	2090
8	4850
9	3080

Figure 1.4 represents the customer spending based on the number of transactions and the amount of a customer.

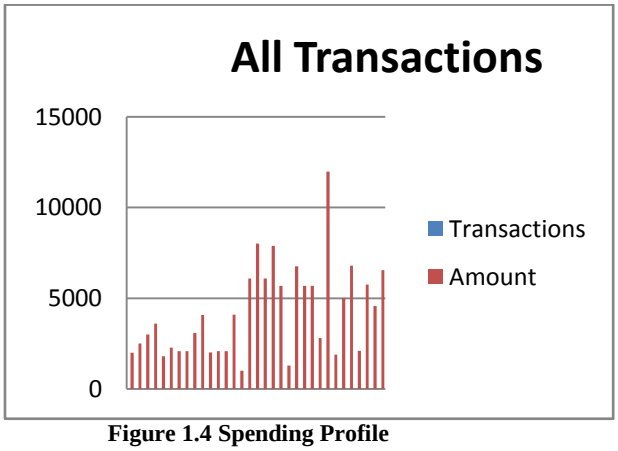

Figure 1.4 Spending Profile

The output of KMEAN clustering algorithm as shown be-low:

Table1.3: Output of KMean Clustering

Cluster Mean	C_1	C_m	C_h
Mean Value	1780	4925	43570
Percentage of total transaction	40%	50%	10%

Figure 1.5 represents the KMEAN Clustering which is based on customer spending profile. Here X attribute represents number of transactions and Y attribute represents amount.

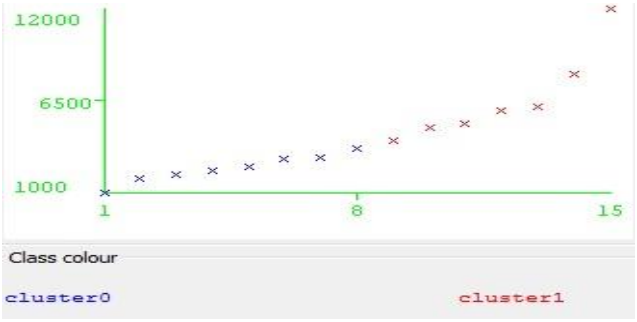

Figure 1.5 Clustering of transaction amount

Figure 1.6 represents income and number of customers of a bank:

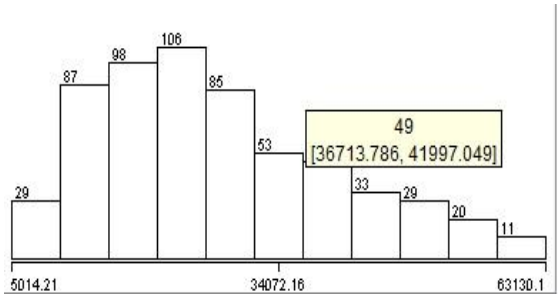

Figure1.6 Income of customers

Through IP address, the user geographical location can be identified. The machine address refers to the system address i.e. unique to each computer. IP address is a logical address i.e. unique and assigned to each computer system on the network. IP address is required to identify the user from where the user is performing the transaction. Suppose the user IP address is 206.190.36.45, so the location that is identified from IP Address is US, California as shown in Figure 1.7.

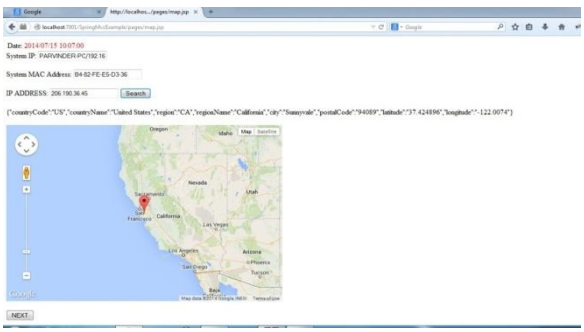

Figure 1.7Details of Current Location

Figure 1.8 represents customer current geographical location is compared with the Previous Location of customer. Previous Location refers to the location from where the customer has performed the previous transaction successfully.

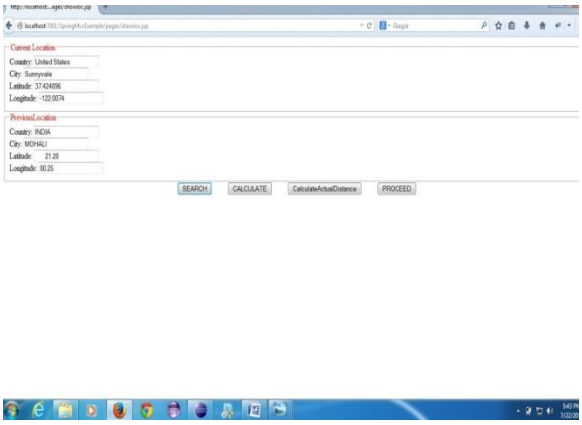

Figure1.8 Compare Current and Previous Transaction Location

The distance between two geographical locations is analyzed. The latitude and longitude is required to calculate the distance between two geographical locations as shown in Figure 1.9.

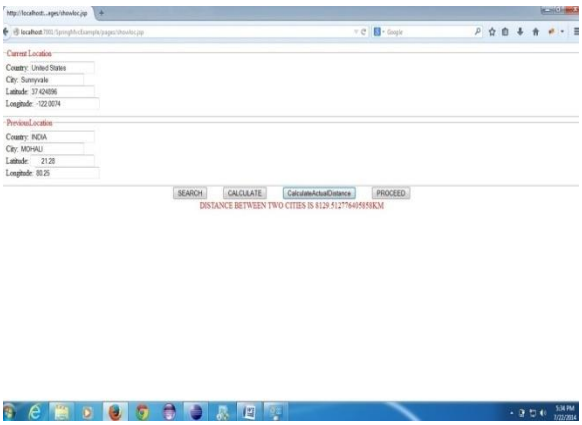

Figure1.9 Distance between two geographical locations

The customer actual distance is that distance that customer can cover in the gap between the previous time and current time as depicted in Figure 1.10. If calculated distance between two location is greater than the actual distance then transaction is considered to be suspicious.

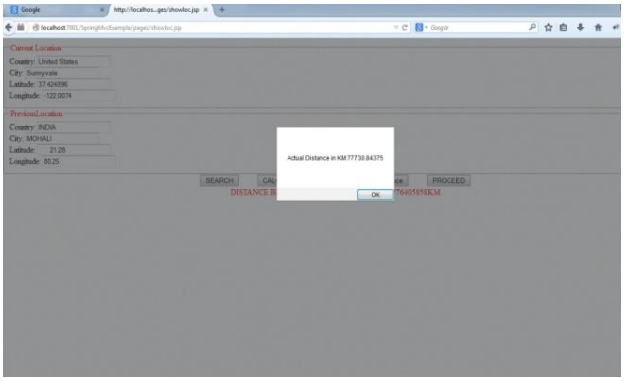

Figure 1.10 Calculation of actual distance

TRSGM (Transaction Risk Score Generation Method) is used to calculate risk score. The risk score represents the probability that the transaction is considered to be fraudulent.

The risk score is computed based on Spending Profile & Geographical Location. The risk score is computed by comparing the amount with the normal behavior of customer in terms of amount and recent location is compared with the previous location in which the distance between two geographical location is identified and actual distance is identified that user can travel during that time frame.

Table 1.4: Risk Score for different value of Amount and Locations

Customer _ID	Amount	Computed Distance	Actual Distance	Risk Score
ID12101	8000	8129.6358	576.4128	0.63825
ID12101	6000	7298.5742	9897.3685	0.17583
ID12101	8000	8173.5127	7730.8475	0.59478

If transaction is suspicious or fraudulent then security mechanism will be activated that ask the user to enter the four digit random number that appears on the screen. Only the genuine user will enter the number in a correct pattern. Figure 1.11 represents Security Mechanism. Suppose the number appears on the screen is 2695, the correct pattern of this number is 2+6+9+5=22 or other patterns are 9652 i.e. in ascending order or the multiplication of numbers 2*6+9*5=57 i.e. in logical manner. Only the genuine customer of a bank knows the pattern

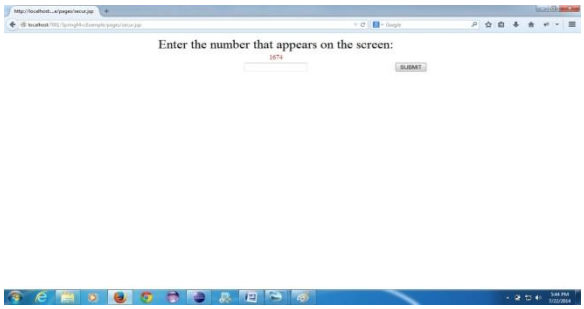

Figure 1.11 Security Mechanism

6. CONCLUSIONS

The purpose is to propose a financial cybercrime detection model that can detect online fraud and to describe various techniques to detect Financial Fraud Detection based on data mining such as neural network, regression, genetic algorithm, Bayesian Learning Approach etc. and compare these techniques. Initially, we present an approach for every online transaction; transaction risk score generation method (TRSGM) calculates the risk based on two parameters: Spending behavior and geographical Location. The risk score is calculated by Bayesian Theorem which is based on the posterior probability. The application is implemented for online transactions. The Data Mining algorithm KMEAN Clustering algorithm is implemented to identify fraudulent transactions based on the spending behavior of a customer. The geographical location of a customer is identified by identifying its IP address and compares the current geographical location with the previous location and identify whether the user can cover entire distance in that time period. If transaction is found to be fraudulent then security system will activate and four digit random numbers appears on the screen. For the authentication of user the user has to enter that pattern in a correct pattern. Unsupervised Learning approach is used to detect new types of financial fraud. Therefore this security mechanism is not stick to only past fraudulent transaction set, but also new type of fraud can be detected by this mechanism. Baye's theorem is used in the model, so the model adapts to changing behavior of genuine customer as well as fraudster. Fraudster behavior changes constantly, so security process needs to be updated regularly. Future work will be directed towards to add more parameters in addition to spending behavior & geographical location.

7. REFERENCES

[1] Bhattacharya.S , Jha.S, Tharakunnel.K and Westland.C.J, "Data mining for credit card fraud", Science Direct, Decision Support System pp 602-613, 2010.

[2] Cao.L, Zhang.H, Zhao.Y, Luo.D and Zhang.C, "Combined Mining: Discovering Informative Knowledge in complex data", IEEE Transactions Vol. 41 No.3 pp 699-712, 2011.

[3] Chuang.K.T, Lin.K.P and Chen.M.S, "Quality Aware Sampling and its application in Incremental Data Mining", IEEE Transactions on Knowledge and Data Engineering Vol. 19 No.4, pp 468-484, 2007.

[4] Cao.L, "Social Security and Social Welfare Data Mining: An Overview", IEEE Transactions on Systems, Man and Cybernetics-Part C: Applications & Reviews Vol. 42 No.6, pp 837-853, 2012.

[5] Chang.W.H and Chang.J.S, "An early fraud detection methods for online auctions", Science Direct Electronic Commerce Research and Applications, pp 346-360, 2012.

[6] Clifton.P, Kate.S.M, Lee.S.C.V and Gaylor.R , " Resilient Identity Crime Detection", IEEE Transactions Volume 24 No.3, March 2012.

[7] Dharwa.J.N and Patel.A.R, "A Data Mining With Hybrid Approach Based Transaction Risk Score Generation Method for Fraud Detection of Online Transaction", International Journal of Computer Applications Volume 16 No.1, pp 18-25, 2011.

[8] Drezewski.R, Spielak.J and Filipowski.W, "System Supporting Money Laundering Detection", Science Direct Digital Investigations, pp 8-21, 2012.

[9] Duman.E. and Ozcelik.H.M , "Detecting credit card fraud by genetic algorithm and scatter search", Science Direct, Expert System with Applications 38 , pp 13057-13063,2011.

[10] Edge.E.M and Sampaio.F.P.R, "The design of FFML: A rule-based policy modeling language for proactive fraud management in financial data streams", Science Direct, Expert System with applications 39, pp 9966-9985, 2012.

[11] Farvaresh.H and Sepehri.M.M, "A data mining framework for detecting subscription fraud in telecommunication", Science Direct, Engineering Applications of Artificial Intelligence 24, pp 182-194, 2010.

[12] GodBole.N, "Information System Security" Wiley Publishers, pp 15-100, 2012.

[13] He.Z, Xu.X, Huang.Z.J and Deng.S, "Mining class outliers: Concepts, Algorithms and Applications in

CRM", Science Direct, Expert System with applications Volume-27 pp 681-697, 2004.

[14] Han.J and Kamber.M, Data Mining: Concepts and Techniques, Second edition, Morgan Kaufmann Publishers, 2006, pp 285-464.

[15] Huang.R, Tawfik.H and Nagar.A.K, "A Novel Hybrid Artificial Immune Inspired Approach for Online Break-in Fraud Detection", International Conference on Computer Science, Science Direct, pp 2733-2742, 2012.

[16] Hajian.S and Ferrer.J.D, "A methodology for direct and indirect discrimination prevention in data mining", IEEE transactions on Knowledge and Data Engineering Vol. 25 No.7, pp 1445-1459, 2013.

[17] Jha.S, Guillen.M and Westland.C.J , "Emp-loying Transaction aggregation strategy to detect fraud", Expert System with applications 39 pp 12,650-12,657,2012.

[18] Joris.M, Celin.V and Greet.V.B,"Outlier detection in relational data: A case study in geographical information systems", Science Direct, Expert System with applications 39 pp 4718-4728, 2012.

[19] Kim.K, Choi.Y and Park.J, "Pricing fraud detection in online shopping malls using a finite mixture model" Science Direct Electronic Commerce Research and Applications, pp195-207, 2013.

[20] Kundu.A,Suvasini.P,Sural.S and Majum-dar.A.K,"BLAST-SSAHA Hybridization for Credit Card Fraud Detection" IEEE Transactions On Dependable and Secure Computing, VOL. 6, NO. 4, Dec 2009.

[21] Lee.H.D, Park.H.W and Kim.S, "Fraud and financial crime detection using malware forensics", Springer Multimedia Tools Appl., pp 479-496, 2013.

[22] Li.J, Wei.W, Yuming.O and Chen.J, "Effective detection of sophisticated online banking fraud on extremely imbalanced data", Springer World Wide Web, pp 449-475, 2012.

[23] Lin.S and Brown.D.E, "An outlier-based data association method for linking criminal incidents", Science Direct, Decision Support System 41, pp 604-615 , Oct, 2004.

[24] Panigrahi.S, Kundun.A, Sural.S and Majum-dar.A.K, "Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning" Science Direct, pp 354-363, 2009.

[25] Pozollo.D.A, Caelen.O, Borgne.Y.A.L, Wa-terschoot.S and Bontempi.G, "Learned lessons in credit card fraud detection from a practitioner perspective", Science Direct Expert System with applications 41, pp 4915-4928, 2014.

[26] Quah.J.T.S and Sriganesh.M, "Real-time credit card fraud detection using Computational Intelligence", Science Direct Expert System with applications 35 pp 1721-1732, 2007.

[27] Sahin.Y, Bulkan.S and Duman.E, "A cost-sensitive decision tree approach for fraud detection", Science Direct, Expert System with Applications 40 pp-5916-5923, 2013.

[28] Sanchez.D, Cerda.L, Serrano.J.M and Vila-.M.A, "Association Rules applied to Credit Card Fraud Detection", Science Direct Expert System with applications 36 pp 3630-3640, 2009.

[29] Srivastava.A, Kundu.A, Sural.S and Maju-mdar.A.K, "Credit Card Fraud Detection Using Hidden Markov Model", IEEE Transactions on Dependable & Secure Computing Vol. 5, March 2008.

[30] Sun.X,Hu.Y, Chen.Y, Ngai.E.W.T and W-ong.Y.H, " The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature" , Science Direct, pp 559-569 ,2010.

[31] Wu.X.D andZhu.X,"Mining With Noise Knowledge: Error-Aware Data Mining" IEEE transactions ON Cybernetics, Vol. 38 No.4, July, 2008.

[32] Wu.S and Wang.S, "Information-Theoretic Outlier Detection for Large-Scale Categorical Data", IEEE Volume 25 No.3, March 2013.

[33] Yeh.Y.R, Lee.Y.J and Wang.Y.C.F, "An-omaly Detection via Online Oversampling Principal Component Analysis", IEEE Transactions on Knowledge and Data Engineering Vol. 25 No.7, pp 1460-1470, 2013.

[34] ZhongL.J and Ghorbani.A,"Improved comp-etitive learning neural networks for network intrusion and fraud detection", Science Direct, pp 135-145 Aug, 2012.