Investigation of Credit Card Fraud Recognition Techniques based on KNN and HMM

N. Malini
M.Phil student
PG & Research department of Computer Science
Quaid-E-Millath Government College For Women(A),
Annasalai,Chennai-600 002,
Tamilnadu,India

M. Pushpa, PhD
Assistant Professor
PG & Research department of Computer Science
Quaid-E-Millath Government College For Women(A),
Annasalai, Chennai-600 002,
Tamilnadu, India

ABSTRACT
Popular payment mode accepted both offline and online is credit card that provides cashless transaction. It is easy, convenient and trendy to make payments and other transactions. Demonetization process operated by India's Prime Minister Narendra Modi seems to be taken major changes in cashless economy. Credit card fraud is also growing along with the development in technology. It can also be said that economic fraud is drastically increasing in the global communication improvement. It is being recorded every year that the loss due to these fraudulent acts is billions of dollars. These activities are carried out so elegantly so that it is similar to genuine transactions. Hence simple pattern related techniques and other less complex methods are really not going to work. Having an efficient method of fraud detection has become a need for all banks in order to minimize chaos and bring order in place. There are several techniques like Machine learning, Genetic Programming, fuzzy logic, sequence alignment, etc are used for detecting credit card fraudulent transactions. Along with these techniques, KNN algorithm and HIDDEN MARKOV MODEL is implemented to optimize the best solution for the problem. This approach is proved to minimize the false alarm rates and increase the fraud detection rate. Moreover the behaviour analysis process of the HMM method helps in minimizing the fraud rates thus retaliate further fraudulent activities more efficiently.

Keywords
Classification, Fraud Detection, K-Nearest Neighbor Algorithm, Hidden Markov Model.

1. INTRODUCTION
In day-to-day usage of credit card transactions the procurement of products and services assists online transactions or card swiping procurements. This leads to increase in online transactions using credit and debit cards evolving to a world of effortless expenditure. Frauds involved in the credit card section have caused severe damage to the users and the service provider and is said to be even worse in coming days. Fraudsters observe and adapt to the quick changes in the technology and find clever ways to involve in illegal activities. Frauds caused due to these smart hackers are hazardous and dangerous. A well-educated fraudster can create several identities and conduct credit card transactions without being caught. Many new types of frauds are emerging making the fraud detection difficult and hard.

Talking in terms of e-commerce transactions the major problem faced due to these fraudulent activities is so similar to legal ones. Hence having an efficient and complex fraud detection system must to prevent these fraudulent activities. The challenging section is to detect frauds in a huge dataset where the legal transactions are more and the fraudulent transactions are bare minimum or close to negligible.

There are very few papers on credit card fraud detection methods due to the fact that these methods cannot be tested without a dataset. Hence it's difficult to prove the robustness or even the probability of success ratio of the methods. As we know credit card information is confidential, the bank owners and service providers do not encourage in sharing these data for experiments as well. Hence KNN is used in large for finding the final results obtained in ranking the research methodology. KNN is proved to deliver accurate detection rate of fraudulent transaction and minimum number of false alert.

Although credit card payment be so convenient method for purchasing things it has created more fraudulent activities by which the card holders suffers financial loss. Determination of different ways to detect the credit card scams. Bankruptcy is one of those methods that are very tricky to identify. Phishing is the common technique that is used to get the privacy details of the card holder without their knowledge. They also guess the card passwords and other details providing false information thus accomplishing the fraudulent activities. Out of all the most common method will be the lost or stolen credit card or the leaked card details will be given the try over online by the fraudster. Thus along with KNN algorithm we require more strong algorithm like Hidden Markov model for detecting and minimizing the fraud rates by repeating the operations. Thus the implementation of all these methodologies despite of detection and protection of the credit card fraudulent can also scrutinize the functionalities that prevent further fraudulent attacks.

2. CREDIT CARD FRAUD
Credit card is pre-approved credit amount that can be used for purchasing goods and service, payment of that purchase is collected later with agreed charges. *Credit card fraud* is a situation when an individual uses someone else's credit card information to charge purchases, or removing funds for personal reasons from the account without owner's authorization. According to CEO of Rippleshot Cahn, who have spent over 15 years' experience in credit card fraud detection stated that 30-40 % of credit card fraud loss can be reduced by early detection.

Credit card fraudsters were committed in the following ways,

- Theft of actual cards,
- Misrepresentation of account or personal information,
- Illegal or unauthorized use of account for personal gain

- An act of criminal deception by use of unauthorized account or personal information.

Credit card fraud can occurs online as well as offline.

- When unauthorized users make use of credit card with the PIN is called online fraud. Using physical card for transactions eg.resturants,buying electronic goods,etc.,
- When an unauthorized user make use of credit card without the PIN is called offline fraud transaction, Eg.through shopping websites, phone transaction etc.,

3. PROBLEM STATEMENT

Classification consists of assigning a class label into a set of unclassified cases [2]. Classification schemes are generally performed on the two approaches namely: statistical and syntactic [4]. Statistical classification is performed on statistical characterizations of patterns which are generated using probabilistic system; whereas structural classification depends on the structural interrelationships of features [5]. There are several methods for pattern recognition namely Bayesian classifiers and much more complex methods like the neural networks. Supervised classifications are a process in which we gather the information of interest from the dataset and consider it as training data. The statistical classification of each training set is obtained. After the statistical classification is performed the best match from the training data is considered and the class it resembles is chosen. In unsupervised classification, large dataset is considered and divided into several parts or groups based on the natural grouping present in the dataset. Unsupervised classification does not require any labeling of data, all it needs is just that the data within a class should be as close as possible in the measurement space and its distance with other classes should be as high as possible.

Researchers are going on to generalize the rule of selecting the appropriate rule of classification for different kind of problems. The purpose of this paper is to study the implementation of K-Nearest Neighbor (KNN) and the performance results of the same when applied on credit card approval system. Moreover further deployments of credit card scams will be minimized using Hidden Markov model. It uses selection, mutation and crossover process for obtaining different combinations of fraudulent detection and enhances the prevention too.

4. CREDIT CARD FRAUD DETECTION

Credit card fraud detection is considered highly confidential and not disclosed to public. Hence, we have discussed a few available methods below.

Table 1. Various Techniques used and its Discussion

Techniques Used for Identifying Fraudulent	Advantage	Disadvantage
Logistic Regression	It produces a simple probability formula for classification. It works well with linear data for credit card fraud detection.	1. It cannot be applied on non-linear data. 2. It is not capable of handling fraud detection at the time of transaction.
Decision Tree	This method can handle non-linear data as well.	1. It involves complex algorithm and even a small change in the data can distract the structure of the tree. Choosing splitting criteria is also complex. 2. It cannot detect fraud at the time of transaction.
Artificial Neural Network	This method is capable of detecting the fraudulent activity at the time of transaction.	1. The number of parameters is to be set before the training is initiated. No rules as such are in place to set the parameters. 2. The network depends on the interconnection between the neurons. Till date there are no methods to determine the optimal topology for a given problem.
Hidden Markov Model.	This method is capable of detecting the fraudulent activity at the time of transaction. The HMM based models reduce the False Positive (FP) transactions predict as fraud though they are really genuine customer.	1. It cannot detect the fraud in initial few transactions.
Support vector machine	This method is capable of detecting the fraudulent activity at the time of transaction.	1. Sometimes it fails to detect fraud cases.

K-Nearest Neighbor Algorithm	There is no requirement of predictive model before classification.	1. The accuracy of the method depends on the measure of distance. 2. It cannot detect the fraud at the time of transaction.

5. FRAUD DETECTION AND PREVENTION METHODS

Fraud detection is a highly complex function to be performed where there is no system which can guarantee a 100% satisfaction result rate [5]. All the existing methods can likely predict fraud transactions and not assure you about the prevention of further scam attacks. Let's consider the properties of good fraud detection method:

1. It must be able to identify the frauds accurately.
2. It must quickly detect fraud cases.
3. At any case a genuine transaction should not be considered as fraud.

It enables scrutiny that constantly checks for fraud prevention and reports proceeding attacks or card misuse. The various other methods for detecting the credit card fraudulent other than finding the nearest neighbor methods are there. The data mining inculcating genetic operation for identification of the credit card misusing and handling the fraudulent transactions carried over. K-Nearest Neighbor algorithms detects using learning techniques of credit card fraudsters while the HMM operations will process based on credit card user's behaviour pattern selection and check over the upcoming hoax. Despite of detection prevention and scrutinizing the advancements in fraudulent activities is must. Markov model with finite probabilistic fraud detection methods verify each and every transactions.

6. K-NEAREST NEIGHBOR ALGORITHM

K-nearest neighbor algorithm is used largely in detection systems. It is also proved that KNN works extremely well in credit card fraud detection systems using supervised learning techniques. In this method the new instance query will be classified depending on the KNN category. This method was first utilized by Aha, Albert and Kibler in the year 1991.
The results of KNN depend on the below three factors:

1. The distance metric used to decide the nearest neighbors.
2. The distance rule that is used for the classification from K-nearest neighbor.
3. The number of neighbors considered to classify the new sample.

When we study on various credit card fraud detection methods based on supervised statistical pattern recognition, KNN achieves high performance rate without using the priori assumptions about the distributions. The KNN based credit card fraud detection techniques need two major things to be estimated namely the distance or similarity measure between two data instances. In KNN any incoming transaction will be calculated for its nearest point to new incoming transaction. So if the incoming transaction is fraudulent then the algorithm indicates it to be fraud. In such cases the value of K is considered small and odd like 1, 3 or 5 to break the ties. Larger K values help in reducing the noise in the data set. The distance between two data instances can be computed using different methods. In place of continuous attributes, Euclidean distance is used. In case of categorical attributes, an easy matching coefficient is used. In place of multivariate data, the distance is calculated for every attribute and then combined. The KNN algorithm can be optimized using better distance metrics. In this method both legitimate and fraudulent examples are to be fed in order to train the data sets. This method is fast with minimum false alerts.

7. K-NEAREST NEIGHBOR CLASSIFICATION FOR FRAUD DETECTION

In order to perform the classification using KNN let's consider a new object along with several known examples. Refer the Fig. 1 for a better understanding on the concept. The query point is in red circle along with plus and minus nearest points. Now our job is to get the outcome of the query point depending on the selected number of nearest neighbors. We are supposed to find if the query will hold plus or minus sign.

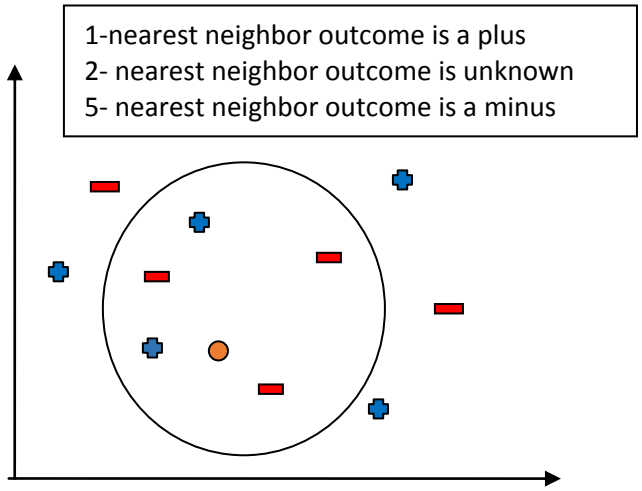

Fig. 1 Implementation of KNN

Consider the outcome of KNN for 1 nearest neighbor. So in this case the KNN will provide the query with plus sign. Now increase the number of nearest neighbor to 2. Now there is a plus and minus points close to the query point. Hence both will have equal number of votes. The next step will be considered with 5 KNN points. Now we can define the nearest neighbor point. In this case there are 2 plus and 3 minus signs; hence the outcome will have minus sign.

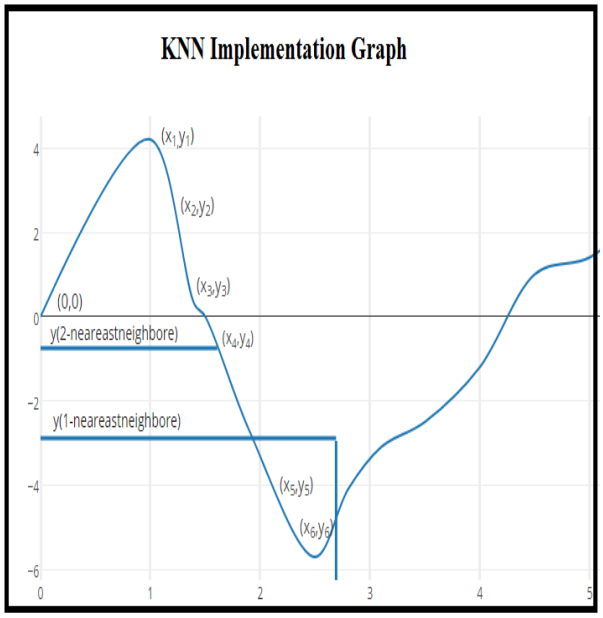

Fig. 2 KNN Implementation Graph

8. HIDDEN MARKOV MODEL IMPLEMENTATION FOR FRAUD PREVENTION

HMM is one of the data mining techniques that involve detection of financial fraud and account attack identification encountered by experts. Among the millions of everyday transaction it involves huge database and many highly efficient techniques to store and manage banking transactions. The accurate transaction techniques will be generated and efficient predictions of legitimate transactions and fraud detections, previous transaction histories will be stored and indulged. Fraudulent intrusions will be scrutinized before it emphasizes in the transaction. The tricksters uses various techniques and change their way of stealing money frequently as they are very careful in abstaining during fraudulent activity. Development of detection technologies easily recognize the methodologies used by scammers and apply barricades for such moves.

HMM use prognostic methodologies for identifying the fraud transactions. It uses training data and examples from prevailing fraudulent activities on credit card transactions. The model uses the hypothetical patterns for arbitrary change in assuming the process relies on the predictive and probabilistic detection of attacks. It maintains the arbitrary values of changes along various time periods and determines the precise hidden model. The sequence of algorithms in hidden Markov designs the probable sequences with observations uses transition functions. It uses transition process that chose among approximate controlling applications over partial state in random and hierarchical models and thus enhances the process of predicting the various fraudulent activities involved in credit card transactions. These operations act as firewall for credit card transactions. The fig 3. Exhibits the flow of the HMM module for credit card fraud detection.

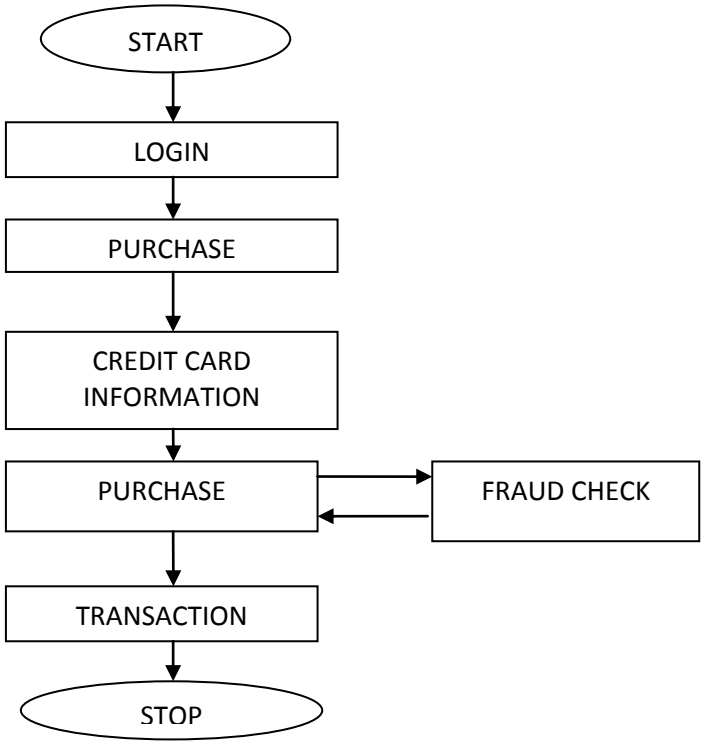

Fig. 3 HMM flow chart for credit card fraud detection

HMM trains the models of previous attacks and supervises the fraudulent in card and amount transactions. It mainly trained in checking the address variation, exceptions identification, change in system IP address thus to deliver the products purchased will be sent to the legitimate user purchased with genuine account transactions and payment done through credit card. When the fraudster deflect the concentration of card holder and tries to deceive his profile it will be recognized as exceptions. Thus it will process the crossover detection process and verifies their card details with their previous genuine transactions.

9. LITERATURE REVIEW

A.J. Graaff et al [1] compare it with the natural immune system and create an artificial immune system as a classifier with positive and negative patterns. AIS method is used to detect network intrusion virally infected files etc. The results obtained using AIS model in Iris plant dataset is also presented in this paper. The theoretical approach of AIS delivers best results in detecting illegitimate patterns. In future the AIS model needs to be trained on legitimate calls and then tested using calling patterns using both legitimate and illegitimate calling patterns.

Abhinav Srivastava et al [2] the author uses the ranges of transaction amount as an attribute in the HMM. The author has suggested method for finding the spending profile of cardholders. It is also discussed how the HMM can identify the fraudulent transactions. The simulation results show the advantages of using HMM and learning the profile of the cardholder plays an important role in analyzing fraudulent cases. The result also shows that 80% of the results are accurate and the system is scalable for large data set as well.

Divya.Iyer et al [3] the author uses Hidden Markov Model (HMM) to detect credit card transaction frauds. The training set is tuned with the normal behavior of the card holder. So if

a credit card transaction is rejected by the trained HMM then that transaction is said to be fraudulent. Care is to be taken that valid and genuine transactions are not considered as fraud. The author also compares various methods with the proposed methods to prove that HMM are much preferred than the other methods.

K.RamaKalyani et al [4] creates a test data and through which the fraudulent activities are detected. This algorithm is also called as an optimization technique based on genetic and natural selection in high computational problems. The author proposes a method to detect credit card fraud and the results are validated using principles of this algorithm. The purpose of detecting fraud cases is to declare it to the client and the service provider.

Renu et al [5] proposed a fraud detection method which involves monitoring the activities of populations to observe and predict undesirable behavior. Undesirable behavior is a set of several habits like intrusion, fraud, delinquency and defaulting. This research speaks on several credit card fraud detection and telecommunication fraud and different techniques which help in resolving the discussed problems.

10. CONCLUSION

Credit card scam has become much more extensive. To progress safety measures of the monetary transaction systems in a habitual and effectual way, structure a precise and well-organized credit card scam detection system is one of the essential functions for money transactions. By performing over sampling and extracting the principal direction of the data we can use our KNN method to determine the anomaly of the target instance. Hence the KNN method can suit for detecting fraud where there is limitation for memory and computation. In fact implementation of Markov models assists prediction and prevention of the card transaction fraudulent. Although when compared with power methods and other known anomaly detection methods, experimental results prove that the KNN method is accurate and efficient.

11. REFERENCES

[1] A.J. Graaff A.P. Engelbrecht Agraaff "The Artificial Immune System For Fraud Detection In The Telecommunications Environment" 20 November 2014

[2] Abhinav Srivastava, Amlan Kundu, Shamik Sural, And Arun K. Majumdar" Credit Card Fraud Detection Using Hidden Markov Model" VOL. 5, NO. 1, JANUARY-MARCH 2008

[3] Divya.Iyer,Arti Mohanpurkar, Sneha Janardhan, Dhanashree Rathod,Amruta Sardeshmukh" Credit Card Fraud Detection Using Hidden Markov Model " 978-1-4673-0126-8/11/$26.00_C 2011 IEEE

[4] K.Ramakalyani, D.Umadevi" Fraud Detection Of Credit Card Payment System By Genetic Algorithm" Volume 3, Issue 7, July-2012.

[5] Renu, Suman" Analysis On Credit Card Fraud Detection Methods" Volume 8 Number 1– Feb 2014

[6] Ekrem Duman, M. Hamdi Ozcelik "Detecting Credit Card Fraud By Genetic Algorithm And Scatter Search". Elsevier, Expert Systems With Applications, (2011). 38; (13057–13063).

[7] S. Benson Edwin Raj, A. Annie Portia, "Analysis On Credit Card Fraud Detection Methods", International Conference On Computer, Communication And Electrical Technology – ICCCET2011, 18th & 19th March, 2011

[8] Y. Sahin And E. Duman, "Detecting Credit Card Fraud By Decision Trees And Support Vector Machines", International Multiconference Of Engineers And Computer Scientists March, 2011.

[9] S. Benson Edwin Raj, A. Annie Portia "Analysis On Credit Card Fraud Detection Methods". "IEEE-International Conference On Computer, Communication And Electrical Technology"; (2011). (152-156).

[10] Eswari.M,Navaneetha,Krishnan.M." Survey On Various Types Of Credit Card Fraud And Security Measures", International Journal Of Advanced Research In Computer Science And Software Engineering , Volume 4, Issue 1, January 2014, Pg.1235 – 1238.

[11] Anika Nahar,Sharmistha Roy,"A Survey On Different Approaches Used For Credit Card Fraud Detection", International Journal Of Applied Information Systems (IJAIS) – Foundation Of Computer Science FCS, New York, USA Volume 10 – No.4, January 2016,Pg.29 – 34

[12] Priya Ravindra Shimpi, Prof. Vijayalaxmi Kadroli ,"Survey On Credit Card Fraud Detection Techniques", International Journal Of Engineering And Computer Scienc,Volume 4 Issue 11 Nov 2015, Page No. 15010-15015

[13] Anshul Singh, Devesh Narayan "A Survey On Hidden Markov Model For Credit Card Fraud Detection". International Journal Of Engineering And Advanced Technology (IJEAT), (2012). Volume-1, Issue-3; (49-52).

[14] V. Dheepa, Dr. R .Dhanapal "Analysis Of Credit Card Fraud Detection Methods". International Journal Of Recent Trends In Engineering, (2009). Vol 2, No. 3; (126-128)