A review of Fraud Detection Techniques: Credit Card

Khyati Chaudhary,
Dept. of Computer Science
GCET, Greater Noida,
INDIA

Jyoti Yadav
Dept. of Computer Science
GCET, Greater Noida,
INDIA

Bhawna Mallick
Dept. of Computer Science
GCET, Greater Noida,
INDIA

ABSTRACT
In present scenario when the term fraud comes into a discussion, credit card fraud clicks to mind so far. With the great increase in credit card transactions, credit card fraud has increasing excessively in recent years. Fraud detection includes monitoring of the spending behavior of users/ customers in order to determination, detection, or avoidance of undesirable behavior. As credit card becomes the most prevailing mode of payment for both online as well as regular purchase, fraud relate with it are also accelerating. Fraud detection is concerned with not only capturing the fraudulent events, but also capturing of such activities as quickly as possible. The use of credit cards is common in modern day society. Fraud is a millions dollar business and it is rising every year. Fraud presents significant cost to our economy worldwide. Modern techniques based on Data mining, Machine learning, Sequence Alignment, Fuzzy Logic, Genetic Programming, Artificial Intelligence etc., has been introduced for detecting credit card fraudulent transactions. This paper shows how data mining techniques can be combined successfully to obtain a high fraud coverage combined with a low or high false alarm rate.

General Terms

Data Mining, Neural Networks, LR, Clustering techniques.

Keywords
Fraud detection; Electronic Commerce; Credit card fraud, Spending pattern; Credit card, fraud detection techniques, On-line banking

1. INTRODUCTION
Fraud refers to obtaining goods/services and money by illegal way. Fraud deals with events which involve criminal motives that, mostly, are difficult to identify. Credit cards are one of the most popular objective of fraud but not the only one. Credit card fraud, a wide-ranging term for theft and fraud committed or any similar payment mechanism as a fraudulent resource of funds in a transaction. Credit card fraud has been expanding issue in the credit card industry. Detecting credit card fraud is a difficult task when using normal process, so the development of the credit card fraud detection models has become of importance whether in the academic or business organizations currently. Furthermore, role of fraud has been changed suddenly during the last few decades along with advancement of technologies. Credit Card Fraud is one of the biggest threats to business and commercial establishments today. Simply, Credit Card Fraud is defined as, "when an individual uses another individuals' 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." A number of systems/models, process and preventive measures will help to stop credit card fraud and reduce financial risks. Banks and credit card companies have gathered large amounts of credit card account transactions. The Credit Card is a plastic card issued to number of users as one of the mode of payment. It allows cardholders to purchasing goods and services based on the cardholder's promise. In China, credit card users are growing rapidly, but only a very few credit card holders use credit cards for paying for day-to-day purchase comparatively with confidence and a sense of security. Reason is, credit card holder has no enough confidence to trust upon the payment system. Secure credit services of banks and development of E-business a reliable fraud detection system is essential to support safe credit card usage, Fraud detection based on analyzing existing purchase data of cardholder (current spending behavior) is a promising way for reducing the rate of credit card frauds. Fraud detection systems come into scenario when the fraudsters exceed the fraud prevention systems and start fraudulent transactions. Along with the developments in the Information Technology and improvements in the communication channels, fraud is spreading all over the world with results of large amount of fraudulent loss. Anderson (2007) has identified and described the different types of fraud. Credit card frauds can be proceed in many different ways such as simple theft, counterfeit cards, Never Received Issue (NRI), application fraud and online/Electronic fraud (where the card holder is not present). Credit card fraud detection is dreadfully difficult, but also common problem for solution. As there is limited amount of data with the transactions being confided, for example, transaction amount, merchant category code (MCC), acquirer number and date and time, address of the merchant. Various techniques in Knowledge Discovery, such as decision tree, neural network and case based reasoning have broadly been used for forming several fraud detection systems/ models. These techniques usually need adequate number of normal transactions and fraud transactions for learning fraud patterns. However, the ratio of fraudulent transactions to its normal transactions is low extremely, for an individual bank.

2. TYPES OF FRAUD
Various types of frauds in this paper include credit card frauds, telecommunication frauds, and computer intrusions, Bankruptcy fraud, Theft fraud/counterfeit fraud, Application fraud, Behavioral fraud,

2.1. Credit Card Fraud: Credit card fraud has been divided into two types: Offline fraud and On-line fraud.

2.1.1. Offline fraud is committed by using a stolen physical card at call center or any other place.

2.1.2. On-line fraud is committed via internet, phone, shopping, web, or in absence of card holder.

2.2. Telecommunication Fraud: The use of telecommunication services to commit other forms of fraud. Consumers, businesses and communication service provider are the victims.

Hansen, McDonald, Messier, and Bell (1996) used a powerful generalized response model to predict management fraud. Model includes the "probit and logit" techniques.

At first, this paper introduces Credit Card, its various types, then related work and possible techniques, models for detecting fraudulent/legal transactions.

2.3. Computer Intrusion: Intrusion Is Defined As The Act Of Entering Without Warrant Or Invitation; That Means "Potential Possibility Of Unauthorized Attempt To Access Information, Manipulate Information Purposefully. Intruders May Be From Any Environment, An Outsider (Or Hacker) And An Insider Who Knows The Layout Of The System.

Computer intrusion can be classified into three categories: misuse intrusions, network intrusions and host intrusions**. Misuse intrusions** analyze the information gather and compare it to large databases of attack signatures. **Network intrusions,** individual packets flowing through a network are analyzed.

Passive intrusions, detects a potential security breach, logs the information and signals an alert.

2.4. Bankruptcy Fraud: This column focuses on bankruptcy fraud. Bankruptcy fraud means using a credit card while being absent. Bankruptcy fraud is one of the most complicated types of fraud to predict. Some methods or techniques may help in fraud prevention. The bank will send its users/customers an order to pay. However, the users will be recognized as being in a state of personal bankruptcy and not able to recover their unwanted loans. The bank will have to cover the losses itself. One of the possible ways to prevent bankruptcy fraud is by doing a pre-check with credit bureau in order to be informed about the past banking history of its customers. Foster & Stine (2004) presented a model to forecast personal bankruptcy among users of credit card**.**

2.5. Theft Fraud/ Counterfeit Fraud: In this section, we focus on theft and counterfeit fraud, which are related to one other. Theft fraud refers using a card that is not yours. As soon as the owner give some feedback and contact the bank, the bank will take measures to check the thief as early as possible. Likewise, counterfeit fraud occurs when the credit card is used remotely; where only the credit card details are needed. Firstly, use of your copied card number and codes via various web-sites, where no signature or physical cards are required. **Pago Report** issues (2005), although in European E-commerce seems to be quite low, at only 0.83 percent along with the average charge-back ratio, significant concerns are notified in detailed analysis. For the listed credit card, the customers are contacted and if they do not react within certain time limit than the card is blocked.

2.6. Application Fraud: When someone applies for a credit card with false information that is termed as application fraud. For detecting application fraud, two different situations have to be classified. When applications come from a same user with the same details, that is called duplicates, and when applications come from different individuals with similar details, that is termed as identity fraudsters. **Phua et al. (2006)** describes application fraud as "demonstration of identity crime, occurs when application form(s) contain possible, and synthetic (identity fraud), or real but also stolen identity information (identity theft)." In most of the banks, eligibility for a credit card, applicants need to complete an application form. Application form is mandatory except for social fields. The bank would also ask for certain details as contact details, such as e-mail address, mobile phone number and land-line number. Confidential information will be the password.

2.7. Behavioral Fraud: Behavioral fraud occurs when sales are made on a 'cardholder present' basis and details of legitimate cards have been obtained fraudulent basis.

3. CREDIT CARD FRAUD DETECTION

In this section, we present some conceptual views of credit card, problems and some real world problems.

3.1. Terms

3.1.1. Credit Card: Credit card is a medium of selling goods or services without having cash in hand. A credit card is a simple way of offering credit to a consumer automatically. Today, almost every credit card carries an identifying number that helps in shopping transactions rapidity.

3.1.2. Fraud: Fraud is an intentional deception made for personal gain or to damage another user/individual; is fraudulent. Legal definition varies by legal jurisdiction for fraud. Fraud is a civil law violation and also a crime. Defrauding people or entities of money is a common purpose of fraud**.**

3.1.3. Credit Card Fraud

Wondering United States, with its high number Credit Card transactions has minimum fraud rate. Ukraine tops the list with staggering 19% fraud rate closely followed by Indonesia at 18.3% fraud rate amongst the high risk countries facing Credit Card Fraud threat, some other countries are Yugoslavia (17.8%), Malaysia (5.9%). and Turkey (9%).Authorized users are permitted for credit card transactions by using the parameters such as credit card number, signatures, card holder's address, expiry date etc. Illegal use of card or card information without the knowledge of the owner itself and thus is an act of criminal deception refers to Credit card fraud. Credit card fraud detection is quite confidential and is not much disclosed publicly. Commonly used fraud detection methods are, rule-induction techniques, decision trees, Support Vector Machines (SVM), LR, ANNs and meta-heuristics such as, k-means clustering, genetic algorithms and nearest neighbor algorithms. Fraud is some kind of human behavior that relates to stealing, misunderstanding, misrepresenting, cheating, cunning false suggestions etc. Sometimes companies deal with millions of external parties, it is cost-prohibitive to check the majority of the external parties' activities and identity manually. Certainly, for investigating each suspicious transaction, they incur a direct overhead cost for each of them. If in case, transaction amount is smaller than overhead cost, investigating is not worthwhile even if it is suspicious.

4. VARIOUS TECHNIQUES OF CREDIT CARD FRAUD

4.1. Neural networks:

Neural network is defined as a set of interconnected nodes designed to represent functioning of the human brain. Each node has a weighted connection to several other linked nodes in adjacent layers. Single node take input received from linked nodes and use the weights of the connected nodes together with easy function for computation of output values. Neural networks can be created for supervised and/or unsupervised learning. The user specifies the number of hidden layers along with the number of nodes within a specific hidden layer. The output layer of the neural network may contain one or several nodes depending upon the application. Recently, neural network researchers have several associated methods from statistics and numerical analysis into their networks. From the

given cases, nonlinear mapping relations from the input space to output space. Neural networks can learn and summarizes the internal assumptions of data even without knowledge of the potential data principles in advance. According to **Rumelhart,** (1986), Neural networks topologies, or architectures, formed by organizing nodes into layers and attach layers of neurons with modified weighted interconnections And it can match its own behavior to the new environment along with the results of formation of evolution capability from present environment to the new possible situation. Statistical methods are sometime unusual in the practice research even though the common advantages of the neural networks in application of credit card fraud detection. On the other side, there are still many disadvantages for the neural networks, such as

(1) Difficulty to confirm the structure,
(2) Excessive training,
(3) Efficiency of training and so on.

For Example, We use multi-layer neural network model and back propagation (BP) algorithm runs on the network. Back propagation (BP) learrns by iteratively processing a trained dat set of tuples A= {a1, a2……, an}, and comparing the network's prediction for each tuple with the actual known target value. Each training tuple has weights that are changed so as to minimize the mean square error between the network's prediction value and the actual target value. Such adaptations are made in the backwards direction, that is, from the output layer,

B= {b1, b2 . …… bn}, through each hidden layer down to the first hidden layer. In this study, a sigmoid function is used for the available nodes in the hidden layers and the output layer. The learning rate "l" is set to countdown of the number of entries in training data involved in the change of weight. **Dorronsoro et al. (1997)** developed a accessible online fraud detection system technically which has some base on a neural classifier. Somewhat, the main limitation is that data need to be clustered by type of account. Neural networks are also recommended frequently for fraud detection. Similarly, some concepts are: (**Aleskerov** et al., 1997) CARD WATCH; (**Ghosh & Reilly, 1994**) FDS; (**Kim & Kim**, 2002). improving detection efficiency "misdetections"; (**Maes** et al., 2002) Back-propagation of error signals; (**Quah & Sriganesh, 2008**; **Zaslavsky & Strizkak**, 2006) SOM; (**Brause** et al**.,** 1999a**; Brause** et al., 1999b) Data mining tools, such as **'Clementine'** allows the use of neural network technologies, which have been used in credit card fraud. On the other hand, **(Ezawa & Norton,** 1996**) Bayesian networks** are also one technique to detect fraud, and have been applied to detect fraud in the telecommunications industry and also in the credit card industry (**Maes** et al., 2002). This technique results are assured in nature. However, one main disadvantage of such a technique is the time constraint. Likewise, by using a rule-based expert system (**Leonard**, 1995), expert systems have also been used in credit card fraud as well. Still, it does not matter that the statistical techniques chosen would fulfill some conditions as the fraud detection system will call for. The system will have to handle skewed distributions of the data for the number of fraudulent transactions which is much less than the total number of transactions. Otherwise, for less skews distribution, the data needs to split into training samples, (**Chan** et al., 1997). (**Fawcett & Provost**, 1997) the system has to be capable of handling noise in the data and be accurate with actual performing classifiers. The suggested solution is to clean the data. As fraudsters reinvent new techniques constantly, for this the system needs to be adaptive and evaluated regularly. The system should also be able to handle fraudulent transactions may be similar to normal transactions. For avoidance of spending time on uneconomic cases, a cost profit analysis is too a must in fraud detection**.** A proposal would be to rely on credit bureau score in order to control fraud and for avoiding of expected losses to help new issuing banks. Generic scoring systems are usually based on a sample from the past behavior of several lenders. Generic systems are sold to those creditors who believe they will find them useful. (**Thomas** et al., 2004) the systems are often available on purchase as well as a basis transaction. The most chief generic models are those that available through the major credit bureau, and affect most credit decisions made by major creditors. As fraud & default are strongly correlated even though these scorecards are basically used to predict defaulting customers, one could use them to detect fraud. A credit bureau score may be included in the credit report of them usual or as a stand-alone product. Credit Bureau has its own models and the competition is increasing high. Generic models consist of only information from a single credit bureau and used in model development. Generic models have data of sample sizes typically ranges from the hundreds of thousands to over a million files. In general, a credit bureau scorecard is developed into a model by using the characteristic data available for that applicant and forecasting the payment behavior of an applicant. Classically, credit bureau scores are centered on external data which have been calibrated in such a way that, with regard to age and gender, for example, they reflect the population.

4.2. Decision Tree:

After introducing the concept of learning system, decision tree method has been developed, C4.5 (**Quinlan,** 1993) and ID3 method (**Quinlan,** 1986) that can deals with continuous data. The decision tree is a table of tree shape with connecting lines to available nodes. Each node is either a branch node followed with more nodes or only one leaf node assigned by classification. With this strategic approach of separating and resolving, decision tree usually detach the complex problem into many simple ones and resolves the sub-problems through repeatedly using, data mining method to discover training various kinds of classifying knowledge by constructing decision tree. The basis of decision tree model is how to construct a decision tree with high precision and small scale.

There are many advantages of Decision tree method. At first the high flexibility that it is a non-parameter method without any notion for the data distribution. Good haleness on the other side. Nearby, it is explainable, which is also the reason of its varied utilization. After that, the conception of a similarity tree using decision tree logic has been developed. (**Kokkinak**i, 1997) a similarity tree refers to edges are labeled with values of attributes and pertaining nodes that are labeled with attribute names, that satisfy some condition and 'leaves' , an intensity factor which implies as the ratio of the number of transactions that satisfy these condition(s) over the total number of legitimate transaction in the particular behavior The advantage of the similarity tree method is that it is suggested that it is easy to implement, to display and to understand. Still, system has some disadvantages that, the requirements to check each transaction one by one. Yet, [**Fan** et al. (2001), similarity trees have given proven results that worked on decision trees and especially on another type of fraud, inductive decision tree in order to establish an intrusion detection system.

4.3. Logistic Regression:

(Altman, Marco 1994; **Flitman,** 1997) Data mining tasks has more and more statistical model that involves discriminant analysis, regression analysis, multiple- logistic regression, etc. Logistic regression (LR) is useful for situations in which

we want to be able to predict the presence or absence of a characteristic or outcome based on values of a set of predictor variables. It is similar to a linear regression model but is suited to models where the dependent variable is dichotomous. Logistic regression coefficients can be used to estimate odds ratios for each of the independent variables in the model and it is applicable to a broader range of research situations than feature analysis. (**Ohlson,** 1980; **Martin,** 1997) estimating the odds of a firm's failure with probability

4.4. Genetic algorithms:

For predictive purposes, algorithms are often acclaimed as a means of detecting fraud. In order to establish logic rules which is capable of classifying credit card transactions into suspicious and non-suspicious classes, one algorithm that has been suggested by **Bentley** et al. (2000) that is based on genetic programming. However, this method follows the scoring process. In the experiment as described in their study, the database was made of 4,000 transactions along with 62 fields. As for the similarity, tree, training and testing samples were employed. For this purpose, different types of rules were tested with the different fields. The best rule among these is with the highest predictability. Their method has proven results for real home insurance data and could be one best method against credit card fraud. **Chan** et al. (1999) has developed an algorithm for prediction of suspect behavior. Origin of their research is that cost model evaluated and rated b whereas other studies use evaluation based on their prediction rate/the True Positive Rate (TPR) and the error rate/the False NegativE Rate (FNR). **Wheeler & Aitken** (2000) formed the idea of combining different algorithms to maximize the power of prediction. Article by, **Wheeler & Aitken**, presents different algorithms: diagnostic algorithms, diagnostic resolution strategies, , best match algorithms, density selection algorithms, probabilistic curve algorithms and negative selection algorithms. As a conclusion from their investigation that probabilistic algorithms and neighborhood-based algorithms have been taken to be appropriate techniques for classification, and further it may be improved using additional diagnostic algorithms for decision-making in borderlines cases as well as for calculation of confidence measures and relative risk measures. The inspiration for GANN, by combining genetic algorithms with neural networks comes from nature. In GANN, the genetic algorithm is used to find some parameters. Main query is how exactly Genetic Algorithm and Neural Network can be combined. Neural Network has been encoded in the genome of the Genetic Algorithm. In GANN the procedure involves generation of number of random individuals. Designing of neural network is according to the genome information which helps in evaluation of parameter strings. Performance can be easily determined after back-propagation training. To find an optimal network, few **GANN** strategies rely only on the GA. In this case no training set takes place which are further evaluated and ranked according to parameter performance. **Genetic Algorithm (GA)** is a search heuristic that copies the process of natural evolution and is used to generate useful and appropriate solutions for optimization problems and search problems. Genetic algorithms (GA) belongs to the larger class of Evolutionary Algorithms (EA), generate solutions to optimization problems using some techniques such as mutation, inheritance, selection, and crossover.

4.5. Clustering techniques:

Two clustering techniques have been suggested for behavioral fraud by **Bolton & Hand** (2002). Peer group analysis is a system that allows identifying accounts which are behaving differently from others at one moment in time whereas previously, they were behaving the same. These certain accounts are then flagged as suspicious. Then fraud analysts have been used to uncover those cases. Hypothesis behind peer group analysis is that if accounts that were behaving the same for a certain period of time and then one account, still behaving significantly differently, then this account has to be notified. Another approach, Breakpoint analysis uses a different hypothesis which states that if a change of card usage is notified on an individual basis, the account must be investigated. Or we can say that based on the transactions of a single card, the break-point analysis can identify suspicious behavior/pattern. Signals of suspicious behavior are a sudden transaction for a high amount, and a high frequency of usage without any knowledge to cardholder(s).

4.6. Outlier Detection:

Outliers are a basic form of non-standard attention that can be used for fraud detection. An observation that deviates much from other observations that arises suspicion that it was generated by a different mechanism is known as outlier. Unsupervised learning approach **is** employed by this model. Generally, the result of unsupervised learning **is** a new explanation or representation of the observed data, which will then lead to improved future decisions. Unsupervised methods do not need the prior knowledge of fraudulent and non-fraudulent transactions in historical database, but instead unsupervised learning detect changes in behavior and/or unusual transactions. These methods involve modeling of baseline distribution that represents normal behavior and then detects observations that show deviation from this norm. On other side, supervised methods, models are trained to discriminate between fraudulent and non-fraudulent transaction so that new observations can be assigned to classes. In supervised methods, they require accurate identification of fraud. In historical databases fraudulent transactions, can only be used to detect frauds of a type that have previously occurred. Advantages of using unsupervised methods over supervised methods that previously undiscovered types of fraud may be detected. **Supervised methods** are only trained to differentiate between legitimate transactions and previously known fraud. Some unsupervised credit card fraud detection techniques have been proposed by **Bolton and Hand** with the help of using behavioral outlier detection techniques. Spending behavior abnormally and frequency of transactions will be identified as outliers, which are likely fraud cases.

5. SOME NUMBERS: COST OF FRAUD

Several research studies on credit card phenomenon report shocking numbers, as fraud is a millions of dollar business. In contrast with internal fraud, two specific surveys, one conducted in the United States by the **ACFE,** (2006), and one PricewaterhouseCoopers (**PwC** 2007), worldwide yield some following information about Corporate fraud: In a survey, Forty-three percent of companies worldwide (**PwC**-survey) has fallen sufferer to economic crime in the respective years 2006 and 2007. PwC survey, analyzes the average financial damage to companies was US$ 2.42 million per company over the past two years. No industry seems to be bigger and safe companies seem to be more accessible to fraud than smaller ones. ACFE study participants estimated a loss of 5 percent of a company's annual revenues to fraud. United States Gross Domestic Product (UGDP) of US$ 13,246.6 billion applied in 2006, would translate approximately US$ 662 billion in fraud losses for the United States only. Numbers mentioned above are all concerned forms of Internal Fraud. However, large costs from external fraud have been

involved. Four important domains afflicted by fraud are regularly reported such as health care, telecommunications, automobile insurance, and credit cards. We found the following numbers on these domains: Generally about US$ 55 billion has been estimated in **telecommunications fraud** (**Abidogum** 2005) Second domain is the **automobile insurance** fraud problem, **Brockett** *et al*. (1998) reference an estimation of the National Insurance Crime Bureau (NICB) that the annual cost in the United States is about US$ 20 billion. We read at the website of the NICB that: "In Insurance industry studies, 10 percent or more of casualty/property insurance claims are fraudulent." (NICB2008). However **health care insurance claims fraud**, the United States National Health Care Anti- Fraud Association (NHCAA) estimates constantly that of the nation's annual health care, at least 3 percent is lost to outright fraud, is $68 billion. Another estimate by government and law enforcement agencies announces the loss as high as 10 percent of their annual expenditure. (NHCAA 2008) Next Fourth domain concerning the credit card fraud, **Bolton and Hand** (2002) cited estimates of US$ 10 billion losses worldwide for Visa/Master card only.

6. RELATED WORK ON CREDIT CARD FRAUD DETECTION

Researchers developed many credit card fraud detection techniques based on data mining approach. **Ghosh and Rilly** have proposed credit card fraud detection with a three-layer approach, feed-forward neural network (FFNN), which requires long training time. **CARDWATCH:** presented by **Aleskerov** et al. proposed that a neural network based database mining system which was a prototype for database mining system developed for credit card fraud detection application and is concerned that it requires one network per customer. **Amalan Kundu** et al suggested a model BLAST-SSAHA Hybridization technique of credit card fraud by online detection. BLAST-SSAHA approach improves the fraud detection by combining both peculiarities as well as misuse detection techniques. **Phua** et al have done a major survey of existing data mining based Fraud Detection System (FDSs). **Chiu** et al have introduced web-services based collaborative scheme for fraud detection in the Banks. The proposed scenario supports the sharing of knowledge about fraud pattern with the participant banks in a heterogeneous and distributed environment. **Abhinav srivastava** et al have proposed Hidden Markov model (HMM) for credit card fraud detection which shows 80% accuracy over a large variation in the input data. **Syeda** et al have improved the speed by using parallel granular neural network of data mining and knowledge discovery process (KDP) for credit card fraud detection and achieve reasonable speed up to 10 processors only & more number of processors introduces load imbalance problem. Markov Model and time series are not scalable to large size data sets due to their time complexity. **Fan** et al recommend the application of distributed data mining in credit card fraud detection and improve the efficiency of highly distributed databases and detection system as this approach uses Boosting algorithm name Ada Cost. Ada Cost uses large number of classifiers and requires more computational resources during detection. **Brause** et al combine advanced data mining techniques and neural network algorithms**. Stolfo** et al intimate a credit card fraud detection system using various meta-learning techniques to learn models of fraudulent credit card transactions**.** To achieve high fraud detection along with low false alarm **Elkan** et al suggest Naïve Bayesian approach for credit card fraud detection. Further, **Elkan and Witten** presents that NB algorithm is very effective in many real world data sets as well as extremely capable in linear attributes. Bayesian networks were faster and accurate to train but are slower when applied to new instances/occurrence In a online system **Vatsa** et al. have currently proposed a game-theoretic approach to credit card fraud detection. . **Wen-Fang** et al have suggested a research on credit card fraud detection model which is based on outlier detection mining on distance sum, which shows that it can detect credit card fraud better than anomaly detection based on clustering. **Jianyun** et al have shows framework for detecting fraudulent transactions. In his paper work describes an FP tree based method to effectively create user profile for detection of fraud. But on the other hand, this technique doesn't recognize unusual patterns i.e. short term behavioral changes of genuine card holders. Today, some of the existing credit card fraud detection techniques which use labeled data to train the classifiers are unable to detect new kinds of frauds. Supervised learning has some disadvantage, that they require human involvement to optimize parameters. On another hand, decision tree do not require any parameter setting from the user and can build faster compared to other techniques.

7. CONCLUSION

Currently, building a precise, accessible and simple handling credit card risk monitoring system is one of the key tasks for the merchant banks, organization to improve merchants' risk management level in an automatic, scientific and adequate way. In this paper, we demonstrate various techniques used in credit card fraud detection and their advantages with data mining techniques including neural networks, and confidence value calculation. Further more studies are encouraged to improve the fraud detection basis to set more suitable weight and cost factor with both good tested accuracy and detection accuracy. More efficient credit card fraud detection system/ model an important requirement for any card issuing bank. Credit card fraud detection has drawn number of techniques, system, and models that have been proposed to counter credit fraud and lot of interest from the research community. The neural network based CARDWATCH shows much great accuracy in fraud detection and processing speed is also high but it is limited to one-network per customer. The Fuzzy Darwinian fraud detection systems (FDFDS) improve accuracy of the system. Since The Fraud detection rate (FDR) of Fuzzy Darwinian fraud detection systems in terms of true positive (TPR) is 100% and presents good results in detecting fraudulent transactions. The Fraud detection rate (FDR) of Hidden Markov model (HMM) is very low as compared to other existing methods. Processing speed of BLAST-SSAHA is fast enough to enable on-line detection of credit card fraud. All the techniques of credit card fraud detection discussed in this survey paper have its own strengths as well as weaknesses and advantages along with disadvantages. Survey of such kind will enable us to build a hybrid approach for fraudulent credit card transactions identification. In daily life, every field of the daily life, credit card fraud has become much more important and popular. Building an accurate and efficient credit card fraud detection system to improve security of the financial transaction is one of the key tasks for the financial institutions. In this paper, we determine 13 classification methods were used to build fraud detecting models/ system. This work demonstrates the advantages of applying the data mining techniques including ANN and LR, BN techniques to the credit card fraud detection problem for the purpose of reducing the bank's or financial risks. Yet, as the distribution of the training data sets become more biased, then the performance of all models decrease in catching the fraudulent transactions.

As a future work; instead of making performance, the cost based ones comparisons just over the prediction accuracy and TPR/FPR, these comparisons will be extended to include the comparisons over other performance metrics.

8. ACKNOWLEDGMENTS

I would like to express my utmost gratitude to my guide Asst. Bhawna Mallick for introducing me to Credit Card Fraud Detection via including cost factor for detection purpose. And giving me chance to work in this field. I shall always be grateful to her for her valuable guidance and encouragement.

Khyati Chaudhary[1]
Jyoti Yadav[2]

9. REFERENCES

[1] Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arun K. Majumdar, "BLAST-SSAHA Hybridization for Credit Card Fraud Detection," *IEEE Transactions On Dependable And Secure Computing*, vol. 6, Issue no. 4, pp.309-315, October-December 2009S.

[2] A. Shen, R. Tong, and Y. Deng, "Application of classification models on credit card fraud detection," June 2007.

[3] Dahl, J.: Card Fraud. In: Credit Union Magazine (2006).

[4] Ghosh and D.L. Reilly, "Credit Card Fraud Detection with a Neural-Network," Proc. 27th Hawaii International Conference on System Sciences: Information Systems: Decision Support and Knowledge-Based Systems, vol. 3, pp. 621-630, 1994.

[5] Kaiyong Deng, Ru Zhang, Hong Guo,Kaiyong Deng,R Zhang, Dongfang Zhang,WenFeng Jiang,Xinxin Niu "Analysis and Study on Detection of Credit Fraud in E-commerce 2011.

[6] Leila Seyedhossein, Mahmoud Reza Hashemi Mining Information from Credit Card Time Series for Timelier Fraud Detection International Symposium on Telecommunications 2010.

[7] Linda Delamaire (UK), Hussein Abdou (UK), John Pointon (UK), "Credit card fraud and detection techniques: a review"2009.

[8] M.F. Gadi, X. Wang, and A.P. Lago, "Comparison with parametric optimization in credit card fraud detection, 2008.

[9] Md Delwar Hussain Mahdi, Karim Mohammed Rezaul, Muhammad Azizur Rahman "Credit Fraud Detection in the Banking Sector in UK: A Focus on E-Business**."** 2010.

[10] Mirjana Pejic-Bach, "Profiling intelligent systems applications in fraud detection and prevention: survey of research articles", 2010.

[11] Raghavendra Patidar, Lokesh Sharma, "Credit Card Fraud Detection Using Neural Network" 2011.

[12] S. Benson Edwin Raj, A. Annie Portia "Analysis on Credit Card Fraud Detection Methods" 2011.

[13] Sahin, Y., Duman, E.: An overview of business domains where fraud can take place, and a survey of various fraud detection techniques. In: Proceedings of the 1st International Symposium on Computing in Science and Engineering, Aydin, Turkey (2010).

[14] Tej Paul Bhatla, Vikram Prabhu & Amit Dua "Understanding Credit Card Frauds," 2003.

[15] Y. Sahin, E. Duman "Detecting Credit Card Fraud by ANN and Logistic Regression" 2011.