NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY
[IJIERT] ISSN: 2394-3696 Website: ijiert.org
VOLUME 8, ISSUE 5, May.-2021

CREDIT CARD FRAUD DETECTION AND ANALYSIS USING MACHINE LEARNING ALGORITHMS

Vodala Chakshu
Department of ECE, Matrusri Engineering College , Hyderabad, India
chakshuvodala8@gmail.com

G. Sai Chand
Department of ECE, Matrusri Engineering College , Hyderabad, India
saichandgupta33@gmail.com

ABSTRACT
Increase in the number of customers since last decade for credit card usage. The customer's without prior approval many frauds have been detected. The unethical use of credit cards by hackers or credit cards users unwilling to pay back the amount are known as the major credit card frauds. The credit card transactions should be detected earlier by statistical methods if any fraud has been detected. The various algorithms in the machine learning are used to analyze the patterns and frauds. This helps the bank to eliminate the frauds by declining suspected transactions.

Keywords: Fraud, Transaction, Algorithms.

INTRODUCTION
We are using credit card daily for our expenses. In a physical-card based purchase, the card holder presents his card physically to a merchant for making a payment. If the card holder does not realize the loss of card, it can lead to a substantial financial loss to the credit card company. Credit Card Misrepresentation is one of the greatest dangers to business and business foundations today. Just, Master card Misrepresentation is characterized as, "when an individual uses another individuals". Fraud detection involves monitoring the activities of populations of users in order to estimate, perceive or avoid objectionable behaviour, which consist of fraud, intrusion, and defaulting. This problem is particularly challenging from the perspective of learning, as it is characterized by various factors such as class imbalance. The number of valid transactions far outnumber fraudulent ones. Also, the transaction patterns often change their statistical properties over the course of time. In order to minimize costs of detection it is important to use expert rules and statistical based models to make a first screen between genuine and potential fraud and ask the investigators to review only the cases with high risk. When a fraud cannot be prevented, it is desirable to identify. The number of domain constraints and characteristics exaggerate the problem of detection and prevention. Customer irritation is to be avoided. Most banks considers huge transactions, among which very few is fraudulent, often less. Also, only a limited number of transactions can be checked by fraud investigators, i.e. we cannot ask a human person to check all transactions one by one if it is fraudulent or not.
The model is in most cases a parametric function, which allows predicting the likelihood of a transaction to be fraud, given a set of features describing the transaction. They have suggested SVM and Logistic Regression classification models are helpful to improve the performance in detecting the cards. Here the training datasets become more biased and the efficiency of all models decreased in catching fraudulent transactions. Misrepresentation discovery frameworks come into situation when the fraudsters surpass the extortion aversion frameworks and begin false exchanges. Alongside the advancements in the Data Innovation and upgrades in the correspondence channels, misrepresentation is spreading everywhere throughout the world with after effects of vast measure of false misfortune.

IMPLEMENTATION
The credit card dataset is taken and it is pre-processed using suitable means algorithm and the dataset is thoroughly checked if any null element is removed or not. After data pre-processing is done, the credit card rows and columns are analyzed. The dataset should be thoroughly understood in order to start using appropriate algorithms. Fraud act as the unlawful or criminal deception intended to result in financial or

personal benefit. It is a deliberate act that is against the law, rule or policy with an aim to attain unauthorized financial benefit. A comprehensive survey conducted by Clifton Phua and his associates have revealed that techniques employed in this domain include data mining applications, automated fraud detection, adversarial detection. Outlier mining is a field of data mining which is basically used in monetary and internet fields. It deals with detecting objects that are detached from the main system the transactions that aren't genuine. They have taken attributes of customer's behaviour and based on the value of those attributes they've calculated that distance between the observed value of that attribute and its predetermined value. There have also been efforts to progress from a completely new aspect. Attempts have been made to improve the alert feedback interaction in case of fraudulent transaction. In case of fraudulent transaction, the authorised system would be alerted and a feedback would be sent to deny the ongoing transaction. The flowchart of implementing the credit card fraud detection using machine learning algorithm is given below.

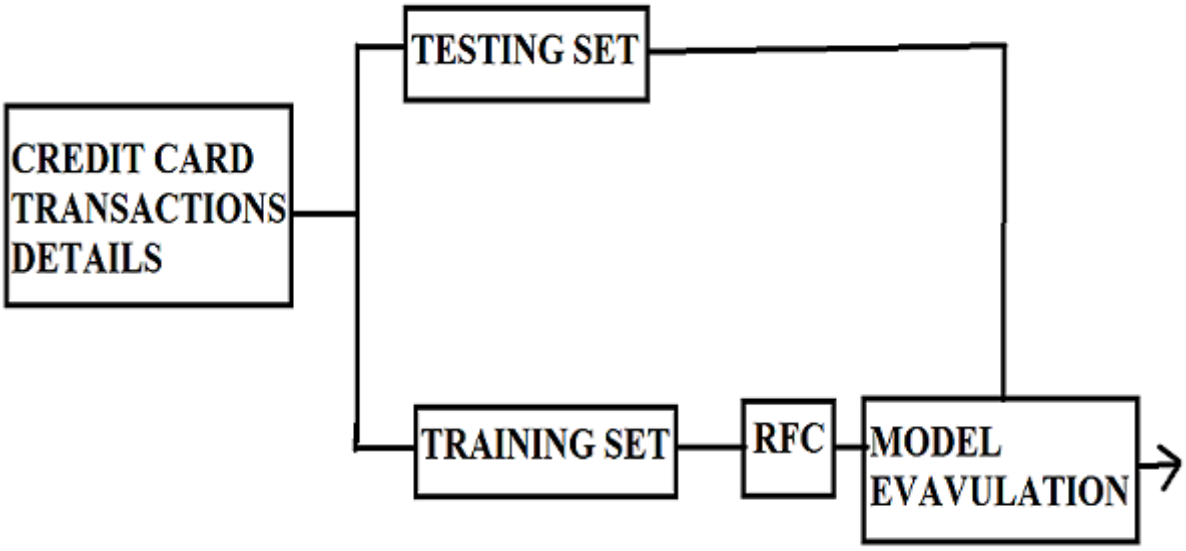

Figure 1. Flowchart of credit card

This process involves some types of iterative algorithms that minimize the error. The aim of the algorithm depends on parameters that control how it works. Because most machine of the learning problems are non-convenes, which means the model depends on our choice of the parameters. So a model may depend on parameters value. By changing these parameters we can find a better model. The dataset was pre-processed for the purpose of improving the performance of the classifiers and reducing their training and operating time. The pre-processing includes investigating the dataset feature space and handling the imbalanced nature of the dataset. This paper would thus be able to help scientist's .what's more, professionals to outline and actualize information digging based frameworks for misrepresentation location or comparative issues.

METHODOLOGY

We are using supervised learning techniques to extract the information as a part of the fraud analysis. The dataset used is a binary classification. Fraud detection is a binary classification task in which any transaction will be predicted and labelled as a fraud or legit. In this paper state of the art classification techniques were tried for this task and their performances were compared. The following subsections briefly explain these classification techniques, data set and metrics used for performance measure.

Support vector machines (SVM) - SVM is introduced in 1992 to solve binary classification problems and then they are extended to nonlinear regression problems. SVMs are based on structural risk minimization unlike ANNs which is based on empirical risk minimization. SVM map the data to a predetermined very high- dimensional space via a kernel function and finds the hyper plane that maximizes the margin between

the two classes. The solution is based only on those data points, which are at the margin. These points are called support vectors.

Naive Bayes Algorithm - Naive Bayes is based on two assumptions. Firstly, all features in an entry that needs to be classified are contributing evenly in the decision Secondly, all attributes are statistically independent, meaning that, knowing an attribute's value does not indicate anything about other attributes' values which is not always true in practice. The process of classifying an instance is done by applying the Bayes rule for each class given the instance. In the fraud detection task, the following formula is calculated for each of the two classes and the class associated with the higher probability is the predicted class for the instance.

Logistic Regression - Logistic regression also does not require independent variables to be linearly related, nor does it require equal variance within each group, which also makes it a less stringent procedure for statistical analysis. As a result, logistic regression was used to predict the probability of fraudulent credit cards. Assumptions and Limitations of Logistic Regression. Logistic regression analysis uses maximum likelihood estimation to predict group membership. However, to interpret the results of the prediction of group membership with precision and accuracy, a preliminary analysis of the cleaned dataset was conducted to observe if the assumptions of logistic regression were met.
The above techniques are used for the detection of fraud transactions involved with the banks.
After the classification, the intensity of the individual cards is inquired and calculated, which prepares and classifies the credit card trend and spending based anomalies. The regression models are used to perform the operation on the given data stream obtained from the credit card company for the detection of the credit card frauds by analysing the spending behaviour of the customers. The structural anomaly modelling is used to detect the credit card frauds with flexibility and variability relationship and payment based relation building, which predicts the spending based anomalies, which leads towards the credit card fraud based decisions.
In order to eliminate the problems in the existing model, the proposed model will be designed with the unbalanced metric normalization methods, where the combination of averages and floating averages (such as Mean, Median etc.) can be utilized to create the state-of-art system in order to minimize the feature unbalance in the feature matrix. In addition to averaging factor based feature description, the flexible and robust feature scaling practices can be utilized, which may vary from column to column in the given data according to its volatility and overall variance, to precise the features in order to create the high accuracy based credit card fraud detection model. The model with best feature selection with probabilistic classification for the purpose of credit card fraud detection would be deployed with certain improvements or enhancements during the proposed model implementation. The best feature selection method will incorporate the selection of the features on the basis of their compatibility, which can be measured with the column or feature variance.

RESULTS

The graphs of the particular algorithms have been obtained and it has been found out that radical basis function obtains 97.2% accuracy using ROC graph. The best algorithm suitable for credit card fraud detection is SVM kernel. The code prints out the number of false positives it detected and compares it with the actual values. This is used to calculate the accuracy score and precision of the algorithms. The fraction of data we used for faster testing is 10% of the entire dataset. The complete dataset is also used at the end and both the results are printed. These results along with the classification report for each algorithm is given in the output as follows, where class 0 means the transaction was determined to be valid and 1 means it was determined as a fraud transaction. The unbalanced data matrices make the credit card fraud detection task tougher by adding the volatility over the given set of data. The unbalanced data distribution requires the balanced feature description in order to normalize the unexpectedly high or low frequencies, which can further improve the performance of the credit card fraud detection models. The accuracy must be improved in order to realize the state-of-art model for the credit card fraud detection.

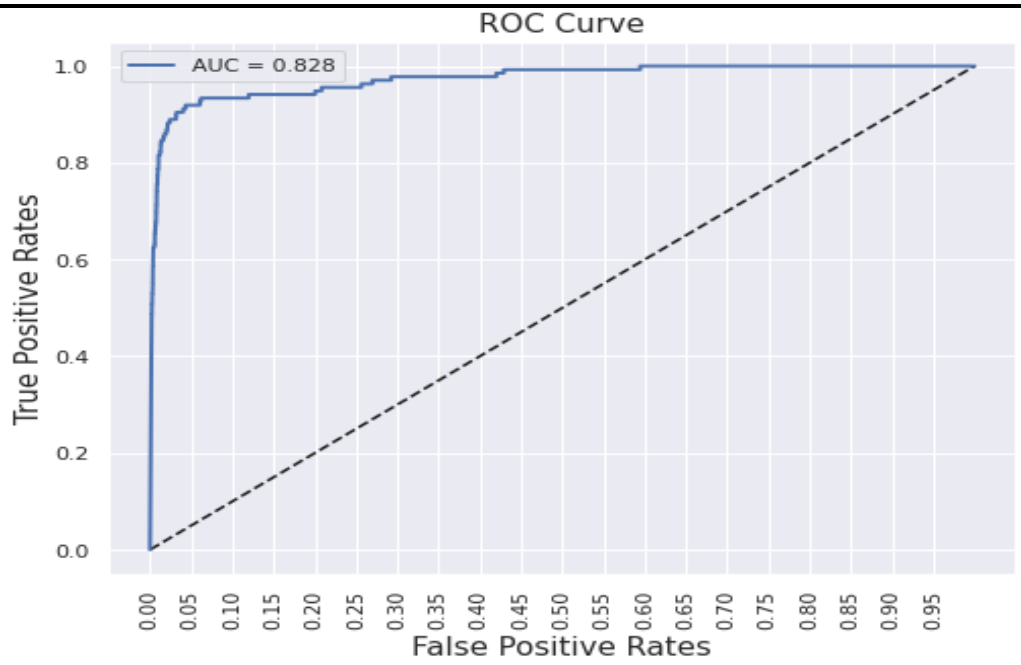

Figure 2. Gaussian accuracy

In this paper we are using the segment portrays a structure that tends to both the issues of stream information mining, idea float and class irregularity for Visa danger appraisal. For adjusting the information, the proposed system executes sub-gatherings with sacking alongside exchange blunder measures. The beneficiary is then deceived into clicking a pernicious connection, which can prompt the establishment of malware, the solidifying of the framework as a component of a ransom ware assault or the noteworthy of delicate data. By comparing the local values of a sample to that of its neighbours, one can identify samples that are substantially lower than their neighbours. These values are quite huge and they are considered as outliers.

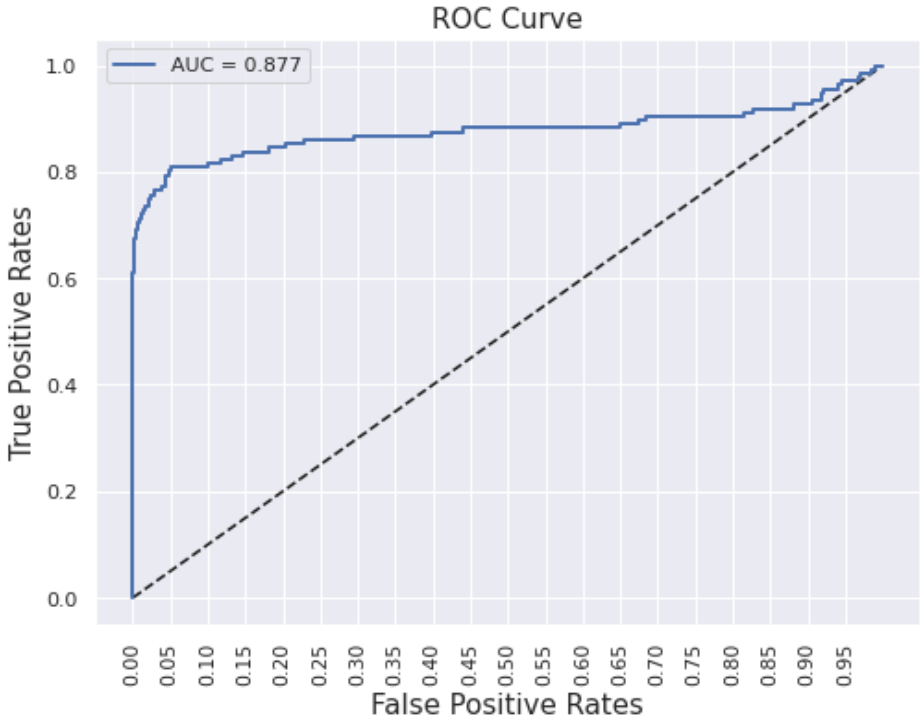

Figure 3. Linear regression accuracy

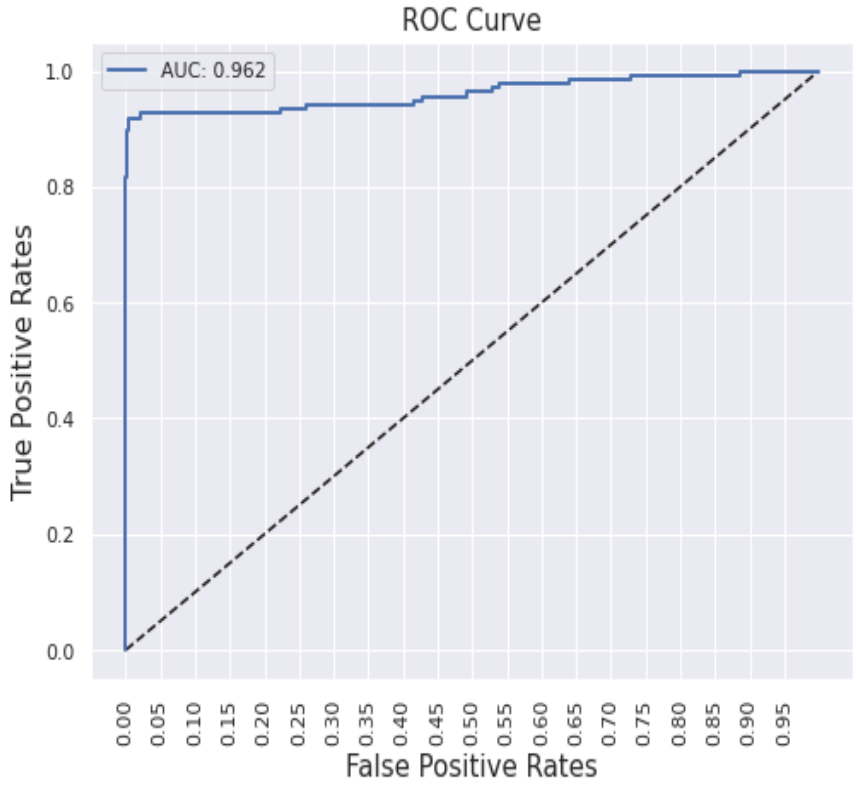

Figure 4. Linear svm graph

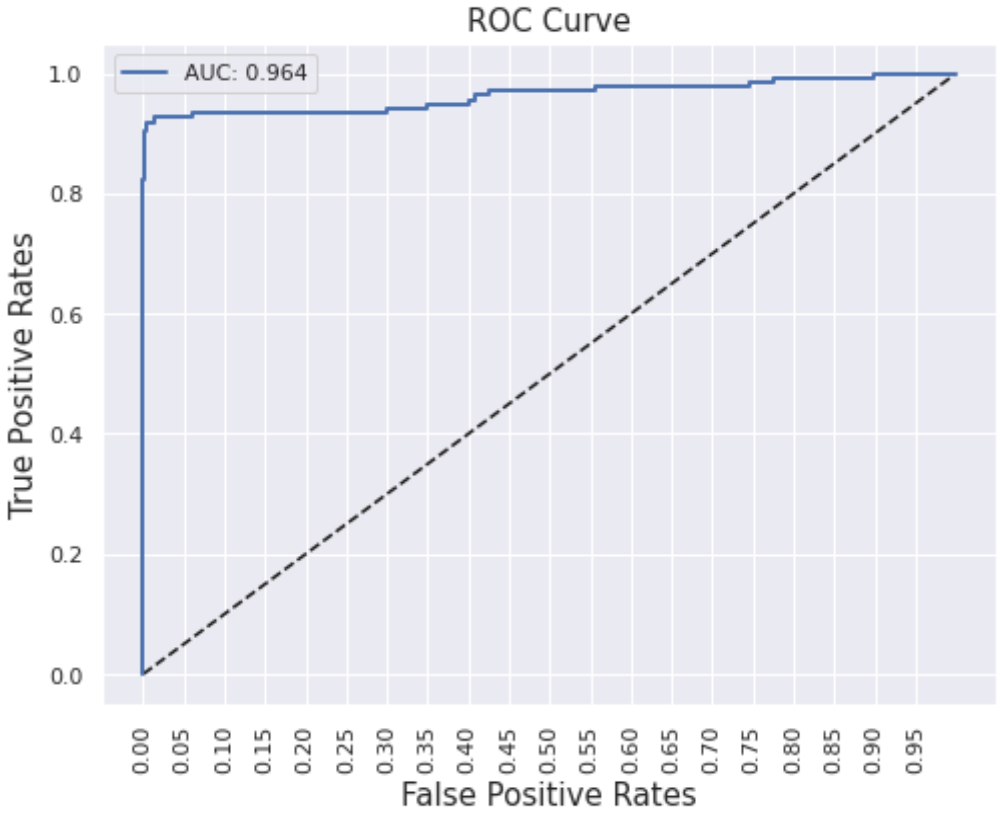

Figure 5. Polynomial svm accuracy

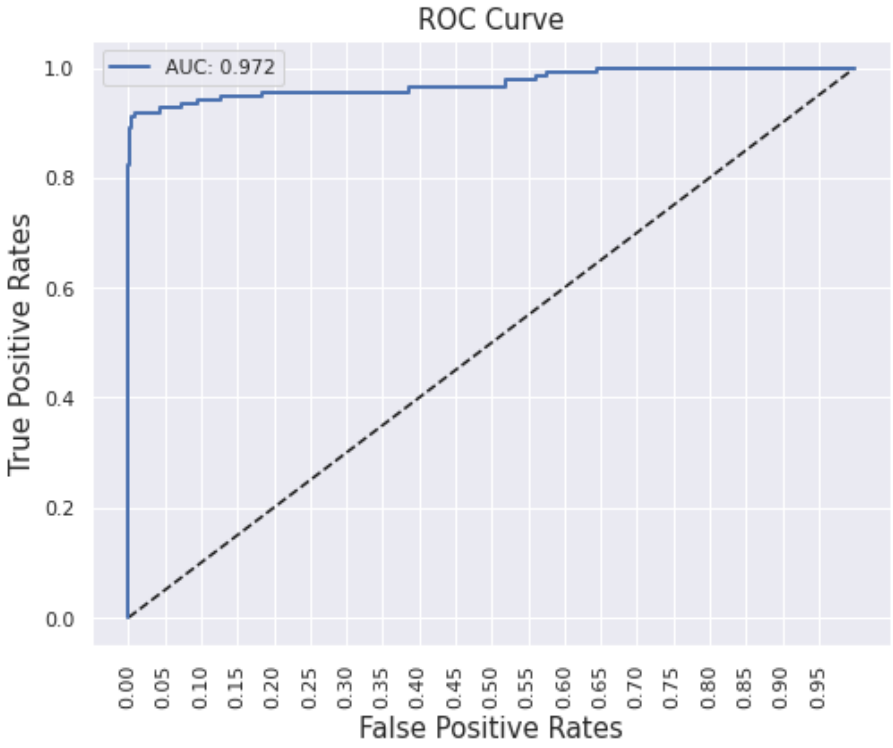

Figure 6. Rbf svm accuracy

CONCLUSION

SVM algorithm, Linear Regression, Gaussian algorithm were used in developing four fraud detection models to classify a transaction as fraudulent or legitimate. Three metrics were used in evaluating their performances. The results showed that there is no data mining technique that is universally better than others. Performance improvement could be achieved through developing a fraud detection model using a combination of different data mining techniques. Credit card fraud is without a doubt an act of criminal dishonesty. This article has listed out the most common methods of fraud along with their detection methods and reviewed recent findings in this field. This paper has also explained in detail, how machine learning can be applied to get better results in fraud detection along with the algorithm, code, explanation its implementation and experimentation results. While the algorithm does reach over 97.2% accuracy, its precision remains only at 25% when a tenth of the data set is taken into consideration. However, when the entire dataset is fed into the algorithm, the precision rises to 30%. This high percentage of accuracy is to be expected due to the huge imbalance between the number of valid and number of genuine transactions.

The credit card fraud detection methods have gained the popularity in the past decade with the evolution of the statistical models. These models are used to automate the process of pattern recognition, which takes comparatively lesser time and can handle a large number of transactions per day. The malicious patterns are also known as outlier or anomaly, which must be detected correctly in order to minimize the bank losses caused by fraudulent transactions. The performance of the proposed model would be evaluated using the precision, recall, F1-measure and accuracy based parameters. By using the supervised learning techniques we are trying to improve the accuracy of the credit card fraud detection based on the data classification method. As the world is becoming digital day by day as a result there is increasing rate of fraud transactions. There are many ways of detection of credit card fraud. If one of these or combination of algorithm is applied into bank credit card fraud detection system, the probability of fraud transactions can be predicted soon after credit card transactions by the banks and a series of anti-fraud strategies can be adopted to prevent banks from great losses before and reduce risks. This paper gives contribution towards the effective ways of credit card fraudulent detection. In this study, three classification methods were used to a deep analysis of the credit cards history business information and have built the fraud detecting models.

FUTURE SCOPE

While we couldn't reach out goal of 100% accuracy in fraud detection, we did end up creating a system that can, with enough time and data, get very close to that goal. As with any such project, there is some room for improvement here. The very nature of this project allows for multiple algorithms to be integrated together as modules and their results can be combined to increase the accuracy of the final result. This model can further be improved with the addition of more algorithms into it. However, the output of these algorithms needs to be in the same format as the others. Once that condition is satisfied, the modules are easy to add as done in the code.

REFERENCES

1) "Survey Paper on Credit Card Fraud Detection by Suman" , Research Scholar, GJUS&T Hisar HCE, Sonepat published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 3, March 2014.
2) "Research on Credit Card Fraud Detection Model Based on Distance Sum – by Wen-Fang YU and Na Wang" published by 2009 International Joint Conference on Artificial Intelligence.
3) "Credit Card Fraud Detection through Parenclitic Network AnalysisBy Massimiliano Zanin, Miguel Romance, Regino Criado, and SantiagoMoral" published by Hindawi Complexity Volume 2018, Article ID 5764370, 9 pages.
4) Tej Paul Bhatla, Vikram Prabhu, and Amit Dua. Understanding credit card frauds. Cards business review, 1(6), 2003.
5) Christopher M Bishop et al. Pattern recognition and machine learning, volume 4. Springer New York, 2006.
6) Van Vlasselaer, Véronique, Cristián Bravo, Olivier Caelen, Tina Eliassi-Rad, Leman Akoglu, Monique Snoeck, and Bart Baesens. "APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions." Decision Support Systems 75 (2015): 38-48.
7) Prakash, A., and C. Chandrasekar. "An optimized multiple semi-hidden markov model for credit card fraud detection." Indian Journal of Science and Technology 8, no. 2 (2015): 165-171.
8) Zareapoor, Masoumeh, and Pourya Shamsolmoali. "Application of credit card fraud detection: Based on bagging ensemble classifier." Procedia Computer Science 48 (2015): 679-685.