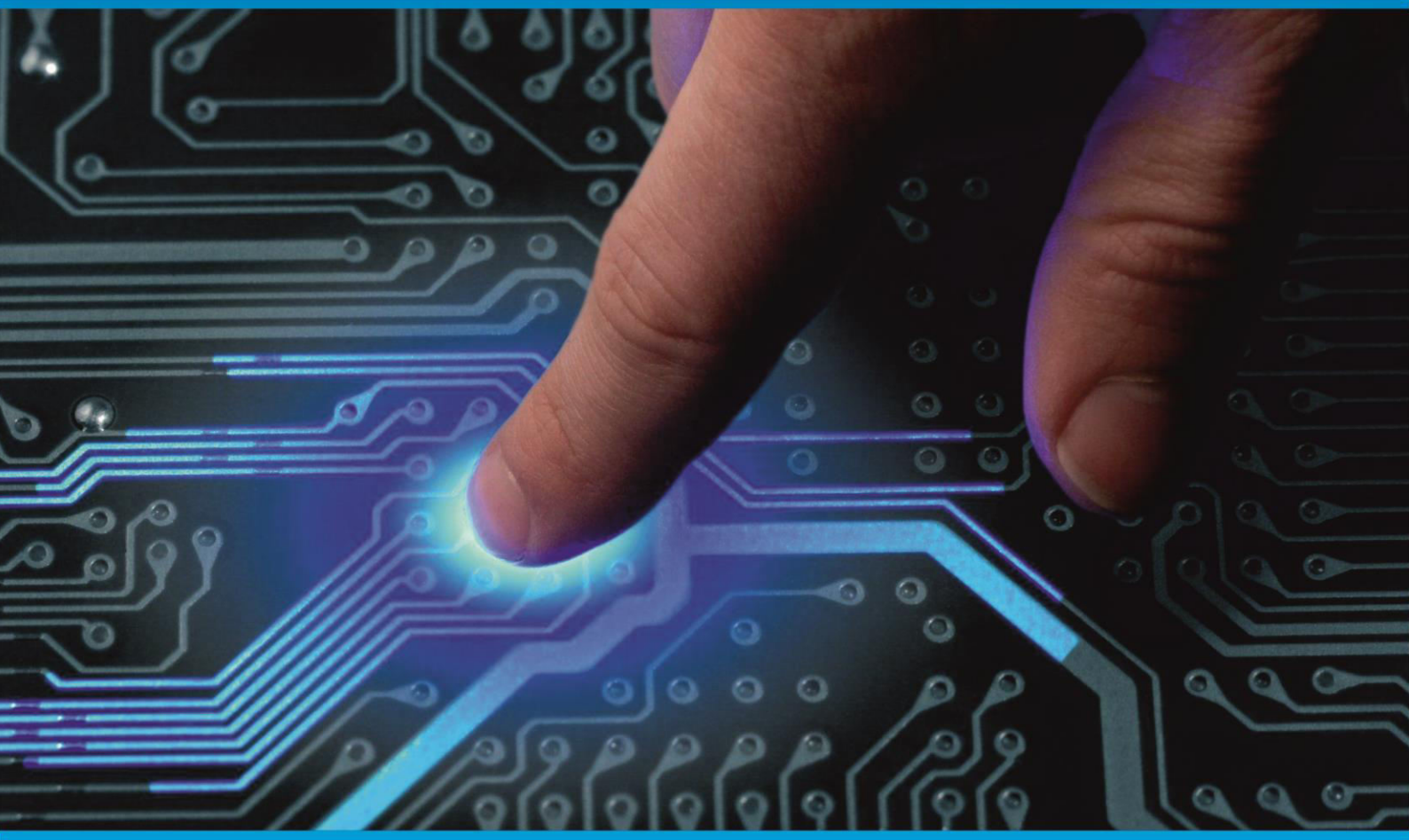
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 3, March 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903210|

Credit Card Fraud Detection Based on Machine Learning

Poongganesh M, Pradeep R, Pranavkumar S, Rahulraja.J, Mr.V.Pradeep Anand, M.Tech

Department of Computer Science and Engineering, Paavai College of Engineering, Nammkal, Tamil Nadu, India

Assistant Professor, Department of Computer Science and Engineering, Paavai College of Engineering, Nammkal, Tamil Nadu, India

ABSTRACT: Fraud is any malicious activity that aims to cause financial loss to the other party. As the use of digital money or plastic money even in developing countries is on the rise so is the fraud associated with them. Frauds caused by Credit Cards have costs consumers and banks billions of dollars globally. Even after numerous mechanisms to stop fraud, fraudsters are continuously trying to find new ways and tricks to commit fraud. Thus, in order to stop these frauds we need a powerful fraud detection system which not only detects the fraud but also detects it before it takes place and in an accurate manner. We need to also make our systems learn from the past committed frauds and make them capable of adapting to future new methods of frauds. In this paper we have introduced the concept of frauds related to credit cards and their various types. We have explained various techniques available for a fraud detection system such as Support Vector Machine (SVM), Artificial Neural Networks (ANN), Bayesian Network, K- Nearest Neighbour (KNN), Hidden Markov Model, Fuzzy Logic Based System and Decision Trees. An extensive review is done on the existing and proposed models for credit card fraud detection and has done a comparative study on these techniques on the basis of quantitative measurements such as accuracy, detection rate and false alarm rate. The conclusion of our study explains the drawbacks of existing models and provides a better solution in order to overcome them.

KEYWORDS: Neural Network, Genetic Algorithm, Support Vector Machine, Bayesian Network, K- Nearest Neighbour, Hidden Markov Model, Fuzzy Logic Based System, Decision Trees.

I. INTRODUCTION

Due to rise and acceleration of E- Commerce, there has been a tremendous use of credit cards for online shopping which led to High amount of frauds related to credit cards. In the era of digitalization the need to identify credit card frauds is necessary. Fraud detection involves monitoring and analyzing the behavior of various users in order to estimate detect or avoid undesirable behavior. In order to identify credit card fraud detection effectively, we need to understand the various technologies, algorithms and types involved in detecting credit card frauds. Algorithm can differentiate transactions which are fraudulent or not. Find fraud, they need to passed dataset and knowledge of fraudulent transaction. They analyze the dataset and classify all transactions.

Credit card fraud is a major problem that involves payment card like credit card as illegal source of funds in transactions. Fraud is an illegal way to obtain goods and funds. The goal of such illegal transaction might be to get products without paying or gain an unauthorized fund from an account. Identifying such fraud is a troublesome and may risk the business and business organizations. In the real world FDS [1], investigator are not able to check all transactions. Here the Fraud Detection System monitors all the approved transactions and alerts the most suspicious one. Investigator verifies these alerts and provides FDS with feedback if the transaction was authorized or fraudulent. Verifying all the alerts everyday is a time consuming and costly process. Hence investigator is able to verify only few alerts each day. The rest of the transactions remain unchecked until customer identifies them and report them as a fraud. Also the techniques used for fraud and the cardholder spending behavior changes over time. This change in credit card

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903210|

transaction is called as concept drift [1] [7]. Hence most of the time it is difficult to identify the credit card fraud. Machine Learning is considered as one of the most successful technique for fraud identification. It uses classification and regression approach for recognizing fraud in credit card. The machine learning algorithms are divided into two types, supervised [14][18] and unsupervised [16] learning algorithm. Supervised learning algorithm uses labeled transactions for training the classifier whereas unsupervised learning algorithm uses peer group analysis [23] that groups customers according to their profile and identifies fraud based on customers spending behavior Many learning algorithm have been presented for fraud detection in credit card which includes neural networks [14][19][21][22], Logistic Regression [3], decision tree [4][15], Naive Bayes [6], Support Vector Machines [5], K-Nearest Neighbors [6] and Random Forest [1][2]. This paper examines the performance of above algorithms based on their ability to classify whether the transaction was authorized or fraudulent and then compares them. The comparison is made using performance measure accuracy, specificity and precision. The result showed that Random Forest algorithm showed better accuracy and precision than other techniques.

Today use of Credit Card even in developing countries has become a common scenario. People use it to shop, pay bills and for online transactions. But with increase in number of Credit Card users, the cases of fraud in Credit Card have also been on rise. Credit Card related frauds cause globally a loss of billions of dollars. Fraud can be classified as any activity with the intent of deception to obtain financial gain by any manner without the knowledge of the cardholder and the issuer bank. Credit Card fraud can be done in numerous ways. By lost or stolen cards, by producing fake or counterfeit cards, by cloning the original site, by erasing or modifying the magnetic strip present at the card which contains the user's information, by phishing, by skimming or by stealing data from a merchant's side. Fraud detection deals with finding a fraud activity amongst thousands of genuine ones, which in fact puts forward a challenge. With continued advancement in fraudulent strategies it is important to develop effective models to combat these frauds in their initial stage only, before they can take to completion. But the major challenge in developing such a model is that the number of fraudulent transactions among the total number of transaction is a very small number and hence the work of finding a fraudulent transaction in an effective and efficient way is quite bothersome.

Credit card frauds can be of following types:

1. Application Frauds: When the fraudster gains control of the application system by accessing sensitive user details like password and username and open a fake account. It generally happens in relation to the identity theft. When the fraudster applies for credit or a new credit card altogether in the name of the card holder. The fraudster steals the supporting documents in order to support or substantiate their fraudulent application.
2. Electronic or Manual Credit Card Imprints: When the fraudster skims information that is placed on the magnetic strip of the card. This information is very confidential and by accessing it the fraudster may use it for fraudulent transactions in future.
3. CNP (Card Not Present): When the fraudster knows the expiry date and account number of the card, the card can be used without its actual physical possession.
4. Counterfeit Card Fraud: It is generally attempted through the process of skimming. A fake magnetic swipe card is made and it holds all the details of the original card. The fake card is fully functional and can be used to commit transactions in future.
5. Lost and Stolen Card Fraud: In cases when the original card holder misplaces their card, it can get to the hands of fraudsters and they can then use it to make payments. It is hard to do this through machine as a pin number is required however; online transactions are easy enough for the fraudster.
6. Card ID Theft: This fraud is similar to application frauds. In ID theft the fraudster acquires the details of the original card to make use of a card or to open a new account. This type of fraud is the most difficult to identify.
7. Mail Non-Receipt Card Fraud: When a customer applies for a card, it takes some time for all the procedural formalities. If fraudster intercepts in the middle of the delivery, they may register the card in their name and may use it to make purchases. This fraud is also known as never received issue fraud.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903210|

II. LITERATURE REVIEW

You Dai, et. al [2] In this paper, they describe Random forest algorithm applicable on Find fraud detection. Random forest has two types, i.e. random tree based random forest and CART based random forest. They describe in detail and their accuracy 91.96% and 96.77% respectively. This paper summarise second type is better than the first type. Suman Arora [3] In this paper, many supervised machine learning algorithms apply on 70% training and 30% testing dataset. Random forest, stacking classifier, XGB classifier, SVM, Decision tree, naïve Bayes and KNN algorithms compare each other i.e. 94.59%, 95.27%, 94.59%, 93.24%, 90.87%, 90.54% and 94.25% respectively. Summarise of this paper, SVM has the highest ranking with 0.5360 FPR, and stacking classifier has the lowest ranking with 0.0335. Kosemani Temitayo Hafiz [4] In this paper, they describe flow chart of fraud detection process. i.e. data Acquisition, data pre-processing, Exploratory data analysis and methods or algorithms are in detail. Algorithms are K- nearest neighbour (KNN), random tree, AdaBoost and Logistic regression accuracy are 96.91%, 94.32%, 57.73% and 98.24% respectively.

There are different supervised and unsupervised learning algorithms used for fraud detection in credit card. Some important are described below. The author [1] has proposed a paper where they have first explained the proper performance measures which is used for fraud identification. The authors have structured a novel learning technique that can solve concept drift, verification latency, and class imbalance issues. The paper also showed effect of above issues in true credit card transactions. Here in paper [2] authors presented two types of classifier using random forests which are used to train the behavior features of transactions. The authors have compared the two random forests and have analyzed their performance on fraud identification in credit card. In paper [3] authors presented a FDS for credit card using Artificial Neural Network and Logistic Regression. The system used to monitor each transaction separately using classifier and then classifier would generate score for each transaction and label this transaction as legal or illegal transaction. A decision tree method was proposed in paper [4]. The method decreased overall misclassification costs and selected splitting property at each node. The author also compared the decision tree method for fraud identification with other models and proved that this approach performs well using performance measure like accuracy and genuine positive rate. The author [5] developed a FDS for credit card transaction using support vector machines and decision tree. This study built seven alternative models that were created using support vector machines and decision tree. The author also compared this classifiers performance using performance measure accuracy. The study also showed that as size of training dataset increases the number of fraud detected by SVM are less than fraud identified by decision tree method. Here in [6] author presented fraud detection system using a Naive Bayes K-Nearest Neighbors method. The main aim of proposed system was to improve accuracy. Naive Bayes Classifier predicts probabilities of fraud in transaction while KNN classifier predicts how near the undefined sample data is to kth training dataset. The author compared both this classifier and showed that both work differently for given dataset. Most of predictive model used for detecting fraud in credit card transaction faces the issue of concept drift. The author [7] presented two FDS based on sliding window and ensemble learning and showed that classifier need to be trained separately using feedback and delayed samples. The outcome of the two was than aggregated to improve the alert precision in FDS. Thus the author showed that to solve the issue of concept drift, the feedback and delayed samples are to be handled separately.

III. PROPOSED SYSTEM

Today modern society is using credit cards for variety on reasons. Similarly fraud in credit card transactions has been growing in recent years. Each year, a huge amount of financial losses are caused by the illegal credit card transactions. Fraud may occur in variety of different forms and may be limited. Therefore there is need to solve the issues of fraud detection in credit card. Additionally, with the development of new technologies criminals finds new ways to commit fraud. To overcome this problem the proposed system for fraud detection in credit card transactions will be designed using ML technique that will provide investigator a small reliable fraud alerts.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903210|

IV. VARIOUS TECHNIQUES OF CREDIT CARD FRAUD DETECTION

We know that all fraudulent transaction follow a similar pattern and by using any pattern recognition system such as Support Vector Machine (SVM), Artificial Neural Networks, Naïve Bayesian Network, K- Nearest Neighbour (KNN), Hidden Markov Model, Fuzzy Logic Based System or Decision Trees we can classify transactions as fraudulent whose working is explained below.

Artificial Neural Network It combines the thinking power of human brain with computational power of machine. It makes use of neurons as the deciding sites and the edges between neurons to calculate the contribution of each neuron in the previous layer in the decision and result at the current neuron. It is based on pattern recognition. Previous year's data is fed into the network and then based upon that data it recognises a new incoming transaction to be a fraud or genuine one. Its training can either be supervised i.e. the outcome is already known for a given transaction and the expected output is compared with actual to train the system or it can be unsupervised where we have no actual results to compare it with and thus are not sure about the results. [1]

Decision Tree It is a computational tool for classification and prediction. A tree comprises of internal nodes which denote a test on an attribute, each branch denotes an outcome of that test and each leaf node (terminal node) holds a class label. It recursively partitions a dataset using either depth first greedy approach or breadth first greedy approach and stops when all the elements have been assigned a particular class. For the partition rule to be efficient it must separate the data into groups where a single class predominates in each group. In other words, the best partition will be the one in which the subsets do not overlap i.e. they are clearly disjoint to a maximum amount. [2]

Fuzzy Logic It is used in the cases when we do not have discrete truth values i.e., they are continuous. It is a multivalued logic. There are certain set of rules based on which a transaction is classified as a genuine or fraud one. There are three important components in fuzzy logic that need to be executed in the stated order: [3] • Fuzzification • Rule Based • Defuzzification In fuzzification we classify an incoming transaction in the categories of high, low or medium based upon the monetary value associated with the transaction. Rule based deals with drafting the rules based on the customer behaviour. The transaction is allowed to occur if it satisfies given set of rules. In Defuzzification, if a transaction does not comply with the predefined set of rules it isn't allowed to occur. It is immediately stopped and then cross checked with the customer that whether it should be granted the permission to continue or be aborted.

Support Vector Machines It is a supervised learning algorithm in which given a dataset it separates them into different classes using a hyperplane. The goal of SVM is to find this hyperplane. There could be many hyperplanes but we are determined to find an optimal hyperplane. The points closest to the hyperplane in the different classes are known as support vectors and these support vectors are used to predict the classes of new data points. A new incoming point is put on the equation of the hyperplane and then is classified as to which class it belongs on the basis of which side of hyperplane it falls on the vector space. To train our machine we feed supervised data i.e. data with results already known. It learns the behaviour of fraud and genuine transactions and then it can classify new transaction as to which class it belongs. [4]

Bayesian Network It is based upon the Bayes Theorem of conditional probability; hence it is a probabilistic model that is used for automated detection of various events. It consists of nodes and edges, wherein the nodes represent the random variables and the edges between the nodes represent the relationships between these random variables and their probabilistic distribution. We calculate predefined minimum and maximum value of probabilities of a transaction being fraud or legal. Then for a new incoming transaction we see that whether it's probability of being legal is less than the minimum defined value for legal transaction and is greater than the maximum defined value for a fraud transaction. If true then the transaction is classified as a fraud. [5]

K- Nearest Neighbour It is one of the most used algorithms for both classification and regression predictive problems. Its performance depends on three factors: the distance metrics, the distance rule and the value of K. Distance metrics gives the measure to locate nearest neighbours of any incoming data point. Distance rule helps us to classify the new data point into a class by comparing its features with that of data points in its neighbourhood. And the value of K decides the number of neighbours with whom to compare. The important question is how do we choose the factor K? In order to obtain the optimal value of K, the training and validation is segregated from the initial dataset. Now a graph based on the validation error curve is plotted to achieve the value of K. This value of K should be used for all

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903210|

predictions. We calculate the dominant class in the vicinity of any new transaction and classify the transaction to belong to that dominant class. [6]

Hidden Markov Model There is a change of state with time hence the name markov. The states are hidden hence cannot be observed directly. But something correlated to them can be observed and based on that sequence of observations we predict the order of state changes. We first train our model based upon given set of parameters like spending habit of cardholder. Initial set of probabilities are chosen based on this profile. Then any new incoming transaction is analysed by our model and classified as fraudulent if it varies from the general profile and behaviour of a cardholder by more than a threshold value and hence it cannot be accepted by the states in hidden markov model. [7]

Logistic Regression To combat the anomalies of linear regression where it gave values greater than 1 and less than 0, logistic regression comes into play. Despite the name being regression, LR is used for classification problems for predicting binomial and multinomial outcomes, having the goal of estimating the values of parameter's coefficients using the sigmoid function. Logistic regression is used for clustering and when a transaction is ongoing it examines the values of its attributes and tells whether the transaction should proceed or not. [8]

V. CONCLUSION

This paper has reviewed various machine learning algorithm detect fraud in credit card transaction. The performances of all this techniques are examined based on accuracy, precision and specificity metrics. We have selected supervised learning technique Random Forest to classify the alert as fraudulent or authorized. This classifier will be trained using feedback and delayed supervised sample. Next it will aggregate each probability to detect alerts. Further we proposed learning to rank approach where alert will be ranked based on priority. The suggested method will be able to solve the class imbalance and concept drift problem. Future work will include applying semi-supervised learning methods for classification of alert in FDS.

REFERENCES

[1] Heta Naik, "Credit card fraud detection for Online Banking transactions", International Journal for Research in Applied Science & Engineering Technology, pp 4573- 4577, 2018https://www.ijraset.com/fileserve.php?FID=16732

[2] You Dai, Jin Yan, Xiaoxin Tang, Han Zhao and Minyi Guo, "Online Credit CardFraud Detection: A Hybrid Framework with Big Data Technologies", IEEE TrustCom/BigDataSE/ISPA , pp 1644 -1651, 2016

[3] Suman Arora , "Selection of Optimal Credit Card Fraud Detection Models Using a Coefficient Sum Approach" , International Conference on Computing, Communication and Automation (ICCCA2017), pp 482 - 487, 2017

[4] Kosemani Temitayo Hafiz, Dr. Shaun Aghili and Dr. Pavol Zavarsky, "The Use of Predictive Analytics Technology to Detect Credit Card Fraud in Canada",

[5] N.Malini and Dr.M.Pushpa , "Analysis on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection" , 3rd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEEICB17) , 2017

[6] Anusorn Charleonnan , "Credit Card Fraud Detection Using RUS and MRN Algorithms" , The 2016 Management and Innovation Technology International Conference (MITiCON-2016) , pp 73 - 76, 2016

[7] John Richard D. Kho and Larry A. Vea, "Credit card Fraud detection based on transaction Behavior", IEEE Region 10 Conference (TENCON), Malaysia, pp 1880 – 1884 , November 2017

[8] Fahimeh Ghobadi and Mohsen Rohani, "Cost Sensitive Modeling of Credit CardFraud Using Neural Network Strategy" , IEEE ICSPIS 2016, Dec 2016

[9] S Md. S Askari and Md. Anwar Hussain, "Credit Card Fraud Detection Using Fuzzy ID3" , International Conference on Computing, Communication and Automation (ICCCA2017), pp 446 - 452 , 2017

[10] Sarween Zaza and Mostafa Al-Emran, "Mining and Exploration of Credit Cards Data in UAE", Fifth International Conference on e-Learning , pp 275-79 , 2015

[11] Krishna Keerthi Chennam and Lakshmi Mudanna, "Privacy and Access Control for Security of Credit Card Records in the Cloud using Partial Shuffling", IEEE International Conference on Computational Intelligence and Computing Research, 2016

[12] Rajeshwari U and Dr B Sathish Babu, "Real-time credit card fraud detection using Streaming Analytics", 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), pp 439 – 444, 2016

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903210|

[13] John O. Awoyemi, Adebayo O. Adetunmbi and Samuel A. Oluwadare, "Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis", IEEE , 2017
[14] Mukesh Kumar Mishra and Rajashree Dash, "A Comparative Study of Chebyshev Functional Link Artificial Neural Network, Multi-Layer Perceptron and Decision Tree for Credit Card Fraud Detection", International Conference on Information Technology, pp 228 -233, 2014
[15] Pornwatthana Wongchinsri and Werasak Kuratach, "A Survey - Data Mining Frameworks in Credit Card Processing", IEEE, 2016

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details