INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Credit Card Fraud Detection for Online Banking Transactions

Heta Naik[1]

[1]*M. Tech (Student), NMIMS University, Mumbai*

Abstract: Online transactions are increasing day by day. It is very helpful in daily routine life. When online transactions are increasing, fraud transactions also increases. Some algorithm and techniques are applied in online backing for detection of fraudulent transaction and reducing it. In this paper various algorithm like K- Nearest Neighbour, Random Tree, AdaBoost and Logistic Regression are used and a comparative study is shown between them. A methodology to detect fraudulent transactions is proposed in this system. There are various challenges which includes distinguishing between normal and fraudulent transaction seems it looks very similar to each other. Parameters to detect such transactions are Time, Amount and Transaction Frequency.
Keywords: Credit card, Fraud Detection, K- Nearest Neighbour, Random Tree, AdaBoost, Logistic Regression

I. INTRODUCTION

Now a days, India is becoming digital India. In 2016, Demonetization took place happened. After that cashless transactions were suddenly increased through applications such as Paytm, Bhim, Hike, etc. and they have their e-wallets as well. All things can be done by online transactions such as Ticket Booking, Money transfer, shopping, etc.

For Online Transaction through any E-wallet, Credit cards are connected with the accounts details such as Credit card Number, Name, CVV, Expiry date, etc. Online transaction is very helpful in our daily routine life but with the ease there comes a risk of fraud generation. Many techniques have been applied for detection of fraud in credit card such as, Machine Learning, Neural Network, K- nearest neighbour algorithm, Outlier algorithm, Decision tree algorithm, K –means algorithm, Random tree algorithm, J48 algorithm, SVM, Big data Technology, Sum approach, RUSMRN algorithm, Naïve Bayes algorithm, RUSboost algorithm, Adaboost algorithm, Fuzzy ID3 algorithm, etc. This paper seeks to carry out comparative analysis of credit card fraud detection using K- Nearest Neighbour, AdaBoost, Random Tree and Logistic Regression on specific dataset.

II. LITERATURE SURVEY

N. Malini et al [1] the author described advantages and disadvantages of Logistic Regression, Decision tree, Artificial Neural Network, Hidden Markov model, Support vector machine and KNN algorithms. The author gives classification between fraudulent and non – fraudulent transactions with use of KNN. Also gives details about outlier detection algorithms with their types, supervised and unsupervised outlier detection. They proved that KNN algorithm gives accurate and efficient result and outlier detection algorithm works faster and gives better results for online large datasets. It uses less memory. Conclude that KNN algorithm extremely well in CCFD (Credit Card Fraud Detection). A similar research domain was presented by Anusorn Charleonnan [2] they proposed a new algorithm "RUSMRN algorithm" based on three classifiers which are RUS, MRN and Naïve Bayes algorithms. The author gives a basic description about these three algorithms and proposed an algorithm step by step. After the implementation of proposed algorithm they analyse and conclude that RUSMRN is appropriate for predicting the data because it has the best classification performance in terms of accuracy and sensitivity. When compared to the sensitivity of other algorithms namely RUSBoost, AdaBoost and naïve Bayes (50.3%, 31.4% and 40.2% respectively), the sensitivity of this algorithm is much higher, of 53.36%. Coming to accuracy, the accuracy of the proposed algorithm is 79.73 % which is also higher than the accuracy of RUSBoost, AdaBoost and naïve Bayes algorithm (77.8%, 57.73% and 70.13% respectively). Another similar domain literature is from John Richard D. Kho and Larry A. Vea [3] they gives proper flow or methodology for detecting the fraudulent transaction such as Data preparation, data Analysis, modelling and testing. It states that J48 and Random tree is higher accuracy 93.50% and 94.32% respectively. In this paper these two algorithm analyzed with different splits and skews. The highest accuracy rate produced over J48 is that of The Random Tree. Although, by seeing the randomness of the dataset, it is concluded that tight bound id yielded by J48 with respect to its variance in accuracy values. In the paper of S Md. S Askari and Md. Anwar Hussain [6], proposed an algorithm for fraud detection which was based on Fuzzy-ID3. Using the attributes which have the most information gain, they split the intermediate nodes. For the purpose of this branching, FuzzITree algorithm is used. The transactions are then classified as fraud, doubtful or normal on the basis of the leaf node.

There are six different types of credit card fraud namely Triangulation Fraud b. Cramming/Salami Attack c. Behavioural Fraud. d. Bankruptcy Fraud. e. Application Fraud. f. Theft Fraud/Counterfeit Fraud. Masoumeh [4] showed a comparative study between different techniques for credit card fraud detection.They have used various methods like decision tree, NN , Support vector machines, KNN , genetic algorithms and thus created a hybrid approach for reducing misclassification costs. Accuracy , speed and cost were taken into consideration. Comparison table was then prepared to show the difference between their accuracy results. It was noted that the algorithms having fast processing speeds were giving medium to low accuracy. HMM and bayesian network can be taken as examples of these. AIS (Artificial Immune System) showed good accuracy with notable speed. Genetic algorithms and HMMs are inexpensive models to showcase fraud detection with good accuracy and speed.

III. PROPOSED WORK

A. Data Acquisition

Data acquisition is the first step of any process for data analysis. Here, we have taken the data from UCI repository. The data contains raw information about the transactions occurring through credit card .It also contains missing values as well as outliers which can be handled in upcoming steps.

B. Data Preprocessing

In data preprocessing we apply different methodologies to clean the data, handling the missing values ,noisy data making the data into one form ,making the data into a single range putting all this things together makes a single step of data preprocessing

C. Exploratory Data Analysis

In exploratory data analysis we plot different graphs to view information in the pictorial form. From this information we interpret some results that will be helpful for further data mining process. It is a simple way of viewing the data.
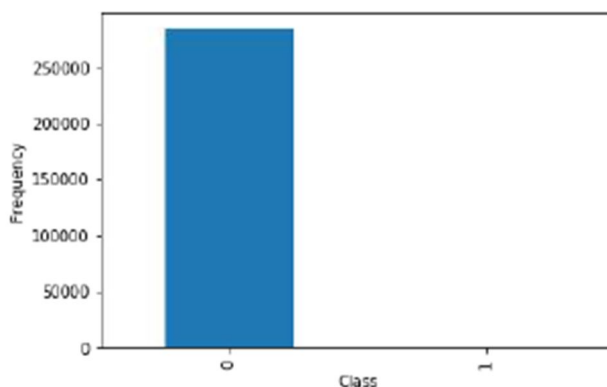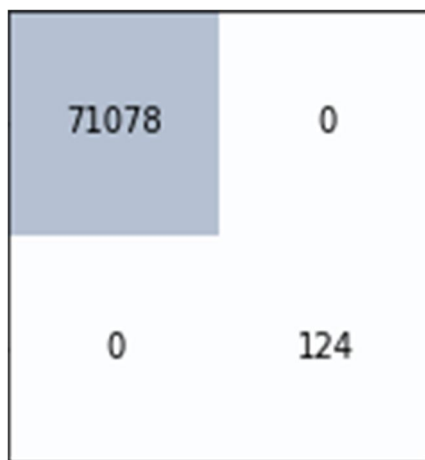

Fig. 1 Imbalance Data

Fig. 2 The Confusion Matrix

D. Methods/ Algorithms

1) *K- Nearest Neighbour (KNN):* In KNN model, the outcome is class oriented. An input object is set into different groups by max voting of its neighbours, with the input object being placed to the class having most similar properties as that of its K neighbours (k is a small positive integer). If k = 1, then the input object is directly assigned to the class of the only nearest neighbour [1] [12].

2) *Random Tree:* Random forests or random decision forests are a group learning technique like regression, classification, tasks for construction of decision trees preparing time and yielding the class that is the method of the classes (classification) or mean expectation (regression) of the individual trees. Random decision forests are appropriate for decision trees; a habit of over-fitting to their training samples [3].

3) *AdaBoost Algorithm:* AdaBoost is best used to boost the performance of any machine learning algorithms. Mainly use for Binary classification problems. AdaBoost Algorithm is very less sensitive and high Specificity i.e. 31.4& and 86.16% respectively [2].

4) *Logistic Regression*: Logistic regression is a factual method for analysis of dataset which has at least one self-dependent variable that determine an output. The resultant output is with a classification variable (Which comprises of only two possible outputs).

E. Comparing the Results

|	KNN [12]	Random Tree[3]	AdaBoost Algorithm [2]	Logistic Algorithm [12]
Accuracy	0.9691	0.9432	0.5773	0.9824
Sensitivity	0.8835	-	0.3140	0.9767
Specificity	0.9711	-	0.8616	0.9824
Limitations	Cannot detect the fraud at the time of transaction	Not Suitable for Randomness in dataset	Cannot detect the fraud at the time of transaction	Not Applied for non - Linear data

Table 1. Comparing Algorithms
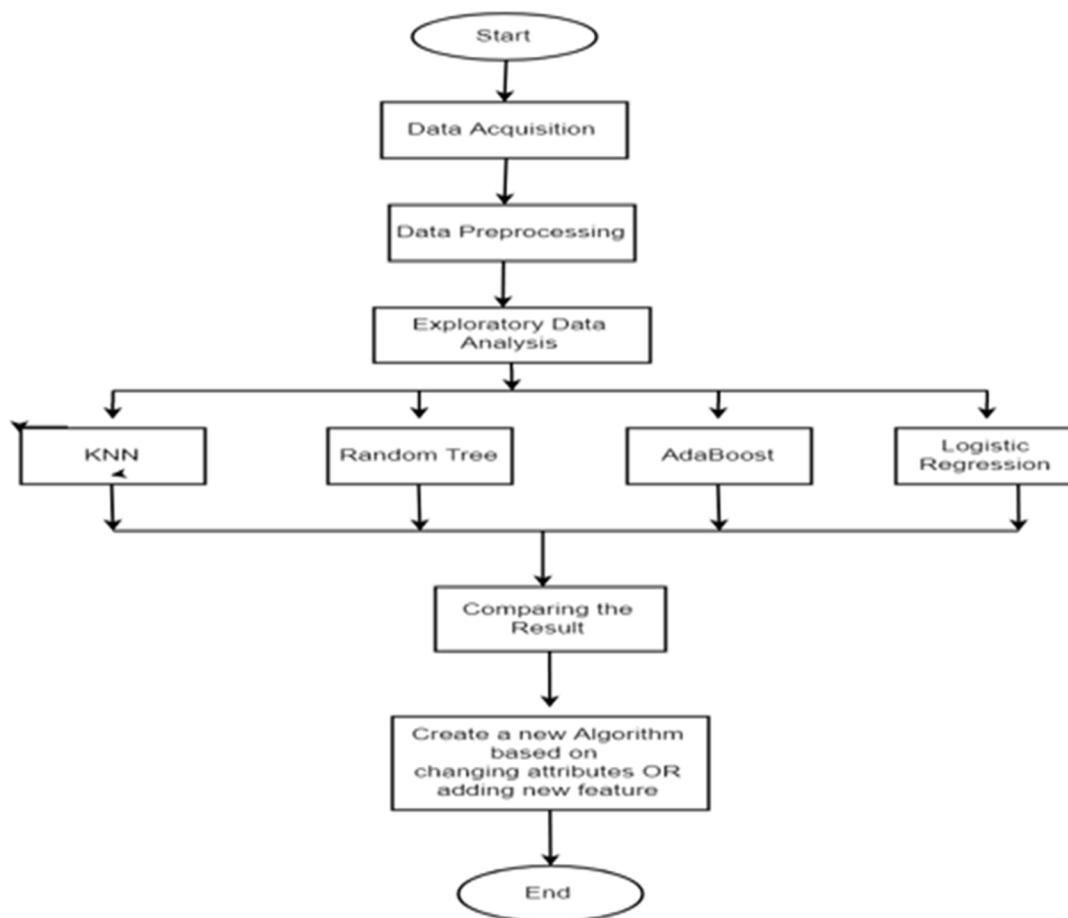
Fig. 3 Flow chart for Fraud detection

IV. CONCLUSION

In this paper, Different four algorithms KNN, AdaBoost, Random tree and Logistic regression are compared for fraud detection mechanism. Logistic regression is better as compared to other algorithms. This model used for imbalanced credit card fraud data. These all algorithms are not applicable for fraud detection at the time of transaction.

REFERENCES

[1] 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

[2] 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

[3] 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 201

[4] Masoumeh Zareapoor, et.al., "Analysis of Credit Card Fraud Detection Techniques: based on Certain Design Criteria", International Journal of Computer Applications (0975 – 8887), vol - 52, pp 35 - 42, August 201

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

[6] 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

[7] 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

[8] 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, 201

[9] 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

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

[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, 20

[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, 20

[13] Awoyemi, Adebayo O. Adetunmbi and Samuel A. Oluwadare, "Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis", IEEE , 201

[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, 201

[15] Pornwatthana Wongchinsri and Werasak Kuratach, "A Survey - Data Mining Frameworks in Credit Card Processing", IEEE, 2016

[16] Yufeng Kou, .et. al. , "Survey of Fraud Detection Techniques", International Conference on Networking, Sensing & Control, pp 749 – 754, 2004

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)