International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 09, September - 2017

Comparative Analysis of Various Classification Algorithms in the Case of Fraud Detection

Ankur Rohilla
DIT University, Dehradun

***Abstract* – Credit card payment has become very popular today. Credit card is an easiest way to pay directly through your bank account. But we all know that everything have some pros as well as some cons. In the case of credit card, fraudsters are the main intruder. These intruders can access some unauthorised transactions. It is very important to prevent your account transaction from these intruders. In this paper we used three different classification algorithms (KNN, Neural network, C5.0) for fraud detection.**

Keywords: Fraud detection, Classification, Neural network, KNN, C5.0

I. INTRODUCTION

Real time credit card is most suitable mode of pension for both online as well as daily purchases. Mostly credit card fraud uses credit card details like cardholder names, credit card pin, credit card cvv etc. Credit card frauds can be perpetrated in many different ways such as simple theft, copy of card information through skimmers or gathering sensitive information through phishing (duplicate websites), counterfeit cards, application fraud, never received issue (NRI), or from unethical employees of credit card companies[1]. Credit card dealing is mainly of two types: The first one is Offline fraud. This fraud detects the stolen credit card and uses the personal details. Second is Online fraud. This fraud detected via internet, mobile phone, e-mails, web, shopping [2].

Mainly Fraud detection detects the data streams of transactions and learns the fraud's patterns. A fraud shows a small fraction of the daily transactions. Their distribution evolves over time because of seasonality [3]. Techniques used for fraud detection are Artificial intelligence & Statistical techniques. Examples of statistical techniques are computing user profiles and classification & clustering to find the patterns & associations among groups of data [4]. AI techniques used for fraud management includes Machine learning to automatically identify the fraud characteristics. These machine learning algorithms detects the pattern close to the classes, clusters or patterns of doubtful behaviour either automatically (unsupervised) or to match a given input. Expert system to encode expertise for detecting fraud in the form of rules [4]. In this paper, three classification methods are tested for their dataset in fraud detection, i.e. decision tree C5.0, neural networks and k-nearest neighbours. These are three methods that can be used to be compared in term of their predictive accuracy in fraud detection [5].

Artificial neuron network (ANN) is a computational model based on the function and design of biological neural networks. Neural network changes according to the information flows through the network that is based on the input and output. Neural network is not programmed, it works on self-learning and experience. It is use where the feature detection becomes tough.

Rest of this paper is organized as follow. The section II provides the literature study about our work. Section III presents an overview of methods used in our work. Section IV provides the description of the data set which is used in this paper. Section V tells about the tool which we used in our work. Then section VI describes the experiments and results. At last section VII tells about the conclusion our work.

II. LITERATURE REVIEW

Raghavendra Patidar.et.al present their work to detect transaction that is fraud through the use of neural network along with the genetic algorithm. To observe the behaviour of artificial neural network they used supervised learning feed forward back propagation algorithm which can be used in future analysis [6].

Aman Srivastava.et.al proposed a system that is being considered as the part of payment gateway that will check whether a transaction is fraud free or not. This system worked on the data that was provided by merchant and the data that is present in payment gateway. Predictions of the transactions were done using neural network to analyse the data of the card holder and also the transaction location and purchase details so that transaction can be detected as genuine. [7].

In [8] analysis of credit card fraud detection has been done through three classification models on two datasets. The approaches were compared according to their accuracy and elapsed time. The gave an idea of working principle of a Chebyshev FLANN in credit card fraud detection. The comparison of its performance was done with two approaches like decision tree for fraud detection and multilayer perceptron network.

Azeem Ush Shan.et.al proposed an algorithm named Simulated Annealing algorithm that was used to train the Neural Networks for detection of credit card frauds in real-time scenario. They proposed a technique that can be used for the detection of fraud in credit card. They presented all the detailed experimental results to show the effectiveness of this technique. Their technique and algorithm was beneficial for the individual users and also for the organizations in terms of cost and time efficiency [9].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 09, September - 2017

Shikha Agrawal et al. represented a survey on irregularity detection use the data mining techniques. If any user is using a fake account or stolen account and wants to perform any fraud activity then an alarm will be generated by their proposed anomaly detection system [10].

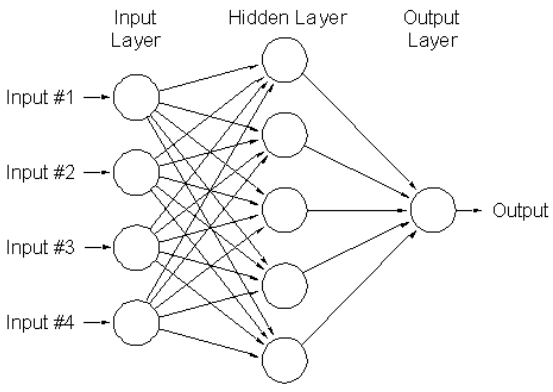
In paper [11] they studied data mining technology to anti automobile insurance fraud. They improved a method named outlier detection method that is based on nearest neighbour with pruning rules. This was applied to automobile insurance fraud and a model was created named anti insurance fraud identification model. They established a corresponding algorithm. They used association rules to mine the law of auto insurance fraud. They performed experimental analysis to verify the method. Their results show that the improved algorithm had the advantages of high accuracy, low time complexity, low impact and high recognition on the k value of algorithm.

Suvasini Panigrahi et al. proposed a novel approach for credit card fraud detection, which combines evidences from current and past behaviour. Proposed fraud detection system (FDS) consists of four components, namely, rule-based filter, Dempster–Shafer adder, transaction history database and Bayesian learner. In the rule-based component, we determine the suspicion level of each incoming transaction based on the extent of its deviation from good pattern. Dempster–Shafer's theory is used to combine multiple such evidences and an initial belief is computed [12].

Alowais et al. proposed credit card fraud detection using personalized or aggregated model. Banks suffer loss in very high amount because of credit card fraud. Here, models are created for each card holder known as personalized model and compared with aggregated model. The work was good but accuracy of personalized models is found to be better than personalized model. This can be improved by using Credit Scoring Model [13].

III. PROPOSED WORK

Artificial neural network



Artificial Neural Network functions similar as a human brain does. Human brain is a collection of neurons that are connected with each other. Similarly ANN is collection of artificial neurons (also called nodes in network) connected to each other. In 1943, Warren S.McCulloch presented ANN as a data processing unit for prediction or classification problems [14]. In 1997 Dorronsoro "et al." developed a system that can detect credit card fraud by using neural network. ANN is under the category of machine learning. The most important thing to note about ANN is that it can be used both as supervised or unsupervised method of learning [15]. ANN is extensively used in fraud detection because it has the ability to detect the hidden pattern in a large and complex data. A computer is capable to think due to neural network technology. It is same as human brain because human brain learns from past experience and is capable in making decision in everyday problem. The same technique is used for credit card fraud detection. When any consumer uses its credit card, a fixed pattern is used that is made by the way consumer uses the credit card [16]. As shown in figure, neural network is a type of train about the information of various things about the card holder such as income, information about large purchase, occupation of card holder, etc. Classification can be done using these patterns of credit card through neural network. One can classify weather a transaction is fraudulent or genuine. We had taken zero as genuine and one as fraudulent. When any unauthorized user uses the credit card, the neural network based fraud detection system checks for the pattern and compares it with the pattern of original card holder on which neural network has been trained. If both the patterns matches, ANN declares the transaction is ok means it is used by valid user.

KNN

K Nearest Neighbor algorithm is used for classification to classify the objects on the basis of distance. KNN is used in predictive analysis, image recognition, text categorization, data mining etc [17]. KNN consists of two processes : distance ranking and distance computing. There are various phases of KNN algorithm such as training phase, testing phase, classification phase. In training phase, the algorithm only stores the feature vectors and corresponding class labels. In testing phase, decisions are made by the algorithm of the basis of training data set. In classification phase, a single number is given to k, which decides how many neighbors influence the classification. The value of k can be large or small. If k=1, then it is called nearest neighbor algorithm. If value of k is large, it reduces the effect of noise on classification. Euclidean distance is commonly used distance metric in KNN algorithm [18].

C5.0

C5.0 is an algorithm that is based on C4.5 algorithm. Its idea is same as that of C4.5 of generating decision tree using recursive process. But there are many new technologies in C5.0 algorithm such as importing boosting and cost matrix. C5.0 is a top-down algorithm. . This algorithm builds a decision tree using information gain as splitting criteria. The advantage of C5.0 is that it has noticeable lower error rates [19]. Therefore it is more accurate and much faster than C4.5 and CART algorithm [20].

IV. DATA SET DESCRIPTION

The data set of credit card fraud detection from UCI Machine Learning repository is used in our work. The total

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 09, September - 2017

numbers of attributes are 15 in this data set. All the attributes are used as input to the neural network. The default payment next attribute is class identifier with value "0" indicates no fraud and value "1" indicates presence of fraud.

Predictable attributes:

Label

Value: 0 = No Fraud

Value: 1= Having Fraud

Input attributes:

1. Limit Bal (amount of given credit :It include both the individual consumer credit & his\her family)
2. Sex (Value 1=male , Value 2=female)
3. Education(1=graduate school, 2=university, 3=high school, 4=others)
4. Marital Status(1=married, 2=single, 3=others)
5. Age(year)

   Pay 0 to Pay 3=History of past payment & tracked monthly payment records.

6. Pay 0= Repayment in September.
7. Pay 2= Repayment in August.
8. Pay 3=Repayment in July.

Repayment status
-1=pay duly (one time period)
1=payment delay for 1 month
2=payment delay for 2 month
3= payment delay for 3 month

9. Bill Amt=amount of bill statement in September.
10. Bill Amt= amount of bill statement in August.
11. Bill Amt= amount of bill statement in July.
12. Pay Amt= amount paid in September.
13. Pay Amt= amount paid in August.
14. Pay Amt= amount paid in July.

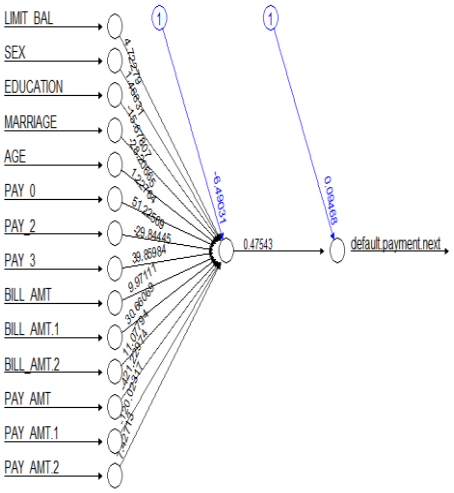
V. TOOL USED

Rstudio software is used in our work for the analysis of various algorithms. In Rstudio it is very easy to install required packages because of its user friendly behaviour. It is an open source integrated development environment (IDE) for R programming language. R language provides a wide variety of statistical and modern graph techniques. It is very easy to understand and implanting a code with this tool.

At backend NoSql is used for storing and processing the database. NoSql provides highly reliable, flexible and available data management services and play an important role in database world.

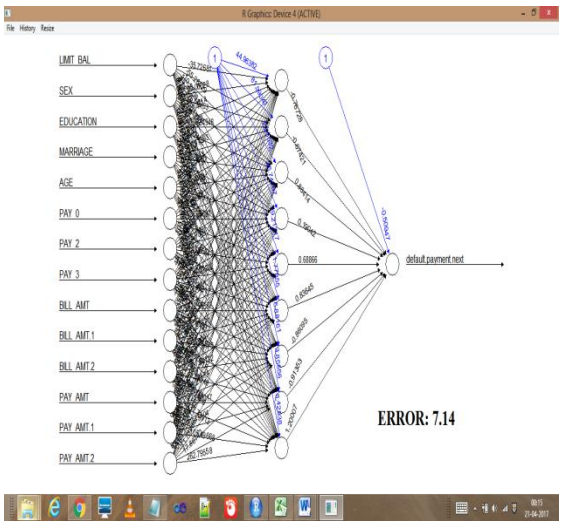
VI. EXPERIMENT AND RESULTS

In our work we used three different classification algorithms for the fraud detection. They are neural network, C5.0 and KNN and then analysed their results. First we used artificial neural network for the classification of fraud detection. All the attributes are used as an input to the neural network except ID. Neural network with single node at hidden layer gave the error rate of 25.08.



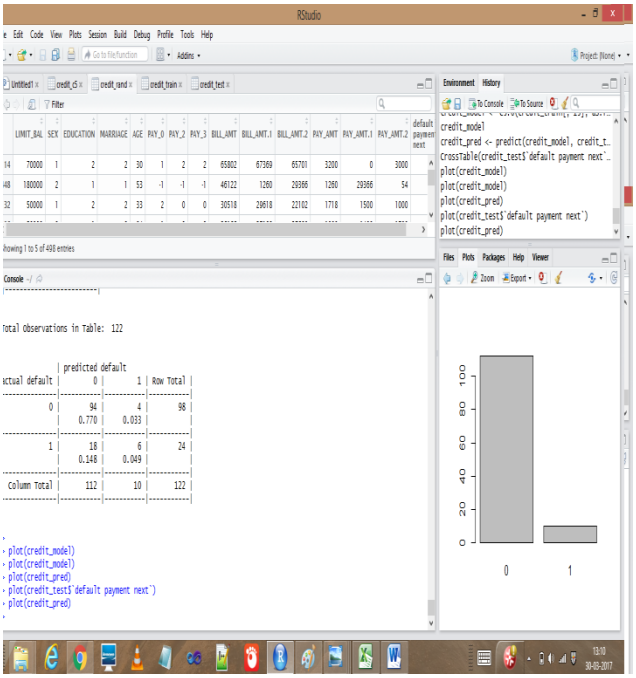
ERROR: 25.08

For improving performance we increased the number of node from 1 to 9 at hidden layer and the error rate was 7.14.



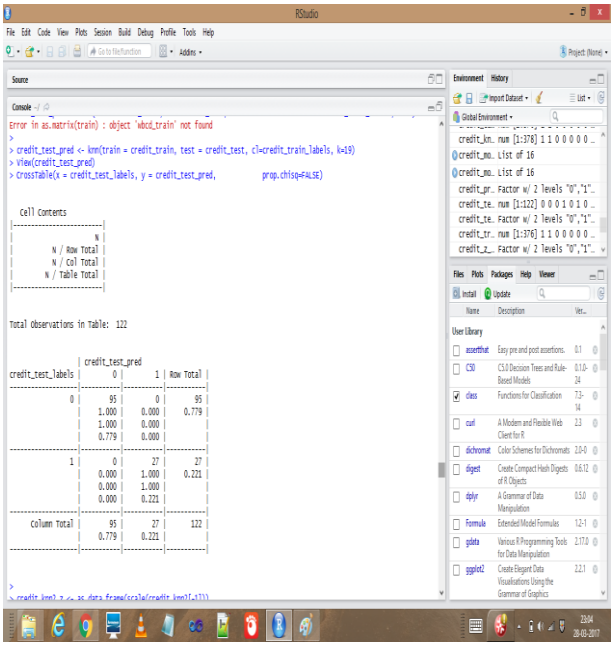
Then we used C5.0 for classification and it gives the result with higher accuracy of 91.8%. It shows an excellent result.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 09, September - 2017

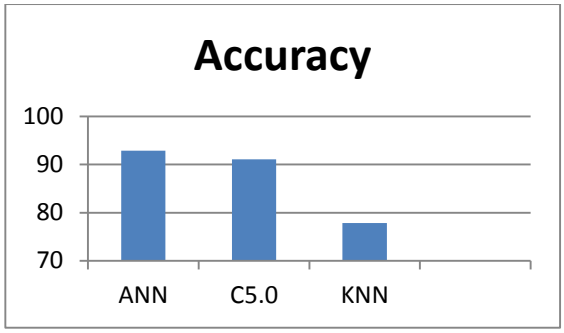


Name of Algorithm	Accuracy (%)
ANN	92.86
C5.0	91.08
KNN	77.86

At last we used KNN algorithm for classification. It also performed well towards result. It gave the good accuracy of 77.86%.



Accuracy graph of algorithms:



VII. CONCLUSION:

The objective of our work is to detect the presence of fraud in credit card database with the help of three classification algorithms and then compared the result of all the three algorithms. In our work first we used neural network and got the accuracy of 92.86%. Which is very good result. Then we used KNN for the classification of fraud. It gave the accuracy of 91.8%. This is also good. At last we used C5.0 and got the accuracy of 77.86%. It also gave a good result. So when we compared the result of these three algorithms we can see that the neural network gives better result for the classification of fraud. So it indicates that neural network can give better result than other algorithms in the case of classification problem.

REFERENCES:

[1] Duman, Ekrem, Ayse Buyukkaya, and Ilker Elikucuk. "A novel and successful credit card fraud detection system implemented in a turkish bank." Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on. IEEE, 2013.

[2] Gaikwad, Jyoti R., et al. "Credit Card Fraud Detection using Decision Tree Induction Algorithm." International Journal of Innovative Technology and Exploring Engineering (IJITEE) 4 (2014).

[3] Dal Pozzolo, Andrea, et al. "Credit card fraud detection and concept-drift adaptation with delayed supervised information." Neural Networks (IJCNN), 2015 International Joint Conference on. IEEE, 2015.

[4] Patidar, Raghavendra, and Lokesh Sharma. "Credit card fraud detection using neural network." International Journal of Soft Computing and Engineering (IJSCE) 1.32-38 (2011).

[5] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain", Psychological review, 65(6):386, 1958.

[6] Patidar, Raghavendra, and Lokesh Sharma. "Cr edit card fraud detection using neural network." International Journal of Soft Computing and Engineering (IJSCE) 1.32-38 (2011).

[7] Srivastava, Aman, et al. "Credit card fraud detection at merchant side using neural networks." Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on. IEEE, 2016.

[8] Mishra, Mukesh Kumar, and Rajashree Dash. "A Comparative Study of Chebyshev Functional Link Artificial Neural Network, Multi-layer Perceptron and Decision Tree for Credit Card Fraud Detection." Information Technology (ICIT), 2014 International Conference on. IEEE, 2014.

[9] Khan, Azeem Ush Shan, Nadeem Akhtar, and Mohammad Naved Qureshi. "Real-time credit-card fraud detection using artificial neural network tuned by simulated annealing algorithm." Proceedings of International Conference on

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 09, September - 2017

Recent Trends in Information, Telecommunication and Computing, ITC. 2014.

[10] Shikha Agrawal, Jitendra Agrawal, " Survey on Anomaly Detection using Data Mining Techniques": 19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems, Procedia Computer Science 60, page no. 708 – 713,Year of publication 2015

[11] Yan, Chun, and Yaqi Li. "The Identification Algorithm and Model Construction of Automobile Insurance Fraud Based on Data Mining." Instrumentation and Measurement, Computer, Communication and Control (IMCCC), 2015 Fifth International Conference on. IEEE, 2015.

[12] Panigrahi, Suvasini, et al. "Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning." Information Fusion 10.4 (2009): 354-363.

[13] Chan, Philip K., and Salvatore J. Stolfo. "Toward Scalable Learning with Non-Uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection." KDD. Vol. 1998. 1998.

[14] Khan, Azeem Ush Shan, Nadeem Akhtar, and Mohammad Naved Qureshi. "Real-time credit-card fraud detection using artificial neural network tuned by simulated annealing algorithm." Proceedings of International Conference on Recent Trends in Information, Telecommunication and Computing, ITC. 2014.

[15] Muhammad Waseem Khan, Imran Qureshi, "Neural Network based Software Effort Estimation: A Survey", Int. J. Advanced Networking and Applications Volume: 05, Issue: 04, Pages: 1990-1995 (2014) ISSN: 0975-0290

[16] Patidar, Raghavendra, and Lokesh Sharma. "Credit card fraud detection using neural network." International Journal of Soft Computing and Engineering (IJSCE) 1.32-38 (2011).

[17] Peng Y, Kou G, Shi Y, Chen ZX (2008) A descriptive framework for the field of data mining and knowledge discovery. Int J Inf Technol Decis Mak 7(4):639–682

[18] George, Treesa, Sumi P. Potty, and Sneha Jose. "Smile detection from still images using KNN algorithm." Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014 International Conference on. IEEE, 2014.

[19] http://rulequest.com/see5-comparison.html.

[20] Pashaei, Elnaz, Mustafa Ozen, and Nizamettin Aydin. "Improving medical diagnosis reliability using Boosted C5. 0 decision tree empowered by Particle Swarm Optimization." Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE. IEEE, 2015.