ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2014

FRAUD DETECTION IN MOBILE TELECOMMUNICATION

Fayemiwo Michael Adebisi[1]* and Olasoji Babatunde O[1].

[1]Department of Mathematical Sciences, College of Natural and Applied Science, Oduduwa University, Ipetumodu,

P.M.B. 5533, Ile-Ife, Osun State, Nigeria.

*Corresponding Author: mfayemiwo@gmail.com

Abstract: Fraud has been very common in our society, and it affects private enterprises as well as public entities. However, in recent years, the development of new technologies has also provided criminals more sophisticated way to commit fraud and it therefore requires more advanced techniques to detect and prevent such events. The types of fraud in Telecommunication industry includes: Subscription Fraud, Clip on Fraud, Call Forwarding, Cloning Fraud, Roaming Fraud, and Calling Card. Thus, detection and prevention of these frauds is one of the main objectives of the telecommunication industry.

In this research, we developed a model that detects fraud in Telecommunication sector in which a random rough subspace based neural network ensemble method was employed in the development of the model to detect subscription fraud in mobile telecoms.

This study therefore presents the development of patterns that illustrate the customers' subscription's behaviour focusing on the identification of non-payment events. This information interrelated with other features produces the rules that lead to the predictions as earlier as possible to prevent the revenue loss for the company by deployment of the appropriate actions.

Keywords: Fraud detection, Telecommunication industry, Neural Network Ensemble, Data Mining.

I. INTRODUCTION

Fraud has been very common in our society, and affects private enterprises as well as public entities. However, in recent years, the development of new technologies has also provided criminals more sophisticated way to commit fraud and has required more advanced techniques to detect and prevent such events [6].

Telecommunication Company worldwide suffers from customers who use the provided services without paying. The estimated losses amount to several billions of dollars in uncollectible debt per day [3]. Even though this is a small percentage comparing to the Telecom Operators' revenue, it is still a significant loss.

Detection and prevention of frauds is one of the main objectives of the telecommunication industry. However, the volume of data being generated nowadays is increasing at phenomenal rate. So, extracting useful knowledge from such data collections is an important and challenging issue. In order to build such a non-trivial model, many researches were carried out on the feasibility of using the Data Mining (DM) techniques which comes from the need of analyzing high volumes of data collected by the telecommunication companies (customer data, unbilled calls, etc.) and related to different kinds of transactions between the company and its customers. Other techniques that have also been used include: Bayesian Network Technique, Distance-Based Method, Time Series Analysis, Rule-Base Approach to Fraud Detection, Neural Network Based Approach, Neural networks with supervised learning, B-Number analysis tool, Multiagent System, Agent-Based Knowledge Discovery in Data, User Profiling, Ensemble Neural Networks, etc. Our main goal is to reveal the behavioural patterns of the customers' subscription, that is, customers who obtain a telecommunication account with no intention of paying for the bill. This study therefore presents the development of patterns that illustrate the customers' subscriptions behaviour focusing on the identification of non-payment events. This information interrelated with other features produces the rules that lead to the predictions as earlier as possible to prevent the revenue loss for the company by deployment of the appropriate actions.

II. Literature Review

There are currently two types of mobile phone communication available. One operates on analogue networks, in which speech is sent as a continuous wave signal, and the other uses digital transmission, in which the traffic is in the form of a rapid stream of binary pulses. With both types, at the beginning of transmission an identifying signal is transmitted with the voice data to enable access onto the telecommunications network. With digital systems this identification signal is encrypted. The identifying signal contains information about the mobile phone account, allowing the network equipment to keep track of the call, maintain the connection between the two parties whilst the users are mobile, and also to perform billing activities. Analogue systems are more vulnerable to fraud as it is not possible to encrypt the identification codes that are transmitted along with the voice data [1].

Reference [2] presented a rule-based approach to detect anomalous telephone calls. The method described used subscriber usage CDR (call detail record) data sampled over two observation periods: study period and test period. The study period contains call records of customers' non-anomalous behaviour. Customers are first grouped according to their similar usage behaviour (like, average number of local calls per week, etc.). Reference [2] developed a probabilistic model to describe their usage for customers in each group. Next, maximum likelihood estimation (MLE) was used to estimate the parameters of the calling behaviour. Then the thresholds were determined by calculating acceptable change within a group. MLE was used on the data in the test period to estimate the parameters of the calling behaviour.

Reference [4] detected a fraudulent transaction through the neural network along with the genetic algorithm. Genetic algorithm was used for making the decision about the network topology, number of hidden layers, and number of nodes that were used in the design of neural network for the problem of credit card fraud detection.

In this work we used a neural network ensemble to develop a model to detect subscription fraud in mobile telecoms. Neural network ensemble is a learning paradigm where many neural networks are jointly used to solve a problem [7].

III. Categories of subscriber

Fraud cases would generally be detected online triggered by traffic measures by the commercial fraud detection system, and confirmed later on as such during the billing process. In order to generate a database of known fraudulent/legitimate cases, it was necessary to formalize the definition of subscribers' categories. Consequently the following four categories of subscribers were defined [5]:

A. Subscription fraudulent: Most of the users in this category do not pay their bills at all, suspicious behaviour in long distance calls within 6 months after the installation date.
B. Otherwise fraudulent: Subscribers for more than a year who present a sudden change but if they do, the debt/payment ratio is very high. The line is typically blocked due to having two or more unpaid bills. This category includes new customers that have never paid their bills but whose monthly expenditures are similar to average residential lines.
C. Normal: Customers with their bills up to date or at most a single unpaid bill for less in their calling behaviour, generating an abnormal rise in their newer billing accounts.
D. Insolvent: Subscribers with a total debt of less than 10 times their monthly payments, than 30 days after the due date.

However, in this research, we focused on the first category - subscription fraudulent.

IV. Data Description

In a telecommunications company the operator keeps record of every event processed by the system. These events are recorded in CDRs (Call Detail Records), generated automatically and are used for billing purposes. Each CDR has information regarding a set of events, voice calls or SMSs, for example. Typically, the CDR is a text file containing information structured by a predefined set of ordered fields separated by a predefined character. Each line of the CDR file is an event processed in the operators system. The structure of the CDR (the number of fields, the order of the fields and the separator character) is defined by the telecommunications company, so the CDR structure varies from operator to operator. However, there is a set of fields that, due to their importance, for billing and rating purposes, are usually common to all CDR structures:

A. A Number - identifies the originator of the event;
B. B Number - identifies the receiver of the event;
C. Event Date - the date the event started;

D. Event Type - identifies the type of the event, for example: 1 (Voice), 2 (SMS), 3 (MMS), 4 (Data);

E. Event Amount - measure of the event, for example, in a voice call the event amount is 124 seconds, in a SMS the event amount is 45 characters;

F. Cell ID - identifies the network cell that processed the event.

The information contained in the CDRs will be the input for all the future work. The study of the contents of CDRs is not a novelty. They were first created with billing purpose, but know they are used with different purposes of great importance to the operators, for instance, discovering user communities.

V. Development of Neural Network Ensemble

A random rough subspace based neural network ensemble method is employed in the development of the model to detect subscription fraud in mobile telecoms. The method involves creating a number of training subsets from the original training set.

For this study, four different training subsets were used to create four classifiers. A predicted target is obtained by averaging the outputs of the four classifiers. Fig 1 and 2 show the block diagrams of the ANN ensemble.

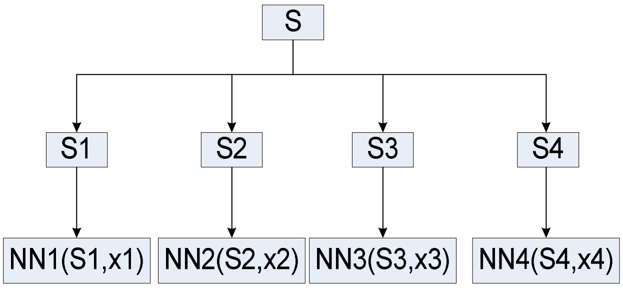

Fig 1: ANN Ensemble Training

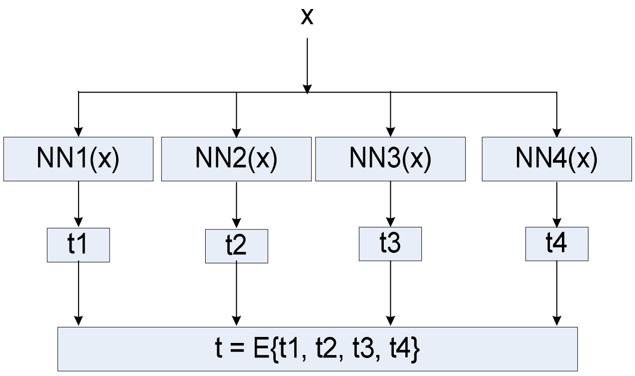

Fig 2: ANN Ensemble Testing

VI. System Model for the Subscription Fraud Detection

The subscription fraud detection system model is achieved by using sequences of call detail records (CDRs), which contain the details of each post-paid users on the network. The information produced for billing also contains usage behaviour information valuable for fraud detection. The proposed system model is presented in Fig 3.
The CDR consists of the following variables: National ID (9 digit number), Address (LGA/CDA number), Age, Income (monthly salary), Phone Number (Customer number given by the network provider), Gender (Male=1, Female=0), Marital (Married=1; Single=0), Retire (Yes=1, No=0), Phone blocked flag (blocked=1; active=0) (PBF), Number of days with unpaid bills (in Days) (NDU), Line account balance (in Naira) (LBA), Maximum debt with international carriers (in Naira) (MDC), Time elapsed between installation date and blocking date (in Days) (TBI), Debt/payment ratio (in Percent) (DPR).

VII. FRAUD PREDICTION (DETECTION)

Subscription fraud prevention is achieved by using customers' commercial antecedents which have been modelled in the ANN object. Some of the assumptions for predicting a fraudulent subscription are as follows:

- Applicant's ID is similar to that of a fraudster
- Applicant's contact phone number is similar to a fraudster
- Applicant's address, age, gender and marital are similar to a fraudster

A fraudster often times uses the first line he used in committing fraud as the contact phone number during application for a new line. A particular number is supplied several times.

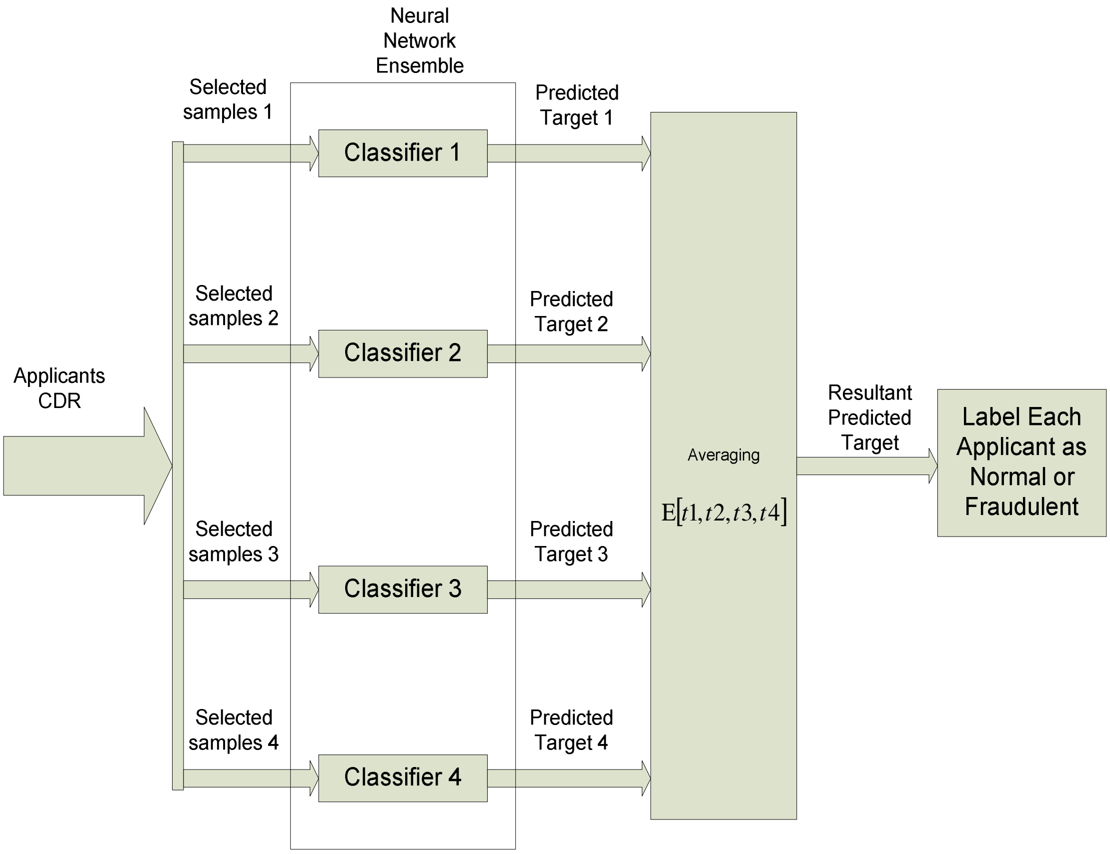

Fig3: Subscription Fraud Detection Model

VIII. TESTING SET

The developed ANN model is tested with a different 100 samples that are not part of the training set used to create the ANN model. Out of the testing samples, 80 samples are fraudulent applications while 20 samples are normal applications. The inputs to the ANN model are CDR variables 1 to 8. The model is used to simulate the inputs to the predict type of application (Fraud=1; Normal=0). The predicted ANN flags are compared with the actual flags to compute:

- True-Positives (TP): fraud samples classified as fraud
- False-Negatives (FN): fraud samples classified as normal
- True-Negatives (TN): normal samples classified as normal
- False-Positives (FP): normal samples classified as fraud

The detection rates are computed as follows:

$$\mathrm{TP}(\%) = \frac{\text{fraud samples classified as fraud}}{\text{Total number of fraud samples}} \times 100$$

$$\mathrm{TN}(\%) = \frac{\text{normal samples classified as normal}}{\text{Total number of normal samples}} \times 100$$

$$\mathrm{FP}(\%) = \frac{\text{normal samples classified as fraud}}{\text{Total number of normal samples}} \times 100$$

$$\mathrm{FN}(\%) = \frac{\text{fraud samples classified as normal}}{\text{Total number of fraud samples}} \times 100$$

The accuracy of the model is computed as:

$$\text{Accuracy} = \frac{\text{Number of correct prediction}}{\text{Total number of test samples}} \times 100$$

IX. PERFORMANCE OF THE NEURAL NETWORK ENSEMBLE TRAINING

Four different NN classifiers were developed using four different data subsets; and the ensemble classifier was obtained as the mean of all the four classifiers. For the validation of the trained NN classifiers, the error performance metrics considered are sum squared error (SSE) and root mean squared error (RMSE). The values of the SSE, MSE and RMSE obtained reveal how well the NN model has been able to learn the training data. In other words, the lower SSE, MSE and RMSE the better the performance of the NN model will be.

Fig 4 shows the comparison of the SSE obtained from the training of the four NN classifiers. The SSE values of 4.0, 24.0, 8.0 and 20.0 were for NN classifier 1, NN classifier 2 NN classifier 3 and NN classifier 4 respectively. The result shows that the NN classifier 1 gives the best SSE performance while NN classifier 2 gives the worst SSE performance. Fig 5 presents the RMSE results of 0.2357, 0.5774, 0.3333 and 0.5279 for NN classifier 1, NN classifier 2, NN classifier 3 and NN classifier 4 respectively. The results show that NN classifier 1 gives the best RMSE performance while NN classifier 2 gives the worst RMSE performance.

X. PERFORMANCE OF THE NEURAL NETWORK ENSEMBLE ON TESTING SET

The performances of the developed NN classifiers were tested with a new data set that was not part of the training set. The testing set consists of the demographic data of some applicants for the post-paid mobile lines. Some of the applicants with the intention of committing fraud could supply former number used to commit fraud, the same address, bank account details or identity card similar to the one supplied in the previous fraud. The testing data have been labelled to distinguish the potential fraud applicants from the normal applicants. Each of the NN classifiers was used to simulate the testing data to predict the label for all the applicants in the data. The predicted labels of the NN ensemble classifier are obtained by averaging the labels obtained from the four NN classifiers. The number of correct classifications and wrong classifications are computed for each of the classifiers in terms of TP, FN, TN and FP.

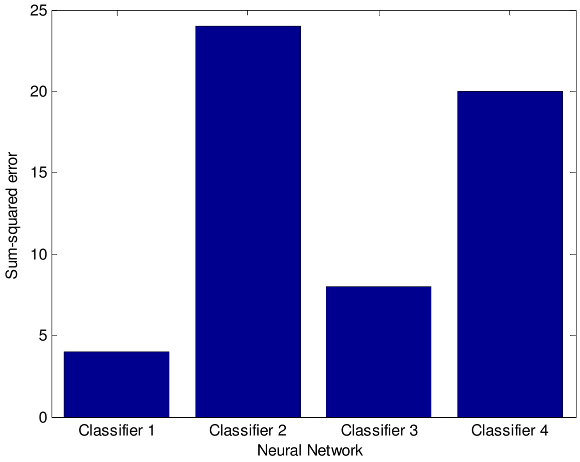

Fig 4:Sum squared error of the Neural Network Classifiers

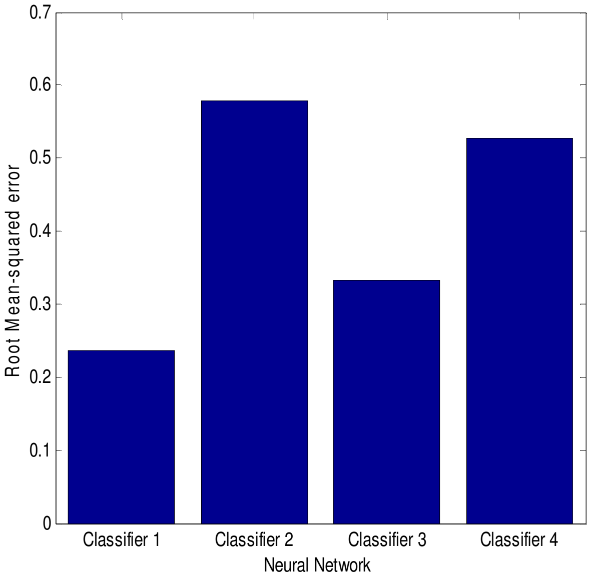

Fig 5: Root mean squared error of the Neural Network Classifiers

XI. Result

The detection errors of the NN classifiers showed that NN classifier 1 gave 7 wrong classifications, NN classifier 2 gave 12 wrong classifications, NN classifier 3 gave 1 wrong classification and NN classifier 4 gave 7 wrong classifications. The detection performances of the NN classifier and NN ensemble are presented in Table 1 and Fig 6 to 10 in terms of percentage TP, FN, TN, FP and Accuracy. The results show that the NN classifier 3 gives the best detection performance with 99% accuracy followed by the NN ensemble with 97% accuracy. This shows that the NN ensemble outperforms three out of the four NN classifiers, which helps to improve the efficiency of the proposed model for fraud prevention.

TABLE I
Detection Performance of the NN Classifiers

	NN Classifier 1	NN Classifier 2	NN Classifier 3	NN Classifier 4	NN Ensemble
fraud samples classified as fraud (TP)	92.500000%	86.250000%	98.750000%	91.250000%	95.000000%
fraud samples classified as normal (FN)	1.000000%	0.000000%	0.000000%	0.000000%	0.000000%
normal samples classified as normal (TN)	95.000000%	90.000000%	100.000000%	100.000000%	95.454545%
normal samples classified as fraud (FP)	6.000000%	11.000000%	1.000000%	7.000000%	3.000000%
Accuracy	93.000000%	87.000000%	99.000000%	93.000000%	97.000000%

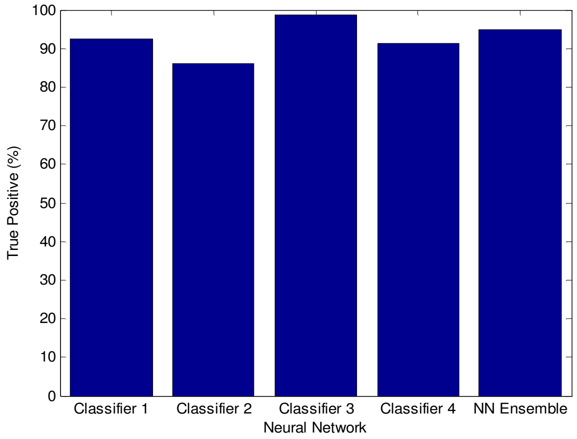

Fig 6: Percentage True Positives of the Neural Network Classifiers

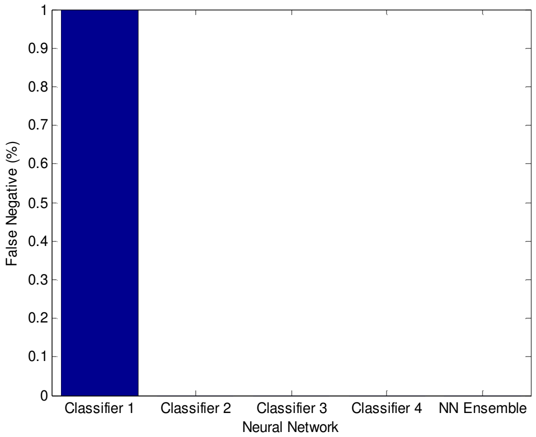

Fig 7: Percentage False Negatives of the Neural Network Classifiers

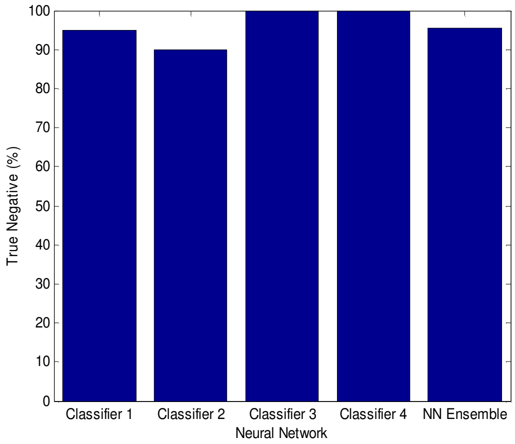

Fig 8: Percentage True Negatives of the Neural Network Classifiers

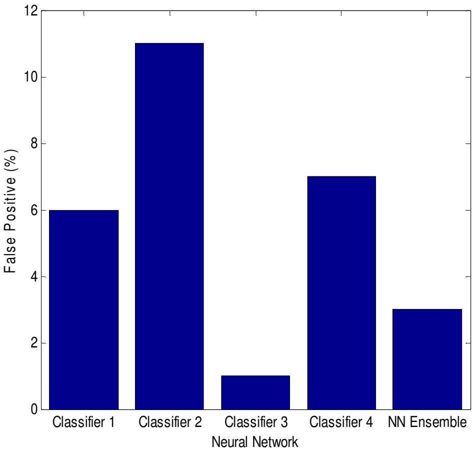

Fig 9: Percentage False Positives of the Neural Network Classifiers

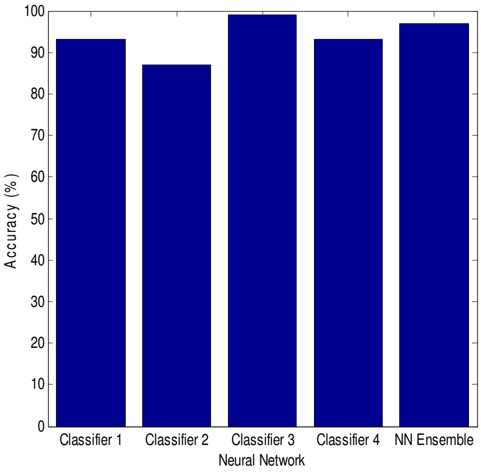

Fig 10: Percentage Accuracies of the Neural Network Classifiers

XII. CONCLUSION

In this research, we developed a model that detects fraud in Telecommunication sector in which a random rough subspace based neural network ensemble method was employed in the development of the model to detect subscription fraud in mobile telecoms.

This study presented the development of patterns that illustrate the customers' subscription's behaviour focusing on the identification of non-payment events. This information interrelated with other features produces the rules that lead to the predictions as earlier as possible to prevent the revenue loss for the company by deployment of the appropriate actions.

ACKNOWLEDGMENT

We would like to thank the people of Intelligent Systems Research Group, in the Department of Mathematical Sciences, College of Natural and Applied Science, Oduduwa University, Ipetumodu, Ile-Ife, Nigeria, for the invaluable support towards the successful completion of this research.

REFERENCES

[1] Barson, P., Field, S., Davey, N., McAskie, G., and Frank, R. "The Detection of Fraud in Mobile Phone Networks". School of Information Sciences, University of Hertfordshire, Hatfield, Herts. AL10 9AB. BNR Europe Limited, London Road, Harlow, Essex, UK, CM17 9NA., 1999.

[2] Gopal, R. K. and Meher, S. K. "A Rule-based Approach for Anomaly Detection in Subscriber Usage Pattern". International Journal of Engineering and Applied Sciences. 3:7., 2007.

[3] Moudani, W. and Chakik, F. "Fraud Detection in Mobile Telecommunication". Notes on Software Engineering, Vol. 1, No. 1., 2013.

[4] Patidar, R. and Sharma, L. "Credit Card Fraud Detection Using Neural Network". International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-NCAI, 2011.

[5] Rajani, S. and Padmavathamma, M.,"A Model for Rule Based Fraud Detection in Telecommunications". International Journal of Engineering Research and Technology (IJERT), Vol. 1. ISSN: 2278-0181., 2012.

[6] Serrano, A. M. R., da Costa, J. P. C. L., Cardonha, C. H., Fernandes, A. A., and de Sousa Jnior, R. T. "Neural Network Predictor for Fraud Detection: A Study Case for the Federal Patrimony Department". IBM Research Sao Paulo, BRAZIL, (DOI: 10.5769/C2012010)., 2010.

[7] Zhou, Z.-H., Wu, J., and Tang, W."Ensembling Neural Networks: Many Could Be Better Than All. Artificial Intelligence" Elsevier, 137(1-2):239 - 263., 2002.