Informative Feature Trained Classification System For Credit Card Fraud Detection

Akshita Gautam, Suraj Yadav

Abstract : Credit Card Fraud is a broad word used as a fraudulent source of funds in a transaction for theft and fraud committed using or with a payment card, for example a credit or debit card. Moreover the traditional systems lacks in various factors such as consuming complexity, process delay etc. in detecting the credit card fraud. In this paper, a novel approach for credit card detection is developed with the amalgamation of neural network and fuzzy inference system. Infinite Feature selection is used to reduce the complexity of the model and selection of informatics features from available dataset. Simulation is done in MATLAB software and analyzed in term of quality factors as accuracy, precision and recall. The simulation results shows that proposed approach provides effective results in terms of all performance factors when compared with existing approaches.

Index Terms : Credit card fraud, Data Mining, Fuzzy interface system, Feature selection, MATLAB simulation.

1. INTRODUCTION

WITHIN this framework of the card issuer industry, the actions carried out by unsought elements to collect undeserved rewards are referred to as a fraud that leads to financial loss to the economical services industry. Here, the author demonstrates the efforts or attempts made by fraudsters to use embezzled identity information and credit card to steal money, goods or services. It becomes easy and popular for people to transact and spend money by using technology as card issuers has a constant aim to increase their operations by launching aggressive campaigns so as to achieve larger parts of the market share. Although, the facility of spending by using technology is expanding, but unfortunately, it also offered a podium for the fraudulent activities which are also increasing rapidly. These activities have, therefore, piercingly augmented since the 1990's, and the growth in credit card fraud is almost costing literally billions of dollars to the card issuer industry per annum. Thus, the industry gets encouraged from this concern to use increasingly more effective mechanisms to triumph over the fraud in the credit card industry.In a simple way, fraud detection is an act of identifying falsified/fake behavior at the instant it occurs [2]. Fraud detection and fraud prevention are two different activities. In fraud prevention, the deployment of strategies is made in order to make it progressively more complicated for people to commit fraud in the first place. One might think that prevention is better than cure but this principle would also overcome here; but it is a fact that prevention is not always helpful enough in order to control the high fraud rate that creates nuisance in the credit card industry; therefore, it cues the use of fraud detection mechanisms.Fraud detection can be taken as a prevention strategy in the future if processing power is enhanced. Credit cards can be easily targeted for fraud due to the fact of using credit cards in abandoned environments. A huge amount of money can be embezzled in a moment and fraudster could not be even traced. The card issuer should get a chance to limit the damage by identifying fraud the moment possible after it occurs. After the transaction is flagged like a possible fraud, the confirmation can be taken from the card holder to know whether the transaction was legitimate or not and the card can be blocked if necessary. The processes of fraud investigation and the chargeback are expensive and resources get much stress in these processes. The detection of fraud occurring quickly is possible but if it involved huge amount such as sometimes thousands of transactions per second, makes it complicated to detect in real-time or even infeasible..

2 LITERATURE

Many interesting applications seek to solve business problems complexity. The establishment of technology has introduced Artificial Neural Networks (ANN) which is capable of simplifying the programming effort and algorithm design used in conventional processes. Systems built on Fuzzy Logic (FL) have the tendency to cope with uncertainty in the environment in which any business flourishes. ANN and FL have been successfully applied in many application fields whether individually or harmonizing strength of each other. From researchers working in different domains, great inclination has been observed in combined neuro-fuzzy approach. Arora, N. & Saini, J.R. (2014) suggested a comprehensive study of existing work in different areas using soft computing methodologies which particularly focus on neural networks and fuzzy logic. Arora & Saini (2013) proposed an ANFIS model that utilized time series in order o predict bankruptcy. The author suggested that the financial status of company affects all the stakeholders. Some idea of current financial stand of the company can be obtained from Scrutiny of financial ratios, but it is not capable of revealing its status in future. In order to calculate the corresponding Z scores, a set of historical data in the form of financial ratios was considered by the proposed model and the Z score is predicted for future time period. This can foresee whether the company would be bankrupted in coming years or not. This work was further polished in another model anticipated by Arora & Saini (2014) which considered financial distressed companies for prediction of status of bankruptcy in future. The anticipated model used Independent Component Analysis (ICA) so as to select input parameters which should not be dependent on each other. 10 financial ratios were comprised in the initial dataset while ICA returned 5 ratios. The selected ratios created a dataset which produced actual data to train Fuzzy Support Vector Machines

- *Akshita Gautam is currently pursuing masters degree in computer science engineering in Jagannath University, Jaipur, India. She is specialized in Data Mining and neural networks and highly interested in credit card fraud detection approaches. E-mail: gautam.akshita@gmail.com*
- *Suraj Yadav is an assistant professor in Computer Science deparment in Jagannath University, Jaipur, India. He is specialized in Data Mining and has high interest in Evolutionary Computation and Mimetic. E-mail: er.surajyadav@gmail.com*

(FSVM). Tuning was executed on various factors like cost, kernel function and curvature of curve in order to determine the correct choice of values. The analysis performed by comparing of FSVM with ANFIS resulted in enhanced clustering potential of FSVM.

3 NEURO FUZZY INFERENCE SYSTEM

Neural network based fraud detection methods are most popular. An interconnected group of artificial neurons is contained in artificial neural network. The functions of the brain especially associative memory and pattern recognition are responsible to motivate the principle of neural network. In neural network similar patterns are identified, future values or events are predicted which are based upon the associative memory of the patterns. Neural network has widely useful in classification and clustering. It has main advantage over other techniques: the neural network model is capable of learning from the past and thus, results can be improved as time passes. Also, the rules can be extracted in this model. Moreover the future activity can be predicted on the basis of the present situation. By utilizing neural networks, efficiently, it will become easy for the banks to detect fraudulent use of a card in faster and more efficient way. Amongst the study of credit card fraud that has been reported, it was observed that most have paid attention on using neural networks. Nonlinear statistical data modeling tools are neural networks. Through these networks, complicated input-output relations can be easily depicted. A neuro-fuzzy system uses a learning algorithm that is derived from neural network theory. This system determine parameters : Fuzzy set and fuzzy rules by processing data samples. This is a fuzzy based model developed in the theory of neural networks using a learning algorithm. The heuristic learning process is based on local data and only creates local changes in the underlying fuzzy structure. This can be regarded as a neural feedback network of three layers. The first layer is input variable, the center layer is hidden and it represents fuzzy rules and the third layer is output variable. Fuzzy sets are represented as connection weights. A fuzzy system like this should not be depicted in order to use a learning algorithm. However, it can be useful, as it constitutes the information flow of input processing within the model. In order to learn the processes of neuro-fuzzy system, the semantic characteristics of the fuzzy system must be takes into account. This contributes to the modifications in system parameter being restricted. Amalgamation of fuzzy logic and neural networks results in Neuro-fuzzy technology (Jin, 2003, pp. 111-140). Each fuzzy logic and neural network has its own set of advantages and faults, and most approaches used to merge these two technologies aim at using the strengths of each technique to address other faults. Neural networks can self-learn, classify and combine inputs with outputs. Neural networks can also become an approximator for universal functions. Given enough information about an unknown continuous function, such as its inputs and outputs, it can be approximated by training the neural network approximate it. The disadvantages this network are that they are not guaranteed to converge, that is to be trained properly, and after their training they cannot give any information about why they take a specific course of action when given a particular input.

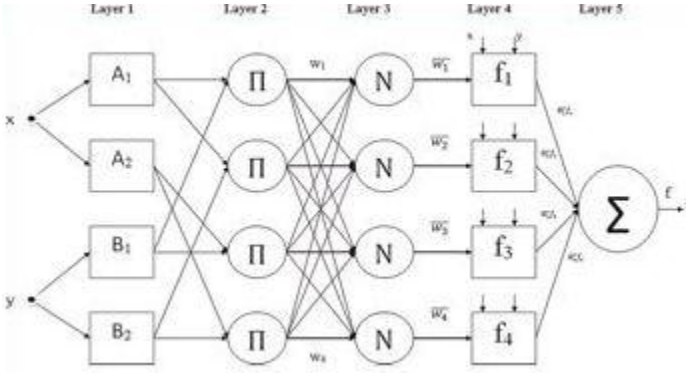

Fig. 1. *The ANFIS structure that has two inputs and one output*

The ANFIS structure having two inputs (x and y) and one output (f) is represented in fig.1. There are five layers in the ANFIS structure. The Fuzzy logic Inference systems can give human readable and understandable information about why a specific course of action was taken because it is governed by a series of IF THEN rules. Fuzzy logic method can adapt in a way that their fuzzy set's rules associated with those rules can be changed to meet some criteria. However fuzzy logic systems lack the capability for self-learning, and must be modified by an external entity. Another salient feature of fuzzy logic systems is that they are, like artificial neural networks, capable of acting as universal approximators.The common feature of being able to act as a universal approximator is the basis of most attempts to merge these two technologies. Not only it can be used to approximate a function but it can also be used by both neural networks, and fuzzy logic systems to approximate each other as well. (Pal et al., 1999, pp. 66)Universal approximation is the ability of a system to replicate a function to some degree. Both neural networks and fuzzy logic systems do this by using a non-mathematical model of the system (Jang et al., 1997, pp. 238; Pal et al., 1999, pp. 19). The term approximate is used as the model does not have to match the simulated function exactly, although it is sometime possible to do so if enough information about the function is available. In most cases it is not necessary or even desirable to perfectly simulate a function as this takes time and resources that may not be available and close is often good enough.

4 PROBLEM FORMULATION

Credit card fraud can be delineated as "Unauthorized account activity performed by a person who is not the genuine account holder. Also it is an incident in which preventive actions are taken to stop the abuse in progress and incorporate risk management practices to protect against similar actions in the future".Numerous researches have been performed in this field for detection and prevention of the credit card frauds. This section contains a brief overview to the few of the researches and thus studied that the credit card fraud detection has become a prominent topic among the researchers. In a traditional work [1], an approach was developed by the author so as to detect the credit card frauds and the whale algorithm was used for the optimization of the neural network credit card. Initially, input of the neural network algorithm is taken as the credit card fraud feature parameters, and the network error is used like the fitness function. Then, initial network weight is optimized by using the whale algorithm. Following is the list of loopholes in traditional work:

1) There was an increase in count of iterations which resulted in the high processing time caused by the implementation of BP neural network for data classification.

TABLE 1
OUTPUT OF PARAMETERS WITH RESPECT TO DIFFERENT TRAINING SAMPLES

Sr. no	Training Samples	Accuracy	Precision	Recall
1	10	94.25802	72.16495	85.88957
2	15	97.91201	91.92547	90.79755
3	20	98.11133	83.42246	95.70552
4	25	99.00596	92.81437	95.09202
5	30	98.50858	87.27273	88.34356
6	35	99.36614	94.5122	95.09202
7	40	98.71757	82.7027	93.29268

2) Due to large number of iterations, a lot of time was consumed by the system for data training process.
3) The complexity level rises in system by implementing whale network and BP neural network.
4) There is delay in data processing because of the high processing time.

5 PROPOSED WORK

Medical image processing is one of the prominent topics that From the above discussion, it is observed that traditional credit card fraud detection system lacks at different factors such as high time consuming, complexity, delay in processing etc. The performance of any system get influenced by the all those factors. Therefore, in this study, the author decided to develop an approach in this direction. In order to develop the new credit card fraud detection approach, the following factors are taken into account:

1) The amalgamation of fuzzy inference system and neural network is decided for implementing rather than traditional BP neural network.
2) The main aim of anticipated work is to reduce the model of large number of iterations through implementation of the neuro-fuzzy optimization system.
3) The benefit of using neuro-fuzzy optimization system is that it is a system based on rules and hence it eradicates the requirement of iterations to evaluate the fitness function. Together with this, the process of data training has become easier than the traditional one, because of defined rules
4) The complete data had been used for training purpose in traditional work resulting in the complexity during training process. Therefore, the proposed work put focus on training the data on the basis of the feature selection. The process provides benefit of easily understating, data categorization, and less complex training process..

6 RESULTS ANALYSIS

The simulation of the proposed work has been done to analyze its efficiency. Credit card fraud detection is done by implementing feature extraction, feature selection and their classification by using LDA, infinite feature selection technique and neuro-fuzzy logic respectively. The results were obtained in terms of Accuracy, Precision and Recall. The results were analyzed with respect to different number of data sets and the comparison is made with the outcomes of existing techniques. The evaluation took place on the basis of different parameters such as type of cluster used, membership functions, inputs and output.

The proposed system was analyzed on behalf of the data set for credit card fraud detection.
Evaluation on the basis of parameters:

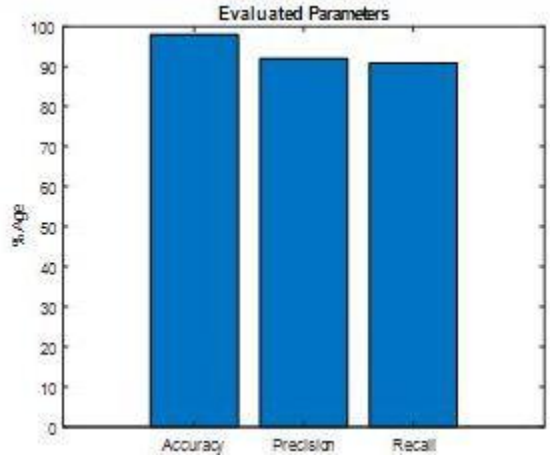

Fig. 2. *Evaluation based on parameters*

It is stated above that three parameters are considered for the evaluation purpose. When this system is analyzed, the simulation results obtained are presented in the form of graph. The graph is shown in fig. 2. The bar graph illustrated the accuracy, precision and recall in terms of percentage. The values of accuracy, precision and recall are 98%, 90% and 88% respectively. This analysis is performed only by considering a single data set.Further, the proposed system is analyzed separately in terms of above mentioned three parameters. These parameters are analyzed with respect to different number of data sets (training samples). The output of this evaluation is shown in the tabular form in table 1.
There are total 40 training samples and the output is obtained after every 5 samples.The graphs obtained the results in which x-axis is representing the training samples and y axis shows the different parameters.

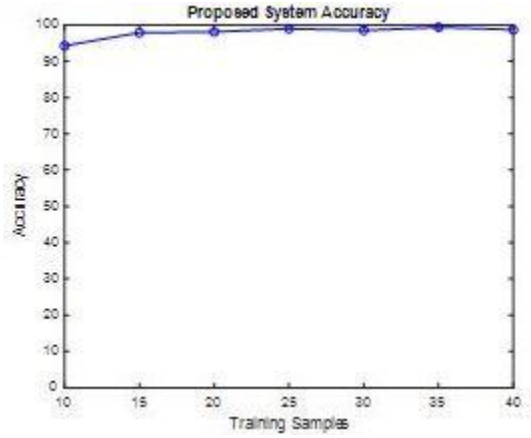

Fig. 3. *Accuracy of proposed system*

The evaluation of the accuracy of proposed system is shown in fig. 3. From the above graph the highest accuracy is 99.00596 %. It is obtained when the number of training samples is 35. For each training sample the accuracy is more than 94%. Thus the proposed system is effective.

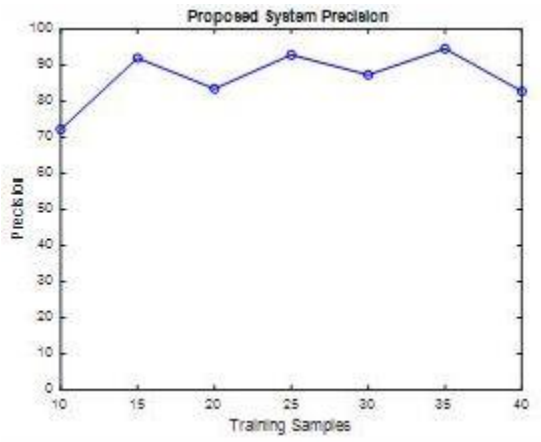

Fig. 4. *Precision of proposed system*

In fig. 4, the graph illustrates about the precision of the proposed system. The precision value varies when number of training samples is increases. The proposed system has 72.16495 % of precision during 10 training samples. However the highest precision is gained for 35 samples i.e. 94.5122%. The precision values for the whole process lies between 72% and 94%. The output for other data sets is shown in table 1.

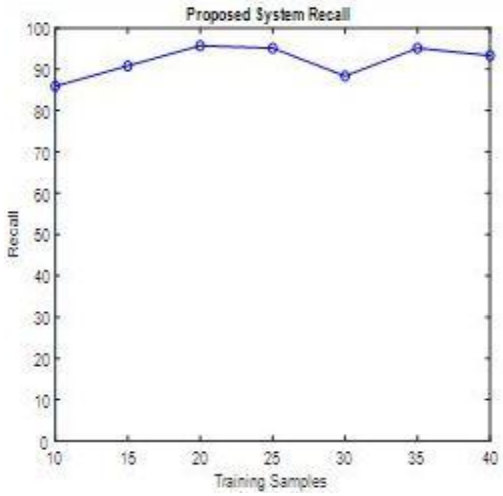

Fig. 5. *Recall of proposed system*

The recall of the proposed system is shown in fig. 5.7. This values lies between the range 85% to 96%. The highest recall value is 95.70552% for 20 training samples. More the recall value, more efficient will be the projected work. Thus, the proposed system has been proved better with these out comes. The corresponding values are shown in table

5.3. Comparative analysis of the parameters on the basis of the proposed scheme and other optimization algorithms:

The simulation results of proposed system are very effective. However these results are compared with the existing techniques. Most of the existing techniques made use of the optimization algorithms. Here in this comparison, parameters are analyzed in terms of following schemes:

- Whale optimization algorithm (WOA-BP)
- Genetic Algorithm (GA-BP)
- Particle Swarm Optimization (PSO-BP)
- CO-BP

Simulation results of all the schemes are framed on table 2 with respect to each parameter. The comparison of different approaches (WOA-BP, GA-BP, CO-BP) for credit card fraud detection is illustrated in the form of bar graphs showing the percentage of achieved parameters with respect to each method.

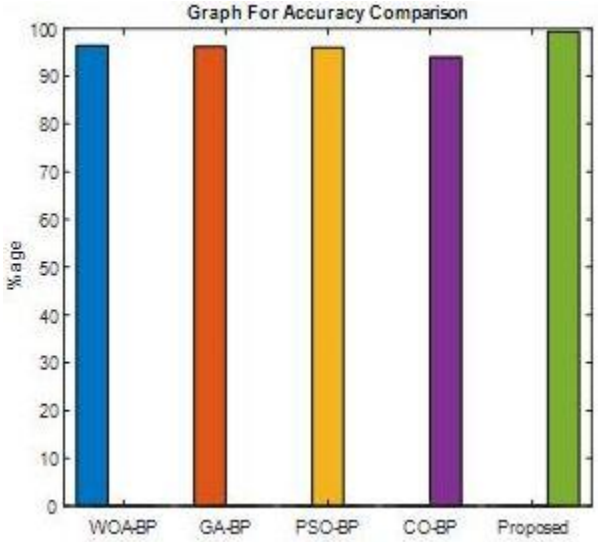

Fig. 6. *Comparison of proposed and existing schemes in terms of accuracy*

Fig.6 is demonstrating about the accuracy of the different methods applied for detecting the credit card fraud. Though, all the methods give more than 90% accurate results but it is clear from the graph that the proposed system has 99.36614% of accuracy. Thus, the proposed system is efficient among all other systems which used optimization algorithms.

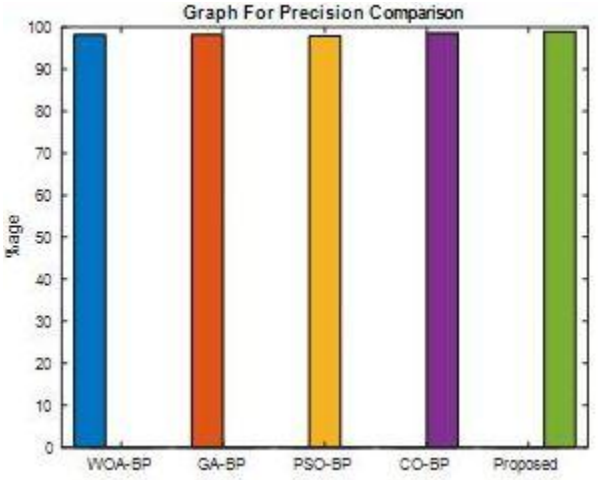

Fig. 7. *Comparison of proposed and existing schemes in terms of precision*

TABLE 2
RESULTS OF EXISTING AND PROPOSED APPROACHES IN TERMS OF PARAMETERS

Factors	WOA-BP	GA-BP	PSO-BP	CO-BP	Proposed
Accuracy	96.4	96.2	96	94	99.36614
Precision	98.25	98.25	97.83	98.63	99.00596
Recall	97.83	97.61	97.83	94.78	98.71757

The simulation results of precision value for the optimization algorithms (WOA-BP, GA-BP, CO-BP) and proposed scheme shows that the results are very much close to each other; however, the proposed system is highly précised as it has nearly 99.00596% of precision. Though detecting fraud using PSO-BP is not much effective and has 97% précised. The results are shown in fig. 7. The corresponding results of other methods are presented in table 2. In fig. 8, the graph is demonstrating about the simulation results of proposed and existing methods (WOA-BP, GA-BP, CO-BP) for detecting frauds in credit cards.

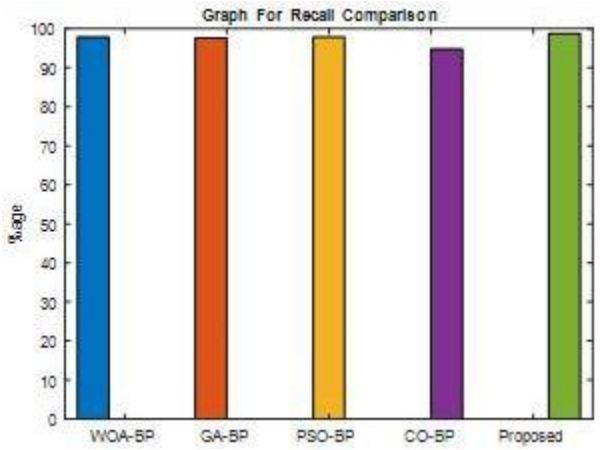

Fig. 8. *Comparison of proposed and existing schemes in terms of recall*

It can be seen in the graph of fig.8 that the proposed system has higher recall value. (98.71775%). However the system with least recall value to detect the fraud is the system with CO-BP algorithm. The table 2 is showing the results for all the methods used for comparison with the novel method.From the comparison of existing and proposed technique in terms of accuracy, precision and recall, the proposed system is proved to be effective and efficient amongst all the systems. This novel technique outperformed the existing systems in which optimization algorithms have been used.

4 CONCLUSION

In this paper, the proposed work aims to reduce the complexity, delay in process etc. for credit card fraud detection. This model is amalgamation of fuzzy inference system and neural networks which created neuro-fuzzy system. The complexity of the model is reduced as unlike traditional systems, it does not include the iterations; instead it uses the fuzzy rules. Feature extraction is done using LDA, Feature selection is performed by infinite feature selection technique. Classification of the features is done by using neuro-fuzzy system. This process provides benefit of easily understating, data categorization, and less complex training process Along with this, the proposed work is analyzed on the basis of three factors accuracy, precision and recall. The simulation results of proposed system are compared with the traditional systems in which optimization algorithms are used. From the results, the proposed system is proved to be as an efficient system from all the traditional systems.

REFERENCES

[1] Chunzhi Wang, Yichao Wang, Zhiwei Ye, Lingyu Yan, Wencheng Cai, Shang Pan , “Credit Card Fraud Detection Based On Whale Algorithm Optimized Bp Neural Network”, IEEE, pp 614-617, 2018

[2] Deshe ,Wang, Bintong Chen, Jing Chen, “Credit Card Fraud Detection Strategies with Consumer Incentives”, ELSEVIER, 2018, https://doi.org/10.1016/j.omega.2018.07.001

[3] Johannes Jurgovsky, Michael Granitzer, Konstantin Ziegler, “Sequence Classification for Credit-Card Fraud Detection”, ELSEVIER, vol 100, pp 234-245, 2018, https://doi.org/10.1016/j.eswa.2018.01.037

[4] Alex G.C.de Sá Adriano , C.M. Pereira Gisele, L.Pappa, “A Customized Classification Algorithm for Credit Card Fraud Detection”, ELSEVIER, vol 72, pp 21-29, 2018, https://doi.org/10.1016/j.engappai.2018.03.011

[5] Fabrizio Carcillo, Andrea Dal Pozzolo, “SCARFF: A Scalable Framework for Streaming Credit Card Fraud Detection with Spark”, ELSEVIER, vol 41, pp 182-194, 2018, https://doi.org/10.1016/j.inffus.2017.09.005

[6] Ugo Fiore, Alfredo De Santis, “Using Generative Adversarial Networks For Improving Classification Effectiveness In Credit Card Fraud Detection”, ELSEVIER, 2017, https://doi.org/10.1016/j.ins.2017.12.030

[7] Suraj Patil, VarshaNemade, Piyush Kumar Soni, “Predictive Modelling For Credit Card Fraud Detection Using Data Analytics”, ELSEVIER, vol 132, pp 385-395, 2018, https://doi.org/10.1016/j.procs.2018.05.199

[8] Nuno Carneiro, Gonçalo Figueira, Miguel Costa, “A Data Mining Based System For Credit-Card Fraud Detection In E-Tail”, ELSEVIER, vol 95, pp 91-101, 2017, https://doi.org/10.1016/j.dss.2017.01.002

[9] Vieira J, Dias FM, Mota A (2004) “Neuro-Fuzzy Systems: A Survey”, In: 5th WSEAS NNA international conference

[10] Jang JSR (1993) “ANFIS: Adaptive-Network-Based Fuzzy Inference System”, IEEE Trans Syst Man Cybern 23:665–685. https://doi.org/10.1109/21.256541

[11] Kar S, Das S, Ghosh PK (2014), “Applications Of Neuro Fuzzy Systems: A Brief Review And Future Outline”, Appl Soft Comput 15:243–259

[12] Barada S, Singh H (1998), “Generating Optimal Adaptive Fuzzy-Neural Models Of Dynamical Systems With Applications To Control”, IEEE Trans Syst Man Cybern Part C Appl Rev 28:371–391. https://doi.org/10.1109/5326.704574

[13] Xu A, Gao F, Wang J, Zhang L (2004), “Collision Avoidance Algorithm Based On ANFIS”, (Beijing Hangkong Hangtian Daxue Xuebao). J Beijing Univ Aeronaut Astronaut 30:670–673