Credit Card Anomaly Detection in Graph Database Using Controlled Mutual Information and Fuzzy Decision Rule

Mr. Mohamed Farook Ali N[1], Dr. Sasirekha N[2]

[1]*Ph.D Research Scholar, PG and Research Department of CS, Vidyasagar College of Arts and Science*

[2]*Assistant Professor, PG and Research Department of CS, Vidyasagar College of Arts and Science*

Abstract - **In recent years, e-commerce grows an important credential for global trade, the observation of anomaly detection which identifies the abnormal behavior in fraud detection of credit card transactions has turned into an interesting field of research. This paper emphases on automatic credit card fraudulent transaction detection with the graph related features. This work does two important tasks they are determining significant features using controlled mutual information and vagueness of credit card information handing is achieved by using fuzzy decision rule. The main objective of the feature subset selection is to increase the maximization of relevancy and to reduce the redundancy among attributes to attributes. The proposed controlled mutual information uses the attribute-to-attribute relationship along with the class label in a supervised manner to improve the relevancy rate of selected feature sets and the class attribute. The proposed fuzzy decision rule is used to infer the knowledge about pattern of credit card dataset and establish a classification model which can effectively handle the inconsistency in determining anomalous which is known as fraudulent transaction even in the imbalance dataset. From the performance analysis it is observed that the proposed model produces better results in credit card fraudulent detection.**

Index Terms - **Credit card fraudulent, anomaly detection, graph, controlled mutual information, fuzzy decision rule, supervised.**

I.INTRODUCTION

In this technical era, all the transactions are done through internet services and e-payment are very flexible and robust to perform secure transaction. The usage of the credit card is also increased, and it improves the lifestyle and the need of the business entities to common persons. But the security threating is one of the toughest challenges which has to be detected more accurately. There are two different categories in detection of credit card fraud transactions. They are user behavioral analysis known as anomaly detection and misuse detection achieved by fraud analysis.

The misuse detection is accomplished by supervised learning model in the transaction level, in this approach each transaction holds predefined labels which represent whether the concern transaction is normal or fraudulent based on analyzing the previous transaction history. These labeled datasets are involved in training process of the classification model, to learn about the normal transaction patterns and fraudulent transaction patterns. During the testing phase, the unknown new transaction without label is classified with the learning knowledge from historical dataset. The conventional classification models are decision tree, neural network and rule induction [1]. This model is recognized to detect in a reliable manner the fraudulent tricks which is examined so far and it is termed as misuse detection [2].

The anomaly detection is done by unsupervised model which is deployed based on the user behavior. An incoming transaction is treated as fraudulent if it is divergent form the normal behavior of the users This is because the behavior of the fraudsters will not be the same as the account owner or they do not have the knowledge about the behavior pattern of the concern owner [3]. The legitimate transaction of owners behaviour is extracted for each account and they detect any deviation from it and treated as fraudulent activities. While new behaviors are examined using this model various type of fraud activities can be detected. For each account their transaction activities are profiled, and it is comprised of type of merchant, location, amount and type of transaction. This method is coined as anomaly detection [4].

It is essential to differentiate among user behaviour analysis and fraud analysis models. In fraud analysis it can be able to detect known fraud attacks with low false detection rate. When a transaction is executed, this model searches previous historical transaction data of concern owner, if there is any match then it is considered to be normal, if know such genuine signature is found then fraud alarm will be raised [5]. Thus, the fraud analysis model works for specific and limited fraud records. Behavioral analysis addresses novel fraud detection by handling different fraud activities by comparing with similarity of normal and fraudulent transaction [6].

This paper concentrates on handling the uncertainty in detecting the fraudulency in credit card transaction. The feature extraction is done using controlled mutual information and graph-based anomaly detection is accomplished by constructing fuzzy decision rule to produce more accuracy in fraudulent detection.

II. RELATED WORK

Raghavendraet al. [7] in their work constructed a three layered artificial neural network which uses the genetic algorithm for optimizing its performance during training phase. The backpropagation-based learning model is used for detection the credit card detection.

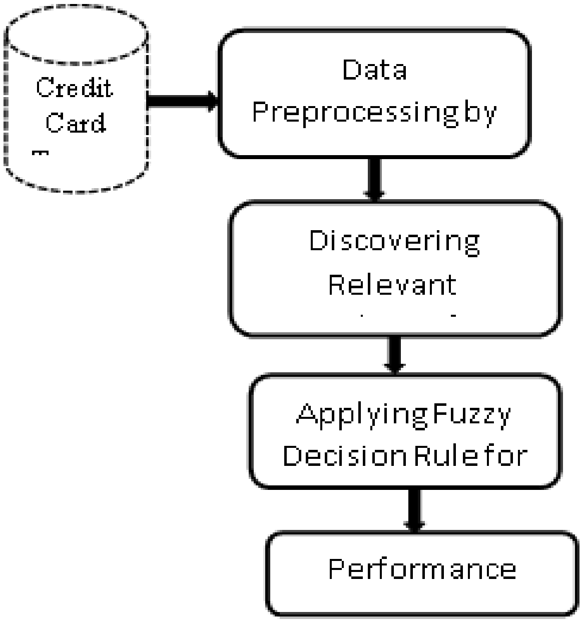
Awoyemi et al [8] they developed a nature adapted artificial model which is used for predicting the credit card fraudulent detection. Genetic algorithm is used for making decision which involves in deciding the number of hidden layers, neurons in each layer and the topology of the network

Mohammed[9] designed a machine learning approaches for credit card fraudulent detection. The used three different layers of auto associative network structure which synthesize both the training and testing process.

In [10] they designed a Random Forest which develops set of decision tree performing the pattern recognition in credit card fraudulence. The pattern recognition is done by random orest by appropriate training set and unlabeled transaction are analyzed and classified during the testing phase [11, 12].

RavneetKauraet al [13] in their work constructed a bidirectional neural model which handles the cell phone transaction with large volume of data offered by the credit card company. They used the rule-based classifier to detect the fraudulent and anomaly transaction.

Syeda, Mubeenaet al [14] presented a granular neural network is used to increase the computation time in prediction of credit card fraud detection, it is a kind of fuzzy neural network. The dataset used in this work is a Visa card transaction fraudulent detection.



Overall Framework of Credit Card anomaly detection in graph databaseusing Fuzzy Decision Rule

The objective of this proposed work is to discover credit card Anomaly in graph database using Fuzzy Decision Rule. The Credit Card dataset consists of 284, 807 transaction with 31 fields and 492 types of fraud. As an initial process, the dataset values are preprocessed using min-max normalization to make the attribute values fall under same range. The dataset consists of 31 fields among them the relevant fields or nodes involved in discovery of credit card anomaly transactions is determined using information gain. Finally, the classification of normal and anomaly transactions is done by developing an enhanced fuzzy decision rule system, which acts as an inference system, based on the gained knowledge about transaction patterns, it classifies each transaction as either normal or anomaly.

III. DATASET DESCRIPTION

The datasets contain transactions made by credit cards in September 2013 by European cardholders [17]. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features V1, V2, … V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependent cost-sensitive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

IV. MIN MAX NORMALIZATION

Before performing the process of classifying each transaction as fraudulent or non-fraudulent during credit card transactions, the value of instances in the dataset has to be in the same range of values. Using raw dataset values for classification often results in inaccurate process of detection, to overcome this issue in this research work Min-Max normalization is applied to maintain the attribute value ranges between 0 to 1. This formulated as

$$\text{Norm}(x) = \frac{x - \min(X)}{Max(X) - \min(X)}$$

Where x is the value of a specific instance's attribute, the min and max are the minimum and maximum value of the overall instances of the respective attribute field.

V. SIGNIFICANT FEATURE SUBSET SELECTION USING CONTROLLED MUTUAL INFORMATION

This work proposed a controlled mutual information [15] approach to maximize the relevancy among feature-class and to minimize the redundancy among feature-feature relationship. This process is helpful in determining the most significant attributes which highly influences the process of classification task is examined and selected for further processing. The workflow of the proposed Controlled mutual information method is depicted is the figure 2.

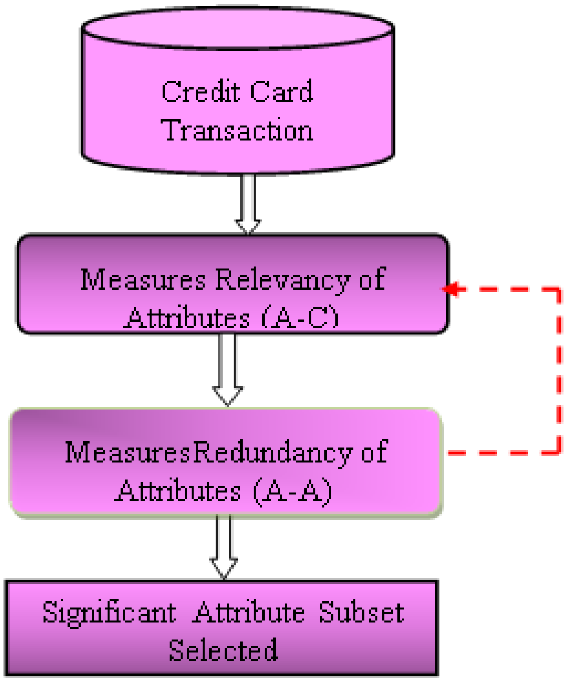

Credit Card Transaction
Measures Relevancy of Attributes (A-C)
MeasuresRedundancy of Attributes (A-A)
Significant Attribute Subset Selected

Figure- Work flow of Controlled Mutual Information based Feature Subset Selection

In general, during the process of feature selection, determining the relevancy and redundancy of attributes is the toughest challenge. The measure of relevancy involves in determining amount of information that can be gained from candidate attributes corresponding to class variable is obtained. In this work to discover the relevancy measure, Mutual Information among attributes and class variable is calculated as follows

$$\text{CMI}(Att,Cl) = \sum_{Att\in ATT} \sum_{Cl\in Class} prob(Att,Cl) log \frac{Prob(Att,Cl)}{Prob(Cl)Prob(Att)}$$

Where Attrefers to the attributes of the credit card dataset, cl refers to the class variable of the dataset. Prob() refers to the probability of events

The redundancy consists of the information among the candidate attributes and the selected attributes. The redundancy measure is also done by mutual information (CMI(Atti,Atts)). However, in general the class information is not integrated during for redundancy measure but this work includes the information of class variable also. The calculated of conditioned Mutual Information to determine the redundancy among attributes are formulated as follows:

$$E(\mathrm{x}) = -\sum_{x\epsilon X}^{n} prob(x)\log_2 prob(x)$$

$$INF(Att_i,Att_s/Cl) = E(Att_s/Cl) - E(Att_s/Att_i,Cl)$$

$$RDF(Att_i,Att_s) = INF(Att_i,Att_s) - E(Att_s/Cl) - E(Att_s/Att_i,Cl)$$

$$Cnd\text{-}MI = \frac{2*INF(Att_i,Cl)}{E(Att_i)+E(Cl)} - \frac{1}{|S|}\sum_{Att_s\in S}\left[\frac{2*RDF(Att_i,Att_s)}{E(Att_i)+E(Att_s)}\right]$$

VI. FUZZY INFERENCE SYSTEM

A fuzzy inference model is a system which uses theory of fuzzy set which maps the input to outputs which the help of fuzzy knowledge base and fuzzy inference system[16]. The input of the dataset is fed to the fuzzy inference model which process them based on predefined rules to generate the outputs. The inputs and outputs involved in fuzzy based classification is represented in real values whereas the processing of the system is depended on Fuzzy rules and its arithmetic

With the given inputs, the FIS computes the output by following six steps as follows:

1. Defining a set of fuzzy rules
2. Using membership function the crisp input values are converted to fuzzy values and this process is known as fuzzification
3. Merging the fuzzified inputs corresponding to the fuzzy rules to accomplish the strength of the rule
4. To obtain output distribution, the consequences are combined
5. Defuzzification is done to convert the fuzzified value to crisp output

VII. FUZZY DECISION RULE

The rules generated by fuzzy inference system are in the form of If then statements and a sample rule of Mamdani FIS with crisp input

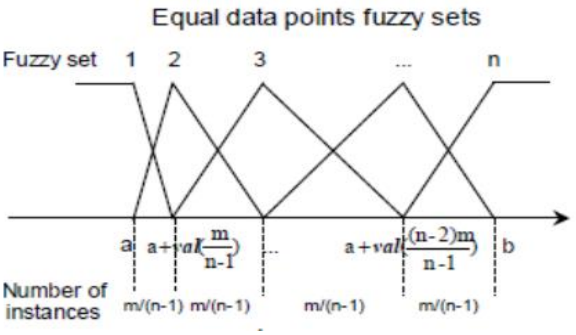


Figure fuzzy partition

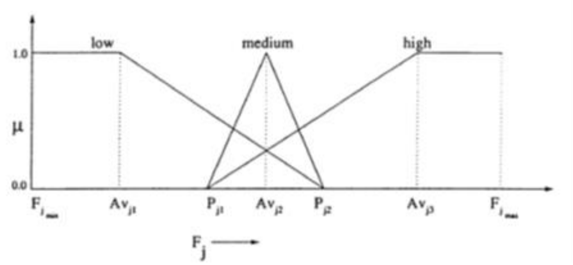


Figure Linguistic Representation

$$\mu_{lw}\left(ds_j\right)(\mathrm{H_i}) = \begin{cases} 1, & \text{for} H_{ij} {}_{<} Bv_{j1} \\ \dfrac{Q_{j2}-H_{ij}}{Q_{j2}-Bv_{j1}}, & \text{for } Bv_{j1} \leq H_{ij} < Q_{j2} \\ 0, & otherwise \end{cases}$$

$$\mu_{med}\left(ds_j\right)(\mathrm{H_i}) = \begin{cases} 0, & \text{for} H_{ij} {}_{<} Q_{j1} \\ \dfrac{Bv_{j2}-H_{ij}}{Bv_{j2}-Q_{j1}}, & \text{for } Q_{j1} \leq H_{ij} < Bv_{j2} \\ \dfrac{Q_{j2}-H_{ij}}{Q_{j2}Bv_{j2}}, & \text{for} Bv_{j2} \leq H_{ij} < Q_{j2} \\ 0, & otherwise \end{cases}$$

$$\mu_{hh}\left(DS_j\right)(\mathrm{H}_i) = \begin{cases} 0, & \text{for} H_{ij} {}_{<} Q_{j1} \\ \dfrac{H_{ij}-Q_{j1}}{Bv_{j3}-Q_{j1}}, & \text{for } Q_{j1} \leq H_{ij} < Bv_{j3} \\ 1, & otherwise \end{cases}$$

Assume if there are n number of attributes in the credit card dataset and m is the target variable, then the fuzzy rule is denoted as,

IF X_1 is high and X_2 is medium and X_3 is high then Yis abnormal transaction

IF X_1 is high and X_2 is medium and X_3 is medium then Y is abnormal transaction

IF X_1 is low and X_2 is medium and X_3 is low then Yis normal transaction

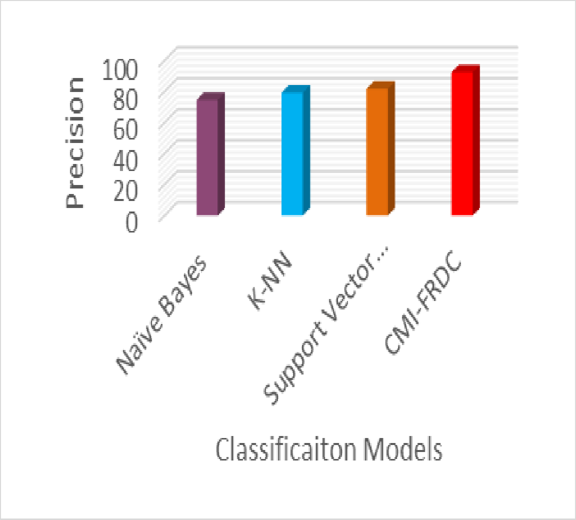
Where X_1, X_2 and X_3 are input variables and Y is the class label for the concern condition

VIII. SIMULATION RESULTS

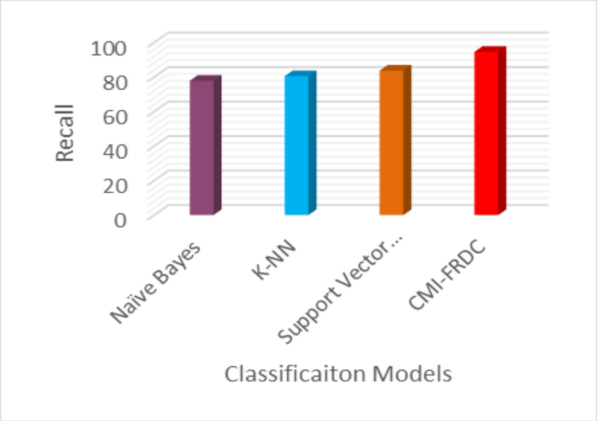
The proposed model Controlled Mutual Information based Fuzzy Rule Decision Classifier (CMI-FRDC) is simulated using Python. The dataset used in this work is collected from Kaggle repository [13]. The 284807 transactions are recorded with time, amount and Pca vectors as the input for finding the anomaly in credit card transaction. The proposed model performance is compared with three classification models Naïve Bayes, K-NN and support vector machine. The evaluation metrics used for investigating its performance are Precision, Recall and F-Measure.

Table: Performance Analysis of Four different classification Models for Anomaly detection in Credit Card Transaction

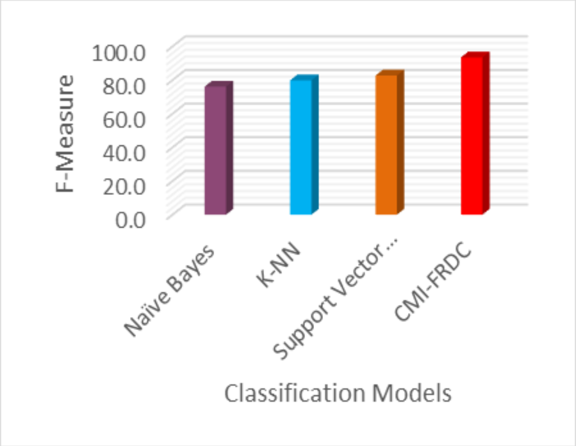
Classification Model	Precision	Recall	F-Measure
Naïve Bayes	74.5	77.8	76.1
K-NN	79.2	80.5	79.8
Support Vector Machine	81.7	83.7	82.7
CMI-FRDC	92.4	94.6	93.5



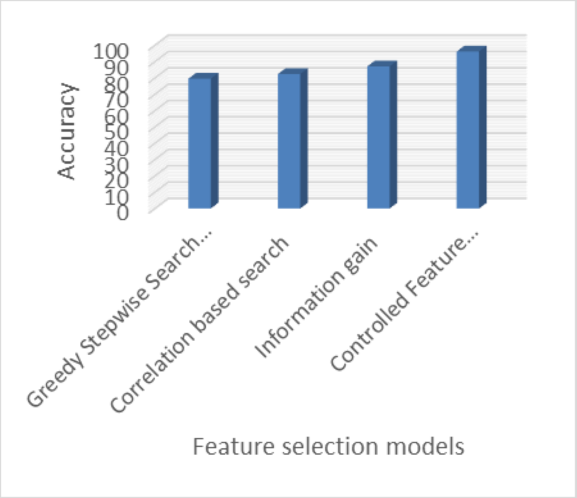
From the above graph it is observed that the performance of proposed Controlled Mutual Information based Fuzzy Rule Decision Classifier (CMI-FRDC) produces better precision while comparing with ANN, SVM and DBN. This is because the Naïve Bayes, KNN and SVM models fails to handle the problem of optimization when there is high degree of ambiguity in credit card fraud detection is graph based anomaly detection. While using FRDC with reduced attributed set by applying CMI-FRDC it encompasses two main advantage they are feature selection is done using improved mutual information and the fuzzy rule decision model handles the vagueness and ambiguity in anomaly detection of credit card transactions.



From the above graph it is observed that the performance of proposed Controlled Mutual Information based Fuzzy Rule Decision Classifier (CMI-FRDC) produces better recall value while comparing with Naïve Bayes, KNN and SVM. The support vector machine occupies next worst performance as it does not have clear definition of classification as it works on the probability and they won't be suitable for large datasets. The use of fuzzy decision making produces better result in ambiguity handling for anomaly detection. These two factors greatly influence the performance of CMI-FRDC to produce the highest precision rate during anomaly detection in credit card transaction.



From the above graph it is observed that the performance of proposed Controlled Mutual Information based Fuzzy Rule Decision Classifier (CMI-FRDC) produces better F-measure value while comparing with Naïve Bayes, KNN and SVM. It is revealed form the figure that as the precision and recall value of proposed CMI-FRDC is high, which is reflected in the F-measure also. This presented work produces better result because as the features selected using improved mutual information



The above graph illustrates the performance of the controlled feature selection model based Fuzzy decision rule classifier produces more accuracy while comparing other feature selection models. The main objective of CFS is to maximize the relevancy among features and class variables and to minimize the redundancy among features to overcome the dependency and to utilize the attribute correlation information in an effective way.

Performance Comparison of Four different Feature subset selection using Proposed Fuzzy Rule Decision Classifier

Feature subset Selection	Accuracy
Greedy Stepwise Search (GSW)	79.6
Correlation based search	82.4
Information gain	87.1
Controlled Feature Selection (CFS)	96.2

The table illustrates the performance of the proposed fuzzy decision rule classifier after performing feature subset selection using four different feature selection algorithms. With these selected set of attributes for detection of anomalous transaction the accuracy of fuzzy decision rule is determined. From the obtained results it is examined that the accuracy of fuzzy decision rule-based classifier produces high accuracy rate of 96.2% while using the proposed controlled mutual information feature subset selection. Hence, this work uses this model for feature subset selection while other feature subset selections fails to achieve the accuracy prominently.

IX. CONCLUSION

The foremost objective of this paper is using graph-based anomaly detection as a specific interest for analyzing credit card fraudulent detection. The richness of such datasets poses major challenges to previously developed anomaly detection schemes to continue the balance among interpretability, accuracy and scalability. The focal point of this paper is to devise an efficient fuzzy based classifications model which provide accurate results in credit card fraudulent detection. This work introduces a fuzzy decision rule-based anomaly detection in credit card transactions which effectively handles the inconsistency in detection of abnormal transactions. To reduce the time and computation complexity the feature subset selection is applied by proposing controlled mutual information. The simulation results are done with various evaluation metrics for both feature subset selection and classification model performance investigation. The outcome of the experimental results produces proves about the efficacy of this proposed model while comparing with other existing models.

REFERENCES

[1] KhyatiChaudhary, JyotiYadav, BhawnaMallick, “A review of Fraud Detection Techniques: Credit Card”, International Journal of Computer Applications Volume 45– No.1 2012.

[2] Michael Edward Edge, Pedro R, Falcone Sampaio, “A survey of signature-based methods for financial fraud detection”, journal of computers and security, Vol. 28, pp 3 8 1 – 3 9 4, 2009

[3] Salvatore J. Stolfo, David W. Fan, Wenke Lee and Andreas L. Prodromidis; "Credit Card Fraud Detection Using Meta-Learning: Issues and Initial Results"; Department of Computer Science-Columbia University; 1997.

[4] Maes S. Tuyls K. Vanschoenwinkel B. and Manderick B.; "Credit Card Fraud Detection Using Bayesian and Neural Networks"; Vrije University Brussel – Belgium; 2002.

[5] W.- Andreas L. Prodromidis and Salvatore J. Stolfo; "Agent-Based Distributed Learning Applied toFraud Detection"; Department of Computer Science- Columbia University; 2000.

[6] Salvatore J. Stolfo, Wei Fan, Wenke Lee and Andreas L. Prodromidis; "Cost-based Modeling

for Fraud and Intrusion Detection: Results from the JAM Project"; 0-7695-0490-6/99, 1999 IEEE.

[7] Raghavendra Patidar, Lokesh Sharma, “Credit Card Fraud Detection Using Neural Network”, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, 2011.

[8] Awoyemi, John O., et al. “Credit Card Fraud Detection Using Machine Learning Techniques: A Comparative Analysis.” 2017 International Conference on Computing Networking and Informatics (ICCNI), 2017.

[9] Mohammed, Emad, and Behrouz Far. “Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study.” IEEE Annals of the History of Computing, IEEE, 1 July 2018.

[10] Xuan, Shiyang, et al. “Random Forest for Credit Card Fraud Detection.” 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), 2018.

[11] S.J. Nowlan, “Max likelihood competition in RBP networks,” Technical Report CRG-TR-90- 2, Dept. of Computer Science, University of Toronto, Canada, 1990

[12] Masoumeh Zareapoora,PouryaShamsolmoali, Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier, Procedia Computer ScienceVolume 48, 2015, Pages 679-685

[13] RavneetKaura, MankiratKaura, Sarbjeet Singha, A Novel Graph Centrality Based Approach to Analyze Anomalous Nodes with Negative Behavior, International Conference on Information Security & Privacy (ICISP2015), 11-12 December 2015, Procedia Computer Science 78 (2016) 556 – 562.

[14] Syeda, Mubeena& Zhang, Yan-Qing & Pan, Yi. (2002). Parallel granular neural networks for fast credit card fraud detection. IEEE International Conference on Fuzzy Systems. 1. 572 - 577. 10.1109/FUZZ.2002.1005055.

[15] Czabanski R., Jezewski M., Leski J. (2017) Introduction to Fuzzy Systems. In: Prokopowicz P., Czerniak J., Mikołajewski D., Apiecionek Ł., Ślęzak D. (eds) Theory and Applications of Ordered Fuzzy Numbers. Studies in Fuzziness and Soft Computing, vol 356.

[16] Rui Zhao, Yang Gao, Pieter Abbeel, Volker Tresp, Wei Xu, Mutual Information-based State-Control for Intrinsically Motivated Reinforcement Learning, 15 pages, Jun 2020

[17] https://www.kaggle.com/mlg-ulb/creditcardfraud