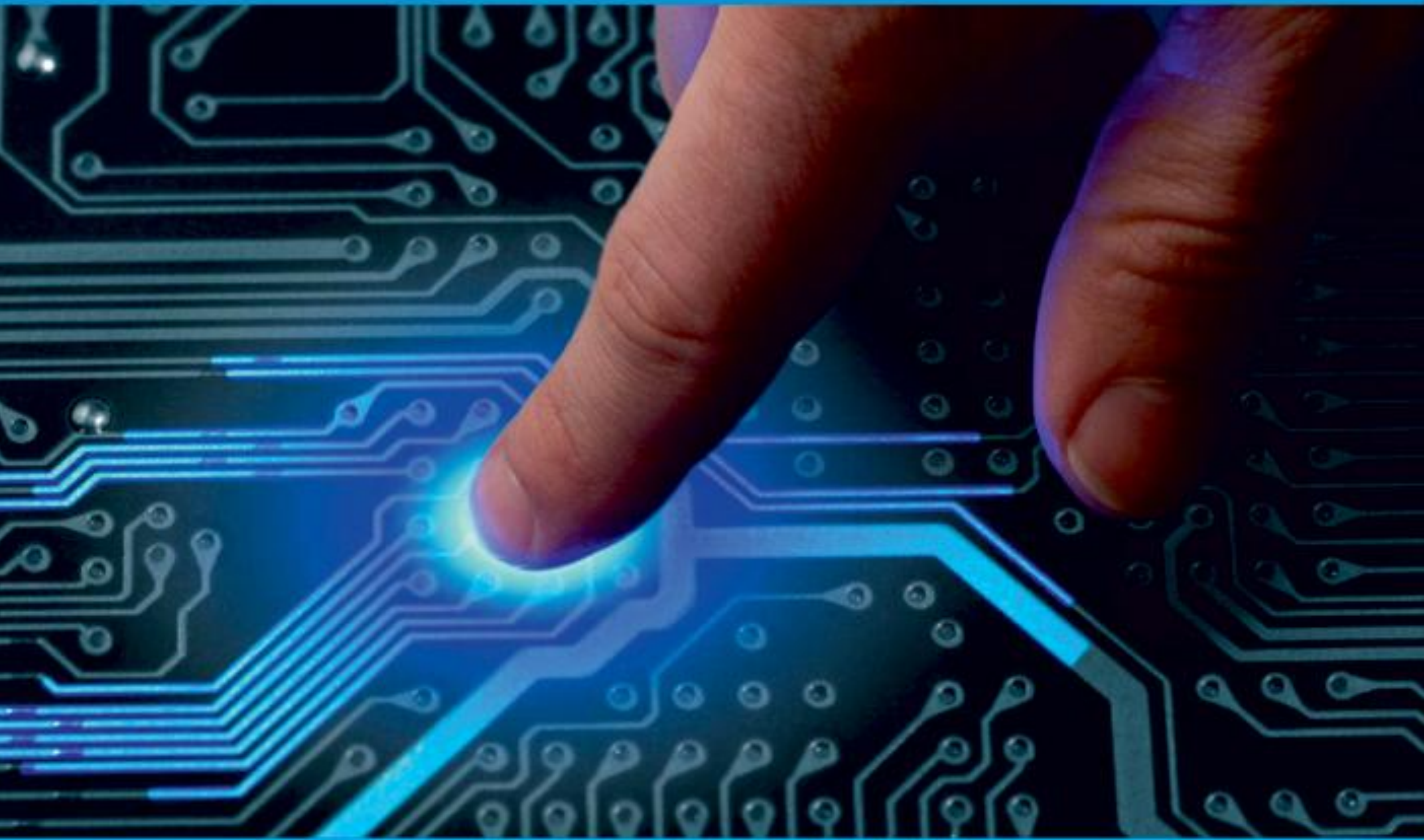
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 6, June 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906114 |

Credit Card Fraud Detection using Machine Learning

Mayuri Mane, Anita Kamble, Anuradha Bhadrashette, Ms. M. A. Rane

Department of Information Technology, Bharati Vidyapeeth's College of Engineering for Women, Pune, India

ABSTRACT: Credit card plays a very important rule in today's economy. It becomes an unavoidable part of household, business and global activities. Although using credit cards provides enormous benefits when used carefully and responsibly, significant credit and financial damages may be caused by fraudulent activities. Many techniques have been proposed to confront the growth in credit card fraud. However, all of these techniques have the same goal of avoiding the credit card fraud; each one has its own drawbacks, advantages and characteristics. In this paper, after investigating difficulties of credit card fraud detection, we seek to review the state of the art in credit card fraud detection techniques, datasets and evaluation criteria. The advantages and disadvantages of fraud detection methods are enumerated and compared. Furthermore, a classification of mentioned techniques into two main fraud detection approaches, namely, misuses (supervised) and anomaly detection (unsupervised) is presented. Again, a classification of techniques is proposed based on capability to process the numerical and categorical datasets. Different datasets used in literature are then described and grouped into real and synthesized data and the effective and common attributes are extracted for further usage. Moreover, evaluation employed criterions in literature are collected and discussed. Consequently, open issues for credit card fraud detection are explained as guidelines for new researchers

KEYWORDS: Credit card, Electronic commerce, Fraud detection, machine learning

INTRODUCTION

A credit card is a thin handy plastic card that contains identification information such as a signature or picture, and authorizes the person named on it to charge purchases or services to his account - charges for which he will be billed periodically. Today, the information on the card is read by automated teller machines (ATMs), store readers, bank and is also used in online internet banking system. They have a unique card number which is of utmost importance. Its security relies on the physical security of the plastic card as well as the privacy of the credit card number. There is a rapid growth in the number of credit card transactions which has led to a substantial rise in fraudulent activities. Credit card fraud is a wide-ranging term for theft and fraud committed using a credit card as a fraudulent source of funds in a given transaction. Generally, the statistical methods and many data mining algorithms are used to solve this fraud detection problem. Most of the credit card fraud detection systems are based on artificial intelligence, Meta learning and pattern matching.

At the current state of the world, financial organizations expand the availability of financial facilities by employing of innovative services such as credit cards, Automated Teller Machines (ATM), internet and mobile banking services. Besides, along with the rapid advances of e-commerce, the use of credit card has become a convenience and necessary part of financial life. Credit card is a payment card supplied to customers as a system of payment. In spite of all mentioned advantages, the problem of fraud is a serious issue in e-banking services that threaten credit card transactions especially. Fraud is an intentional deception with the purpose of obtaining financial gain or causing loss by implicit or explicit trick. Fraud is a public law violation in which the fraudster gains an unlawful advantage or causes unlawful damage. The estimation of amount of damage made by fraud activities indicates that fraud costs a very considerable sum of money. Credit card fraud is increasing significantly with the development of modern technology resulting in the loss of billions of dollars worldwide each year. statistics from the Internet Crime Complaint Centre show that there has been a significant rising in reported fraud in last decade. Financial losses caused due to online fraud only in US, was reported $3.4 billion in 2011. Fraud detection involves identifying scarce fraud activities among numerous legitimate transactions as quickly as possible. Fraud detection methods are developing rapidly in order to adapt with new incoming fraudulent strategies across the world. But, development of new fraud detection techniques becomes more difficult due to the severe limitation of the ideas exchange in fraud detection. On the other hand, fraud detection is essentially a rare event problem, which has been variously called outlier analysis, anomaly detection, exception mining, mining rare classes, mining imbalanced data etc. The number of fraudulent transactions is usually a very low fraction of the total transactions. Hence the task of detecting fraud transactions in an accurate and efficient manner is fairly difficult and challengeable. Therefore, development of efficient methods which can distinguish rare fraud activities from billions of legitimate transaction seems essential. Although, credit card fraud detection has gained attention and extensive study especially in recent years and there are lots of surveys about this kind of fraud such as [1], [2], [3],neither classify all credit card fraud detection techniques with analysis of datasets and attributes. Therefore in this paper, we attempt to collect and integrate a complete set of researches of literature and analyse them from various aspects.

II.LITERATURE REVIEW

The credit card fraud detection techniques are classified in two general categories: fraud analysis (misuse detection) and user behaviour analysis (anomaly detection). The first group of techniques deals with supervised classification task in transaction level. In these methods, transactions are labelled as fraudulent or normal based on previous historical data. This dataset is then used to create classification models which can predict the state (normal or fraud) of new records. There are numerous model creation methods for a typical two class classification task such as rule induction [1], decision trees [2] and neural networks [3].This approach is proven to reliably detect most fraud tricks which have been observed before [4], it also known as misuse detection. The second approach deals with unsupervised methodologies which are based on account behaviour. In this method a transaction is detected fraudulent if it is in contrast with user's normal behaviour. This is because we don't expect fraudsters behave the same as the account owner or be aware of the behaviour model of the owner [5].To this aim, we need to extract the legitimate user behavioural model (e.. user profile)for each account and then detect fraudulent activities according to it. Comparing new behaviours with this model, different enough activities are distinguished as frauds. The profiles may contain the activity information of the account; such as merchant types, amount, location and time of transactions, [6].This method is also known as anomaly detection. It is important to highlight the key differences between user behaviour analysis and fraud analysis approaches. The fraud analysis method can detect known fraud tricks, with a low false positive rate. These systems extract the signature and model of fraud tricks presented in oracle dataset and can then easily determine exactly which frauds, the system is currently experiencing. If the test data does not contain any fraud signatures, no alarm is raised. Thus, the false positive rate can be reduced extremely. However, since learning of a fraud analysis system (i.e. classifier) is based on limited and specific fraud records, It cannot detect novel frauds. As a result, the false negatives rate may be extremely high depending on how ingenious are the fraudsters.

User behavior analysis, on the other hand, greatly addresses the problem of detecting novel frauds. These methods do not search for specific fraud patterns, but rather compare incoming activities with the constructed model of legitimate user behavior. Any activity that is enough different from the model will be considered as a possible fraud. Though, user behavior analysis approaches are powerful in detecting innovative frauds, they really suffer from high rates of false alarm. Moreover, if a fraud occurs during the training phase, this fraudulent behavior will be entered in baseline mode and is assumed to be normal in further analysis[7].In this section we will briefly introduce some current fraud detection techniques which are applied to credit card fraud detection tasks, also main advantage and disadvantage of each approach will be discussed. An artificial neural network (ANN) is a set of interconnected nodes designed to imitate the functioning of the human brain [9]. Each node has a weighted connection to several other nodes in adjacent layers. Individual nodes take the input received from connected nodes and use the together with a simple function to compute output values. Neural networks come in many shapes and architectures. The Neural network architecture, including the number of hidden layers, the number of nodes within a specific hidden layer and their connectivity, most be specified by user based on the complexity of the problem. ANNs can be configured by supervised, unsupervised or hybrid learning methods.

In supervised learning, samples of both fraudulent and non-fraudulent records, associated with their labels are used to create models. These techniques are often used in fraud analysis approach. One of the most popular supervised neural networks is back propagation network (BPN). It minimizes the objective function using a multi-stage dynamic optimization method that is a generalization of the delta rule. The back propagation method is often useful for feed-forward network with no feedback. The BPN algorithm is usually time-consuming and parameters like the number of hidden neurons and learning rate of delta rules require extensive tuning and training to achieve the best performance [10]. In the domain of fraud detection, supervised neural networks like back-propagation are known as efficient tool that have numerous applications [11], [12], [13].

Raghavendra Patidar, et al. [14] used a dataset to train a three layers back propagation neural network in combination with genetic algorithms (GA)[15]for credit card fraud detection. In this work, genetic algorithms was responsible for making decision about the network architecture, dealing with the network topology, number of hidden layers and number of nodes in each layer.

Also, Aleskerovet al. [16] developed a neural network based data mining system for credit card fraud detection. The proposed system (CARDWATCH) had three layers auto associative architectures. They used a set of synthetized data for training and testing the system. The reported results show very successful fraud detection rates. In [17], a P-RCE neural network was applied for credit card fraud detection. P-RCE is a type of radial-basis function networks [18, 19]that usually applied for pattern recognition tasks. Krenkeret al. proposed a model for real time fraud detection based on bidirectional neural networks [20]. They used a large data set of cell phone transactions provided by a credit card company. It was claimed that the system outperforms the rule based algorithms in terms of false positive rate. Again in [21] a parallel granular neural network (GNN) is proposed to speed up data mining and knowledge discovery process for credit card fraud detection. GNN is a kind of fuzzy neural network based on knowledge discovery (FNNKD).The underlying dataset was extracted from SQL server database containing sample Visa Card transactions and then pre-processed for applying in fraud detection. They obtained less average training errors in the presence of larger training dataset.

PROPOSED SYSTEM DESIGN

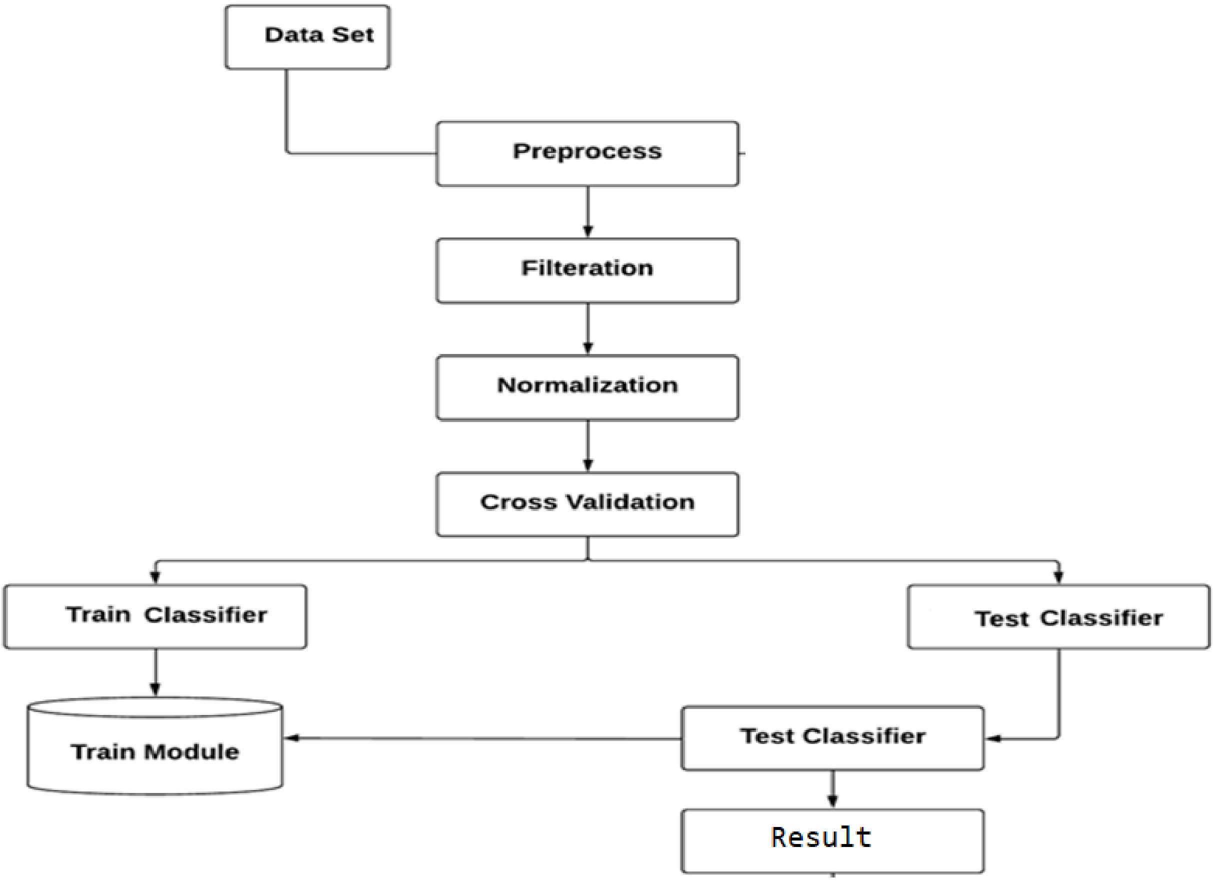


Figure 1 : Proposed system design

Data Preprocessing:

The basic data mining processes has done in first phase, Basically real time data sometimes hold some misclassified instances and null values, using data acquisition, data normalization, feature extraction etc. The normalize data does contains actual values of each attributes.

Classification using Machine Learning:

Once data pre-processing has done system deals with classification module of system, according the supervised learning approach we apply 10 fold and 15 fold cross validation on entire dataset for training and testing respectively. The classification results will shows actual disruption on roads based on real time data.

Analysis

In final phase we shows the predicted classification accuracy of system as well as comparative analysis of proposed system with various machine learning algorithm show display the effectiveness of proposed system.

Algorithm Design

Training Algorithm

Input: Training dataset TrainData[], Various activation functions[], Threshold Th

Output: Extracted Features Feature_set[] for completed trained module.

Step 1: Set input block of data d[], activation function, epoch size,

Step 2 : Features.pkl ← ExtractFeatures(d[])

Step 3 : Feature_set[] ← optimized(Features.pkl)

Step 4 : Return Feature_set[]

Testing Classification

Input: Training dataset TestDBLits [], Train dataset TrainDBLits[] and Threshold Th.

Output: Resulset <class_name, Similarity_Weight> all set which weight is greater than Th.

Step 1: For each testing records as given below equation

$$testFeature(k) = \sum_{m=1}^{n} (.featureSet[A[i] \ldots \ldots .. A[n] \leftarrow \text{TestDBLits})$$

Step 2 : Create feature vector from $testFeature(m)$ using below function.

$$\text{Extracted\_FeatureSet\_x } [\text{t}\ldots\ldots\text{n}] = \sum_{x=1}^{n}(\text{t}) \leftarrow testFeature \text{ (k)}$$

Extracted_FeatureSet_x[t] holds the extracted feature of each instance for testing dataset.

Step 3: For each train instances as using below function

$$trainFeature(l) = \sum_{m=1}^{n} (.featureSet[A[i] \ldots \ldots .. A[n] \leftarrow \text{TrainDBList})$$

Step 4 : Generate new feature vector from $trainFeature(m)$ using below function

.

$$\text{Extracted\_FeatureSet\_Y}[\text{t}\ldots\ldots\text{n}] = \sum_{x=1}^{n}(\text{t}) \leftarrow TrainFeature\ (l)$$

Extracted_FeatureSet_Y[t] holds the extracted feature of each instance for training dataset.

Step 5 : Now evaluate each test records with entire training dataset

$$weight = calcSim\ (\text{FeatureSetx} \,||\, \sum_{i=1}^{n} \text{FeatureSety[y]})$$

| DOI: 10.15680/IJIRCCE.2021.0906114 |

Step 6 : Return Weight

IV.RESULTS AND DISCUSSIONS

We evaluate our research provides better classification accuracy for various cross fold validation, the Naïve Bayes algorithms provides better classification accuracy than other classification results. Various supervised learning algorithms has used for comparative analysis like ANN [12], Random Forest [13], J48 [14] and K-Means [15] and out of them proposed NB provides best accuracy than other classification approaches.

The figure 2 to figure 5 shows the similar experiment analysis of proposed system with some existing systems. Finally this graphs proved the proposed classification provides better accuracy than other machine learning based classification algorithms.

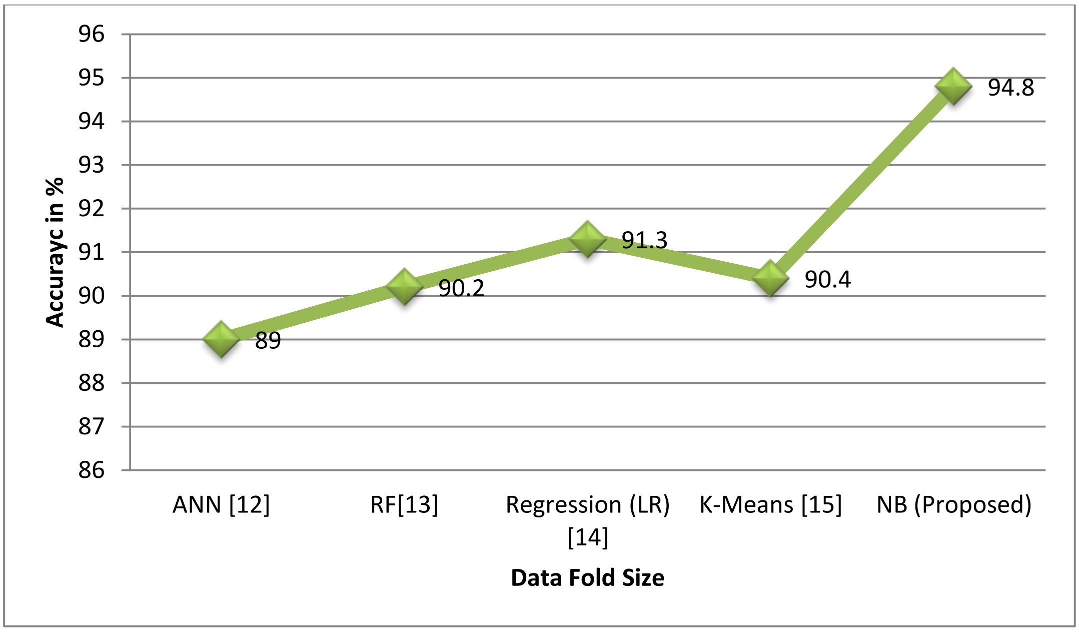


Figure 2: Classification Accuracy of existing machine learning with proposed Naïve Bayes Algorithm when 10 Fold Cross data Validation

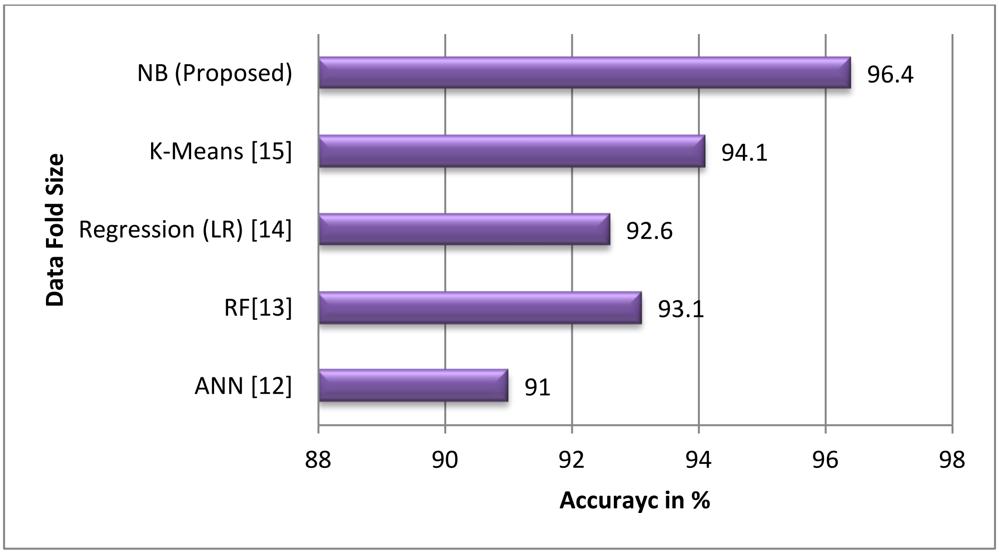


Figure 2: Classification Accuracy of existing machine learning with proposed Naïve Bayes Algorithm when 15 Fold Cross data Validation

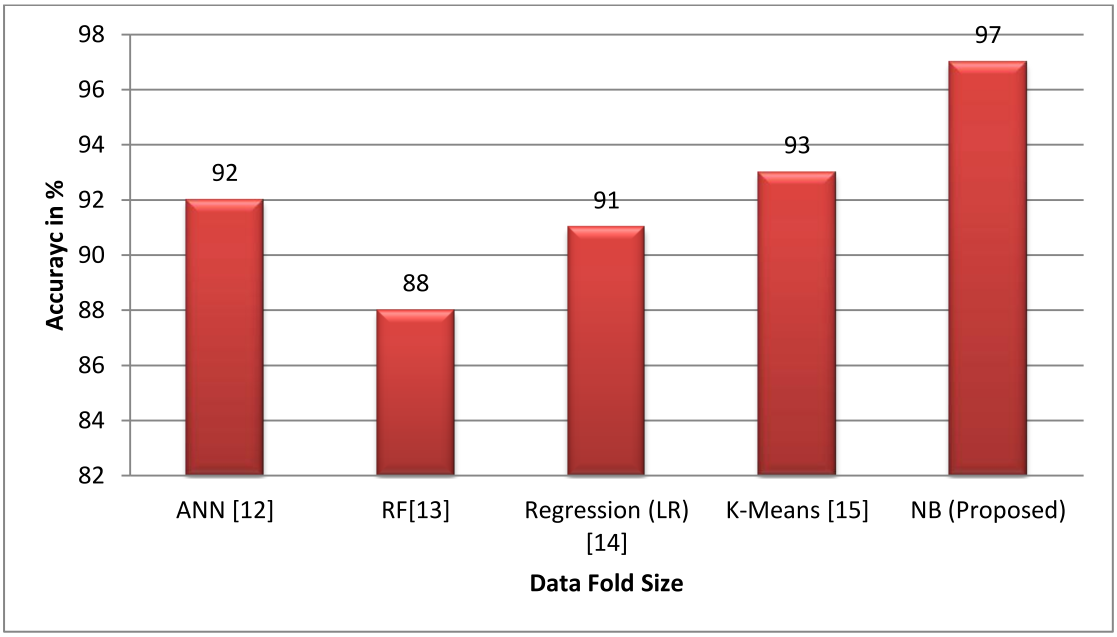


Figure 3: Classification Accuracy of existing machine learning with proposed Naïve Bayes Algorithm when 20 Fold Cross data Validation

V.CONCLUSION

This method proves accurate in finding out the fraudulent transactions and minimizing the number of false alert. Machine learning is appropriate in such kind of application areas. The use of this algorithm in credit card fraud detection system results in detecting or predicting the fraud probably in a very short span of time after the transactions has been made. This will eventually prevent the banks and customers from great losses and also will reduce risks.

REFERENCES

[1] Khyati Chaudhary, Jyoti Yadav, Bhawna Mallick, “ A review of Fraud Detection Techniques: Credit Card”, International Journal of Computer Applications Volume 45– No.1 2020.
[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, 2019.
[3] Linda Delamaire, Hussein Abdou, John Pointon, “Credit card fraud and detection techniques: a review”, Banks and Bank Systems, Volume 4, Issue 2, 2019.
[4] 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 ScienceColumbia University; 1917.
[5] Maes S. Tuyls K. Vanschoenwinkel B. and Manderick B.; "Credit Card Fraud Detection Using Bayesian and Neural Networks"; Vrije University Brussel – Belgium; 2002.
[6] Andreas L. Prodromidis and Salvatore J. Stolfo; "Agent-Based Distributed Learning Applied to Fraud Detection"; Department of Computer Science- Columbia University; 2020.
[7] 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, 2019 IEEE.
[8] Soltani, N., Akbari, M.K., SargolzaeiJavan, M., “A new user-based model for credit card frauddetection based on artificial immune system,” Artificial Intelligence and Signal Processing (AISP), 2019 16th CSI International Symposium on., IEEE, pp. 029-033, 2012.
[9] S. Ghosh and D. L. Reilly, “Credit card fraud detection with a neural-network”, Proceedings of the 27th Annual Conference on System Science, Volume 3: Information Systems: DSS/ KnowledgeBased Systems, pages 621-630, 2018. IEEE Computer Society Press.
[10] MasoumehZareapoor, Seeja.K.R, M.Afshar.Alam, ”Analysis of Credit Card Fraud Detection Techniques: based on Certain Design Criteria”, International Journal of Computer Applications (0975 – 8887) Volume 52– No.3, 2019.
[11] Fraud Brief – AVS and CVM, Clear Commerce Corporation, 2018, http://www.clearcommerce.com.
[12]All points protection: One sure strategy to control fraud, Fair Isaac, http://www.fairisaac.com, 2007.

[13] Clear Commerce fraud prevention guide, Clear Commerce Corporation, 2002, http://www.clearcommerce.com.
[14]RaghavendraPatidar, Lokesh Sharma, "Credit Card Fraud Detection Using Neural Network", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, 2011.
[15] Holland, J. H. "Adaptation in natural and artificial systems." Ann Arbor: The University of Michigan Press. (1975).
[16] E. Aleskerov, B. Freisleben, B. Rao, „CARDWATCH: A Neural Network-Based Database Mining System for Credit Card Fraud Detection", the International Conference on Computational Intelligence for Financial Engineering, pp. 220-226, 2017.
[17] SushmitoGhosh, Douglas L. Reilly, Nestor, "Credit Card Fraud Detection with a NeuralNetwork", Proceedings of 27th Annual Hawaii International Conference on System Sciences, 2015.
[18] Moody and C. Darken, "Learning with localized receptive fields." in Proc. of the 1988 Connectionist Models Summer School, D.S. Touretzky, G.E. Hinton and T.J. Sejnowski, eds., Morgan Kaufmann Publishers, San Mateo, CA, 2019, pp. 133-143.
[19] S.J. Nowlan, "Max likelihood competition in RBP networks," Technical Report CRG-TR-90- 2, Dept. of Computer Science, University of Toronto, Canada, 1990.
[20] A. Krenker, M. Volk, U. Sedlar, J. Bester, A. Kosh, "Bidirectional Artificial Neural Networks for Mobile-Phone Fraud Detection," Journal of Artificial Neural Networks, Vol. 31, No. 1, pp. 92-98, 2019.
[21]MubeenaSyeda, Yan-Qing Zbang and Yi Pan," Parallel granular neural networks for fast credit card fraud detection", international conference on e-commerce application, 2020.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details