International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST

A Novel Learning Strategy for Credit Card Fraud Detection

Prof. P.S.Gayake[1], Ikhe Varsha[2], Guldagad Rutuja[3], Bhore Pranjal[4], Labade Sujata[5]

[1] Assistant Professor and [2,3,4,5] Students

Department Information Technology, Dr. Vithalrao Vikhe Patil College of Engineering, Ahmednagar, Savitribai Phule Pune University, Pune, Maharashtra, India

Article Info
Volume 9, Issue 3
Page Number : 244-250
Publication Issue
May-June-2022
Article History
Accepted : 10 May 2022
Published : 30 May 2022

ABSTRACT

Card fraud detection is presently the most frequently occurring problem in the present world. This is due to the rise in both online transactions and e-commerce platforms. Credit card fraud generally happens when the card was stolen for any of the unauthorized purposes or even when the fraudster uses the credit card information for his use. In the present world, we are facing a lot of credit card problems. Credit card fraud detection is presently the most frequently occurring problem in the present world. This is due to the rise in both online transactions and e-commerce platforms. Credit card fraud generally happens when the card was stolen for any of the unauthorized purposes or even when the fraudster uses the credit card information for his use. In the present world, we are facing a lot of credit card problems. To detect the fraudulent activities the credit card fraud detection system was introduced. To detect the fraudulent activities the credit card fraud detection system was introduced. This project aims to focus mainly on machine learning algorithms. The algorithms used are Logistic regression algorithm.

Keywords : Training Module, Machine Learning, Logistic Regression, Authentications, Confidentiality, Transactions, Security System, Verification System, Credentials, etc.

I. INTRODUCTION

Financial fraud is an ever growing menace with far reaching consequences in the finance industry, corporate organizations, and government. Fraud can be defined as criminal deception with intent of acquiring financial gain. High dependence on internet technology has enjoyed increased credit card transactions. As credit card transactions become the most prevailing mode of payment for both online and offline transaction, credit card fraud rate also accelerates. Credit card fraud can come in either inner card fraud or external card fraud. Inner card fraud occurs as a result of consent between cardholders and bank by using false identity to commit fraud while the external card fraud involves the use of stolen credit card to get cash through dubious means. A lot of researches have been devoted to detection of

external card fraud which accounts for majority of credit card frauds. Detecting fraudulent transactions using traditional methods of manual detection is time consuming and inefficient, thus the advent of big data has made manual methods more impractical. However, financial institutions have focused attention to recent computational methodologies to handle credit card fraud problem.

Data mining technique is one notable methods used in solving credit fraud detection problem. Credit card fraud detection is the process of identifying those transactions that are fraudulent into two classes of legitimate (genuine) and fraudulent transactions [1]. Credit card fraud detection is based on analysis of a card's spending behaviour. Many techniques have been applied to credit card fraud detection, artificial neural network [2], genetic algorithm [3, 4], support vector machine [5], frequent itemset mining [6], decision tree [7], migrating birds optimization algorithm [8], naïve bayes [9]. A comparative analysis of logistic regression and naïve bayes is carried out in [10]. The performance of bayesian and neural network [11] is evaluated on credit card fraud data.

Decision tree, neural networks and logistic regression are tested for their applicability in fraud detections [12]. This paper [13] evaluates two advanced data mining approaches, support vector machines and random forests, together with logistic regression, as part of an attempt to better detect credit card fraud while neural network and logistic regression is applied on credit card fraud detection problem [14]. A number of challenges are associated with credit card detection, namely fraudulent behaviour profile are dynamic, that is fraudulent transactions tend to look like legitimate ones; credit card

transaction datasets are rarely available and highly imbalanced (or skewed); optimal feature (variables) selection for the models; suitable metric to evaluate performance of techniques on skewed credit card fraud data. Credit card fraud detection performance is greatly affected by type of sampling approach used, selection of variables and detection technique(s) used. This study investigates the effect of hybrid sampling on performance of fraud detection of naïve bayes, k-nearest neighbour and logistic regression classifiers on highly skewed credit card fraud data.

This paper seeks to carry out comparative analysis of credit card fraud detection using naive bayes, k-nearest neighbor and logistic regression techniques on highly skewed data based on accuracy, sensitivity, specificity and Matthews's correlation coefficient (MCC) metrics. This paper extends the handling of highly imbalanced credit card fraud data in [33]. The imbalanced dataset used in this study which contains about 0.172% of fraud transactions is sampled in a hybrid approach. The positive class (fraud) is oversampled while the negative class (legitimate) is under-sampled by the same number of times to achieve two distributions of 34:66 and 10:90. The three techniques are applied to the data. The performance comparison of the three techniques is analysed based on accuracy, sensitivity, specificity, Matthews Correlation Coefficient (MCC) and balanced classification rate.

II. RELATED WORK

"Credit Card Fraud Detection Using Machine Learning" [1] This project aims to focus mainly on machine learning algorithms. The algorithms used are random forest algorithm and the algorithm The Credit Card Fraud Detection Problem includes modelling past credit card transactions with the data of the ones that turned out to be fraud. This model is then used to recognize whether a new transaction is fraudulent or not. Our objective here is to detect 100 of the fraudulent transactions while minimizing the incorrect fraud classifications. Credit Card Fraud Detection is a typical sample of classification. In this process, we have focused on analysing and pre-processing data sets as well as the deployment of multiple anomaly detection algorithms such as Local Outlier Factor and Isolation Forest algorithm on the PCA transformed Credit Card Transaction data.

Credit card fraud detection using machine learning techniques: A comparative analysis"[2] This paper investigates the performance of na¨ıve bayes, k-nearest neighbor and logistic regression on highly skewed credit card fraud data. Dataset of credit card transactions is sourced from European cardholders containing 284,807 transactions. A hybrid technique of under-sampling and oversampling is carried out on the skewed data. The three techniques are applied on the raw and pre-processed data. The work is implemented in Python. The performance of the techniques is evaluated based on accuracy, sensitivity, specificity, precision, Matthews correlation coefficient and balanced classification rate. The results shows of optimal accuracy for na¨ıve bayes, k-nearest neighbor and logistic regression classifiers are 97.92, 97.69 and 54.86 respectively. The comparative results show that k-nearest neighbour performs better than na¨ıve bayes and logistic regression techniques.

"Deep Learning Approach for Credit Card Fraud Detection"[3]. As technology evolves rapidly, the world is using credit cards instead of cash in its everyday lives, opening up a new way for fraudulent people to abuse them. Credit card fraud losses reached approximately 28.65 billion in 2019, according to Nilsson's report, and global card fraud is expected to reach around 32.96 billion by 2023. Providers should therefore develop an efficient model to detect and prevent fraud early. In this paper, we used deep learning techniques as an effective way to detect fraudsters in credit card transactions. Therefore, we present a model for predicting legitimate transactions or fraud on Kaggle's credit card dataset. The proposed model is OSCNN (Over Sampling with Convolution Neural Network) which is based on over-sampling pre-processing and CNN (convolution neural network). The MLP (Multi-layer perceptron) was also applied to the dataset. Comparing the MLPOSCNN results, they proved that the proposed model achieved better results with 98 accuracy.

"A Review of Credit Card Fraud Detection Techniques"[4]. Credit card plays a significant standard in the present wealth. It turns into a necessary piece of the family unit, business, and worldwide exercises. Although utilizing credit cards gives might profits when used carefully and dependably, huge credit and monetary effects might be imported by deceitful practices by fraudsters attributable to the ubiquity of electronic asset moves. Financial institutions try to enhance continuously their fraud detection systems, but fraudsters are at the same time hack into the systems with new techniques and tools. Such cheats cause a danger to the protection of humankind, bringing about monetary misfortunes. There is a requirement for planning progressed extortion discovery answers to limit the perils of these fakes. For the detection of deceits, many machine learning algorithms can be utilized. This note paper first discusses the statistics of credit card frauds in the world and primarily in India, then the type of frauds and gives a diagram to analyse the presentation of a few machine learning algorithms by doing a relative report that can be utilized for classifying transactions as misrepresentation or a real one. It also mentions the currently used state of the art techniques to counter these attacks and highlights its limitations along with proposing a solution for it.

ALGORITHMS AND METHODLOGY

Classification of credit card transactions is mostly a binary classification problem. Here, credit card transaction is either as a legitimate transaction (negative class) or a fraudulent transaction (positive class). Fraud detection is generally viewed as a data mining classification problem, where the objective is to correctly classify the credit card transactions as legitimate or fraudulent [6].

Credit Card Fraud:

Credit card frauds have been partitioned into two types: inner card fraud and external fraud [12, 15] while a broader classification have been done in

three categories, that is, traditional card related frauds (application, stolen, account takeover, fake and counterfeit), merchant related frauds (merchant collusion and triangulation) and Internet frauds (site cloning, credit card generators and false merchant sites) [16].

Credit card transactions data are mainly characterized by an unusual phenomenon. Both legitimate transactions and fraudulent ones tend to share the same profile. Fraudsters learn new ways to mimic the spending behaviour of legitimate card (or cardholder). Thus, the profiles of normal and fraudulent behaviours are constantly dynamic. This inherent characteristic leads to a decrease in the number of true fraudulent cases identified in a pool of credit card transactions data leading to a highly skewed distribution towards the negative class (legitimate transactions).

Feature (Variables) Selection:

The basis of credit card fraud detection lies in the analysis of cardholder's spending behavior. This spending profile is analyzed using optimal selection of variables that capture the unique behavior of a credit card. The profile of both a legitimate and fraudulent transaction tends to be constantly changing. Thus, optimal selection of variables that greatly differentiates both profiles is needed to achieve efficient classification of credit card transaction. The variables that form the card usage profile and techniques used affect the performance of credit card fraud detection systems. These variables are derived from a combination of transaction and past transaction history of a credit card. These variables fall under five main variable types, namely all transactions statistics, regional statistics, merchant type statistics, time based amount statistics and time-based number of transactions statistics [9].

The variables that fall under all transactions statistics type depict the general card usage profile of the card. The variables under regional statistics type show the spending habits of the card with taken into account the geographical regions. The variables under merchant statistics type show the usage of the card in different merchant categories. The variables of time based statistics types identify the usage profile of the cards with respect to usage amounts versus time ranges or frequencies of usage versus time ranges. Most literature focused on cardholder profile rather than card profile. It is evident that a person can operate two or more credit cards for different purposes. Therefore, one can exhibit different spending profile on such cards. In this study, focus is beamed on card rather than cardholder because one credit card can only exhibit a unique spending profile while a cardholder can exhibit multiple behaviours on different cards.

Credit card Fraud Detection:

As credit card becomes the most general mode of payment (both online and regular purchase), fraud rate tends to accelerate. Detecting fraudulent transactions using traditional methods of manual detection are time consuming and inaccurate, thus the advent of big data had made these manual methods more impractical. However, financial institutions have turned to intelligent techniques. These intelligent fraud techniques comprise of computational intelligence (CI)-based techniques. Statistical fraud detection methods have been divided into two broad categories: supervised and unsupervised [22]. In supervised fraud detection methods [13], models are estimated based on the samples of fraudulent and legitimate transactions to classify new transactions as fraudulent or legitimate while in unsupervised fraud detection, outliers' transactions are detected as potential instances of fraudulent transactions. A detailed discussion of supervised and unsupervised techniques is found in [23]. Quite a number of studies on a range of techniques have been carried out in solving credit card fraud detection problem. These techniques include but not limited to; neural network models (NN), Bayesian network (BN), intelligent decision engines (IDE), expert systems, meta-learning agents,

machine learning, pattern recognition, rule-based systems, logic regression (LR), support vector machine (SVM), decision tree, k-nearest neighbor (kNN), meta learning strategy, adaptive learning etc. Some related works on comparative study of credit card fraud detection techniques are presented.

III.PROPOSED SYSTEM

Afterwards, the proposed solution will be implemented with all essential input and output parameters. Then the implementation will undergo a thorough performance analysis and detailed comparison with the existing models. We will train module on dataset of credit card transaction. Then we will pass some transaction to module to predict weather transaction is fraud or not

First we collect dataset of credit card transaction from kaggle.com. Then pre-processing of dataset to handle missing value and unwanted data. Data pre-processing is vital in any information mining process as they straightforwardly sway achievement pace of the task. ....

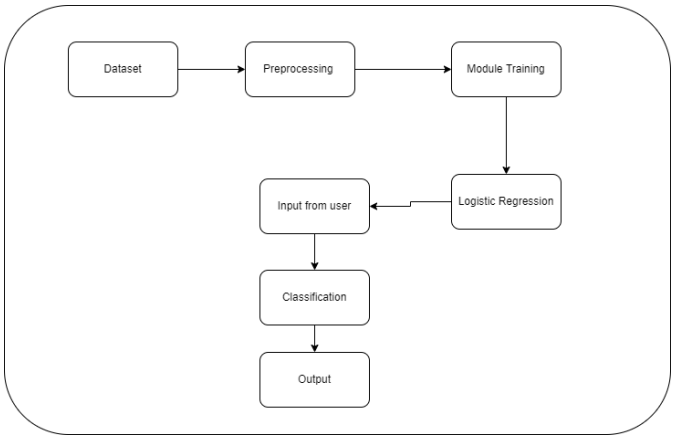

Fig. 1: System Architecture

Information is supposed to be messy in case it is missing quality, trait esteems, contain commotion or exceptions and copy or wrong information. Presence of any of these will debase nature of the outcomes. In this dataset two type of transaction is available.

1. Fraud Transaction.
2. Correct transaction.

System will train on this dataset by applying logistic regression algorithm.

After training of module we will test our module accuracy.

We will divide our dataset into two parts.

1. Training Data.
2. Testing data.

So we will divide dataset into part percent Training and 30 percent for testing. We utilize the preparation information to fit the model and testing information to test it. The models produced are to anticipate the outcomes obscure which is named as the test set. As you brought up, the dataset is isolated into train and test set to really look at exactness's, precisions via preparing and testing it on it.

IV.RESULT ANLYSIS

This research work is to implement a web-based application for the healthcare community to prevent various attacks of patients as well as user's confidential data records storage and transmission time. The result analysis is done based on the following parameters is as follows:

- Time consumption
- Response Time
- Computation Cost
- Performance accuracy

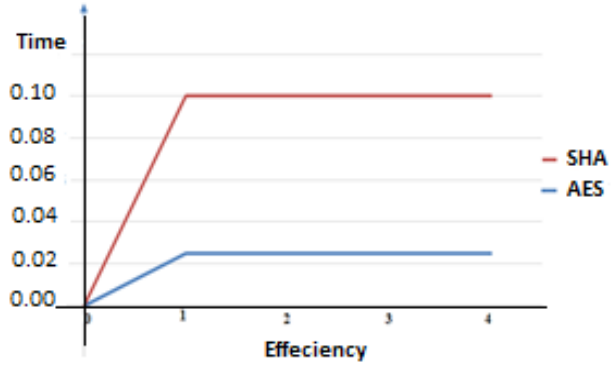

Fig.8: Time and Efficiency Chart

Here, Whole System took more attributes for the input purpose but here mainly concentrates on the Time and Performance of the system. In existing system required more time, space and security issues so we first focus on those things. Supported a couple

of attributes we'll get the subsequent analytical result for our proposed system.

Parameter	Existing	Proposed
A	10	4
B	10	5
C	8	8
D	10	3
E	8	2

Table 1: Result Table

Where,

A = Time Consumption.

B = Response Time.

C = Computation Cost.

D = Performance accuracy

E = Scalable & User Friendly.

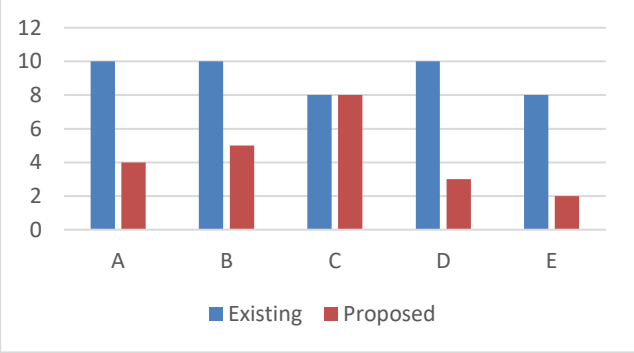

Fig.9: Time line chart of Result Analysis

V. CONCLUSION

After doing the comparative study, it has been found that every algorithm contains its focal points and weaknesses. There are many fraud detection techniques available but any of the techniques is not able to detect fraud when it is happening it detects when fraud is already done because a very less number of transactions are fraudulent. So the solution to this problem is that we need a technology that can detect fraud when fraud is happening. So the significant task of the time being is to construct a correct and quick recognizing model that can identify not just fraud occurring on the internet but also tampering with credit cards themselves.

VI. REFERENCES

[1] P. Chan and S. Stolfo, “Toward scalable learning with non-uniform class and cost distributions: A case study in credit card fraud detection,” KDD, 1998.

[2] SamanehSorournejad, Z. Zojaji, R. E. Atani, and A. H. Monadjemi, “A survey of credit card fraud detection techniques: Data and technique-oriented perspective,” arXiv [cs.CR], 2012.

[3] Rohilla, Anju, and Ipshita Bansal. “Credit Card Frauds: An Indian Perspective.” Volume 2, 2015: 591-597.

[4] Maes, Sam, et al. “Credit card fraud detection using Bayesian and neural networks.” Proceedings of the 1st international naiso congress on neuro-fuzzy technologies. 2002.

[5] Save, Prajal, et al. “A novel idea for credit card fraud detection using a decision tree.” International Journal of Computer Applications 161.13 (2017).

[6] F. Braun, O. Caelen, E. N. Smirnov, S. Kelk, and B. Lebichot, “Improving card fraud detection through suspicious pattern discovery,” in Advances in Artificial Intelligence: From Theory to Practice, Cham: Springer International Publishing, 2017, pp. 181– 190.

[7] Kiran, Sai, et al. “Credit card fraud detection using Na¨ıve Bayes model-based and KNN classifier.” International Journal of Advanced Research, Ideas, and Innovations in Technology 4.3 (2018).

[8] Campus, Kattankulathur. “Credit card fraud detection using machine learning models and

collating machine learning models." International Journal of Pure and Applied Mathematics 118.20 (2018): 825-838.

[9] Lakshmi, S. V. S. S., and S. D. Kavilla. "Machine learning for credit card fraud detection system." Int. J. Appl. Eng. Res. 13.24 (2018):

[10] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and C. Jiang, "Random forest for credit card fraud detection," in 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), 2018, pp. 1–6.

[11] Divakar, Kavya, and K. Chitharanjan. "Performance evaluation of credit card fraud transactions using boosting algorithms." Int. J. Electron. Commun. Comput. Eng. IJECCE 10.6 (2019): 262-270.

[12] U. Porwal and S. Mukund, "Credit card fraud detection in Ecommerce," in 2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), 2019, pp. 280–287.

[13] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic, and A. Anderla, "Credit card fraud detection - machine learning methods," in 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH), 2019, pp. 1–5.

[14] X. Zhang, Y. Han, W. Xu, and Q. Wang, "HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture," Inf. Sci. (Ny), 2019

[15] A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga, and N. Kuruwitaarachchi, "Real-time credit card fraud detection using machine learning," in 2019 9th International Conference on Cloud Computing, Data Science Engineering (Confluence), 2019, pp. 488–493.

Cite this article as :

Prof. P. S. Gayake, Ikhe Varsha, Guldagad Rutuja, Bhore Pranjal, Labade Sujata, "A Novel Learning Strategy for Credit Card Fraud Detection", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 9 Issue 3, pp. 244-250, May-June 2022.
Journal URL : https://ijsrst.com/IJSRST229356