INTERNATIONAL RESEARCH JOURNAL OF ENGINEERING & APPLIED SCIENCES
ISSN: 2322-0821(O) ISSN: 2394-9910(P)
VOLUME 12 ISSUE 3
Jul 2024 - Sep 2024
www.irjeas.org

Original Article

Credit Card Fraud Detection using Gradient Boosting Machine Learning Techniques

Shivam Singh[1], Pankaj Pandey[2]

[1]*Research Scholar, Dept. of Computer Science & Engineering, Jai Narain College of Technology (JNCT), Bhopal, INDIA*
[2]*Asst. Professor, Dept. of Computer Science & Engineering, Jai Narain College of Technology (JNCT), Bhopal, INDIA*

Corresponding Author: pankaj.cse@jnctbhopal.ac.in

DOI –10.55083/irjeas.2024.v12i03002

Abstract: As more people use credit cards, fraudulent usage has become more widespread to keep up with the demand. The disadvantage is that it depends on developing an accurate CCFDS. In order to increase the security of financial transaction systems in an independent and effective manner, financial institutions need to construct a CCFDS that is accurate and efficient. During the process of detecting fraudulent activity involving CC, even if seemingly little aspects of the fraud are neglected, there is a dangerous possibility that these errors may turn out to be significant obstacles down the road. The detection process is a very tricky task since its dataset consists of substantial class imbalances. It biases classifier models towards the validation set, resulting in high train and validation accuracies but inordinately large false positives or false negatives. Financial institutions such as banks, etc. use credit card fraud detection are algorithms for matching transactional statistics and guesstimating whether the upcoming transaction is a fraudulent one or not. Moreover, institutions' resources might be directed toward more questionable transactions in order to reduce fraud levels. This paper is used to GB based ML technique for CCFDS and achieve good accuracy.

Keyword: Credit Card Fraud Detection System (CCFDS), Credit Card (CC), Gradient Boosting (GB), Machine Learning (ML).

I. INTRODUCTION

Fraudulent actions have been on the rise in a variety of businesses around the world, particularly in the field of the financial sector. Financial fraud (FF) is something when someone harms your financial health. The method that is applied to give you financial loss can be deceptive, misleading, or any other illegal practice. For any FF [1], detecting FF in the early stages is essential to minimize loss. Along with detecting any FF, reporting the fraud to the required and appropriate agencies is also advisable. We can observe various kinds of FF [2], such as identity theft, investment-related fraud, mortgage and lending fraud, mass-marketing fraud, and CC fraud (CCF). CCF is the most serious and prevalent in financial institutions, and it must be prevented or detected as quickly as possible. Fraud detection approaches must investigate and strictly handle CCF to limit its effects drastically. For detecting and preventing fraudulent transactions, various ML and artificial intelligence approaches give excellent results [3]. Using these approaches, previous transactions are used to train fraud detection systems to predict future fraud transactions. The primary issue in developing any efficient model that can be used to detect fraudulent transactions is datasets, as most available datasets are highly imbalanced. The number of fraudulent cases is substantially lower than in normal circumstances regarding fraud detection. In a skewed dataset, one class of dataset

ISSN(E): 2322-0821, ISSN(P): 2394-9910

contains a vast number of instances, while the other class has a minimal number of them. ML algorithms, on the other hand, operate best when there is a balanced distribution of classes. In the past years, different techniques have been used to address the issue of skewed datasets. The techniques that are used to generate balanced datasets are SMOTE or ADASYN tries to overcome the challenges of these techniques and can propose new methods to convert the imbalanced dataset to a balanced dataset.

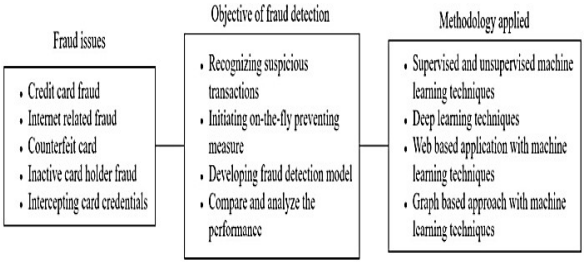


Fig. 1: Various Frauds

Bank crime does have a tremendous effect on the investment business as well as everyday life. Fraud is considered like any misleading activity committed by an individual with the intention of acquiring something illegally. In fraud, the offender deceives the victim to get an advantage or value. Most fraud happens in the real estate industry, especially when selling and buying, and in altering tax and insurance records. Such behaviour is unusual, although it is routinely done out by individuals, groups, and even companies. Fraud may affect industry trust, savings, and living costs. Financial institutions deploy several antifraud techniques. Fraudsters are adaptable and develop new ways to circumvent security safeguards. Economic crime occurs despite attempts by banks, law enforcement, and the government. Today's fraudsters may be a very creative, brilliant, and speedy group. This thesis compares fraud-detection methods like machine learning. It's also utilised to uncover hidden truths in large data sets [7].

Establishing a robust security system to protect theft in these CC transactions stillremains a challenge as it places hard earned money of the diligent workers at stake [8, 9]. In most cases, fraud happens when an unauthorized party obtains the credit card user's credentials and uses it to make transactions [10]. The financial institutions have placed various methods including artificial intelligence and machine learning based solutions for early frauddetection. Even Fintechcompanies and regulating authorities are hoping forward to some comprehensive and trustworthy solution for analysing CCF [11, 12].

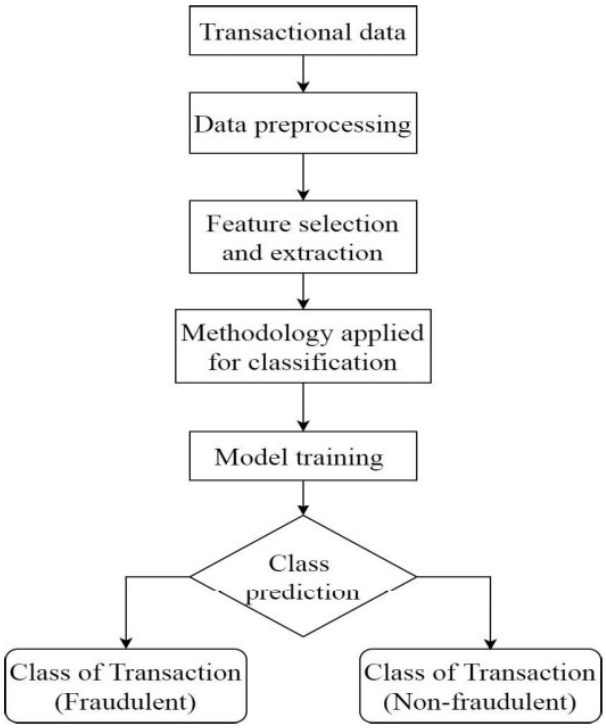


Fig. 2: Approach for Fraud

II. TYPES OF CARDS

Payment cards are the part of payment system i.e. issued by payment organization or bank. There are number of different type of cards available in market [13]. But most common card used by customer is either credit card or debit card. Consumer used it for payment purpose. It can be used either for physical payment or virtual payment. It removes the concept of carry paper money [14, 15].

Fig. 3: Type of cards

Advantage of digital cash is no need to carry cash and merchant cannot refuse to accept it. Following chart shows the various types of card available in Mauritius market. Each and every card has its unique features. Customer can issue card from legal organization as per their needs [16].

1. Credit card: It is a plastic material card, issued by bank or payment organization. It gives credit to customer for purchase goods or services. There is spending limit on card.

2. Debit card: It is plastic material card, issued by bank to their account holder. It is different from credit card, it directly withdrawal money from customer account. Normally used for cashless transaction.

3. Charge Card: Plastic cards are distinct from credit cards. It is provided to customers by a payment processor, and they are responsible for paying for the card. It permits client to do on the web or actual spending. Credit on the card is unlimited [17].

4. Brilliant Card: It is a chip based plastic card given by bank to their clients. There is no credit cap on this card. The cardholder will occasionally pay for their statement. Clients need to pay charges to card giving organization.

5. Business Card: A business card is similar to plastic material credit card. Card holder name, job title, business address, phone number, etc. information printed on card. Business card is used for business expenses at your home or abroad. It can be business debit card or credit card [18, 19].

6. Prepaid Card: It is not like credit or debit card. Unlike credit or debit card, for prepaid card u does not require a bank account. The amount on the card must be entered beforehand. You can burn through cash that is stacked in your card for installment. It is reusable card; ones stacked sum is utilized, customer can reload or toss the card [20].

7. Other Cards: Apart from all these types of cards, there are many different types of cards like cheque card, gift card, cash card, reward card, etc.

III. PROPOSED METHODOLOGY

However, implementation of these solutions still remains a challenge due to the practical problems associated with these transactions like the CC transactions data set is highly imbalanced, there is lack of sufficient data and large stake in false negative rates. Majorityof the machine learning models trained on such data sets perform well on accuracy-based performance criteria however, do not produce good ROC-AUC performance due to excessive false positive or false negative rates. The models tend to get biased towards the majority class, in such scenario missing even few transactions of minority class may be disastrous. Many researchers have been working in this area and finding new ways and techniques to mitigate this problem of occurrence of fraud in online transactions however, the fundamental problem for the process of detecting fraudulent CC transactions lies in resolving this issue.

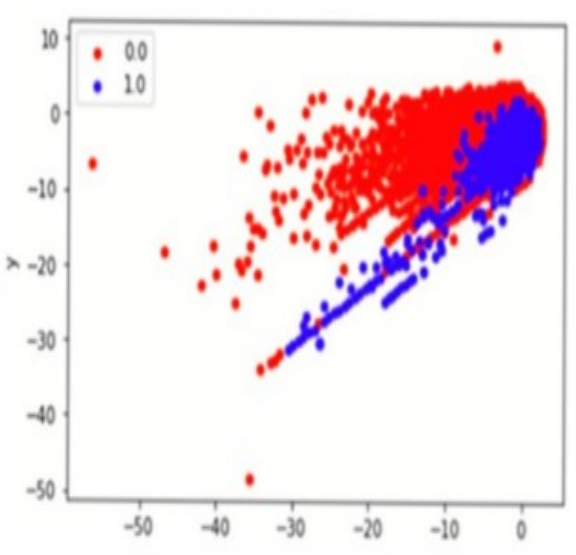


Fig. 4: Credit Card dataset

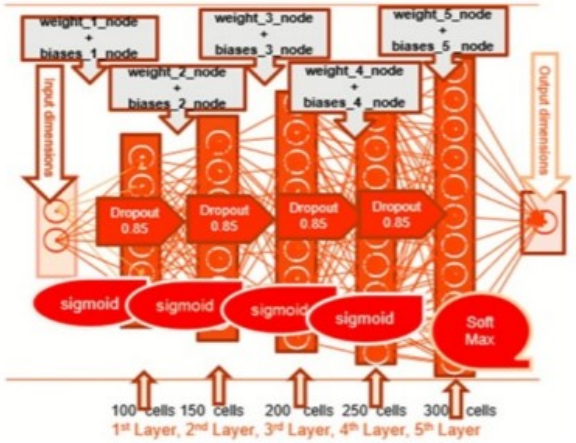


Fig. 5: Visual representation of AIFD

As first step of the research methodology, available literatures related to Supervised ML and its different classification techniques were studied in detail. Human input and output are necessary for supervised learning algorithms, as well as evaluation on prediction accuracy during the training process. It was studied that there is no requirement of training in unsupervised learning as it usually differs from supervised learning

approaches. However, supervised learning systems are easier to implement than unsupervised learning approaches. The thesis examines the supervised learning algorithms that are commonly employed in data classification. A comparative study of SML classification algorithms was performed on the above mentioned imbalanced credit card transactions dataset for retrieving the data from the data frame for finding fraudulent and legitimate transactions. From the results, it was observed that prediction accuracy score of LR and SVM is higher than the other classifiers used in the imbalanced training dataset. Aligning with the research objective and to identify a model to improve results and reduce False Negative Rate, an experimental model of first making the imbalanced dataset into a balanced data set by applying the technique of class balancing was developed. A classifier was trained on the dataset using a non-linear variant of support vector machine. Proposed model showed an improved AUC score. The extensive experiments reveal that the suggested strategy is reliable and can identify examples of suspicious transactions with higher objective functions than a conventional SVM methodology. The developed experimental model of class balancing was further improved by developing another selective hybrid class balancing technique with new RF algorithm to construct an optimized classifier and analyse the results. From this experimental result, it was observed that this proposed GB method performs better than the previous model and produces good AUC score value and optimizes False Negatives.

Major improvements in the current study begin with the limitations of the payment card database and its restrictions to certain banking institutions only.

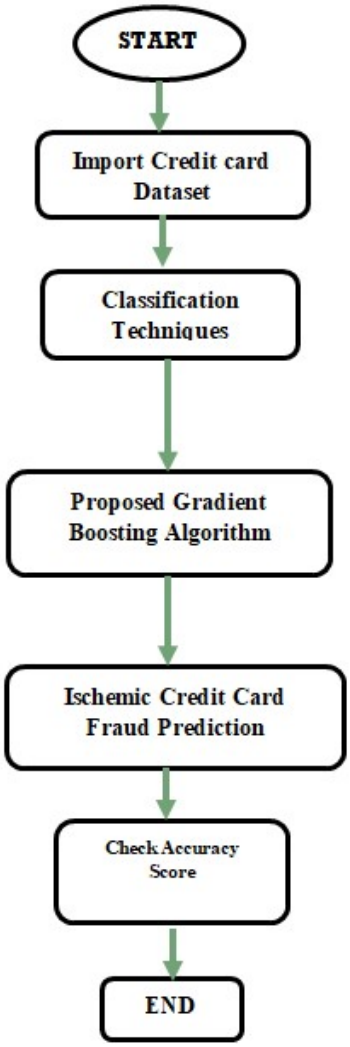


Fig. 6: Flow Chart of Proposed Methodology

Since the majority of the credit card transactions occur in Asia, the major drawback or limitation of real-world data challenges researcher to proceed further in solving the increasing fraud cases each day. What is required here on first hand is as much real-world data as possible, most probably from the locations outside Asia. It must also be noticed that collection of worldwide transactional data is itself not enough. The difference is also seen in the payment attributes as well as spending habits that usually differ from one consumer to the other and also with different locations.

Algorithm steps:

Input: DATA = {(x1,y1), (x2,y2),…….(xN,yN)},

$$L\{O(y),O(x)\}$$

Where: $L\{O(y),O(x)\}$ is the approximate loss function

Begin

Initialize: $$(x)=\frac{\arg\min}{w}\sum_{i=1}^{n}L(y_i,w)$$

$\boldsymbol{form}$=**1**:$\boldsymbol{M}$

Where, $r_{im} = \frac{\partial L(y_i, O(x_i))}{\partial O(x_i)}$

r_{im} = Training Data

r_{im} = Testing Data

$C_m(x)$

Train weak learner $\boldsymbol{C_m(x)}$ on training data

Calculate$\boldsymbol{w}$:

$$w_m = \arg\min \sum_{i=1}^{N} L(y_i, O_{m-1}(x_i) + wC_m(x_i))$$

$$O_m(x) = O_{m-1}(x) + wC_m(x)$$

Update:

End for

End

Output: $\boldsymbol{O_m(x)}$

IV. SIMULATION PARAMETER

While going through the literatures related to the subject, it has been identified that though there has been much work done by researchers in the field of 'false positive' however, there has been very little work done in the field of 'false negative'. Though the numbers of transactions that are passed as false negative are not that high in number, due to the checks and rules already placed by the systems, however the amount of transactions involved is significant. If a model is developed to identify these false negative transactions, then there will be huge savings in terms of money for both the banks and the merchants.

$$\Pr ecision = \frac{TP}{TP + FP} \times 100$$

Re*call* = (TPD/ TPD + FDP) x 100

Accuracy = (TPD + TDN /

TPD + TND + FPD + FND) x 100

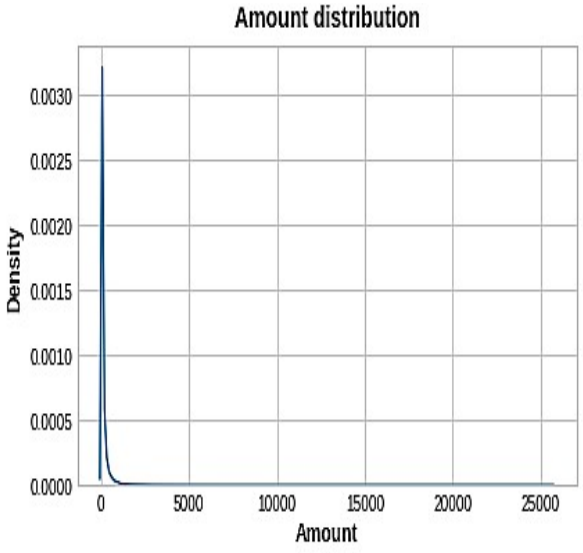
$$F1 - Score = \frac{2(\Pr ecision \times \mathrm{Re}\, call)}{\Pr ecision + \mathrm{Re}\, call} \times 100$$

V. SIMULATION RESULTS

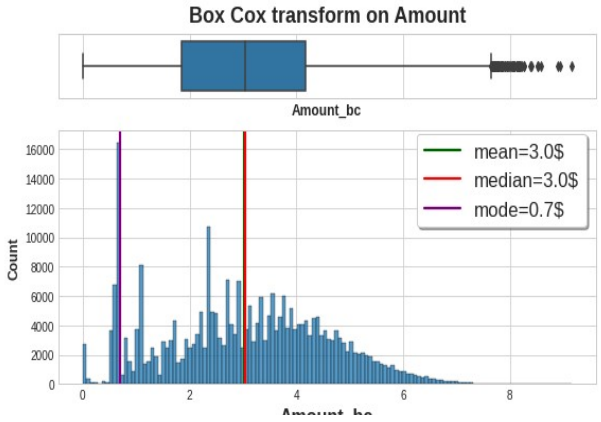
The foremost limitation that the researcher faced while doing this study was availability of a single dataset which limited in scope as well as in size. As a result, a limited number of classifiers, data sampling treatment methods, and ensemble methods could be applied. Further, the data set was masked due to privacy concerns, 28 out of 31 features in the data set were masked. This masking of dataset, restricted getting complete understanding of the various features of the dataset as well as understanding what impact these features might have created on the model's performance. Nonetheless, this limitation was advantageous which helped the researcher to perform the experimental method and further apply feature selection on the basis of the classifier used in the experiment.

Data preprocessing

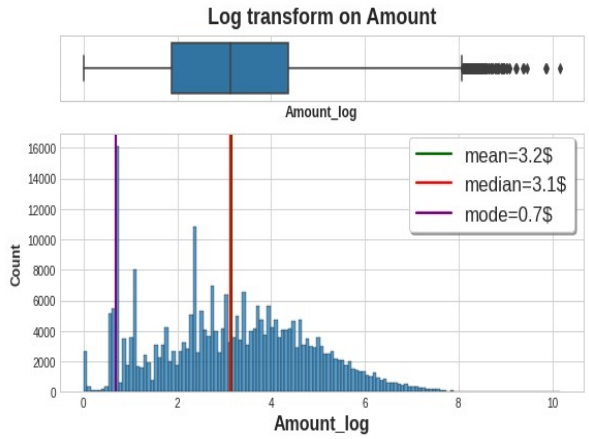
1. Check Duplicate values in pandas data frame and found 8063 duplicate data, all related to non fraudole nt transactions. I will drop them.
2. Amount Distribution



3. Box Cox and Log Transformation on Amount



ISSN(E): 2322-0821, ISSN(P): 2394-9910



Different ML Algorithm simulation parameters are shown in table 1. Recall, precision, F1-score and accuracy are represented by four ML algorithm in table 1. 84.6% of recall, 78% of precision. 69.9% of F1-Score and 97.7% of accuracy are achieved by Naïve Bayes. 61.2% of recall, 82.1% of precision. 70.1% of F1-Score and 98.8% of accuracy are achieved by LR. 70.4% of recall, 86.1% of precision. 77.5% of F1-Score and 98.8% of accuracy are achieved by K-NN. 75.7.2% of recall, 93.7% of precision. 91.6% of F1-Score and 99.3% of accuracy are achieved by GB.

Table 1: Simulation Result for Different Technique

Results	Naïve Bayes	Logistic Regression	K-NN	Proposed Technique
Recall	84.6%	61.2%	70.4%	75.7%
Precision	78%	82.1%	86.1%	93.7%
F1 Score	69.9%	70.1%	77.5%	91.6%
Accuracy	97.6%	98.8%	98.8%	99.3%

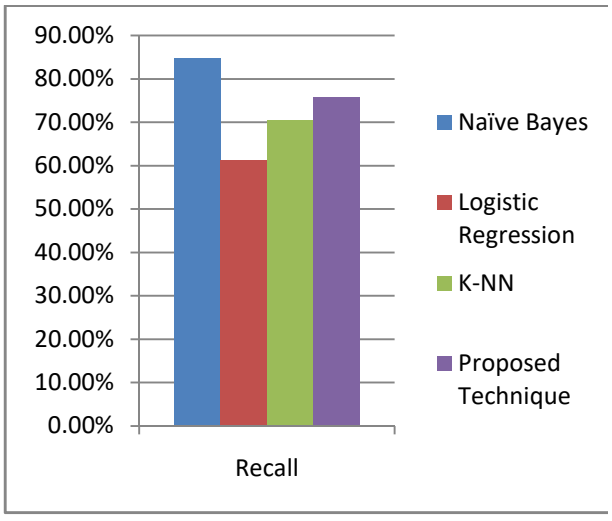


Figure 6: Graphical Represent of Recall

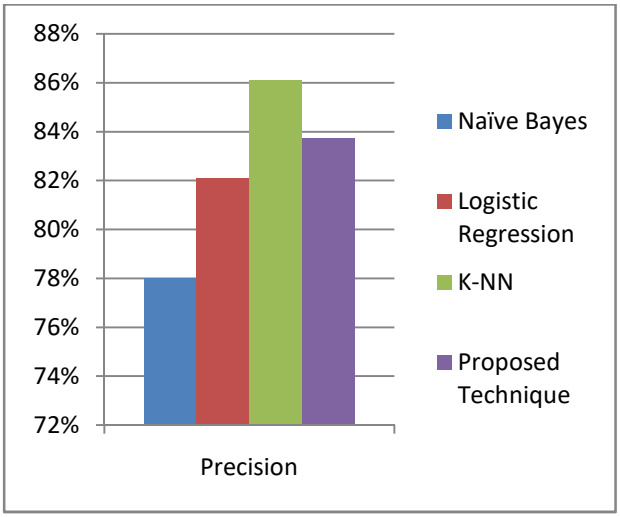


Figure 7: Graphical Represent of Precision

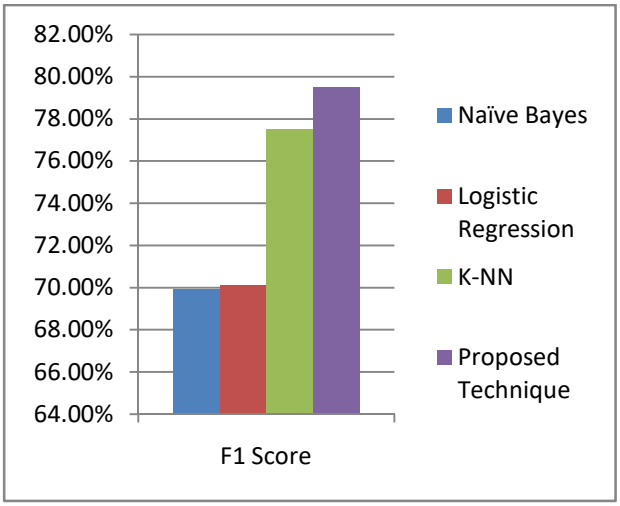


Figure 8: Graphical Represent of Recall

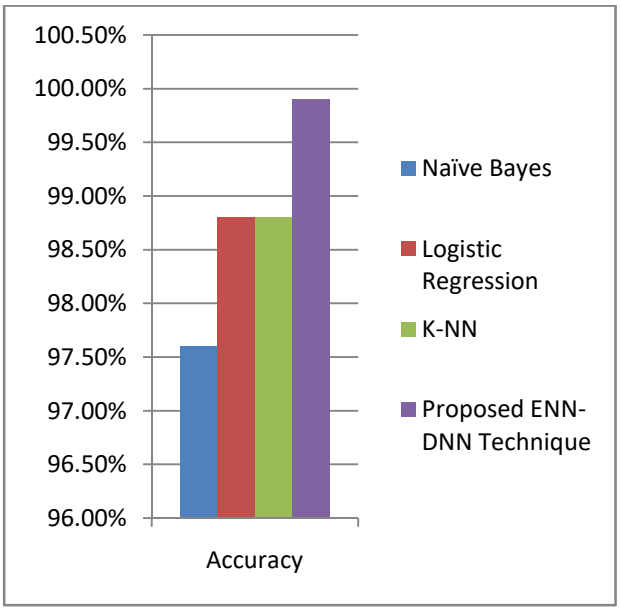


Figure 9: Graphical Represent of Accuracy

Comparison results for previous and proposed technique are displayed in table 2. CNN-SVM is facilitated by the use of CC data in the previous method. A 27.8% improvement in accuracy is achieved by the proposed GB technique against previous CNN-SVM. Figure 10 exhibits a bar graph depicting the accuracy of training and testing.

Table 2: Comparison Result

Results	Previous Technique	Proposed Technique
Recall	90.34%	75.7%
Precision	90.50%	93.7%
F1 Score	90.41%	91.6%
Accuracy	91.08%	99.3%

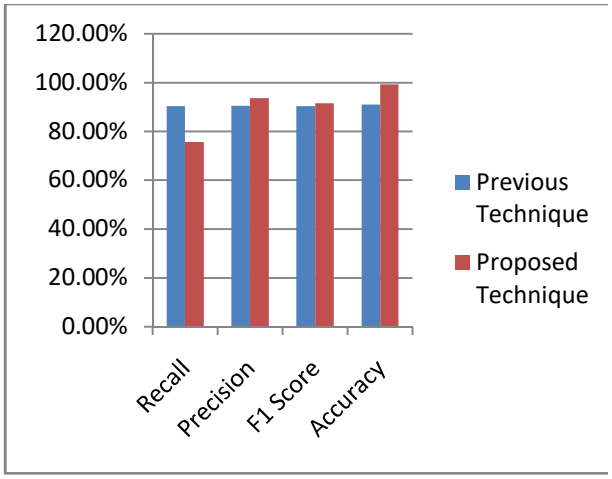


Figure 10: Graphical Represent of Accuracy

VI. CONCLUSION

Electronic payment has two options users either choose online or offline. From the simple analysis report, the most intended payment method for hacker is CC.

Today, for the business companies or banking sectors the CCF become one of the biggest issues that requires more safety and security. As technologies are increasing, facilities are increasing with that other hand CCF scams are also rising. Consumers wants secure channel to complete the transaction. There are many different types of CCF and we can detect the fraud transaction by available different fraud detection techniques.

The present model developed through the research will be able to identify the fraudulent transactions as well as the non-fraudulent transactions more accurately and efficiently in a supervised environment. However, the model can be considered as robust when it is able to perform equally efficiently, accurately and is able to predict similar accurate results in an un-supervised dataset as well as reinforcement environment. Hence, the future scope of work of this model relates to working upon and developing this model for an un-supervised dataset.

REFERENCES

[1] M J Madhurya, H L Gururaj, B C Soundarya, K P Vidyashree and A B Rajendra, "Exploratory analysis of credit card fraud detection using machine learning techniques", Global Transitions Proceedings 3 (2022) 31–37.

[2] JaberJemai, AnisZarrad and Ali Daud, "Identifying Fraudulent Credit Card Transactions Using Ensemble Learning", IEEE Access 2024.

[3] Fahdah A. Almarshad, GhadaAbdalazizGashgari and Abdullah I. A. Alzahrani, "Generative Adversarial Networks-Based Novel Approach for Fraud Detection for the European Cardholders 2013 Dataset", IEEE Access 2023.

[4] IbomoiyeDomorMienye, and Yanxia Sun, "A Deep Learning Ensemble With Data Resampling for Credit Card Fraud Detection", IEEE Access 2023.

[5] AyoubMniai ,MounaTarik , And Khalid JebariLma, Fstt, "A Novel Framework for Credit Card Fraud Detection", IEEE Access 2023.

[6] Ebenezer Esenogho, IbomoiyeDomorMienye, Theo G. Swart, KehindeAruleba and George Obaido, "A Neural Network Ensemble With Feature Engineering for Improved Credit Card Fraud Detection" IEEE Access 2022.

[7] Wang Ning, Siliang Chen, Songyi Lei and Xiongbin Liao, "AMWSPLAdaboost Credit Card Fraud Detection Method Based on Enhanced Base Classifier Diversity", IEEE Access 2023.

[8] AsmaCherif, ArwaBadhib, HeyfaAmmar, SuhairAlshehri, ManalKalkatawi, Abdessamad Imine, "Credit card fraud detection in the era of disruptive technologies: A systematic review", Journal of King Saud University – Computer and Information Sciences 35 (2023) 145–174

[9] Mustafa Abdul Salam, Khaled M. Fouad, Doaa L. Elbably, Salah M. Elsayed, "Federated learning model for credit card fraud detection with data balancing techniques", Neural Computing and Applications (2024) 36:6231–6256.

[10] Yuanming Ding, Wei Kang, JianxinFeng, Bo Peng, and Anna Yang, "Credit Card Fraud Detection Based on Improved

Variational Autoencoder Generative Adversarial Network", IEEE Access 2023.

[11] Rejwan Bin Sulaiman, VitalySchetinin and Paul Sant, "Review of Machine Learning Approach on Credit Card Fraud Detection", Human-Centric Intelligent Systems (2022) 2:55–68.

[12] Jonathan Kwaku Afriyie, Kassim Tawiah, Wilhemina AdomaPels, Sandra Addai-Henne, Harriet Achiaa Dwamena, Emmanuel OdameOwiredu, Samuel Amening Ayeh and John Eshun, "A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions", Decision Analytics Journal, Elsevier 2023.

[13] Petros Boulieris, John Pavlopoulos, Alexandros Xenos and Vasilis Vassalos, "Fraud detection with natural language processing", Machine Learning, Springer 2023.

[14] Fuad A. Ghaleb, Faisal Saeed, Mohammed Al-Sarem, Sultan Noman Qasem and Tawfik Al-Hadhrami, "Ensemble Synthesized Minority Oversampling-Based Generative Adversarial Networks and Random Forest Algorithm for Credit Card Fraud Detection", IEEE Access 2023.

[15] Fawaz Khaled Alarfaj, Iqra Malik, HikmatUllah Khan, NaifAlmusallam, Muhammad Ramzan and Muzamil Ahmed, "Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms", IEEE Access 2022.

[16] S Geetha, Yusuf Mohammed Khan, RohanSujay, SaiPavanYoganand and Rohan B, "Fraudulent URL and Credit Card Transaction Detection Syste m Using Machine Learning", International Conference on Advances in Electronics, Communication, Computing and Intelligent Information Systems (ICAECIS), IEEE 2023.

[17] Narendra Kumar, Kunal Tomar, Tushar Sharma, PiyushJyala, Dhruv Malik and Ishaan Dawar, "Customer behavior-based fraud detection of credit card using a random forest algorithm", International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1), IEEE 2023.

[18] Ana Jessica, Febi Vincent Raj and Janani Sankaran, "Credit Card Fraud Detection Using Machine Learning Techniques", 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN), IEEE 2023.

[19] Rupali Aggarwal, Pradeepta Kumar Sarangi and Ashok Kumar Sahoo, "Credit Card Fraud Detection: Analyzing the Performance of Four Machine Learning Models", International Conference on Disruptive Technologies (ICDT), IEEE 2023.

[20] A. Cherif A. Badhib H. Ammar S. Alshehri M. Kalkatawi and A. Imine "Credit card fraud detection in the era of Disruptive Technologies: A systematic review" Journal of King Saud University - Computer and Information Sciences vol. 35 no. 1 pp. 145-174 2023.

[21] P. Gupta A. Varshney M. R. Khan R. Ahmed M. Shuaib and S. Alam "Unbalanced credit card fraud detection data: A machine learning-oriented comparative study of balancing techniques" Procedia Computer Science vol. 218 no. 1 pp. 2575-2584 2023.

[22] SalomiHurriyaAnjum and GeetaPatil, "Cheat Detection for Credit Cards Using Artificial Intelligence", North Karnataka Subsection Flagship International Conference (NKCon), IEEE 2022.

[23] Aanchal Gupta, Kanishka Singh, Nonita Sharma and ManikRakhra, "Machine Learning For Detecting Credit Card Fraud", IEEE North Karnataka Subsection Flagship International Conference (NKCon), IEEE 2022.

[24] Divya Sharma and Sandeep Singh Kang, "Analysis of Credit card fraud detection techniques using Machine Learning", International Conference on Futuristic Technologies (INCOFT), IEEE 2022.

Conflict of Interest Statement: *The authors declare that there is no conflict of interest regarding the publication of this paper.*

Cite this Article
Shivam Singh et.al. Credit Card Fraud Detection using Gradient Boosting Machine Learning Techniques. International Research Journal of Engineering & Applied Sciences (IRJEAS). 12(3), pp. 0 7 - 1 5 , 2024.
https://doi.org/10.55083/irjeas.2024.v12i03002