Credit Card Fraud Transaction Detection Using Outlier Detection Models With Neural Network

Megha Nayak[1], Prof. Satendra Sonare[2]
[1, 2] Dept of CSE
[2]Professor, Dept of CSE
[1, 2] Gyan Ganga Institute of Technology and Sciences, Jabalpur, Madhya Pradesh, India.

Abstract- *Recent developments in e-commerce and telecommunications have increased the use of credit cards in both online and daily transactions. However, credit card fraud is on the rise, causing huge losses for financial institutions every year. Developing effective fraud detection systems is critical to mitigating these losses, but is difficult as most credit card datasets are highly unstable. In addition, credit card fraud using traditional machine learning algorithms is ineffective because its design includes a static map from input vectors to output vectors. As a result, they cannot change the purchasing habits of their credit card customers.*

This paper presents an effective method for credit card fraud detection using a neural network classifier and a data resampling method. The cluster group was adopted using neural network as a learner based on adaptive learning. The effectiveness of the proposed method is demonstrated using publicly available real-world credit card transaction datasets. The performance of the proposed approach is benchmarked against the following algorithms: Logistic Regression, support vector machine (SVM), multilayer perceptron (MLP), decision tree, traditional XGBoost and MLP.

Keywords- Detection of fraud; tracking of fraud; Fraud transaction understanding, Neural Network, Adaptive Learning.

I. INTRODUCTION

The volume of electronic exchange has risen altogether in later a long time, due to the popularization of e-commerce such as online retailers (e.g., Amazon, Ebay and Alibaba). Credit/debit cards are broadly utilized in electronic exchange. As of late, cardless exchanges [1] in credit card operations ended up increasingly prevalent by web installment doors (e.g. Paypal and Alipay). The worldwide e-commerce showcase is anticipated that it'll be worth a stunning us$ 24 trillion by 2019 [2]. In any case, there has been a concurrent development in false exchanges [3] which comes about in a emotional affect on the economy. A overview of over 160 companies uncovers that the number of online fakes is 12 times higher than that of offline fakes ever year [4]. Since a physical card isn't required within the e-commerce situation and as it were the data almost the card is sufficient for a exchange, a fraudster as it were needs this data for a extortion. For case, after the fraudster takes the account and watchword of a card from its honest to goodness cardholder, they utilize them to buy a few products. Fraudsters more often than not get card data by means of a assortment of ways: capture attempt sends containing recently issued cards, replicating and reproducing card data through skimmers, or gathering delicate data through phishing (cloned websites) or from unscrupulous representatives of credit card companies [5]. Due to the complexity of the environment and people's foundation, it is difficult to avoid all the veritable cardholder' account from being stolen. The promising way to identify this kind of extortion is to analyze the expending designs on each account and to figure out any errors with regard to the "usual" exchange designs [6].

Credit card extortion (CCF) could be a sort of personality burglary in which somebody other than the proprietor makes an illegal exchange employing a credit card or account subtle elements. A credit card that has been stolen, misplaced, or falsified might result in extortion. Card-not-present extortion or the utilize of your credit card number in e-commerce exchanges has moreover ended up progressively common as a result of the increment in online shopping. Expanded extortion, such as CCF, has brought about from the extension of e-banking and a few online installment situations, coming about in yearly misfortunes of billions of dollars. In this era of computerized installments, CCF discovery has ended up one of the foremost imperative objectives. As a trade proprietor, it cannot be debated that long haul is heading towards a cashless culture. As a result, normal installment strategies will now not be utilized within the future, and thus they will not be accommodating for extending a trade. Clients will not continuously visit the commerce with cash in their pockets. They are presently putting a premium on charge and credit card installments. As a result, companies will have to be upgrading their environment to ensure that they can take all sorts of installments. Within another a long time, this circumstance is anticipated to end up much more extreme [7].

In 2020, there were 393,207 cases of CCF out of around 1.4 million add up to reports of personality burglary [8]. CCF is presently the moment most predominant sort of character burglary recorded as of this year, as it were taking after government archives and benefits extortion [9]. In 2020, there were 365,597 frequencies of extortion executed utilizing unused credit card accounts [10]. The number of character burglary complaints has climbed by 113% from 2019 to 2020, with credit card character robbery reports expanding by 44.6% [11]. Installment card burglary taken a toll the worldwide economy $24.26 billion final year. With 38.6% of detailed card extortion misfortunes in 2018, the joined together states are the foremost defenceless nation to credit robbery. As a result, budgetary educate should prioritize preparing themselves with a mechanized extortion location framework.

The objective of administered CCF discovery is to make a machine learning (ml) show based on existing value-based credit card installment information. The show ought to recognize between false and no false exchanges, and utilize this data to choose whether an approaching exchange is false or not. The issue includes a assortment of principal issues, counting the system's fast response time, fetched affectability, and include pre-processing. Ml may be a field of artificial insights that employments a computer to create forecasts based on earlier information patterns ml models have been utilized in numerous considers fathoming various challenges. Profound learning (dl) calculations connected applications in computer organize, interruption discovery, managing an account, protections, portable cellular systems, wellbeing care extortion discovery, restorative and malware discovery, discovery for video observation, area following, android malware discovery, domestic robotization, and heart illness forecast.

We investigate the practical application of ml, especially dl calculations, to recognize credit card burglaries within the keeping money industry. For information categorisation challenges, the back vector machine (svm) may be a directed ml method. It is utilized in a assortment of spaces, counting picture acknowledgment [12], credit rating [9], and public security [13]. Svm can handle direct and nonlinear twofold classification issues, and it finds a hyperplane that isolates the input data within the bolster vector, which is predominant to other classifiers. Neural systems were the primary strategy utilized to distinguish credit card robbery within the past [8]. As a result, (dl), a department of ml, is currently focused on dl approaches. In later a long time, profound learning approaches have gotten critical consideration due to considerable and promising results in different applications, such as computer vision, normal dialect handling, and voice. In any case, as it were some thinks about have inspected the application of profound neural systems in distinguishing CCF [7]. It employments a number of profound learning calculations for recognizing CCF. In any case, in this think about, we select the NN show and its layers to decide in case the first extortion is the typical exchange of qualified datasets. A few exchanges are common in datasets that have been named false and illustrate flawed exchange conduct. As a result, we center on directed and unsupervised learning in this investigate. The lesson awkwardness is the problem in ml where the entire number of a lesson of information (positive) is distant less than the entire number of another lesson of information (negative). The classification challenge of the uneven dataset has been the subject of a few considers. A broad collection of ponders can give a few answers. Hence, to the leading of our information, the issue of course lopsidedness has not however been fathomed. We propose NN show by including the extra layers for highlights extraction and the classification of credit card exchanges as false or something else. The best traits from the arranged dataset are positioned utilizing highlight determination procedures. After that, CCF is classified utilizing a few directed machine-driven and profound learning models.

II. RELATED WORK

The progress in advances such as e-commerce and budgetary innovation (fintech) applications has started an increment within the number of online card exchanges that happen on a everyday premise. As a result, there has been a spike in credit card extortion that influences card issuing companies, shippers, and banks. It is subsequently fundamental to create components that guarantee the security and judgment of credit card exchanges. In inquire about [14], we actualize a machine learning (ml) based system for credit card extortion location employing a genuine world imbalanced datasets that were produced from European credit cardholders. To illuminate the issue of lesson awkwardness, we re-sampled the dataset utilizing the engineered minority over-sampling procedure (destroyed). This system was assessed utilizing the taking after ml strategies: back vector machine (svm), calculated relapse (lr), irregular woodland (RF), extraordinary angle boosting (XGBoost), choice tree (DT), and additional tree (ET). These ml calculations were coupled with the versatile boosting (AdaBoost) procedure to extend their quality of classification. The models were assessed utilizing the exactness, the review, the accuracy, the Matthews relationship coefficient (mcc), and the range beneath the bend (auc). Additionally, the proposed system was actualized on a exceedingly skewed engineered credit card extortion dataset to advance approve the comes about that were obtained in this investigate. The test results illustrated that utilizing the AdaBoost includes a positive effect on the execution of the

proposed strategies. Assist, the comes about gotten by the boosted models were predominant to existing methods.

Recent progressions in electronic commerce and communication frameworks have altogether expanded the utilize of credit cards for both online and standard exchanges. Be that as it may, there has been a consistent rise in false credit card exchanges, costing budgetary companies tremendous misfortunes each year. The improvement of successful extortion location calculations is crucial in minimizing these misfortunes, but it is challenging since most credit card datasets are profoundly imbalanced. Moreover, utilizing routine machine learning calculations for credit card extortion discovery is wasteful due to their plan, which includes a inactive mapping of the input vector to yield vectors. Subsequently, they cannot adjust to the energetic shopping behavior of credit card clients. This paper [15] proposes a productive approach to identify credit card extortion employing a neural arrange gathering classifier and a crossover information resampling strategy. The ensemble classifier is gotten employing a long short-term memory (lstm) neural organize as the base learner within the versatile boosting (AdaBoost) procedure. In the interim, the cross breed resampling is accomplished utilizing the engineered minority oversampling strategy and altered closest neighbor (smote-enn) strategy. The viability of the proposed strategy is illustrated utilizing freely accessible real-world credit card exchange datasets. The execution of the proposed approach is benchmarked against the taking after calculations: back vector machine (svm), multilayer perceptron (mlp), choice tree, conventional AdaBoost, and lstm. The test comes about appear that the classifiers performed way better when prepared with the resampled information, and the proposed lstm gathering outflanked the other calculations by getting a affectability and specificity of 0.996 and 0.998, respectively.

The progress in advances such as e-commerce and monetary innovation (fintech) applications has started an increment within the number of online card exchanges that happen on a everyday premise. As a result, there has been a spike in credit card extortion that influences card issuing companies, vendors, and banks. It is in this manner basic to create components that guarantee the security and astuteness of credit card exchanges. In this investigate [16], we actualize a machine learning (ml) based system for credit card extortion location employing a genuine world imbalanced datasets that were created from European credit cardholders. To illuminate the issue of course awkwardness, we re-sampled the dataset utilizing the manufactured minority over-sampling strategy (destroyed). This system was assessed utilizing the taking after ml strategies: back vector machine (svm), calculated relapse (lr), arbitrary timberland (RF), and extraordinary angle boosting (XGBoost), choice tree (DT), and additional tree (AT). These ml calculations were coupled with the versatile boosting (AdaBoost) strategy to extend their quality of classification. The models were assessed utilizing the precision, the review, the accuracy, the Matthews relationship coefficient (MCC), and the zone beneath the bend (AUC). Additionally, the proposed system was actualized on an exceedingly skewed engineered credit card extortion dataset to encourage approve the comes about that were obtained in this investigate. The exploratory results illustrated that utilizing the AdaBoost features a positive effect on the execution of the proposed strategies. Assist, the comes about gotten by the boosted models were predominant to existing strategies.

Credit cards play an essential role in today's digital economy, and their usage has recently grown tremendously, accompanied by a corresponding increase in credit card fraud. Machine learning (ML) algorithms have been utilized for credit card fraud detection. However, the dynamic shopping patterns of credit card holders and the class imbalance problem have made it difficult for ML classifiers to achieve optimal performance. In order to solve this problem, this paper [17] proposes a robust deep-learning approach that consists of long short-term memory (LSTM) and gated recurrent unit (GRU) neural networks as base learners in a stacking ensemble framework, with a multilayer perceptron (MLP) as the meta-learner. Meanwhile, the hybrid synthetic minority oversampling technique and edited nearest neighbor (SMOTE-ENN) method is employed to balance the class distribution in the dataset. The experimental results showed that combining the proposed deep learning ensemble with the SMOTE-ENN method achieved a sensitivity and specificity of 1.000 and 0.997, respectively, which is superior to other widely used ML classifiers and methods in the literature.

Modern propels in electronic commerce frameworks and communication advances have made the credit card the possibly most prevalent strategy of installment for both customary and online buys; in this way, there's altogether expanded extortion related with such exchanges. False credit card exchanges taken a toll firms and shoppers expansive money related misfortunes each year, and fraudsters persistently endeavor to discover modern innovations and strategies for committing false exchanges. The discovery of false exchanges has gotten to be a noteworthy calculates influencing the more noteworthy utilization of electronic payment. In this way, there's a require for proficient and viable approaches for identifying extortion in credit card exchanges. This paper proposes and cleverly approach for recognizing extortion in credit card exchanges utilizing an optimized light angle boosting machine (OLightGBM). Within the proposed approach [18], a Bayesian-based hyperparameters

optimization calculation is intellectual's coordinates to tune the parameters of a light angle boosting machine (LightGBM). To illustrate the adequacy of our proposed OLightGBM for identifying extortion in credit card exchanges, tests were performed utilizing two real-world open credit card exchange information sets comprising of false exchanges and authentic ones. Based on a comparison with other approaches utilizing the two information sets, the proposed approach outflanked the other approaches and accomplished the most elevated execution in terms of exactness (98.40%), region beneath collector working characteristic bend (auc) (92.88%), exactness (97.34%) and f1-score (56.95%).

Widespread utilize of web moreover had the considerable affect on the increment of the online card transactions especially with the starting of the final decade. At the side the increment of online exchanges, the around the world keeping money segment was constrained to bargain with or to come across an unexpected number of false exercises, however. Consequently, rule-based frameworks were outlined to stamp the high-risk exchanges and let the specialists to affirm the false nature of such exchanges. As a countermeasure, inactive nature of rule-based frameworks was abused by the most recent assaults to go undetected. Hence, analysts pointed at planning versatile extortion location frameworks utilizing basically machine learning strategies with the exceptionally later application of profound learning. Be that as it may, they were centered on detecting fraudulent exercises but, to the finest of our information, none of them dug into the superior understanding the characteristics of false card exchanges in arrange to deliver more versatile models. Hence, in this think about, we built the greatest information set ever utilized in a investigate, comprising of 4b non-fraud and 245k extortion exchanges contributed to by the 35 banks in turkey. Subsequently, we present and look at the execution of profile-based extortion discovery models, specifically card-type based show, exchange characteristics based demonstrate, and amount-based demonstrate. Too, we made transient and spatial examination on our information set to appear the strength of the proposed models against maturing and zero-day attacks.

III. PROPOSED WORK

The proposed method to detect the fraudulent financial statements using the credit card dataset with various models such as Logistic Regression, Random Forest, XGBoost Model (GBM), SVM, MLP and Neural Network Model. The purpose of the thesis is to construct a valid and rigorous fraudulent financial statement detection model. Analysis of financial details is one of those simple methods to identify frauds. The aim of the research is to distinguish financial details, the values of which could indicate the fraud in financial statements. Moreover, the Multi-Layer Neural Network model of fraud detection in financial statements has been developed.

In the first stage, for the variable selection process appropriate deviations are used with certain constrains. The preprocessing is applied to remove the unwanted information.

In second stage Outlier Detection models Like IQR Method, Standard Deviation method and Z-score methods are applied to detect and remove the outliers present in the dataset.

In the third stage Random Under sapling is applied to resample the dataset. Because after removal of the outliers the dataset reached to imbalanced state. In such situation the classifiers will give priority to higher element types while training the models which will impact the accuracy of the models.

In final stage the models are trained using Random forest or random decision forest method will be used which is an ensemble learning method for classification, regression and other types like SVM, MLP and neural network model.

3.1 Modules Description

A. Random Forest

Random forests is an ensemble learning method for classification and regression, that operate by constructing a multitude of decision tree at training time and outputting the class that is the mode of the class or mean prediction of the individual trees .Random forest uses full decision trees. It has low bias and high variance where variance can be defined as the error from the small fluctuations in the training data set and bias can be defined as the error from the spontaneous assumptions. The prediction describes a method of building a forest of uncorrelated trees using a CART like procedure, combined with randomized node optimization and bagging.
Important points are,

The algorithm working is explained as follows,

1. The input for the Random Forest Algorithm will be the credit card dataset.
2. The data will be split into training and test data. Then, the Random Forest model is constructed and then the testing of data follows.
3. The confusion matrix will be built from which the accuracy will be calculated.

B. Gradient Boosting Machine

Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. The trees will be built with the help of previously built trees.

The algorithm working is explained as follows,

1. The credit card dataset will be given as the input.
2. The data will be split into three parts such as train, test and valid data.
3. In training data, the model will be built. In test data, the output will be predicted with the help of training data. The validation data will be used to validate whether the given output is valid or not.

C. Artificial Neural Network

Neural computing refers to a pattern recognition methodology for machine learning. The resulting model from neural computing is often called an Artificial Neural Network (ANN) or a neural network. Figure 5.1 represents that the ANN is composed of multiple nodes such as X1, X2…Xi, The result of these operations is passed to other neurons.

The output at each node is called its activation or node value**.** Each link is associated with weight W1j, W2j…Wij. ANNs are capable of learning, which takes place by altering weight values. Neural networks have been used in many business applications for pattern recognition, forecasting, prediction, and classification. Neural network computing is a key component of any data mining tool kit.

3.2 Proposed Algorithm

The algorithm working is explained as follows,

1. The credit card dataset will be given as the input for Artificial Neural Network.
2. The data will be split into three parts such as train, test and valid data.
3. Tuning will be performed to reduce the error rate.
4. In training data, the model will be built. In test data, the output will be predicted with the help of training data. The validation data will be used to validate whether the given output is valid or not.
5. The number of hidden layers is increased to predict the improved accuracy.

IV. RESULTS WORK

4.1 Performances-Evaluation Measures

Traditional methods of estimating ML classifiers can use confusion metrics relating to the difference between the rock bottom dataset truth and the model's prediction where TP, TN, FP, and FN denote true positive, true negative, false positive and false negative, respectively.

i) ACCURACY: Accuracy is used to measure the performance in the evidence domain recovery and processing of the data. The fraction of the results that are successfully classified can be represented by equation (9) as follows:
Accuracy = (TP + TN)/ (TP + FP + TN + FN)

ii) PRECISION: Precision is a performance assessment that measures the ratio of correctly identified positives and the total number of identified positives. This can be seen as follows:
Precision = TP/ (TP + FP).

iii) F-MEASURE/F1-SCORE: The f-measure considers both the precision and the recall. The f-measure may be assumed to be the average weight of all values, which can be seen as follows:
F = (2X precision × Recall)/ (precision + Recall).

iv) RECALL: The recall is also referred to as the sensitivity, which is the ratio of connected instances retrieved over the total number of retrieved instances and can be seen as follows:
Recall = TP/ (TP + FN).

Table 4.1: Accuracy Comparisons.

Implemented Model	**Accuracy in Percentage**
Logistic Regression	94.93
Support Vector Machine	94.93
Random Forest Classifier	94.93
XGBoost Classifier	95.94
MLP Classifier	93.91
Multilayer NN	96.05

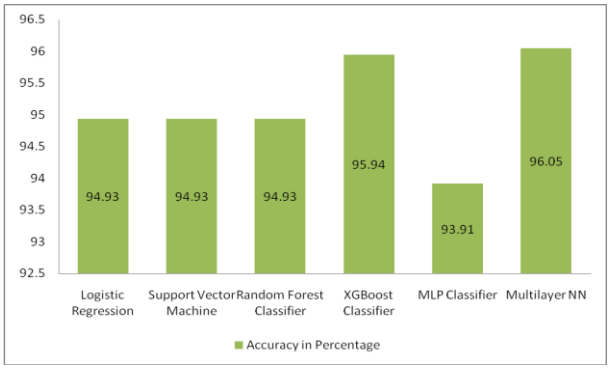

Figure 4.1: Accuracy Comparison Chart.

REFERENCES

[1] Gupta S and Johari R. A new framework for credit card transactions involving mutual authentication between cardholder and merchant. In Communication Systems and Network Technologies, pages 22–26. IEEE, 2011.

[2] C. Arun. Fraud: 2016 & its business impact. Technical report, 11 2016.

[3] Vronique Van Vlasselaer, Cristin Bravo, Olivier Caelen, Tina Eliassi-Rad, Leman Akoglu, Monique Snoeck, and Bart Baesens. Apate: A novel approach for automated credit card transaction fraud detection using network-based extensions. Decision Support Systems, 75:38–48, 2015.

[4] Suvasini Panigrahi, Amlan Kundu, Shamik Sural, and A. K. Majumdar. Credit card fraud detection: A fusion approach using dempstercshafer theory and Bayesian learning. Information Fusion, 10(4):354–363, 2009.

[5] Jon T. S Quah and M Sriganesh. Real-time credit card fraud detection using computational intelligence. Expert Systems with Applications an International Journal, 35(4):1721–1732, 2007.

[6] Abhinav Srivastava, Amlan Kundu, Shamik Sural, and Arun Majumdar. Credit card fraud detection using hidden markov model. Dependable & Secure Computing IEEE Transactions on, 5(1):37–48, 2007.

[7] Y. Abakarim, M. Lahby, and A. Attioui, ‘‘an efficient real time model for credit card fraud detection based on deep learning,’’ in Proc. 12th Int. Conf. Intell. Systems: Theories Appl., Oct. 2018, pp. 1–7, doi: 10.1145/3289402.3289530.

[8] A. O. Balogun, S. Basri, S. J. Abdulkadir, and A. S. Hashim, ‘‘Performance analysis of feature selection methods in software defect prediction: A search method approach,’’ Appl. Sci., vol. 9, no. 13, p. 2764, Jul. 2019, doi: 10.3390/app9132764.

[9] B. Bandaranayake, ‘‘Fraud and corruption control at education system level: A case study of the Victorian department of education and early childhood development in Australia,’’ J. Cases Educ. Leadership, vol. 17, no. 4, pp. 34–53, Dec. 2014, doi: 10.1177/1555458914549669.

[10] V. N. Dornadula and S. Geetha, ‘‘Credit card fraud detection using machine learning algorithms,’’ Proc. Comput. Sci., vol. 165, pp. 631–641, Jan. 2019, doi: 10.1016/j.procs.2020.01.057.

[11] K. He, X. Zhang, S. Ren, and J. Sun, ‘‘Deep residual learning for image recognition,’’ 2015, arXiv: 1512.03385

[12] D. Molina, A. LaTorre, and F. Herrera, ‘‘SHADE with iterative local search for large-scale global optimization,’’ in Proc. IEEE Congr. Evol. Comput. (CEC), Jul. 2018, pp. 1–8, doi: 10.1109/CEC.2018.8477755.

[13] J. Kim, H.-J. Kim, and H. Kim, ‘‘Fraud detection for job placement using hierarchical clusters-based deep neural networks,’’ Int. J. Speech Technol., vol. 49, no. 8, pp. 2842–2861, Aug. 2019, doi: 10.1007/s10489-019-01419-2.

[14] E. Ileberi, Y. Sun and Z. Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost," in IEEE Access, vol. 9, pp. 165286-165294, 2021, doi: 10.1109/ACCESS.2021.3134330.

[15] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba and G. Obaido, "A Neural Network Ensemble With Feature Engineering for Improved Credit Card Fraud Detection," in IEEE Access, vol. 10, pp. 16400-16407, 2022, doi: 10.1109/ACCESS.2022.3148298.

[16] E. Ileberi, Y. Sun and Z. Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost," in IEEE Access, vol. 9, pp. 165286-165294, 2021, doi: 10.1109/ACCESS.2021.3134330.

[17] I. D. Mienye and Y. Sun, "A Deep Learning Ensemble With Data Resampling for Credit Card Fraud Detection," in IEEE Access, vol. 11, pp. 30628-30638, 2023, doi: 10.1109/ACCESS.2023.3262020.

[18] A. A. Taha and S. J. Malebary, "An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine," in IEEE Access, vol. 8, pp. 25579-25587, 2020, doi: 10.1109/ACCESS.2020.2971354.