DETECTION OF CREDIT CARD FRAUD TRANSACTION USING HYBRID MACHINE LEARNING ALGORITHMS

[1]BHUKYA DHARMA, [2]DR.D. LATHA

[1]Research Scholar , Department of Computer Science and Engineering, Adikavi Nannaya University, Rajamahendravaram, India
[2] Asst. Professor, Department of Computer Science and Engineering, Adikavi Nannaya University, Rajamahendravaram, India

Email: dharmaaknu9@rediffmail.com

ABSTRACT

In developed countries, credit card transactions are now the main method of payment, and their utility is growing quickly in developing. As a result, frauds are becoming a more serious issue, resulting in financial losses and a decline in customer trust. Firstly, the both real and fraudulent actors continually change their conduct, and secondly, is that datasets are wildly biased. There have been several suggestions for methods to handle the increasing number of credit card fraud transactions. To effectively identify fraudulent transactions, there has been use of machine learning techniques. This analysis explains the way to detect credit card fraud using a hybrid machine learning algorithm. The dataset utilized in September 2013 was the record of credit card transactions done by European cardholders over a duration of two days. Random Forest (RF) and Support Vector Machine (SVM) machine learning models are combined in hybrid categorization. The results of the Hybrid Machine Learning model are based on Accuracy, Sensitivity, Specificity, and Precision. Described model achieves Accuracy as 98%, Sensitivity as 96%, Specificity as 97%, and Precision as 96%. The outcomes from using the hybrid classification model have shown to be much more successful than those from using separate classification methods

Keywords: *Hybrid Machine Learning, Credit Card (CC), Fraud transactions, SVM, RF, Precision, Accuracy.*

I. INTRODUCTION

Financial fraud is a persistent problem that has far-reaching effects on the financial industry, corporate organizations, and government [1]. Credit card transactions are increasing significantly as business moves towards e-commerce.

In growing economies like India, cashless transactions and digital payments are increasing. Legitimate and fraudulent transactions fall into two types [2]. Due to several weaknesses in the developed e-banking systems, the fraudulent transactions increased. The majority of transaction data is made up of several properties, such as the recipient, the transaction's value, and the credit card's number. Since transaction records are frequently characterized by a high number of samples, multiple dimensions, and online updates, automatic algorithms are crucial since human analysts cannot always easily or quickly identify fraudulent trends in these datasets [3]. Additionally, whenever it regards reporting card theft, loss, or fraudulent use, the cardholder is not reliable. There are two types of card fraud: interior and exterior fraud. If a bank employee is linked to a client who used a fake identity, there has been internal fraud. Credit cards that have been stolen and used by fraudsters to make purchases constitute outer fraud. The two most effective ways for preventing fraud are detection and prevention [4]. Preventing attacks from fraudsters by acting as an additional layer of defense. Once prevention has failed, detection is carried out. [5]. Therefore, detection makes it possible to identify and warn as soon as a fraudulent transaction is initiated.

Both retailers and customers are experiencing financial loss as a result of financial theft involving credit card transactions. It is a major problem; banks card manufacturing companies have to spend a lot of money to fix [6]. However, they can't ignore the financial losses, which also rise with e-commerce. E-payment is made quick, convenient, seamless, and simple to use through online purchases and payment services. It invites thieves to take part in an innovative type of fraud. Banks and organizations use effective security measures to deal with these problems, but fraudsters constantly adapt their delicate methods. Therefore, it is crucial to improve detection and preventive methods.

Monitoring and analyzing the transactional conduct of various users in order to estimate the identification of undesired behaviour is necessary for fraud detection [7]. They are interested in learning about the various technologies, algorithms, and types used to identify credit card fraud in order to successfully detect.

If suitable precautions are taken and research is done on the behavior of fraudulent operations, the likelihood of this abuse occurring in the future may be reduced. Another way to put it is when someone uses someone else's credit or debit card for their personal gain while the cardholder and the individuals that authorized it don't know anything about the company. The development of computational algorithms that can identify fraudulent transactions based on the volume and timing of those transactions uses machine learning techniques [8]. Algorithms based on machine learning identify future records observed in the domain by using a large amount of example data from the underlying domain [9]. For the algorithms in the class of supervised learning algorithms, the example data classes must first be labelled. Other classes
of algorithms, but depend on unsupervised learning, in which the input is grouped into identical units and designated as belonging to a single class [10].

The selection of an algorithm for the model is dependent on the performance of each algorithm under classification. The selection of the wrong algorithm can result in overfitting or underfitting. The balance of bias and variance are the driving forces behind the selection of the algorithm. In this analysis, credit card fraud transactions are detected using a hybrid machine learning model. This type of technology aims to stop fraud by identifying fraudulent transactions before they are committed to the database. A hybrid fraud detection system should also reduce false positives, which interrupt normal transactions and are inconvenient for the end user. This technique is advantageous both for the organization and for the customer.

The remaining sections are arranged as follows: Literature review details are provided in Section II. In Section III, the methodology of the hybrid machine learning algorithm process is discussed. The performance analysis of credit card fraud transaction detection is presented in Section IV, and Section V generates the work to a conclusion.

2. LITERATURE SURVEY

Lebichot B., He-Guelton L., Verhelst T., Oblé F. Le Borgne Y. -A., and Bontempi G., et. al. [11] addresses the development, use, and for the purpose of identifying credit card fraud in online transactions, transfer learning algorithms are evaluated. They provide two contributions: first, they demonstrate that the quantity of labelled samples in the target domain has a significant impact on many transfer techniques performance, and second, On the basis of self-supervised and semi-supervised domain adaptation classifiers, they propose an ensemble method to this problem. The extensive experimental evaluation demonstrates that this approach is not only very accurate but also not exactly dependent on the quantity of labelled samples.

S. Han, K. Zhu, M. Zhou and X. Cai, et. al. [12] Any intelligent optimizers are now available, can be benefit from the suggested strategy to increase their efficiency while solving MMOPs (Multimodal Multiobjective Optimization Problems). This is supported by experimental findings from resolving 12 scalable unbalanced distance minimization issues and 22 of these Congress on Evolutionary Computation (CEC) 2019 challenges using a number of optimizers. In order to demonstrate the suggested method's usefulness, they finally apply it to challenges involving credit card fraud detection. Z. Li, G. Liu and C. Jiang, et. al. [13] concentrate on using a deep neural network's loss function to build deep feature representations of lawful and fraudulent transactions. Full centre loss (FCL), a novel type of loss function that takes angle and distance relationships between features taken into consideration is suggested by the author as a way to more thoroughly supervise deep representation learning. In order to illustrate the detection performance of described model by contrasting FCL with other advanced loss functions, the author conducts several tests on two enormous data sets of credit card transactions, one of which is private and the other is public. The outcomes show that FCL performs better than its competitors.

Zheng L., Zhou M., Liu G., Jiang C., Yan C.,and Li M., et. al. [14] introduces Improved TrAdaBoost and its Use in Transaction Fraud Detection. AdaBoost has been expanded TrAdaBoost (Transfer AdaBoost) which allow for the transfer of knowledge from one domain to another. Depending on the distribution distance between a target domain and an inaccurately categorized instance, it modifies the weight of the instance in the source domain (raising or lowering it). The distance is calculated using the theory of resembling kernel Hilbert space. Five different data sets are used in a series of studies, and the outcomes show the advantage of TrAdaBoost

Can B., Karsligil E. M. Yavuz A. G., and Guvensan M. A., et. al. [15] intended to create systems for adaptive fraud detection that mostly used machine learning techniques, with the more recent use of deep learning. They created the largest data collection ever utilized in a research project for this study, containing 245 thousand fraudulent transactions and 4 billion non-fraudulent transactions, with contributions from 35 Turkey banks. So, they demonstrate and analyze the performance of profile-based fraud detection algorithms, such as the amount-based model, the card-type model and the model for transaction attributes. In order to demonstrate the suggested models resistance to aging and zero-day attacks, they also performed temporal and spatial analyses on our data set.

Zhang Z., Liu Q. Chen L., and Wang P., et. al. [16] making use of the transaction status and current transaction group behaviour, a novel approach for creating individual behaviour has been developed, can be improve the accuracy of low-frequency user behavior. To establish a benchmark for the user's own transaction behaviour, they initially analyze the optimal technique for setting risk thresholds along with the user's only previous transactions. Therefore, To create the common behaviour of the current transaction group, the density-based spatial clustering of applications with noise (DBSCAN) clustering technique is used to recover the behaviour characteristics of all recent normal samples and fraud samples. Using a sliding window method, the current transaction state is then retrieved from prior transaction records. Low-frequency users can benefit from the plan recommended in this analysis, according to the analysis, have a low percentage of evaluation mistakes for real transactions and can accurately identify fraudulent transactions.

A. A. Taha and S. J. Malebary, et. al. [17] explains that to use an optimized Light gradient boosting machine (OLightGBM) to identify fraud in credit card transactions. Two real-world public credit card transaction data sets with a mix of fraudulent and real transactions were used in trials to show the efficacy of described OLightGBM for spotting fraud in credit card transactions. By using the two data sets, the proposed approach was contrasted with earlier techniques, the area under the receiver operating characteristic curve (AUC) (92.88%), and the greatest accuracy (98.40%), F1-score (56.95%) and precision (97.34%), were achieved by this device, which outperformed the others. Makki S., Taher Y., Assaghir Z., Haque R., Hacid M. -S. and Zeineddine H., et. al. [18] gives a thorough experimental investigation and fixes for the imbalance categorization issue. Combined with the fraud detection machine learning algorithms, the author examined these solutions. The author summarized the results using a dataset with credit card fraud labels and pointed out their weaknesses. This study concludes that techniques to unbalanced classification are useless, especially when there is a significant amount of inconsistency in the data. The methods used as recently, this analysis demonstrates a lot of false alarms, which are expensive for financial organizations.

Randhawa K., Loo C. K., Seera M., Lim C. P. and Nandi A. K., et. al. [19] detects credit card fraud using machine learning techniques. First, they utilize standard models. Then, hybrid techniques that use AdaBoost and majority voting techniques are used. A publicly accessible credit card data set is utilized to assess the model's effectiveness. The results of the experiment strongly suggest that the majority voting approach detects the credit card fraud situations with excellent accuracy rates. Zheng L., Yan C., Liu G., and Jiang C., et. al. [20] To express the logical relationship of characteristics in transaction records, offer the Logical graph of behavior profile (LGBP) an extensive order-based model. The path-dependent transition probability from one feature to another may be determined using the LGBP and users transaction histories. To represent the diversity in user transaction behaviors, they define a coefficient of variation based on information entropy. They have tested described method using real data sets, and the results show that it outperforms the three advanced methods.

3. DETECTION OF CREDIT CARD FRAUD TRANSACTION

Figure 1 demonstrates the block diagram of detection of Credit card fraud transaction utilizing hybrid machine learning algorithm.

The input dataset was the record of credit card transactions done over the duration of two days in September 2013 by cardholders across Europe. Of the 284,807 transactions in the sample, only 492 are fraudulent. There are just 492 fraudulent transactions in the dataset's 284,807 total transactions. It includes data on transactions totaling 284,807. 0.17% of the transaction data are in the positive class (fraud instances). They take advantage of quantity and time by applying exploratory data analysis.

By checking the input data for abnormal entries and missing entries, data pre-processing is carried out. Format, clean, and sample from chosen record to organize information. Data cleaning refers to the removal or restoration of incomplete or empty data. Additionally, there can be instances of data that are partial and lacks the information that they should have to stop these occurrences.

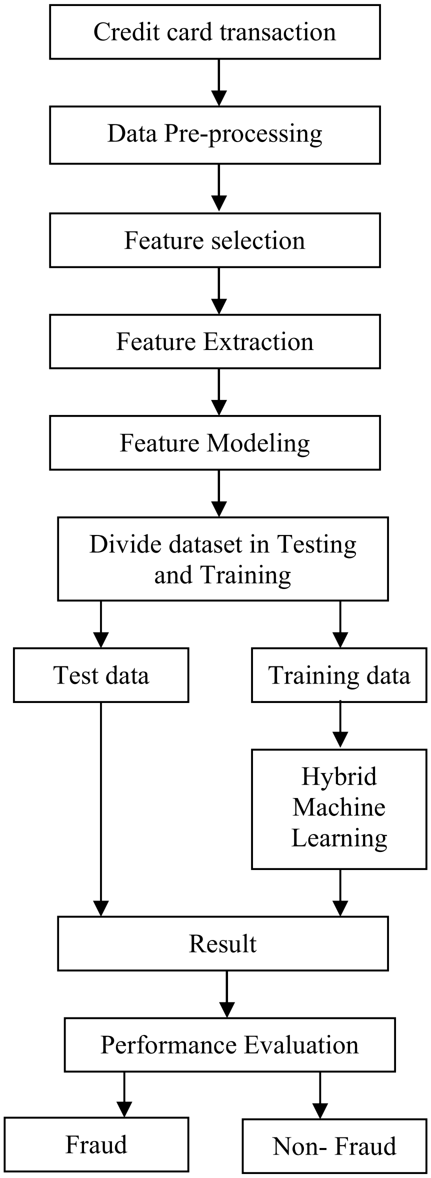

Fig. 1: Block Diagram Of Detection Of Credit Card Fraud Transaction

When there are many characteristics, choosing relevant and crucial ones is essential for the efficient detection of credit card fraud. Only numerical input variables from a Principal Component Analysis (PCA) transformation are present in the dataset. They are unable to provide the data's original attributes or further context due to privacy concerns. Features from V1 to V28 were the primary components discovered using PCA for the dataset. PCA just keeps the characteristics "Time" and "Amount" the similar. The 'Time' function takes into consideration the seconds that have passed between any transaction and the start the transaction for the dataset. Cost-sensitive learning, can make use of the 'Amount' function, which is the Amount transaction. The response variable, "Class," has a value of 1 in the case of fraud and 0 in all other cases.

The data are divided half in an 80:20 ratio, with 80% operating towards training a hybrid machine learning model and 20% operating towards testing. Two classification methods are used in hybrid machine learning training models: random forest (RF) and support vector machine (SVM). Next, they create both models.

The linear problem is converted into a higher dimensional feature space through support vector machines. This enables the resolution of complicated, non-linear problems like the identification of credit card fraud transaction implementing linear classification without expanding the complexity of the computation. To restructure the dataset, a kernel function is used. Between a space point in higher dimensions and the input space point, it is regarded as a mapping.

Among the popular supervised learning algorithms is the Random Forest algorithm. This can be applied to classification and regression purposes. But categorization issues are this approach is most frequently utilized. As a forest frequently consists of trees, the Random Forest algorithm builds decision trees from sample data and generates predictions from each sample data Consequently, an ensemble technique is the random forest algorithm. Since it averages the results, this method performs better than single decision trees at preventing over-fitting.

Sensitivity, accuracy, Precision, and are specificity performance parameters are used to quantify the analyzed model performance and determine whether a transaction is fraudulent or real.

4. RESULTS

Data about credit card fraud was obtained from a European credit card provider. The information includes the transactions completed during the previous two days. The data collection includes 284,807 transactions, of which 492 are fraudulent. The PCA transformation turns the dataset containing the input variable into numerical values. Due to confidentiality concerns, this is done. They divided the entire dataset into two groups: (20%) is the test set and (80%) is the training set. They need to evaluate criteria like accuracy, sensitivity, specificity, and precision in order to compare different algorithms. These have the following expressions:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \ldots.(1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \ldots.(2)$$

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \ldots.(3)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \ldots.(4)$$

1. True Positive, this quantifies the volume of fraudulent transactions that the system actually detects.

2. True Negative, This represents the overall amount of legal transactions that the system even recognizes as legal.

3. Genuine transactions that are mistakenly classified as fraudulent activity are known as false positives.

4. A fraudulent transaction that is incorrectly labelled as legitimate is known as a false negative.

In Table 1, the performance parameters of hybrid machine learning for credit card fraud transaction detection are compared with those of individual machine learning classifier-based credit card fraud transaction detection.

Table 1: Performance Of Different Classifiers

Parameters	Accuracy	Precision	Sensitivity	Specificity
RF	84	86	85	85
SVM	90	91	91	90
Hybrid model (SVM+RF)	98	96	96	97

Fig. 2 shows the comparative graphical representation of accuracy and precision parameters for described hybrid model and individual models.

Fig. 3 shows the comparative graphical representation of Sensitivity and Specificity parameters for described hybrid model and individual models.

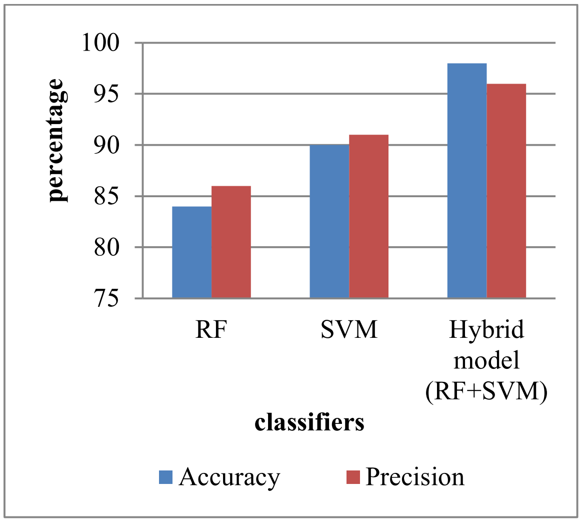

Fig. 2: Comparative Analysis Of Accuracy And Precision Parameters

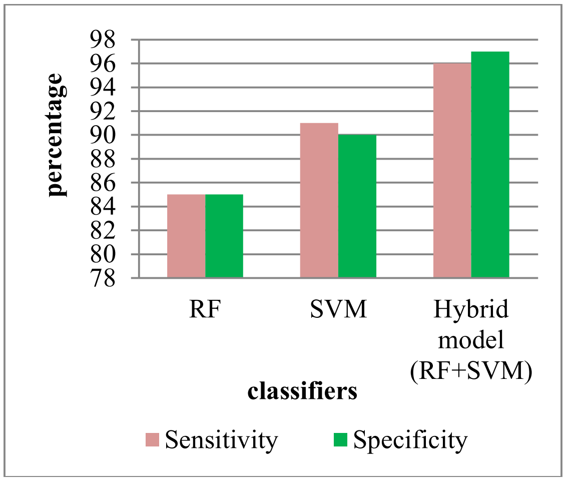

Fig. 3: Comparative Analysis Of Sensitivity And Specificity Parameters

According to Table 1, hybrid machine learning performs better than individual categories at detecting credit card fraud transactions. Described model achieves Accuracy as 98%, Sensitivity as 96%, Specificity as 97%, and Precision as 96%.

5. CONCLUSION

In this paper, Detection of Credit card fraud Transaction Utilizing hybrid machine learning Algorithm is described. Monitoring and analyzing the transactional conduct of various users in order to estimate the identification of fraudulent behaviour involves the detection of fraud. The dataset used was the record of credit card transactions done over a duration of two days in September 2013 by cardholders throughout Europe. Two categories are used in hybrid machine learning training models: random forest (RF) and support vector machine (SVM). The complete dataset has been divided into two parts: training set (80%) and test set (20%). They must evaluate criteria like accuracy, sensitivity, specificity, and precision in order to compare different algorithms. The results show that hybrid machine learning performs better at detecting credit card fraud transactions than individual categories. Described model achieves Accuracy as 98%, Sensitivity as 96%, Specificity as 97%, and Precision as 96%. They intend to use deep learning algorithms in the future to identify credit card fraud.

REFERENCES

[1] R. Li, Z. Liu, Y. Ma, D. Yang and S. Sun, "Internet Financial Fraud Detection Based on Graph Learning," in *IEEE Transactions on Computational Social Systems*, vol. 10, no. 3, pp. 1394-1401, June 2023, doi: 10.1109/TCSS.2022.3189368.

[2] Y. Xie, G. Liu, C. Yan, C. Jiang and M. Zhou, "Time-Aware Attention-Based Gated Network for Credit Card Fraud Detection by Extracting Transactional Behaviors," in *IEEE Transactions on Computational Social Systems*, vol. 10, no. 3, pp. 1004-1016, June 2023, doi: 10.1109/TCSS.2022.3158318.

[3] S. K. Hashemi, S. L. Mirtaheri and S. Greco, "Fraud Detection in Banking Data by Machine Learning Techniques," in *IEEE Access*, vol. 11, pp. 3034-3043, 2023, doi: 10.1109/ACCESS.2022.3232287.

[4] 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.

[5] F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan and M. Ahmed, "Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms," in *IEEE Access*, vol. 10, pp. 39700-39715, 2022, doi: 10.1109/ACCESS.2022.3166891.

[6] S. Han, K. Zhu, M. Zhou and X. Cai, "Competition-Driven Multimodal Multiobjective Optimization and Its

Application to Feature Selection for Credit Card Fraud Detection," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 12, pp. 7845-7857, Dec. 2022, doi: 10.1109/TSMC.2022.3171549.

[7] N. Nguyen, "A Proposed Model for Card Fraud Detection Based on CatBoost and Deep Neural Network," in *IEEE Access*, vol. 10, pp. 96852-96861, 2022, doi: 10.1109/ACCESS.2022.3205416.

[8] S. K. Hashemi, S. L. Mirtaheri and S. Greco, "Fraud Detection in Banking Data by Machine Learning Techniques," in *IEEE Access*, vol. 11, pp. 3034-3043, 2023, doi: 10.1109/ACCESS.2022.3232287.

[9] 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.

[10] P. Roy, P. Rao, J. Gajre, K. Katake, A. Jagtap and Y. Gajmal, "Comprehensive Analysis for Fraud Detection of Credit Card through Machine Learning," *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Pune, India, 2021, pp. 765-769, doi: 10.1109/ESCI50559.2021.9397029.

[11] B. Lebichot, T. Verhelst, Y. -A. Le Borgne, L. He-Guelton, F. Oblé and G. Bontempi, "Transfer Learning Strategies for Credit Card Fraud Detection," in *IEEE Access*, vol. 9, pp. 114754-114766, 2021, doi: 10.1109/ACCESS.2021.3104472.

[12] S. Han, K. Zhu, M. Zhou and X. Cai, "Information-Utilization-Method-Assisted Multimodal Multiobjective Optimization and Application to Credit Card Fraud Detection," in *IEEE Transactions on Computational Social Systems*, vol. 8, no. 4, pp. 856-869, Aug. 2021, doi: 10.1109/TCSS.2021.3061439.

[13] Z. Li, G. Liu and C. Jiang, "Deep Representation Learning With Full Center Loss for Credit Card Fraud Detection," in *IEEE Transactions on Computational Social Systems*, vol. 7, no. 2, pp. 569-579, April 2020, doi: 10.1109/TCSS.2020.2970805.

[14] L. Zheng, G. Liu, C. Yan, C. Jiang, M. Zhou and M. Li, "Improved TrAdaBoost and its Application to Transaction Fraud Detection," in *IEEE Transactions on Computational Social Systems*, vol. 7, no. 5, pp. 1304-1316, Oct. 2020, doi: 10.1109/TCSS.2020.3017013.

[15] B. Can, A. G. Yavuz, E. M. Karsligil and M. A. Guvensan, "A Closer Look Into the Characteristics of Fraudulent Card Transactions," in *IEEE Access*, vol. 8, pp. 166095-166109, 2020, doi: 10.1109/ACCESS.2020.3022315.

[16] Z. Zhang, L. Chen, Q. Liu and P. Wang, "A Fraud Detection Method for Low-Frequency Transaction," in *IEEE Access*, vol. 8, pp. 25210-25220, 2020, doi: 10.1109/ACCESS.2020.2970614.

[17] 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.

[18] S. Makki, Z. Assaghir, Y. Taher, R. Haque, M. -S. Hacid and H. Zeineddine, "An Experimental Study With Imbalanced Classification Approaches for Credit Card Fraud Detection," in *IEEE Access*, vol. 7, pp. 93010-93022, 2019, doi: 10.1109/ACCESS.2019.2927266.

[19] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim and A. K. Nandi, "Credit Card Fraud Detection Using AdaBoost and Majority Voting," in *IEEE Access*, vol. 6, pp. 14277-14284, 2018, doi: 10.1109/ACCESS.2018.2806420.

[20] L. Zheng, G. Liu, C. Yan and C. Jiang, "Transaction Fraud Detection Based on Total Order Relation and Behavior Diversity," in *IEEE Transactions on Computational Social Systems*, vol. 5, no. 3, pp. 796-806, Sept. 2018, doi: 10.1109/TCSS.2018.2856910.

.