Anomaly Detection Classifiers for Detecting Credit Card Fraudulent Transactions

Prerna Singh
Department of Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad-Meerut Road, Ghaziabad-201206, India
prernasingh51002@gmail.com

Khyati Singla
Department of Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad-Meerut Road, Ghaziabad-201206, India
khyatisingla1805@gmail.com

Prince Piyush
Department of Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad-Meerut Road, Ghaziabad-201206, India
prince.piyush.2019@gmail.com

Bharti Chugh
Department of Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad-Meerut Road, Ghaziabad-201206, India
bharti.cse@kiet.edu

***Abstract*—The internet and e-commerce have grown quickly, which has increased credit card use but also, regrettably, increased credit card fraud. To address this, Anomaly Detection has emerged as a crucial method for identifying unusual events and data in datasets. It uses advanced algorithms to detect deviations from normal patterns, helping authorities proactively combat fraudulent activities. While digital advancements offer convenience, they also expose vulnerabilities. Anomaly Detection offers a modern defense, safeguarding financial systems by early spotting of anomalies. In this study, we employed two algorithms - the Isolation Forest (IF) and the Local Outlier Factor (LOF) for identification of anomalies. To improve the performance of these models, we also employed a variety of resampling strategies. Specifically, we used techniques like Random Undersampling, AllKNN, Synthetic Minority Oversampling Technique (SMOTE), and Synthetic Minority Oversampling Technique - Edited Nearest Neighbor (SMOTE-ENN) to balance the European Credit Card Fraudulent transactions dataset and the German Credit Card fraud dataset. Out of the different configurations, the Isolation Forest classifier demonstrated the highest accuracy, reaching 99.81%, when applied to the initially imbalanced European credit card fraudulence dataset. On the other hand, the German credit card dataset achieved a remarkable accuracy of 70.60% through the implementation of the LOF classifier, coupled with the Random Undersampling technique to address its imbalanced nature.**

***Index Terms*—Anomaly Detection, IF, LOF, Credit Card Fraud, Random Undersampling, AllKNN, SMOTE, SMOTE-ENN**

I. Introduction

Hackers and fraudsters have targeted credit cards for years, and this is unlikely to change anytime soon. Unfortunately, as technology advanced, so did the most prevalent scams, so con artists are constantly coming up with new ruses to con unsuspecting victims. In recent years, there has been a massive increment in the use of online banking, high-tech payments, and buying using debit/credit cards as a result of the proliferation of businesses, online services, and internet users. While this offers convenience and efficiency, it has also given rise to financial frauds like money laundering, card theft, and identity theft. Despite benefits such as cashless transactions and time saving, these frauds pose serious risks modern tech advancements have fueled more sophisticated fraud techniques, causing substantial losses to businesses and users.

Fraud detection systems employing data mining and machine learning have been developed, but challenges remain. There exist two distinct categories of credit card fraudulent activities. The first involves the unlawful acquisition of the physical credit card itself, while the second pertains to the illicit procurement of sensitive card-related information, encompassing the card number, CVV code, card type, and related data. Through the unlawful acquisition of credit card information, an unauthorized person can potentially gain access to substantial financial resources or execute significant transactions prior to the cardholder's awareness of such activities. Fraudulent activities include unauthorized transactions, credit card theft, and more, affecting customer trust and business stability credit card fraud, both through application and behavior, is a major concern, leading to billions in losses.

In 2017, global credit card fraud losses reached $22.8 billion, and rise to $31 billion in 2020. Detection of these frauds is crucial due to their impact on financial transactions and institutions. Sixty-five percent of cardholders, up from 58 percent the year before, have experienced fraud of some kind. As the second-largest category for identity theft fraud, the Federal Trade Commission claims that it received roughly 390,000 allegations of credit card identity theft fraud in 2021. By new credit card accounts in 2020, identity theft climbed by 48%. With reference to Nilson report, credit card theft is increasing and is expected to touch an astounding figure of

$38.5 billion by 2027.

Credit card fraud detection has been aided by the use of artificial intelligence, data mining and machine learning techniques, among others. Unfortunately, these initiatives have not produced any noteworthy results. One major obstacle to using ML methodologies for credit card fraud detection is the personal and confidential nature of that data. Therefore, datasets with anonymized properties are used in the development of ML models for credit card fraud detection. Therefore, identifying fraudulent transaction is very tough because new patterns, deviations and features emerge with each fraudulent transaction.

A. Motivation

The increasing prevalence of credit card fraud has prompted a compelling need for robust fraud detection mechanisms. As financial transactions shift towards digital platforms, the vulnerability to fraudulent activities escalates. The motivation behind this endeavor lies in preventing substantial financial losses, ensuring customer trust, and upholding the integrity of electronic payment systems. By working in this domain seeks to address this critical issue, safeguarding both consumers and financial institutions. By harnessing state-of-the-art technical automation such as ML, artificial intelligence and data analytics, the aim is to create proactive and accurate systems that identify and mitigate fraudulent transactions, ultimately fostering a more secure and resilient financial landscape.

B. Literature Review

The substantial financial losses incurred as a result of fraudulent activities have spurred researchers to seek a remedy capable of pre-emptively identifying and thwarting such illicit actions. Numerous approaches have been posited and subjected to testing in pursuit of this objective.

Ref [1] used the SMOTE technique to balance the dataset. They then ran the Multilayer Perceptron (MP), Random Forest (RF), Naive Bayes (NB), and Logistic Regression (LR) algorithms. Their research showed that the Random Forest algorithm produced the best outcomes. Ref [2] conducted an investigation involving the utilization of SVM, KNN and Artificial Neural Network (ANN) techniques to forecast instances of fraud. A comparative analysis was conducted between ML algorithms and ANN. The results indicated that the implementation of an ANN yielded an accuracy level approaching 100%. Ref [3] employed the RF and AdaBoost algorithms, utilizing the confusion matrix to generate ROC curves. The outcomes derived from both algorithms yielded congruent conclusions. When evaluating precision, recall, and the F1-score, RF algorithm exhibited the highest values.

Ref [4] executed GridSearchCV and random search for hyperparameter optimization. The Decision Tree classifier achieved 72.1% accuracy on the imbalanced German credit card dataset. LDA reached an impressive 98.6% accuracy on the imbalanced European credit card fraud dataset. Additionally looked into were SMOTE, Naive Bayes, and Logistic Regression. As a result, the Decision Tree model performs better than the LR, LDA, and Naive Bayes algorithms. Ref [5] crafted an ML-based framework fusing LR, SVM, RF, NB, XGBoost, and ET algorithms with AdaBoost. The amalgamation aimed to heighten classification efficacy, evaluated via metrics like accuracy, recall, precision, MCC, and AUC. Ref [6] introduces a resilient deep-learning methodology incorporating LSTM and GRU neural networks as foundational classifiers within a stacking ensemble structure and a MLP functions as the higher-level learner. To address class distribution imbalance, the SMOTE-ENN technique is utilized in the dataset.

Ref [7] conducts a performance assessment of the effectiveness of 9 distinct classifier models on credit card fraud detection data. The models are Logistic Regression, K-Nearest Neighbors, Random Forest, Naive Bayes, Multi Layer Perceptron, AdaBoost, quadrant discriminative analysis, ensemble learning and pipelining. To rectify dataset imbalance, the ADASYN technique is used. In conclusion, Pipelining method demonstrated superior performance across diverse metrics. Ref [8] introduced a financial fraudulence detection scheme utilizing a Deep Convolution Neural Network (DCNN) and algorithms of deep learning to detect fraudulent transactions. Within a 45-second timeframe, an accuracy rate of 99% in detection was achieved. Ref [9] employed the GA for the purpose of feature selection. The devised detection mechanism incorporated a selection of ML classifiers - DT, RF, LR, ANN, and NB. Notably, the genetic algorithm was incorporated within the RF's fitness function. This amalgamation, termed GA-RF (utilizing version 5), yielded a remarkable overall accuracy of 99.98%. A combination of data resampling strategy and a neural network ensemble classifier are introduced by Ref [10]. LSTM neural network is implemented as the primary trainee to create the ensemble classifier within the AdaBoost framework. Simultaneously, the hybrid resampling strategy incorporates the SMOTE-ENN method. According to their results, using the SMOTE-ENN data resampling method in tandem with the boosted LSTM classifier is a promising strategy for spotting fraudulent acts in financial transactions using credit cards.

Ref [11] compared the LOF and IF algorithms using Python to identify fraudulent transactions, yielding a 97% accuracy for LOF and 76% for IF. Ref [12] implemented the IF model and utilized the Area Under the Precision-Recall curve (AUC), which demonstrated superior outcomes compared to the Area Under the ROC curve. The efficiency of their approach in a fraud detection model was noted as 98.72%. Ref [13] employed the one-class support vector machine classifier (OCSVM) for outlier detection, optimizing hyperparameters γ and ν through grid search based on maximizing the AUC.

The GS-OCSVM exhibited superior fraud detection to the isolation forest, with higher true negative rates for both German and European credit card datasets. Ref [14] utilized Python to apply the IF and LOF algorithms. After evaluating the datasets, they observed that the IF algorithm delivered a higher precision rate compared to the LOF algorithm. Ref [15] discusses the IF algorithm's role in credit card fraud detection and contrasts it with the LOF algorithm. IF outperforms LOF in terms of error, accuracy-score and recall for two algorithms. The rate of fraud detection is about 27% with Isolation Forest, In comparison to LOF's mere 2%. IF model boasts 99.774% precision, surpassing LOF's 99.65%. Ref [16] employed the random under-sampling approach on imbalanced data and applied three machine learning algorithms - LR, NB, KNN. The algorithms efficacy was assessed using metrics encompassing accuracy-score, F-measure, sensitivity, precision, specificity and AUC.

II. Preliminaries

A. EDA (Exploratory Data Analysis)

In this study, we conducted an evaluation on two distinct datasets. The European credit card fraud detection dataset and the German credit card fraudulent dataset, both sourced from Kaggle. These datasets are presented in CSV format (.csv).

The dataset pertaining to European credit card transactions encompasses an extensive count of 284,807 instances, reflecting the payments conducted by consumers across Europe in the month of September 2013. Remarkably, out of this extensive pool, a mere 492 transactions are identified as fraudulent, constituting an exceedingly minute proportion of approximately 0.17%. This stark discrepancy in statistics emphasizes how unbalanced the European credit card fraud dataset is. There are two categories for these transactions: legitimate and fraudulent. The original functionality and other contextual information are removed from the training data to meet security issues. Using only integral attributes as its emphasis, the Principal Component Analysis (PCA) transformation's outcomes are presented. The PCA-derived principal components are denoted by the variables V1, V2, V3,....,V28. Notably, the PCA conversion process has not been applied to the 'Time' and 'Number' properties.

The original German credit card dataset comprises a total of 21 distinct features, with 13 of them being categorical in nature and the remaining 7 being numerical attributes. This dataset encompasses a total of 1000 credit card transactions. Among these transactions, 700 are classified as normal transactions, while the remaining 300 are categorized as fraudulent transactions, resulting in a proportion of 0.42% representing the occurrence of fraudulent activities.

B. Data Preprocessing

Data preprocessing stands as a pivotal initial stage in the journey of refining raw data into a more structured and suitable form. Given the dataset's combination of categorical and numerical attributes, the technique of label encoding emerges as a practical choice to translate all features into a uniform numerical representation. Within this encoding scheme, normal transactions are denoted by '0', whereas fraudulent transactions are signified as '1'. Within the context of the European dataset, it was uncovered that a total of 1081 instances exhibited duplicated transaction records. In order to uphold precision and fairness in predictive outcomes, a strategic decision was taken to eliminate these instances of duplication from the dataset. In conjunction with this, the process of data scaling was introduced, involving the meticulous removal of any instances featuring null or missing values. This diligent effort contributes to data integrity and the credibility of analysis outcomes. Furthermore, a critical step involved the process of normalization, whereby data points are standardized to a common scale. This procedure fosters consistency across the dataset and facilitates a more meaningful and accurate analysis, ultimately enhancing the quality of insights derived from the data.

C. Data Balancing

Both of the datasets originally display a compelling class imbalance, with preponderance of transactions are categorized as normal occurrences. Given this substantial disparity, addressing the imbalance becomes an imperative undertaking to ensure the efficient and unbiased training of predictive models. Commonly employed methodologies to rectify the skewed class distribution encompass strategies such as equalizing the class proportions through techniques like decreasing the scale of the majority class (undersampling), magnifying the scale of the minority class (oversampling), or implementing a combination of both strategies. To address the imbalance, two distinct techniques for undersampling are employed - the Random Undersampler and the AllKNN technique. In contrast, for oversampling, the SMOTE methodology is harnessed to perform data balancing. Furthermore, a hybrid approach termed the SMOTE-ENN is incorporated. This hybrid approach merges the benefits of both undersampling and oversampling, aiming to restructure the dataset's distribution.

The overarching goal of these methodologies is to redress the inherent class imbalance, thereby fortifying the model training process by enabling accurate learning of both the minority and majority classes. This comprehensive approach lays the foundation for more resilient and effective model outcomes.

III. Machine Learning Models

In the scope of this study, we embark on a comprehensive comparison between two distinct algorithms - the LOF and the IF. These algorithms hold a pivotal role in the domain of anomaly detection, a field aimed at identifying unusual or unexpected occurrences within datasets.

A. Isolation Forest

The Isolation Forest algorithm, in particular, stands out as an unsupervised methodology meticulously designed for the explicit purpose of outlier detection. Its core principle revolves around the concept of ”isolation”, a procedure where an individual data instance is segregated from the rest of the dataset. This distinctive approach sets it apart from conventional methods, introducing a fresh perspective to anomaly detection. What distinguishes the Isolation Forest algorithm is its departure from the customary reliance on distance and density metrics. Instead, it innovatively incorporates the concept of isolation, thereby elevating the efficacy and efficiency of anomaly identification. This strategic shift leads to improved accuracy and a more nuanced understanding of anomalies within the data. An additional advantage of the Isolation Forest algorithm is its resource efficiency. It demands minimal memory allocation, making it suitable for diverse computational environments. Furthermore, its computational overhead is remarkably low, thanks to its intrinsic linear time complexity.

B. Local Outlier Factor

To identify outliers, the LOF algorithm compares a data point’s local variances to those of its neighbors. It gauges outliers based on local density, influenced by proximity to nearest neighbors. Calculating the data point’s density against neighboring points enables identification of regions with akin densities, singling out those with notably lower densities. LOF measures the extent of density deviation of data point from its neighbors; if significantly lower, it’s considered an outlier. By emphasizing local relationships and densities, LOF unveils nuanced anomalies often unnoticed by conventional methods. This approach enriches anomaly detection’s granularity, offering a more thorough comprehension of irregularities within the dataset.

C. Implementation

There are three distinct subsets of the dataset - training set, validation set, and testing set. The classifier is initially taught on the training dataset, then hyperparameter tuning is performed on the validation dataset, and finally its performance is evaluated on the testing data. Notably, the model is solely exposed to the training and validation datasets. A crucial point of emphasis is the strict prohibition of utilizing the test dataset for classifier training.

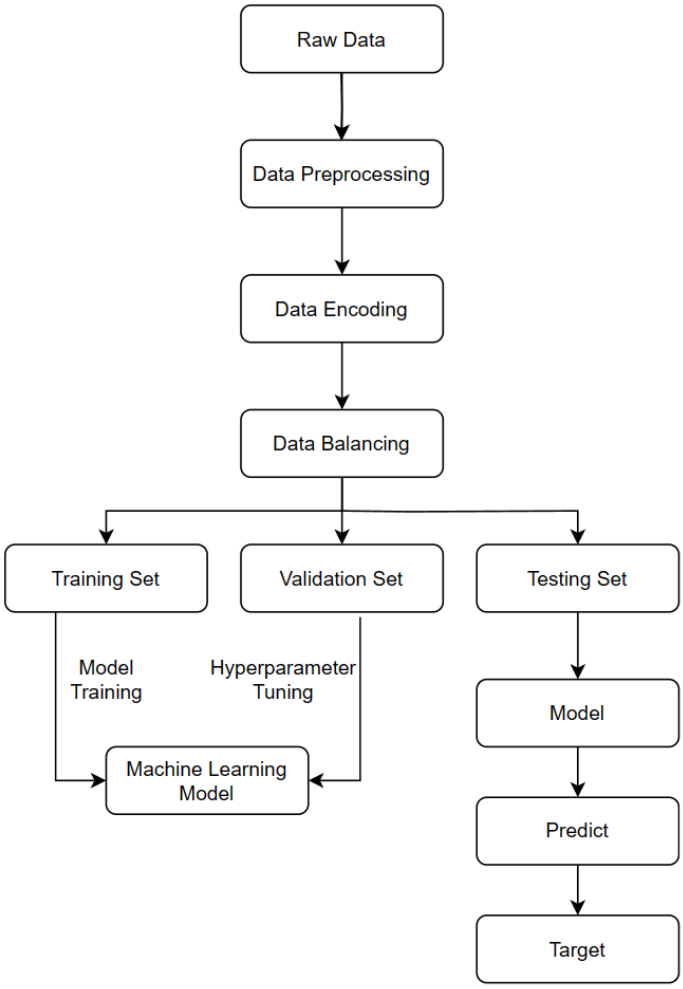

Fig. 1: Block Diagram of the Proposed Model

As depicted in Fig. 1 above, the block diagram delineates the configuration of the planned model. The model’s operational sequence comprises several key steps - data collection, data processing, identification of the suitable model based on data characteristics, subsequent model training and testing, and culminating in a comprehensive evaluation. This stage is pivotal in driving the refinement of machine learning model accuracy.

In unsupervised learning, the training process involves using only the variables present in the dataset, without any predefined labels. The goal is to build models that can identify patterns or deviations within the data. Accuracy rate, precision, recall, and F1-score are just some of the criteria used to assess these models’ efficacy, the detection of outliers, and a confusion matrix, which includes metrics like true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). These models are then put to the test by making predictions during the testing phase.

In our research project, we have two sets of data. For

both of these datasets, we begin by evaluating the models without making any adjustments to the class distribution. Then, we proceed to apply resampling techniques on both the European and German credit card fraud transaction datasets. These techniques include random undersampling, AllKNN, SMOTE, and SMOTE-ENN. After each resampling technique is applied, we record and analyze the evaluation metrics to understand how they have changed. This procedure allows us to evaluate the effect of the various resampling techniques on the accuracy of the models applied to both data sets.

IV. Results and Discussions

To evaluate the most efficient algorithm for detecting fraudulent transactions, various criteria for comparing algorithms have been implemented. Accuracy, recall, and precision are the uttermost popular metrics used for the evaluation of the success of machine learning algorithms. All of these measurements can be computed using a Confusion matrix, which is a structured way of evaluating model performance. The assessment of how well a model performs was carried out based on these metrics. The models were subjected to testing using both the original dataset and a dataset that had been resampled. The findings clearly demonstrated that in most of the cases the procedure of resampling plays a significant aspect in influencing the performance of the models.

On the imbalanced European and German credit card dataset, two prominent anomaly detection models, IF and LOF were employed. The aim was to improve their performance, reduce false negatives, and enhance the total accuracy of the models. Additionally, various resampling techniques were explored to tackle the challenge posed by imbalanced data.

Table I: Model performance on European Dataset without resampling.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	No. of Outliers
IF	99.81	63	53	54	107
LOF	99.81	50	50	50	104

Table II: Model performance on German Dataset without resampling.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	No. of Outliers
IF	70.50	35	50	41	59
LOF	70.50	35	50	41	59

To diminish the impact of class imbalance, four resampling techniques were applied - Random Undersampling, AllKNN, SMOTE, and SMOTE-ENN. These techniques aimed to balance the class distribution and enhance the models ability to capture patterns from both majority and minority classes. The results of the IF and LOF models using a variety of resampling strategies on European and German dataset are shown in Tables 3, 4, 5, and 6.

Table III: IF Model performance on European Dataset with various resampling techniques.

Resampling Technique	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	No. of Outliers
Random under-sampling	98.98	49	50	50	101
AllKNN	98.80	53	78	55	682
SMOTE	49.76	25	50	33	57130
SMOTE-ENN	48.77	25	49	33	58201

Table IV: LOF Model performance on European Dataset with various resampling techniques.

Resampling Technique	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	No. of Outliers
Random under-sampling	98.99	49	50	50	100
AllKNN	98.96	50	50	50	591
SMOTE	49.98	67	50	34	56880
SMOTE-ENN	49.53	42	50	34	57341

Table V: IF Model performance on German Dataset with various resampling techniques.

Resampling Technique	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	No. of Outliers
Random under-sampling	70.56	35	50	41	53
AllKNN	51.61	48	48	48	45
SMOTE	52.85	47	50	36	132
SMOTE-ENN	43.80	22	50	30	59

Table VI: LOF Model performance on German Dataset with various resampling techniques.

Resampling Technique	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	No. of Outliers
Random under-sampling	70.60	35	50	41	52
AllKNN	54.83	55	55	54	42
SMOTE	51.42	26	48	34	136
SMOTE-ENN	43.80	47	50	33	59

A comparison of all models employed on European and German dataset, along with resampling techniques, is conducted to discover the optimal model, and the results are summarized in Table 7. Whereas Table 8 illustrates the

comparative study of our proposed approach with the already published research findings.

Table VII: Best Model on respective dataset

Dataset	Best Model
European	IF
German	LOF - Random under-sampling

Table VIII: Comparative analysis with previous reports on European Dataset.

Reference	Model	Accuracy (%)
Ref [11]	LOF	97
Ref [12]	IF	98.72
Ref [14]	IF	99.72
Proposed	IF	99.81

To effectively prevent credit card fraud and protect the security of financial transactions, this research highlights the need of employing both modern algorithms and data preparation methodologies.

Moreover, the positive impact of resampling was evident in the complete elimination of false negatives (FN), as well as a reduction in the count of false positives. This indicates that fraudulent transactions were not only effectively identified but also distinguished from normal transactions, thus minimizing both types of misclassifications.

V. Conclusion

Through the use of a dataset built from German credit card transactions and European credit card transactions, this study compares the performance of various machine learning classification techniques. Identifying specific transactions as fraudulent is the main goal. The central aim is to discern the fraudulent nature of specific transactions. Both the credit card dataset underwent a series of stages, including importation, preprocessing, encoding, and configuration, in preparation for model training facilitated by the machine learning workflow mechanism. Subsequently, the dataset was subjected to training, deployment, and evaluation for each classification model, encompassing the utilization of multiple resampling techniques.

In particular, when conducting a comparative analysis between the IF model and the LOF model for the European credit card fraudulent transactions dataset, the IF model outperforms the LOF model in terms of recognizable performance, with an accuracy of 99.81%.
On the German credit card fraudulent transaction dataset, the LOF algorithm performs at its highest accuracy of 70.60% over the IF model when executed in integration with the random undersampling methodology.

References

[1] Varmedja, Dejan, Mirjana Karanovic, Srdjan Sladojevic, Marko Arsenovic, and Andras Anderla. "Credit card fraud detection-machine learning methods." In 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH), pp. 1-5. IEEE, 2019.

[2] Asha, R. B., and Suresh Kumar KR. "Credit card fraud detection using artificial neural network." Global Transitions Proceedings 2, no. 1 (2021): 35-41.

[3] Sailusha, Ruttala, V. Gnaneswar, R. Ramesh, and G. Ramakoteswara Rao. "Credit card fraud detection using machine learning." In 2020 4th international conference on intelligent computing and control systems (ICICCS), pp. 1264-1270. IEEE, 2020.

[4] Chugh, Bharti, and Nitin Malik. "Machine Learning Classifiers for Detecting Credit Card Fraudulent Transactions." In Information and Communication Technology for Competitive Strategies (ICTCS 2021) ICT: Applications and Social Interfaces, pp. 223-231. Singapore: Springer Nature Singapore, 2022.

[5] Ileberi, Emmanuel, Yanxia Sun, and Zenghui Wang. "Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost." IEEE Access 9 (2021): 165286-165294.

[6] Mienye, Ibomoiye Domor, and Yanxia Sun. "A Deep Learning Ensemble With Data Resampling for Credit Card Fraud Detection." IEEE Access 11 (2023): 30628-30638.

[7] Bagga, Siddhant, Anish Goyal, Namita Gupta, and Arvind Goyal. "Credit card fraud detection using pipeling and ensemble learning." Procedia Computer Science 173 (2020): 104-112.

[8] Chen, Joy Iong-Zong, and Kong-Long Lai. "Deep convolution neural network model for credit-card fraud detection and alert." Journal of Artificial Intelligence and Capsule Networks 3, no. 2 (2021): 101-112.

[9] Shirodkar, Nikita, Pratikesh Mandrekar, Rohit Shet Mandrekar, Rahul Sakhalkar, KM Chaman Kumar, and Shailendra Aswale. "Credit card fraud detection techniques–A survey." In 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), pp. 1-7. IEEE, 2020.

[10] Esenogho, Ebenezer, Ibomoiye Domor Mienye, Theo G. Swart, Kehinde Aruleba, and George Obaido. "A neural network ensemble with feature engineering for improved credit card fraud detection." IEEE Access 10 (2022): 16400-16407.

[11] John, Hyder, and Sameena Naaz. "Credit card fraud detection using local outlier factor and isolation forest." Int. J. Comput. Sci. Eng 7, no. 4 (2019): 1060-1064.

[12] Gupta, Swati, Sanjay Patel, Surender Kumar, and Goldi Chauhan. "Anomaly detection in credit card transactions using machine learning." (2020).

[13] Kittidachanan, Kittikun, Watha Minsan, Donlapark Pornnopparath, and Phimphaka Taninpong. Anomaly detection based on GS-OCSVM classification. In 2020 12th International Conference on Knowledge and Smart Technology (KST), pp. 64-69. IEEE, 2020.

[14] Vijayakumar, V., Nallam Sri Divya, P. Sarojini, and K. Sonika. "Isolation forest and local outlier factor for credit card fraud detection system." International Journal of Engineering and Advanced Technology (IJEAT) 9 (2020): 261-265.

[15] Rajeev, Haritha, and Uma Devi. "Detection of credit card fraud using isolation forest algorithm." In Pervasive Computing and Social Networking: Proceedings of ICPCSN 2021, pp. 23-34. Springer Singapore, 2022.

[16] Itoo, Fayaz, Meenakshi, and Satwinder Singh. "Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection." International Journal of Information Technology 13 (2021): 1503-1511.

[17] Liu, Fei Tony, Ting, Kai Ming and Zhou, Zhi-Hua. "Isolation forest." Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on

[18] Liu, Fei Tony, Ting, Kai Ming and Zhou, Zhi-Hua. "Isolation-based anomaly detection." ACM Transactions on Knowledge Discovery from Data (TKDD) 6.1 (2012): 3

[19] Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000, May). LOF: identifying density-based local outliers. In ACM sigmod record.