Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 04, April-2023

Credit Card Fraud Detection: A Survey

Kavyashree S
Information Retrieval Lab,
Department of Computer Science and Engineering
Presidency University, Bangalore

Dr. Zafar Ali Khan
Associate Professor and Program Head CSE,
Information Retrieval Lab, Department of Computer
Science and Engineering
Presidency University, Bangalore

Abstract: The most common electronic payment method is a credit card, which is more prone to fraud due to the amount of daily online transactions that are taking place. As more people use credit cards for e-commerce as well as other purchases in the latest days, credit card fraud has grown in prevalence. Almost all firms of business nowadays, in both small and large sectors, accept credit cards as a transaction method. The COVID-19 epidemic has profoundly impacted individual interactions. The rising usage of credit cards is one of these massive changes. Due to the epidemic, e-commerce has become crucial for meeting customer expectations. Due to card fraud, credit card firms have incurred significant losses. At the moment, the most frequent problem is the identification of credit card fraud. Credit card firms are searching for the best procedures and technology to identify and minimize fraudulent credit card transactions. In this study, many approaches for detecting credit card fraud have been examined, highlighted, and contrasted with drawbacks and benefits for each.

Keywords: Detection Techniques, Machine Learning, Fraud Detection

1. INTRODUCTION

Fraud is an unethical act committed by a third party by defrauding the innocent. Card fraud works both ways: with card fraud, which is now less common, as well as without card fraud (more common). Numerous methods of compromise are possible, and they frequently take place without the cardholder's awareness. Database security flaws have become more expensive because of the internet; millions of accounts have been exposed in certain situations. Finding the origin of the attack might be difficult since a fraudster may have a stolen card's details for months even before the loss. Cardholders can quickly report stolen cards. The cardholder might not become aware of unauthorized use until they receive a statement. By routinely checking their accounts to make sure there are no unapproved or suspicious transactions, cardholders may lower their risk of fraud. The COVID-19 epidemic has led to an upsurge in global e-commerce volumes. Digital transactions involving credit cards are becoming a frequent occurrence. Through the internet, the advancement of technology has substantially altered how we communicate with our environment. People's everyday routines have been impacted by simple tasks like Internet surfing for information and online patient diagnosis. Spam via SMS and email is the most preferred way to commit online fraud.

Compared to the previous year, which saw 2,545 instances and frauds totaling Rs 119 crore, the number of cases involving internet-related frauds climbed to 3,596 cases and Rs 155 crore during the year ending March 2022. Overall, there have been more bank frauds, although the amount involved has decreased. Off-balance sheet frauds totaled Rs 1,077 crore in 21 instances, as opposed to Rs 535 crore in 23 cases, according to the RBI Annual Report in March 2022 [26]. However, as of March 22, 2022, the value of all frauds recorded by the banking sector had decreased to Rs 60,414 crore with 9,103 incidents, down from Rs 138,211 crore with 7,359 cases the year before. Of this, Rs. 40,282 crore in frauds and Rs. 17,588 crore in private banks' reports [26].

Example: Random Credit Card Numbers- Fraudsters will employ computer programs that create card numbers at random. To identify which number combinations actually belong to a credit card, the card numbers will be tested with minor "test charges." The thieves know they have a true credit card number when they receive payment approval. After that, they'll try using the account to make significantly greater purchases. Fortunately, the solution to this problem is as simple as carefully going over your credit card statements. Many people we know just give their credit card statement a cursory scan when it arrives or when they log into their account to make a payment. You may make sure you don't pay something fraudulent or unforeseen by carefully reviewing every transaction on your statement. If you discover anything unfamiliar, notify your credit card company right away. Additionally, your bank can allow you to register for transaction alerts.

Phishing- One of the online shopping scams that have existed essentially since the beginning of the internet is this one. It entails sending a link offering a coupon or promotion via email (or, in modern times, publishing on social media). Another possibility is a distressing email informing you that your account has been suspended. The same guidelines that applied to spoof websites also apply to phishing scams. Never click on a suspicious email link. By placing your mouse over the link and examining the address that appears on your screen, you may verify the URL. Don't click on anything that doesn't seem real. Likewise, you can visit the website on your own, login, and verify the status of your account or change your password if you are concerned that there might be an issue with your account.

2. LITERATURE SURVEY

2.1 Machine Learning Methods

Mary Isangediok and Kelum Gajamannage [1] Investigated four cutting-edge machine learning (ML) strategies that are appropriate for dealing with unequal groups to increase extreme gradient boost and precision, decision trees, random forests, and logistic regression all work together to minimize false positives. First, two original benchmark unbalanced fraud detection datasets are used to train these classifiers: websites with phishing URLs and credit card transaction fraud, using the frameworks sampling of the Random Under Sample, Synthetic Minority Oversampling Method, and SMTE corrected the Nearest Neighbor, the synthetically balanced datasets are modeled for each original dataset (SMOTEENN). Extreme Gradient Boosting (XGB) & Random Forest (RF) outperform Logistic Regression (LR) & Decision Tree (DT). The precision & Area Under the Curve of Precision and Recall (AUC PR), Extreme Gradient Boosting (XGB) outperforms Random Forest (RF) and shows more robustness.

Asha RB and Suresh Kumar KR [2] aimed to use Numerous Machine Learning, including the Support Vector Machine (SVM), K- Nearest neighbor (KNN), and Artificial Neural Network (ANN), are used forecast the likelihood of scam levels in order to distinguish between false and lawful transactions, they then conducted a distinction between the actually achieved Deep Learning and Supervised Machine Learning methodologies. Precision, recall, and accuracy are computed and the outcome is evaluated using the confusion matrix. It has two classes: projected class and actual class. When it comes to detecting fraud in credit card transactions, artificial neural networks are more accurate than support vector machines and K-Nearest Neighbor methods.

Anuruddha Thennakoon et al [9], Emmanuel Ileberi et al [15], and Aisha Mohammad Fayyomi et al [18] proposed best-suited algorithms to recognize four different fraud-related transaction patterns and by addressing pertinent difficulties brought up by earlier study on banking credit cards scam detection and prevention, a different and unique detection model is created. Genuine credit card scam prevention is handled using Predictive Analytics and a module with the API, and the client is alerted via the Graphical User Interface as soon as an illegal activity has taken. Logistic Regression (LR), Naïve Bayes (NB), Decision Tree (DT), and Support Vector Machine (SVM) are the machine learning systems that correctly recognized scam patterns.

Dejan Varmedja et al [19] For the purpose of detecting fraudulent transactions, compare specific machine learning techniques. Accordingly, several comparisons were made, and it was shown that the Random Forest Algorithm returns good values, that is, better distinguishing between fraudulent and legitimate transactions. This was determined using a number of parameters, including recall, accuracy, and precision.

Vaishnavi Nath Dornadula and Geetha S [20] designed and developed a splitting fraud detection technique that uses streaming transaction data with the goal of deriving behavioral patterns from customer historical transaction details, wherein cardholders are classified based on each activity the money spent. The transactions completed by cards from different classes are then grouped based on the sliding window procedure, allowing the behavior of the groupings to be obtained. The groups are trained individually using several classifiers. In the end, Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF) are the classifiers that produced the higher values. One of the best methods to anticipate fraud can be chosen as the classifier with the highest rating score.

B. N. V. Madhubabu et al [22] proposed a system identifying using a random forest algorithm, to identify fraudulent transactions and their accuracy Random Forest Algorithm (RFA). This method uses decision trees to classify the dataset according to a supervised learning algorithm. Following the grouping of the dataset, a Confusion Matrix is gathered. The Confusion Matrix is employed to assess RFA’s effectiveness.

Noor Saleh Alfaiz and Suliman Mohamed Fati [23] proposed 66 machine-learning models in all were proposed made up of two phases of evaluation. Each model employs Stratified K-Fold Cross Validation together with a real-world dataset of European cardholders with the purpose of catching credit card fraud. In the initial round, nine machine learning techniques are assessed for their ability to recognize illegal activities. In the secondary level, the top triple classifiers are picked to be used once more, and each of the top three algorithms is evaluated using 19 resampling strategies. In the 330 assessment criteria that taken roughly a month to complete, AIIKNN undersampling strategy combined with CatBoost (AIIKNN-CatBoost) is declared to be the better system offered.

Prerak Parekh et al [24] focuses on comparing the performance of several supervised machine-Learning classifiers for recognizing illegal activities when working with a highly skewed dataset by applying sampling approaches and implementing unique problem-solving techniques. Problems include class imbalance, sample methods, and overfitting. They were able to get around it via resampling strategies like Synthetic Minority Oversampling Technique (SMOTE) and random under-sampling. With a 0.97 Receiver Characteristic Curve- Area Under Curve (ROC-AUC) score, the random forest with oversampling approach delivered better results in this study

2.2 Deep Learning Methods

Rajesh Sabari [3] Each Generative adversarial network (GAN) was trained for 5000 rounds before the results were analyzed. The starting Generative adversarial network (GAN) puts the Discriminator Network against the Generator Network using the system's model connections will be retained by the cross entropy loss from the discriminator. On this basis, it was possible to determine the original data's shape and range. To eventually get a softmax prediction y identifying whether the augmented image was fraudulent or not, they additionally processed it through a Convolutional Layer 1x29 and completely linked compact variations layers (300/100 Neurons for each of them).

J Femila Roseline et al [5] a Long Short Term Memory- Recurrent Neural Network (LSTM-RNN) was proposed for detecting the incidence of fraud; the mechanism has been added to boost performance even more. Vectors with complex connected features make up the information sequence for fraud detection; it has been demonstrated that models with this structure will perform very well. Other algorithms like Support Vector Machine (SVM), Naive Bayes (NB), and Artificial Neural Networks (ANN) are contrasted for LSTM-RNN (ANN). Results from experiments show that their suggested model has a greater level of accuracy and generates effective results. When, Artificial Neural Networks (ANN), Support Vector Machine (SVM), and Naive Bayes (NB) were evaluated for superior prediction results, the LSTM-RNN classifier with boosting technique outscored them all.

Sahil Negi et al [6] The models used in this project were Extreme Gradient Boosting (XGB), Multi-layer Perceptron (MLP), and Logistic Regression (LR). These models were trained and tested using the Kaggle dataset. The results of the suggested system models were verified using the classification metrics accuracy, confusion matrix, AU the ROC Curve, F1 Score, Recall, and Precision. The Multi-layer Perceptron (MLP) surpasses Extreme Gradient Boosting (XGB) and Logistic Regression (LR), according to this study's findings.

Fawaz Khaled Alarfaj et al [7] utilized the most recent advancements in deep learning algorithms to do this. To get effective results, a comparative examine deep Learning and machine Learning classifiers done. Detecting the fraud utilizing collection of European card bench mark thorough empirical investigation is conducted. For situations involving credit card detection. The demonstrated system performs better in State-Of-The-Art ml and deep Learning methods.

Ruttala Sailusha et al [10] and D. Tanouz et al [11] both AdaBoost technique and the Random Forest (RF) algorithms are utilized. The outcomes of the two algorithms are based on many criteria such as f1-Score, Accuracy, Precision, and recall. The Receiver Operating Characteristic Curve (ROC) is drawn using the Confusion Matrix as a foundation. The procedure with the greater accuracy, recall, precision, and f1- Score tends to be the great one for detecting scams when the algorithm from Random Forest and AdaBoost are differentiated.

Jeonghyun Hwang and Kangseok Kim [12] for fraud detection, an effective domain-adaptation strategy is suggested. In order to reduce the divergence that happens when a general characteristic is changed between two domains, the suggested method makes use of the discriminative properties found in Feature Maps and Generative Adversarial Networks (GANs). Implementing backward classes for the Data augmentation strategy, which makes use of a GAN, generates fresh data samples in sequence to address the problem of imbalance class and improve the model's detection accuracy. Using classification performance assessment indicators, they assessed the demonstrated domain-adaption model’s Convolutional Neural Network (CNN) models. The Support Vector Machine (SVM) performed well in terms of classification, but as the dataset size was increased, it needs a greater training period than a convolutional neural network. As a result, it was demonstrated that the suggested domain adaption model can simultaneously categorize two datasets with related domains while taking less time to learn than the Support Vector Machine (SVM) and Convolutional Neural Network (CNN).

Tushar Patil [13] developed a hybrid approach that combines the feature selection method SelectKBest with the data modeling approach CT-GAN to manage the class imbalance data. All other classifiers are outperformed by Random Forest paired with SelectKbest and CT-GAN based on F1-Score and Recall quantities (LR, XGB). All of the algorithms have demonstrated a considerable increase in all-around performance prediction with a decrease in bugs trained after being on supplemented data.
Sairamvinay Vijayaraghavan et al [21] presented an alternative approach to address the class imbalance dataset problem by not just following a plain unintelligent oversampling, but also allowing a more generalizable data augmentation method. Not only GAN is a good approach to address the class imbalance problems but it can be a better approach than normal random oversampling by focusing on a more generalizable sample generation which allows more intelligently induced synthetic samples. Hence, this data augmentation approach is much more practical and can be used in many applications that involve class imbalance which can be effectively resolved with synthetic examples generated by GAN to improve predictive performance.

2.3 Neural Networks

Rejwan Bin Sulaiman et al [14] compared machine training methods for credit card scam detection and Confidentiality of the info in the literature review and then provided a combined system model employing the Artificial neural network in an interconnected framework training. It has been discovered to be an effective method to raise CCFD Accuracy while Anonymity maintenance. An interconnected training model that uses an artificial neural network can increase the machine learning architecture's ability to recognize illegal activities. The proposed hybrid strategy, which makes use of real-world datasets, has the potential to significantly modify how CCFD is used while opening up new opportunities for the banking and financial sector.

2.4 Artificial Intelligence

Munira Ansari [16] et al used artificial intelligence to find credit card fraud. A publicly available data in the credit card set is used to assess the system's effectiveness. There is a potential for fraud because the system's prediction level and precision of fraud detection are not perfect. A class of genuine data in a card from a banking institution is utilized to continue analysis. Vibrations are also mixed with the sampling data in order to grant access to the algorithm's robustness further. Based on the investigations performed, the greater count experiment is best accurate in detecting occurrences of theft of credit cards.

2.5 Other Methods

Ayoub Mniai and Khalid Jebari [4] designed a hybrid methodology for detecting credit card fraud. Support Vector Data Description (SVDD) and Particle Swarm Optimization are combined in the hybrid solution (PSO). For instance, the efficiency of SVDD is determined by the random selection of two parameters, c and. The suggested model searches for the fastest possible solution to these two parameter optimization problems in order to improve accuracy. The findings indicate that the PSO-SVDD approach outperformed SVM in terms of performance accuracy.

Gayan K. Kulatilleke [8] analyzed sampling techniques as a practical pre-processing mechanism to manage encoded data that is susceptible to imbalance and proposed a data-driven classifier selection strategy for typical highly unbalanced fraud detection data sets. The selection strategy's model outperforms comparable models while operating under more realistic circumstances, demonstrating the strategy's usefulness. By outperforming even the most sophisticated generative models while employing the entire hugely uneven real data distribution, the strategy's effectiveness is validated.

S P Maniraj et al [17] While the algorithm does reach over 95% accuracy when one-tenth of the data set is considered, its precision remains only at 28%. The algorithm focuses on the application of multiple anomaly detection techniques, data analysis, pre-processing, the Isolation forest classifier, and Local outline Factor on the Pca-Transformed credit card transactions input. The precision rises to 33% when the proposed model is fed to the complete dataset, though. This high accuracy rate is anticipated in light of the stark disparity between the volume of genuine transactions and the volume of genuine transactions.

Sl.no	Author	Methodology Used					Dataset
		ML	DL	NN	AI	Others	
1	Chee Bhagyani et al [9]	✓					Risky Merchant Category Codes (MCC), Transactions larger than $100, risky ISO Response codes, and unknown web addresses [9]
2	S P Maniraj, Aditya Saini, Swarna Deep Sarkar, Shadab Ahmed [17]					✓	Kaggle dataset [17]
3	Dejan Varmedja et al [19]	✓					European dataset [19]
4	Vaishnavi Nath Dornadulaa, Geetha S [20]	✓					Credit card dataset [20]
5	Ruttala Sailusha et al [10]		✓				Credit card fraud [10]
6	Jeonghyun Hwang, Kangseok Kim [12]		✓				Credit card dataset, financial dataset [12]
7	Rajesh Sabari, rajeshsa [3]		✓				Credit card fraud
8	Asha RB, Suresh Kumar KR [2]	✓					European credit card dataset [2]
9	D. Tanouz et al [11]		✓				European card dataset [11]
10	Tushar Patil [13]		✓				European credit card dataset [13]
11	Munira Ansari, Hashim Malik et al [16]				✓		Financial dataset [16]

12	Aisha Mohammad Fayyomi, Derar Eleyan, Amina Eleyan [18]	✓					European credit card dataset [18]
13	B. N. V. Madhubabu, T. Vyshnavi, K. Ashok [22]	✓					Credit card dataset
14	Prerak Parekh et al [24]	✓					European credit card dataset [24]
15	Mary Isangediok, Kelum Gajamannage [1]	✓					Phishing website URLs, credit card fraud [1]
16	Ayoub Mniai, Khalid Jebari [4]					✓	European credit card, Synthetic Financial dataset [4]
17	J Femila Roseline et al [5]		✓				Kaggle source [5]
18	Sahil Negi, Sudipta Kumar Das, Rigzen Bodh [6]		✓				Credit card fraud [6]
19	Fawaz khaled alarfaj et al [7]		✓				European card dataset, The Brazilian dataset, Commercial banks in China, Cardholders dataset of Europe [7]
20	Gayan K. Kulatilleke [8]					✓	European dataset
21	Rejwan Bin Sulaiman, Vitaly Schetinin, Paul Sant [14]			✓			Australian credit card, German credit card dataset [14]
22	Emmanuel Ileberi, Yanxia Sun, Zenghui Wang [15]	✓					European credit card dataset [15]
23	Sairamvinay Vijayaraghavan, Terry Guan, Jason [21]		✓				Kaggle dataset [21]
24	Noor Saleh Alfaiz, Suliman Mohamed Fati [23]	✓					European credit card dataset [23]

Table1: Literature Summary

By studying the above literature, we came to the conclusion that some of them have used ml algorithms, some deep learning, and others artificial neural networks, where no one had used the GAN- LSTM combined model. So, we are using the Generative Adversarial Network and Long short-term memory for good accuracy output. The below diagram shows the proposed model of our project:

3. PROPOSED SYSTEM

The planned GAN-LSTM based on credit card fraud detection is described in the below section:

Dataset - The training datasets for this project from Kaggle. There are 31 characteristics in the Kaggle dataset, 28 of which have been made anonymous and are designated V1 through V28. The transaction's timing and value, as well as a label indicating whether or not it was fraudulent, make up the final three features [3]. In our research, the Python programming language, machine Learning, and several pieces of equipment were utilized to contrast multiple systems and suggest a credit fraud detection model [6].

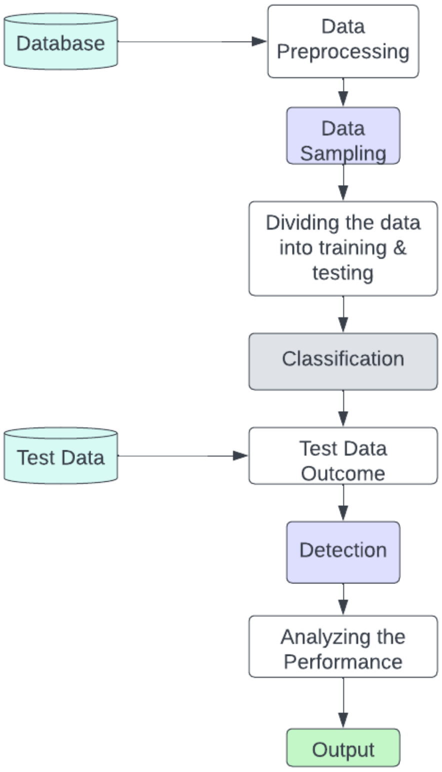

Figure1: Proposed Model

4. CONCLUSION

In recent years, credit card theft has increased dramatically. Methods for detecting fraud are always being improved to help lawbreakers counteract their ability to change their tactics. As fraud detection systems advance, it becomes simpler and quicker to identify fraud as soon as feasible once it has been detected. The methods that were examined using GAN-LSTM here are the ability to rapidly and effectively track down credit card fraud and reducing the credit card fraud.

REFERENCES:

[1] Mary Isangediok and Kelum Gajamannage, “Fraud Detection Using Optimized Machine Learning Tools Under Imbalance Classes,” arXiv, September 2022

[2] Asha RB and Suresh Kumar KR, “Credit card fraud detection using artificial neural network,” Elsevier, Global Transitions Proceedings, pp.35-41, January 2021

[3] Rajesh Sabari and rajeshsa, “Improving Credit Card Fraud Detection based on CNN/GAN”

[4] Ayoub Mniai and Khalid Jebari, “Credit Card Fraud Detection by Improved SVDD,” Proceedings of the World Congress on Engineering, Scopus, July 2022

[5] J Femila Roseline, GBSR Naidu, Dr. V. Samuthira Pandi, S Alamelu alias Rajasree and Dr.N. Mageswari, “Autonomous credit card fraud detection using machine learning approach,” Elsevier, June 2022

[6] Sahil Negi, Sudipta Kumar Das and Rigzen Bodh,” Credit Card Fraud Detection using Deep and Machine Learning,” International Conference on Applied Artificial Intelligence and Computing (ICAAIC 2022), September 2022

[7] Fawaz khaled alarfaj, iqra malik et al, “Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms,” IEEE Access, April 2022

[8] Gayan K. Kulatilleke, ” Credit card fraud detection - Classifier selection strategy,” arXiv, August 2022

[9] Anuruddha Thennakoon, Chee Bhagyani et al, “Real-time Credit Card Fraud Detection Using Machine Learning,” IEEE, 2019

[10] Ruttala Sailusha, V. Gnaneswar et al, “Credit Card Fraud Detection Using Machine Learning,” Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020) IEEE Xplore, June 2020

[11] D. Tanouz, R Raja Subramanian et al, “Credit Card Fraud Detection Using Machine Learning,” Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021) IEEE Xplore, July 2021

[12] Jeonghyun Hwang and Kangseok Kim, “An Efficient Domain-Adaptation Method using GAN for Fraud Detection,” (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 11, no. 11, 2020

[13] Tushar Patil, “Credit Card Fraud Detection Using Conditional Tabular Generative Adversarial Networks (CT-GAN) and Supervised Machine Learning Techniques,” National College of Ireland, September 2021

[14] Rejwan Bin Sulaiman, Vitaly Schetinin and Paul Sant, “Review of Machine Learning Approach on Credit Card Fraud Detection,” Human-Centric Intelligent Systems, Springer, March 2022

[15] Emmanuel Ileberi, Yanxia Sun and Zenghui Wang, “A machine learning based credit card fraud detection using the GA algorithm for feature selection,” Ileberi et al. Journal of Big Data, Springer Open, 2022

[16] Munira Ansari, Hashim Malik et al, “Credit Card Fraud Detection,” International Journal of Engineering Research & Technology (IJERT), vol.9, 2021

[17] S P Maniraj, Aditya Saini, Swarna Deep Sarkar and Shadab Ahmed, “Credit Card Fraud Detection using Machine Learning and Data Science,” International Journal of Engineering Research & Technology (IJERT), vol.8, September 2019

[18] Aisha Mohammad Fayyomi, Derar Eleyan and Amina Eleyan, “A Survey Paper On Credit Card Fraud Detection Techniques,” international journal of scientific & technology research, vol.10, September 2021
[19] Dejan Varmedja, Mirjana Karanovic et al, “Credit Card Fraud Detection - Machine Learning methods,” 18th International Symposium INFOTECH-JAHORINA, March 2019
[20] Vaishnavi Nath Dornadulaa and Geetha S, “Credit Card Fraud Detection using Machine Learning Algorithms,” international conference on recent trends in advanced computing, Elsevier, 2019
[21] Sairamvinay Vijayaraghavan, Terry Guan and Jason, “GAN-based Data Augmentation to Resolve Class Imbalance,” June 2022
[22] B. N. V. Madhubabu, T. Vyshnavi and K. Ashok, “Credit Card Fraud Detection Algorithm using Decision Trees based Random Forest Classifier,” Turkish Journal of Computer and Mathematics Education, vol.12, no.01, 2021
[23] Noor Saleh Alfaiz, Suliman Mohamed Fati, “Enhanced Credit Card Fraud Detection Model Using Machine Learning,” 2022
[24] Prerak Parekh et al, “credit card fraud detection with resampling techniques”
[25] https://resources.additionfi.com/examples-of-credit-card-fraud
[26] https://indianexpress.com/article/business/bank-fraud-card-internet-frauds-increase-with-rs-155-crore-in-3596-cases7940446/
[27] https://developers.google.com/machine-learning/gan/gan_structure