International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

DEEP LEARNING ALGORITHMS USING FRAUDULENT DETECTION IN BANKING DATASETS

S. Swathiga[1], J. Thanushya[2], A.Fatima[3]
[1,2]Assistant Professor, Department of CSE, PET Engineering College, Vallioor
[3]PG Student, Department of CSE, PET Engineering College, Vallioor

ABSTRACT

In today's world, high dependency on internet technology has enjoyed increased credit card transactions, but credit card fraud has also accelerated as an online and offline transaction. Financial fraud is a growing concern with far-reaching consequences for the government, corporate organizations, and the finance industry. The implementation of fraudulent detection in banking datasets using deep learning algorithms. It was developed by the Python Jupyter software. Initially, the input dataset is initiated by a preprocessing technique. The process is handled by data cleaning, which helps clean the datasets and, additionally, handles the missing values. Under pre-processing, the process started with data visualization process. Next, the data splitting process handles the data set and divides it for the purpose of the regression process. Deep learning algorithms are used in this work. The deep learning technique that handles the MLP algorithm is to predict fraudulent transactions and normal transactions. The final step is predicting whether the output of identification for fraudulent transactions is achieved in the model evaluation process.

Keywords- Credit card transactions, Finance industry, Fraudulent detection, Preprocessing technique, Python Jupyter software, Data cleaning, Data visualization process, Data splitting process, MLP algorithm.

I. INTRODUCTION

Credit card fraud is a huge ranging term for theft and fraud committed using or involving at the time of payment by using this card. The purpose be to purchase goods without paying, or to transfer unauthorized funds from an account. A credit card that has been stolen, lost, or counterfeited might result in fraud. Card-not-present fraud, or the use of your credit card number in e-commerce transactions has also become increasingly common as a result of the increase in online shopping. Increased fraud, such as CCF, has resulted from the expansion of e-banking and several online payment environments, resulting in annual losses of billions of dollars. In this era of digital payments, CCF detection has become one of the most important goals. As a business owner, it cannot be disputed that the future is heading towards a cashless culture. As a result, typical payment methods will no longer be used in the future, and therefore it will not be helpful for expanding a business. Customers will not always visit the business with cash in their pockets. It are now placing a premium on debit and credit card payments. As a result, companies will need to update their environment to ensure that it can take all types of payments. In the next years, this situation is expected to become much more severe. Credit card fraud is add on to identity theft. As per the information from the United States Federal Trade Commission, the theft rate of identity had been holding stable during the mid-2000s, but it was increased by 21 percent in 2008.

II. LITERATURE SURVEY

Alarfaj F.K, 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" (2022). Credit card frauds cause significant financial losses for both credit card holders and financial companies. In this research study, the main aim is to detect such frauds, including the accessibility of public data, high-class imbalance data, the changes in fraud nature, and high rates of false alarm.[1]

Ding Y, W. Kang, J. Feng, B. Peng and A. Yang, "Credit Card Fraud Detection Based on Improved Variational Autoencoder Generative Adversarial Network" (2023) The method is tested on an open credit card dataset, with the experimental results demonstrating that the oversampling method utilizing the improved VAEGAN is superior to the oversampling method of Generative Adversarial Network (GAN), Variational Auto encoder (VAE), and Synthetic. ‘‘Time’’ and ‘‘Amount’’, that have not undergone PCA conversion [2].

Ghaleb F.A, F. Saeed, M. Al-Sarem, S. N. Qasem and T. Al-Hadhrami, "Ensemble Synthesized Minority Oversampling-Based Generative Adversarial Networks and Random Forest Algorithm for Credit Card Fraud Detection" (2023). Most credit card frauds are conducted online by illegally obtaining payment credentials through data breaches, phishing, or scamming. Many solutions have been suggested to address the credit card fraud problem for online transactions.[3]

Ileberi E, Sun and Z. Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost" (2021). In this research, implement a machine learning (ML) based framework for credit card fraud detection using a real world imbalanced datasets that were generated from European credit cardholders. The models were evaluated using the accuracy, the recall, the precision, the Matthews Correlation Coefficient (MCC), and the Area under the Curve (AUC) [4].

Mienye I.D and Y. Sun, "A Deep Learning Ensemble with Data Resampling for Credit Card Fraud Detection"(2023). 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 [5].

III. EXISTING SYSTEM

Fraudulent transactions have become a growing problem in online banking, and fraud detection has always been challenging. Along with credit card development, the pattern of credit card fraud has always been updated.

Fraudsters do their best to make it look legitimate, and credit card fraud has always been updated. Fraudsters do their best to make it look legitimate. They try to learn how fraud detection systems work and continue to stimulate these systems, making fraud detection more complicated.

Therefore, researchers are constantly trying to find new ways or improve the performance of the existing methods.

Disadvantages of Existing System

1. Card-not-present fraud, or the use of your credit card information in e-commerce transactions, has also grown more widespread as online purchasing has increased.

2. Inaccurate Prediction. Machine learning algorithms require a large amount of high-quality data to be effective.

3. Difficult To Interpret.

IV. PROPOSED SYSTEM

In this work, Deep learning models and algorithms are used to detect and prevent online or banking frauds by analyzing large amounts of data to identify patterns and anomalies that may indicate fraudulent activity. This data can include transaction data, customer behaviour data, and device data.

Advantages of Proposed System

□ Detection of anomalies faster

□ Better Predictions

□ Saves Time and Money

V. SYSTEM ARCHITECTURE

First, the input dataset is allowed to undergo preprocessing. Which involves cleaning and analyzing the input dataset and handling the missing values. Following data pre-processing, the next stage of data visualization Matplotlib and Seaborn tools are used in the data visualization process. These modules are used for observing, exploring, and understanding the data in detail. Next, the data splitting process handles the data set and divides training and testing data for the purpose of the regression and identification process. Then the datasets are allowed to be classified using the classification techniques of deep learning. The deep learning technique handles the MLP algorithm. It works to handle robustness, missing data, and feature importance analysis against over fitting. Which makes it a popular option for a variety of uses, including prediction and detection issues. Additionally, the classification techniques provide excellent specificity and sensitivity. Following that, data validation is employed to improve integrity. Finally, the predicted fraudulent transaction is identified in the evaluation block. Overall, the proposed project was developed in Python, employing the Jupyter software.

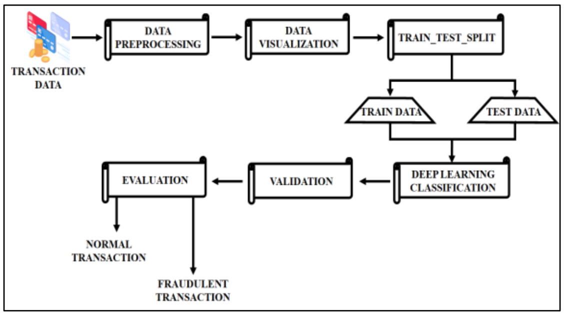


Fig 5.1 System Architecture

VI. SYSTEM IMPLEMENTATION

6.1 MODULES

1. Data Preprocessing
2. Data visualization
   2.1 Matplotlib
   2.2 Seaborn
3. Classification
4. Validation and evaluation

6.2 MODULE DESCRIPTION

6.2.1 DATA PREPROCESSING

Data Preprocessing, a component of data preparation, describes any type of processing performed on raw data to prepare it for another data processing procedure. It has traditionally been an important preliminary step for the data mining process.

The need for data Preprocessing is there because good data is undoubtedly more important than good models and for which the quality of the data is of paramount importance.

6.2.2 DATA VISUALIZATION

Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from the main goal of data visualization is to make it easier to identify patterns, trends and outliers in large data sets.

Data visualization is a field in data analysis that deals with visual representation of data. It graphically plots data and is an effective way to communicate inferences from data.

These processes are handled in two modules: Matplotlib and Seaborn.

6.2.2.1 Matplotlib

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible.

6.2.2.2 Seaborn

Seaborn is a library for making statistical graphics in Python. It builds on top of Matplotlib and which work on exploring and understanding the data.

6.2.3 CLASSIFICATION

Data classification is the process of organizing data into categories that make it easy to retrieve, sort and store for future use. A well-planned data classification system makes essential data easy to find and retrieve. Data classification is broadly defined as the process of organizing data by relevant categories so that it may be used and protected more efficiently. On a basic level, the classification process makes data easier to locate and retrieve.

6.2.4 VALIDATION AND EVALUATION

Validation and evaluation are related processes that can be used in a variety of contexts, including machine learning, training, and software testing. Validation in machine learning is like an authorization or authentication of the prediction done by a trained model. A validation data set is used in deep learning to compare the performance of different trained models. Data validation means checking the accuracy and quality of source data before using, importing or otherwise processing data. While on the other hand, evaluation in machine learning refers to assessment or test of entire machine learning model and its performance in various circumstances. The method of evaluation focuses on the accuracy of the model in predicting the end outcomes. Even though there are several stages, the stage of evaluating a

DL model is the most crucial because it gives us an idea of the accuracy of model prediction.

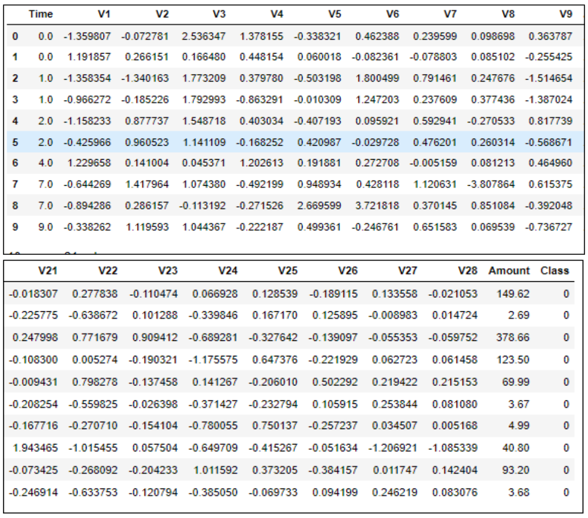
VII. RESULTS AND DISCUSSION

The chosen language for this project was Python. Python is a general-purpose, high-level programming language. Python is a popular high-level, general-purpose programming language. In 1991, Guido van Rossum invented it, and the Python Software Foundation continued to develop it. Programmers may communicate their ideas in less lines of code because to its syntax, which was developed with readability of code as a primary focus. For many reasons, this was an easy decision. The Python language is supported by a large community. A visit to Stack Overflow can easily fix any issues that may arise.

It lets you install packages without changing how Python is installed on your system. The Anaconda program facilitates the creation of environments for several Python and package versions. In your project environments, Anaconda is also used for package installation, removal, and upgrades. With the Jupyter Notebook, an open-source online tool, you can create and distribute documents with narrative text, equations, live code, and visualizations. Python is a programming language that lets you work quickly and integrate systems more efficiently.

Input Dataset

This section presents results obtained from analyzing several banking samples to ascertain whether the byte sequences extracted by the proposed method provide useful information for manual analysis. Figures 7.1 display the input dataset of the fraud detection.



	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739
5	2.0	-0.425966	0.960523	1.141109	-0.168252	0.420987	-0.029728	0.476201	0.260314	-0.568671
6	4.0	1.229658	0.141004	0.045371	1.202613	0.191881	0.272708	-0.005159	0.081213	0.464960
7	7.0	-0.644269	1.417964	1.074380	-0.492199	0.948934	0.428118	1.120631	-3.807864	0.615375
8	7.0	-0.894286	0.286157	-0.113192	-0.271526	2.669599	3.721818	0.370145	0.851084	-0.392048
9	9.0	-0.338262	1.119593	1.044367	-0.222187	0.499361	-0.246761	0.651583	0.069539	-0.736727

V21	V22	V23	V24	V25	V26	V27	V28	Amount	Class
-0.018307	0.277838	-0.110474	0.066928	0.128539	-0.189115	0.133558	-0.021053	149.62	0
-0.225775	-0.638672	0.101288	-0.339846	0.167170	0.125895	-0.008983	0.014724	2.69	0
0.247998	0.771679	0.909412	-0.689281	-0.327642	-0.139097	-0.055353	-0.059752	378.66	0
-0.108300	0.005274	-0.190321	-1.175575	0.647376	-0.221929	0.062723	0.061458	123.50	0
-0.009431	0.798278	-0.137458	0.141267	-0.206010	0.502292	0.219422	0.215153	69.99	0
-0.208254	-0.559825	-0.026398	-0.371427	-0.232794	0.105915	0.253844	0.081080	3.67	0
-0.167716	-0.270710	-0.154104	-0.780055	0.750137	-0.257237	0.034507	0.005168	4.99	0
1.943465	-1.015455	0.057504	-0.649709	-0.415267	-0.051634	-1.206921	-1.085339	40.80	0
-0.073425	-0.268092	-0.204233	1.011592	0.373205	-0.384157	0.011747	0.142404	93.20	0
-0.246914	-0.633753	-0.120794	-0.385050	-0.069733	0.094199	0.246219	0.083076	3.68	0

Fig7.1 Input Dataset

Data Distributions

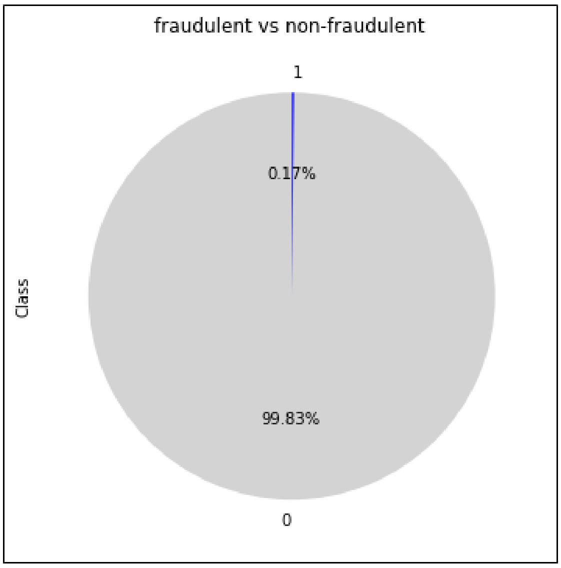


Fig 7.2 Data distributions for each class

Figure 7.2 displays the data distribution of the fraudulent data and non-fraudulent.

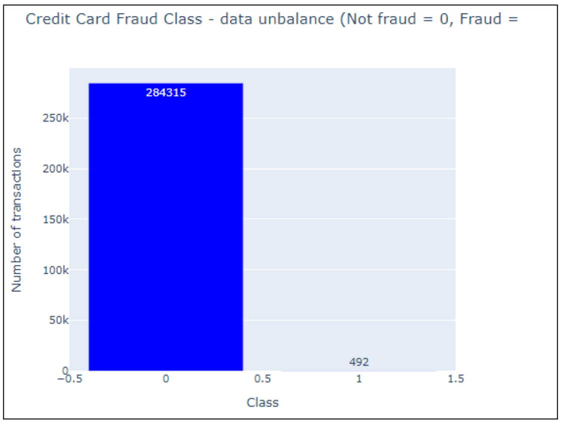


Fig 7.3 Class distributions

Data distributions for each class between normal transaction and fraud transaction image is displayed in Figure 7.3.

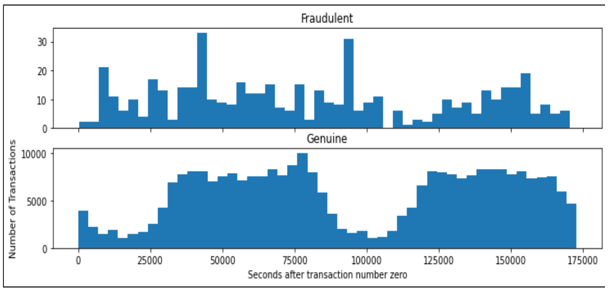


Fig 7.4 Exploration of transaction

Figure 7.4 shows illustrations of the exploration of transactions. These histogram images are deep exploration of the dataset based on the number of transactions.

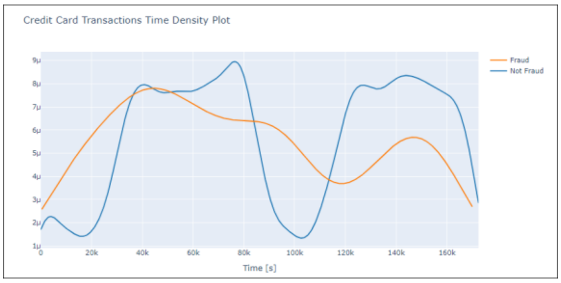


Fig 7.5 Time density plot

The time density plot is shown in Figure 7.5. In this figure, the blue color line represents the normal transaction, and the orange color line represents the fraud transaction based on time.

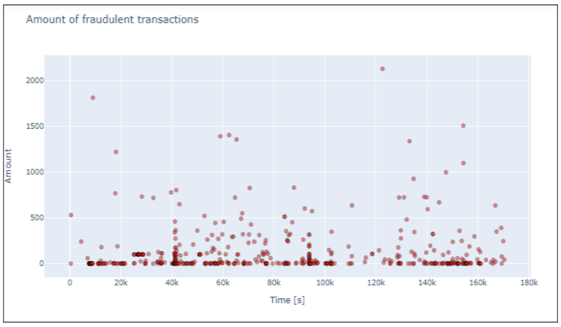


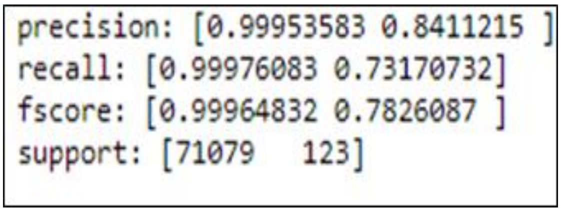
Fig 7.6 Fraudulent Transaction

Figure 7.6 establishes the fraudulent transaction.

Multilayer Perceptron

A multilayer perceptron is a neural network connecting multiple layers in a directed graph, which means that the signal path through the nodes only goes one way. Each node, apart from the input nodes, has a nonlinear activation function. This figure represents that performance of MLP. Figure 7.7 shows that performance of MLP.

Performance of MLP



```
precision: [0.99953583 0.8411215 ]
recall: [0.99976083 0.73170732]
fscore: [0.99964832 0.7826087 ]
support: [71079   123]
```

Fig 7.7 Performance of MLP

Confusion Matrix of MLP

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

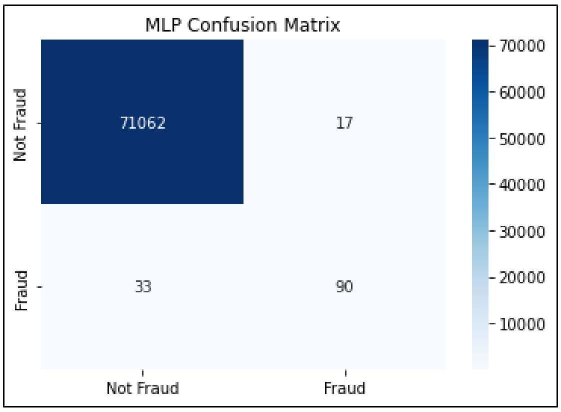


Fig 7.8 Confusion matrix for MLP

Figure 7.8 represents that confusion matrix for MLP. In this figure shows that value for malign and benign. It is often used to measure the performance of classification models, which aim to predict a categorical label for each input instance.

Roc Curve of MLP

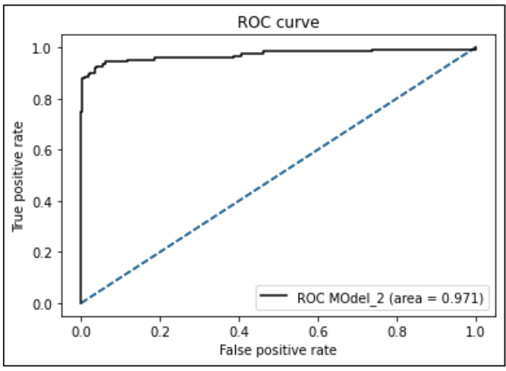


Fig 7.9 ROC Curve for MLP

Figure 7.9 represents a multilayer perceptron algorithm ROC Curve. An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters: True Positive Rate. False Positive Rate.

Precision_Recall Curve for MLP

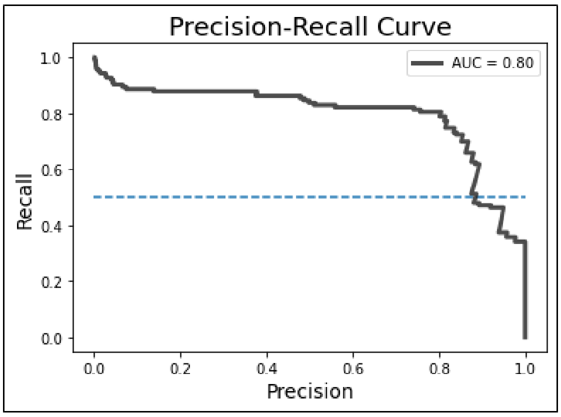


Fig 7.10 Precision_Recall Curve for MLP

A multilayer perceptron algorithm (MLP) precision-recall curve is shown in Figure 7.10. A precision-recall curve is a plot of the precision (y-axis) and the recall (x-axis) for different thresholds, much like the ROC curve. A no-skill classifier is one that cannot discriminate between the classes and would predict a random class or a constant class in all cases.

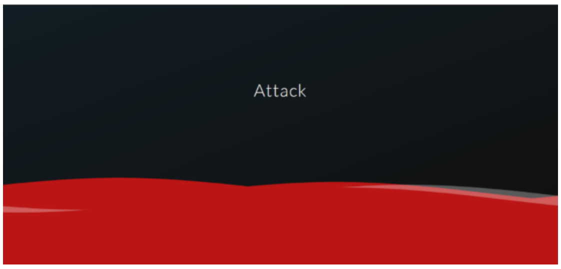

Attack

Fig 7.10 Fraud Detection

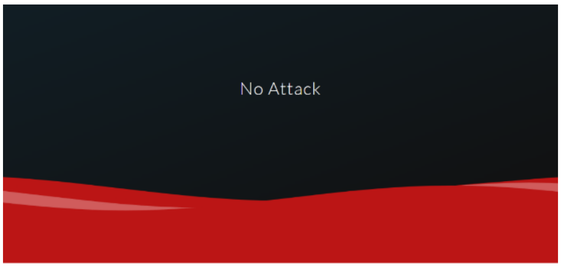

No Attack

Fig 7.11 No Fraud Detection

VIII.CONCLUSION AND FUTURE ENHANCEMENTS

The detection of credit card fraud is a vital research field. This is because of the increasing number of fraud cases in financial institutions. This project opens the door for employing machine learning to build systems that can detect fraud. Building an automated-based system to detect fraud requires a database to train the system (or classifier). This work demonstrates the advantages of applying deep learning techniques, including MLP classification techniques, to the credit card fraud detection problem for the purpose of reducing the bank's financial risks. Finally, the proposed classifier is evaluated based on its accuracy, and the MLP classifier generates the best results. Using these methods for the detection of credit cards yields better performance than traditional algorithms.In future work, an efficient hybrid deep learning method will be developed for enhanced accuracy and performance compared to traditional algorithms.

REFERENCES

[1] Alarfaj.F.K, 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.

[2] Bhanusri, A., Valli, K.R.S., Jyothi, P., Sai, G.V. and Rohith, R., 2020. Credit card fraud detection using Machine learning algorithms. Journal of Research in Humanities and Social Science, 8(2), pp.04-11.

[3] Ding.Y, W. Kang, J. Feng, B. Peng and A. Yang, "Credit Card Fraud Detection Based on Improved Variational Autoencoder Generative Adversarial Network," in IEEE Access, vol. 11, pp. 83680-83691, 2023.

[4] Ghaleb.F.A, F. Saeed, M. Al-Sarem, S. N. Qasem and T. Al-Hadhrami, "Ensemble Synthesized Minority Oversampling-Based Generative Adversarial Networks and Random Forest Algorithm for Credit Card Fraud Detection," in IEEE Access, vol. 11, pp. 89694-89710, 2023.

[5] Ileberi.E, 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.

[6] Lunghi.D, G. M. Paldino, O. Caelen and G. Bontempi, "An Adversary Model of Fraudsters' Behavior to Improve Oversampling in Credit Card Fraud Detection," in IEEE Access, vol. 11, pp. 136666-136679, 2023.

[7] Lebichot.B, 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.

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

[9] Ning.W, S. Chen, S. Lei and X. Liao, "AMWSPLAdaboost Credit Card Fraud Detection Method Based on Enhanced Base Classifier Diversity," in IEEE Access, vol. 11, pp. 66488-66496, 2023.

[10] Roy.A, J. Sun, R. Mahoney, L. Alonzi, S. Adams and P. Beling, "Deep learning detecting fraud in credit card transactions," 2018 Systems and Information Engineering Design Symposium (SIEDS), pp. 129-134, 2018.

[11] San Miguel Carrasco.R and M. -Á. Sicilia-Urbán, "Evaluation of Deep Neural Networks for Reduction of Credit Card Fraud Alerts," in IEEE Access, vol. 8, pp. 186421-186432, 2020.

[12] Tingfei.H, C. Guangquan and H. Kuihua, "Using Variational Auto Encoding in Credit Card Fraud Detection," in IEEE Access, vol. 8, pp. 149841-149853, 2020.

[13] Trivedi, N.K., Simaiya, S., Lilhore, U.K. and Sharma, S.K., An efficient credit card fraud detection model based on machine learning methods. International Journal of Advanced Science and Technology, 29(5), pp.3414-3424, 2020.

[14] Thennakoon.A, C. Bhagyani, S. Premadasa, S. Mihiranga and N. Kuruwitaarachchi, "Real-time Credit Card Fraud Detection Using Machine Learning," 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence),pp. 488-493, 2019.

[15] Yee, O.S., Sagadevan, S. and Malim, N.H.A.H., Credit card fraud detection using machine learning as data mining technique. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 10(1-4), pp.23-27, 2018.