ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03(04), April 2024, pp. 41-53

Article Title: Implementation of A Deep Learning Algorithm for Fraud Detection Based on Credit Transactions

Implementation of A Deep Learning Algorithm for Fraud Detection Based on Credit Transactions

R. Sahila Devi

Assistant Professor, Department of Computer Science and Engineering, Rohini College of Engineering and Technology, Palkulam, India, 1r.sahiladevi@gmail.com

ABSTRACT

Online transactions using credit cards are being more reliable and efficient method followed by most of the users. In the same way, the fake transaction and scam occurring during the credit card transactions are increasing day by day. This scam not only steal user's money but also misuse their personal details. Manual detection of these scam may be a tedious process, so this detection must be a deep learning method for the fastest and effective method of detection. This paper proposes a methodology of deep learning algorithm involving Multi-Layer Perceptron (MLP) to prevent the users from fraud transaction. The data which is trained form a neural network and the test data is evaluated to define the classification, whether the data is fraud or verified data. This work involves data processing, data cleaning and data visualization of the data transaction. At the end of the process, the detected results undergo validation and model evaluation. Data visualization uses libraries from Matlab to learn the data, which is used in the data classification.

Keywords: Deep Learning (DL), Artificial Intelligence (AI), Data processing, Data visualization, Data cleaning.

1 Introduction

As the use of internet and digital money increased, occurrence of scam is also getting its highest peak. Especially in the area of money transaction, there are lots of scammers involved. There are lot of fraud detection techniques involved during the detection of these scams. The process of detection becomes more efficient when deep learning is involved in the process and the method must be reliable to execute [1]. The data for which the result must be fraudulent is already trained for reference, which generate set of transactions which is considered as the scam. These techniques will improve the effectiveness of detection machines [2]. The data must be a normalised dataset, so that the final detection is made much more accurate. This normalization is carried out by the process of pre-processing, where the features from the transaction information are obtained and further processed [3].

Thus to make the detection more accurate, the data preprocessing is the vital stage in the deep learning process. For the process of reliable detection the system must have fast pre-processing time for the further preprocessing of the data. By implementing fast pre-processing execution, the size is optimized without reducing the key features. Additionally, the data cleaning process

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03(04), April 2024, pp. 41-53

Article Title: Implementation of A Deep Learning Algorithm for Fraud Detection Based on Credit Transactions

is continued by measuring the accuracy of the data present, so that the needed information will not be removed [4]. This continuous preprocessing makes the data to be renewed with only the relevant information in comparison to the fetched data. For the quick transaction, the processing time is analysed to be 6 milliseconds and as average it is 0.8 milliseconds [5].

After the preprocessing, the transaction data undergo data visualization, which makes the data more reliable to handle and understand. In data visualization, the information of the data gathered with be more reliable and that makes even a larger data to easily handle. The data processing speed must be fast to make the data visualization more accessible. As it deals with the information, this is one of the important tools in the data processing [6]. Data visualization deals with the data by giving a visual impact to the task transferred. This deals with the two sets of data namely source data and the target data [7]. All the data thus collected and then processed are also revised during the data visualization, which makes the detection more perfect and also increase the effective processing while dealing with the fraud transaction datasets [8].

When the features of the data are made reliable and easy to handle, the data undergo classification. This classification is carried out in google colab to make the classification task more agile. This classification is mainly concentrated to make the model more reliable and to increase the quality. The deep learning method becomes more efficient which involving the binary classification tasks [9]. As fraud detection must be carried out in real-time, the binary classification will give more accuracy. The classification process is done with the classifier based on specific metrics like accuracy and the other feature which gives more efficiency to the detection. This classifier as the final result, classifies the output as fraudulent or not [10].

This paper propose the identification of fraud transactions in banking datasets using deep learning methods. Under pre-processing, the process is handled by data cleaning, which helps to clean the datasets and additionally, handles the missing values. By data visualization, the patterns in the dataset is clearly understood and integrity is improved. The data splitting process divides the processed dataset for the regression process. Deep learning algorithms gives more accurate results as it form the neural network from the trained dataset and process beyond the available dataset. This work is developed by using the Matlab libraries.

2 Recent works

Wang *et al* [11] [2023] have developed a credit card fraud detection method to increase the robustness of the method already being used. To improve the already used AdaBoost technique, Adaptive hybrid learning algorithm is implemented. This methods is a self-paced algorithm which follows a threshold level to find the accurate algorithm. This self-paced algorithm is then embedded to the existing algorithm. Various results of AdaBoost algorithms are implemented and the results are compared to find the effective threshold value to be applied. By comparing all the result values, the accurate threshold value gives good AUC value and make the detection

to be faster than the other methods. The direction of the decision making remains unchanged, which reduces the efficiency of the detection.

Zhaohui *et al* [12] [2020] have demonstrated an approach to make the low-frequency data transaction more efficient. This is done by divide the data into high frequency data and the low frequency data, which is followed by a threshold determination method. The transaction data is obtained by the sliding window operation. As many results obtained, the Naïve bayes algorithm is used in the process of comparing and calculating the probability of the fraud detection. This exhibit good effective detection even for the low frequency data transactions and reduces the misjudgement of the normal transactions. By detecting both the low and the high frequency data transactions, the disturbance during the transaction is reduced. The dividing of the high and low frequency transaction reduces the efficiency of the system and make the process more complex and time-consuming.

Yuanming *et al* [13] [2023] have analysed the generator section of Variational Autoencoder Generative Adversarial Network (VAEGAN). This method also introduce a new sampling method, as the traditional sampling method is not much efficient in detecting the transaction in faster rate. By introducing new sampling along with the XGBoost methodology, the process of classification is make faster and simpler. This classification is carried out by considering the base sample dataset from the already trained dataset. The merit is that it is easily extended to other available applications to make the operation more manageable and reduces the class imbalance. The disadvantage of this methodology is, the recall at certain ratio is not significant compared to other methods and this will increase the complexity.

Haibo *et al* [14] [2022] have implemented a proposed framework applying the Quantum Machine Learning (QML) algorithm and using the Support Vector Machine (SVM). Various machine learning techniques are implemented on the different datasets which are highly imbalanced or moderately imbalanced. By involving the QML algorithm, the potential of solving the critical problem occurring in the classification of the data is reduced. The merit is, by using the SVM technique, both the accuracy and speed of the detection is increased and the wise methodology is chose for different types of transaction datasets. The demerit is as the method deals with the time bases series of datasets, the non-time based datasets cannot be computed with high accuracy and efficiency.

Ebenezer *et al* [15] [2022] have introduced effective method to detect the credit card fraudulent using neural network classifier and implementing data resampling method. A method which proves more efficient than using convolutional networks is deployed. The classifier is built by Long-Short Term memory (LSTM) neural network and AdaBoost as the base learner. The dataset being processed must be balances, this is carried out by the SMOTENN algorithm. This algorithms and methods altogether forms to enhance the robustness of the deep learning model. The advantage of using this algorithm is, the sensitivity

and the specificity is increased. The disadvantage is, this process is time consuming and complex.

3 Proposed work explanation

This work proposes the identification of fraud transactions in banking datasets using deep learning methods. Figure 1 represents the architecture of the proposed methodology to implement the process of fraud detection.

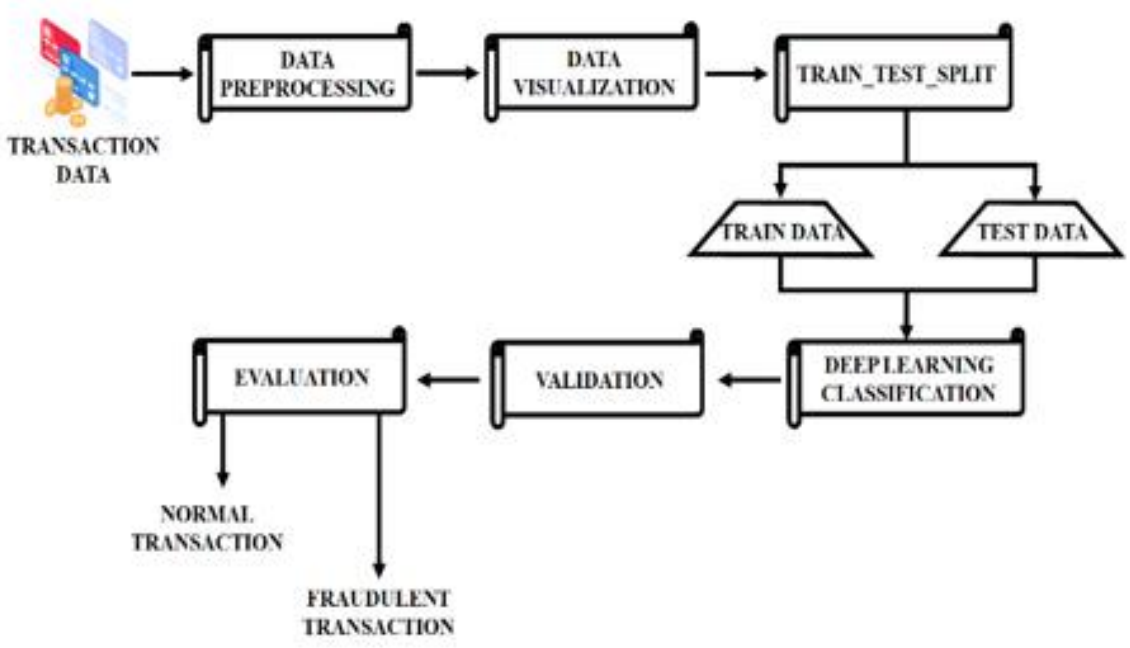

Figure 1: *Block diagram of proposed methodology*

The data preprocessing is done for making the main features of the data to be more visible and accurate. The data cleaning is the major step involved in the data processing. In data visualization, the important data in the large dataset is specified with any different way of visualization for better understanding of the data fed as input. The deep learning technique handles the MLP algorithm. It works to handle robustness, missing data, and feature importance analysis against overfitting. Additionally, using MLP as the classification techniques improves the specificity, sensitivity and integrity. The proposed work is developed in Matlab software.

4 Modelling

The modelling of this proposed system comprises of the following processes, Data processing, Data cleaning, Data visualization, Deep learning classification, validation and model evaluation. When the data of transaction is read by the machine, the preprocessing is assigned. During the preprocessing all the unwanted information and distortion present in the data will

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03(04), April 2024, pp. 41-53

Article Title: Implementation of A Deep Learning Algorithm for Fraud Detection Based on Credit Transactions

be removed by the method of data cleaning. Then the preprocessed image undergo data visualization, where the data present are represented in the manner by which the information are easily understandable and accessible for further classification. During the data classification, MLP classification algorithm is used and gather knowledge beyond the data provided. After the completion of the validation and the model evaluation, the data obtained is concluded that whether it is fraudulent or not.

4.1 Data Pre-processing

Data preprocessing is the process of processing raw data so that it is suitable for further processing and analysis of deep learning. This provide training to the dataset and assist the system to make more accuracy by enhancing the features inbuilt in the data transaction details. Data preprocessing is necessary in deep learning as in most cases, raw data has various issues like missing values, irrelevant feature variables, inconsistent column values, etc. the input dataset is allowed to undergo preprocessing, which involves cleaning and analyzing the input dataset and handling the missing values. Usually, this process is used to conduct the extraction of the data by reducing the noise distortions. By doing so, the main features in the data will be clearer and this makes the process of visualization simpler and efficient.

4.2 Data Cleaning

The data is not in a specific range, so that the preprocessing technique is used to fit the dataset into an exact range. This helps in improving the quality of data so as to make the analysis of the classification simpler. Among them, the data cleaning process is essential. In this work, the data cleaning technique is used for cleaning the data, removing null values, and handling the missed values of the datasets. Data cleaning mainly focus on the data quality enchantment, which is highly essential in the data processing methods. The unwanted information and distortion in the input data is removed for the further classification of the data.

4.3 Data Visualization

Data visualization depicts the visual representation of the data gathered and analyzed. Even large datasets is easily processed and learned by visually representing them. The main objective of this method is to find the main data from the large dataset. This prevent from the mismatch of data for the classification purpose. These processes are handled in two modules in the Matlab. These modules are used for observing, exploring, and understanding the data in detail.

4.4 Deep Learning Classification

The deep learning concept is a developed form of the machine learning concept that extract the features of the data fed automatically and process with those data to gather advanced information. The deep learning technique involves forming an artificial neural network to understand the data beyond the trained data by learning the existing dataset. The methodology

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03(04), April 2024, pp. 41-53

implemented is multilayer perceptron (MLP). The datasets are allowed to be classified using the classification techniques of deep learning. This process is done to train the models by using the MLP approach that deals with larger datasets. The data splitting process is type of classification, which handles the data set and divides training and testing data for the purpose of the regression and identification process.

MLP:

A multilayer perceptron (MLP) is made up of multiple layers, each of which is connected to the next. This algorithm uses Back propagation algorithm, where a strong artificial neural network is built. Each node is a processing element or a neuron that has a nonlinear activation function except the input nodes. It employs back propagation, a supervised learning technique, to train the network. It is a feed forward artificial neural network model that maps input sets to output proper sets. The final stage is the prediction process, output is generated by the output layer using the data processed in the hidden layer.

4.5 Validation

While developing a system, the data validation and evaluation is vital. Data validation is used in deep learning to compare the performance of different trained models. Data validation deals with checking the accuracy and quality of source data before using, importing or otherwise being processed. It gives a brief validation of how the data need to be processed and classified with giving a comparison dataset. This gives an effective approach of operational validation of the fraud detection. They concludes where the data processed and the result obtained is valid.

4.6 Evaluation

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. More accurate representation of the system is presented by the operational evaluation, even for larger transaction datasets. It concluded whether the data classification is accurate or not. They involve real time dataset for the further processing and improvement in the processing.

5 Results and Discussion

The result shows the output obtain by implementing the proposed method in the detection of fraud transactions using deep learning method.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03(04), April 2024, pp. 41-53

Article Title: **Implementation of A Deep Learning Algorithm for Fraud Detection Based on Credit Transactions**

	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

Figure 2: *Input Datasets*

Figure 2 represent the input data collected from the transaction of data during the credit card transactions

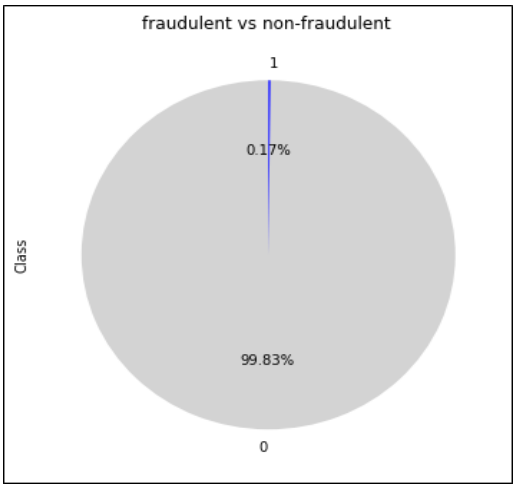

Figure 3: *Data Distribution*

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03(04), April 2024, pp. 41-53

Figure 3 is the pie chart representing the distribution of the data to detect the data whether it is fraudulent or not.

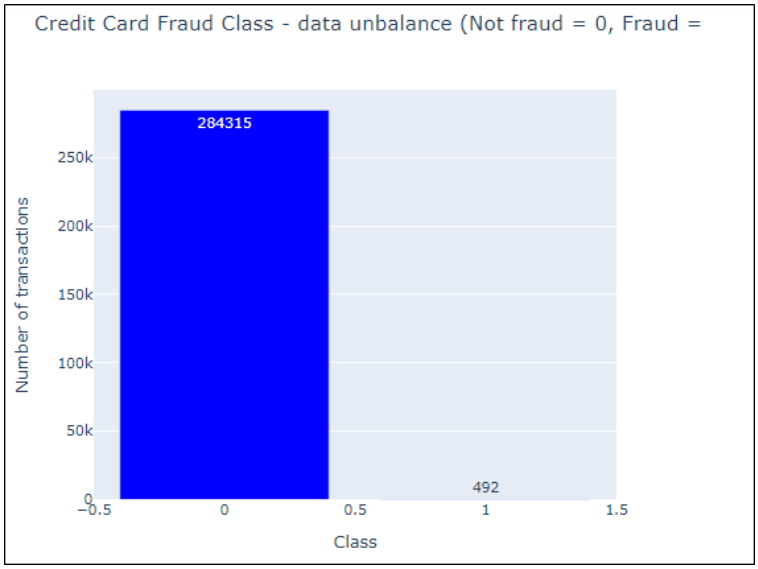

Figure 4: *Class Distribution*

Figure 4 represents the distribution of class based on number of transactions and the class already classified.

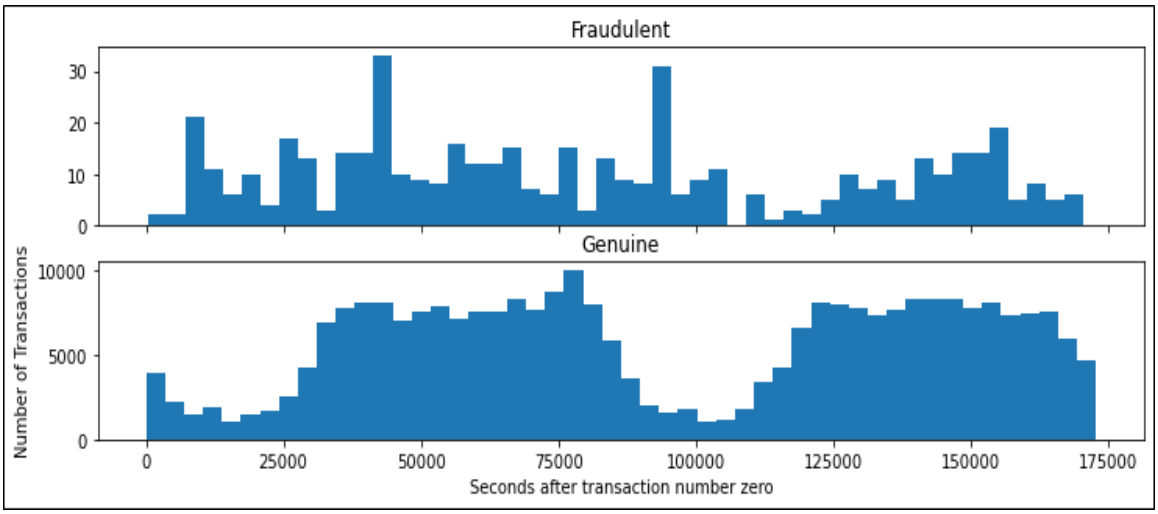

Figure 5: *Exploration of transaction*

Figure 5 represents the transaction exploration.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03(04), April 2024, pp. 41-53

Article Title: Implementation of A Deep Learning Algorithm for Fraud Detection Based on Credit Transactions

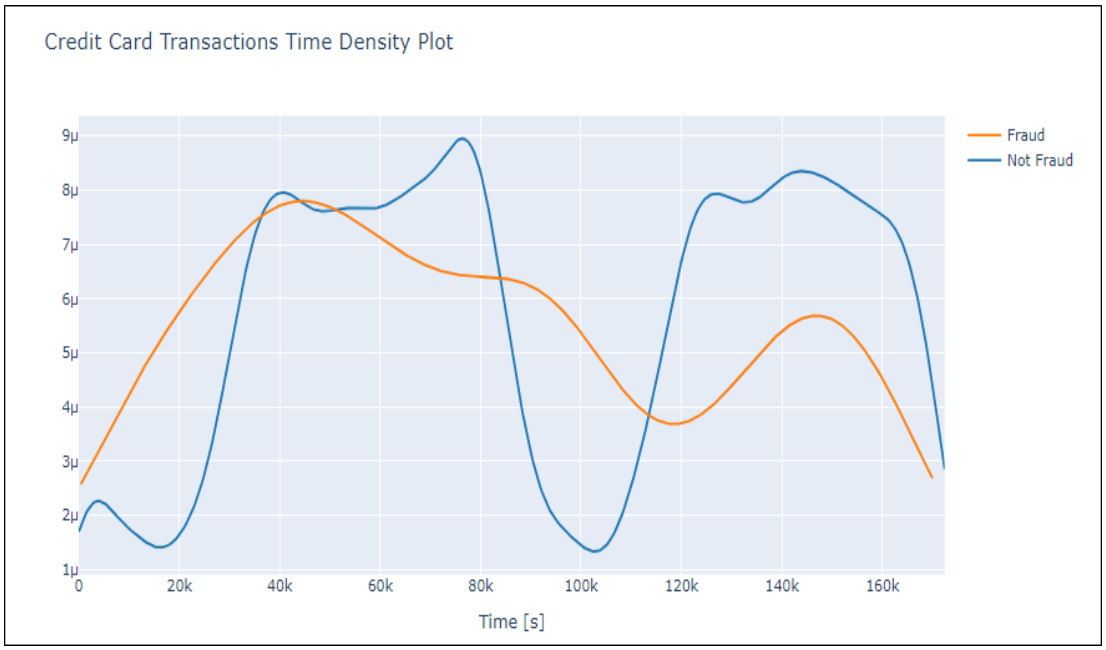

Figure 6: *Transaction time density plot*

Figure 6 represents the transaction time density plot. This density plot is the representation of the numerical variable change between the time and transactions being captured.

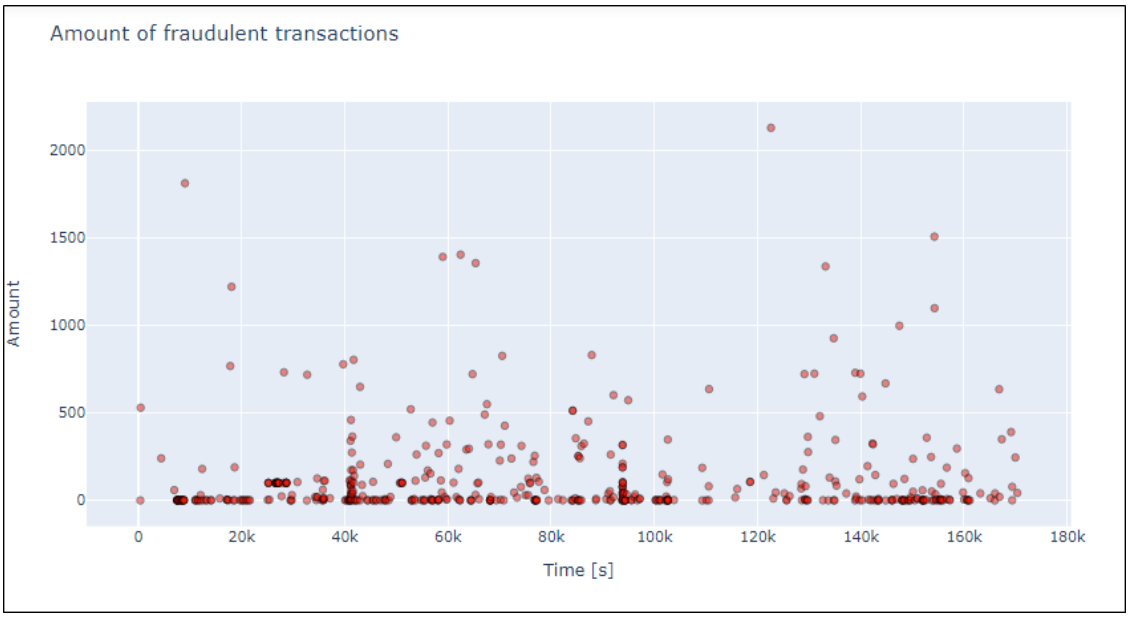

Figure 7: *Fraudulent Transaction*

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03(04), April 2024, pp. 41-53

Article Title: Implementation of A Deep Learning Algorithm for Fraud Detection Based on Credit Transactions

Figure 7 represents the fraudulent transaction for a particular mentioned time duration. This chart represents the increase in the amount of fraud transactions as the time varies.

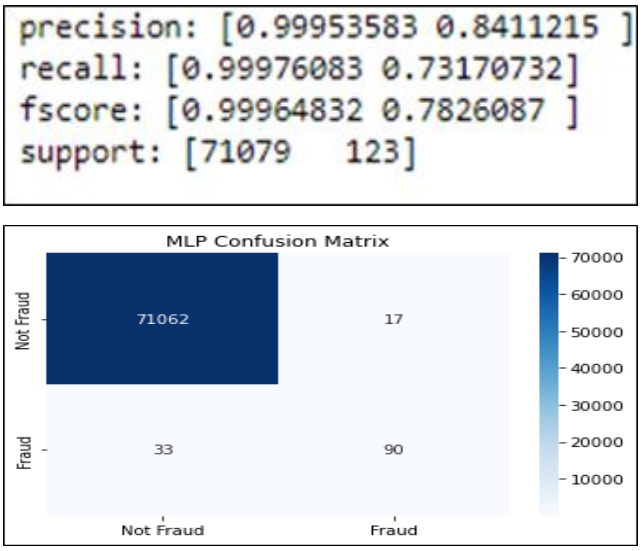

Figure 8: *Confusion matrix of MLP*

Figure 8 represents the confusion matrix, in which the classifier's performance is measured by comparing the different classes and categories specified.

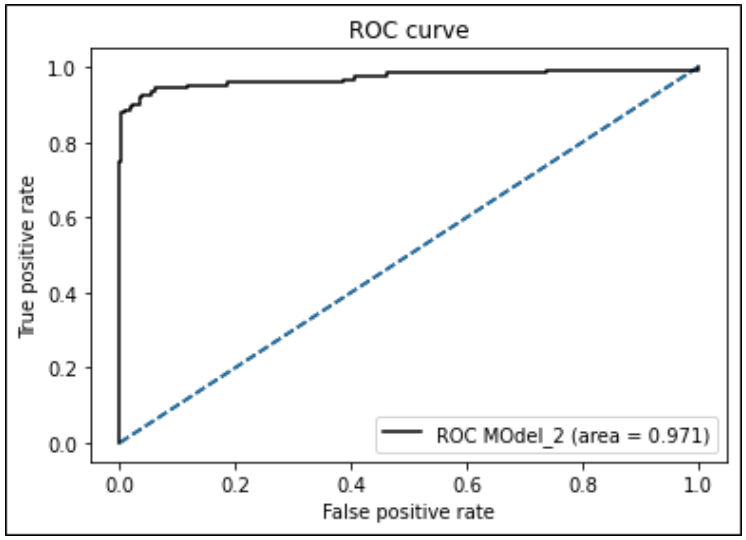

Figure 9: *ROC curve of MLP*

Figure 9 represents the ROC curve, which is especially used for the binary classifications, that compares the values between the false positive value and the true positive value.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03(04), April 2024, pp. 41-53

Article Title: Implementation of A Deep Learning Algorithm for Fraud Detection Based on Credit Transactions

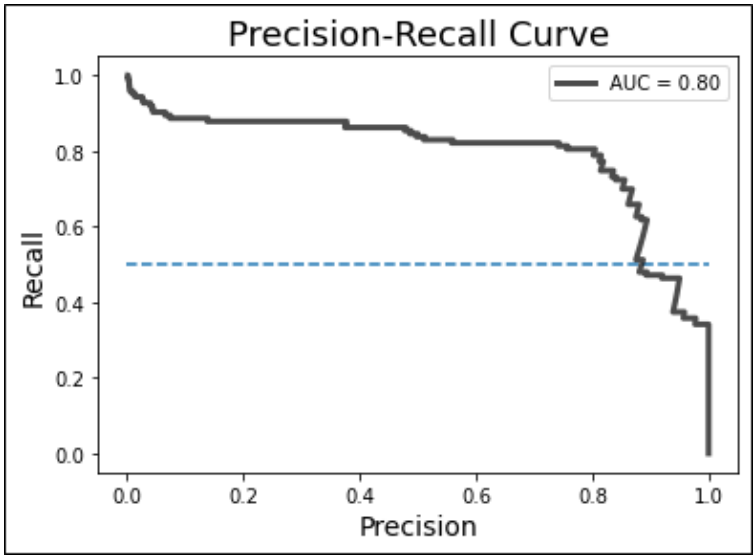

Figure 10: *Precision recall curve of MLP*

Figure 10 represents the Precision recall curve is obtained as a result of comparing the precise value and the recall value.

6 Conclusion

The detection of credit card fraud is essential in data sharing field. This work is based on deep learning to build systems that detects the fraud transactions. The deep learning technique handles the MLP algorithm used for the classification. To build an automated-based system to detect fraud requires a database to train the system. It works to handle robustness, missing data, and feature importance analysis against overfitting. This work demonstrates the advantages of applying deep learning techniques, including MLP classification techniques for the reliability and accuracy of the detection. 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 the other algorithms.

References

1. Ileberi Emmanuel; Yanxia Sun; Zenghui Wang, Year: 2022, "A machine learning based credit card fraud detection using the GA algorithm for feature selection", Journal of Big Data, Vol: 9, no: 1, pp. 24.

2. Lunghi Daniele; Gian Marco Paldino; Olivier Caelen; Gianluca Bontempi, Year: 2023, "An Adversary Model of Fraudsters' Behavior to Improve Oversampling in Credit Card Fraud Detection", Vol: 11, pp. 136666-136679.

3. Ghaleb Fuad; Faisal Saeed; Mohammed Al-Sarem; Sultan Noman Qasem; Tawfik Al-Hadhrami, Year: 2023, "Ensemble Synthesized Minority Oversampling based Generative Adversarial Networks and Random Forest Algorithm for Credit Card Fraud Detection", IEEE Access.

4. Alarfaj Fawaz Khaled; Iqra Malik; Hikmat Ullah Khan; Naif Almusallam; Muhammad Ramzan; Muzamil Ahmed, Year: 2022, "Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms", IEEE Access, Vol: 10, pp. 39700-39715.

5. Can Baris; Ali Gokhan Yavuz; Elif Karsligil; Amac Guvensan, Year: 2020, "A closer look into the characteristics of fraudulent card transactions", IEEE Access, Vol: 8, pp. 166095-166109.

6. Yang Ying; Tao Chen, Year: 2019, "Analysis and visualization implementation of medical big data resource sharing mechanism based on deep learning", IEEE Access, Vol: 7, pp. 156077-156088.

7. Lebichot Bertrand; Theo Verhelst; Yann-Ael Le Borgne; Liyun He-Guelton; Frederic Oble; Gianluca Bontempi, Year: 2021, "Transfer learning strategies for credit card fraud detection", IEEE access, Vol: 9, pp. 114754-114766.

8. Fahdah Almarshad; Ghada abdalaziz gashgari; Abdullah Alzahrani, Year: 2023, “Generative adversarial networks-based novel approach for fraud detection for the european cardholder’s 2013 dataset”.

9. Ileberi Emmanuel; Yanxia Sun; Zenghui Wang, Year: 2021, "Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost", IEEE Access, Vol: 9, pp. 165286-165294.

10. Nguyen Nghia; Truc Duong; Tram Chau; Van-Ho Nguyen; Trang Trinh; Duy Tran; Thanh Ho, Year: 2022, "A proposed model for card fraud detection based on Catboost and deep neural network", IEEE Access, Vol: 10, pp. 96852-96861.

11. Ning Wang; Siliang Chen; Songyi Lei; Xiongbin Liao, Year: 2023, "Amwspladaboost credit card fraud detection method based on enhanced base classifier diversity", IEEE Access.

12. Zhang Zhaohui; Ligong Chen; Qiuwen Liu; Pinged Wang, Year: 2020, "A fraud detection method for low-frequency transaction", IEEE Access, Vol: 8, pp. 25210-25220.

13. Almarshad Fahdah; Ghada Abdalaziz Gashgari; Abdullah IA Alzahrani, Year: 2023, "Generative Adversarial Networks-Based Novel Approach for Fraud Detection for the European Cardholders 2013 Dataset", IEEE Access zero.

14. Wang Haibo; Wendy Wang; Yi Liu; Bahram Alidaee, Year: 2022, "Integrating machine learning algorithms with quantum annealing solvers for online fraud detection", IEEE Access, Vol: 10, pp. 75908-75917.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 03(04), April 2024, pp. 41-53

Article Title: Implementation of A Deep Learning Algorithm for Fraud Detection Based on Credit Transactions

15. Esenogho Ebenezer; Ibomoiye Domor Mienye; Theo Swart; Kehinde Aruleba; George Obaido, Year: 2022, "A neural network ensemble with feature engineering for improved credit card fraud detection", IEEE Access, Vol: 10, pp. 16400-16407.