International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 12, Issue 6 [June. 2023] PP: 219-228

Online Fraud Detection Using Integrating Machine Learning Algorithms

P.Lakshmi Prabha*,Dr.R.Umamaheswari*1,Dr.K.Chandramohan*2

** PG Student, Department of Computer science Engineering, Gnanamani College of Technology, Namakkal, Tamilnadu, India*

**1 Professor, Department of Computer science Engineering, Gnanamani College of Technology, Namakkal, Tamilnadu, India*

**2Professor, Department of Computer science Engineering, Gnanamani College of Technology, Namakkal, Tamilnadu, India*

ABSTRACT

The rapid participation in online based transactional activities raises the fraudulent cases all over the world and causes tremendous losses to the individuals and financial industry. Credit card fraud events take place frequently and then result in huge financial losses. This project proposes a credit card fraud analysis using predictive modeling with machine learning technique by the data preparation carried out to manipulate the raw data into a form that can readily and accurately be analyzed. The overall work has been categorized as phase I and phase II separately. In phase I, the exploration stage, the initial patterns, characteristics, and points of interest are uncovered from the data. The accuracy and quality of source data before training a new model version is checked by data validation. Further, to correlate the data train and to get the model's accuracy Cat boost Classifier and CNN classifier is used. However, the performance of the cat boost classifier are better only when the data are properly tuned. Moreover, the CNN classifiers require lot of training data for the effective functioning. Also, they fail to encode the position and orientation of the objects. To overcome the above limitations, the model's accuracy XG boost Classifier and LSTM classifier is used in phase II. The XG boost classifier uses a direct route to minimum error, converging more quickly with fewer steps. The LSTM are much better at handling long-term dependencies. For analyzing the performance of the proposed work, it is implemented in python platform. The accuracy is validated are presented in both techniques.

***Key Words**: XGBoost, Machine Learning. Long short-term memory networks, Gradient boosting Machine.*

Date of Submission: 16-06-2023 Date of acceptance: 02-07-2023

I. Introduction

Financial fraud is a growing concern with far reaching consequences in the government, corporate organizations, finance industry, In Today's world high dependency on internet technology has enjoyed increased credit card transactions but credit card fraud had also accelerated as online and offline transaction. As credit card transactions become a widespread mode of payment, focus has been given to recent computational methodologies to handle the credit card fraud problem. There are many fraud detection solutions and software which prevent frauds in businesses such as credit card, retail, e-commerce, insurance, and industries. Data mining technique is one notable and popular methods used in solving credit fraud detection problem. It is impossible to be sheer certain about the true intention and rightfulness behind an application or transaction. To seek out possible evidences of fraud from the available data using mathematical algorithms is the best effective option. Fraud detection in credit card is the truly the process of identifying those transactions that are fraudulent into two classes: legit and fraud class transactions, several techniques are designed and implemented to solve to credit card fraud detection such as genetic algorithm, artificial neural network, frequent item set mining, machine learning algorithms, migrating birds optimization algorithm, can also perform comparative analysis of random forest, AdaBoost, XGBoost and Light GBM. Credit card fraud detection is a very popular but also a difficult problem to solve.

Firstly, due to issue of having only a limited amount of data, credit card makes it challenging to match a pattern for dataset. Secondly, there can be many entries in dataset with fraud transactions which fits the pattern of legitimate behaviour. Also, the problem has many constraints. Firstly, sensitive data sets are not easily accessible for public and the results of researches are often hidden and censored, making the results inaccessible and due to this it is sometimes challenging to benchmark certain models. Secondly, the improvement of methods is more difficult by the fact that the security concern imposes a limitation to exchange of ideas and methods in fraud detection, and especially in credit card fraud detection. Lastly, the data sets are continuously evolving and changing which makes the profiles of normal and fraudulent behaviors being different from the legit transaction in the present, which may have been fraud in past or vice versa.

In [1] Propose a large volume, wide range, frequency, as well as importance is stored from small to large organizations over the cloud. The whole information is available from massive amounts of sources such as followers on social media, customer order behaviours, likes, and shares.

In [2][3] proposes the application of machine learning techniques spreads widely throughout computer sciences domains such as spam filtering, web searching, ad placement, recommender systems, credit scoring, drug design, fraud detection, stock trading, and many other applications.

In [4] present the the various parameters that are used to construct the model to find the optimal combination of parameters to detect fraudulent activity. Deep Learning algorithms are a class of machine learning algorithms that use multiple non-linear processing units for feature extraction and transformation. These processing units discover intermediate representations in a hierarchical manner. The features discovered in one layer form the basis for processing of the succeeding layer.

Aims at addressing four fraud natures that belong to the CNP fraud category described above and propose a method to detect those frauds real time[5].

In [6]propose the prevailing data mining concerns people with credit card fraud detection model based on data mining. Since the problem is approached as a classification problem, classical data mining algorithms are not directly applicable.

In [7]Propose to improve fraud detection accuracy with growing number of transactions done by user per second. The increase in number of users and online transactions has brought heavy workloads to these systems.

In [8] proposes a case where a person uses someone else's credit card for personal reasons while the owner and the card issuing authorities are unaware of the fact that the card is being used.

In [9][10] proposes some of the key parts of the company's future business. Because of that, companies need to store that data, to process it and what is really important, to keep it safe. Without securing data, a lot of it can be used by other companies or even worse, it can be stolen.

Fraud has been increasing drastically with the progression of state-of-art technology and worldwide communication. Fraud can be avoided in two main ways: prevention and detection. Prevention avoids any attacks from fraudsters by acting as a layer of protection. Detection happens once the prevention has already failed. Therefore, detection helps in identifying and alerting as soon as a fraudulent transaction is being triggered. However, there has been an extreme increase in fraudulent transactions that affect the economy dramatically. Credit card fraud can be classified into several categories. The two types of frauds that can be mainly identified in a set of transactions are Card-not-present (CNP) frauds and Card-present (CP) frauds. Recently, card not-present transactions in credit card operations have become popular among web payment gateways. With the rapid development of economy globalization in recent decades, credit cards are much more popular in commercial transactions. The corresponding problem of the credit card fraud emerges accordingly. Machine learning approaches have been suggested to overcome these challenges.

II. Proposed block diagram

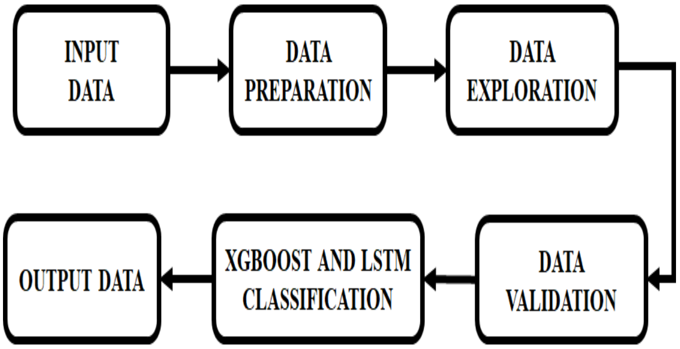


Figure.1 Proposed system Block Diagram

In this paper, online fraud detection using integrated machine learning algorithms is proposed. The data preparation stage begins with an analysis of the input image. The stage of data exploration follows data preparation and involves the exploration of data. Then the accuracy and quality of the source data are checked in the data validation stage. To overcome the existing work limitations, the XG boost classifier and LSTM classifier are used in the proposed work. The XG boost classifier uses a direct route to minimum error, converging more quickly with fewer steps. The LSTM is much better at handling long-term dependencies. Using the existing approach, efficient credit card fault analysis is performed with better accuracy. For analyzing the performance of the proposed work, it is implemented on the Python platform. The accuracy validated is presented in both techniques

OBJECTIVE

- To predict whether a credit card transaction is fraudulent or not.
- To efficiently classify the credit card frauds accurately using XGBoost and LSTM classifier.
- To prevent the losses caused by illegal acts.

III. Machine Learning for credit card fraud detection system

Fraud is as old as humanity itself and can take an unlimited variety of different forms. Moreover, the development of new technologies provides additional ways in which criminals may commit fraud, for instance in e-commerce the information about the card is sufficient to perpetrate a fraud. The use of credit cards is prevalent in modern day society and credit card fraud has kept on growing in recent years. Financial losses due to fraud affect not only merchants and banks (e.g. reimbursements), but also individual clients. If the bank loses money, customers eventually pay as well through higher interest rates, higher membership fees, etc. Fraud may also affect the reputation and image of a merchant causing non-financial losses that, though difficult to quantify in the short term, may become visible in the long period. For example, if a cardholder is victim of fraud with a certain company, he may no longer trust their business and choose a competitor. The actions taken against fraud can be divided into fraud prevention, which attempts to block fraudulent transactions at source, and fraud detection, where successful fraud transactions are identified a posteriori. Technologies that have been used in order to prevent fraud are Address Verification Systems (AVS), Card Verification Method (CVM) and Personal Identification Number (PIN). AVS involves verification of the address with zip code of the customer while CVM and PIN involve checking of the numeric code that is keyed in by the customer. For prevention purposes, financial institutions challenge all transactions with rule based filters and data mining methods as neural networks. Fraud detection is, given a set of credit card transactions, the process of identifying if a new authorized transaction belongs to the class of fraudulent or genuine transactions. A Fraud Detection System (FDS) should not only detect fraud cases efficiently, but also be cost-effective in the sense that the cost invested in transaction screening should not be higher than the loss due to frauds.

In order to minimize costs of detection it is important to use expert rules and statistical based models (e.g. Machine Learning) to make a first screen between genuine and potential fraud and ask the investigators to review only the cases with high risk. Typically, transactions are first filtered by checking some essential conditions (e.g. sufficient balance) and then scored by a predictive model (Figure.2). The predictive model scores each transaction with high or low risk of fraud and those with high risk generate alerts. Investigators check these alerts and provide a feedback for each alert, i.e. true positive (fraud) or false positive (genuine). These feedbacks can then be used to improve the model. A predictive model can be built upon experts' rules, i.e. rules based on knowledge from fraud experts, but these require manual tuning and human supervision. Alternatively, with Machine Learning (ML) techniques. Efficiently discover fraudulent patterns and predict transactions that are most likely to be fraudulent. ML techniques consist in inferring a prediction model on the basis of a set of examples. The model is in most cases a parametric function, which allows predicting the likelihood of a transaction to be fraud, given a set of features describing the transaction. In the domain of fraud detection, the use of learning techniques is attractive for a number of reasons.

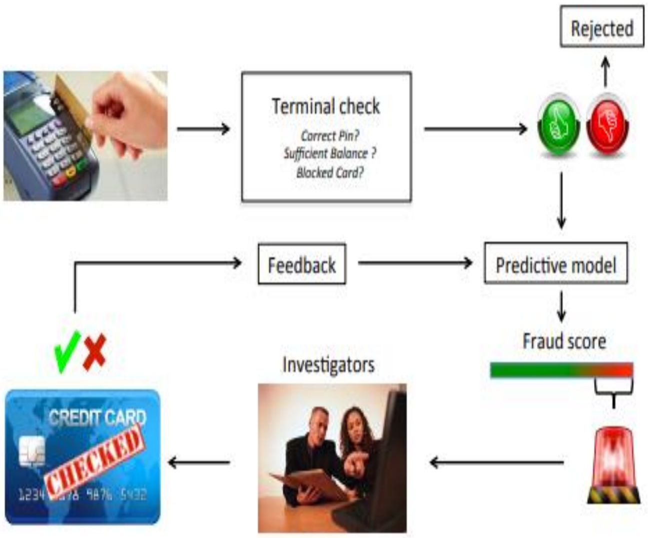
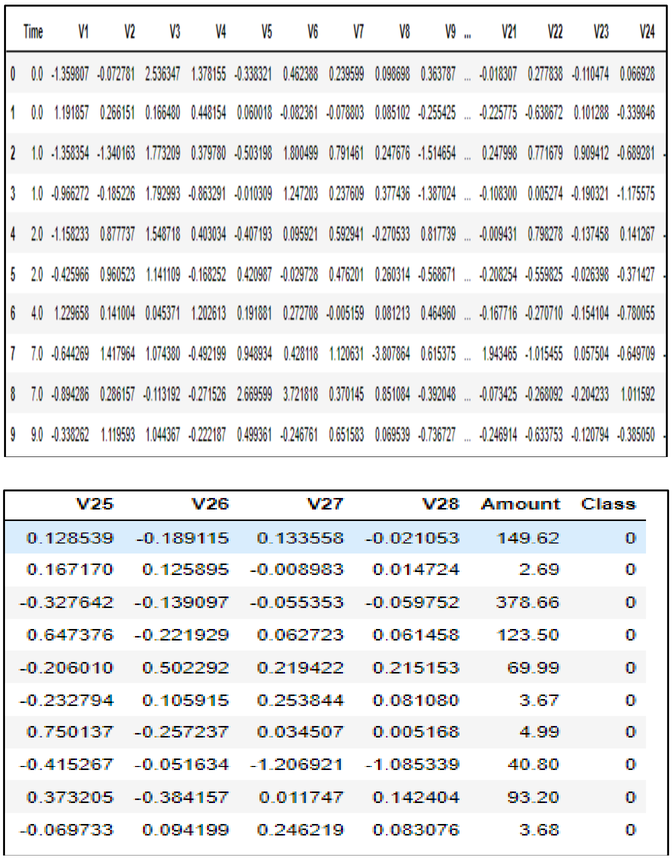


Figure.2 The Credit Card Fraud Detection process

IV. Results and Discussion:

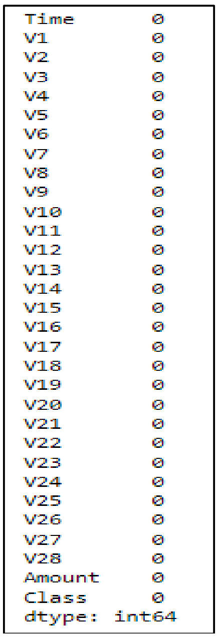
The input dataset of the online transactions reviews is represented in the Figure.3. The condition of the ttransactions, ratings and reviews of the credit card are given in the input dataset



	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V21	V22	V23	V24
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...	-0.018307	0.277838	-0.110474	0.066928
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...	-0.225775	-0.638672	0.101288	-0.339846
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...	0.247998	0.771679	0.909412	-0.689281
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	...	-0.108300	0.005274	-0.190321	-1.175575
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	...	-0.009431	0.798278	-0.137458	0.141267
5	2.0	-0.425966	0.960523	1.141109	-0.168252	0.420987	-0.029728	0.476201	0.260314	-0.568671	...	-0.208254	-0.559825	-0.026398	-0.371427
6	4.0	1.229658	0.141004	0.045371	1.202613	0.191881	0.272708	-0.005159	0.081213	0.464960	...	-0.167716	-0.270710	-0.154104	-0.780055
7	7.0	-0.644269	1.417964	1.074380	-0.492199	0.948934	0.428118	1.120631	-3.807864	0.615375	...	1.943465	-1.015455	0.057504	-0.649709
8	7.0	-0.894286	0.286157	-0.113192	-0.271526	2.669599	3.721818	0.370145	0.851084	-0.392048	...	-0.073425	-0.268092	-0.204233	1.011592
9	9.0	-0.338262	1.119593	1.044367	-0.222187	0.499361	-0.246761	0.651583	0.069539	-0.736727	...	-0.246914	-0.633753	-0.120794	-0.385050

V25	V26	V27	V28	Amount	Class
0.128539	-0.189115	0.133558	-0.021053	149.62	0
0.167170	0.125895	-0.008983	0.014724	2.69	0
-0.327642	-0.139097	-0.055353	-0.059752	378.66	0
0.647376	-0.221929	0.062723	0.061458	123.50	0
-0.206010	0.502292	0.219422	0.215153	69.99	0
-0.232794	0.105915	0.253844	0.081080	3.67	0
0.750137	-0.257237	0.034507	0.005168	4.99	0
-0.415267	-0.051634	-1.206921	-1.085339	40.80	0
0.373205	-0.384157	0.011747	0.142404	93.20	0
-0.069733	0.094199	0.246219	0.083076	3.68	0

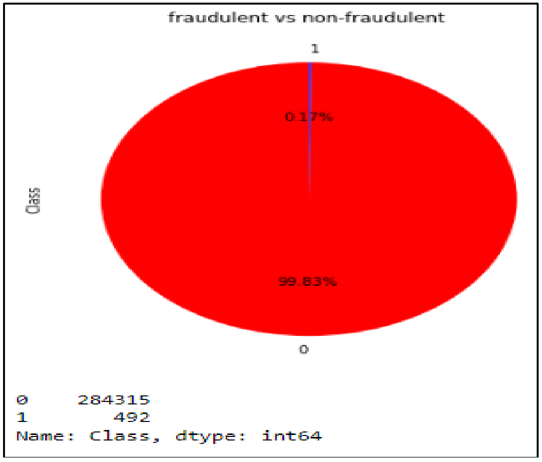
Figure.3 Input Dataset



```
Time      0
V1        0
V2        0
V3        0
V4        0
V5        0
V6        0
V7        0
V8        0
V9        0
V10       0
V11       0
V12       0
V13       0
V14       0
V15       0
V16       0
V17       0
V18       0
V19       0
V20       0
V21       0
V22       0
V23       0
V24       0
V25       0
V26       0
V27       0
V28       0
Amount    0
Class     0
dtype: int64
```

Figure.4 Null Values

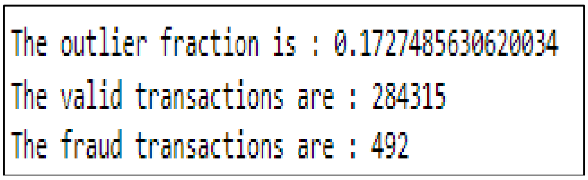
The above output shows Missing values of the reviews in the Figure.4.



```
0     284315
1        492
Name: Class, dtype: int64
```

Figure.5 Fraudulent Vs Non-Fraudulent

The Fraudulent Vs Non-Fraudulent plot in the above pie chart. Both classes 0 and 1 are the online transactions. Fraudulent transactions are 0.17% out of 100% then non-fraudulent transactions are 99.83% out of 100%. The 0 represents the fraudulent transactions and 1 represents the non-fraudulent transactions.



```
The outlier fraction is : 0.1727485630620034
The valid transactions are : 284315
The fraud transactions are : 492
```

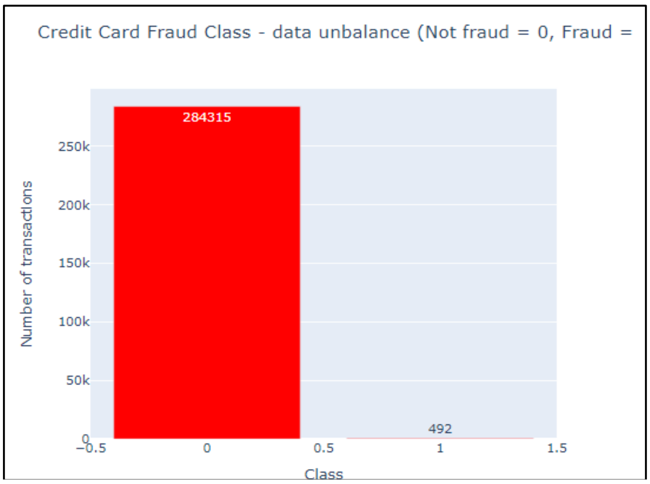


Figure.6 Class

The Bar plot representation of credit card fraud class per number of transactions is shown in Figure.6.
Fraudulent Transactions = 492 and represented in 1
Non-Fraudulent Transactions = 284315 and represented in 0

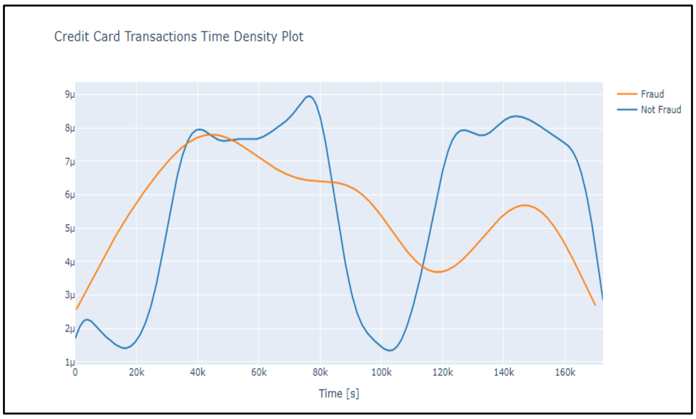


Figure.7 Credit Card Transactions Time Density Plot

Figure.7. represents fraudulent transactions have a distribution more even than valid transactions which are equally distributed in time, including the low real transaction times.

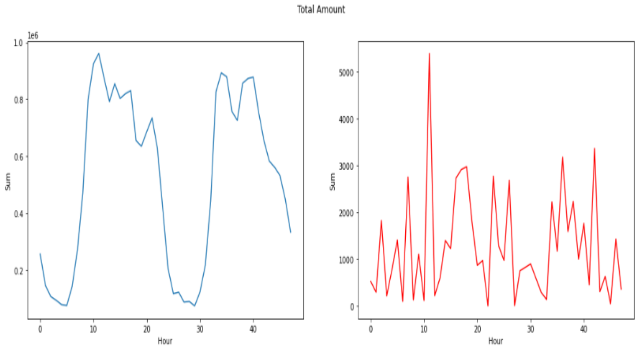


Figure.8 Total Amount of Transactions

Figure.8.represents the total amount of fraudulent transactions for varying time period at different intervals of hour. The total amount of fraudulent transactions shows steady increase and decreases as indicated in the output.

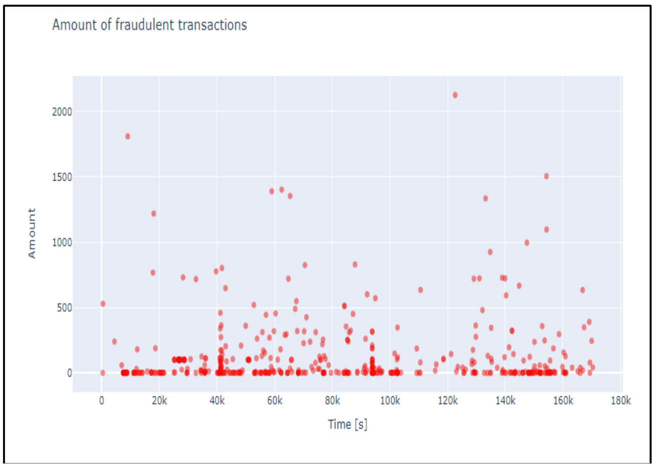
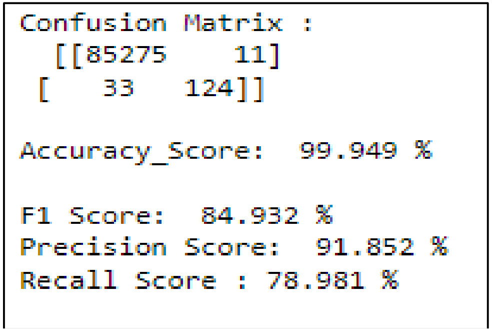


Figure.9 Amount of Fraudulent Transactions

Figure.9 shows the amount of fraudulent transactions at different intervals of time

XGBoost (eXtreme Gradient Boosting)



```
Confusion Matrix :
  [[85275    11]
 [   33   124]]

Accuracy_Score:  99.949 %

F1 Score:  84.932 %
Precision Score:  91.852 %
Recall Score : 78.981 %
```

Figure.10 Performance Matrix for XGBoost

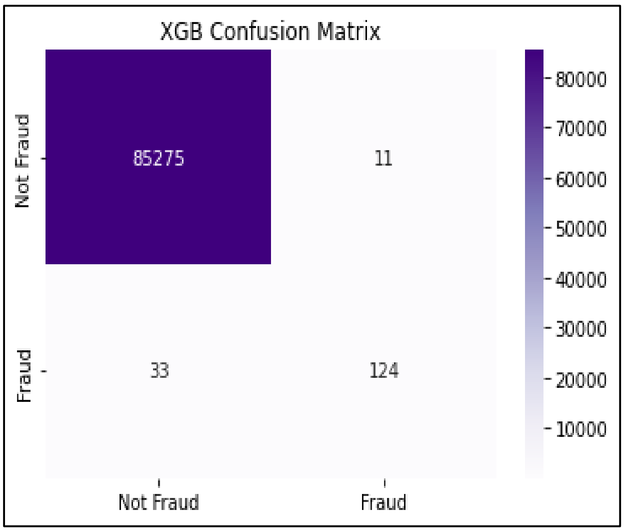


Figure.11 XG Boost

The using of confusion matrix represents the efficiency of the classifications process of different metrics that accounts for selectivity and specificity to minimize the time consumption. Confusion matrix for XG Boost is demonstrated in Figure.11.

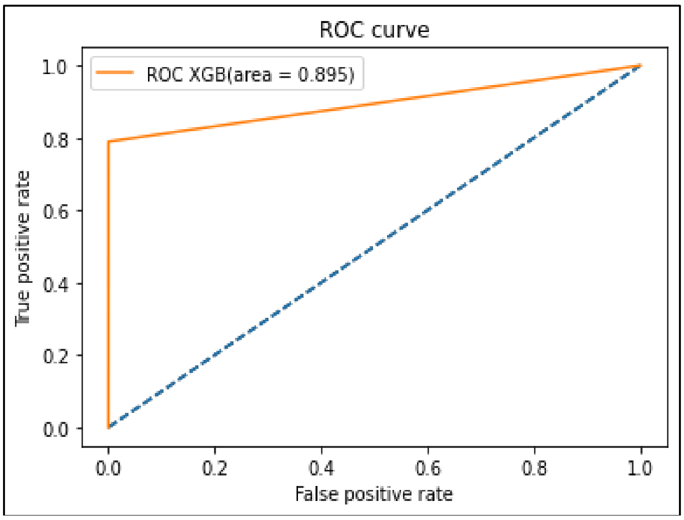


Figure.12 ROC Curve

The graph is plotted between true positive and false positive rating of the fraudulent and non-fraudulent transactions. The Roc curve has an area of 0.895.

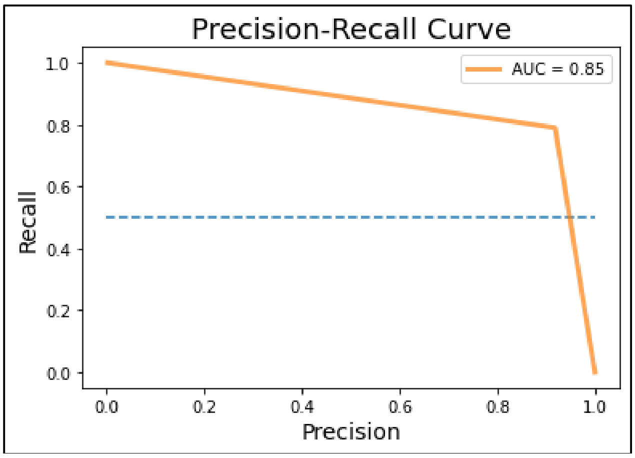
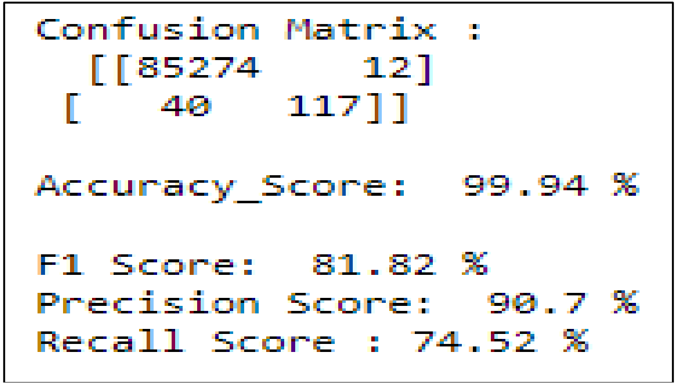


Figure.13 Precision-Recall Curve

The accuracy of the precision-recall curve is shown in the Figure.13.

Accuracy = 99.949
Precision = 91.852
Recall = 78.981

Long short-term memory (LSTM)



```
Confusion Matrix :
  [[85274     12]
 [    40    117]]

Accuracy_Score:   99.94 %

F1 Score:   81.82 %
Precision Score:   90.7 %
Recall Score : 74.52 %
```

Figure.14 Performance Matrix for LSTM

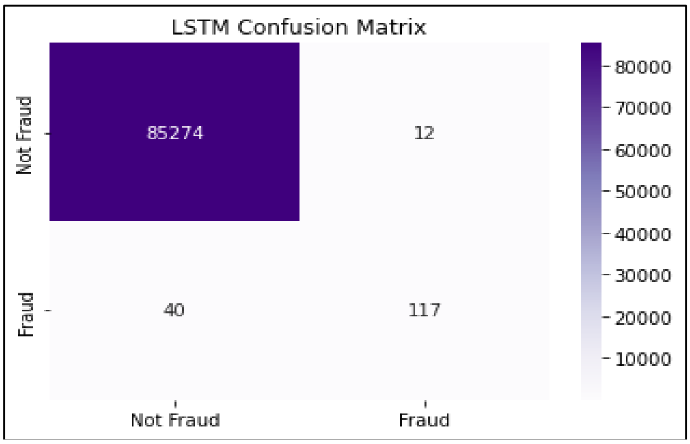


Figure.15 Confusion Matrix for LSTM

The confusion matrix is a systematic way to allocate the predictions to the original classes to which the data originally belonged. The confusion matrix for the LSTM multiclass classifier with attention mechanism is show in Figure.15.

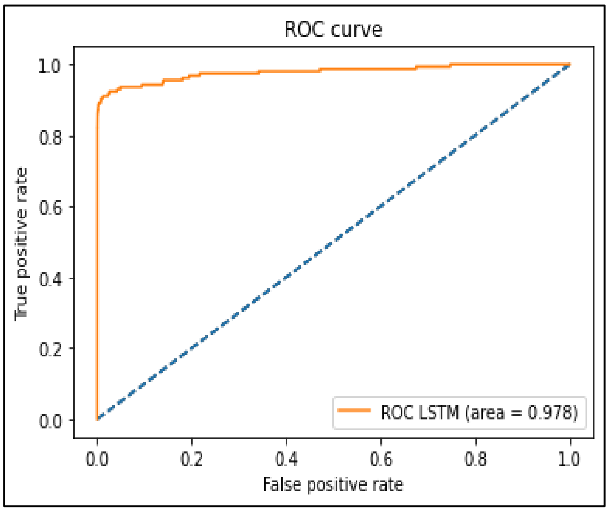


Figure.16 ROC Curve

The graph is plotted between true positive and false rating of LSTM classifier fraudulent and non-fraudulent transactions. The Roc curve has an area of 0.978.

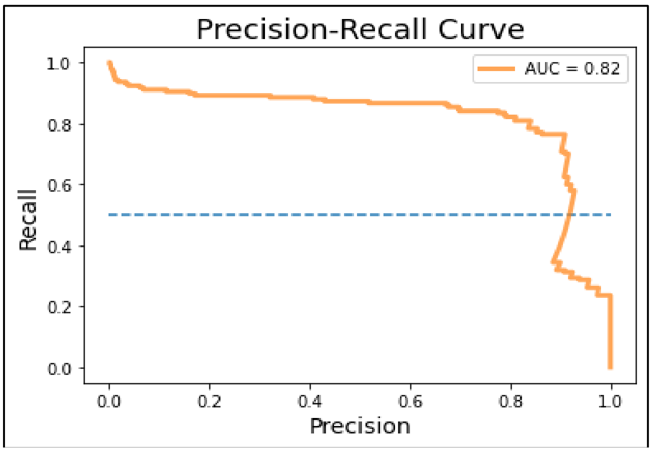


Figure.17 Precision-Recall Curve

The accuracy of the precision-recall curve is shown in the Figure.17.
Precision = 90.7
Recall = 74.52
Accuracy = 99.94

V. CONCLUSION

This paper is implemented online fraud detection using machine learning algorithms. The input image is analyzed in the data preparation stage. Following data preparation is the stage of data exploration, which includes data exploration. Next, during the data validation stage, the reliability and quality of the source data are examined. The XG boost classifier and LSTM classifier are employed in the proposed work to get beyond the existing work constraint. The XG boost classifier converges more quickly and with fewer steps by taking a direct path to the least error. The LSTM handles long-term dependence considerably better. The proposed approach enables more accurate and effective credit card fault analysis. It is implemented on the Python platform in order to evaluate how well the proposed work performs. Both methodologies provide accuracy that has been validated.

REFERENCE

[1]. 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.

[2]. 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.

[3]. A. Roy, 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.

[4]. A. Thennakoon, 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.

[5]. 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.

[6]. Patil, S., Nemade, V. and Soni, P.K., Predictive modelling for credit card fraud detection using data analytics. Procedia computer science, 132, pp.385-395, 2018.

[7]. Maniraj, S.P., Saini, A., Sarkar, S.D., Ahmed, S., Credit Card Fraud Detection using Machine Learning and Data Science. IJERT, 8(9), pp.110-115, 2019.

[8]. Varmedja, D., Karanovic, M., Sladojevic, S., Arsenovic, M., Anderla, A., Credit Card Fraud Detection - Machine Learning methods. 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH), pp. 1-5, 2019.

[9]. Xuan, S., Liu, G., Li, Z., Zheng, L., Wang, S., Jiang, C., Random forest for credit card fraud detection. 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), 2018, pp. 1-6, 2018.

[10]. Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.E., He-Guelton, L., Caelen,O., Sequence classification for credit-card fraud detection. Expert Systems with Applications (2018), 100(15), pp- 234-245, 2018.