International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-10, Issue-3, May 2022
https://doi.org/10.55524/ijircst.2022.10.3.59
Article ID IRPP1208, Pages 375-380
www.ijircst.org

Analyzing Various Machine Learning Algorithms for Blockchain-Based Fraud Detection

S. Giribabu[1], V. Sriharsha[2], Dr. Patan Hussain Basha[3], K. Suresh[4], and M. Sivudu[5]

[1,2,3,4,5]Assistant Professor, Department of Computer Science & Engineering, PACE Institute of Technology and Sciences, Ongole, Andhra Pradesh, India

ABSTRACT- A blockchain network's economics and user confidence can be seriously harmed by fraud. Consensus algorithms like proof of work and proof of stake can verify the legitimacy of a transaction but not the identity of the people who are conducting or verifying it. On a blockchain network, fraud can still occur, as a result of this. One approach to fighting fraud is to make use of machine learning techniques. There are two types of machine learning: supervised and unsupervised. We use a variety of supervised machine learning techniques in this study to distinguish between legitimate and fraudulent purchases. We also compare decision trees, Naive Bayes, logistic regression, multilayer perceptron, and other supervised machine learning techniques in detail for this challenge.

KEYWORDS- Blockchain, Machine, Learning, Machine Learning Algorithms, Fraud Detection.

I. INTRODUCTION

For a very long time, attempts to identify dishonest financial dealings have been studied. Fraudulent transactions discourage potential bitcoin investors and others from putting their faith in blockchain technology. They also hurt the economy. Most of the time, fraud is caused by something about the transaction's nature or the people involved. To ensure that the community and the network's integrity are not compromised, members of a blockchain network strive to quickly identify fraudulent transactions. A few AI approaches have been proposed to resolve this issue, and keeping in mind that a portion of the results look encouraging [4], nobody approach stands apart as obviously predominant. We compare and contrast the effectiveness of several supervised machine learning models, including SVM, Decision Tree, Naive Bayes, Logistic Regression, and a few deep learning models, for identifying fraudulent blockchain transactions in this study.A comparison of this kind will be helpful in deciding which method is best for balancing speed and accuracy.We are going to find out which transactions and users are most likely to commit fraud.

II. LITERATURE SURVEY

Cai, Y., Zhu, D. Fraud detections for online businesses: a perspectivefrom blockchain technology.

CoworkersYuanfeng Cai and discussed both objective and subjective forms of fraud.They end up figuring out that blockchain [1]is good at detecting objective fraud but not subjective fraud, so they use machine learning to fill in the gap.

Xu, J.J. Are blockchains immune to all malicious attacks

Jennifer J. Xu talked about the various types of fraud that blockchain can detect and those that it can't yet.[2] As a consequence of this, hypotheses regarding the elements that a component of machine learning needed to take into consideration began to take shape.She makes it abundantly clear that despite the fact that blockchain essentially adheres to a set of established rules, threats like data fraud and framework hacking are still attainable and difficult to stop[3].

Ostapowicz M., ˙ Zbikowski K. (2019) Detecting Fraudulent Accounts on Blockchain: A Supervised Approach

Micha Ostapowicz et al. used supervised machine learning techniques touncovered fraudulent behavior.[4] The ease with which criminals could steal money through widely used software or forged communications received the majority of their attention.They tested the effectiveness of various classifiers, such as Random Forests, Support Vector Machines, and XGBoost, in identifying these records using a dataset that contained more than 300,000 records.[5]

Podgorelec, B., Turkanovi´c, M. and Karakatiˇc, S., 2020. A Machine Learning-Based Method for Automated Blockchain Transaction Signing Including Personalized Anomaly Detection.

BlazPodgorelec and otherscreated an automated blockchain transaction verification system based on machine learning.As a result, it employs a specific method to identify suspicious financial transactions.[6]

Anomaly detection in bitcoin network using unsupervised learning methods

Thai T. [7] Pham and othersset out to discover a particular anomaly in the networks that handle bitcoin transactions. The Mahalanobis distance, K-means clustering, and unsupervised support vector machines were used to identify transactions and individuals that could be malicious.They made use of data that was divided into two graphs, one for users and the other for transactions.

A Review of Data Mining Methods for Identifying Blockchain

Li Ji et al. [8] anomalies provide a comprehensive look at how data mining anomalies can be detected using deep learning techniques. In addition, they provide comprehensive and specific explanations of the most typical detection strategies. The current methods' advantages and disadvantages, as well as anticipated future developments in this field, are discussed.

The Financial Impact of Money Laundering Supported by Cryptocurrencies

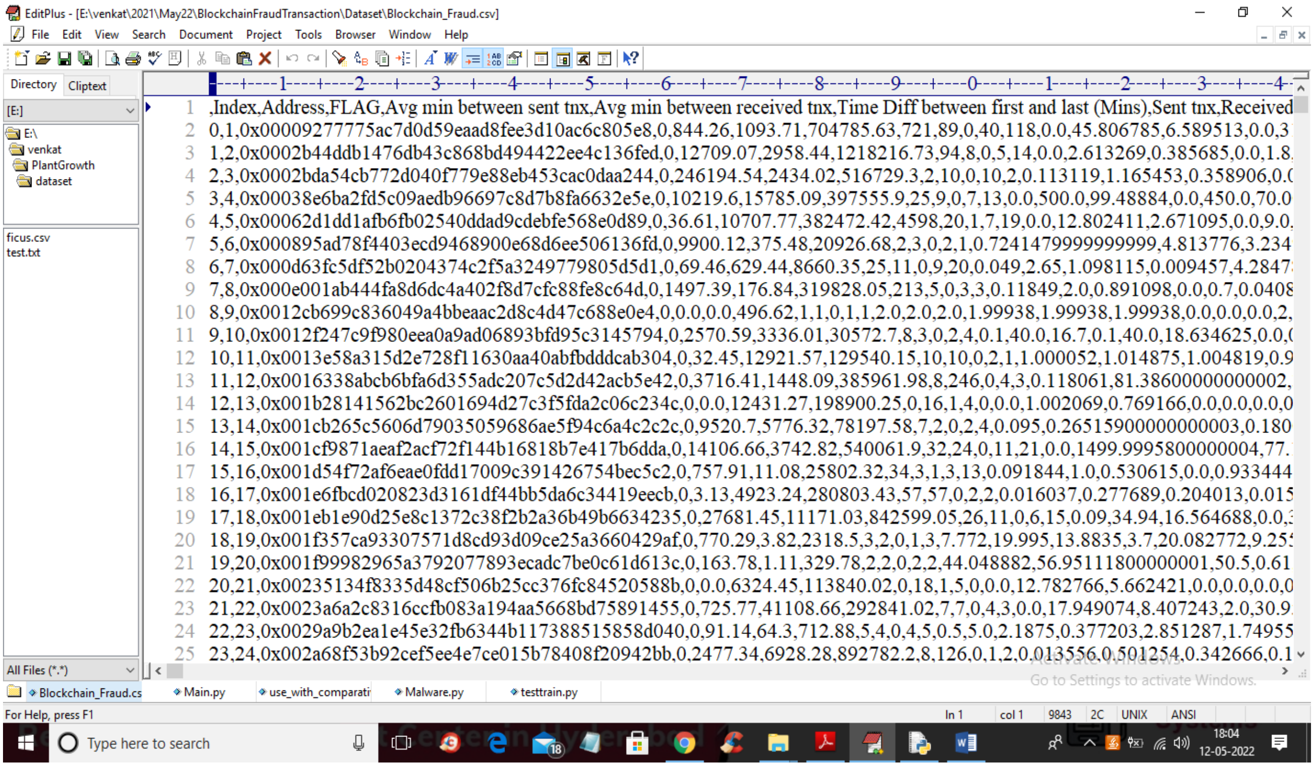
Christian Brenig et al.'[9]s economics of a cryptocurrency-based plan for money laundering. They talk about what money laundering is and how to avoid it. We are now even more motivated to investigate the prevalence of cryptocurrency and blockchain-related fraud as a result of this study[10].

III. PROPOSED SYSTEM

A user's participation in a Blockchain transaction does not guarantee that he will not commit fraud, which can have a negative impact on the economy of any country, despite the fact that Blockchain is thought to be secure against attacks due to its proof of work and transaction validation using hash code. To evaluate these tools' efficacy, the author of this work employs a wide range of machine learning techniques, including Logistic Regression, MLP, SVM, Decision Tree, and many others.

In order to write this article, all user and transaction information was gathered from the Blockchain fraud transaction dataset. The dataset was then processed to normalize values, replace missing values with 0, and remove all non-numerical data.

A. Dataset



```
,Index,Address,FLAG,Avg min between sent tnx,Avg min between received tnx,Time Diff between first and last (Mins),Sent tnx,Received
0,1,0x00009277775ac7d0d59eaad8fee3d10ac6c805e8,0,844.26,1093.71,704785.63,721,89,0,40,118,0.0,45.806785,6.589513,0.0,3
1,2,0x0002b44ddb1476db43c868bd494422ee4c136fed,0,12709.07,2958.44,1218216.73,94,8,0,5,14,0.0,2.613269,0.385685,0.0,1.8
2,3,0x0002bda54cb772d040f779e88eb453cac0daa244,0,246194.54,2434.02,516729.3,2,10,0,10,2,0.113119,1.165453,0.358906,0.0
3,4,0x00038e6ba2fd5c09aedb96697c8d7b8fa6632e5e,0,10219.6,15785.09,397555.9,25,9,0,7,13,0.0,500.0,99.48884,0.0,450.0,70.0
4,5,0x00062d1dd1afb6fb02540ddad9cdebfe568e0d89,0,36.61,10707.77,382472.42,4598,20,1,7,19,0.0,12.802411,2.671095,0.0,9.0,
5,6,0x000895ad78f4403ecd9468900e68d6ee506136fd,0,9900.12,375.48,20926.68,2,3,0,2,1,0.7241479999999999,4.813776,3.234
6,7,0x000d63fc5df52b0204374c2f5a3249779805d5d1,0,69.46,629.44,8660.35,25,11,0,9,20,0.049,2.65,1.098115,0.009457,4.2847
7,8,0x000e001ab444fa8d6dc4a402f8d7cfc88fe8c64d,0,1497.39,176.84,319828.05,213,5,0,3,3,0.11849,2.0,0.891098,0.0,0.7,0.0408
8,9,0x0012cb699c836049a4bbeaac2d8c4d47c688e0e4,0,0.0,0.0,496.62,1,1,0,1,1,2.0,2.0,2.0,1.99938,1.99938,1.99938,0.0,0.0,0.0,2,
9,10,0x0012f247c9f980eea0a9ad06893bfd95c3145794,0,2570.59,3336.01,30572.7,8,3,0,2,4,0.1,40.0,16.7,0.1,40.0,18.634625,0.0,0
10,11,0x0013e58a315d2e728f11630aa40abfbdddcab304,0,32.45,12921.57,129540.15,10,10,0,2,1,1.000052,1.014875,1.004819,0.9
11,12,0x0016338abcb6bfa6d355adc207c5d2d42acb5e42,0,3716.41,1448.09,385961.98,8,246,0,4,3,0.118061,81.38600000000002,
12,13,0x001b28141562bc2601694d27c3f5fda2c06c234c,0,0.0,12431.27,198900.25,0,16,1,4,0,0.0,1.002069,0.769166,0.0,0.0,0.0,0
13,14,0x001cb265c5606d79035059686ae5f94c6a4c2c2c,0,9520.7,5776.32,78197.58,7,2,0,2,4,0.095,0.26515900000000003,0.180
14,15,0x001cf9871aeaf2acf72f144b16818b7e417b6dda,0,14106.66,3742.82,540061.9,32,24,0,11,21,0.0,1499.9995800000004,77.
15,16,0x001d54f72af6eae0fdd17009c391426754bec5c2,0,757.91,11.08,25802.32,34,3,1,3,13,0.091844,1.0,0.530615,0.0,0.933444
16,17,0x001e6fbcd020823d3161df44bb5da6c34419eecb,0,3.13,4923.24,280803.43,57,57,0,2,2,0.016037,0.277689,0.204013,0.015
17,18,0x001eb1e90d25e8c1372c38f2b2a36b49b6634235,0,27681.45,11171.03,842599.05,26,11,0,6,15,0.09,34.94,16.564688,0.0,
18,19,0x001f357ca93307571d8cd93d09ce25a3660429af,0,770.29,3.82,2318.5,3,2,0,1,3,7.772,19.995,13.8835,3.7,20.082772,9.25
19,20,0x001f99982965a3792077893ecadc7be0c61d613c,0,163.78,1.11,329.78,2,2,0,2,2,44.048882,56.95111800000001,50.5,0.61
20,21,0x00235134f8335d48cf506b25cc376fc84520588b,0,0.0,6324.45,113840.02,0,18,1,5,0,0.0,12.782766,5.662421,0.0,0.0,0.0,0
21,22,0x0023a6a2c8316ccfb083a194aa5668bd75891455,0,725.77,41108.66,292841.02,7,7,0,4,3,0.0,17.949074,8.407243,2.0,30.9
22,23,0x0029a9b2ea1e45e32fb6344b117388515858d040,0,91.14,64.3,712.88,5,4,0,4,5,0.5,5.0,2.1875,0.377203,2.851287,1.74955
23,24,0x002a68f53b92cef5ee4e7ce015b78408f20942bb,0,2477.34,6928.28,892782.2,8,126,0,1,2,0.013556,0.501254,0.342666,0.1
```

Figure 1: Names of the dataset columns

In figure 1, the names of the dataset columns can be found in the first row of the screen, and the values of the dataset can be found in the rows that follow. In the dataset, there is a column called FLAG, and its values range from 0 to 1, with 0 representing a normal transaction and 1 representing a fraudulent one.

IV. METHODOLOGY

Figure 1 provides a summary of the fraudulent activity detection workflow. Our proposed system basically begins performing additional checks to determine whether a transaction may be fraudulent after the Blockchain network has approved the transaction after completing all of the basic checks. The Blockchain network itself has the ability to easily invalidate transactions, so this strategy ensures that there is no additional overhead associated with even checking them.

There are three main phases to the work that was done:

- Pre-processing phase
- Building and training various models
- Performance evaluation of all the models

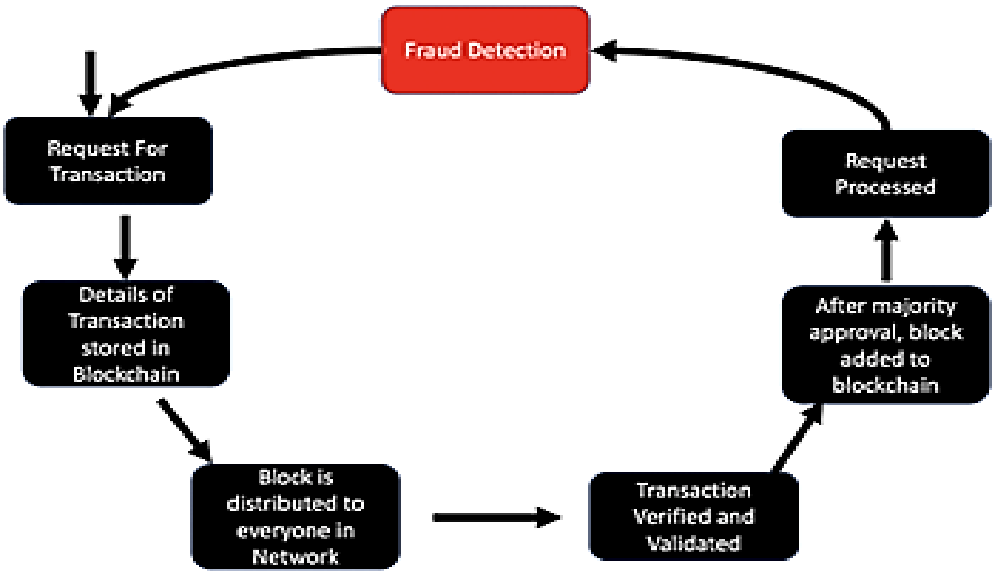
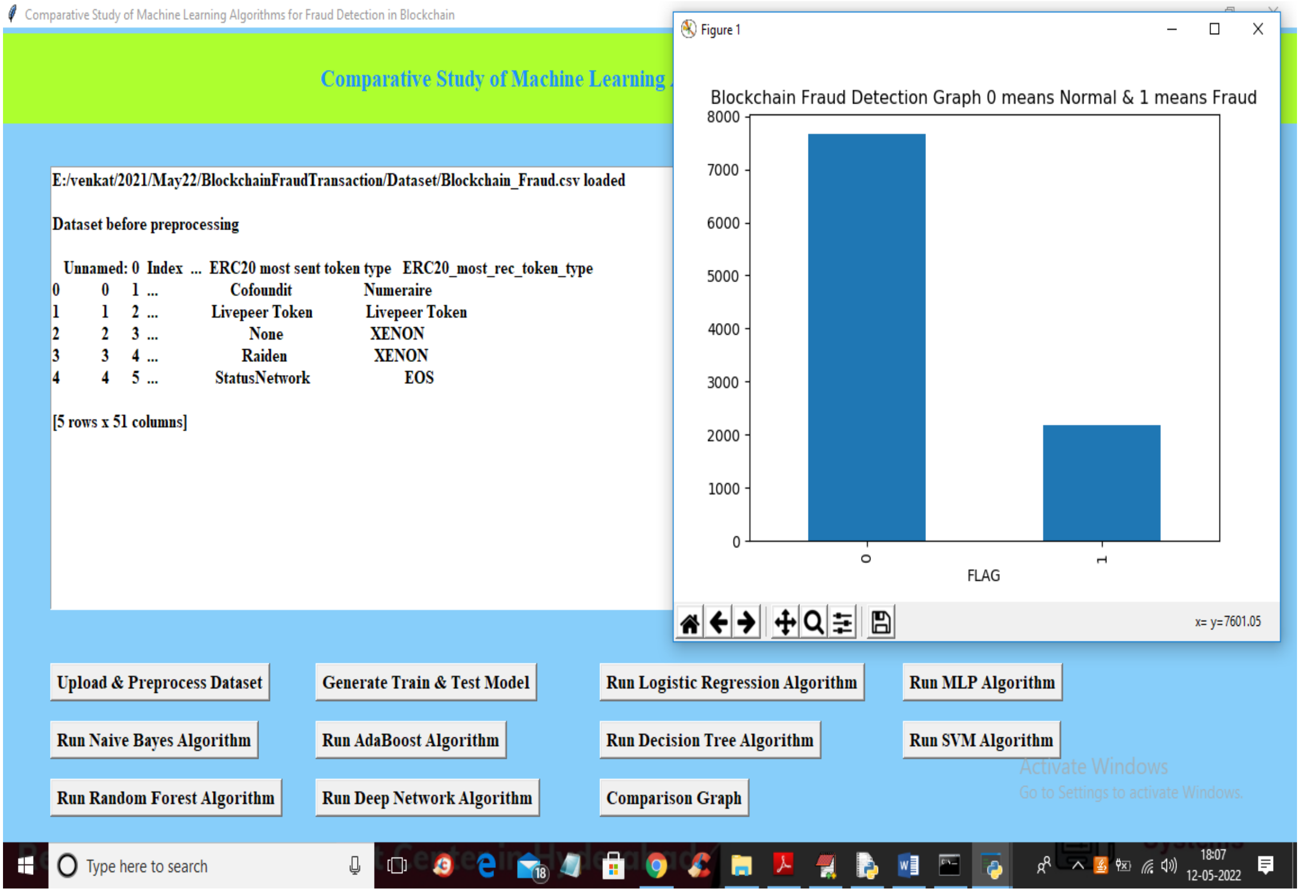

Fraud Detection
Request For Transaction
Details of Transaction stored in Blockchain
Block is distributed to everyone in Network
Transaction Verified and Validated
After majority approval, block added to blockchain
Request Processed

Figure 2: Workflow of applying check for fraud detection

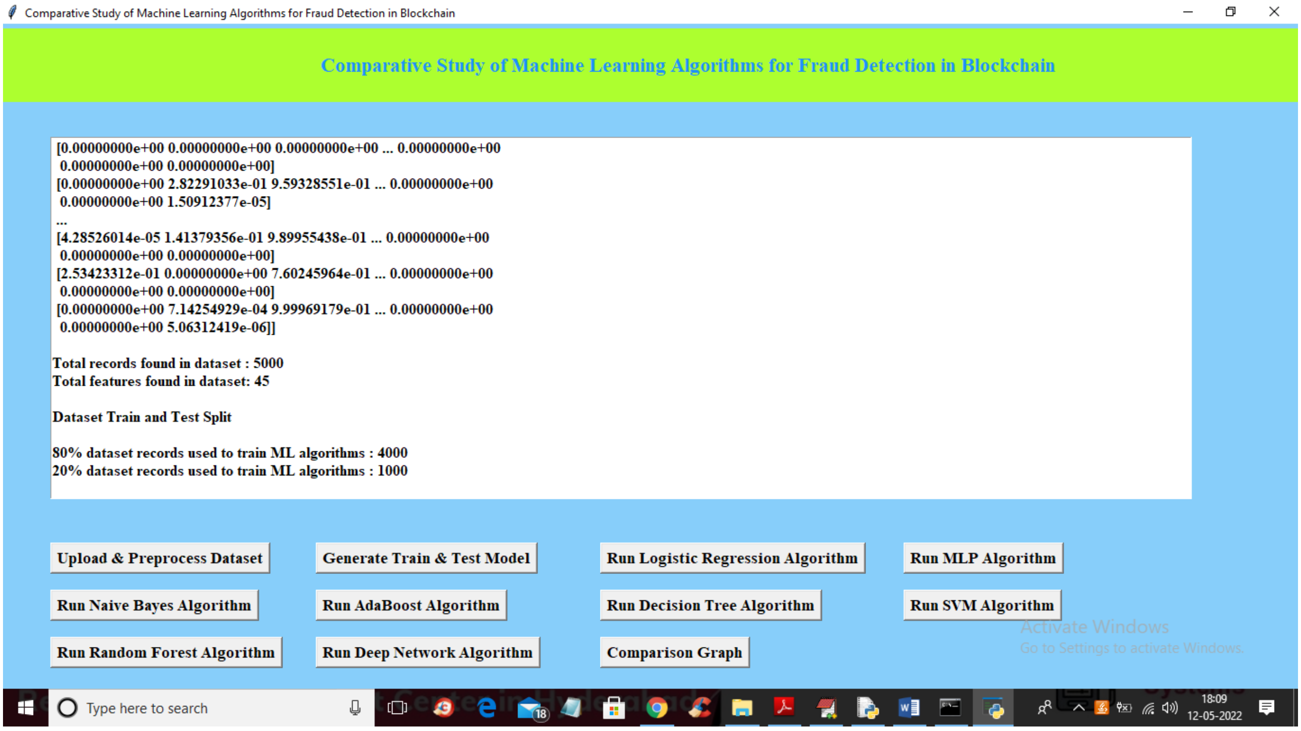
V. RESULTS AND DISCUSSION


Comparative Study of Machine Learning Algorithms for Fraud Detection in Blockchain
Comparative Study of Machine Learning
Figure 1
Blockchain Fraud Detection Graph 0 means Normal & 1 means Fraud
8000
7000
6000
5000
4000
3000
2000
1000
0
0
1
FLAG
x= y=7601.05
Upload & Preprocess Dataset
Generate Train & Test Model
Run Logistic Regression Algorithm
Run MLP Algorithm
Run Naive Bayes Algorithm
Run AdaBoost Algorithm
Run Decision Tree Algorithm
Run SVM Algorithm
Run Random Forest Algorithm
Run Deep Network Algorithm
Comparison Graph
Activate Windows
Go to Settings to activate Windows.
Type here to search
18:07
12-05-2022

```
E:/venkat/2021/May22/BlockchainFraudTransaction/Dataset/Blockchain_Fraud.csv loaded

Dataset before preprocessing

   Unnamed: 0  Index  ...  ERC20 most sent token type  ERC20_most_rec_token_type
0           0      1  ...                   Cofoundit                  Numeraire
1           1      2  ...              Livepeer Token             Livepeer Token
2           2      3  ...                        None                      XENON
3           3      4  ...                      Raiden                      XENON
4           4      5  ...               StatusNetwork                        EOS

[5 rows x 51 columns]
```

Figure 3: Loaded dataset

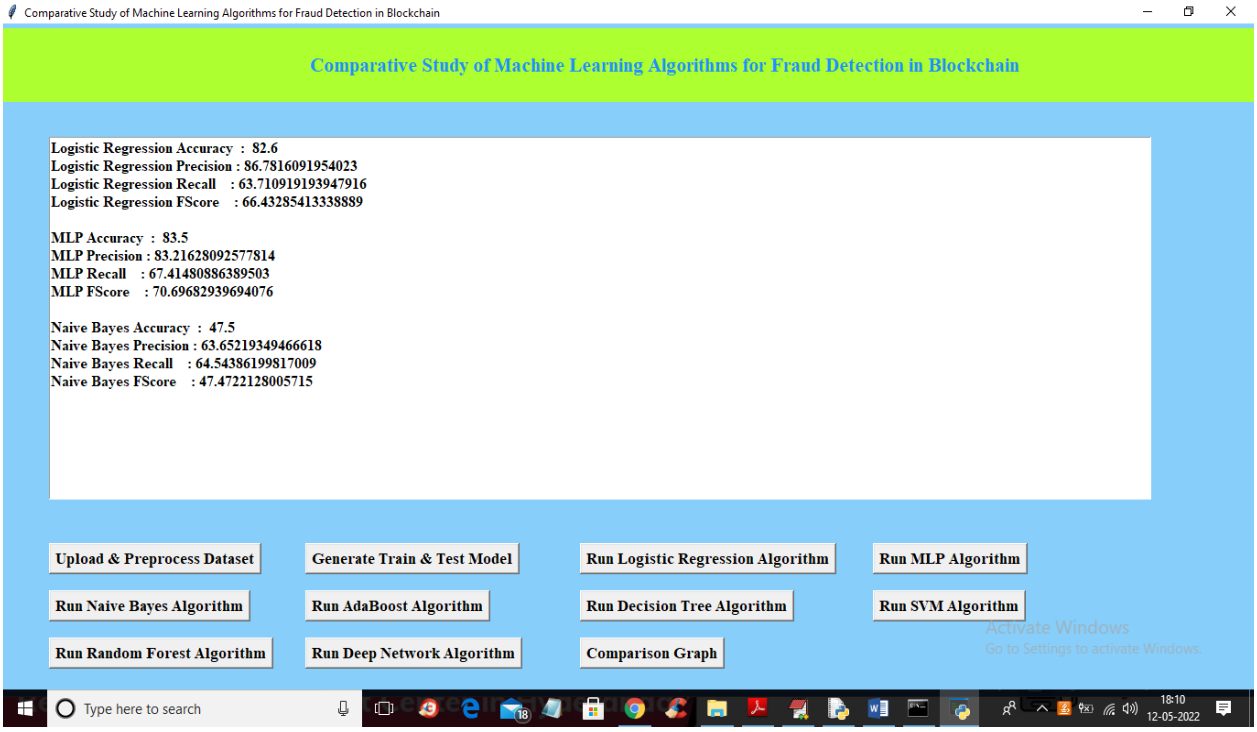
In the above figure 3, dataset loaded and dataset contains some non-numeric data and ML algorithms will not take such data so we need to remove and graph x-axis contains type of transaction and y-axis contains number of records and now close above graph and then click on 'Generate Train & Test Model' button to get below output.


Comparative Study of Machine Learning Algorithms for Fraud Detection in Blockchain
Comparative Study of Machine Learning Algorithms for Fraud Detection in Blockchain
Upload & Preprocess Dataset
Generate Train & Test Model
Run Logistic Regression Algorithm
Run MLP Algorithm
Run Naive Bayes Algorithm
Run AdaBoost Algorithm
Run Decision Tree Algorithm
Run SVM Algorithm
Run Random Forest Algorithm
Run Deep Network Algorithm
Comparison Graph
Type here to search
18:09
12-05-2022

```
[0.00000000e+00 0.00000000e+00 0.00000000e+00 ... 0.00000000e+00
 0.00000000e+00 0.00000000e+00]
[0.00000000e+00 2.82291033e-01 9.59328551e-01 ... 0.00000000e+00
 0.00000000e+00 1.50912377e-05]
...
[4.28526014e-05 1.41379356e-01 9.89955438e-01 ... 0.00000000e+00
 0.00000000e+00 0.00000000e+00]
[2.53423312e-01 0.00000000e+00 7.60245964e-01 ... 0.00000000e+00
 0.00000000e+00 0.00000000e+00]
[0.00000000e+00 7.14254929e-04 9.99969179e-01 ... 0.00000000e+00
 0.00000000e+00 5.06312419e-06]]

Total records found in dataset : 5000
Total features found in dataset: 45

Dataset Train and Test Split

80% dataset records used to train ML algorithms : 4000
20% dataset records used to train ML algorithms : 1000
```

Figure 4: Converted data

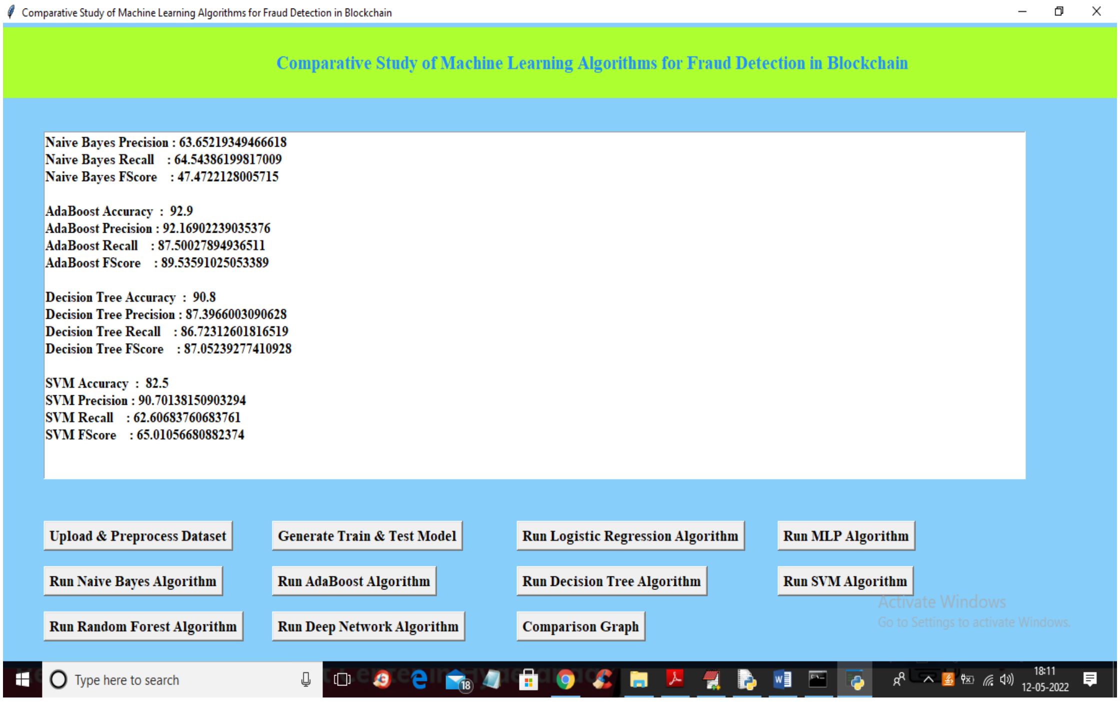
In figure 4, all of the data has been converted to a numerical format, and the total number of records and columns in the dataset can be seen on the above screen. The dataset was then divided into train and test, and the train and test data are now ready. Click on each button to run all algorithms, and the results are as shown below.


Comparative Study of Machine Learning Algorithms for Fraud Detection in Blockchain
Comparative Study of Machine Learning Algorithms for Fraud Detection in Blockchain
Upload & Preprocess Dataset
Generate Train & Test Model
Run Logistic Regression Algorithm
Run MLP Algorithm
Run Naive Bayes Algorithm
Run AdaBoost Algorithm
Run Decision Tree Algorithm
Run SVM Algorithm
Run Random Forest Algorithm
Run Deep Network Algorithm
Comparison Graph
Type here to search
18:10
12-05-2022

```
Logistic Regression Accuracy  :  82.6
Logistic Regression Precision : 86.7816091954023
Logistic Regression Recall    : 63.710919193947916
Logistic Regression FScore    : 66.43285413338889

MLP Accuracy  :  83.5
MLP Precision : 83.21628092577814
MLP Recall    : 67.41480886389503
MLP FScore    : 70.69682939694076

Naive Bayes Accuracy  :  47.5
Naive Bayes Precision : 63.65219349466618
Naive Bayes Recall    : 64.54386199817009
Naive Bayes FScore    : 47.4722128005715
```

Figure 5: Performance or accuracy of algorithm

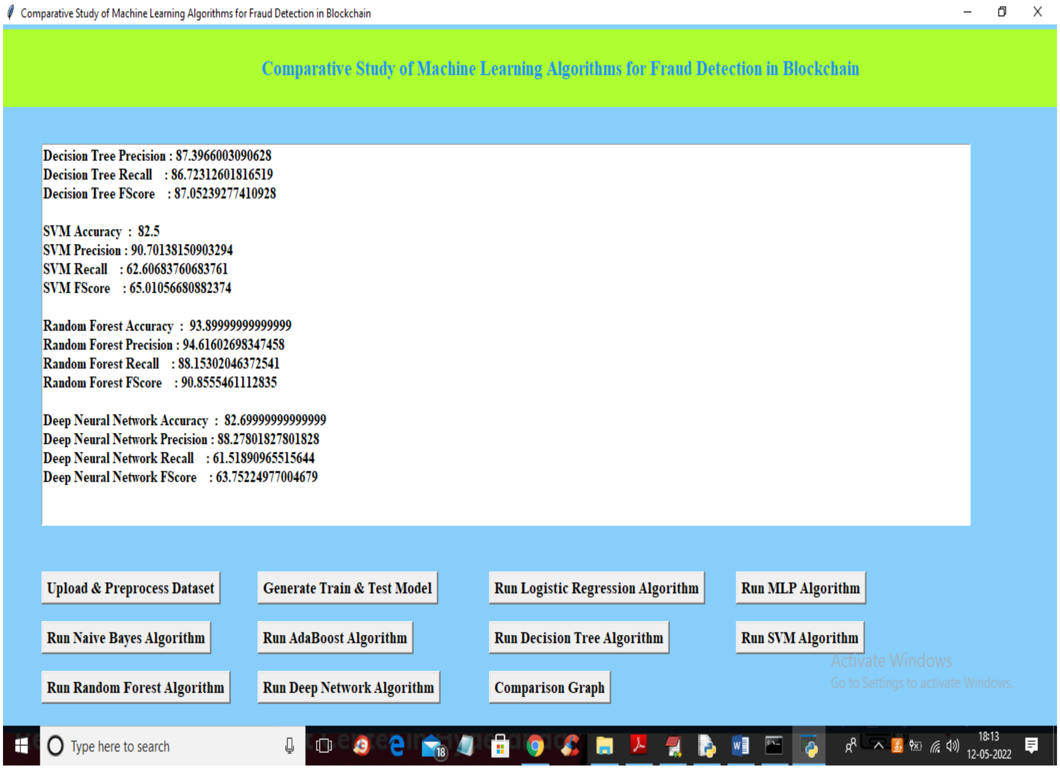
In figure 5, the performance or accuracy of each algorithm can be seen on the screen above, and the remaining algorithm accuracy can be seen below.


Comparative Study of Machine Learning Algorithms for Fraud Detection in Blockchain
Comparative Study of Machine Learning Algorithms for Fraud Detection in Blockchain
Upload & Preprocess Dataset
Generate Train & Test Model
Run Logistic Regression Algorithm
Run MLP Algorithm
Run Naive Bayes Algorithm
Run AdaBoost Algorithm
Run Decision Tree Algorithm
Run SVM Algorithm
Run Random Forest Algorithm
Run Deep Network Algorithm
Comparison Graph
Activate Windows
Go to Settings to activate Windows.
Type here to search
18:11
12-05-2022

```
Naive Bayes Precision : 63.65219349466618
Naive Bayes Recall    : 64.54386199817009
Naive Bayes FScore    : 47.4722128005715

AdaBoost Accuracy  :  92.9
AdaBoost Precision : 92.16902239035376
AdaBoost Recall    : 87.50027894936511
AdaBoost FScore    : 89.53591025053389

Decision Tree Accuracy  :  90.8
Decision Tree Precision : 87.3966003090628
Decision Tree Recall    : 86.72312601816519
Decision Tree FScore    : 87.05239277410928

SVM Accuracy  :  82.5
SVM Precision : 90.70138150903294
SVM Recall    : 62.60683760683761
SVM FScore    : 65.01056680882374
```

Figure 6: Accuracy of AdaBoost, Decision Tree, and SVM

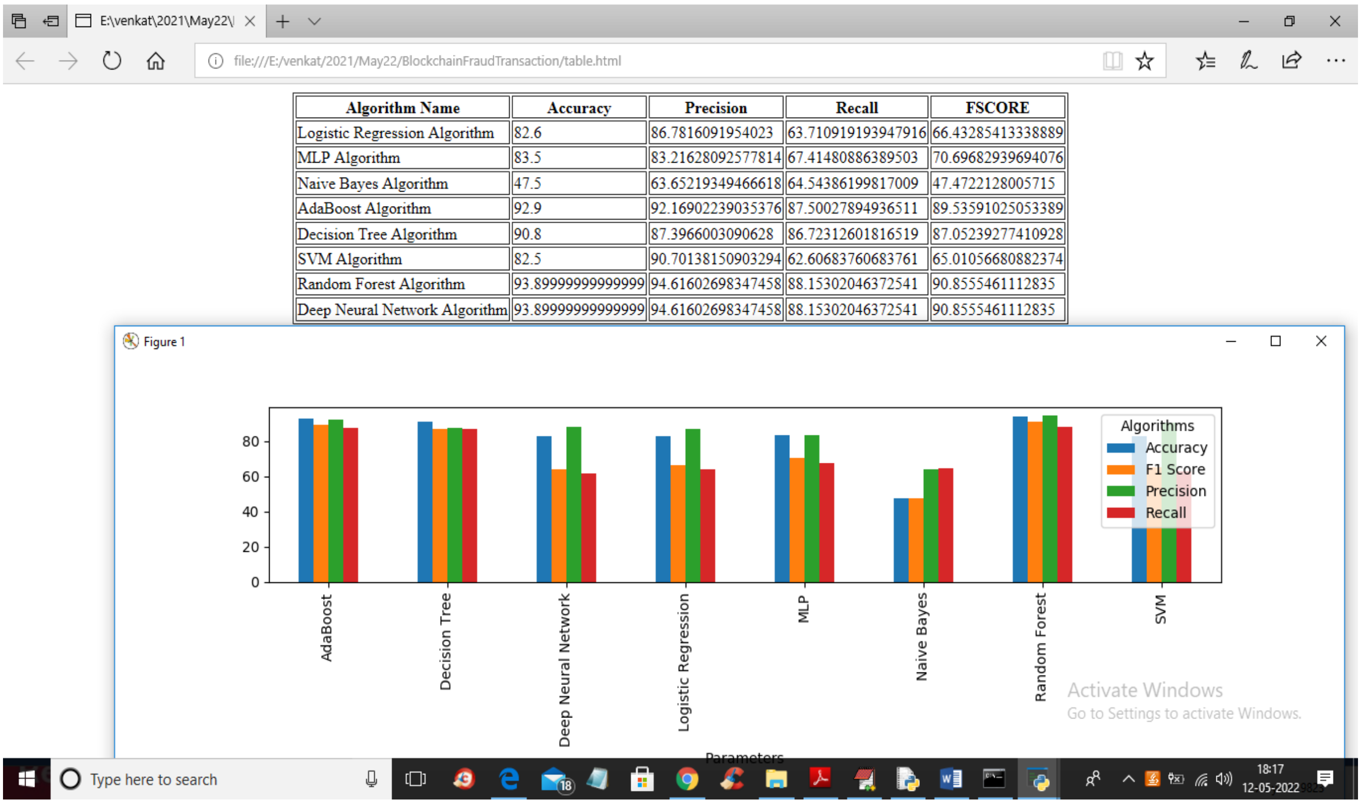
The accuracy of AdaBoost, Decision Tree, and SVM can be seen on the screen above figure 6, while the accuracy of the remaining algorithms can be seen in below figure 7.


Comparative Study of Machine Learning Algorithms for Fraud Detection in Blockchain
Comparative Study of Machine Learning Algorithms for Fraud Detection in Blockchain
Upload & Preprocess Dataset
Generate Train & Test Model
Run Logistic Regression Algorithm
Run MLP Algorithm
Run Naive Bayes Algorithm
Run AdaBoost Algorithm
Run Decision Tree Algorithm
Run SVM Algorithm
Run Random Forest Algorithm
Run Deep Network Algorithm
Comparison Graph
Activate Windows
Go to Settings to activate Windows.
Type here to search
18:13
12-05-2022

```
Decision Tree Precision : 87.3966003090628
Decision Tree Recall    : 86.72312601816519
Decision Tree FScore    : 87.05239277410928

SVM Accuracy  :  82.5
SVM Precision : 90.70138150903294
SVM Recall    : 62.60683760683761
SVM FScore    : 65.01056680882374

Random Forest Accuracy  :  93.89999999999999
Random Forest Precision : 94.61602698347458
Random Forest Recall    : 88.15302046372541
Random Forest FScore    : 90.8555461112835

Deep Neural Network Accuracy  :  82.69999999999999
Deep Neural Network Precision : 88.27801827801828
Deep Neural Network Recall    : 61.51890965515644
Deep Neural Network FScore    : 63.75224977004679
```

Figure 7: Accuracy of Random Forest and Deep Neural Networks

The accuracy of Random Forest and Deep Neural Networks can be seen on the above figure 7, with Random Forest outperforming all other algorithms. Now, select the "Comparison Graph" button to view the output in below figure 8.



Algorithm Name	Accuracy	Precision	Recall	FSCORE
Logistic Regression Algorithm	82.6	86.7816091954023	63.710919193947916	66.43285413338889
MLP Algorithm	83.5	83.21628092577814	67.41480886389503	70.69682939694076
Naive Bayes Algorithm	47.5	63.65219349466618	64.54386199817009	47.4722128005715
AdaBoost Algorithm	92.9	92.16902239035376	87.50027894936511	89.53591025053389
Decision Tree Algorithm	90.8	87.3966003090628	86.72312601816519	87.05239277410928
SVM Algorithm	82.5	90.70138150903294	62.60683760683761	65.01056680882374
Random Forest Algorithm	93.89999999999999	94.61602698347458	88.15302046372541	90.8555461112835
Deep Neural Network Algorithm	93.89999999999999	94.61602698347458	88.15302046372541	90.8555461112835

Figure 8: Accuracy of precision, recall, and FSCORE

The accuracy, precision, recall, and FSCORE of each algorithm and tabular form are displayed in above figure 8, with Random Forest providing superior results for all of them.

VI. CONCLUSION

A method for identifying suspicious blockchain transactions has been presented using machine learning.This method looked at support vector machines, decision trees, logistic regression, dense neural networks, and other supervised learning methods.A comprehensive evaluation and comparison of all available methods is carried out using accuracy.This work could be expanded to include a comparison of clustering and other unsupervised algorithms.In the near future, we hope to conduct a comprehensive investigation of fraudulent activity in a private blockchain.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] Cai, Y., Zhu, D. Fraud detections for online businesses: a perspectivefrom blockchain technology.

[2] Xu, J.J. Are blockchains immune to all malicious attacks

[3] Ostapowicz M., ˙ Zbikowski K. (2019) Detecting Fraudulent Accounts on Blockchain: A Supervised Approach

[4] Podgorelec, B., Turkanovi´c, M. and Karakatiˇc, S., 2020. A Machine Learning-Based Method for Automated Blockchain Transaction Signing Including Personalized Anomaly Detection.

[5] Farrugia S, Ellul J, Azzopardi G. Detection of illicit accounts over the Ethereum blockchain. Expert Systems with Applications

[6] Pham, Thai, and Steven Lee. ”Anomaly detection in bitcoin network using unsupervised learning methods

[7] Li, Ji, et al. ”A Survey on Blockchain Anomaly Detection UsingData Mining Techniques.

[8] Brenig, Christian, and G¨unterM¨uller. ”Economic analysis of cryptocurrency backed money laundering.” (2015)

[9] Monamo, Patrick, VukosiMarivate, and Bheki Twala. ”Unsupervised learning for robust Bitcoin fraud detection.” 2016 Information Security for South Africa (ISSA). IEEE, 2016.

[10] Anomaly Detection in Bitcoin Network Using Unsupervised Learning Methods

[11] Data mining for credit card fraud: A comparative study S. Bhattacharyya, Sanjeev Jha, K. Tharakunnel, J. Westland

[12] The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature E. Ngai, Yong Hu, Y. Wong, Yijun Chen, Xin Sun

[13] Graph based anomaly detection and description: a survey L. Akoglu, Hanghang Tong, Danai Koutra.