CREDIT CARD FRAUD DETECTION USING XGBOOST ALGORITHM

Anumandla Nirosha,H.No: 21S41D5801,Mtech (Cse), Department Of Computer Science, Vaageswari College Of Engineering Thimmapur,Karimnagar, Telangana, India, Email-Id: anunirosha364@gmail.com.

Dr. Gulab Singh , Associate Professor, Department Of Computer Science, Vaageswari College Of Engineering Thimmapur,Karimnagar, Telangana, India,Email-Id: gulsinchu@gmail.com.

ABSTRACT

The primary goal of the research is to identify credit card fraud in the real world. Recent years have seen a sharp increase in fraudulent activity due to the credit card industry's spectacular expansion in transaction volume. The goal is to take items without paying for them or to withdraw money from an account without authorization. For all institutions that issue credit cards, implementing effective fraud detection systems has become essential to reducing losses. The fact that neither the card nor the cardholder must be present for the transaction to be completed is one of the biggest obstacles to the company. Because of this, the merchant is unable to confirm that the client making the transaction is, in fact, the legitimate cardholder. The accuracy of identifying fraud may be enhanced with the suggested scheme's use of the Xgboost algorithm and random forest. Random forest and Xgboost algorithms are used in the classification process to assess the user's current dataset and data collection. Optimise the correctness of the resultant data in the end. The methods' effectiveness is assessed in terms of precision, sensitivity, specificity, and accuracy. The graphical model visualisation is then produced after processing a portion of the supplied characteristics to identify the fraud detection. The methods' effectiveness is assessed in terms of precision, sensitivity, specificity, and accuracy.The Xgboost algorithm outperformed Random Forest in accuracy.

Index Terms:- Feature extraction, Malware,Genetic algorithms,Machine learning,Xgboost,Machine learning algorithms.

I.INTRODUCTION

Many approaches for detecting fraudulent activity in credit card transactions have been used, and researchers have been considering ways to create models based on artificial intelligence, data mining, fuzzy logic, and machine learning. While credit card fraud detection is a common topic to address, it is also a very challenging one. We used machine learning to build the credit card fraud detection feature in our proposed system. With the development of methods for machine learning. It has been determined that machine learning is an effective tool for fraud detection. Online transaction procedures include the transmission of a lot of data, which produces a binary outcome: authentic or fraudulent. Features are created inside the sample fake datasets. These are informational details, such as the credit card's origin and age, as well as the value of the client account. There are hundreds of characteristics, and each one influences the likelihood of fraud to a different degree. It should be noted that the artificial intelligence of the machine, which is powered by the training data, determines how much each attribute contributes to the fraud score; fraud analysts do not decide this. Therefore, in terms of credit card fraud, a transaction using a credit card will have an equally high fraud weighting if it is shown that the use of cards for fraud is widespread. On the other hand, the donation level would parallel if this shrank. To put it simply, these models learn on their own without explicit programming, like manual review. Regression and classification algorithms are used in machine learning to identify credit card fraud. We classify online and offline fraud card transactions using supervised learning algorithms like Random Forest. An enhanced form of the decision tree is the random forest. Comparing Random Forest to other machine learning algorithms, it is more accurate and efficient. Random forest selects just a subsample of the feature space at each split in an attempt to mitigate the correlation problem that was previously highlighted. It basically tries to solve a stopping criterion for node splits, which I will address in more detail later, and make the trees de-correlated and pruned.

II.LITERATURE SURVEY

- Kosemani Temitayo Hafiz, Dr. Shaun Aghili, Dr. Pavol Zavarsky proposed this research paper focuses on the creation of a scorecard from relevant evaluation criteria, features, and capabilities of predictive analytics vendor solutions currently being used to detect credit card fraud. The scorecard provides a side-byside comparison of five credit card predictive analytics vendor solutions adopted in Canada. From the ensuing research findings, a list of credit card fraud PAT vendor solution challenges, risks, and limitations was outlined.

- Amlan Kundu, Suvasini Panigrahi, Shamik Sural described this paper propose to use two-stage sequence alignment in which a profile Analyser (PA) first determines the similarity of an incoming sequence of transactions on a given credit card with the genuine cardholder's past spending sequences. The unusual transactions traced by the profile analyser are next passed on to a deviation analyser (DA) for possible alignment with past fraudulent behaviour. The final decision about the nature of a transaction is taken on the basis of the observations by these two analysers. In order to achieve online response time for both PA and DA, we suggest a new approach for combining two sequence alignment algorithms BLAST and SSAHA.
- Wen-Fang YU, Na Wang describe along with increasing credit cards and growing trade volume in China, credit card fraud rises sharply. How to enhance the detection and prevention of credit card fraud becomes the focus of risk control of banks. It proposes a credit card fraud detection model using outlier detection based on distance sum according to the infrequency and unconventionality of fraud

in credit card transaction data, applying outlier mining into credit card fraud detection. Experiments show that this model is feasible and ac.

- Vijayshree B. Nipane, Poonam S. Kalinge explains with growing advancement in the electronic commerce field, fraud is spreading all over the world, causing major financial losses. In current scenario, Major cause of financial losses is credit card fraud; it not only affects trades person but also individual clients. Decision tree, Genetic algorithm, Meta learning strategy, neural network, HMM are the presented methods used to detect credit card frauds. In contemplate system for fraudulent detection, artificial intelligence concept of Support Vector Machine (SVM) & decision tree is being used to solve the problem. Thus by implementation of this hybrid approach, financial losses can be reduced to greater extend.
- Sitaram patel, Sunita Gond thesis propose the SVM (Support Vector Machine) based method with multiple kernel involvement which also includes several fields of user profile instead of only spending profile. The simulation result shows improvement in TP (true positive), TN (true negative) rate, & also decreases the FP (false positive) & FN (false negative) rate.

III. EXISTING SYSTEM

In existing System, a research about a case study involving credit card fraud detection, where data normalization is applied before Cluster Analysis and with results obtained from the use of Cluster Analysis and Artificial Neural

Networks on fraud detection has shown that by clustering attributes neuronal inputs can be minimized. And promising results can be obtained by using normalized data and data should be MLP trained. This research was based on unsupervised learning. Significance of this paper was to find new methods for fraud detection and to increase the accuracy of results. The data set for this paper is based on real life transactional data by a large European company and personal details in data is kept confidential. Accuracy of an algorithm is around 50%. Significance of this paper was to find an algorithm and to reduce the cost measure. The result obtained was by 23% and the algorithm they find was Bayes minimum risk.

DISADVANTAGES OF EXISTING SYSTEM:

1. In this paper a new collative comparison measure that reasonably represents the gains and losses due to fraud detection is proposed.

2. A cost sensitive method which is based on Bayes minimum risk is presented using the proposed cost measure.

IV PROPOSED SYSTEM:

In proposed System, we are applying random forest algorithm for classification of the credit card dataset. Random Forest is an algorithm for classification and regression. Summarily, it is a collection of decision tree classifiers. Random forest has advantage over decision tree as it corrects the habit of over fitting to their training set. A subset of the training set is sampled randomly so that to train each individual tree and then a decision tree is built, each node then splits on a feature selected from a random subset of the full feature set. Even for large data sets with many features and data instances training is extremely fast in random forest and because each tree is trained independently of the others. The Random Forest algorithm has been found to provide a good estimate of the generalization error and to be resistant to over fitting.

ADVANTAGES OF PROPOSED SYSTEM:

- The importance of variables in a regression or classification problem in a natural way can be done by Xgboost.
- The transaction amount. Feature 'class' is the target class for the binary classification and it takes value 1 for positive case (fraud) and 0 for negative case (not fraud)

V. SYSTEM DESIGN

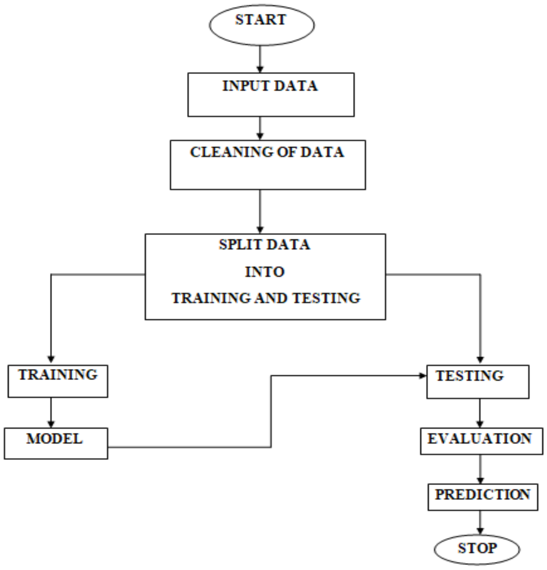


Fig1: Architecture of system.

VI.MODULE DESCRIPTION:

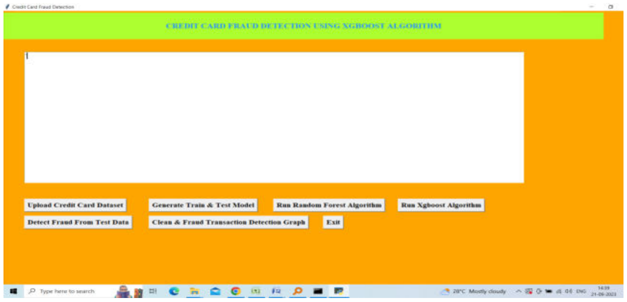
1. **. Upload Credit Card Dataset**
In this module user upload Credit Card Dataset.
2. **Generate Train & Test Model**
In this module user train & test model through dataset.
3. **Run Random Forest Algorithm**
In this module random forest algorithm classify dataset.

4. **Run Xgboost Algorithm**
In this module xgboost algorithm classify dataset.

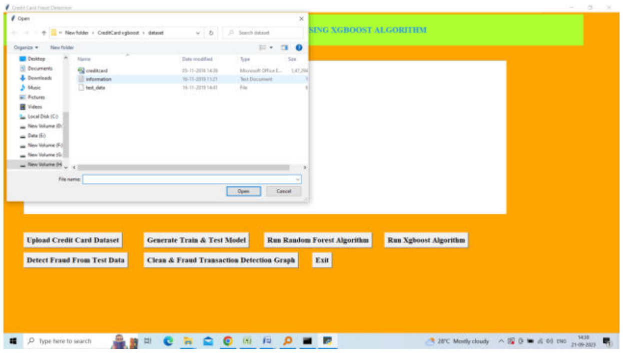
5. **Detect Fraud From Test Data**
In this module fraud is detected from dataset.
6. **Clean & Fraud Transaction Detection Graph**
In this module clean & Fraud Transaction detection graph is shown.

VII. RESULT:

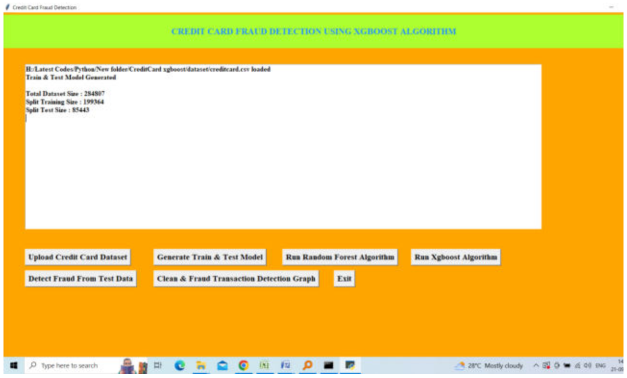
Application Screen


Credit Card Fraud Detection
CREDIT CARD FRAUD DETECTION USING XGBOOST ALGORITHM
Upload Credit Card Dataset
Generate Train & Test Model
Run Random Forest Algorithm
Run Xgboost Algorithm
Detect Fraud From Test Data
Clean & Fraud Transaction Detection Graph
Exit
Type here to search
28°C Mostly cloudy

Upload dataset

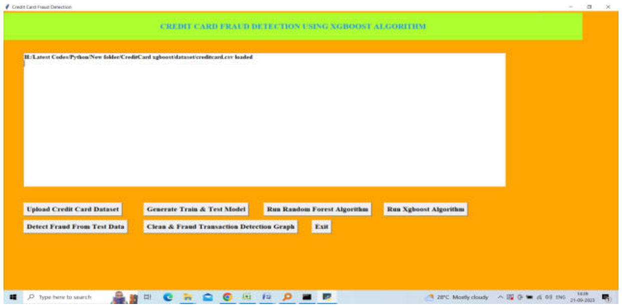


Train and Test


Credit Card Fraud Detection
CREDIT CARD FRAUD DETECTION USING XGBOOST ALGORITHM
Upload Credit Card Dataset
Generate Train & Test Model
Run Random Forest Algorithm
Run Xgboost Algorithm
Detect Fraud From Test Data
Clean & Fraud Transaction Detection Graph
Exit
Type here to search
28°C Mostly cloudy

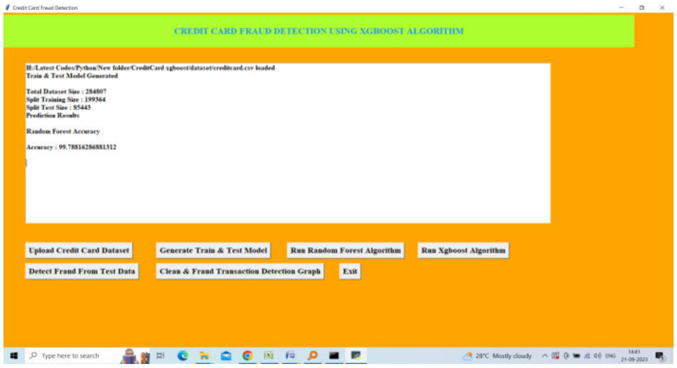
```
H:/Latest Codes/Python/New folder/CreditCard xgboost/dataset/creditcard.csv loaded
Train & Test Model Generated

Total Dataset Size : 284807
Split Training Size : 199364
Split Test Size : 85443
```

Random Forest Algorithm

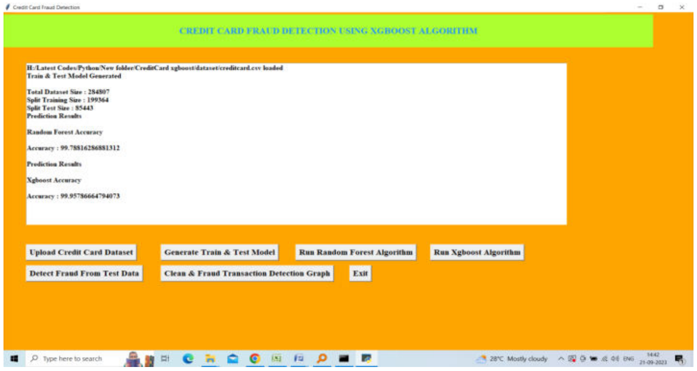

Credit Card Fraud Detection
CREDIT CARD FRAUD DETECTION USING XGBOOST ALGORITHM
Upload Credit Card Dataset
Generate Train & Test Model
Run Random Forest Algorithm
Run Xgboost Algorithm
Detect Fraud From Test Data
Clean & Fraud Transaction Detection Graph
Exit
Type here to search
28°C Mostly cloudy

```
H:/Latest Codes/Python/New folder/CreditCard xgboost/dataset/creditcard.csv loaded
```

Run Xgboost Algorithm

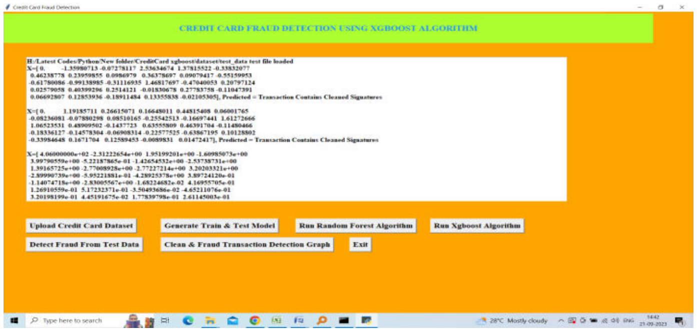

Credit Card Fraud Detection
CREDIT CARD FRAUD DETECTION USING XGBOOST ALGORITHM
Upload Credit Card Dataset
Generate Train & Test Model
Run Random Forest Algorithm
Run Xgboost Algorithm
Detect Fraud From Test Data
Clean & Fraud Transaction Detection Graph
Exit
Type here to search
28°C Mostly cloudy

```
H:/Latest Codes/Python/New folder/CreditCard xgboost/dataset/creditcard.csv loaded
Train & Test Model Generated

Total Dataset Size : 284807
Split Training Size : 199364
Split Test Size : 85443
Prediction Results

Random Forest Accuracy

Accuracy : 99.78816286881312
```


Credit Card Fraud Detection
CREDIT CARD FRAUD DETECTION USING XGBOOST ALGORITHM
Upload Credit Card Dataset
Generate Train & Test Model
Run Random Forest Algorithm
Run Xgboost Algorithm
Detect Fraud From Test Data
Clean & Fraud Transaction Detection Graph
Exit
Type here to search
28°C Mostly cloudy

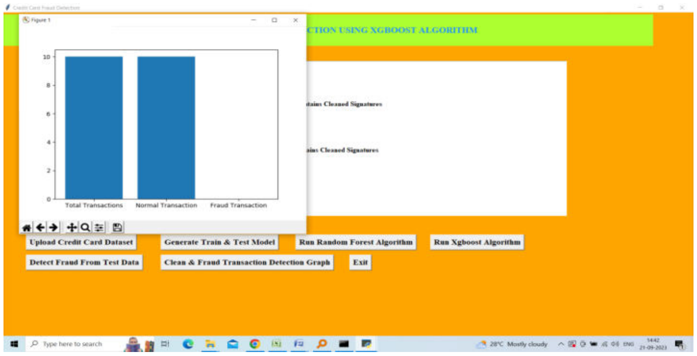
```
H:/Latest Codes/Python/New folder/CreditCard xgboost/dataset/creditcard.csv loaded
Train & Test Model Generated

Total Dataset Size : 284807
Split Training Size : 199364
Split Test Size : 85443
Prediction Results

Random Forest Accuracy

Accuracy : 99.78816286881312

Prediction Results

Xgboost Accuracy

Accuracy : 99.95786664794073
```

Detect Fraud from Test Data


Credit Card Fraud Detection
CREDIT CARD FRAUD DETECTION USING XGBOOST ALGORITHM
Upload Credit Card Dataset
Generate Train & Test Model
Run Random Forest Algorithm
Run Xgboost Algorithm
Detect Fraud From Test Data
Clean & Fraud Transaction Detection Graph
Exit
Type here to search
28°C Mostly cloudy

```
H:/Latest Codes/Python/New folder/CreditCard xgboost/dataset/test_data test file loaded
X=[ 0.         -1.35980713 -0.07278117  2.53634674  1.37815522 -0.33832077
  0.46238778  0.23959855  0.0986979   0.36378697  0.09079417 -0.55159953
 -0.61780086 -0.99138985 -0.31116935  1.46817697 -0.47040053  0.20797124
  0.02579058  0.40399296  0.2514121  -0.01830678  0.27783758 -0.11047391
  0.06692807  0.12853936 -0.18911484  0.13355838 -0.02105305], Predicted = Transaction Contains Cleaned Signatures

X=[ 0.          1.19185711  0.26615071  0.16648011  0.44815408  0.06001765
 -0.08236081 -0.07880298  0.08510165 -0.25542513 -0.16697441  1.61272666
  1.06523531  0.48909502 -0.1437723   0.63555809  0.46391704 -0.11480466
 -0.18336127 -0.14578304 -0.06908314 -0.22577525 -0.63867195  0.10128802
 -0.33984648  0.1671704   0.12589453 -0.0089831   0.01472417], Predicted = Transaction Contains Cleaned Signatures

X=[ 4.06000000e+02 -2.31222654e+00  1.95199201e+00 -1.60985073e+00
  3.99790559e+00 -5.22187865e-01 -1.42654532e+00 -2.53738731e+00
  1.39165725e+00 -2.77008928e+00 -2.77227214e+00  3.20203321e+00
 -2.89990739e+00 -5.95221881e-01 -4.28925378e+00  3.89724120e-01
 -1.14074718e+00 -2.83005567e+00 -1.68224682e-02  4.16955705e-01
  1.26910559e-01  5.17232371e-01 -3.50493686e-02 -4.65211076e-01
  3.20198199e-01  4.45191675e-02  1.77839798e-01  2.61145003e-01
```

Clean and Fraud Transaction Detection Graph


Credit Card Fraud Detection
Figure 1
10
8
6
4
2
0
Total Transactions
Normal Transaction
Fraud Transaction
CTION USING XGBOOST ALGORITHM
tains Cleaned Signatures
ains Cleaned Signatures
Upload Credit Card Dataset
Generate Train & Test Model
Run Random Forest Algorithm
Run Xgboost Algorithm
Detect Fraud From Test Data
Clean & Fraud Transaction Detection Graph
Exit
Type here to search
28°C Mostly cloudy

VIII. CONCLUSION

The xgboost algorithm will perform better with a larger number of training data, but speed during testing and application will suffer. Application of more pre-processing techniques would also help. The SVM algorithm still suffers from the imbalanced dataset problem and requires more preprocessing to give better results at the results shown by SVM is great but it could have been better if more preprocessing have been done on the data.

IX. FUTURE ENHANCEMENT

For future work, the methods studied in this paper will be extended to online learning models. In addition, other online learning models will be investigated. The use of online learning will enable rapid detection of fraud cases, potentially in real-time. This in turn will help detect and prevent fraudulent transactions before they take place, which will reduce the number of losses incurred every day in the financial sector.

X. REFERENCES

[1] Sudhamathy G: Credit Risk Analysis and Prediction Modelling of Bank Loans Using R, vol. 8, no-5, pp. 1954-1966.

[2] LI Changjian, HU Peng: Credit Risk Assessment for ural Credit Cooperatives based on Improved Neural Network, International Conference on Smart Grid and Electrical Automation vol. 60, no. - 3, pp 227-230, 2017.

[3] Wei Sun, Chen-Guang Yang, Jian-Xun Qi: Credit Risk Assessment in Commercial Banks Based On Support Vector Machines, vol.6, pp 2430-2433, 2006.

[4] Amlan Kundu, Suvasini Panigrahi, Shamik Sural, Senior Member, IEEE, “BLAST-SSAHA Hybridization for Credit Card Fraud Detection”, vol. 6, no. 4 pp. 309-315, 2009.

[5] Y. Sahin and E. Duman, “Detecting Credit Card Fraud by Decision Trees and Support Vector Machines, Proceedings of International Multi Conference of Engineers and Computer Scientists, vol. I, 2011.

[6] Sitaram patel, Sunita Gond , “Supervised Machine (SVM) Learning for Credit Card Fraud Detection, International of engineering trends and technology, vol. 8, no. -3, pp. 137- 140, 2014.

[7] Snehal Patil, Harshada Somavanshi, Jyoti Gaikwad, Amruta Deshmane, Rinku Badgujar," Credit Card Fraud Detection Using Decision Tree Induction Algorithm, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.4, April- 2015, pg. 92-95

[8] Dahee Choi and Kyungho Lee, “Machine Learning based Approach to Financial Fraud Detection Process in Mobile Payment System", vol. 5, no. - 4, December 2017, pp. 12-24.

9. C. Alippi, G. Boracchi, and M. Roveri, “Just-in-time classifiers for recurrent concepts,” IEEE Trans. Neural Netw. Learn. Syst., vol. 24, no. 4, pp. 620–634, Apr. 2013.

10. B. Baesens, V. Van Vlasselaer, and W. Verbeke, Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection. Hoboken, NJ, USA: Wiley, 2015.