iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XII **Month of publication:** December 2022

DOI: https://doi.org/10.22214/ijraset.2022.47769

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

Credit Card Fraud Detection

Mrs. M.M. Swami[1], Rushikesh Ghuge[2], Gurshan Singh[3], Harsh Tiwari[4], Rohan Kalaskar[5]
Dept of Computer Science and Engineering, AISSMS College of Engineering, Pune, Maharashtra, India

Abstract: Due to exponential growth in the field of online transactions, credit cards are widely used in most financial aspects and hence there are more risks of fraudulent transactions. These fraudulent transactions can be shown by analysing several behaviours of credit card users from earlier transaction history datasets. If any abnormality is noticed in the behaviour from the existing patterns, there is the possibility of fraudulent transaction. In this project the proposed will use Ensemble Learning Algorithms (XGBoost). By using these models, the proposed system will predict if the transaction is fraudulent or genuine. Therefore, by the implementation of this methodology in fraud detection systems, monetary losses which are caused due to fraudulent transactions can be decreased.

I. INTRODUCTION

With the world moving forward the era of cash lessness is coming, but with every advancement come with its drawbacks. With the increase in use of credit cards, the fraudulent transactions have increased. Credit card fraud may be defined as use of a credit/debit card which has been reported lost, revoked, or stolen to obtain anything of value. The entire consumer credit industry is affected by it. Credit fraud is one of the fastest-growing types of fraud and the most difficult to intercept. This fraud can occur in various scenarios either the website's security was breached or caused due to the negligence of owner.

The soul justification of this research paper is to point out the likeness of the fraudulent transaction while use of a credit card. To achieve this goal, the primary step is to create a fraud detection system by using machine learning, which discovers the fraud transactions with high accuracy and less time. The proposed system will use Ensemble Learning Algorithms like XGBoost. By manoeuvring these models, the system will anticipate if the transaction is genuine/non- fraudulent or fraudulent.

Ensemble learning model uses multiple algorithms to assure better predictive performance that could be acquired from any of the inherent learning algorithms alone. Which make this model more fast, accurate and effective among the rest of the models.

II. MOTIVATION

Now a day the consumer prefers the most welcome payment mode that is via credit card. Credit card are the most fitting way for online shopping, paying bills etc. At the same time, the fraudulent transactions using credit card is a chief issue which must be circumvented. Thereby, there are various techniques at one's disposal to avoid these threats effectively. In this research a system will be modelled using Ensemble learning algorithms to detect the fraudulent transaction.

A. Ensemble Learning

Ensemble learning model uses multiple algorithms to assure better predictive performance than could be acquired from any of the inherent learning algorithms alone. An ensemble-system is acquired by integration of various models (henceforth classifiers). Consequently, such systems are also named as multiple classifier systems. Ensemble learning usually construct more precise solution then a sole model would.

Ensemble learning can be classified into three classes which are generally discussed and applied in practice

1) Bagging

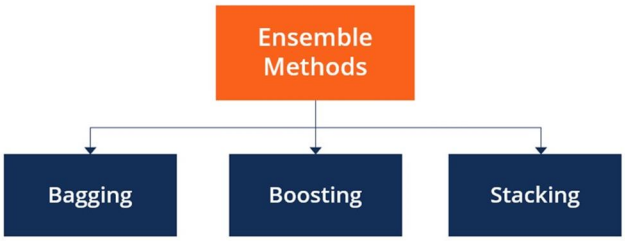
It requires fitting multiple decision trees on different samples of the same dataset and averaging the predictions.

2) Boosting

It requires adding ensemble members consecutively to exact the predictions put together by preceding models and outcomes as a weighted average of the projection.

3) Stacking

It requires fitting several types of models on the matching data and running another model to grasp how to best integrate the predictions.



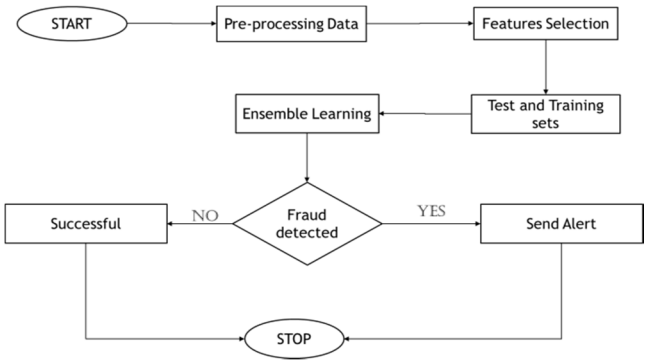
III. LITERATURE REVIEW

TITLE	METHODOLOGY
[1] Credit Card Fraud Detection Using Random Forest Algorithm	Random Forest, Accuracy 90%
[2] Credit Card Fraud Detection Using Deep Learning	Convolutional Neural Network, Accuracy 90%
[3] Credit Card Fraud Detection	Isolation Forest Accuracy 95%
[4] Credit Card Fraud Detection Using Machine Learning	AdaBoost, Accuracy 95%
[5] Detecting Default Payment Fraud in Credit Cards	Support Vector Machine, Accuracy 82%
[6] Credit card fraud identification based on unbalanced data set based on fusion model	Lasso Logistic Xgboost, Accuracy 95%
[7] Real-time Credit Card Fraud Detection Using Machine Learning	Convolutional Neural Network, Accuracy 91%

M. Suresh Kumar et al. implemented Random Forest algorithm in [1]. They used a real-world data set and got the accuracy of 90%. Whereas in [3] isolation forest was used which increased the detection accuracy from 90% to 95%. Dr. Anju Pratap and Anu Maria Babu proposed a model [2] using Convolutional Neural Network with an accuracy of 90%. They used a dataset with 284,807 transactions and 31 columns where 494 transactions were fraud. Dataset is broken into 80: 20 where 80% of data are used for training CNN model while 20% are used for purpose training. In [7] Anuruddha Thennakoon et al. also used CNN with an accuracy of 91% but they used two datasets, a genuine transaction log and fraudulent transaction log. They used SMOTE for oversampling and for under-sampling CNN and RUS were used. Donglin Li in [6] joint lasso logistic and XG boost algorithm to produce an accuracy of 95%. Lasso Logistic is slower while XGBoost is faster but it holds some unimportant variables. Santanu Kumar et al. used SVM in [5] with an accuracy of 82%, the data was gathered from UCI machine learning repository with total 690,000 instances. It has 23 columns and 30,000 rows.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

IV. SYSTEM ARCHITECTURE



A. Data Source and Description

In the proposed system architecture, the system is provided with the dataset which is taken from Kaggle which is an open-source platform. In the dataset there are 284,807 transactions or rows. The dataset contains features from V1 to V28 that are the principal components obtained by PCA transformation. The only attributes which are not have been transformed are Amount and Time.

B. Data Preprocessing

As the dataset is highly unbalanced the proposed system will check for the missing values and it will be filled using KNN algorithm. After checking that the dataset has no missing values, it will standardize the features for easy numerical operations and for the optimization of algorithm.

C. Dividing the training set and test set

After selecting the features, the dataset will be split into two partitions as training dataset and testing dataset. Only the training dataset will be used to build the model and test data will be used for the evaluation of the model. The proposed system will use 70% of the data to train the model and 30% for the testing of the model.

D. Features Selection

The process of reducing the input variables is known as Feature Selection. The proposed system will use backward elimination method for selecting the features. In backward elimination method initially all variables are taken and one by one non-significant variables are removed.

1) *Overfitting and Underfitting:* Overfitting is a situation where the variables are more than required and it causes decrease in accuracy whereas in underfitting variables are less than required which also leads to decrease in accuracy. In case of overfitting we can use regularization.

E. Algorithm

1) *XGBoost:* XGBoost Classifier (Extreme Gradient Boosting) will be applied in the proposed system which is an ensemble learning algorithm. XGBoost is based on GDBT (Gradient Boosting Decision Tree).
2) *Working of XGBoost:* In this algorithm, progressive decision trees are created. After that all independent variables are assigned with weights, which are fed to the decision tree which predicts results. The weights of variables which are predicted incorrect by the tree are increased and fed to second decision tree. These predictions are then ensembled to give a more precise model.

F. Training and Testing procedure

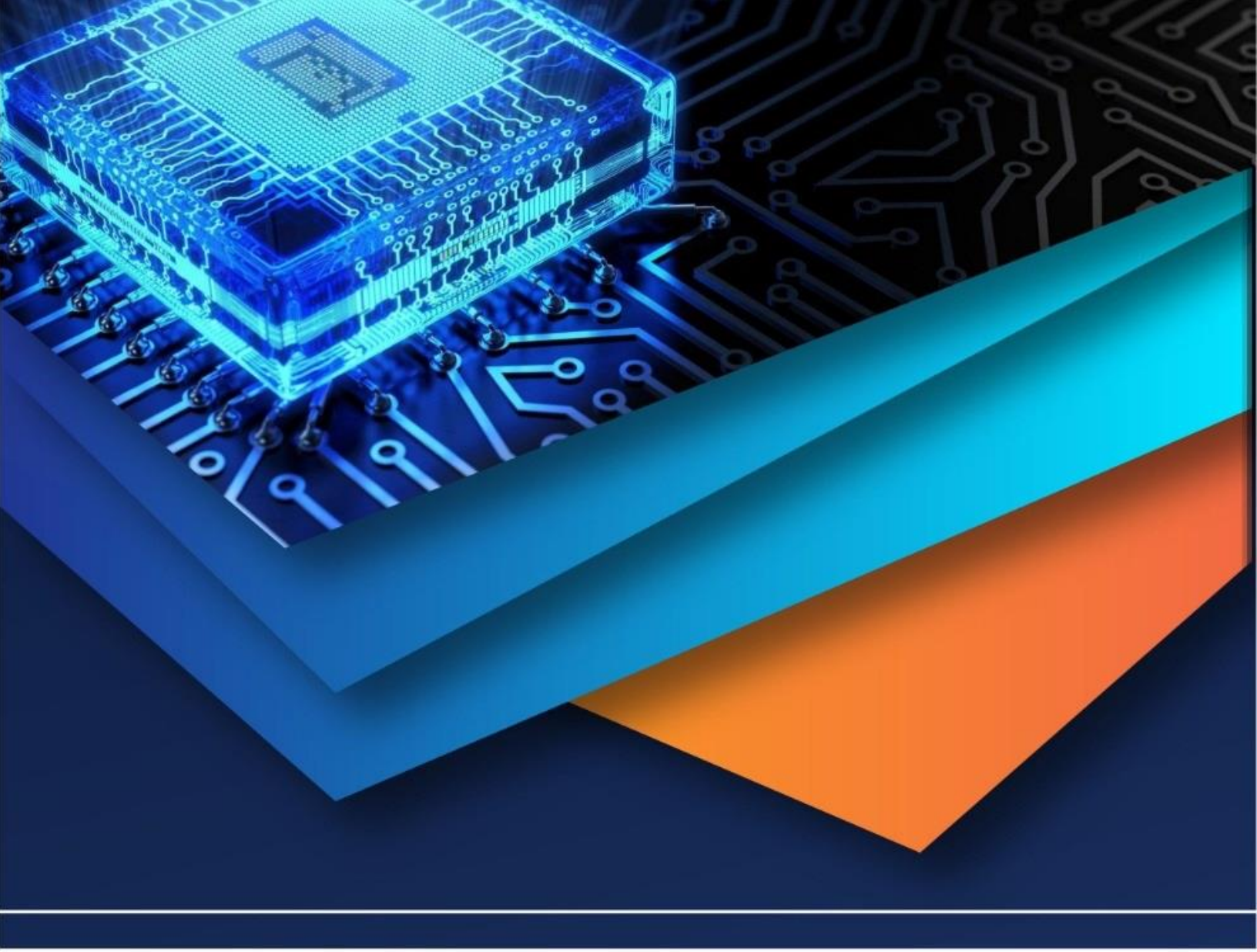
The proposed system will train the model with the help of training dataset which will be created in earlier processes, in the training phase the proposed system will be trained with random samples from training dataset. After successful training of model, the model will be tested with the help of testing dataset. The proposed system will check the predicting output against the known fraud transactions to check the accuracy of the model. We can also obtain confusion matrix which we help to find out the accuracy of the model.

V. CONCLUSION

Credit fraud has become increasingly uncontrolled in recent years. The aim is to find the fraudulent transaction before they cause any harm or loss to consumers. Thus the goal is to model a system with help of ensamble learning algorithms. The hope is to achive a model with higer acuracy and which consumes less time than the indivisual models. With thegoal to lessen the loss caused by credit frauds in each year, this paper put forward an ensemble approach, to combine multpul algorithms to to assure better predictive performance than could be acquired from any of the inherent learning algorithms alone.

REFFERENCES

[1] M. Suresh Kumar, V. Soundarya, S. Kavitha, E.S. Keerthika, E. Aswini" CREDIT CARD FRAUD DETECTION USING RANDOM FOREST ALGORITHM" IEEE 2019
[2] Dr. Anju Pratap, Anu Maria Babu" Credit Card Fraud Detection Using Deep Learning" IEEE 2020
[3] Zaiyyan Khan, Siddhesh Jadhav, Hashim Malik, Munira Ansari "Credit Card Fraud Detection" IJERT 2021
[4] D. Eswar, CH V N M Praneeth, Raja Subramanian, D. Tanouz "Credit Card Fraud Detection Using Machine Learning" IEEE 2021
[5] Santanu Kumar Rath, Debachudamani Prusti, Aditya Sai Kandukuri, S. S. Harshini Padmanabhuni "Detecting Default Payment Fraud in Credit Cards" IEEE 2019
[6] Donglin li "Credit card fraud identification based on unbalanced data set based on fusion model" IEEE 2019
[7] Anuruddha Thennakoon, Chee Bhagyani, Sasitha Premadasa, Shalitha Mihiranga, Nuwan Kuruwitaarachchi "Real-time Credit Card Fraud Detection Using Machine Learning" IEEE 2019
[8] Vehbi Cagri Gungor, Cengiz Gezer, Gokhan Goy "Credit Card Fraud Detection with Machine Learning Methods" IEEE 2020
[9] Hint Wint Khin, Aye Aye Khine "Credit Card Fraud Detection Using Online Boosting with Extremely Fast Decision Tree" IEEE 2020
[10] Abhishek M, Navaneeth A V,Dileep M R" A Novel Approach for Credit Card Fraud Detection using Decision Tree and Random Forest Algorithms" IEEE 2021
[11] Andrew. Y. Ng, Michael. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes", Advances in neural information processing systems, vol. 2, 2002
[12] John Richard D. Kho, Larry A. Vea "Credit Card Fraud Detection Based on Transaction Behaviour" published by Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, IEEE 2017. 5
[13] Yashvi Jain, Namrata Tiwari, ShripriyaDubey, Sarika Jain, "A Comparative Analysis of Various Credit Card Fraud Detection Techniques, Blue Eyes Intelligence Engineering and Sciences Publications 2019"
[14] Learning Robert A. Sowah, Moses A. Agebure, Godfrey A. Mills, Koudjo M. Kaumudi, "New Cluster Under Sampling Technique for Class Imbalance "IJMLC 2016
[15] Baraneetharan, E. "Role of Machine Learning Algorithms Intrusion Detection in WSNs: A Survey." Journal of Information Technology 2, no. 03 (2020).
[16] Mohamed Jaward Bah, Mohamed Hammad "Progress in Outlier Detection Techniques: A Survey" Hongzhi Wang, of the IEEE 2019
[17] A. Bifet and R. Kirkby Massive Online Analysis, Technical Manual, Univ. of Waikato, 2009.
[18] R. Bolton and D. Hand, "Unsupervised Profiling Methods for Fraud Detection," Statistical Science, vol. 17, no. 3, pp. 235255, 2001.
[19] P. Brockett, R. Derrig, L. Golden, A. Levine, and M. Alpert, "Fraud Classification Using Principal Component Analysis of RIDITs," The J. Risk and Insurance, vol. 69, no. 3, pp. 341-371, 2002, doi: 10.1111/15396975.00027.
[20] R. Caruana and A. Niculescu-Mizil, "Data Mining in Metric Space: An Empirical Analysis of Supervised Learning Performance Criteria," Proc. 10th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '04), 2004, doi: 10.1145/1014052.1014063.
[21] P. Christen and K. Goiser, "Quality and Complexity Measures for Data Linkage and Deduplication," Quality Measures in Data Mining, F. Guillet and H. Hamilton, eds., vol. 43, Springer, 2007, doi: 10.1007/978-3-54044918-8.
[22] C. Cortes, D. Pregibon, and C. Volinsky, "Computational Methods for Dynamic Graphs," J. Computational and Graphical Statistics, vol. 12, no. 4, pp. 950-970, 2003, doi: 10.1198/1061860032742.
[23] Experian. Experian Detect: Application Fraud Prevention System, Whitepaper, http://www.experian.com/products/pdf/expe rian_dete ct.pdf, 2008.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY