Credit Card Fraud Detection Using Machine Learning

Akshai Biju
Department of Computer Applications
Amal Jyothi College of Engineering Kanjirappally, India
akshaibiju@mca.ajce.in

Merin Manoj
Asst.professor Department of Computer Application
AmalJyothi College of Engineering Kottayam, India
merinmanoj@amaljyothi.ac.in

Abstract: Because of technological advancements,the growth of Ecommerce industry has been increasedand it leads to the usage of credit card transactions for online purchases. Nowadays peoples are most commonly using online transactions, and it is very comfortable and helpful. The most prevalent payment option nowadays is credit card transactions. As a result of these online transactions, the number of fraud casesis increasing day by day. Thus, it become one of the great challenges for the banks to detect this fraud in transactions. The objective of this work is to find the fraud in credit card accurately. The machine learning helps us to detect these type of fraud activities that occur in credit card transactions in accurate manner. In this paper we have been included the problems that causes and activities of fraud in credit card. Several machine learning algorithms are built by applying boosting techniques to it, which includes logistic regression and random forest utilizing ensemble classifiers on an unbalanced dataset. The algorithms that have been used in this system are random forest classifier and ADA boost algorithm. The both of these algorithms are based on accuracy, precision, and area under curve score. We compare the results that obtained from both these algorithms and choose the one with the greatest accuracy, precision and area under curve score. Based on this we select the best algorithm for detecting fraud in credit card transactions. The conclusion of this research shows how to use supervised techniques to train and analyse the best classifier, resulting in a more accurate solution.

Keywords: Accuracy, f1 score, precision, fraud detection, credit card, Random Forest, ADA boost, CAT boost, XGA boost, Light GBM model.

I. INTRODUCTION

Theft or crimes committed with a credit card are referred to as credit card frauds. The basic goal of these thefts or frauds is to buy something without paying for it, or to withdraw or move funds from an account without authorization. Fraud detection systems are being employed to monitor a twelfth of one percent of all transactions, resulting in billions of dollars in lost revenue. Credit fraud is one of the most serious issues facing today's online enterprises. To avoid the frauds effectively, first we have to understand all the mechanisms that maybe executing by these frauds. These Credit card fraudsters uses a variety of methods to commit frauds. In virtual card purchases the fraudster only need to know the card details of the victim. For avoiding those the carddetails must be kept private. In order to protect the privacy of the credit card the card information's should not be leaked. Some of the methods mainly used by the fraudsters are phishing websites, using fake cards, intercepted cards, stolen cards etc. Usually in these types of cases the cardholders are unaware ofhis/her card information's has been stolen. The easiestway to detect such fraud activities are to check the spending patterns on every card and identify any deviations from the usual spending patterns. The greatest strategy to lower the rate of successful credit card frauds is to detect fraud by examining the cardholder's existing data buy. Because the data sets aren't available, and the outcomes aren't made public, the available data sets, such as logged data and user activity, should be used to detect fraud instances.

II. TYPES OF ALGORITHMS

Random Forest: It is supervised learning technique. The data from the dataset are given to multiple decision tree then those are trained and merges into a single forest. The result from the decision tree are selected by majority in case it is classifier, or selected by mean, median if result is regression.

ADA boost: It is the binary classification boosting technique. Multiple "weak classifiers" are combined into a single "strong classifier" in this common boosting method.

CAT boost: It is a decision tree gradient boosting method. Even without considerable hyper parameter adjustment, it gives good results. It is a fast and scalable and provides GPU support.

XGA boost: It's a scalable and adaptable boosting technique designed specifically for data science tasks.

DOI: 10.5281/zenodo.6364744
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

It is capable of dealing with missing data and regularization.

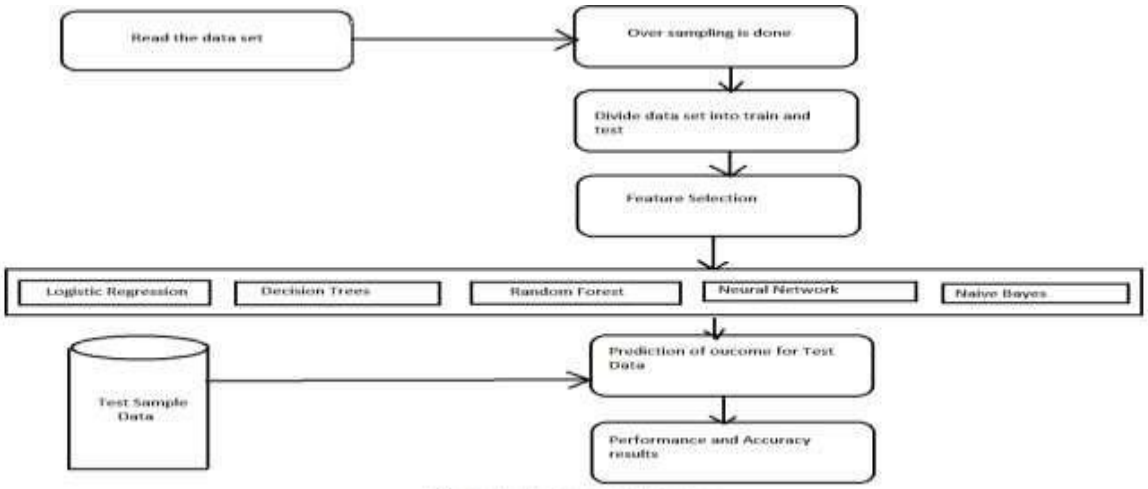


Figure1: System Architecture

III. PROPOSED TECHNIQUE

The proposed techniques for detecting frauds in credit card system are used in this paper. We compare all the results such as accuracy, precision, AUC score obtained from all the algorithms, to determine which one is the best and accurate that can be used by credit card merchants to identify fraud transactions. The figure below shows the architectural diagram to detect the best algorithm for overall system framework.

IV. METHODOLOGY

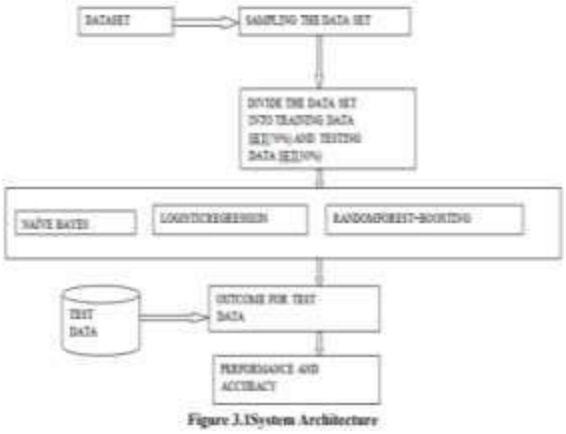
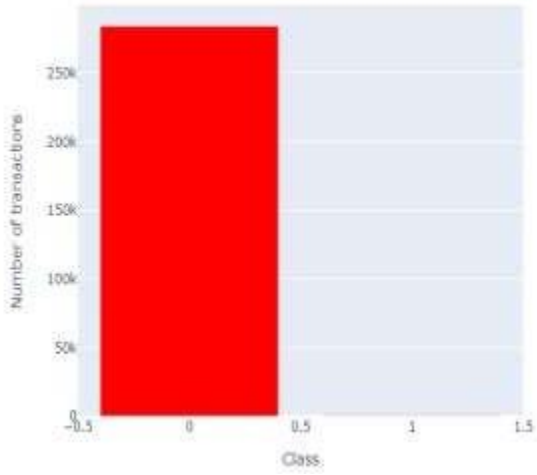


Figure 3.1System Architecture

V. Steps For Finding Best Algorithm

Step 1: Import the dataset
Step 2: Convert the data into data frames format.
Step 3: Do random sampling.
Step 4: Decide the amount of data for training data and testing data.
Step 5: Give 70% data for training and remaining data for testing
Step 6: Assign train datasets to the models.
Step 7: Apply the algorithm among 3 different algorithms and create the model.
Step 8: Make predictions for test datasets for each algorithm.
Step 9: Calculate accuracy of each algorithm
Step 10: Apply confusion matrix for each variable.
Step 11: Compare the algorithms for all the variables and find out the best algorithm.

VI. RESULTS AND DISCUSSIONS



DOI: 10.5281/zenodo.6364744
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

Fig 6.1: Credit card fraud class

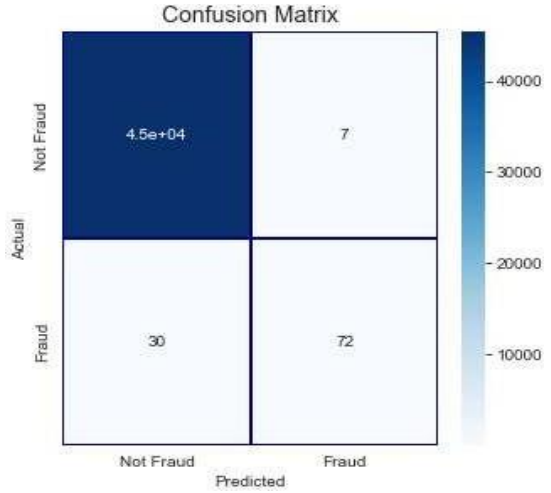


Fig 6.2: Confusion Matrix for Random Forest model

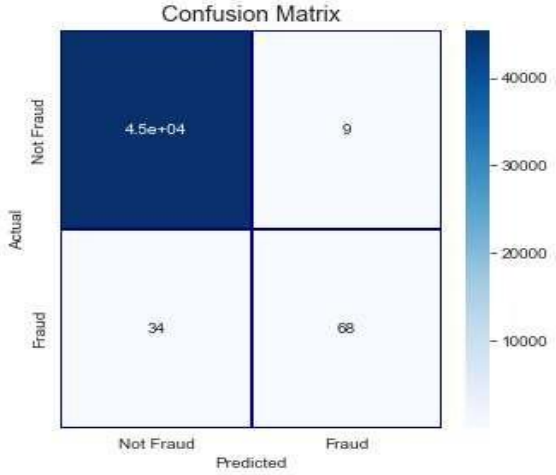


Fig 6.3: Confusion Matrix for ADA Boost model

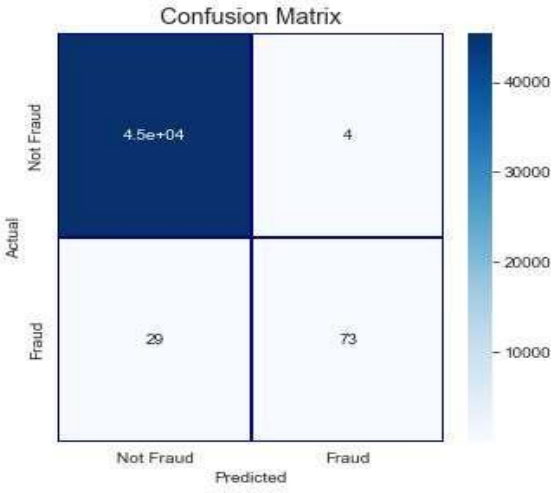


Fig 6.4: Confusion Matrix for CAT Boost model

VII. EXPERIMENTAL RESULTS

Performance analysis for three different algorithms

	Random Forest	ADA Boost	CAT Boost	XG Boost	Light GBM
AUC Score	0.85	0.83	0.86	0.97	0.946

VIII. CONCLUSIONS

In this paper we have been learned the applications of machine learning like Random Forest, ADA Boost, CAT Boost, XG Boost, Light GBM model. Supervised learning algorithms are a first in the literature in terms of application domain. The likelihood of fraudulent transactions can be anticipated quickly after they occurif these algorithms are integrated into a bank's credit card fraud detection system. In addition, a variety of anti-fraud techniques can be implemented to protect banks from large losses and minimize risks. The study's goal was approached differently than other classification issues in that we had a variable penalty of misclassification. The proposed system's performance is assessed using precision, f1-score, andaccuracy. We looked into the data, looking for data imbalances, displaying the features, and figuring out the relationships between them. The proposed system's performance is assessed using precision, f1- score, and accuracy. When predicting the target for thetest set, we got an AUC score of 0.85 with RandomForrestClassifier. When predicting the target for the test set with the ADA Boost Classifier, we got an AUC scode of 0.83When predicting the target for the test set with the CAT Boost Classifier, we got an AUC scode of 0.86. When predicting the target for the test set with the XG Boost Classifier, we got an AUC score of 0.97When predicting the target for the test set, we obtained an AUC scode of 0. 946 using the Light GBM Classifier. We were able to attain an AUC value of 0.93 for the test prediction using cross-validation.

IX. FUTURE SCOPE

The Random Forest with Boosting technique clearly outperforms the other credit card fraud detection techniques in this case, as seen in the above comparative analysis. However, one of the paper's flaws is that we can't use machine learning to distinguish the names of fraud and non-fraud transactions for the supplied dataset when employing

DOI: 10.5281/zenodo.6364744
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

the aforementioned three approaches. We can use a variety of techniques to overcome this challenge for the project's future development.

X. REFERENCES

[1]. https://tel.archives-ouvertes.fr/tel-02951477/document

[2]. https://towardsdatascience.com/credit-card-fraud-detection-using-machine-learning-python-5b098d4a8edc

[3]. https://ieeexplore.ieee.org/document/8123782

[4].https://www.researchgate.net/publication/336800562_Credit_Card_Fraud_Detection_using_Machine_Learning_and_Data_Science

DOI: 10.5281/zenodo.6364744
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam