IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of
Information Technology & Computer Engineering

ISSN 2347–3657

Volume 11, Issue 2, Apr 2023

ADVANCED CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING

V.NAVYA SREE1, J. SUPRAJA2, K. SPHOORTHY SRI 3, K. OJASWANI4

Article Info

Received: 29-01-2023 Revised: 20 -02-2023 Accepted: 2-03-2023

ABSTRACT;

Due to the rapid growth of the E-Commerce industry, the use of credit cards for online purchases has increased dramatically. In recent years, credit card fraud is becoming a major complication for banks as it has become very difficult for detecting fraud in the credit card system. To overcome this hardship Machine learning plays an eminent role in detecting the credit card fraud in the transactions. Modeling prior credit card transactions with data from ones that turned out to be fraudulent is part of the Card Fraud Detection Problem. In Machine learning the machine is trained at first to predict the output so, to predict the various bank transactions various machine learning algorithms are used. The SMOTE approach was employed to oversample the dataset because it was severely unbalanced. This paper the examines and overview the performance of K-nearest neighbors, Decision Tree, Logistic regression and Random forest, XGBoost for credit card fraud detection. The assignment is implemented in Python and uses five distinct machine learning classification techniques. The performance of the algorithm is evaluated by accuracy score, confusion matrix, f1-score, precision and recall score and auc-roc curve as well.

Key Words: Fraud Detection, Machine Learning, Logistic regression, KNN, Decision tree, random forest, SMOTE, XGboost.

INTRODUCTION

In credit card transactions, fraud is defined as the unlawful and unwanted use of an account by someone who is not the account's authorised user. This misuse, as well as the behaviour of such fraudulent operations, can be investigated in order to reduce it and prevent such occurrences in the future. In simple terms, credit card fraud occurs when a person uses another person's credit card for personal gain while the owner and card-issuing authorities are unaware of the transaction. It is currently one of the most serious risks to enterprises. However, to fight the fraud completely, it is essential to first understand the structure of executing a fraud. Credit card fraudsters opt many numbers of ways to commit fraud. Card fraud occurs when a physical card is stolen or when critical account information is stolen, such as the card account number or other information that must be available to conduct a transaction. A major challenge in applying Machine Learning to fraud detection is the presence of highly imbalanced dataset. In many publicly available databases, the vast majority of transactions are lawful, with only a small percentage of them being fraudulent.

1ASSISTANT PROFESSOR, DEPARTMENT OF ECE, MALLA REDDY ENGINEERING COLLEGE FOR WOMEN, HYDERABAD.
2,3&4UG SCHOLAR, DEPARTMENT OF ECE, MALLA REDDY ENGINEERING COLLEGE FOR WOMEN, HYDERABAD

Researchers face a big issue in designing an accurate fraud detection system with fewer fraud transactions compared to legal transactions, allowing them to detect fraudulent behaviour successfully. In our paper, we apply multiple classification approaches such as KNN, Decision Tree, Logistic regression and Random forest, XGBoost. Our aim is to build a classifier which will be able to separate fraud transactions from non-fraud ones. We will be comparing the accuracy and effectiveness of these applied algorithms in detecting fraud transactions.

LITERATURE REVIEW Fraud is defined as an illegal deception intended to gain financial or personal gain. It's a planned conduct that goes against the law or a policy with the goal of gaining unjust financial gain. Data mining applications and adversarial detection are among the strategies used in this domain, according to a comprehensive survey undertaken by Clifton Phua and his colleagues. On a European dataset, classic methods such as Decision Tree, XGboost, random forest, and a mixture of particular classifiers were utilised, resulting in a recall of over 91 percent. Only after balancing the dataset by oversampling the data was high precision and recall achieved.

METHODOLOGY

Proposed Method The proposed techniques emphasizes on detecting Credit Card Fraudulent transactions whether it is a genuine/nonfraud or a fraud transaction and the approaches used to separate fraud and non-fraud are KNN, Decision Tree, Logistic regression, XGBoost, Random forest and Finally we will observe which approach is best for detecting credit card frauds. The system architecture has following steps:

- Import of Necessary Packages
- Read the Dataset
- Exploratory Data Analysis i.e. finding null values, duplicate values etc.
- Selecting Features (X) and the Target (y) columns
- Train Test Split will split the whole dataset into train and test data
- Build the model i.e. Training the model
- Test the model i.e. Model prediction
- Evaluation of the system i.e. Accuracy score, F1- score etc. The figure(Fig-1) below shows the system architecture diagram.

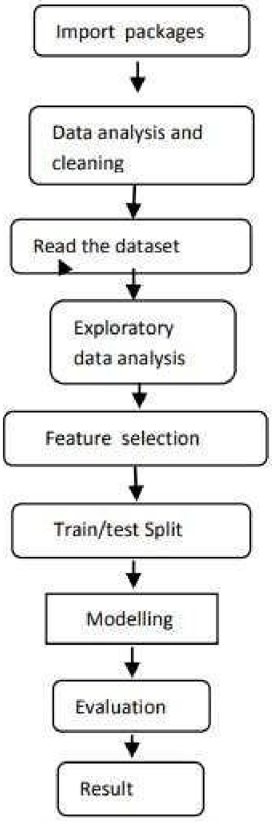

Fig. 1. : Architecture diagram

Machine learning: It is a set of strategies for identifying patterns in data on the fly and then using those patterns to predict future outcomes. Also, provides several algorithms that allow machines to perceive current events and make appropriate judgments based on that perception. It is selfcontained and makes its own decisions. Unsupervised learning and supervised learning are the two main types of machine learning.

Supervised Learning: In this technique, both the input and output are known ahead of time. This is known as supervised learning because it learns from a training data set and builds a model from it, which then predicts results when applied to new data. Supervised learning techniques include Decision Trees, Nave Bayes, and others.

Unsupervised Learning: When we have only input data and no corresponding output variable, we call it unsupervised learning. Unsupervised learning's main task is to automatically create class labels. The association between the data can be discovered using unsupervised learning methods to see if they can be grouped together. Clusters are the name for this type of group. Cluster analyses is another term for unsupervised learning. Unsupervised learning techniques include K Means Clustering, KNN, and others.

Dataset In this work, Kaggle's Credit Card Fraud Detection dataset was employed. The transactions in this dataset were made by European cardholders over the period of two days in September 2013. The dataset has 31 numerical features. The PCA transformation of these input variables was performed to keep these data anonymous due to privacy concerns and some of the input variables contained financial information. Three of the listed characteristics were not altered. The "Time" feature shows the amount of time that has passed between the first and subsequent transactions in the dataset. The "Amount" function shows the total amount of credit card transactions. The "Class" feature displays the label and only allows two values: 1 for fraudulent transactions and 0 for all regular transactions. The dataset included 284,807 transactions, 492 of which were fraudulent and the rest were legitimate. When we look at the numbers, we can see that the dataset is severely skewed, with only 0.173 percent of transactions being classified as fraudulent. Preprocessing the data is critical since the distribution ratio of classes plays such an important role in model accuracy and precision. As a result, it is critical to balance the data, which is accomplished using sampling procedures. The Smote technique was used.

Understanding the dataset Histogram plots and correlation matrix are being used to understand the dataset. Correlation matrix depicts if there is very little or no correlation between individual features and the targeted column. It gives an idea of how features correlate with each other and can help in predicting what features are more relevant for our prediction

Preprocessing Feature selection is a key strategy for determining which variables in a dataset are the most important. Overfitting can be reduced, accuracy can be improved, and training time can be reduced by carefully selecting useful features and deleting the less critical ones. Techniques like visualisation can help with this. When working with data that is highly uneven, some type of balancing is essential in order for the model to be trained efficiently. Changing the class distribution involves undersampling the dominant class, oversampling the minority class, or a mixof the two. SMOTE (Synthetic Minority Oversampling Technique) is a well-known oversampling technique that has been proved to work with unbalanced datasets

CONCLUSION Credit card fraud is a significant commercial issue. These types of scams can result in significant personal and business losses. As a result, businesses are investing an increasing amount of money in creating new concepts and methods for detecting and limiting fraud. The major purpose of this article was to look at a variety of machine learning algorithms for detecting fraudulent transactions. As a consequence of the comparison, it was discovered that the Xgboost algorithm produces the best results, i.e. best classifies whether transactions are fraudulent or not. This was determined using a variety of metrics, including recall, accuracy, and precision, the f1 score, and the AUC-roc curve. For this type of situation, having a high recall value is crucial. It has been established that feature selection and dataset balancing are critical in achieving significant results. Other machine learning techniques, such as evolutionary algorithms and various forms of stacked classifiers, as well as rigorous feature selection, should be studied further to improve results

REFERENCES

[1] Machine Learning For Credit Card Fraud Detection System, Lakshmi S V S , Selvani Deepthi Kavila, november 2018.

[2] Credit Card Fraud Detection using Data science and Machine learning, S P Maniraj, Aditya Saini , Shadab Ahmed, Swarna Deep Sarkar, September 2019.

[3] A. Mishra, C. Ghorpade, "Credit Card Fraud Detection on the Skewed Data Using Various Classification and Ensemble Techniques" 2018 IEEE International Students' Conference on Electronics ,Electrical and Computer Science (SCEECS) pp. 1-5. IEEE.

[4] S. V. S. S. Lakshmi, S. D. Kavilla "Machine Learning For Credit Card Fraud Detection System", unpublished

[5] C. Wang, Y. Wang, Z. Ye, L. Yan, W. Cai, S. Pan, "Credit card fraud detection on basis of whale algorithm optimized BP neural network", 2018 13th International Conference on Computer Science & Education (ICCSE) pp. 1-4. IEEE.