Credit-Card-Fraud-Detection

ABSTRACT:

Credit card fraud is a significant problem, with billions of dollars lost each year. Machine learning can be used to detect credit card fraud by identifying patterns that are indicative of fraudulent transactions. Credit card fraud refers to the physical loss of a credit card or the loss of sensitive credit card information. Many machine learning algorithms can be used for detection. This project proposes to develop a machine-learning model to detect credit card fraud. The model will be trained on a dataset of historical credit card transactions and evaluated on a holdout dataset of unseen transactions.

Keywords: Credit Card Fraud Detection, Fraud Detection, Fraudulent Transactions, K- Nearest Neighbors, Support Vector Machine, Logistic Regression, Decision Tree.

Overview

With the increase of people using credit cards in their daily lives, credit card companies should take special care of the security and safety of the customers. According to (Credit card statistics 2021), the number of people using credit cards worldwide was 2.8 billion in 2019; also, 70those users own a single card. Reports of Credit card fraud in the U.S. rose by 44.7in 2020. There are two kinds of credit card fraud, and the first is having a credit card account opened under your name by an identity thief. Reports of this fraudulent behaviour increased 48to 2020. The second type is when an identity thief uses an existing account you created, usually by stealing the information on the credit card. Reports on this type of Fraud increased 9to 2020(Daly, 2021). Those statistics caught We's attention as the numbers have increased drastically and rapidly throughout the years, which motivated We to resolve the issue analytically by using different machine learning methods to detect fraudulent credit card transactions within numerous transactions.

Project goals

The main aim of this project is the detection of fraudulent credit card transactions, as it is essential to figure out the fraudulent transactions so that customers do not get charged for the purchase of products that they did not buy. Fraudulent Credit card transactions will be detected with multiple ML techniques. Then, a comparison will be made between the outcomes and results of each method to find the best and most suited model for detecting fraudulent credit card transactions; graphs and numbers will also be provided. In addition, it explores previous literature and different techniques used to distinguish Fraud within a dataset.

Data Source:

The dataset was retrieved from an open-source website, Kaggle.com. It contains data on transactions made in 2013 by European credit card users in two days only. Thedataset consists of 31 attributes and 284,808 rows. Twenty-eight attributes are numeric variables that, due to the confidentiality and privacy of the customers, have been transformed using PCA transformation; the three remaining attributes are ”Time”, which contains the elapsed seconds between the first and other transactions of each Attribute, ”Amount” is the amount of each transaction, and the final attribute “Class” which contains binary variableswhere “1” is a case of fraudulent transaction, and “0” is not as case of fraudulent transaction.

Dataset: kaggle Dataset

Algorithm :

1) **Support Vector Machine (SVM)**
2) **Decision Tree (D.T.)**
3) **XGBoost**
4) **Random Forest**
5) **Logistic Regression**

Results:

Conducted A/B testing, achieving the highest AUC of 0.974 with XGBoost and oversampling techniques.

Model	AUC
LR	0.855
SVM	0.910
Random Forest	0.960
XG Boost	0.974

Future Work

There are many ways to improve the model, such as using it on different datasets with various sizes and data types or by changing the data splitting ratio and viewing it from a different algorithm perspective. An example can be merging telecom datato calculate the location of people to have better knowledge of the location of the card owner while his/her credit card is being used; this will ease the detection because if the card owner is in Dubai and a transaction of his card was made in Abu Dhabi, it will easily be detected as Fraud.

Conclusion

In conclusion, the main objective of this project was to find the most suited model for creditcard fraud detection in terms of the machine learning techniques chosen for the project. It was met by building the four models and finding the accuracies of them all; the best in terms of accuracy is KNN and Decision Tree, which scored 100 on credit card fraud and increased the customer’s satisfaction as it will provide themwith a better experience and feeling secure.