Mini Project Proposal (Group - 4)

Members: Abhishek Srivastava - 19BCE10071
Vishesh bhadauria - 19BCE10304
Nandita Jain - 19BCE10314
Aman Gupta - 19BCE10289
V Surya Kumar - 19BCE10286
Akshra Singh - 19BCE10295
Sanjana Singh - 19BCE10274
Rupesh agrawal - 19BCE10080
Soumya Rajadhyaksha - 19BAI10120
Rachita Jha - 19BCE10283"

Background/Overview

Credit card fraud detection has become one of the most important aspects in this era of digital payments. As a merchant, you cannot deny the fact that we all are moving towards a cashless society. Therefore, you cannot stick to traditional payment methods. It will not help you to grow your business. It is a fact that customers will not always enter your shop carrying cash in their pockets. They are now giving importance to debit card payments, credit card payments. Therefore, you will have to change the ecosystem of your business in such a way so that it can accept all forms of payments.

It is important that credit card companies are able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase.

The key objective of any credit card fraud detection system is to identify suspicious events and report them to an analyst while letting normal transactions be automatically processed.

For years, financial institutions have been entrusting this task to rule-based systems that employ rule sets written by experts. But now they increasingly turn to a machine learning approach, as it can bring significant improvements to the process.

Problem Statement

Detect credit card fraudulent transactions

By using classification Algorithm and monitoring cash transaction to Identify the cash transactions just below regulatory reporting thresholds. Also, Identifying a series of cash disbursements by customer number that together exceed regulatory reporting threshold.

Objective of the project :

a) The General objective of this project is to build a model which predicts fraud and fraud-less transactions with respect to the time and amount of the transaction in the best way using machine learning algorithms .

b)

- Predicting credit card transactions is fraud or not based on the transaction related amount , location and other transaction related data .
- Aims to track down credit card transaction data , which is done by detecting anomalies in the transaction data .
- Notifying the cardholder about any suspicious transaction.
- To identify the suspicious events and report them while letting normal transactions be automatically processed .

Approach/Methodology:

To begin with, we took the **creditcard.csv** file as the input data and created user-defined functions that calculate **false positive(fp)**, **false negative(fn)**, **true positive(tp)**, **true negative(tn)**, **True positive rate (tpr)**, **True negative rate(tnr)**, **False positive rate(fpr)** and **False-negative rate(fnr)** from the data set. This helps us in calculating the Receiver Operating Characteristic**(ROC)** and Area Under the Curve**(AUC)**.

To graphically obtain the above function, we import the `gridExtra` library, provide functions that can be used to create additional grobs beyond those provided by the grid package. We formed a colour palette and set values to the x and y-axis to hence plot the two graphs. A confusion matrix is formed to represent fp, fn, tp, tn, tpr, fpr, tnr, tpr.

To get an insight of the data, we use the **sprintf()** function, which gives us the details of the number of rows and columns in the dataset.

Then we will create the **Random Forest** model by using a training set after splitting the data into a training and test set. For the test set, we will use the trained model to predict the Fraud/Not Fraud Class.

Given the class imbalance ratio, we will use the **Area Under the Precision-Recall Curve** to assess accuracy (AUC). The ROC AUC and Precision-Recall AUC scores summaries the curves and can be used to compare classifiers.
For unbalanced classification, confusion matrix accuracy is less usefull because it is critical to use a metric that includes FP and FN evaluation. It is critical to reduce the number of FN (Predicted: Not Fraud and True: Fraud) as much as possible because their cost could be very high.

Literature review:

Fraud detection is a complex task and there is no system that correctly predicts any transaction as fraudulent. The properties for a good fraud detection system are:
1. Should identify the frauds accurately.
2. Should detect the frauds quickly.
3. Should not classify a genuine transaction as fraud.
Outlier detection is a critical task as outliers indicate abnormal running conditions from which significant performance degradation may happen. Techniques used in fraud detection can be divided into two:
1) **Supervised techniques** where past known legitimate/fraud cases are used to build a model which will produce a suspicion score for the new transactions.
2) **Unsupervised** are those where there are no prior sets in which the state of the transactions are known to be fraud or legitimate.

SP Maniraj [1] In this paper, they describe a Random forest algorithm applicable on Find fraud detection. Random forest has two types. They describe in detail and their accuracy 91.96% and 96.77% respectively. This paper summarizes the second type is better than the first type.
Suman Arora [2] In this paper, many supervised machine learning algorithms apply on 70% training and 30% testing dataset. Random forest, stacking classifier, XGB classifier, SVM, Decision tree and KNN algorithms compare each other i.e. 94.59%, 95.27%, 94.59%, 93.24%, 90.87%, 90.54% and 94.25% respectively. Summaries of this paper, SVM has the highest ranking with 0.5360 FPR, and stacking classifier has the lowest ranking with 0.0335.
Kosemani Temitayo Hafiz [3] In this paper, they describe the flow chart of the fraud detection process. i.e. data Acquisition, data pre-processing, Exploratory data analysis and methods or algorithms are in detail. Algorithms are K- nearest neighbor (KNN), random tree and Logistic regression accuracy are 96.91%, 94.32%, 57.73% and 98.24% respectively.

References:

1. Credit Card Fraud Detection Based on Transaction Behavior -by John Richard D. Kho, Larry A. Vea published by Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017
2. L.J.P. van der Maaten and G.E. Hinton, Visualizing High Dimensional Data Using t-SNE (2014), Journal of Machine Learning Research
3. Machine Learning Group — ULB, Credit Card Fraud Detection (2018), Kaggle