Machine Learning Engineer Nanodegree

Capstone Proposal: Credit Card Fraud Detection

Mena Gabara

9.11.2019

Domain Background

Credit card fraud is a wide-ranging term for theft and fraud committed using or involving a payment card, as a fraudulent source of funds in a transaction.The purpose may be to obtain goods without paying, or to obtain unauthorized funds from an account. It is also an adjunct to identity theft. According to the United States Federal Trade Commission, while the rate of identity theft had been holding steady during the mid 2000s, it increased by 21 percent in 2008. However, credit card fraud, that crime which most people associate with ID theft, decreased as a percentage of all ID theft complaints for the sixth year in a row. Although incidences of credit card fraud are limited to about 0.1% of all card transactions, they have resulted in huge financial losses as the fraudulent transactions have been large value transactions. In 1999, out of 12 billion transactions made annually, approximately 10 million—or one out of every 1200 transactions—turned out to be fraudulent. Also, 0.04% (4 out of every 10,000) of all monthly active accounts were fraudulent.

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.

And that is where machine learning comes in handy, such behaviour could be detected and millions of dollars could be saved.

Problem Statement

The goal here is to detect whether a new transaction is fraud or not. This works by feeding the system with previous data -that has unique and common features like: time, class(fraud, not fraud), amount.. etc - to collect data which would help in accurately classifying new transaction.

Datasets and Inputs

The dataset used is the one provided by kaggle:
https://www.kaggle.com/mlg-ulb/creditcardfraud/data#

The datasets contains transactions made by credit cards in September 2013 by european cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

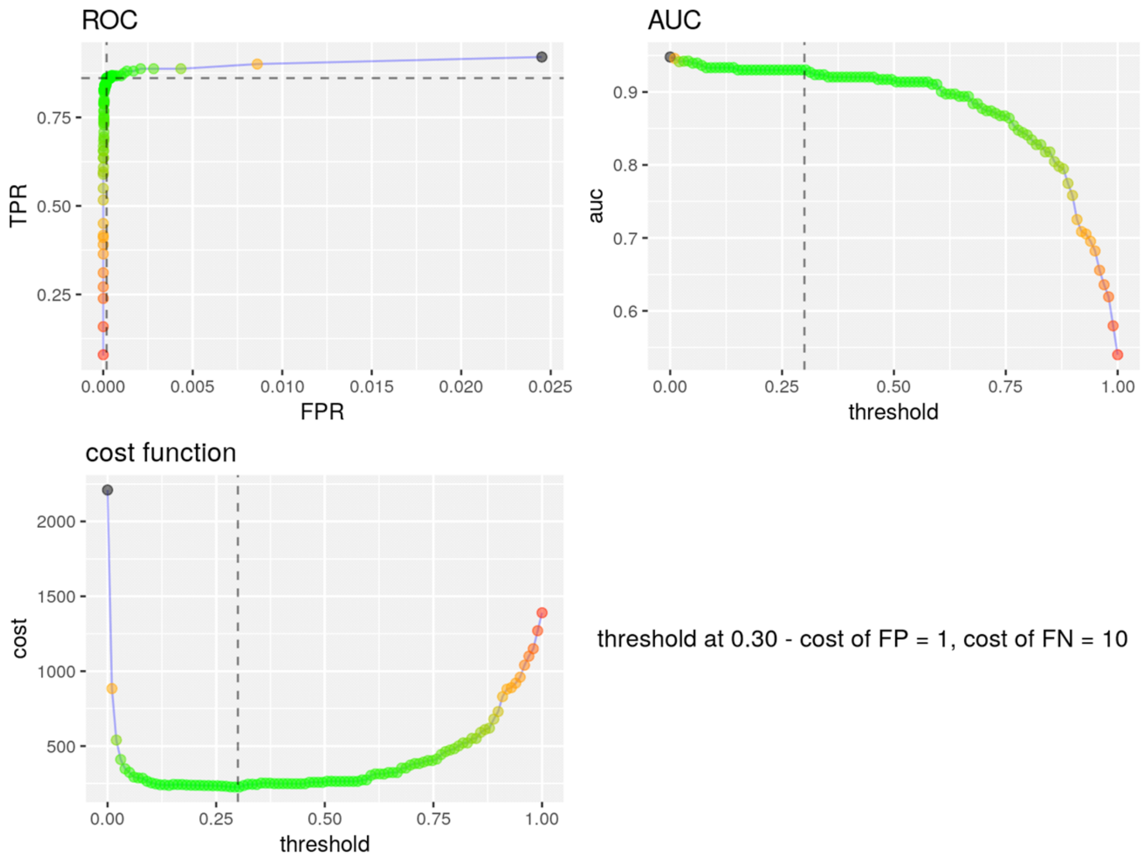
It contains only numerical input variables which are the result of a PCA transformation. Features V1, V2, ... V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependant cost-senstive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

Solution Statement

Hence the data is unblanced a way to re-balance it is needed. Also the problem is considered as a binary classification problem. So most probably a mix of supervised learning algorithms will be applied and will be selected based on their score. Algorithms like (logistic regression, random forest, decision tree, naive bayes ..etc) to produce a prediction model that will be able to reduce the false prediction and classify whether a transaction is fraud or not with the highest possible score and smallest time.

Benchmark Model

In order to be able to mesaure how accurate the model is, some benchmark models need to be taken into consideration and after searching the internet in what other have accomplished especially in kaggle. Recall, F1-score, Accuracy should not be less than 80%. and since the case is very sensetive a lower score can not be taken into consideration as a good model.



threshold at 0.30 - cost of FP = 1, cost of FN = 10

Evaluation Metrics

Hence the goal is to correctly classify fraud transactions. Classifying some fraud cases as normal seems more crutual than classifying a normal case as a fraud. so AUPRC will be used to calculate the accuracy of the score. and the closer the score to 1 the more accurate the model is. the confusion matrix will help in comparing between the implemented model and the benchmarket model.

TP = True Positive. Fraud transactions are predicted as fraud.

TN = True Negative. Normal transactions are predictd as normal.

FP = False Positive. Normal transactions are predicted as fraud.

FN = False Negative. Fraud transactions are predicted as normal.

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$F1Score = 2 * \frac{Recall * Precision}{Recall + Precision}$$

Project Design

- Load the data from Kaggle.
- Exploratory Data Analysis and remove outliers.
- Data Pre-processing if any required like trying to make the data more balanced.
- Build different supervised learning models.
- Compare the implemented models and evaluate the score and select the best one and compare with the benchmarket model.