Machine Learning Engineer Nanodegree

Capstone Proposal: Credit Card Fraud Detection Model

Femi Ogunbode

September 16, 2017

Domain Background

Credit card fraud is the use of a payment card, either debit or credit obtained fraudulently or illegally to carry out fraudulent transactions.

Credit card fraud is a problem that has been plaguing financial industries for years, for example in my country Nigeria, ATM's (Debit Card) had the highest percentage of fraud volume in 2016 with 49.6% which accounts for half of all fraudulent transactions on electronic payment channels.[1]

Solving this problem could bring about reduced volume of fraudulent transactions in financial industries hereby saving them a huge volume of money lost and this problem can be solved using cutting edge predictive analytics where machine learning algorithms are used to detect fraud patterns and determine future probabilities and trends.

Problem Statement

Given a transaction, it is often difficult know which is fraudulent or genuine. For example given a transaction with some features such as the time, location, amount, occupation of the card holder etc. among many other features there is a need to accurately classify which transactions have a very high probability of being fraudulent.

The problem before us is a binary classification problem, because the problem arises from deciding whether or not a transaction is fraudulent.

[1] The Nigeria Electronic Fraud Forum "A Changing Payments Ecosystem: The Security Challenge" *Annual Report, Pg. 16* (2016)

Datasets and Inputs

The dataset to be used in solving this problem is an anonymized set of credit card transactions labelled as fraudulent or genuine. The dataset has been collected and analysed during a research collaboration of Worldline and the Machine Learning Group (http://mlg.ulb.ac.be) of ULB (Université Libre de Bruxelles) on big data mining and fraud detection and has been made public on Kaggle.

This dataset contains transactions made by credit cards in September 2013 by European cardholders. It 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.

The dataset contains only numerical input variables which are the result of a PCA transformation. Due to confidentiality issues, the original features and more background information about the data were not provided. 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. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

Solution Statement

To solve this problem, a prediction model would be developed using several machine learning algorithms on this dataset to see which will give the best performance which is judged against a pre-defined evaluation metric. These algorithms include Naïve Bayes Classifier, Logistic Regression, Multi-Layer Perceptron and Decision Trees.

Since the features made available are a reduced dimension of the original dataset, no other form of reduction would be made going forward. All features provided would be used in building the model.

Benchmark Model

For this project, considering the imbalance class ratio accuracy would not be used to judge the model since it can be misleading. Instead, as a benchmark the model should have an Area Under the Precision-Recall Curve (AUPRC) score of 74% or greater.
This score was gotten from a simple logistic regression classifier I built, this serves as benchmark towards building a better model. The intuition behind using the AUPRC Score is explained in the next section.

Evaluation Metrics

Given the class imbalance ratio, I will measure the accuracy using the Area Under the Precision-Recall Curve (AUPRC) as opposed to accuracy.

The Area Under the Precision-Recall Curve is gotten by plotting the Precision against Recall.[2]

Precision and Recall are defined as follows:

$$\text{Precision} = \frac{TP}{TP+FP} \qquad \text{Recall} = \frac{TP}{TP+FN}$$

- True Positive (TP) – An example where a transaction is fraudulent and is classified correctly as fraudulent.
- False Positive (FP) – An example where a transaction is genuine and is classified incorrectly as fraudulent.
- False Negative (FN) – An example that is genuine but is classified incorrectly as fraudulent.

Since Precision & Recall don't account for true negatives (cases where a transaction is fraudulent but classified as genuine by the model) the choice for using the Area Under the Precision-Recall Curve (AUPRC) seems reasonable given that there are lot more cases of genuine transactions than fraudulent.[3]

The closer to 1 the AUPRC is, the better the model is. A model with AUPRC score of 1 implies a perfect classifier.

[2] Jesse Davis & Mark Goadrich "The Relationship Between Precision-Recall and ROC Curves" *Proceedings of the 23rd International Conference on Machine Learning (ICML) Pg. 2.* (2006)

[3] Jesse Davis & Mark Goadrich "The Relationship Between Precision-Recall and ROC Curves" *Proceedings of the 23rd International Conference on Machine Learning (ICML).* (2006)

Project Design

First of all, descriptive statistics of the dataset would be calculated to have a basic understanding of the distribution and structure of the dataset.

Given the dataset, exploratory data analysis would be carried out to understand how these variables contribute to the outcome, though this approach might be limited and much intuition cant' be drawn since the features aren't original and the actual names of the features or what it represents is hidden.

Outlier detection and removal is up next, this phase seeks to find all points/observations that are irregular or abnormal relative to the dataset.

Next, since all features are numerical, pre-processing steps such as standardization would be applied on the dataset to bring all features on the same scale and make it have a mean and variance of 0 & 1 respectively.

Afterwards, considering the imbalance in the dataset the SMOTE algorithm would be used to even out the imbalance by constructing new points in the minority class (fraudulent cases).
Different binary classification algorithms such as Naïve Bayes, Logistic Regression, Multi-Layer Perceptron and Decision Trees would be applied on the dataset to see which would give the best performance which is defined by the evaluation metric.

After training and validation of different algorithms to create a model. The best performing model is finally selected and optimized using techniques like grid search for hyper-parameter tuning, which is then applied to the test set to predict which transactions are fraudulent or otherwise.