Credit Card Fraud Detection: In this project, we are training classifier to classify transactions as normal or anomaly. It is highly unbalanced data with very low fraction of one class.

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.

Dataset can be downloaded from this link:

https://www.kaggle.com/mlg-ulb/creditcardfraud

We will use following two codes:

1. Semisupervised methods: In his first code, we will use some of the semisupervised methods like fitting a Gaussian distribution or multivariate normal distribution to one class (normal transactions). Then we will classify all datapoints based on the probability of each datapoint coming from this distribution.

2. Undersampling method: In this code, we will try under sampling technique to classify unbalanced data. We will also tune logistic regression model to come up with best classifier.