

Thesis topic Proposal

Application of Machine Learning in Credit Card Fraud Detection

Credit card fraud is a growing problem in the finance world. However, to detect a fraud is not a such an easy task to perform. In order to tackle this problem, we would like to propose topic of **Application of Machine Learning in Credit Card Fraud Detection**. For this topic, a dataset of transactions made by credit cards in September 2013 by European cardholders published on Kaggle, which is highly imbalanced, has been chosen to be experimented on. There are only 492 out of 284807 transactions are labelled as fraudulence. More specifically, we would like to look into how to process and deal with imbalance data using different techniques such as undersampling, oversampling and hybrid sampling. Furthermore, different predictive models such as logistic regression, random forest or boosting along with these processing techniques would be used to fit the dataset and produce prediction whether a transaction is genuine or fraudulent. Combinations of predictive models and sampling techniques would be evaluated using a variety of evaluation metrics such as: precision, recall, f1 score, precision-recall (PR) curve and receiver operating characteristics (ROC) curve. The main objective of this research is to look into multiple approaches in order to solve imbalance datasets and to figure out the best technique and model for the specific problem of credit card fraud.