



COMP5200M Project Specification

Student Name: Shengxuan Ji
Programme of Study: Advanced Computer Science (Data Analytics)
Supervisor Name: David Head
Name of External Company (if any):
Type of Project: Exploratory Software
Provisional Title of Project: Credit Card Fraud Detection
Aim of Project: <ol style="list-style-type: none"> 1. Identify the variables with the highest predictive power for fraudulence, that is, the variables with the most substantial impact on the outcome. 2. Compare various models and determine the model with the highest accuracy, as well as the underlying reasons for its superior performance.

Objectives:

1. Perform data pre-processing include SMOTE (Synthetic Minority Oversampling Technique) on credit card transactions data and carry out a visualization analysis and Z-test to detect any potential patterns and determine the most effective variables.
2. Utilize both supervised and unsupervised machine learning algorithms for detecting transactional fraud, including logistic regression, random forest, SVM (Support Vector Machine), Isolation Forest, and Local Outlier Factor. If time permits, the Naive Bayes Classifier and Autoencoder will also be utilized.
3. Conduct multiple performance evaluation metrics, including accuracy, precision, recall, and ROC curve analysis, to assess the effectiveness of the models, followed by a detailed analysis of the underlying factors contributing to the superior performance of the selected models. During project implementation, if a better approach is discovered, it will also be adopted.
4. Present a summary of the key points related to Anomaly Detection. Highlight the most important takeaways from the analysis. Finally, propose improvement measures.

Deliverables:

1. A report
2. Code for the whole project.