Executive Summary

As credit card fraudulent transactions are on the rise, it is important that credit card companies can recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase. This project intends to create a model for credit card fraud detection using past credit card transactions with the knowledge of the ones that turned out to be a fraud.

Our approach follows the CRISP-DM methodology where we have gathered the necessary data and cleanup/preparation was done on the data as per our needs. Based on graph analysis, transaction time does not have any influence on prediction of fraudulent transaction. Our input dataset contains only numerical input variables which are the result of a PCA Dimensionality reduction to protect user identities and sensitive features. Our dataset is severely imbalanced as most of the transactions are non-fraudulent. We are implementing oversampling technique called SMOTE to handle our imbalanced dataset. Many methodologies and models have been applied on our test data to identify transactions that are likely to be fraudulent. Upon evaluating the accuracy and performance of various models, Random Forest classification model proved to be the best fit to identify fraudulent transactions.