Method Overview

Preprocessing
First we pre-process the mbti and Yelp datasets by scaling, normalising, feature engineering, vectorisation and then creating a bag of words

Predictive Modelling on MBTI
We then apply logistical regression, SVM, Random Forest and Recurrent Neural Network to mbti dataset

Applying Models to Yelp
Then based on accuracy and using cross validation, the best performing model was used on the yelp dataset

Analysis of Yelp Reviewers
The data was then analysed to determine the distrubution of personality types.