For our training data, we have opted to trim the overall dataset provided to remove the candidates who lost in primaries and received less than 10% of the general election vote to help remove noise from candidates who were not truly ‘in the running’ for the general election. This trimming step cut down our candidates across the 5 elections from 13353 to 5116.

For our selection of features we ranked the available features of the class using domain knowledge

Interesting features:

Our first feature selected was Party\_Previous\_Vote\_Share

Incumbent

Raised

Spent

recent\_presidential\_vote