CS 5002 - Senior Design II
Test Plan and Results

Overall Test Plan

We plan to test our application by allocating some data points in the master dataset as test data points. To get the best results we choose the data points randomly in different ratios multiple times and we plan to do this in 5 phases. We used a Kaggle dataset that has about 500 politicians with their classes - democrats and republicans. Then another dataset was used that didn't have anyone from these 500 politicians to test the model. In each phase a confusion matrix was used to compare the TP(True Positive), TN(True Negative), FP(False Positive), FN(False Negative) values

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

+ 1 value means perfect prediction
- 1 value means completely opposite prediction
0 No better than the random prediction

Phase I Testing

In this phase we divided the data set into 70% - Train and 30% - Test ratio. We then used Mathews correlation coefficient as a metric.

	1	0
1	490	0
0	0	10

This gave us 100% accuracy and +1 MCC value

Phase II Testing

In this phase we divided the data set into 75% - Train and 25% - Test ratio. We then used Mathews correlation coefficient as a metric.

	1	0
1	470	1
0	0	29

This gave us 99.8% accuracy and +0.9 MCC value

Phase III Testing

In this phase we divided the data set into 80% - Train and 20% - Test ratio. We then used Mathews correlation coefficient as a metric.

	1	0
1	450	15
0	15	20

This gave us 93% accuracy and +0.8 MCC value

Phase IV Testing

In this phase we divided the data set into 85% - Train and 15% - Test ratio. We then used Mathews correlation coefficient as a metric.

	1	0
1	492	0
0	0	8

This gave us 100% accuracy and +1 MCC value

Phase V Testing

In this phase we divided the data set into 90% - Train and 10% - Test ratio. We then used Mathews correlation coefficient as a metric.

	1	0
1	490	0
0	0	10

This gave us 100% accuracy and +1 MCC value