

For the midterm project, I decided to create a SVM model using LIBSVM on Java. To gain the features, I extracted the 2 sentences and created a class that accepted the 2 strings and created several values between 0 and 1. The features I made are:

- Difference of number of words.
- Whether there is a quote.
- How many numbers are different.
- How many words are different.
- Difference of average length of the words.

To train the data, I saw that there was a 3:1 ratio on the class in the training data. I decided to create 3 models in order to have a 50/50 split for training the model.