

Jewel Chen
Statistical Learning
Final Project

My final approach model was a KNN model. I had to initially normalize the data set, then run cross validation to determine the optimal k value. Because the data set was so large, cross validation occurred over a small, randomly selected subset of the data. After the optimal k value was determined, the model was applied to the entire data set. The results were then unnormalized and written to Excel using `xlsx`.

I initially attempted to code a k -means model, since the k is already known. Had there been more time / guidance, I think this method would have worked better. I also attempted to run LASSO on the variables to see if removing certain features would make the model more powerful. Ultimately, however, removing variables did not do much to the model as a whole.