# Weekly Report 6

Jeremy Lachowicz

2/26/2022

## 1 Accomplished this week

I ran a large grid search on the Word2Vec models to discover which hyperparamters work the best. I used two measures to test this. For the first measure, I iterated through every word of the test document, and then checked if the other documents contained that word. If they did, I found the cosine similarity of that vector with the test model's vector. Then, I calculated the mean of the list. This is a simple measure, so it doesn't inform much; however it produced great results: **Longus: 0.07 Mark: 0.42 Mathew: 0.10 Ignatius: 0.9965**. These are the results from the best grid search model (the test model contains other works of Ignatius, so it was accurate). For the other measure I used, I iterated through every word of the test document, and then checked if the other documents contained that word. If they did, I calculated the cosine similarity of that vector with the test model's vector. From here I counted which author had the largest percentage of similar word vectors to the test document. This measure was fairly accurate, and the best model from the grid search returned these results: **Ignatius: 81.538% Mark: 15.385% Longus: 1.538% Mathew: 1.538%**. Again, this shows the vast majority of words in the document were the most similar to Ignatius, which is correct.

## 2 Accomplish next week

I want to create and code better measures, and tinker more with other preprocessing steps to see if I can further improve the models. Also, I will start the coding for Doc2Vec models.