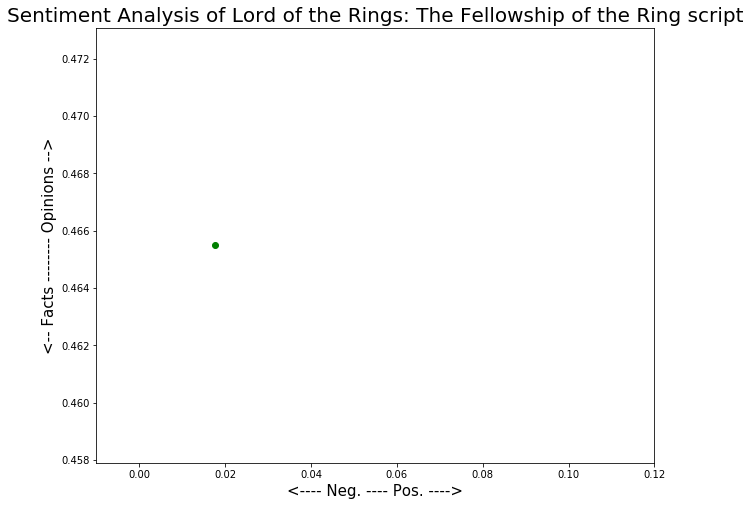
For my original project plan, I wanted to do a text analysis, a sentiment analysis and a topic modeling of the Lord of the Rings: The Fellowship of the Ring movie script and a few individual characters. My goal was to break down the movie script and see how well it matches up with my perception of the characters because I just recently finished the book. I wanted to do a text analysis to review the entire script to tokenize words and phrases and examine lexical density, sentence length, and vocabulary richness. Additionally, I wanted to do the same to several characters who speak the most throughout the script. To run these tests, I wanted use Natural Language Toolkit (nklt). Using TextBlob, I wanted to run a sentiment analysis on the characters to determine who is more positive and negative. Finally, I wanted to do a topic modeling of the script, utilizing genism as my library. Through topic modeling, I hoped to identify the most prevalent themes and their associated words/phrases, along with identifying which character is associated with which theme in the script. Unfortunately, due to some technical difficulties I will cover at the end of the analysis, I was not able to do the individual character analysis.

After I imported my script into a file and cleaned it, I was able to run a few tests to see what words occurred most frequently within the script and these were my results:

|  |  |  |  |
| --- | --- | --- | --- |
| Frodo | 386 | Pippin | 101 |
| Gandalf | 273 | **Merry** | 100 |
| Ring | 150 | **Legolas** | 65 |
| Aragorn | 144 | **Gimli** | 56 |
| Sam | 138 | **Hobbits** | 53 |
| Boromir | 116 | **Saruman** | 52 |
| Bilbo | 104 | **Strider** | 47 |

As I expected, Frodo is the most frequently seen word in the movie script, followed by Gandalf and the Ring. Frodo is the focal point of the entire story, much of world building is done by Gandalf, and the Ring is what brings the characters together, for better or for worse. It also must be noted that “Strider” is an alias for Aragorn used at the beginning of the book, technically making him the third most occurring word or character in the book. The most surprising part about the analysis is that Boromir, a prominent, but not main character, is the sixth highest occurring word in the script. I find this surprising because in the books he does not appear to occur nearly as much as he does in the script, whereas it appears Legolas or Gimli play a more integral role to driving the story line. The only other noteworthy part of the analysis is that Pippin and Merry almost occur the same amount of times, which is not surprising because they are essentially twins.

The next part of my analysis was a sentiment analysis of the script, which measured whether the script used more factual or more opinionated language on a scale from 0, least opinionated, to 1, most opinionated. Additionally, the script was measured on a scale from -1 to 1 as to how negative or positive it was, with -1 being the most negative and 1 being the most positive. Here are my results:

Regarding the polarity of the scale, it seems as if the script leans more towards words that are factual rather than opinionated. While this may not tell us much, I would hypothesize that if we did sentiment analysis of individual characters, we would see difference. For example, since Gandalf does much of the story building, I would say that he would be closer to zero on the scale, as he gives more information than opinion. Regarding the polarity, it seems to be neutral, with a very slight tendency to be positive. Again, I would hypothesize that this might change when doing individual character analysis, as someone like Frodo would be more negative, and someone like Sam would be more positive given my perception of the book.

For my final analysis, I took the genism module and decided to run Latent Dirichlet allocation (LDA), a type of topic modeling associated with genism. I imported my document term matrix, created both a corpus and dictionary of terms, and then ran my analysis. Since I was only looking through one script, I decided that only one topic would be necessary to analyze. Here was my result:

**[(0: '0.032\*"frodo" + 0.023\*"gandalf" + 0.013\*"ring" + 0.012\*"aragorn" + 0.012\*"sam" + 0.010\*"boromir" + 0.009\*"bilbo" + 0.009\*"pippin" + 0.008\*"merry" + 0.006\*"legolas"')]**

As expected, the theme of the book is about these several characters that go on an adventure to destroy the ring, hence “The Fellowship of the Ring”. If you were to run the analysis again, you would find the same names, just in different order.

Overall, much of what I found was not extremely surprising to me and matched with what my perception was of the book. The only issue I had was running the individual character analysis, as I could not find a way to successfully import the text files into Jupyter Notebook without an error. I think the problem stemmed from the fact that I was copying and pasting the lines from the notebook into the text document and that somehow affected my analysis. I believe that this problem is not capable of being solved through my knowledge of coding right now, but I will attach what code I did have to show what I was attempting to do originally. Additionally, I will attach the text document of each individual character.