**Overall Instructions**

* The final document should be a report in text or PDF. You should separately turn in the final analysis code/compiled report. Be sure to spell and grammar check your work! For summary sections, do not simply copy your answers from the assignment. Write a coherent report of the analyses you did and the results you found. The following sections should be included (and make sure each section is clearly labeled). Be sure to write in narrative style, as if you were creating a report summary for other readers.

**Introduction**

* Introduce your text source. Where did you get the text you are using for this project? How big is the text source? Give a brief description of the summary of the text in your own words.[[1]](#footnote-1) Explain why you were interested in analyzing this set of text (1-2 paragraphs).

**Processing Text**

* Summarize your results from the raw text and part of speech analysis section. What parts of the analysis did you struggle with? Do you have suggestions on better ways to implement what you did? What type of normalization procedure should one use for an NLP analysis? Did you have any surprises on the most common words or other parts of the analysis? (2 paragraphs)

**Information Extraction**

* Summarize your results from information extraction section. What kinds of results did you see from spacy for entities? Are there obvious things missing that you might consider training your own model? What else might you do with these techniques? Consider that information extraction is one of the hardest NLP tasks. What could we do to make it easier? (2 paragraphs)

**Text Summarization**

* Summarize the results from the text summarization section. Do you feel that the bag of words methods are useful for text summarization? Can you make a good summary from grabbing sentences from the text or should we explore more of text generation? Do you think enterprise search engines are useful for your work? What other things might you do with the information from text summarization? (2 paragraphs)

**Classification**

* Summarize your results from two sections on classification. What are other things you might considering classifying from your text? Explain what you think the best options are for feature extraction and algorithms. Do you think the systems to explain machine learning models are useful? What else might we do to help understand our models so we can fix them when they go wrong? (2 paragraphs)

**Your Own Analysis**

* **Be sure to turn in the code and compiled document you used to write this summary.**
* Include a description of the analysis, the goals of the analysis, and what you found in your results. Why were you interested in running this analysis? (2 paragraphs)

**Wrap Up**

* Include a conclusion paragraph about the usefulness of NLP techniques to analyze text. Do we need complex models to transform and understand text? (1 paragraph)

1. Note that these documents are checked for copying from internet sources. Please use your own words so none of us must deal with the plagiarism reports at the end of the semester. [↑](#footnote-ref-1)