CS 410 Progress Report

Gabe Coelho

gp17@illinois.edu

Which tasks have been completed?

- I identified the main dataset to be used in the analysis.
- Reviewed the data in the dataset to ensure it contains all the necessary information for the analysis.
- Preprocessed the text data, to clean out any special characters, emoji, stop words, etc. and normalized most of the values (this dataset was particularly clean, which eased the load on the preprocessing).
- Identified the main text mining algorithms to be used.

Which tasks are pending?

- Identify the highest word count in positive reviews.
- Identify the highest word count in negative reviews.
- Generate a bag of words representation.
- Train, build, evaluate the model multiple times over.
- Predict the sentiment as the final step.

Are you facing any challenges?

- Because I am in the initial phase, I haven't encountered as many challenges, but I am expecting to face more difficulty in the model building and training phases, since I have never written such a program before.
- I am not fully aware of libraries that could be helpful for sentiment analysis-related tasks, so this is something I need to research and ponder.