

1. Basically I just tried out patterns that I found from the EDA stage. After trying each pattern individually, I then combined them to see how they work together, and from that I was able to get a pretty good model.
2. The thing that worked the best were the feature words, which contributed a lot to the model. Other features such as the number of exclamation points and word count didn't work as well, but they still add some accuracy to the model.
3. I was surprised at a few words that favored ham/spam emails. For instance, "marketing" for spam and "spam" for ham. Another thing I was a bit surprised at is how the number of question marks aren't that different between spam and ham emails.

