1. For a significant portion of the time I spent on selecting features for my model, I was focused on utilizing the "words_in_texts" method and hoping to come up with the five most distinctive words that could distinguish spam mails from ham mails. I came up with words like 're:' in the subject headings and matching 'html' tags in the email bodies. I ultimately used the tf-idf method to obtain better features for my model.
2. I wanted to find more indicative words by closely inspecting emails and subjects in my training set, hoping to come up with very strong indicator words. I realized that I was overfitting to my training set - obtaining around 89% accuracy on my training set, but a disastrous 76% on my first submission to Kaggle. After trying out the tf-idf method and seeing a high accuracy yield, I cleaned the data a bit more by parsing html commands and produce readable text so tf-idf could be implemented on pure text.
3. I was surprised that what seemed pretty indicative words to me didn't prove to be that effective at all when predicting the test set. I was also surprised how quickly I've fallen into the trap of overfitting my data. I am surprised that I did spend the most time understanding my data, doing data cleaning, and running into problems when there are missing values in my data set that prevented me from applying functions.