IMPROVEMENTS TO THE SPAM DETECTOR 3000

The spam detector will take any word into consideration and test the probability of the word. The problem with this is this method will also include parts of speech that are unnecessary; such as pronouns, conjunction words, article words, etc. Because these words are used in every day speech and text, calculating the probability of these words being spam or not will add more or less probability to the file/email being spam. If this method includes a bag of words that will exclude all these words then the accuracy of the probability will be much better.

Another improvement would be recognizing letter patterns in a word. If the word contains a majority of a letter such as “AAAAAAAAAAAAA” or “IiIiIjJjJjJjJ” then it can instantly be counted as spam since those words do not exist in the English dictionary.