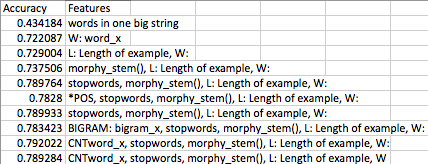
My username in Kaggle is Anas My approach with this problem of feature collection was to get the basic information first and test and record my results. After that, add more features that are not so obvious but can collect like the length of the example. The interesting thing was that adding bigram information did not help the accuracy. Bellow is a graph showing my attempts:



W: word\_x: all the words starting with a “W:

L: Length of example

Morphy\_stem(): a function that stems the word

Stopwords: removed stopwords

BIGRAM: added bigram information with a the mentioned tag

CNTword\_x: all words counts information