

In this program I have used the simplest algorithm of Naive Bayes to classify the data.

Steps:

1. Converted all the input into lower case.
2. Removed special characters.
3. Created a vocabulary of both negative and positive data set.
4. Used Laplace to smooth (both Negative and Positive data set).

While testing:

4. Ignored the words which are not in the training vocabulary.

How the code works?

Tagged the data based on maximum probability between positive and negative classifiers.

Included all the words in both the word dictionaries so that the word count in both dictionaries are equal, using Laplace have given the probabilities of each word dictionaries by smoothing them.

Used Naïve Bayes to classify the data.

Finally, I used K Fold to test the data, but the data was too small to test it.