Part 3: Reflection

Below are the three sentences on which my classifier failed

Sentence	Expected	Actual Result
I didn't like the movie	Negative	Positive
A dirty movie	Negative	Neutral
What a horrible movie	Negative	Neutral

I observed that the result of the classifier mainly depends on the words present in the positive and negative dictionaries which were formed while training.

In the above example, I suppose the words 'didn't', 'dirty' and 'horrible' were not present in both positive and negative dictionaries. Hence, we got the wrong result.

The accuracy of the model also depends on the number of words in pickle file. If the pickle file has more positive and negative words, the accuracy improves.

Part 5. Evaluation

The dataset contains around 43,914 files. Out of which I've trained the models bayes and improved bayes with 80% data i.e., around 38,928 files. The testing was done on the remaining 20% data i.e., 4,986 files.

------------------------Running the Bayes Classifier-----------------

Results Summary:
negative: 669
neutral: 305
positive: 4012
Classification Accuracy: 87.24%
Classification Precision: 94.19%
Classification Recall: 90.35%
Classification F-measure: 92.23%

------------------------Bayes Classifier End--------------------------

------------------------Running Best Bayes Classifier---------------

Best Results Summary:
negative: 1151
neutral: 402
positive: 3433
Best Classification Accuracy: 82.29%

Best Classification Precision: 83.35%
Best Classification Recall: 93.93%
Best Classification F-measure: 88.32%

-------------------------Best Bayes Classifier End-------------------

Comparison of performances of two classifiers:

    1> Out of 4,986 reviews, the bayes classifier achieved 87% accuracy where as best bayes has achieved 82% accuracy.

    2> The Precision obtained by bayes is 94% whereas for the best bayes it has dropped to 83%. I suppose the reason for the reduction is due to consideration of unigrams, bigrams.

    3> The Recall for the best bayes is 3% more than the bayes which is 90%. This is due because the model has got more words in positive and negative review dictionary.

    4> Finally, the F-measure of the bayes is 92% whereas the best bayes is 88%.

    5> The improvement for the bayesbest.py is mainly because of exclusion of words less than 3 letters and also use of bigrams.

Ways to improve the classifier:

Some other ways to improve the classifier is use of punctuations, amount of capitalization, length of document and inculcating reinforced learning such as learning from false positives, false negatives of the test data.