Report – Assignment 3

Student ID: 3718042 (Madhura Gaikwad)

In this assignment, I compared and implemented various approaches for Sentiment Analysis.

After implementation, I found that Naïve Binary Bayes perform better than (vanilla) Naïve Bayes. And Naïve Bayes perform better than Sentiment Lexicon.

To answer the questions:

1. Most-Frequent baseline class performs better than sentiment lexicon.
2. Yes, Naïve Bayes and Naïve Binary Bayes outperform Baseline methods.
3. Yes, naive Bayes with binary features outperform (vanilla) naive Bayes. And (vanilla) Naïve Bayes is second best.
4. Naïve Binary Bayes (NBBIN) is best than Naïve Bayes (NB).
   After comparing the outputs in NB & NBBIN with dev.classes.txt the following observations are made:
   - In NB.output.txt, Some of the ‘positive’ and ‘neutral’ documents were wrongly computed as ‘negative’ and some ‘negative’ documents were wrongly computed as ‘positive’.
   - Similarly in NBBIN.output.txt, Some of the ‘positive’ and ‘neutral’ documents were wrongly computed as ‘negative’ and some ‘negative’ documents were wrongly computed as ‘positive’.
   - For predicting negative classes, NB has performed better than NBBIN in this. NBBIN has predicted most of them as ‘positive’ (Ex Documents : 32, 40, 143)
   - For predicting neutral classes, NBBIN has performed better than NB in this. NB has predicted most of them as ‘positive’ (Ex Documents : 205, 108)
   - The Precision, Recall and F Score for ‘positive’, ‘negative’ and ‘neutral’ classes are slightly higher for NBBIN than NB.
   - If we skip the computation for ‘neutral’ classes, surprisingly we can see a downward performance and sentiment lexicon performs better than NB and NB performs better than NBBIN.
   - Also, in the current dataset, if all the documents are computed as ‘negative’, Naïve Bayes gives 67% of accuracy, so even if the output is completely unacceptable with all documents being classified as ‘negative’, it still performs better than sentiment lexicon.