**BITS F464-MACHINE LEARNING**

**Assignment-3**

Naïve Bayes Classifier

By

Yashodhar Meduri (2015A7PS0060H)

Tarun Raheja (2015A7PS0106H)

Ahraz Rizvi (2015A7PS0012H)

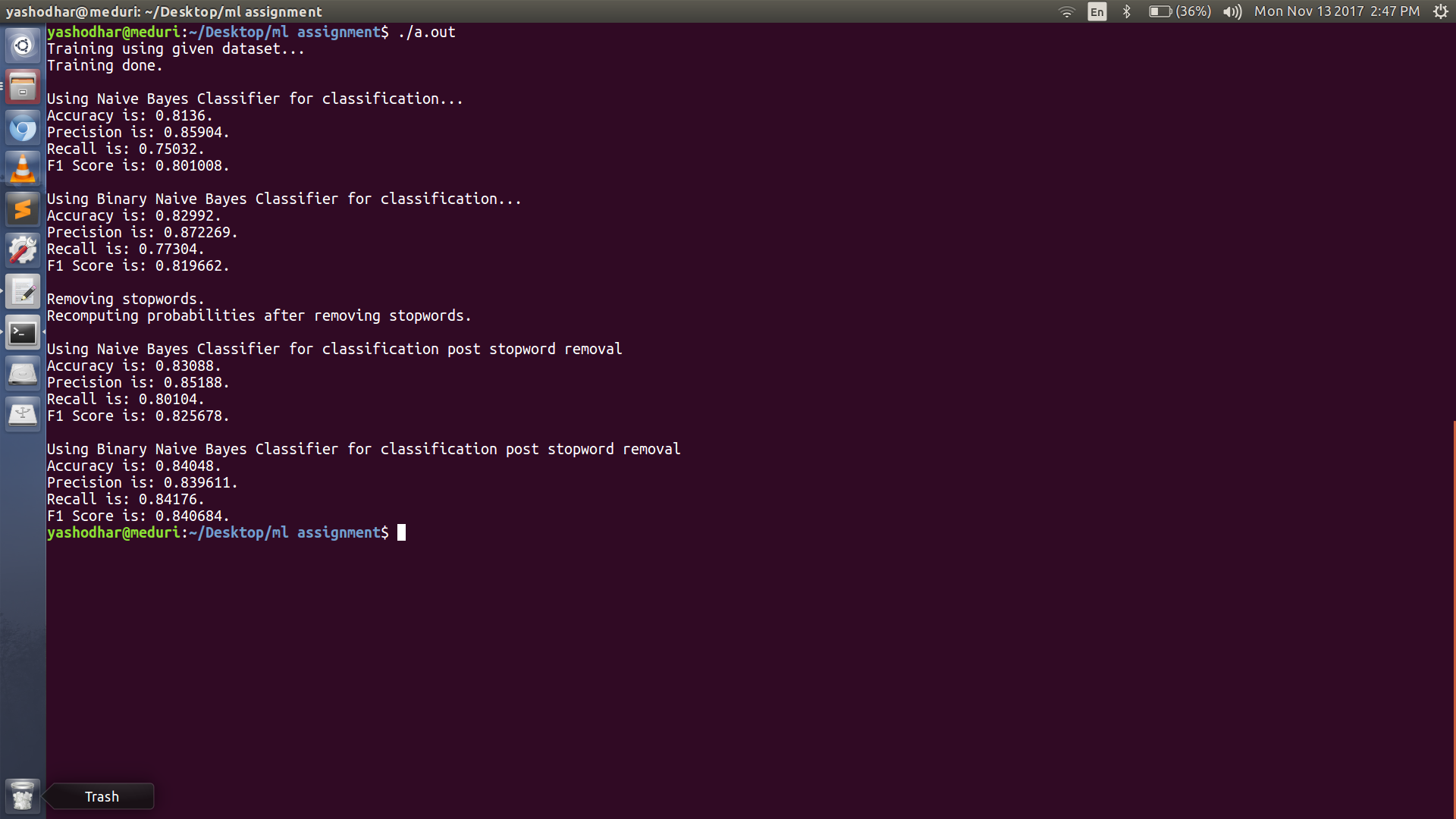
In the assignment, we did the sentiment analysis using Naive Bayes Classifier. We implemented it in Cpp for movie reviews dataset.We extended it to Binary Naive Bayes classifier and calculated the precision and recall.

Later on, we removed the stop words and implemented the above mentioned processes again.

The following were the results and observation:

RESULTS:

|  |  |  |  |
| --- | --- | --- | --- |
| **Classifier** | **PRECISION** | **RECALL** | **F1 MEASURE** |
| Naïve Bayes | 0.85904 | 0.75032 | 0.801008 |
| Binary Naïve Bayes | 0.872269 | 0.77304 | 0.819662 |
| Naïve Bayes after removing stop words | 0.85188 | 0.80104 | 0.825678 |
| Binary Naïve Bayes after removing stop words | 0.839611 | 0.84176 | 0.840684 |



From the above table we observe,

There is usually a trade-off between sensitivity and specificity (or recall and precision). Intuitively, if you cast a wider net, you will detect more relevant documents/positive cases (higher sensitivity/recall) but you will also get more false alarms (lower specificity and lower precision). So when we remove stop words we tend to catch more documents, thereby increasing recall. Also at the same time we get a lot more number of false alarms, which leads to decrease in precision.