* Top 230 each class (OUR parameter)
  + After removing duplicates, feature vector size=343

Running SVM...Total accuracy =93.38917817523206%

Running Decision Tree...Total accuracy =87.52546977586597%

Running NB...Total accuracy =93.45709757754132%

* Top 500 each class
  + After removing duplicates, feature vector size=728

Running SVM...Total accuracy =93.43445777677157%

Running Decision Tree...Total accuracy =88.65745981435363%

Running NB...Total accuracy =94.70228661987774%

* Top 1000 each class
  + After removing duplicates, feature vector size=1445

Running SVM...Total accuracy =92.86846275752774%

Running Decision Tree...Total accuracy =88.54426081050487%

Running NB...Total accuracy =96.40027167760924%

* Top 2000 each class
  + After removing duplicates, feature vector size=2848

Running SVM...Total accuracy =90.8988000905592%

Running Decision Tree...Total accuracy =88.8612180212814%

Running NB...Total accuracy =97.17002490378086%

* Top 3000 each class
  + After removing duplicates, feature vector size=

Running SVM...Total accuracy =89.13289563051845%

Running Decision Tree...Total accuracy =89.08761602897894%

Running NB...Total accuracy =97.44170251301789%

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Top X words from class | Feature vector size after removing duplicates | SVM | Decision Tree | NB |
| 230 | 343 | 93.39 | 87.53 | 93.46 |
| 500 | 728 | 93.43 | 88.66 | 94.7 |
| 1000 | 1445 | 92.87 | 88.54 | 96.4 |
| 2000 | 2848 | 90.9 | 88.86 | 97.17 |
| 3000 | 4275 | 89.13 | 89.09 | 97.44 |