Summary

|  |  |  |  |
| --- | --- | --- | --- |
|  | Naïve Bayes | Decision Tree | KNN |
| Stemmer | 62.14 | 56.88 | 77.93 |
| Stemmer + bigrams/trigram | 63.07 | 56.05 | 74.61 |
| Stemmer + bigrams/trigram + stopword removal | 63.07 | 56.05 | 74.61 |
| No Stemmer | 62.14 | 56.88 | 77.93 |

SVM

No data. Ran 20 minutes before aborting.

# 1

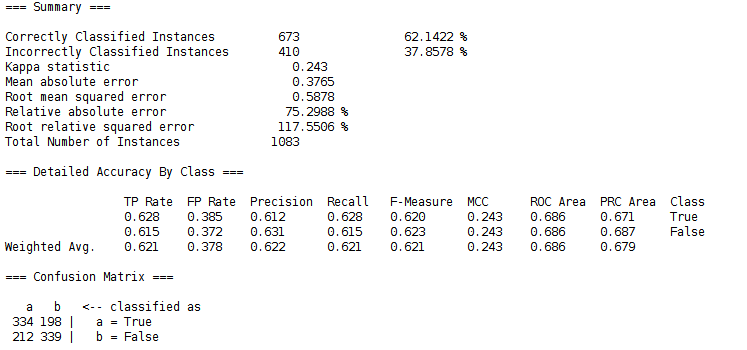
StringToWordVector Settings  
IDFTransform = True

TFTransform = True

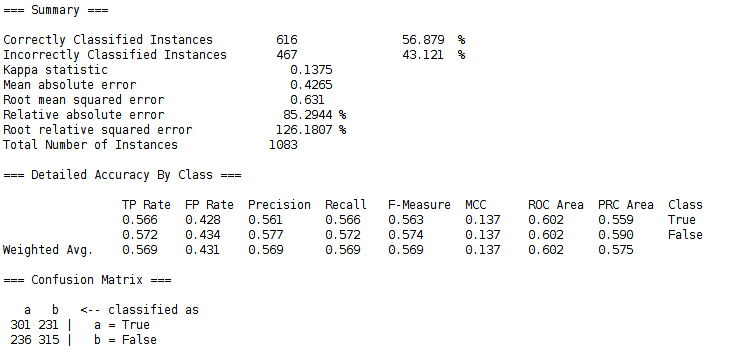
wordsToKeep = 1000

stemmer = SnowballStemmer

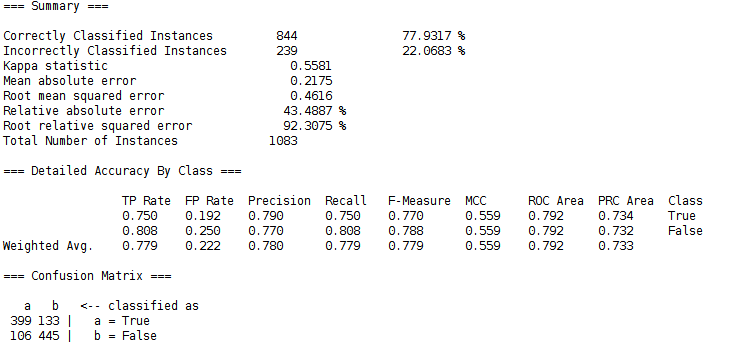
## Naïve Bayes



## J48 (C4.5)



## IBk (K-nearest neighbor)



# 2

StringToWordVector Settings  
IDFTransform = True

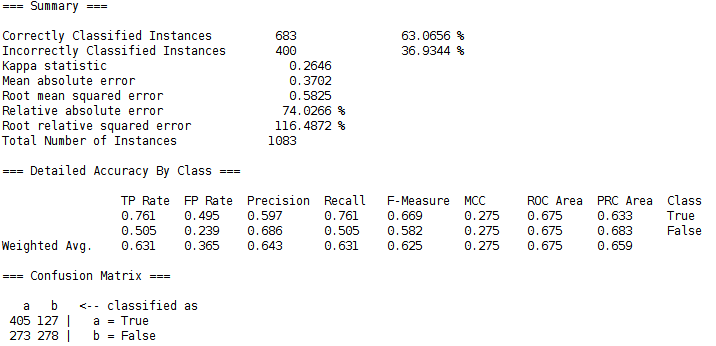
TFTransform = True

wordsToKeep = 1000

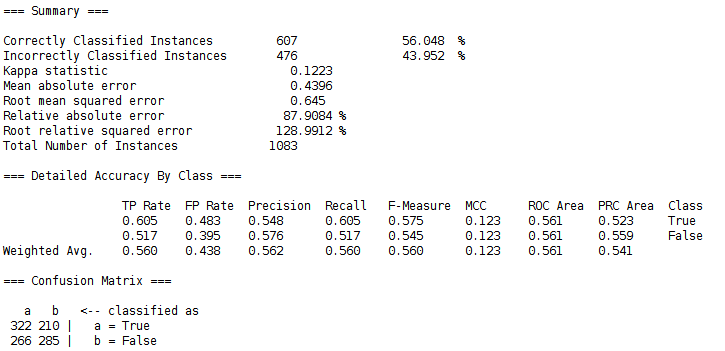
stemmer = SnowballStemmer

nGramTokenizer = min2, max3

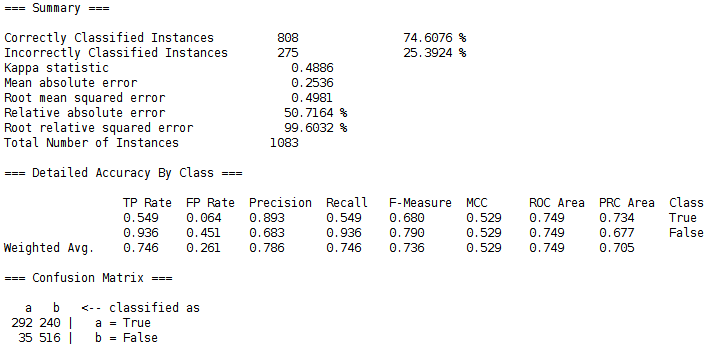
## Naïve Bayes



## J48



## IBk



# 3

StringToWordVector Settings  
IDFTransform = True

TFTransform = True

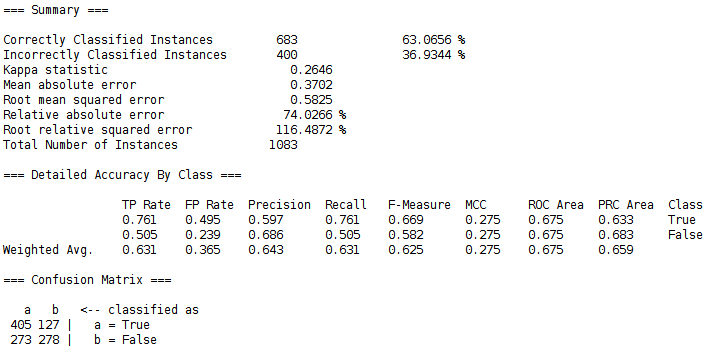
wordsToKeep = 1000

stemmer = SnowballStemmer

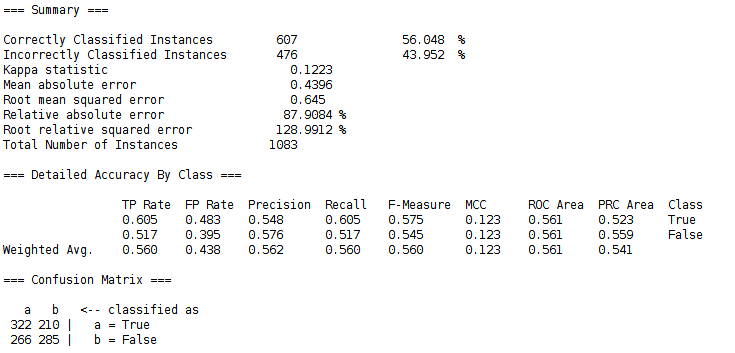
nGramTokenizer = min2, max3

stopwords = Rainbow

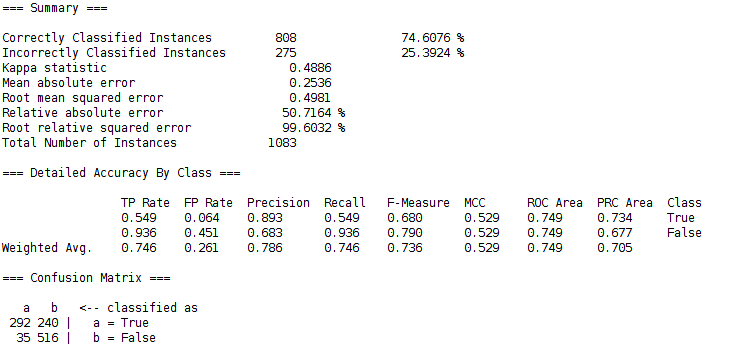
## Naïve Bayes



## J48



## IBk



# 4

StringToWordVector Settings  
IDFTransform = True

TFTransform = True

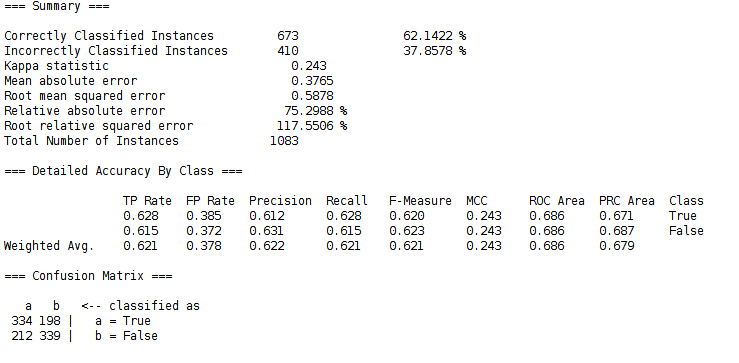
wordsToKeep = 1000

stemmer = Null

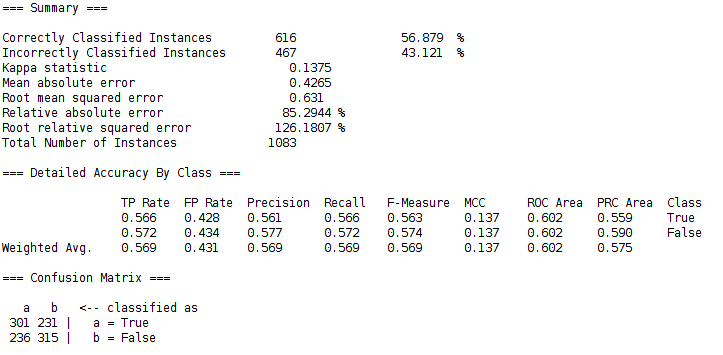
nGramTokenizer = Null

stopwords = Null

## Naïve Bayes



## J48



## IBk

