**CSI4107 Assignment 2: Sentiment Analysis in Twitter Messages**

**Comparison of Results**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Parameters (Yes/No)** | | | |
| **Test #** | **Tokenization** | **Stop Word Removal** | **Stemming** | **Attribute Selection** |
| 1 | No | No | No | No |
| 2 | Yes | No | No | No |
| 3 | Yes | Yes | No | No |
| 4 | Yes | Yes | Yes | No |
| 5 | Yes | Yes | Yes | Yes |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Accuracy of Tests (%)** | | | |
| **Test #** | **Naïve Bayes** | **SVM/SMO** | **Decision Trees/ J48** | **KNN** |
| 1 | 48.4371 | 86.3071 | 79.7234 |  |
| 2 |  |  |  |  |
| 3 |  |  |  |  |
| 4 |  |  |  |  |
| 5 |  |  |  |  |

Test 1 is computed on the provided semeval\_twitter\_data.arff file, and uses the StringToWord attribute filter in Weka to extract words. Complete results can be found in the following files:

* test1\_naivebayes.txt
* test1\_svm.txt
* test1\_decisiontrees.txt