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

**Comparison of Results**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Parameters (Yes/No)** | | | | | **Accuracy of Tests (%)** | | |
| **Test #** | **StringToWord** | **Personal Tokenization** | **Stop Word Removal** | **Stemming** | **Attribute Selection** | **Naïve Bayes** | **SVM/**  **SMO** | **Decision Trees/ J48** |
| 1 | Yes | No | No | No | No | 48.4371 | 86.3071 | 79.7234 |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

Test 1 uses the provided semeval\_twitter\_data.arff file, and uses 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