|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Feature Set** | **# of features selected** | **Model** | **Accuracy** | **Dataset** |
| word\_features | - | DT | 0.59 | Test |
| word\_pos\_features | - | DT | 0.56 | Test |
| word\_pos\_liwc\_features | - | DT | 0.56 | Test |
| word\_features | - | NB | 0.55 | Test |
| word\_pos\_features | - | NB | 0.58 | Test |
| word\_pos\_liwc\_features | - | NB | 0.58 | Test |
| word\_bin | - | DT | 0.50 | Test |
| word\_bin | 128 | NB | 0.79 | Test |
| word\_embedding | - | DT | 0.50 | Test |
| word\_embedding | 50 | NB | 0.48 | Test |
| word\_ALL\_LIWC | - | DT | 0.62 | Test |
| word\_ALL\_LIWC | 2 | NB | 0.65 | Test |

**Table -1**

The way I implemented the feature binning is through the Frequency Distribution method given by the nltk function. All I did was to calculate the total of words through Frequency Distribution, this was also given from the assignment. The number of bins I decided to choose is 5. Therefore, anything that is greater than 5 will consider 5.

**Table-2**

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature Set** | **Model** | **Accuracy** | **Dataset** |
| word\_features | NB | 0.885 | Dev |
| word\_features | NB | 0.619 | Test1 |
| word\_features | NB | 0.643 | Test2 |
| word\_pos\_features | NB | 0.895 | Dev |
| word\_pos\_features | NB | 0.619 | Test1 |
| word\_pos\_features | NB | 0.639 | Test2 |
| word\_bin | NB | 0.89 | Dev |
| word\_bin | NB | 0.609 | Test1 |
| word\_bin | NB | 0.639 | Test2 |
| word\_embedding | NB | 0.795 | Dev |
| word\_embedding | NB | 0.504 | Test1 |
| word\_embedding | NB | 0.596 | Test2 |
| word\_all\_liwc | NB | 0.56 | Dev |
| word\_all\_liwc | NB | 0.561 | Test1 |
| word\_all\_liwc | NB | 0.531 | Test2 |
| word\_features | DT | 0.71 | Dev |
| word\_features | DT | 0.533 | Test1 |
| word\_features | DT | 0.573 | Test2 |
| word\_pos\_features | DT | 0.495 | Dev |
| word\_pos\_features | DT | 0.4952 | Test1 |
| word\_pos\_features | DT | 0.4951 | Test2 |
| word\_bin | DT | 0.625 | Dev |
| word\_bin | DT | 0.4666 | Test1 |
| word\_bin | DT | 0.527 | Test2 |
| word\_embedding | DT | 0.74 | Dev |
| word\_embedding | DT | 0.571 | Test1 |
| word\_embedding | DT | 0.5836 | Test2 |
| word\_all\_liwc | DT | 0.56 | Dev |
| word\_all\_liwc | DT | 0.5619 | Test1 |
| word\_all\_liwc | DT | 0.5311 | Test2 |
| word\_features | SVM | 0.685 | Dev |
| word\_features | SVM | 0.4857 | Test1 |
| word\_features | SVM | 0.534 | Test2 |
| word\_pos\_features | SVM | 0.7 | Dev |
| word\_pos\_features | SVM | 0.552 | Test1 |
| word\_pos\_features | SVM | 0.547 | Test2 |
| word\_bin | SVM | 0.83 | Dev |
| word\_bin | SVM | 0.571 | Test1 |
| word\_bin | SVM | 0.619 | Test2 |
| word\_embedding | SVM | 0.795 | Dev |
| word\_embedding | SVM | 0.504 | Test1 |
| word\_embedding | SVM | 0.567 | Test2 |
| word\_all\_liwc | SVM | 0.56 | Dev |
| word\_all\_liwc | SVM | 0.561 | Test1 |
| word\_all\_liwc | SVM | 0.5311 | Test2 |