

Hate Speech Detection - TiMBL

Ravi Regulagedda, Jieli Liu, Ugonna Virginia Ahumibe

Initial Accuracy with no hyper-parameter tuning: 70.5451%

```
Tested: 1 @ Fri Feb 24 16:48:46 2023
Tested: 2 @ Fri Feb 24 16:48:46 2023
Tested: 3 @ Fri Feb 24 16:48:46 2023
Tested: 4 @ Fri Feb 24 16:48:46 2023
Tested: 5 @ Fri Feb 24 16:48:46 2023
Tested: 6 @ Fri Feb 24 16:48:46 2023
Tested: 7 @ Fri Feb 24 16:48:46 2023
Tested: 8 @ Fri Feb 24 16:48:46 2023
Tested: 9 @ Fri Feb 24 16:48:46 2023
Tested: 10 @ Fri Feb 24 16:48:46 2023
Tested: 100 @ Fri Feb 24 16:48:46 2023
Tested: 1000 @ Fri Feb 24 16:48:47 2023
Ready: 1537 @ Fri Feb 24 16:48:47 2023
Seconds taken: 0.5724 (2685.41 p/s)

overall accuracy: 0.705921 (1085/1537), of which 1533 exact matches
There were 13 ties of which 7 (53.85%) were correctly resolved
```

Figure 1: Result showing the baseline accuracy

After many tests the best results were achieved with `-a 4`. Every other parameter had no affect on the data, or produced worse results. The final accuracy was an improvement of 0.1% bringing the final accuracy upto 70.6571%.

In some cases, setting parameters like `-k 7` etc., improved the performance but they never gave a better performance than the above parameter. Another interesting observation was the test set and dev set had different optimal parameters. The dev set had greatest accuracy at `-a 3 -k 7 -b 9`. For these parameters, the test set had a lower accuracy than the baseline at 70.3%

The output showing the accuracy for the test set is below.

```
Tested: 1 @ Mon Feb 27 16:12:15 2023
Tested: 2 @ Mon Feb 27 16:12:15 2023
Tested: 3 @ Mon Feb 27 16:12:15 2023
Tested: 4 @ Mon Feb 27 16:12:15 2023
Tested: 5 @ Mon Feb 27 16:12:15 2023
Tested: 6 @ Mon Feb 27 16:12:15 2023
Tested: 7 @ Mon Feb 27 16:12:15 2023
Tested: 8 @ Mon Feb 27 16:12:15 2023
Tested: 9 @ Mon Feb 27 16:12:15 2023
Tested: 10 @ Mon Feb 27 16:12:15 2023
Tested: 100 @ Mon Feb 27 16:12:15 2023
Tested: 1000 @ Mon Feb 27 16:12:15 2023
Ready: 1537 @ Mon Feb 27 16:12:15 2023
Seconds taken: 0.4303 (3572.30 p/s)

overall accuracy: 0.706571 (1086/1537), of which 1533 exact matches
There were 13 ties of which 8 (61.54%) were correctly resolved
```

Figure 2: Result showing the baseline accuracy