Featured Trials

Description	Feature Types	nGram Length	Data Revision	Passes	Accuracy
Baseline	Binary	1	8	1	0.8529
Bigrams	Binary	2	8	1	0.8636
Wordcount Features	Wordcount	1	10	1	0.8561
Bigrams with Gramcounts (# of occurrences of each bigram)	Wordcount	2	11	1	0.8452
Removing Stopwords	Binary	1	12	1	0.8516
Bigrams with Stopwords Removed (bigrams containing at least 1 useful word were kept)	Binary	2	13	1	0.8542
Bigrams x7	Binary	2	15	7	0.8668

Commands Used

All CLI commands were run from the included Jupyter notebook, but I have included skeletons here:

```
rm sentiment.model
rm .cache
vw --random_seed 1 --ngram {nGramLength} --l2 0 --cache --final_regressor
sentiment.model --loss_function logistic --passes {nPasses} < {trainDataPath}
&> /dev/null
vw --testonly -i sentiment.model --predictions predictions.txt --binary  <
{testDataPath}
```

To replicate any of the trials listed above, replace the `{nGramLength}` and `nPasses` with the parameters listed in the table. Replace `{trainDataPath}` and `{testDataPath}` with the paths to the datasets in `saved_sets/` that correspond to the data revision number listed in the table.

Example:

To replicate the "Bigrams x7" model:

```
rm sentiment.model
rm .cache
vw --random_seed 1 --ngram 2 --l2 0 --cache --final_regressor sentiment.model -
-loss_function logistic --passes 7 < saved_sets/train-15.vw &> /dev/null
vw --testonly -i sentiment.model --predictions predictions.txt --binary  <
saved_sets/test-15.vw
```