**NLP Lab 8**

Dixitha Kasturi

[dkasturi@syr.edu](mailto:dkasturi@syr.edu)

**Topic: Continuation of classification and feature exploration**

I explored both Option 1 and Option 2.

Option 1:

Accuracy with stop words: 0.76

Accuracy with stop words removal: 0.788

The accuracy did increase, which in a way showed that the stop words were indeed contributing negatively to the model’s overall performance. I removed the stop words in the NOT\_featuresets function

**Top 30 features with stop words:**

Most Informative Features

V\_engrossing = True pos : neg = 19.5 : 1.0

V\_stupid = True neg : pos = 17.8 : 1.0

V\_mediocre = True neg : pos = 17.1 : 1.0

V\_generic = True neg : pos = 15.1 : 1.0

V\_inventive = True pos : neg = 14.9 : 1.0

V\_routine = True neg : pos = 14.4 : 1.0

V\_boring = True neg : pos = 14.2 : 1.0

V\_flat = True neg : pos = 13.5 : 1.0

V\_unique = True pos : neg = 12.9 : 1.0

V\_refreshingly = True pos : neg = 12.9 : 1.0

V\_refreshing = True pos : neg = 12.9 : 1.0

V\_wonderful = True pos : neg = 12.5 : 1.0

V\_90 = True neg : pos = 12.4 : 1.0

V\_warm = True pos : neg = 11.3 : 1.0

V\_stale = True neg : pos = 11.1 : 1.0

V\_mesmerizing = True pos : neg = 10.9 : 1.0

V\_car = True neg : pos = 10.4 : 1.0

V\_mindless = True neg : pos = 10.4 : 1.0

V\_dull = True neg : pos = 10.3 : 1.0

V\_quietly = True pos : neg = 10.3 : 1.0

V\_captures = True pos : neg = 10.1 : 1.0

V\_powerful = True pos : neg = 10.0 : 1.0

V\_annoying = True neg : pos = 9.7 : 1.0

V\_provides = True pos : neg = 9.7 : 1.0

V\_chilling = True pos : neg = 9.6 : 1.0

V\_waste = True neg : pos = 9.5 : 1.0

V\_supposed = True neg : pos = 9.1 : 1.0

V\_tiresome = True neg : pos = 9.1 : 1.0

V\_meandering = True neg : pos = 9.1 : 1.0

V\_unexpected = True pos : neg = 8.9 : 1.0

**Top 30 features after stop words removal:**

Most Informative Features

V\_engrossing = True pos : neg = 19.7 : 1.0

V\_routine = True neg : pos = 14.9 : 1.0

V\_inventive = True pos : neg = 14.4 : 1.0

V\_generic = True neg : pos = 13.6 : 1.0

V\_90 = True neg : pos = 13.6 : 1.0

V\_mediocre = True neg : pos = 13.6 : 1.0

V\_flat = True neg : pos = 13.4 : 1.0

V\_absorbing = True pos : neg = 13.0 : 1.0

V\_refreshing = True pos : neg = 13.0 : 1.0

V\_intimate = True pos : neg = 13.0 : 1.0

V\_NOTenough = True neg : pos = 12.3 : 1.0

V\_warm = True pos : neg = 12.2 : 1.0

V\_wonderful = True pos : neg = 12.2 : 1.0

V\_boring = True neg : pos = 12.1 : 1.0

V\_dull = True neg : pos = 12.0 : 1.0

V\_refreshingly = True pos : neg = 11.7 : 1.0

V\_stupid = True neg : pos = 11.0 : 1.0

V\_tv = True neg : pos = 10.7 : 1.0

V\_beauty = True pos : neg = 10.6 : 1.0

V\_provides = True pos : neg = 10.6 : 1.0

V\_thin = True neg : pos = 10.6 : 1.0

V\_touching = True pos : neg = 10.5 : 1.0

V\_extraordinary = True pos : neg = 10.4 : 1.0

V\_stale = True neg : pos = 10.3 : 1.0

V\_powerful = True pos : neg = 9.9 : 1.0

V\_captures = True pos : neg = 9.8 : 1.0

V\_tiresome = True neg : pos = 9.6 : 1.0

V\_unless = True neg : pos = 9.6 : 1.0

V\_flawed = True pos : neg = 9.4 : 1.0

V\_document = True pos : neg = 9.0 : 1.0

**Option 2:**

When the given function was used with 4 numeric features the following accuracy was obtained

**Text

Description automatically generated**

When the given function was used with 1 numeric feature the following accuracy was obtained

Text

Description automatically generated with medium confidence

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

The accuracy increased when only 1 numeric feature was considered. I wrote 2 versions of the same action to verify if my approach was correct.