

Text Preprocessing:

Following steps were applied in text preprocessing:

- 1) Convert text to lowercase
- 2) Remove punctuations
- 3) Remove word contractions
- 4) Replace all digits with string 'number'
- 5) Replace all familial relationships with the string 'famrel'
- 6) Replace all month names by string "month"
- 7) Replace all days of week by string "dayofweek"
- 8) Replace all friendship (non-family) relations with string "friend"
- 9) Remove Stop words. But these are carefully curated stopwords specifically for this project. We do not use standard stopwords in the nltk library as some of the negations are actually important.
- 10) Lemmatization and word position tagging using the standard nltk library.

Performance of the classes:

The following classes were strongest, and gave the best accuracy when used all by themselves:

0	Newly Discovered
8	Passing Away
9	Remission

Following are some of the observations of different combinations of classes: [square brackets indicate combined classes.

Test Case Classes	SVM	LSVC
980	80	73
9806	72	63
9807	70	68
98076	69	59
98065	62	58
98045	50	62
980452	62	53
9803	60	60
9870[65]	60	60

9870[234561]	61	62
--------------	----	----

Results of the rest of the test cases are not included as they were below 60% accuracy. The highlighted test case was the best case because it includes all classes and gives good accuracy.