Reddit Post Classification

By Daniel Sun

The Subreddits

- r/ProRevenge: text-based story subreddit detailing users achieving their revenge usually through meticulously planned methods

- r/MaliciousCompliance: text-based story subreddit of people conforming to a letter but not the spirit of a request

WARNING: May have some foul language as emotions are generally in play in these two subreddits

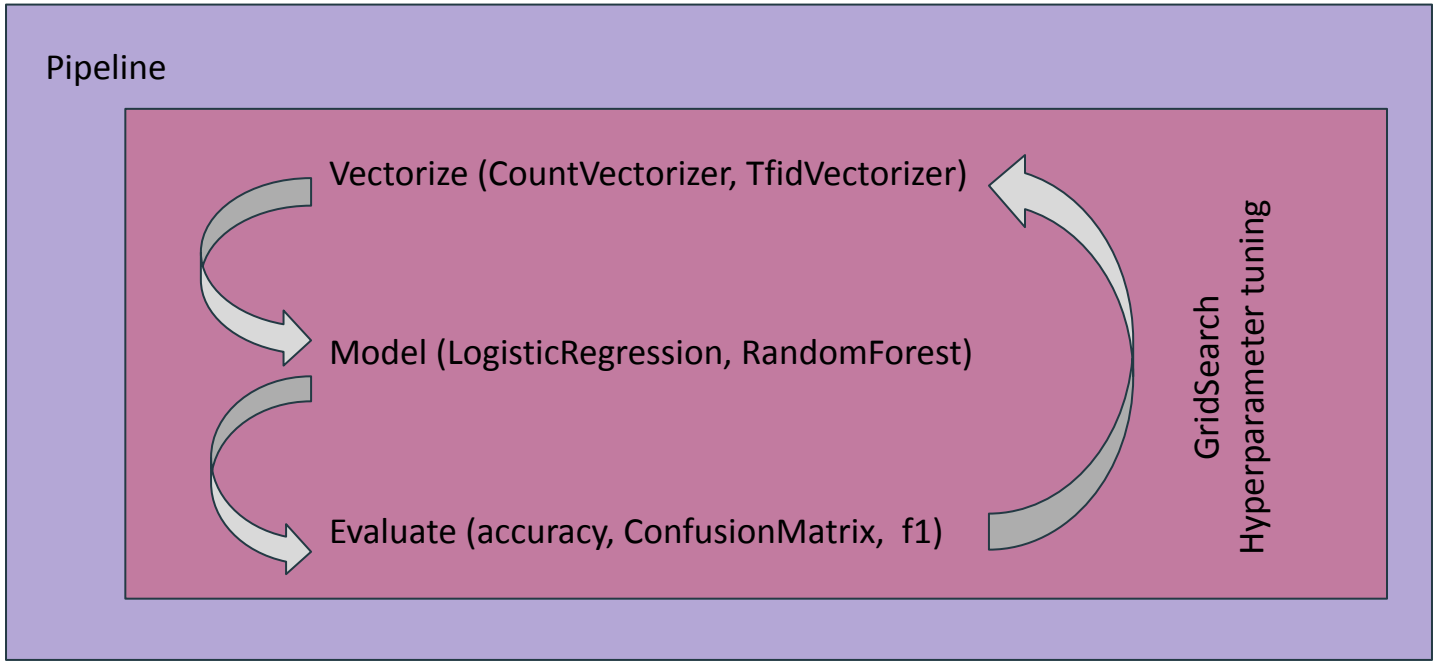
Acquisition and Cleaning

- User /u/stuck_in_the_matrix's Pushshift.io
  - 20,000 total scrapes

- Only title, subtext, and subreddit info from API was used
  - Title: The title of the post (generally a short hook for the contents of the post)
  - Subtext: The body of the post (generally the story the user is telling)
  - Subreddit: The subreddit the post comes from

- Many posts taken by the API were [removed] by moderators for one reason or another. These were eliminated
  - Final dataset:
    - ~6000 r/ProRevenge
    - ~6000 r/MaliciousCompliance

General Procedure

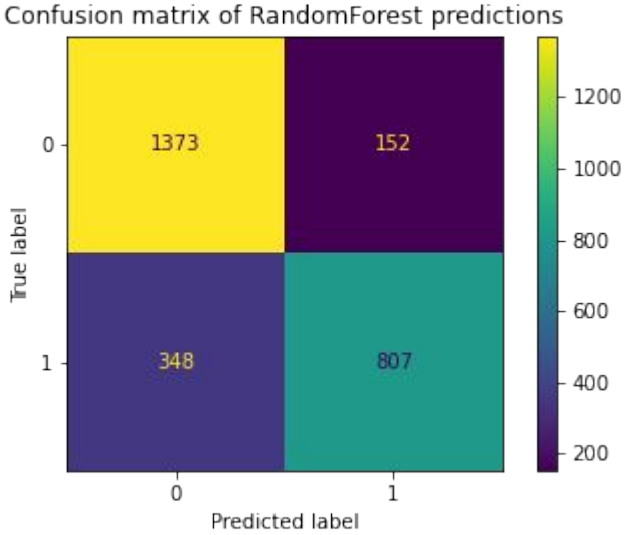


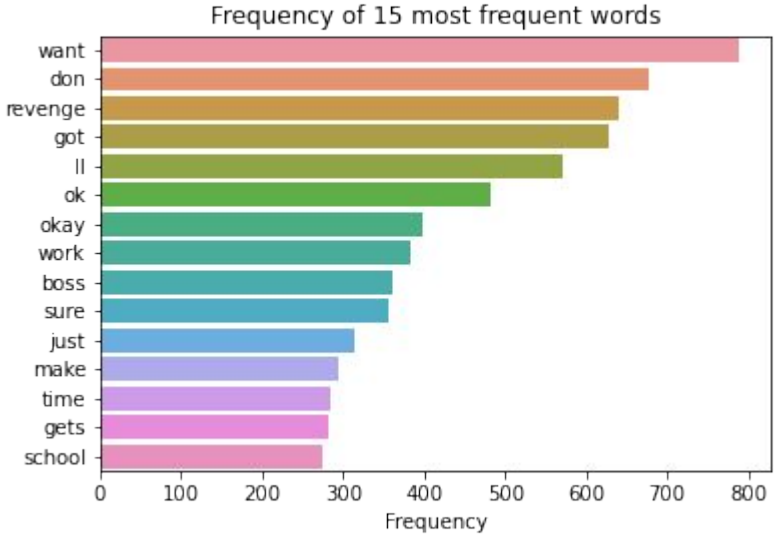
The Models

The Model	F1 Score	Accuracy
Null	—-—	0.55
**RandomForest(cv/selftext)	0.854	0.823
RandomForest(cv/title)	0.846	0.813

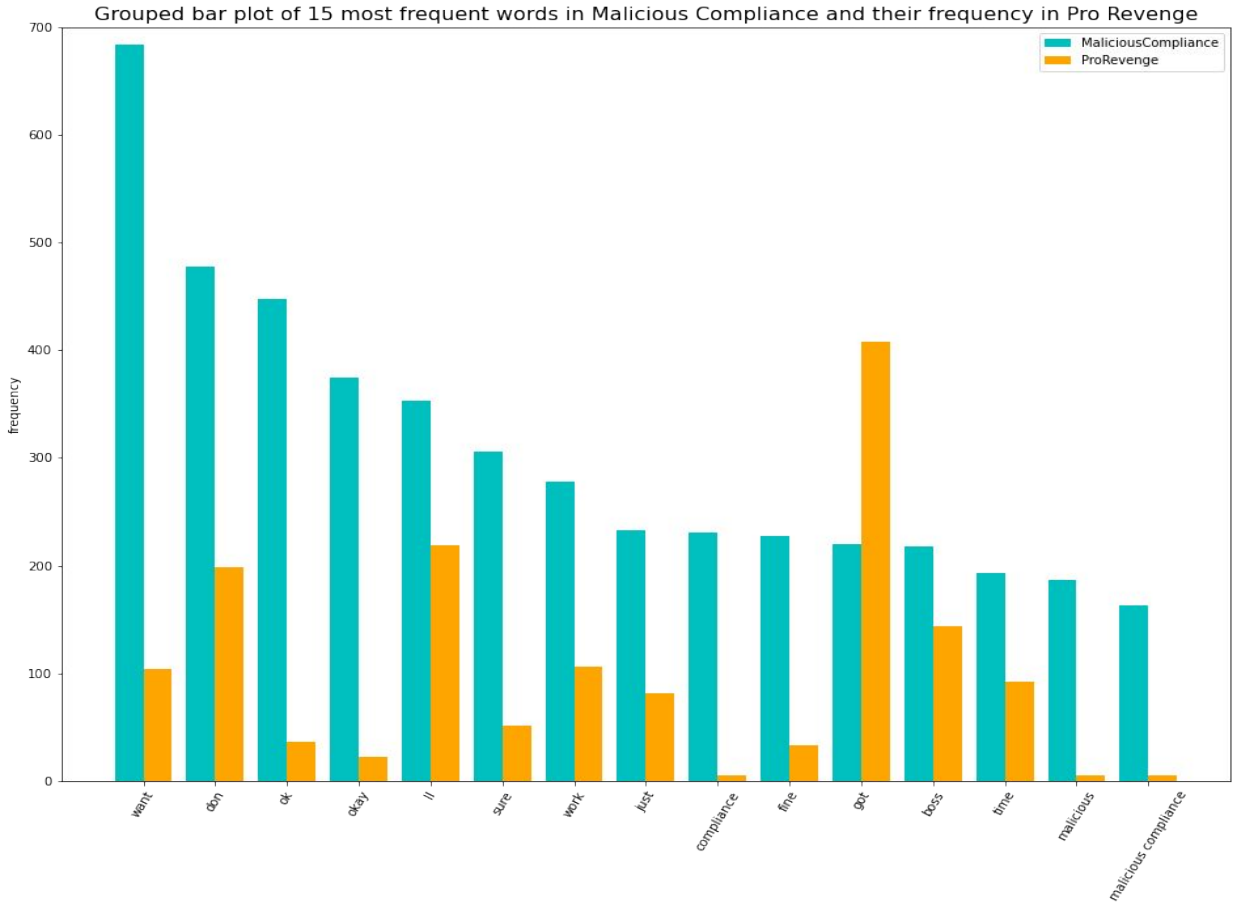
**Due to the best cv parameters including unigram *and* bigrams, the number of vectors was astronomical with selftext. Downstream procedure visualizations were gated by my system's memory

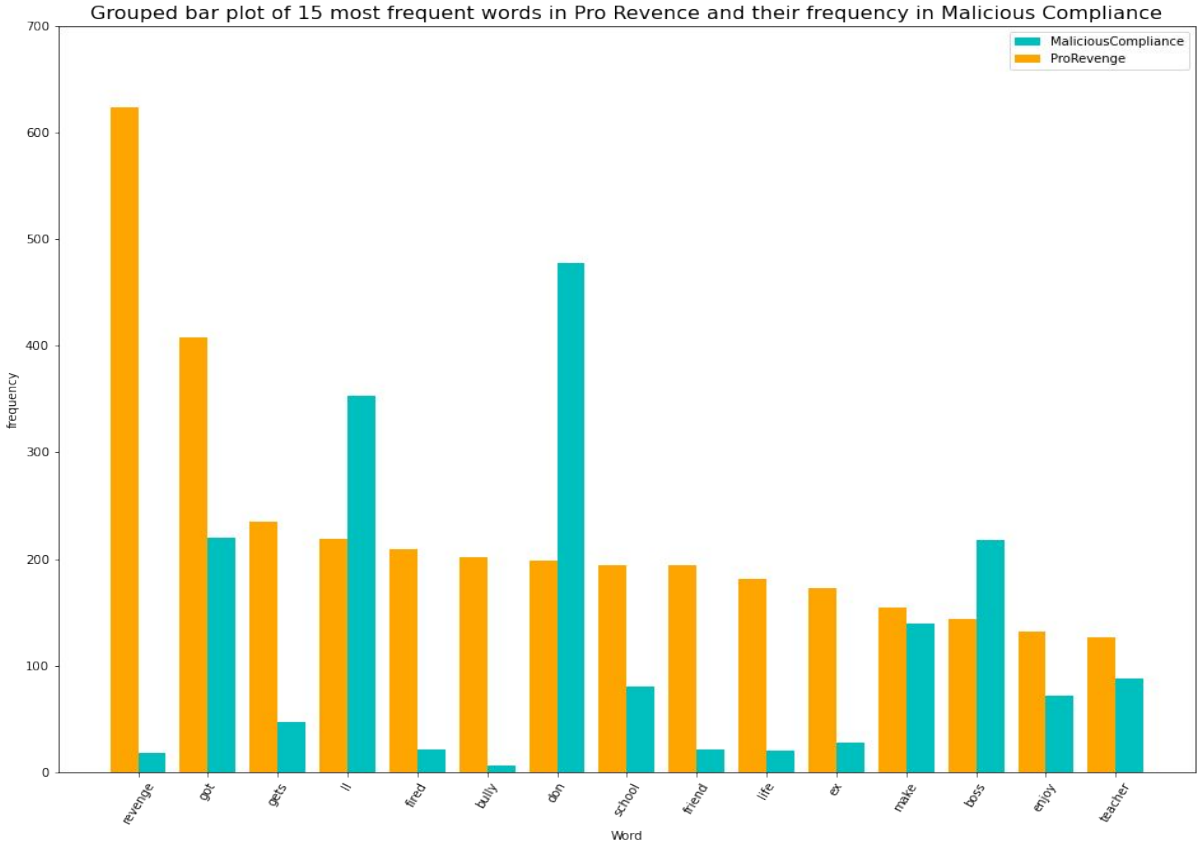
Results of the Model



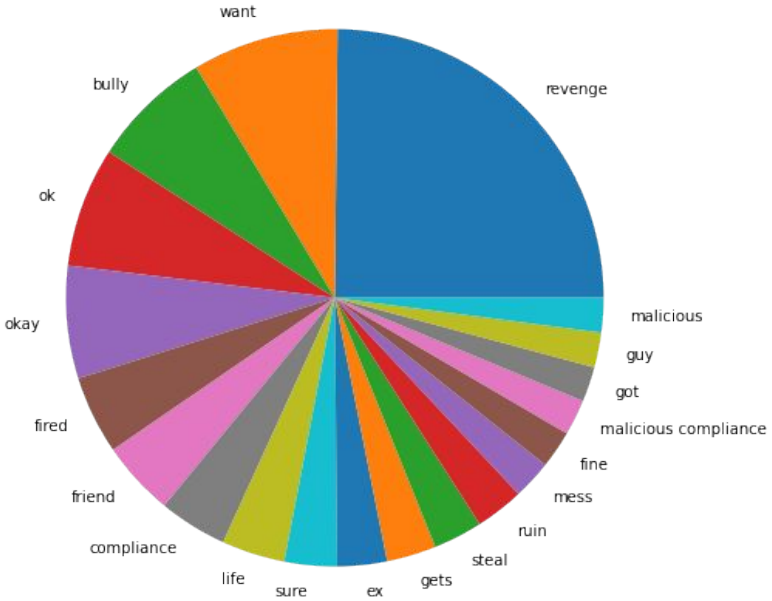


** 1 = ProRevenge, 0 = Malicious Compliance


Grouped bar plot of 15 most frequent words in Malicious Compliance and their frequency in Pro Revenge
MaliciousCompliance
ProRevenge
frequency
0
100
200
300
400
500
600
700
want
don
ok
okay
ll
sure
work
just
compliance
fine
got
boss
time
malicious
malicious compliance


Grouped bar plot of 15 most frequent words in Pro Revence and their frequency in Malicious Compliance
MaliciousCompliance
ProRevenge
frequency
Word
0
100
200
300
400
500
600
700
revenge
got
gets
ll
fired
bully
don
school
friend
life
ex
make
boss
enjoy
teacher

Feature Importance



Conclusion

- Post titles containing the words ‘revenge’, ‘bullied’, and ‘fired’, are more likely to be part of posts from r/ProRevenge while post titles containing the words ‘want’, ‘ok’, and ‘okay’ are more likely to be from r/MaliciousCompliance.

- CV/Randomforest model has some predicting power behind it, but further testing into other hyperparameters and models, like lemmatization, knn, and naive bayes, along with improving memory allocation and usage could be the next step to finding a stronger model

Additional Info

Variable	Importance
revenge	0.04293886794381857
want	0.015193002191328867
bully	0.012536351628589558
ok	0.012479212541265017
okay	0.011704662197512079
fired	0.008170120979295297
friend	0.007756189509767916
compliance	0.007135161971270696
life	0.006568491351676133
sure	0.005456262618502138
ex	0.005166853171952439
gets	0.005134287886308158
steal	0.005112238966125555
ruin	0.005064548440264722
mess	0.004021651990987687
fine	0.003927897451331342
malicious compliance	0.003684309661302677
got	0.003664394426058452
guy	0.003620499676711848
malicious	0.0035843286765251…