

# Revenge.

## Classifying Passive vs. Active Revenge in Related Subreddits

Emily Siegel  
GA DSI Project 3



...

**“Living well is the best revenge.  
And maybe a voodoo doll is a  
good investment.”**

—The Psychology of Revenge (and Vengeful  
People)

# Problem Statement

- Are the differences between *Active* vs. *Passive* revenge classifiable in the subreddits [r/MaliciousCompliance](#) versus [r/pettyrevenge](#) and [r/ProRevenge](#)?
- [r/MaliciousCompliance](#) is *Passive* not premeditated revenge and the other 2 subreddits are more *Active* planned acts of vengeance.
- Moderators for [r/MaliciousCompliance](#) can use this classification model to help filter submissions that might be better suited for another subreddit.

# Subreddit Fast Facts

## **r/MaliciousCompliance**

Created on Jan 4, 2016

1.6 million Members

Following orders of a superior while knowing that compliance with those orders will have an unintended or negative result

## **r/pettyrevenge**

Created Nov 1, 2012

1.1 million Members

Someone has wronged you, but you got your revenge

Petty, nothing criminal or serious

## **r/ProRevenge**

Created Nov 14, 2012

1.1 million Members

Should involve going out of your way and above and beyond to get revenge

# Project Workflow

## Web Scraping

Pushshift API  
550 days of submissions

1

2-3

## Overview & Cleaning

Removing Html & Markdown  
Artifacts, Punctuation etc.  
11,189 Total Submissions leftover

## Lemmatize, Vectorize & Visualize

EDA!!!

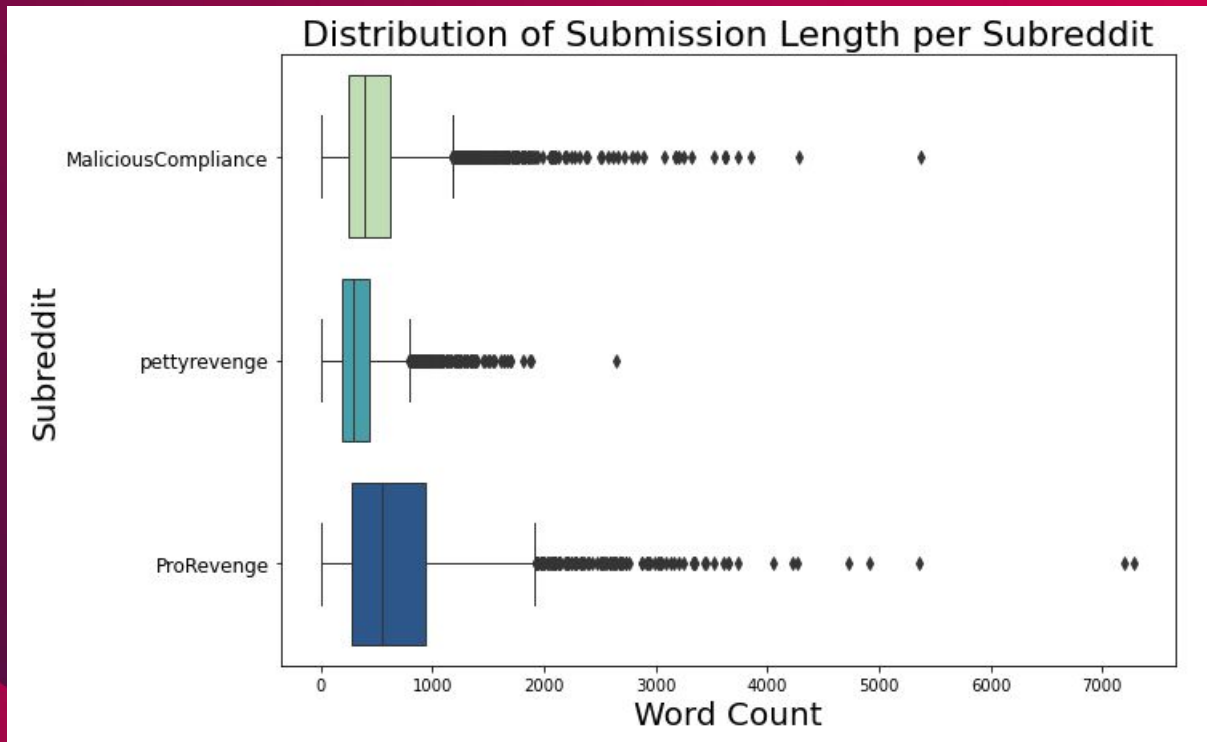
4

5 -6

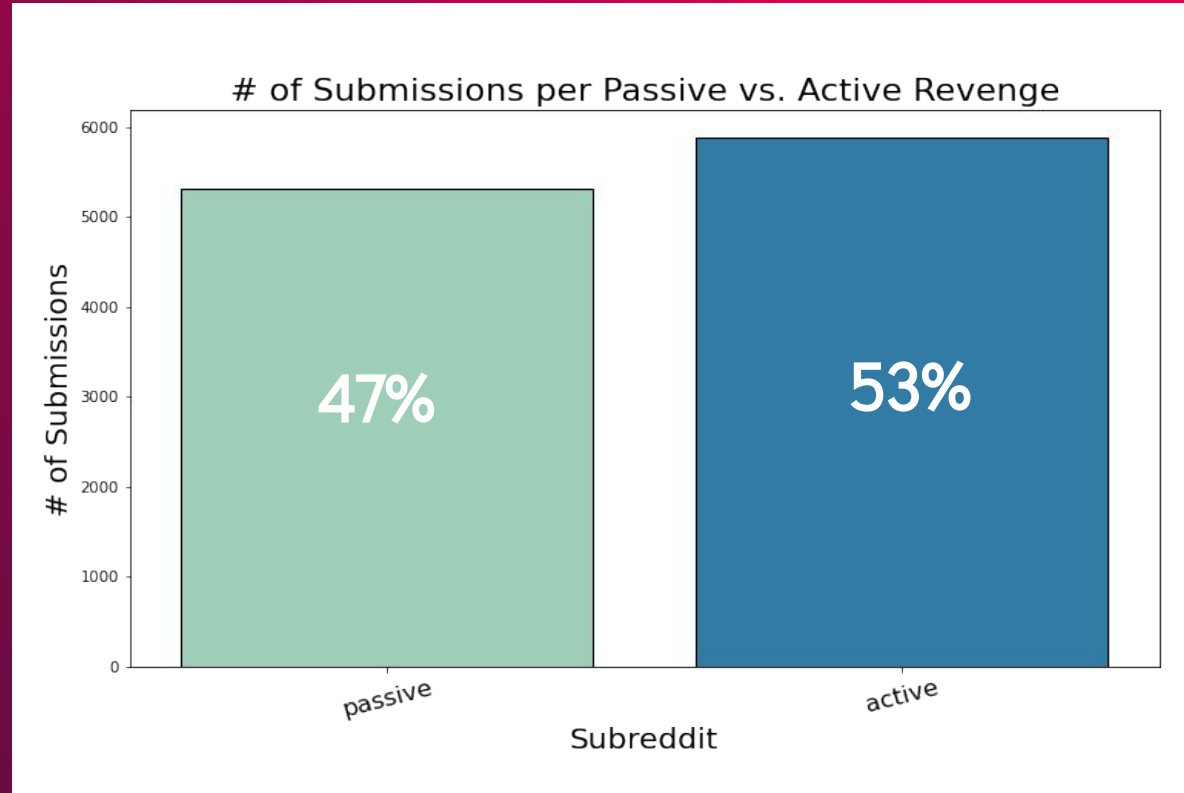
## Modeling & Results

Can we help the  
r/MaliciousCompliance  
Moderators???

# Submission Lengths

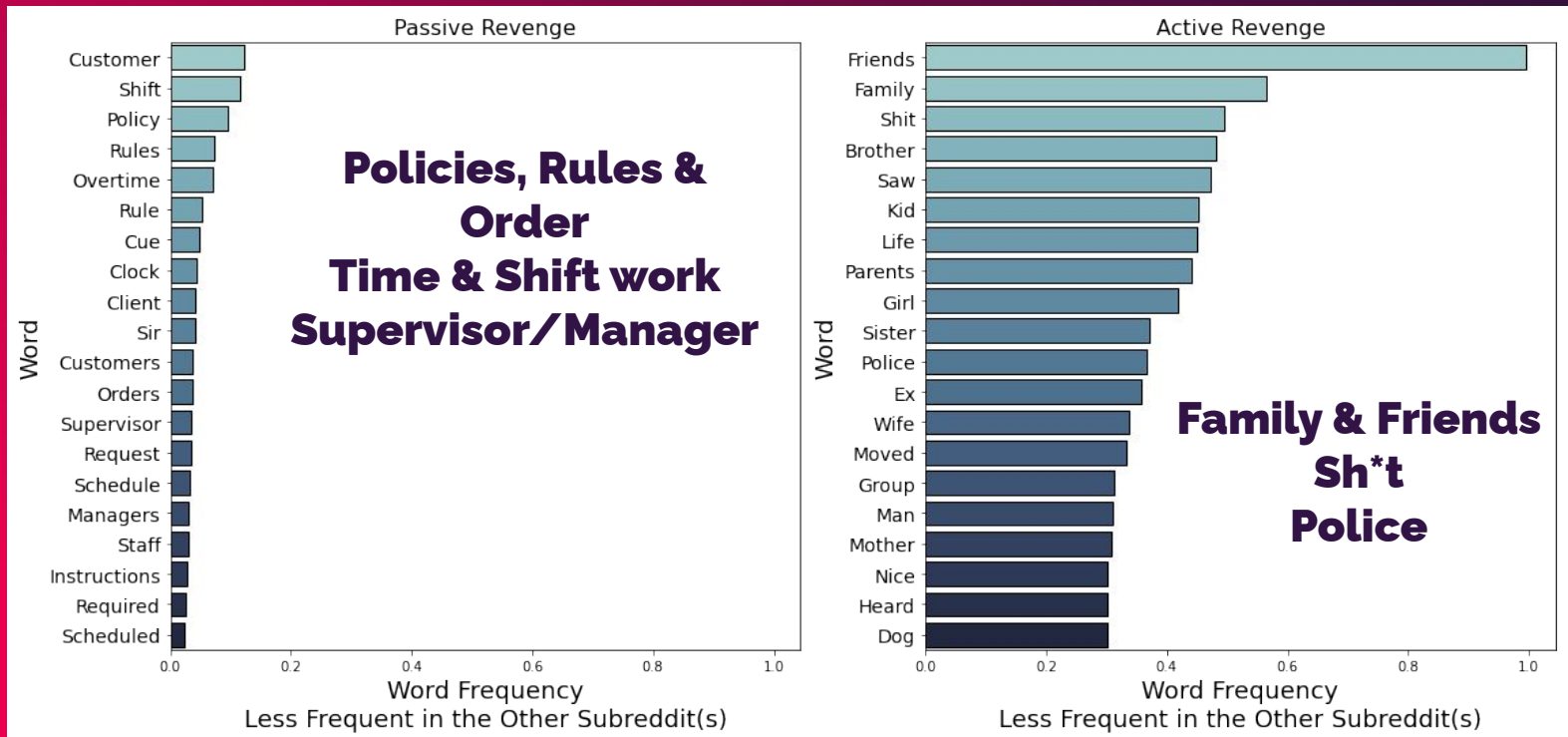


# Baseline



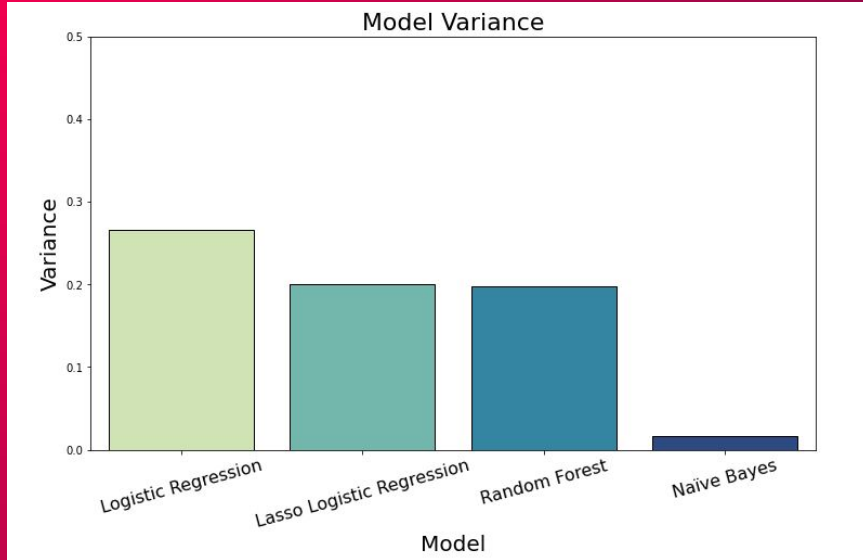
Shoutout to Gwen for this plot inspo!

# Words Most Common to Passive vs. Active Revenge Subreddits





# Model Evaluation

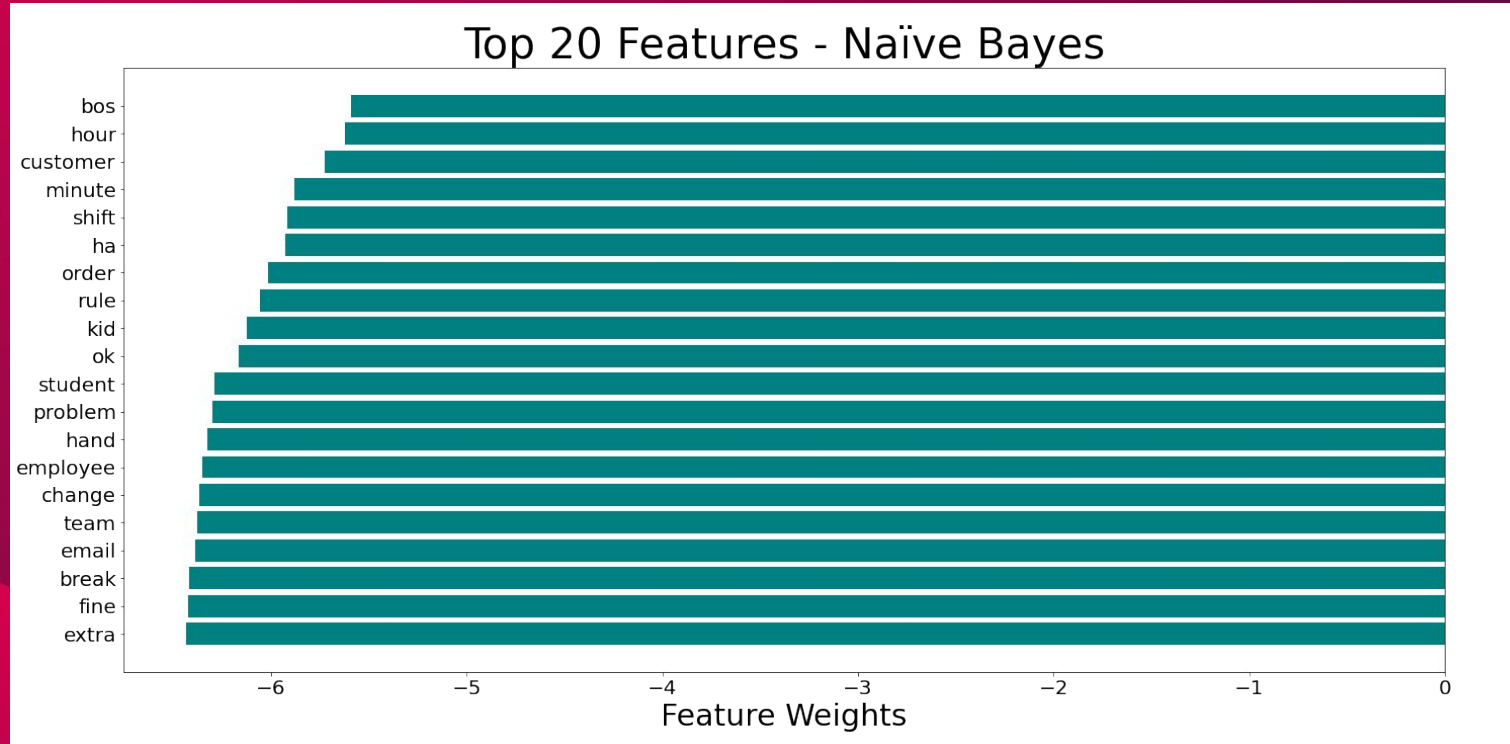


Baseline of 52.6%

	Train Accuracy	Test Accuracy	Variance
Logistic Regression	0.999333	0.733008	0.2663
Lasso Logistic Regression	0.968517	0.768210	0.2003
Random Forest	0.978922	0.780666	0.1983
Naïve Bayes	0.778949	0.762524	0.0164

# Production Model

All **negative** feature words indicate they're better at detecting our Passive Revenge subreddit class



# Conclusions

.....

01

Beat Baseline of 52.6%  
Production model has an  
accuracy of 76%

02

Overfitting was the largest  
hurdle in most models.

03

Would not use this model to auto-mod  
and remove submissions, but it's a useful  
tool for r/MaliciousCompliance mods to  
filter submissions that might be a better  
fit for r/pettyrevenge or r/ProRevenge!

# Questions?.....

• Thank You!

# References

- Data was scraped from Pushshift Reddit API <https://github.com/pushshift/api>
- <https://www.psychologytoday.com/us/blog/tech-support/201707/the-psychology-revenge-and-vengeful-people>
- David S. Chester, C. Nathan DeWall, The pleasure of revenge: retaliatory aggression arises from a neural imbalance toward reward, Social Cognitive and Affective Neuroscience, Volume 11, Issue 7, July 2016, Pages 1173–1182, <https://doi.org/10.1093/scan/nsv082>
- [https://en.wikipedia.org/wiki/Malicious\\_compliance](https://en.wikipedia.org/wiki/Malicious_compliance)
- [https://github.com/qwenrathgeber/subreddit\\_text\\_classification](https://github.com/qwenrathgeber/subreddit_text_classification)
- Slide template provided by Slidesgo