# Reddit Analysis using PushShift API

Presenter: Clifford Cho

reddit

## ABOUT COMMUNITY

A place to exchange stories about your daily experiences in brick & mortar retail.

**640k**
Members

**962**
Online

Created Nov 9, 2011

## ABOUT COMMUNITY

This is a support group for people raised by abusive parents (with toxic, self-absorbed or abusive personality traits, which may be exhibited by those who suffer from cluster B personality disorders). Please share your stories, your questions, your histories, your fears and your triumphs. Significant others and friends are all welcome.

**714k**
Members

**936**
Online

Created Feb 18, 2013

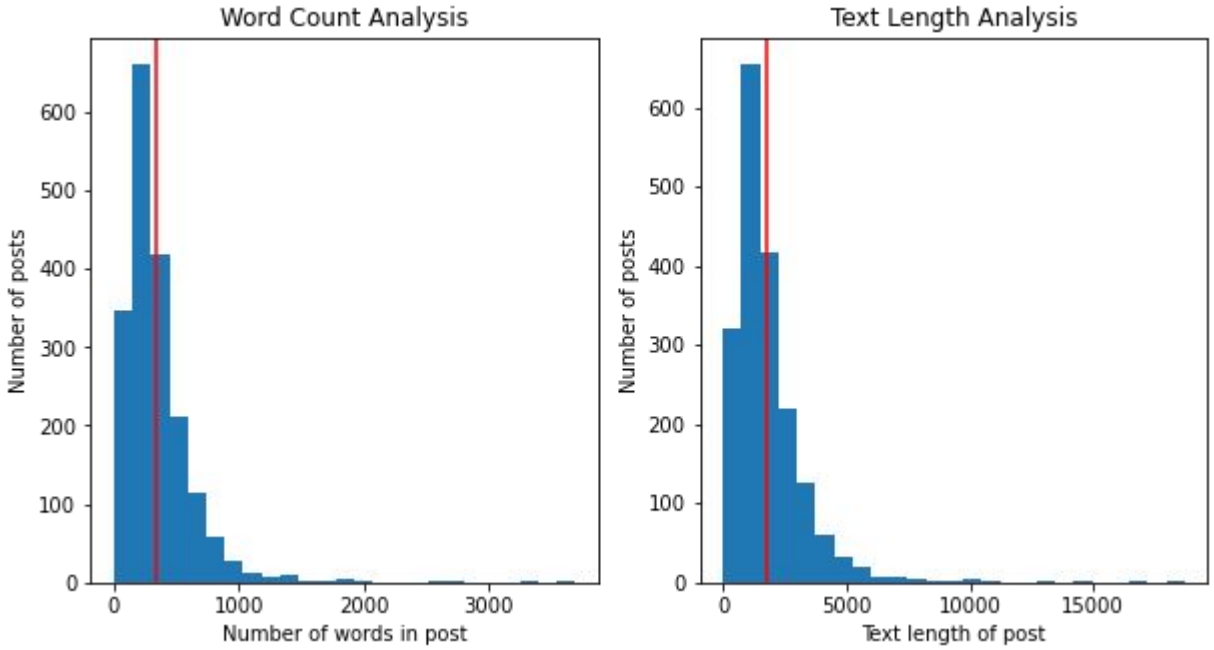
# Problem Statement

While “Karens” are known for being openly obnoxious and disliked by the public, narcissistic parents can be publicly beloved figures. This juxtaposition between public and private abuse led to a question if there was a connection between the two groups.

The main goal of this project is to see what type of vocabulary is seen in both subreddits in order to create a classification model that can predict the subreddit a piece of text comes from. Do “Karens” and narcissistic parents share any overlap?

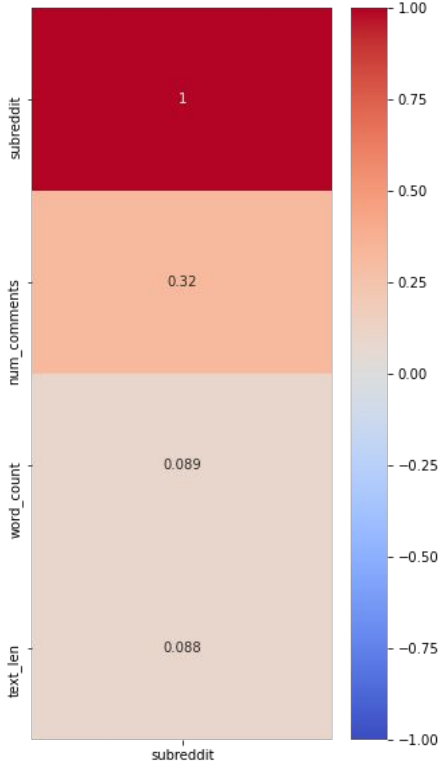
Hypothesis: The two types of people will have significant overlap because abuse is more about the abuse of power, something evident in both situations.

# Exploratory Data Analysis









# EDA using CountVectorizer

r/TalesFromRetail notable words:

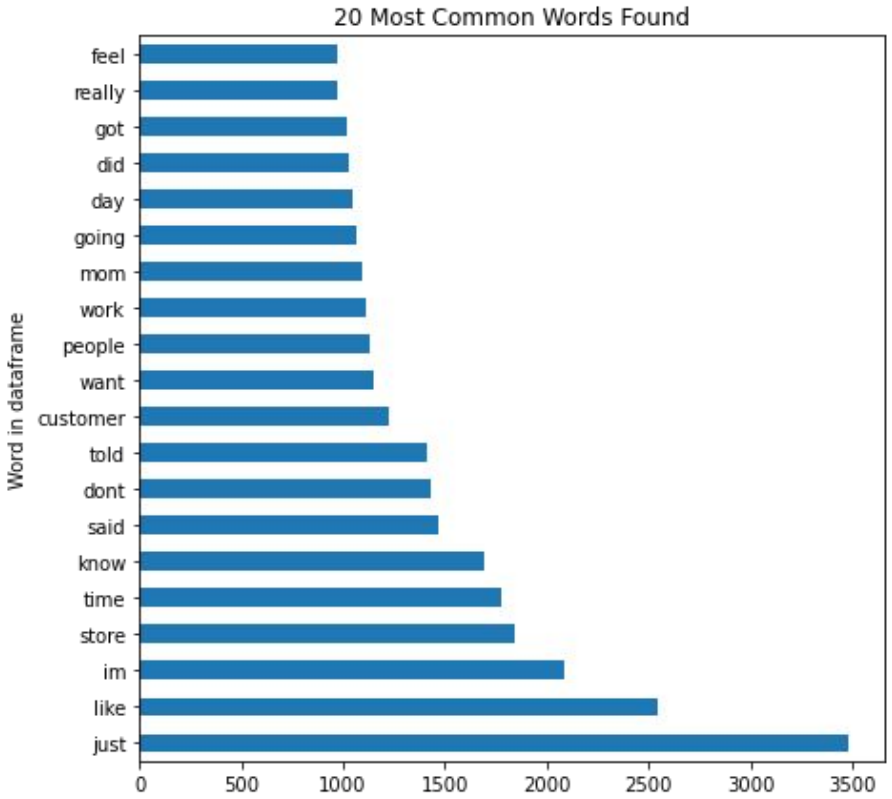
- Store, customer, manager, work, people, lady

r/raisedbynarcissists notable words:

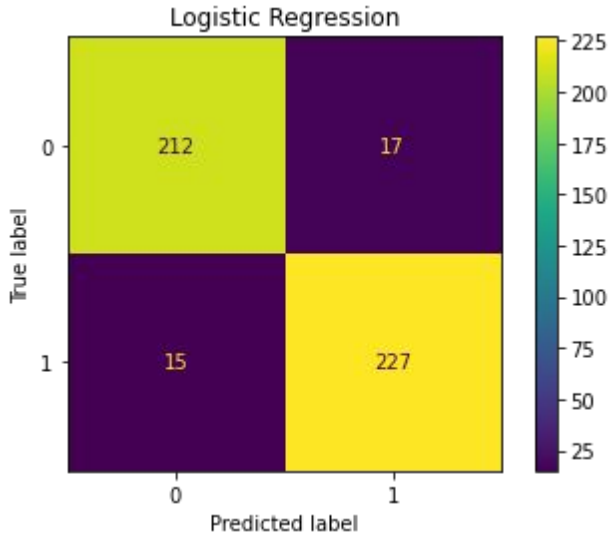
- Mom, feel, parents, dad, life, mother, family, years

“Mom” is said 50% more than “Dad”.

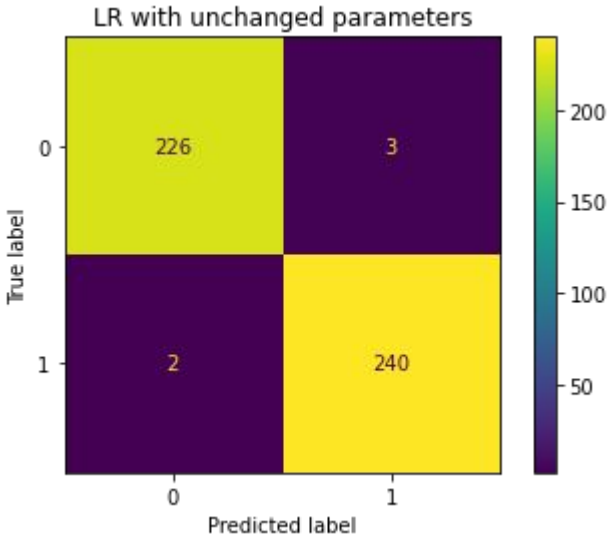
Retail also has more female “Karens” .



# Classification Modeling

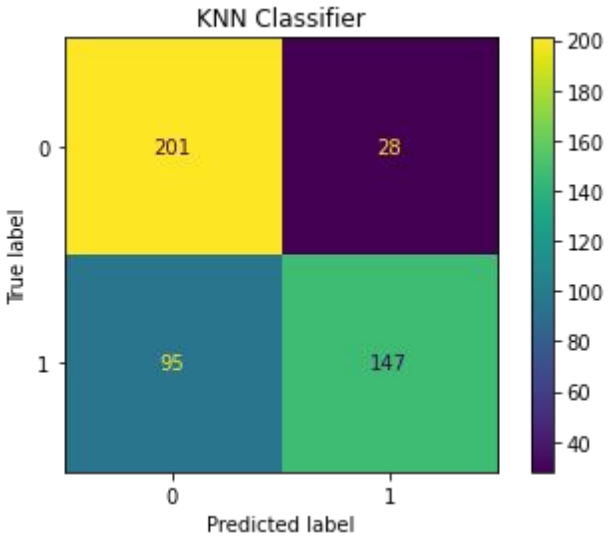



91% accuracy in Logistic Regression.

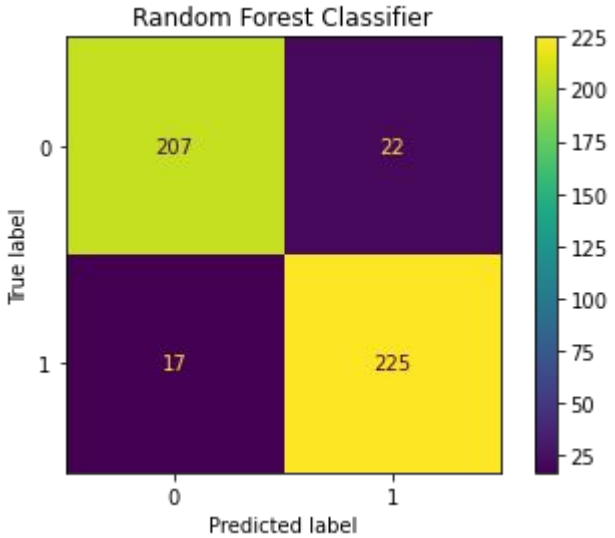



97%+ accuracy in base LR.

# Classification Modeling




72% accuracy in KNN.




90% accuracy in Random Forests

# Conclusion and Recommendations

Conclusion: Inconclusive in finding a major connection between subreddits.

Majority of classifying words were unique vocabulary from both subreddits.

Data did show overall focus on gender and money.

Recommendations: A better image of demographics can be seen by analyzing the numeric data for age of both poster and target.

Further focus on the raisedbynarcissists comments may be worthwhile.

# Thank You! Any Questions?