

Pricing Advertising Space

Using Subreddit Post Popularity for
advertisement

Ad Pricing

How to monetize based on
subreddit posts?

Reddit is starting to sell advertising space in subreddits. They want a tiered structure to pricing based on post popularity. (eg. if a post is more popular, it will cost more for ad space within a subreddit)

Some high level statistics

— — —

How many titles per subreddit?

How many words in total used in all titles?

Average word per title, sliced by subreddit?

	<i>Titles Scraped</i>	<i>Total Words in all Titles</i>	<i>Average Number of words per title</i>
<i>Psychology</i>	857	13575	~16
<i>Mental Health</i>	981	4408	~5

2 Tiered Problem - Part 1

— — —

Problem 1: Find the correct
subreddit - Classifying Psychology
vs Mental Health

Q1. How good are we at predicting
the correct subreddit?

Q2. What methods did we use to
predict the subreddit?

Q3. What is the baseline? **0.5013**

<i>Method</i>	<i>Score on Train</i>	<i>Score on Test</i>
Logistic Regression - CV	0.9956	0.8992
Multinomial NB - CV	0.9509	0.8706
KNN - TDIDF	0.5094	0.5059
Multinomial NB - TDIDF	0.9906	0.8756
Random Forest	0.9993	0.8605

What words are predictive for

Psychology?

1. Study
2. Psychology
3. Brain
4. Research
5. New
6. Psychological
7. Faces
8. Memory
9. Cognitive
10. Boys
11. Ways
12. Men
13. Effect
14. Researchers
15. People



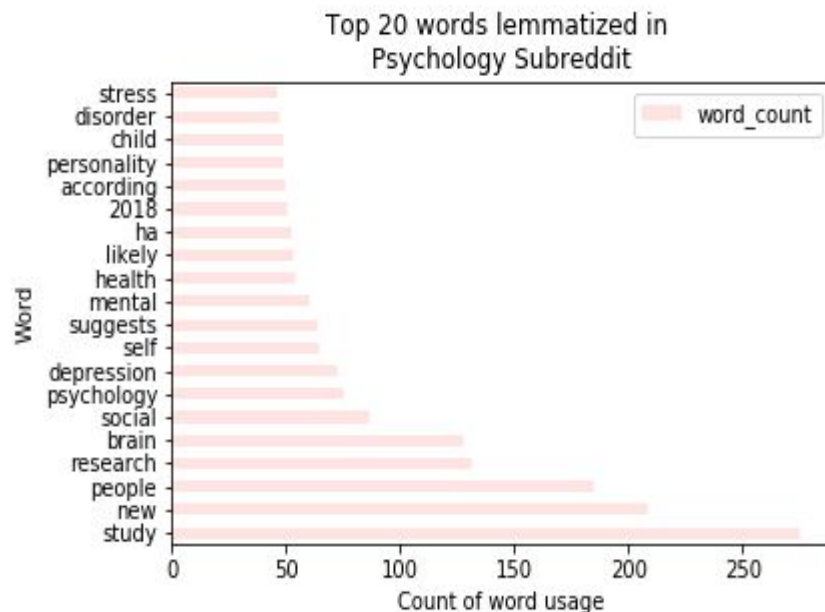
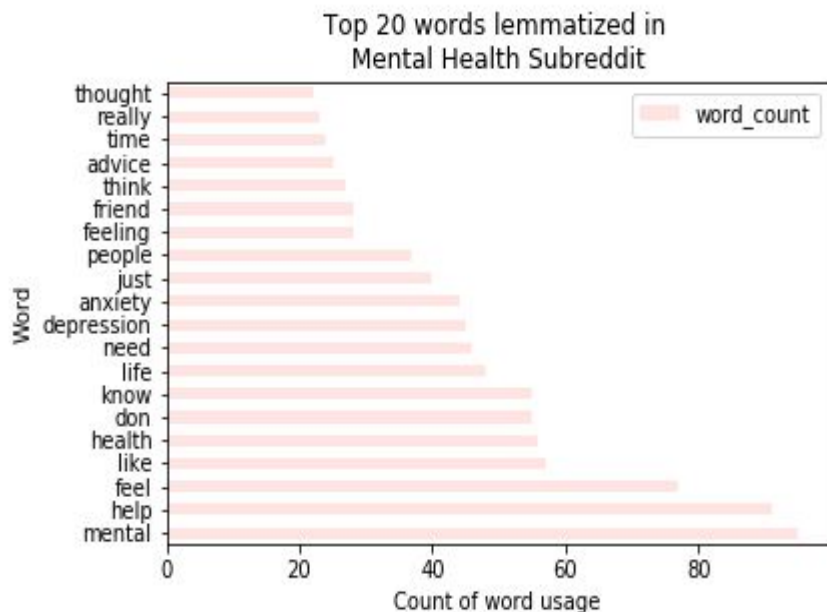
What words are predictive for

Mental Health?

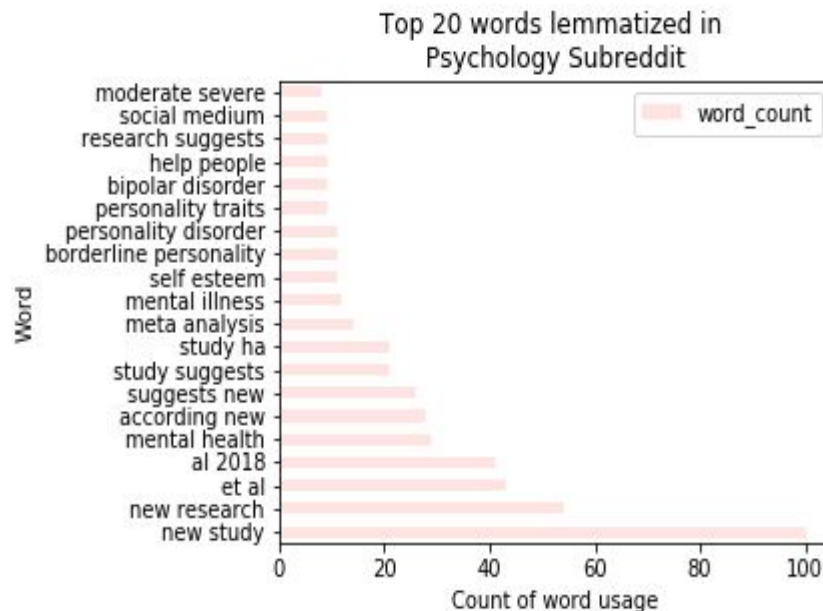
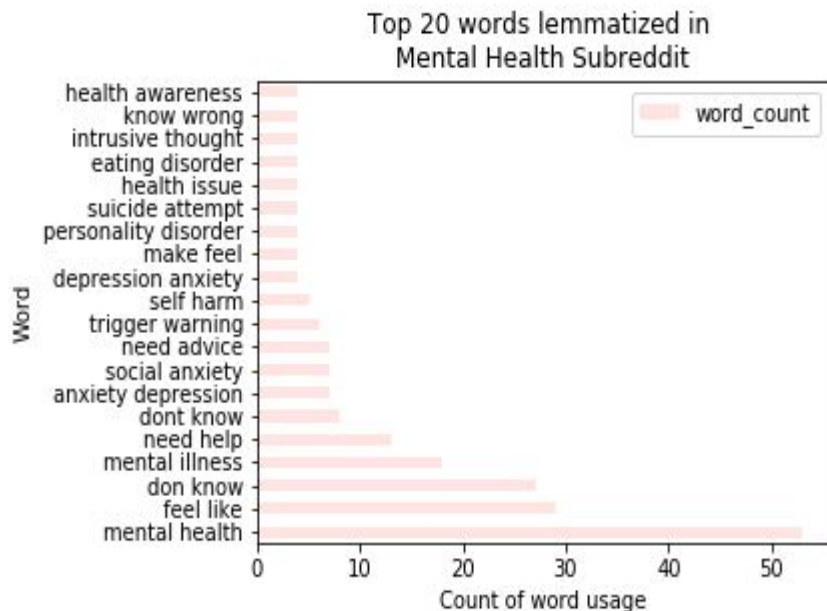
1. Help
2. hate
3. Mental
4. Scared
5. Really
6. Quite
7. Looking
8. Ill
9. Friend
10. Therapist
11. Life
12. Advice
13. Stop
14. Feeling
15. Weighted



What words had the highest frequency?



What about bigrams?



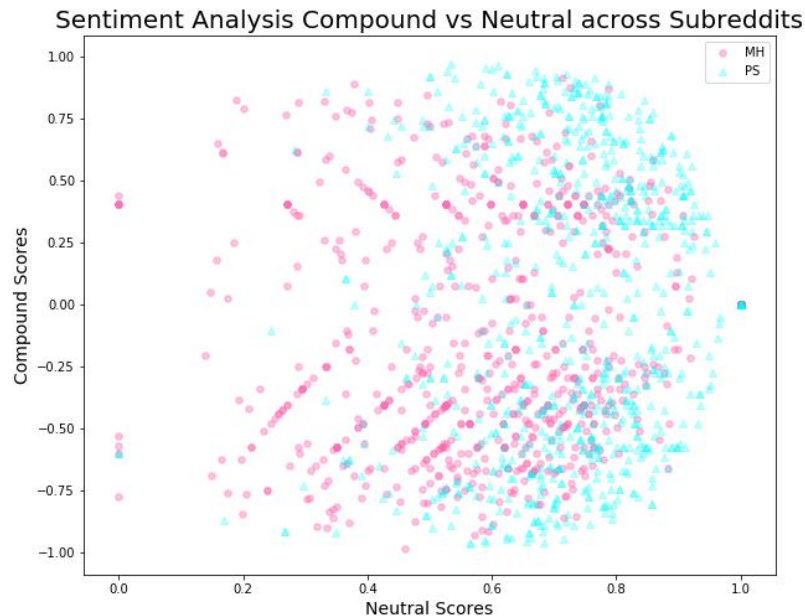
Let's talk about
feelings

308%

Psychology subreddit had 308% more words than the Mental Health subreddit although the latter had 981 distinct titles and the prior had only 857 distinct titles

Surprisingly Mental Health had more negative connotations

Neutral Scores from 0 to 0.5 are actually negative scores from 0 to -.5, while Neutral range 0.5+ relates to positive scores 0 to 0.5



Let's look at the most predictive words for psychology & mental health

— — —

1. Study
2. Psychology
3. Brain
4. Research
5. New
6. Psychological
7. Faces
8. Memory
9. Cognitive
10. Boys
11. Ways
12. Men
13. Effect
14. Researchers
15. People

1. Help
2. hate
3. Mental
4. Scared
5. Really
6. Quite
7. Looking
8. Ill
9. Friend
10. Therapist
11. Life
12. Advice
13. Stop
14. Feeling
15. Weighted

**Perhaps add space will be more effective for
Mental Health Subreddit**

2 Tiered Problem - Part 2

— — —

Problem 2: Given words, how many ups?

Average Ups: **~7**

Median Ups: **2**

Q1. How good are we at predicting the up values?

Q2. What is the baseline? **0.3188**

<i>Method</i>	<i>Score on Train</i>	<i>Score on Test</i>
Random forest	0.3212	0.3239
Adaboost	0.3306	0.3198
Gradient boost	0.5371	0.2753
SVM	0.3171	0.3239

Conclusions and next steps

It seems as though while we are very good at predicting the correct subreddit, we are not very good at predicting the up votes. Perhaps a better marketing strategy would be to predict which subreddits are the most popular on the hot page and price advertisements based on the subreddit.