
NLP for Mental Health Awareness

— Annelise Holverstott —
11/5/2021


Contextual Memes


 ocd_and_positivity · [Follow](#) ...

 avpd-chekov
don't you just love how mental disorders are basically buy one get seven free

 cestneuneblog
I didn't even really want the first one my mum made me buy it

 dorky-lesbian
Mine's a family heirloom, passed down through generations

 ig_adhdlife
Based in: Not shared ...

 **Chuck Wendig** ✓
@ChuckWendig ...

Inside you are two wolves. One has Generalized Anxiety Disorder and the other has Adult ADHD. The one who wins is oh wait I'm just receiving word that they're teaming up. They're a team now. So that's happening.

13:36 · 18 Oct 21 · [Twitter for iPhone](#)

The problem

These models aim to classify posts from subreddits for several overlapping mental health conditions, with the goal of helping people understand their brains (or their loved ones' brains) better.

Disclaimers!

This is a tool for personal use, not as a replacement for an actual, certified mental health professional.

Neurodivergence ≠ Mental Illness

Process

Step 1

Collect Data

Web scraper using Reddit's API, PushShift. Collected 2000 posts from each of 4 subreddits: ADHD, Anxiety, Depression, Aspergers (autism)

Step 2

EDA

Checked post length but did not exclude any outliers. Ended up with 6339 posts, mean length 174 words. Some custom stop words.

Step 3

Modeling

- VADER
- CountVectorizer
- Multinomial Naive Bayes
- Random Forest

Baseline

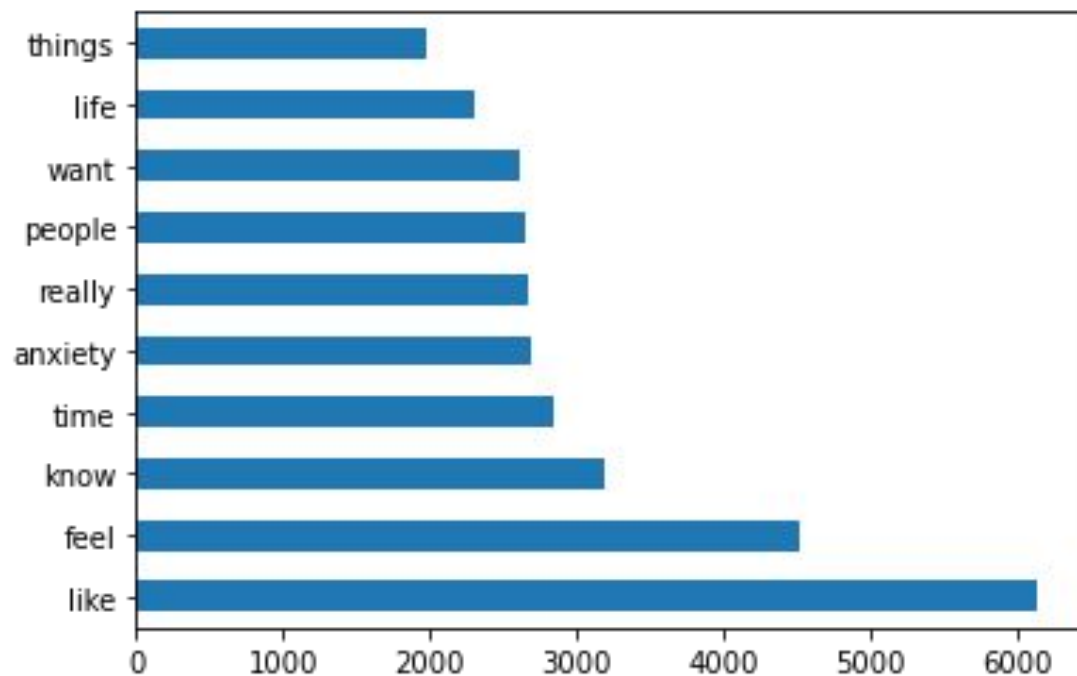
r/Anxiety 29%

r/Depression 26%

r/ADHD 24%

r/Aspergers (autism) 21%

Top 10 Words



Multinomial Naive Bayes Model

With CountVectorizer

Accuracy:

82% Train, 77% Test

Best Parameters:

max_df = 0.9

max_features = 5000

min_df = 3

ngram_range = (1, 2)

Random Forest Classifier

with CountVectorizer
&
VADER Sentiment Analysis

Accuracy:

90% Train, 79.5% Test

Best Parameters:

max_depth = 19

min_samples_leaf = 2

min_samples_split = 5

n_estimators = 1000

Conclusions & Recommendations

- Both models perform better than baseline
- Overfitting is an issue
- Looking at *wrong* results could be a good next step
- If you are interested in your mental health, submitting a writing sample about your mental state to this tool might be able to point you in the right direction
- Doctors are still better than computers!