

# Using *NLP* to Detect & Predict Mental Disorders



Emma Scotson





**World Health  
Organization**

**970 million**  
people living with a mental disorder

**26%, 28%**  
increase respectively in anxiety and major  
depressive disorders in 2020 (COVID-19)

**1 in 3**  
people with depression receive formal  
mental health care



# Data Understanding

Integrated from **9 Kaggle datasets**,  
meticulously curated, cleaned, and  
compiled

Over **51,000** online posts via Reddit,  
Twitter, etc.

7 different mental health statuses

**Normal**

**Depression**

**Suicidal**

**Anxiety**

**Bi-Polar**

**Personality Disorder**

# F-1 Score

(weighted)

**Precision**

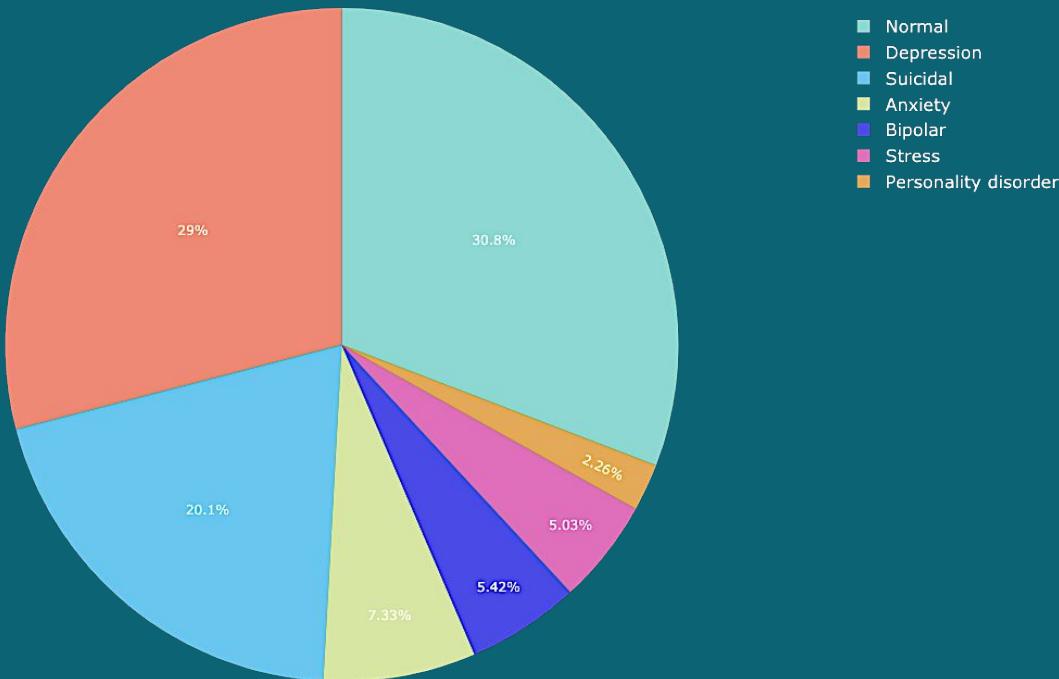
*False Positives*

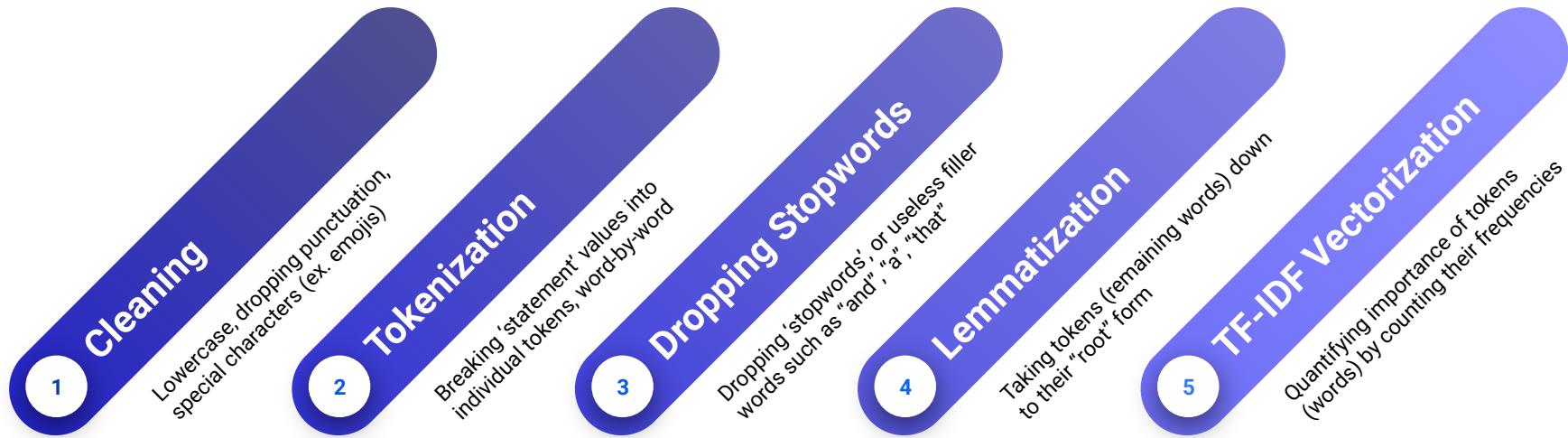
**Recall**

*False Negatives*

# Class Imbalance

Proportion of Each Status Category



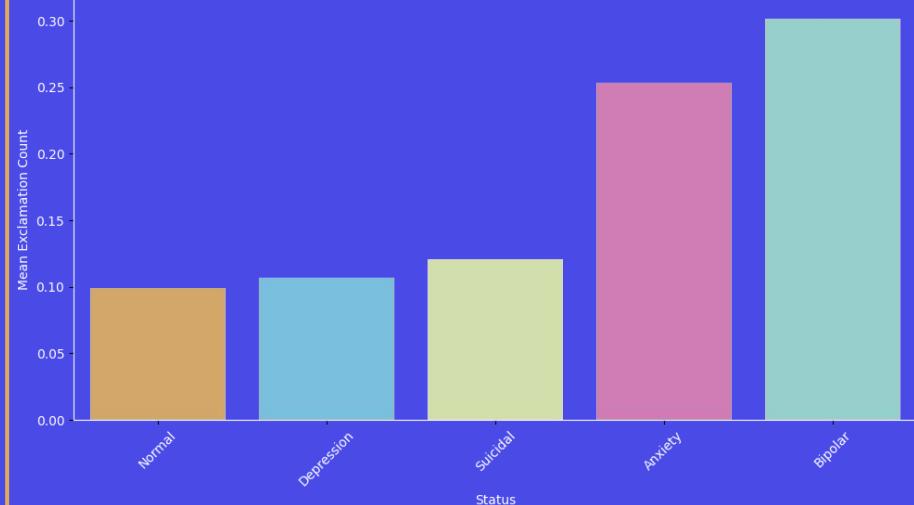


# Numeric Data

Bi-Polar   Depression   Suicidal   Anxiety   Normal

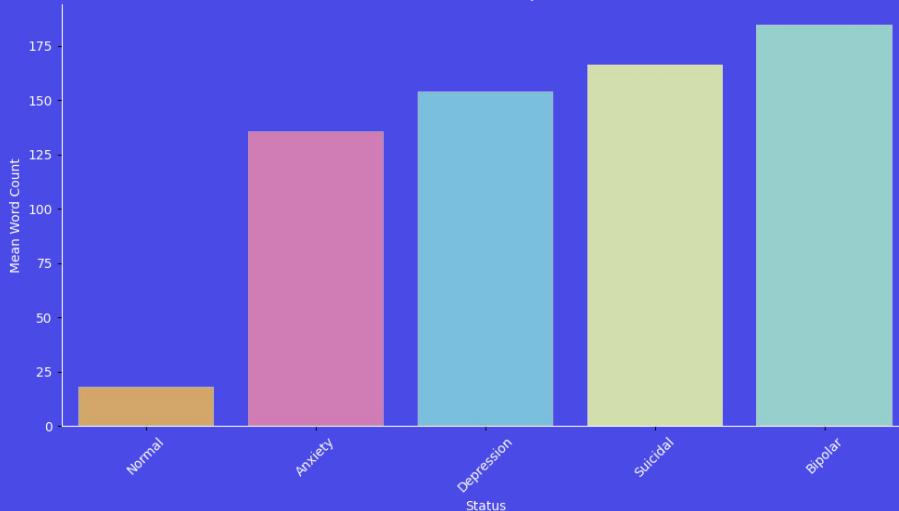
## Exclamation Count

Mean Exclamation Count by Status



## Word Count

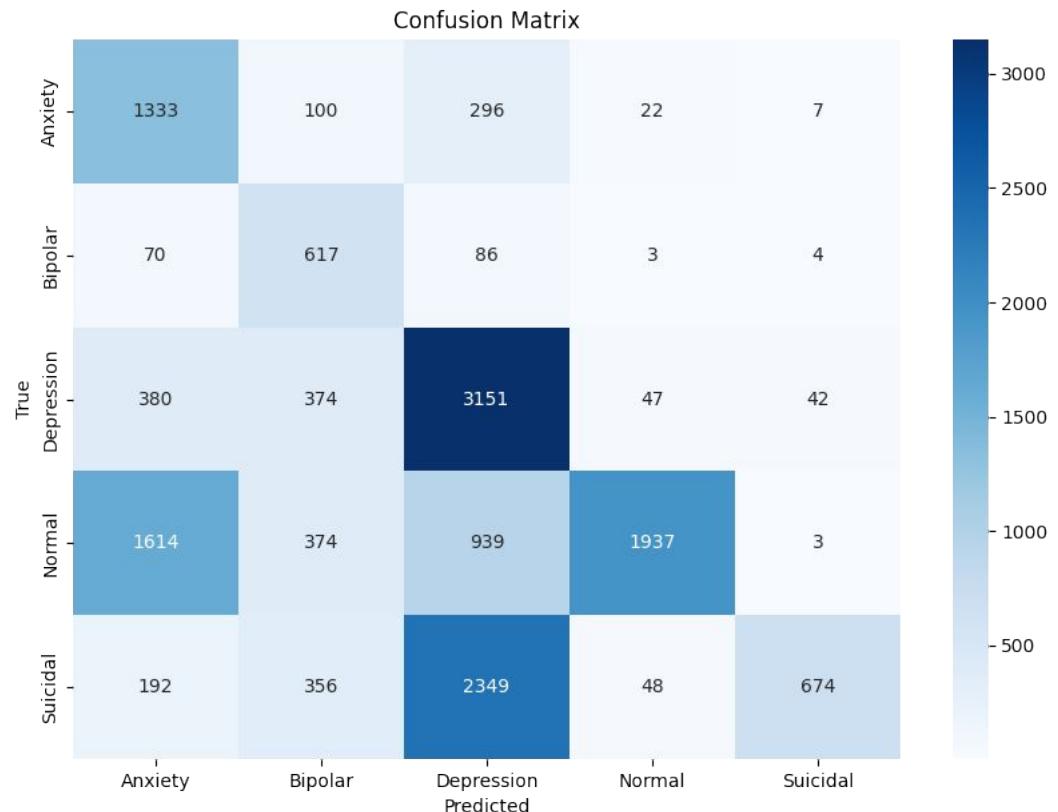
Mean Word Count by Status



# Percentage Uppercase Words

		Average	Maximum
1	<i>Suicidal</i>	7.6%	100%
2	<i>Bipolar</i>	6.0%	25%
3	<i>Anxiety &amp; Depression</i>	5.6%	100%

**Statement 20:** I really hate posting this shit on social media so I am hoping an anonymous account will help. I just needed to get this off my chest. I FUCKING HATE BEING AN ADULT. EVERYTHING IS SO LONLEY, ALL YOU DO IS WORK AND PAY BILLS, NO ONE REALLY ACTUALLY CARES ABOUT YOU, AND you are STUCK DOING EVERYTHING YOURSELF EVEN WHEN YOU REALLY JUST NEED THE COMFORT OF ANOTHER HUMAN BEING, BUT EVERYONE ELSE IS JUST AS DEPRESSED AS YOU SO are not NOBODY GOT TIME FOR THAT. Thank you Adulting.



Relabelling status groups that were incorrectly tagged

**2349**

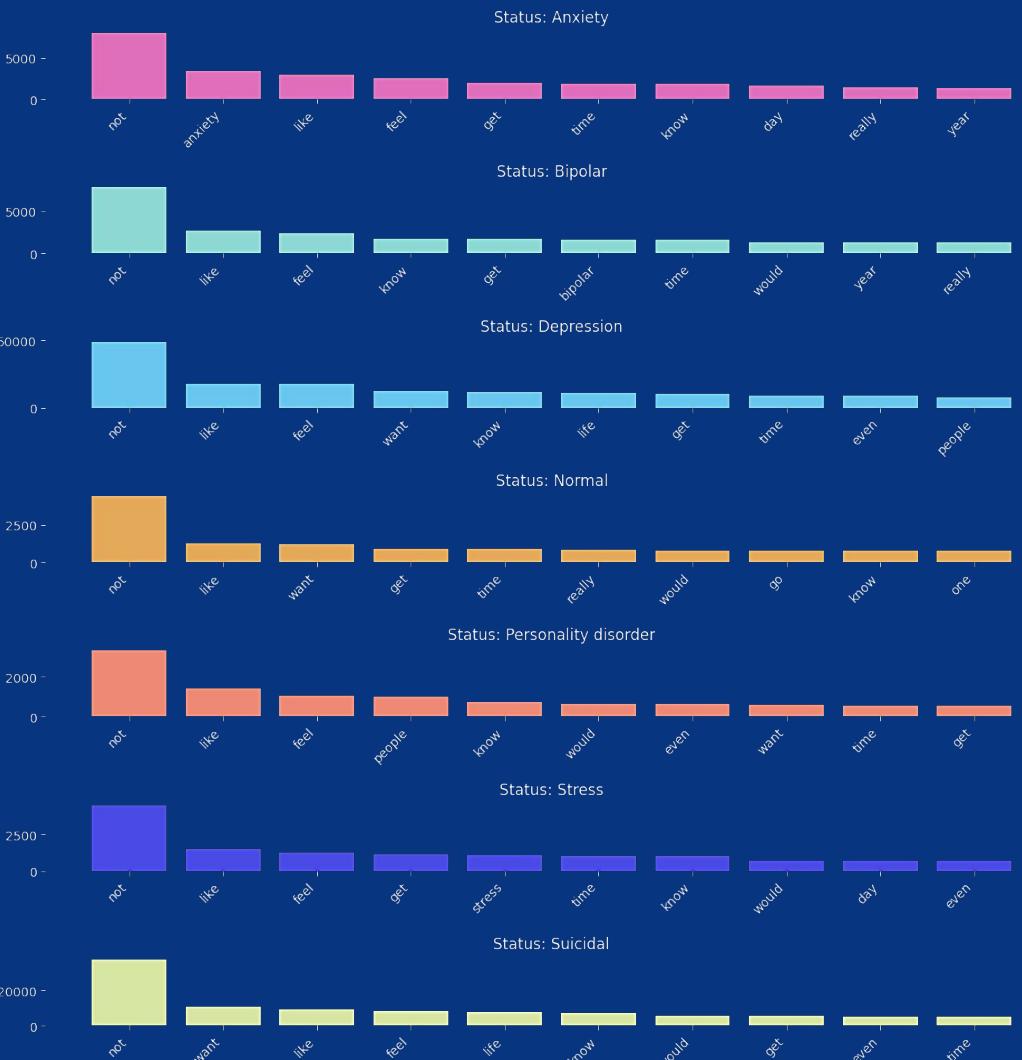
*'Suicidal' incorrectly predicted as 'Depression'*

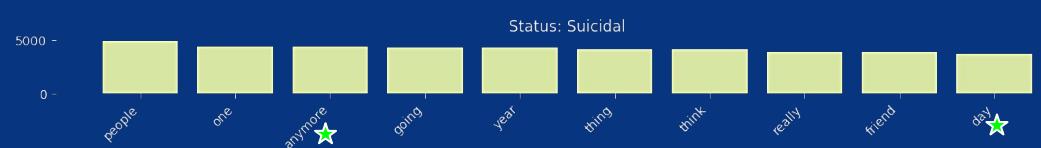
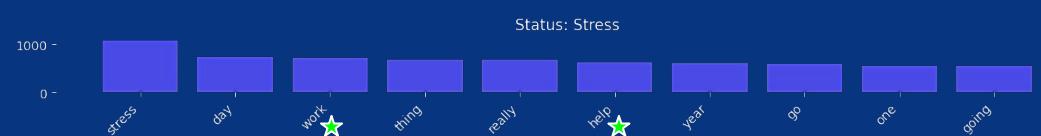
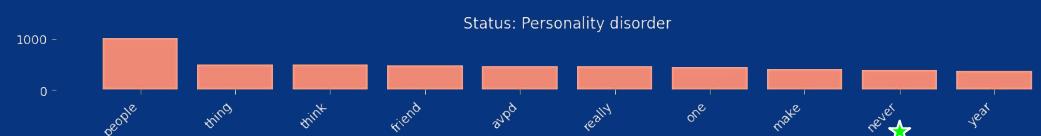
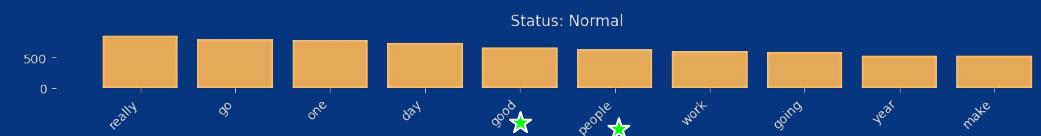
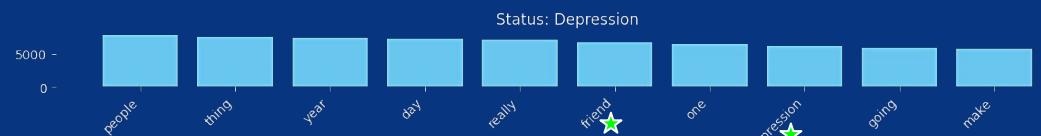
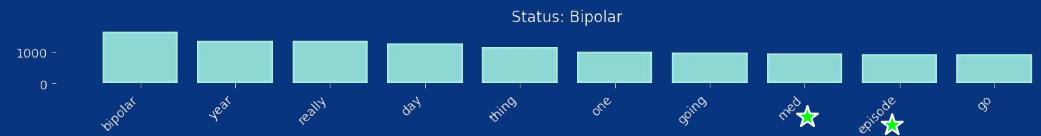
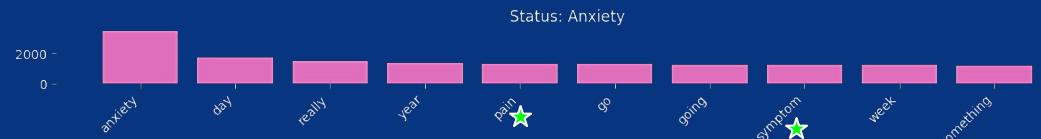
**1614**

*'Normal' incorrectly predicted as 'Anxiety'*

**939**

*'Normal' incorrectly predicted as 'Depression'*





# Conclusions



# Best Model: Random Forest

Numeric & Text Features

F1: 0.4305

Test set score: 0.4659

clf\_alpha: 0.1

tfidf\_max\_features: 20000

F1: 0.5683

Test set score: 0.5734

clf\_alpha: 0.1

tfdif\_max\_features: 10000

select k best: 5000

F1: 0.7487

Test set score: 0.7512

clf\_C: 10.0, clf\_solver: saga

max iterations: 5000

max features tfidf: 20000

F1: 0.7643

Test set score: 0.7668

tfidf\_max\_features: 10000





worth cry reason self best almost mom fucked yet  
week night another pill real hope still  
call else another depression work worse first sorry  
guess person future better made past body  
back therapy come see alone well mean  
told nobody id probably may be old sick  
actuallly job home away started stupid post  
job away money feel pain kid  
wake hurt seem talk happy head start  
hurt look scared alive start  
getting around sometimes world alive care  
good hard next find give  
long kind least tell bad  
sure hard mother pretty month  
relationship sleep ago deserve  
living hell went





# Next Steps

- Further reclassifying **status labels**
  - (ex. Bipolar bigrams mistcategorized as Depression)
- **Domain knowledge:** thorough research as to how and why certain mood disorders are diagnosed
- Unsupervised machine learning to **group words by topics.**
- More model improvements will be needed before developing **resource-providing chatbot**

# Questions?