

DETECTING DEPRESSION VIA TWEETS



DO YOU KNOW....

- Depression is a common mental disorder. Globally, more than 264 million people of all ages suffer from depression
- Depression is a leading cause of disability worldwide and is a major contributor to the overall global burden of disease
- At its worst, depression can lead to suicide.



Depression can affect anyone

MENTAL HEALTH

BELLA HADID OPENS UP ABOUT HER BATTLE WITH DEPRESSION AND ANXIETY

"That's why I am so passionate about sharing with you my newest endeavor, so we can help bring happiness back into the world, through brain health."



Naomi Osaka withdraws from French Open & reveals 'bouts of depression'

⌚ 31 May 2021 | Tennis |



Naomi Osaka has won four Grand Slams, all on hard courts

World number two Naomi Osaka has withdrawn from the French Open after the controversy over her refusal to speak to the media at the tournament.

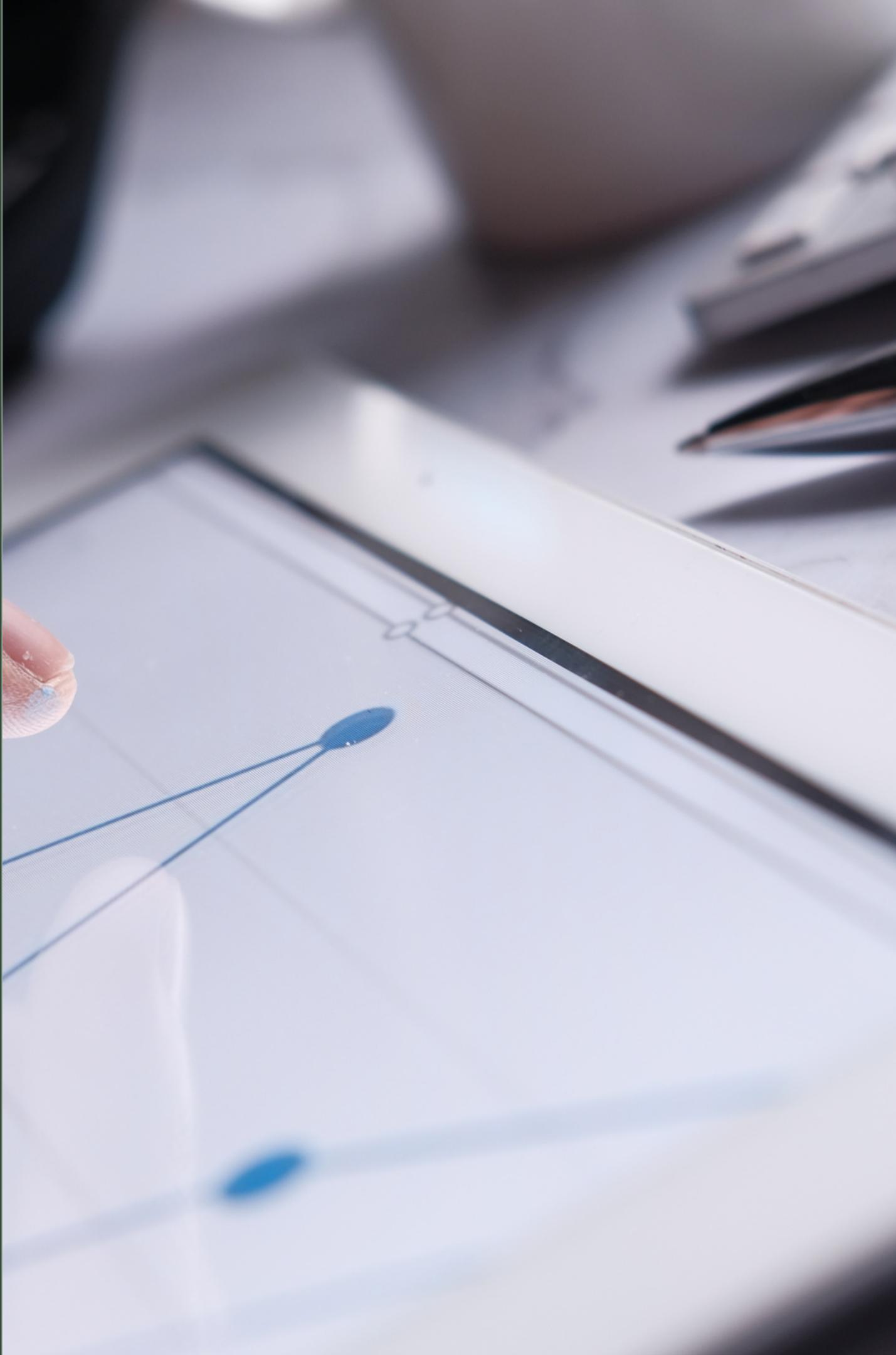
In a tweet announcing the move, Japan's Osaka also said she had "suffered long bouts of depression" since winning her first Grand Slam title in 2018.



To detect signs of depression through tweets so we can encourage these users to seek help

PROBLEM STATEMENT





COLLECTING DATA

DATA

Depressed tweets

viritaromero	Update README.md	db82fd8 on 1 Feb 2019	7 commits
Depression_detection_tweets.ipynb	Add files via upload	3 years ago	
README.md	Update README.md	3 years ago	
depression_detection_tweets.py	Add files via upload	3 years ago	
sentiment_tweets3.csv	Add files via upload	3 years ago	

README.md

Detecting-Depression-in-Tweets

Detecting Depression in Tweets using Baye's Theorem

Depression is a mental illness that is not taken seriously in some countries. I grew up in Mexico and I never got

Harshita9511	Update README.md	c915695 on 14 Aug 2020	4 commits
Depression_Detection_with_Boosti...	boosting	2 years ago	
Depression_Detection_with_Boosti...	boosting	2 years ago	
README.md	Update README.md	13 months ago	
dataset.csv	boosting	2 years ago	
image.jpg	Add files via upload	13 months ago	

README.md

Depression Detection with Boosting Ensemble Learning Classifiers

Positive tweets

 Sentiment140 dataset with 1.6 million tweets
Sentiment analysis with tweets

Marios Mihailidis Kazanova • updated 4 years ago (Version 2)

Data Tasks (1) Code (247) Discussion (18) Activity Metadata Download (239 MB) New Notebook

Usability 8.8 License Other (specified in description) Tags internet, online communities, social networks, linguistics, languages

Description

Context

This is the sentiment140 dataset. It contains 1,600,000 tweets extracted using the twitter api . The tweets have been annotated (0 = negative, 4 = positive) and they can be used to detect sentiment .

Content

It contains the following 6 fields:

1. target: the polarity of the tweet (0 = negative, 2 = neutral, 4 = positive)
2. ids: The id of the tweet (2087)

DATASET

Depressed tweets

- Combine data from both github sources and manually filter to get a cleaner dataset

Combined dataset

- Depressed tweets: 1256
- Positive tweets: 1250

DATA CLEANING

Remove hashtags,
numbers, punctuations,
symbols, stopwords

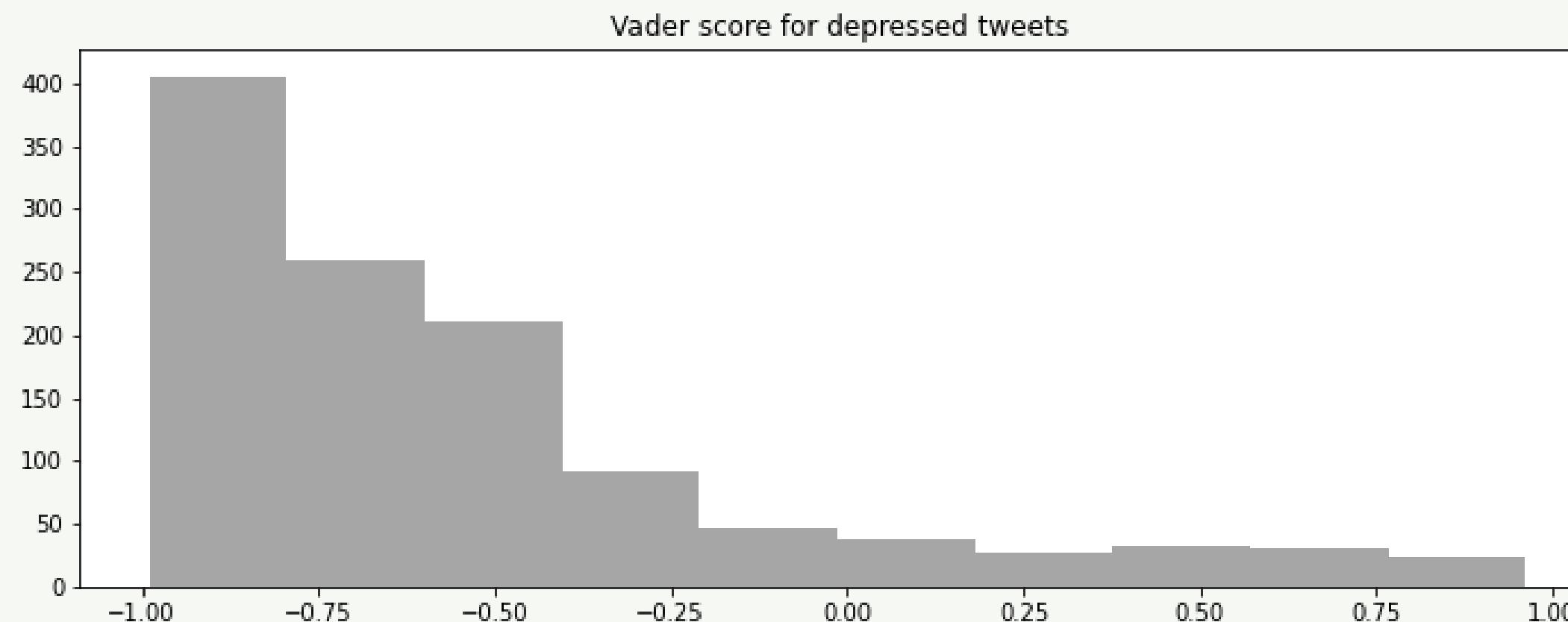
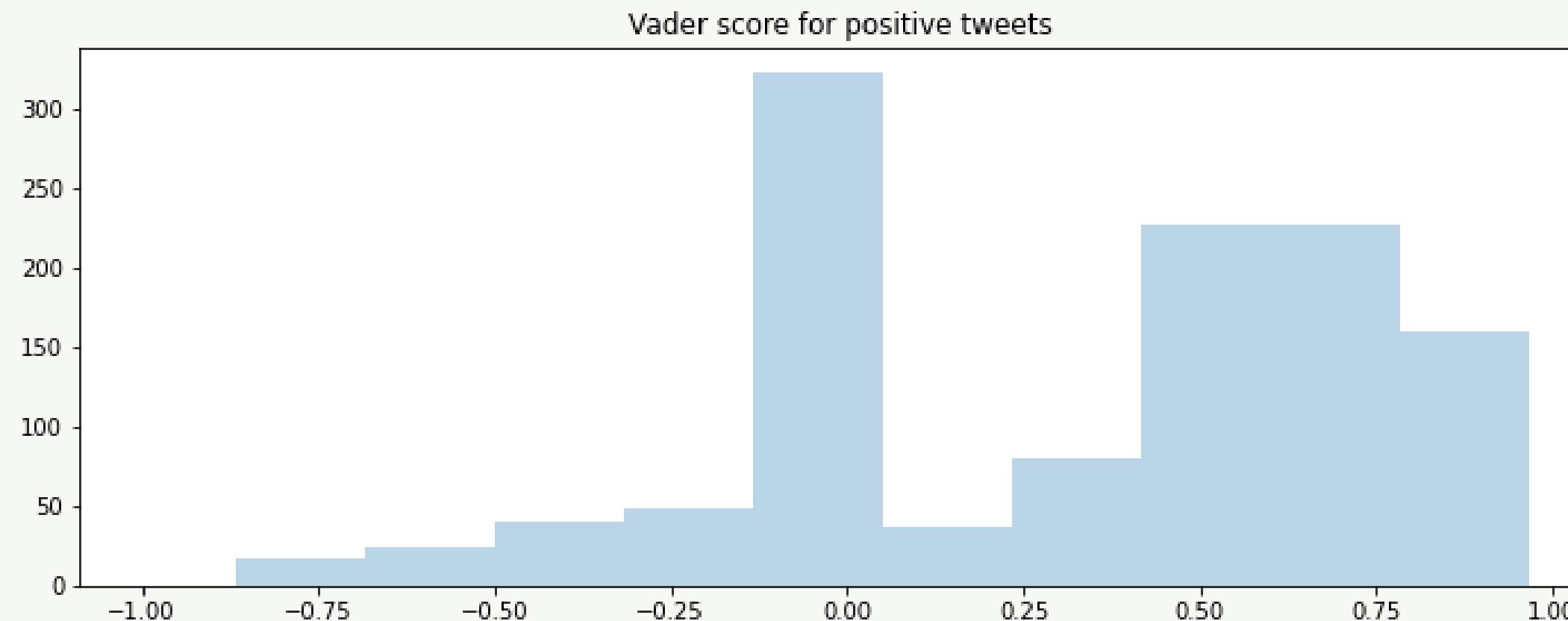
Expand contraction words
eg. I'd --> I would have

Lemmatise to get the
root word to reduce
word duplicates

EXPLORE THE DATA

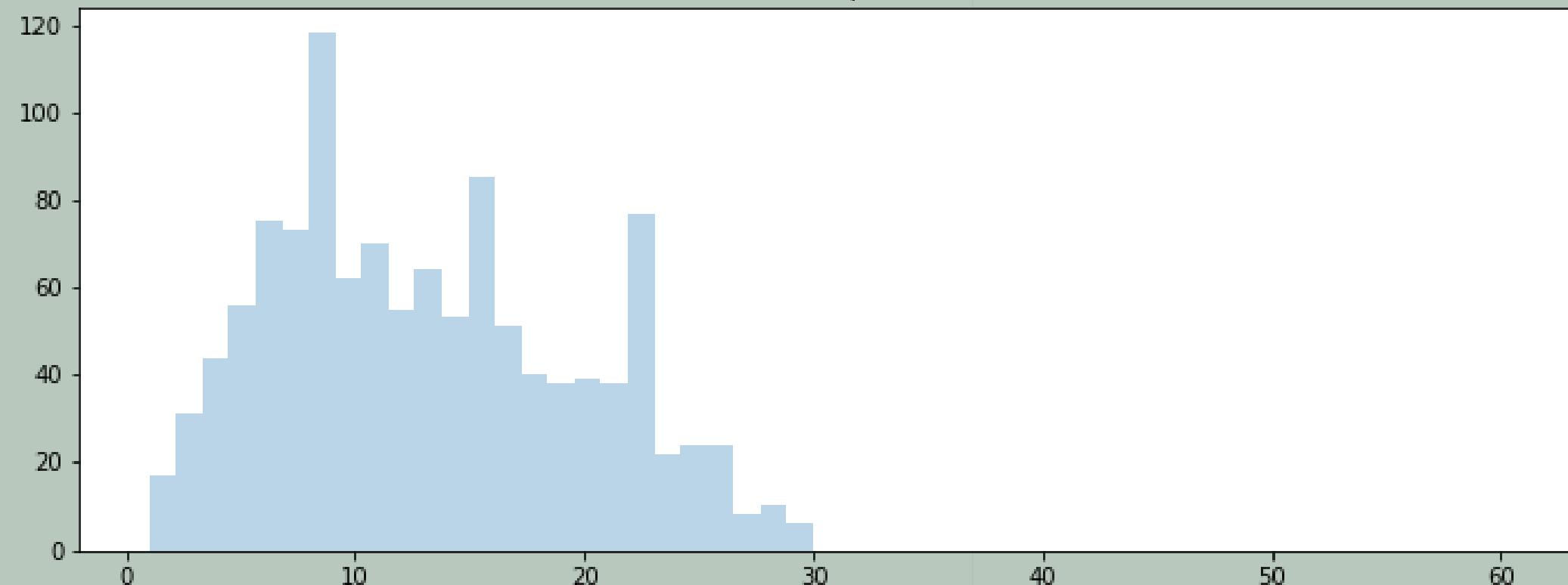


SENTIMENT SCORE

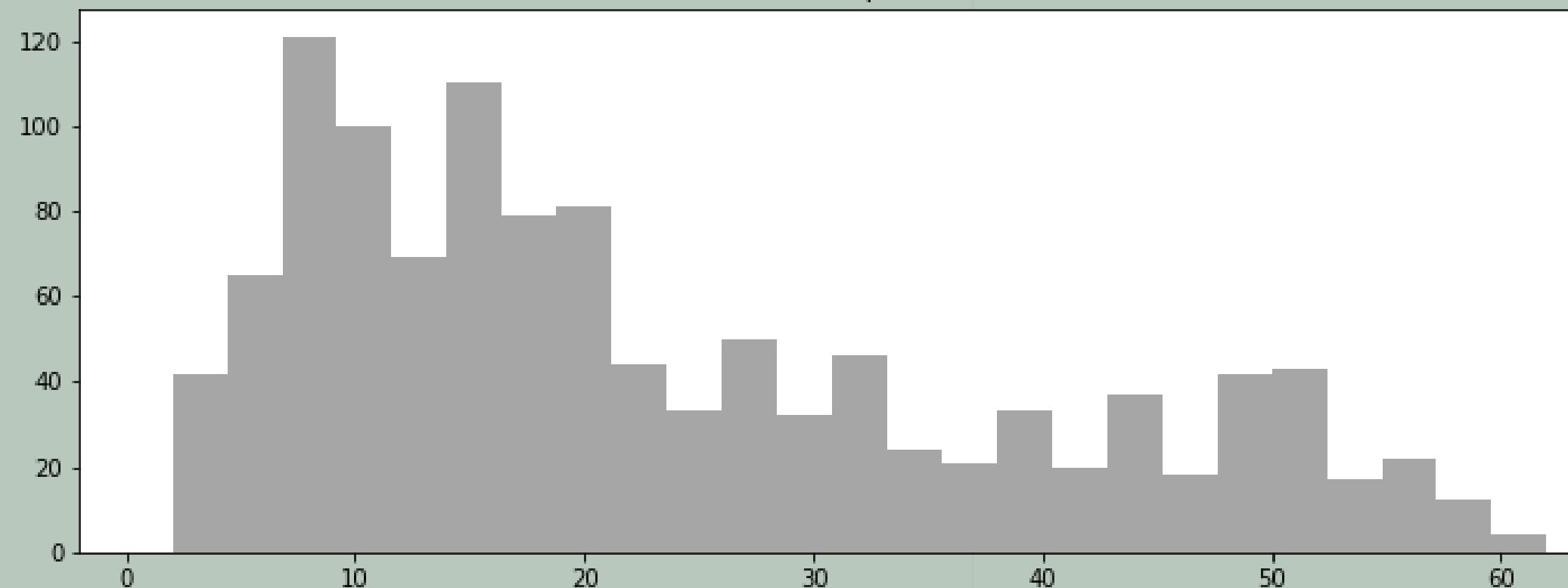


WORD COUNT

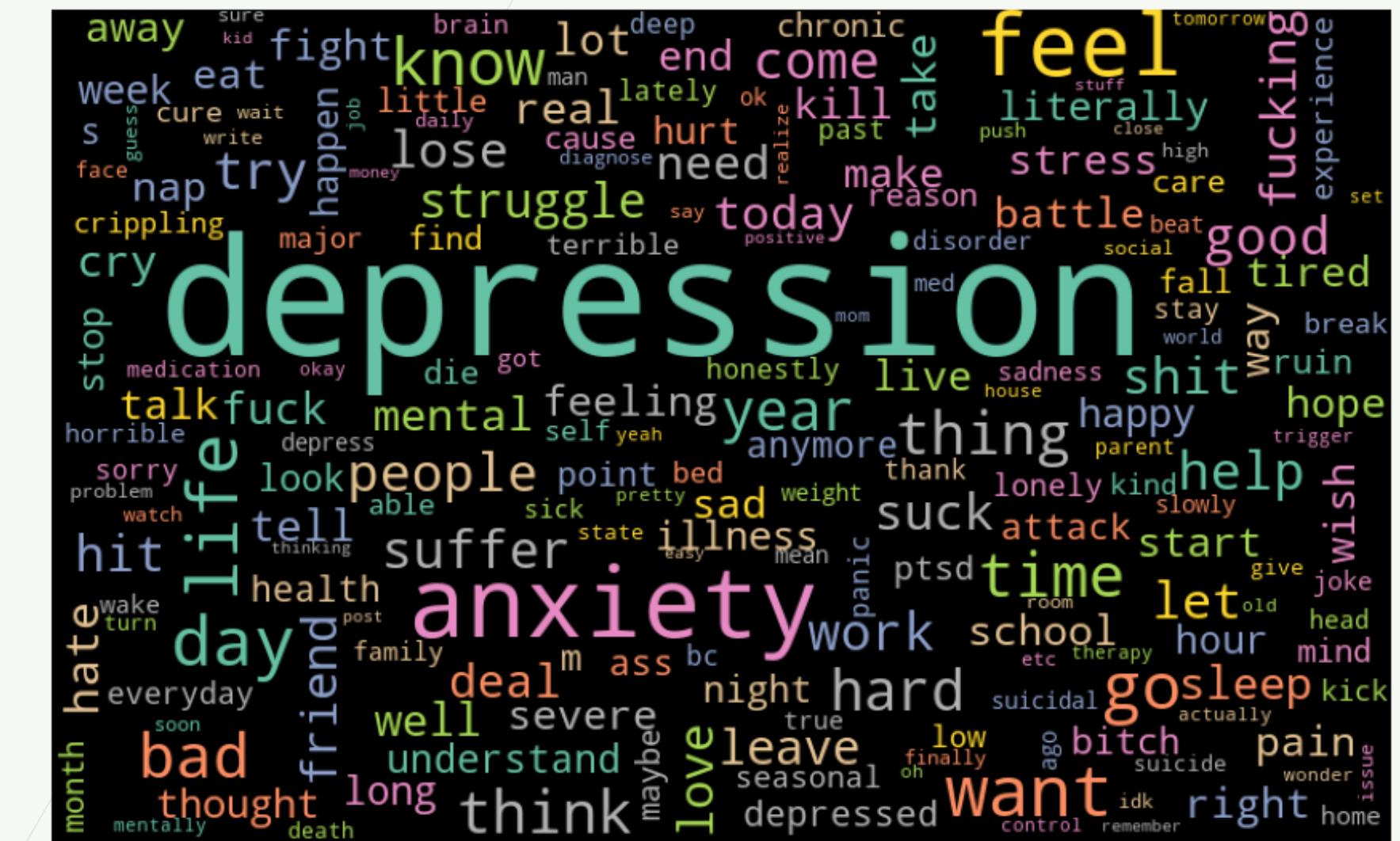
Word count for positive tweets



Word count for depressed tweets

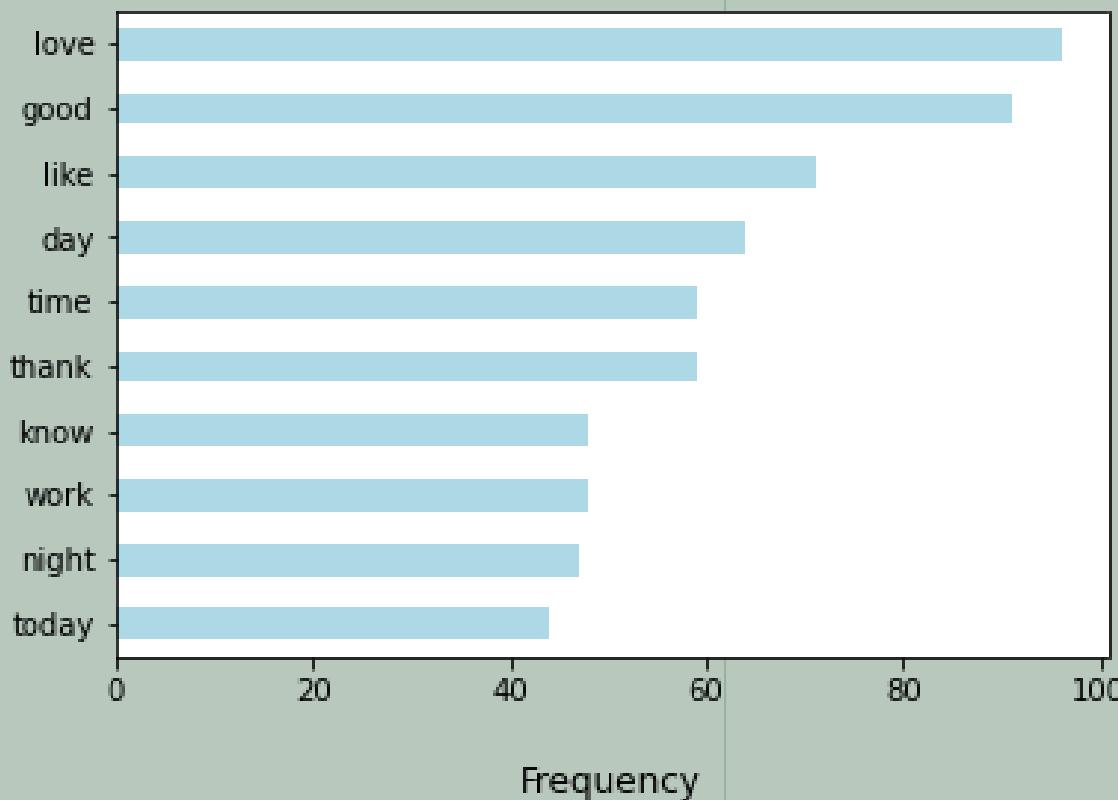


WORD CLOUD

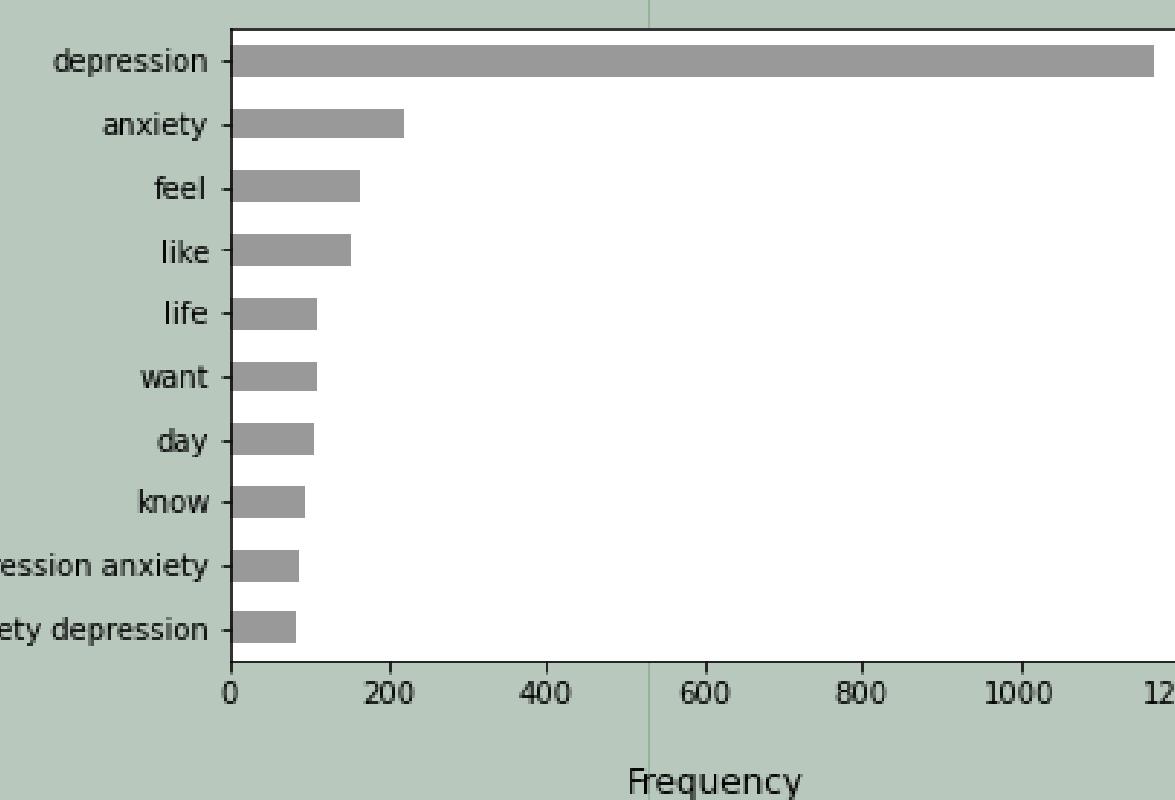


WORD FREQUENCY AND SCORE

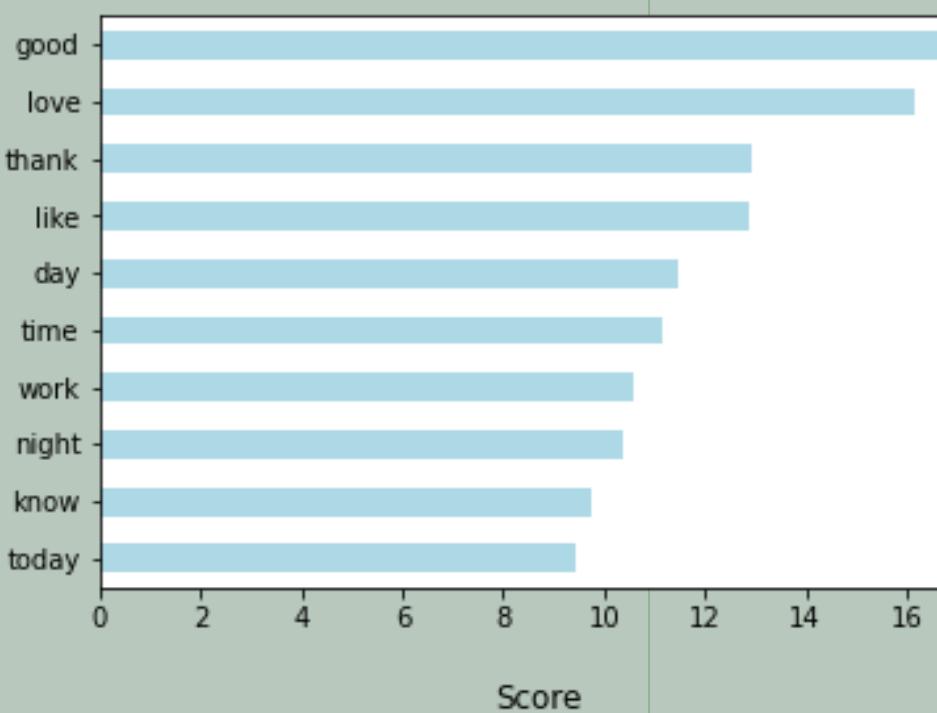
Top 10 occurring words in Positive tweets



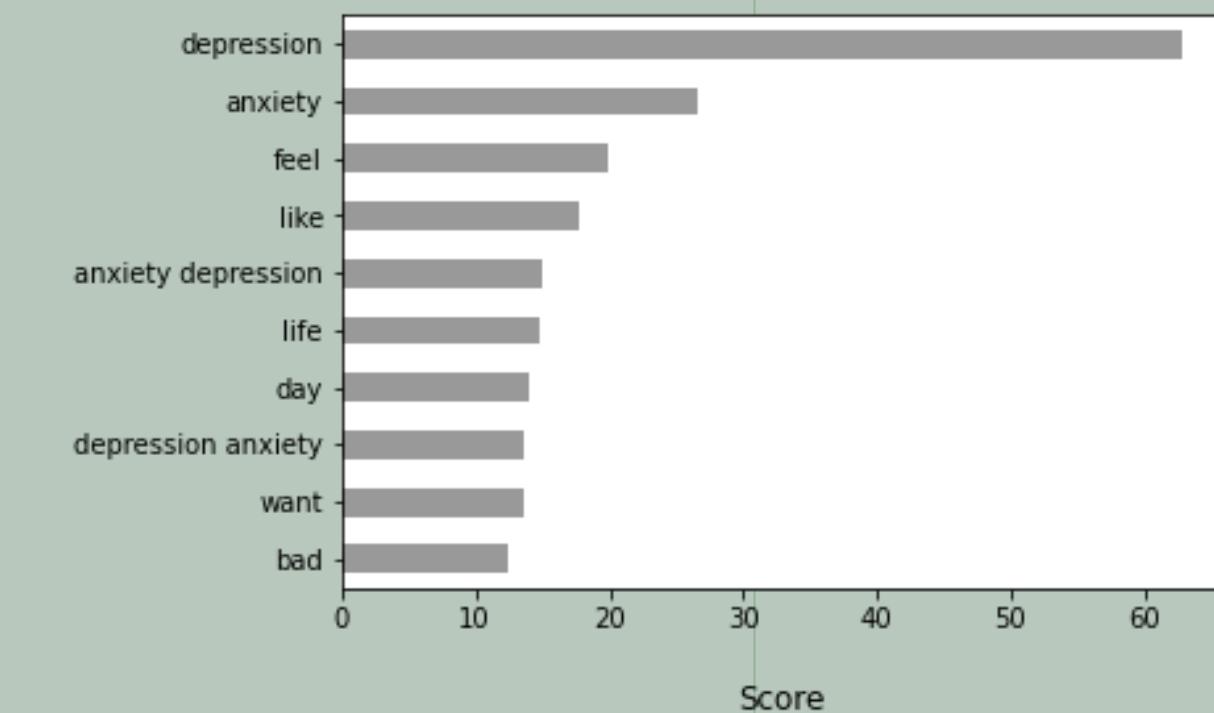
Top 10 occurring words in Depressed tweets



Top 10 occurring words in Positive tweets



Top 10 occurring words in Depressed tweets





MODELING

MODELS USED

Logistic regression

Multinomial
Naive Bayes

Random Forest

LSTM + CNN

MODELING RESULTS

Logistic regression

Train accuracy

0.98

Test accuracy

0.98

Multinomial
Naive Bayes

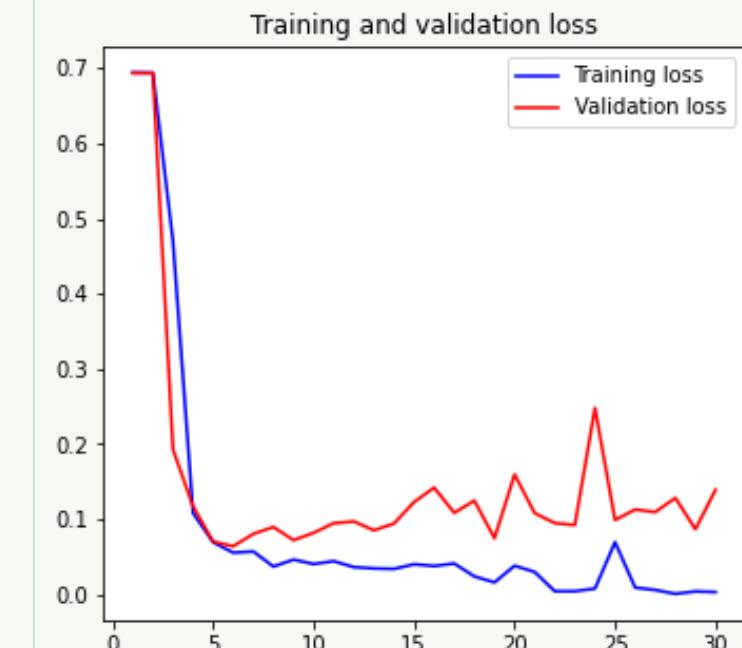
0.95

Random Forest

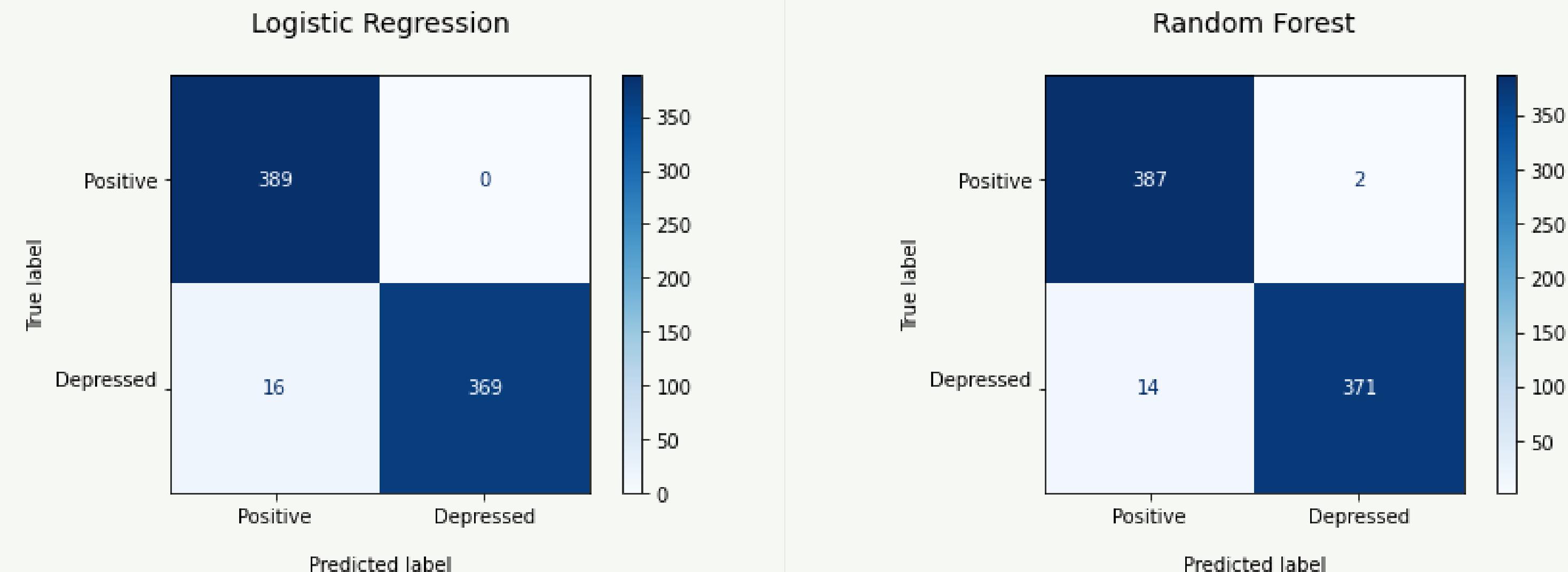
0.98

LSTM + CNN

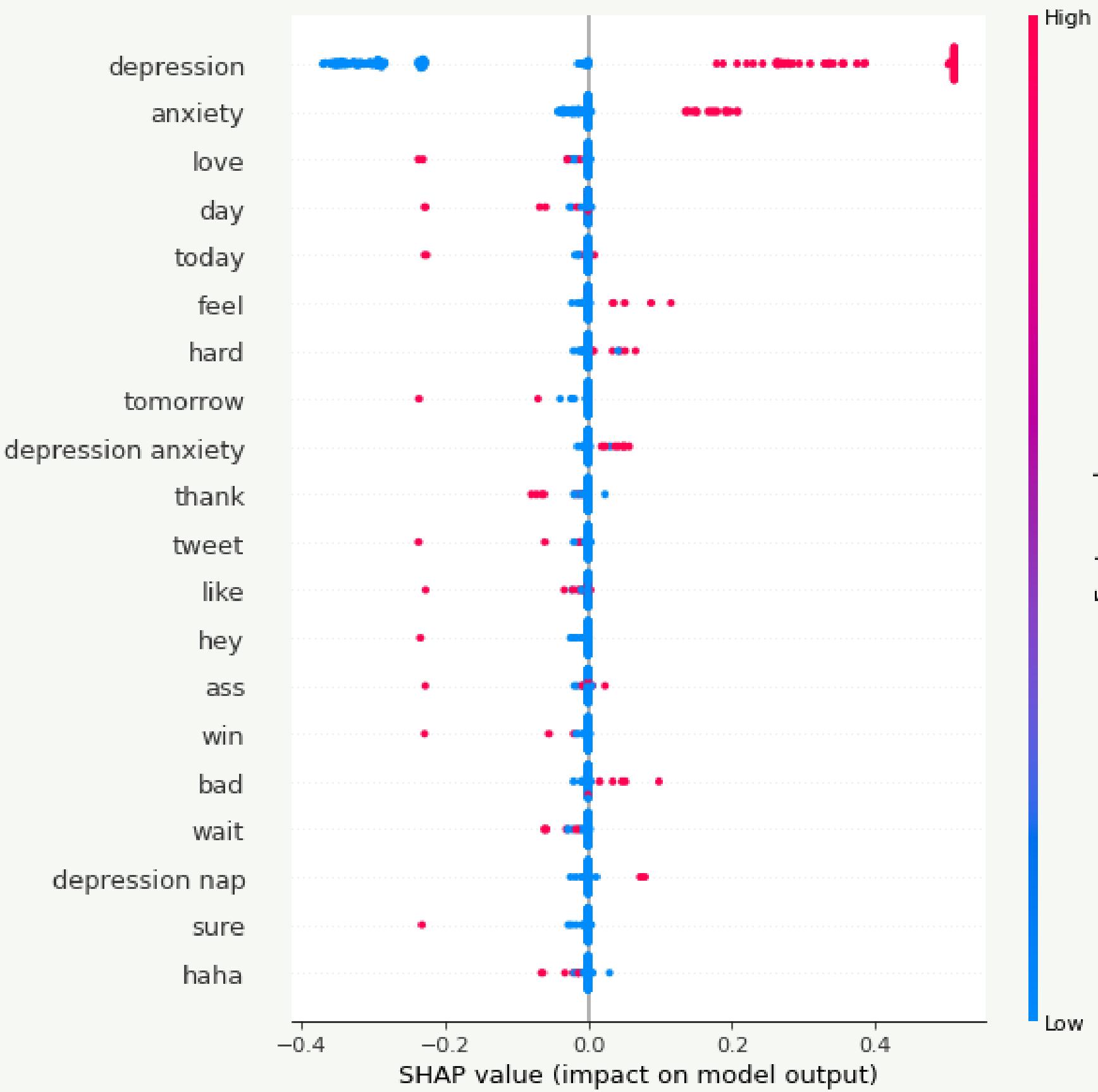
0.99



LOGISTIC REGRESSION VS RANDOM FOREST



RANDOM FOREST - SHAP



MOVING 'DEPRESSION' FROM TWEETS



To evaluate
if the model still works well

RANDOM FOREST SCORE

Train accuracy: 0.80
Test accuracy: 0.78

CONCLUSIONS

- Best model: Random forest
- Look into improving the model without the word 'depression'
- Retrain LSTM + CNN model by using pretrained word embedding such as GloVe
- Explore other mental health condition such as anxiety

THANK YOU!