

COVID-19 Subreddits Classification

r/COVID19_support
r/COVID19positive

Elise Nerissa

Agenda

1. Context
2. Problem Statement
3. Methodology
4. Explanatory Descriptive Analysis (EDA)
5. Model Performance
6. Conclusion & Recommendations

Context

Goal:

Ensure that the posts that appear on the respective subreddits are relevant to the community

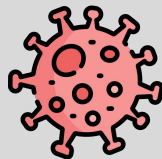


Stakeholders: Reddit Company/ SubReddit Moderators

Subreddits: r/COVID19_support & r/COVID19positive



Problem Statement



1. Identify if a post should belong to COVID-19 Support or COVID-19 Positive Subreddits
2. Identify unique features/words that frequently occur in each subreddit

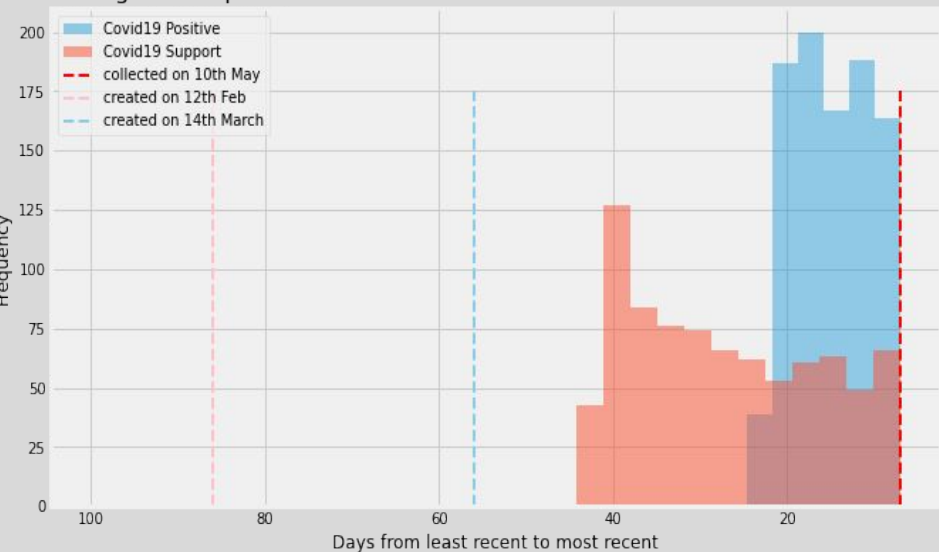
Methodology



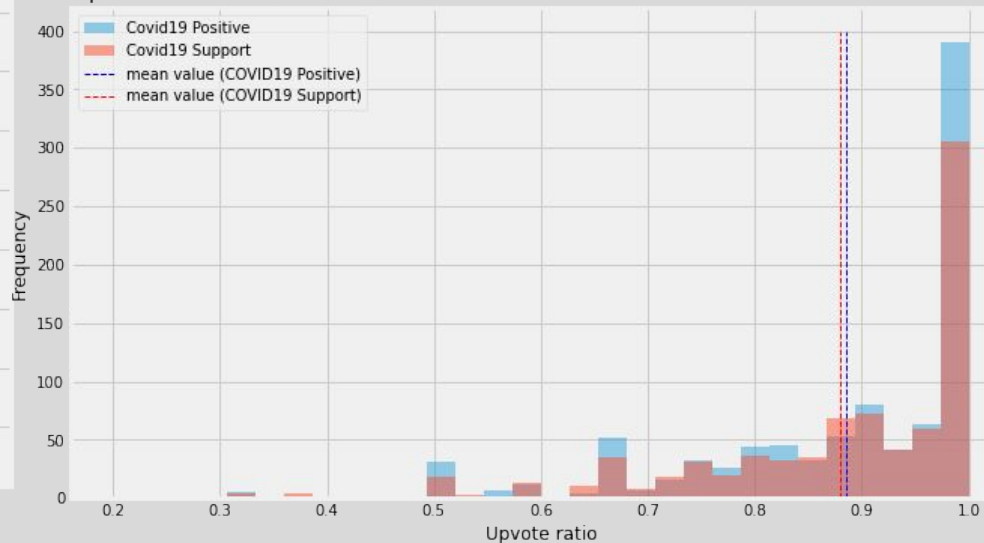
1. Data Collected from **Feb/March 2020 - May 2020**
2. There are about **900 posts from each subreddit**
3. **Title and content of the posts** are used as features for modelling and predictions
4. The text data is processed and parsed using **NLP techniques**
5. **Classification technique** is deployed for modelling and prediction
6. Evaluation metric: **Accuracy score**

Exploratory Data Analysis

The age of the posts



Upvote ratio of each subreddit



Model Performance



Train
Accuracy Score

Test
Accuracy Score

Logistic
Regression

93.1%

90.3%

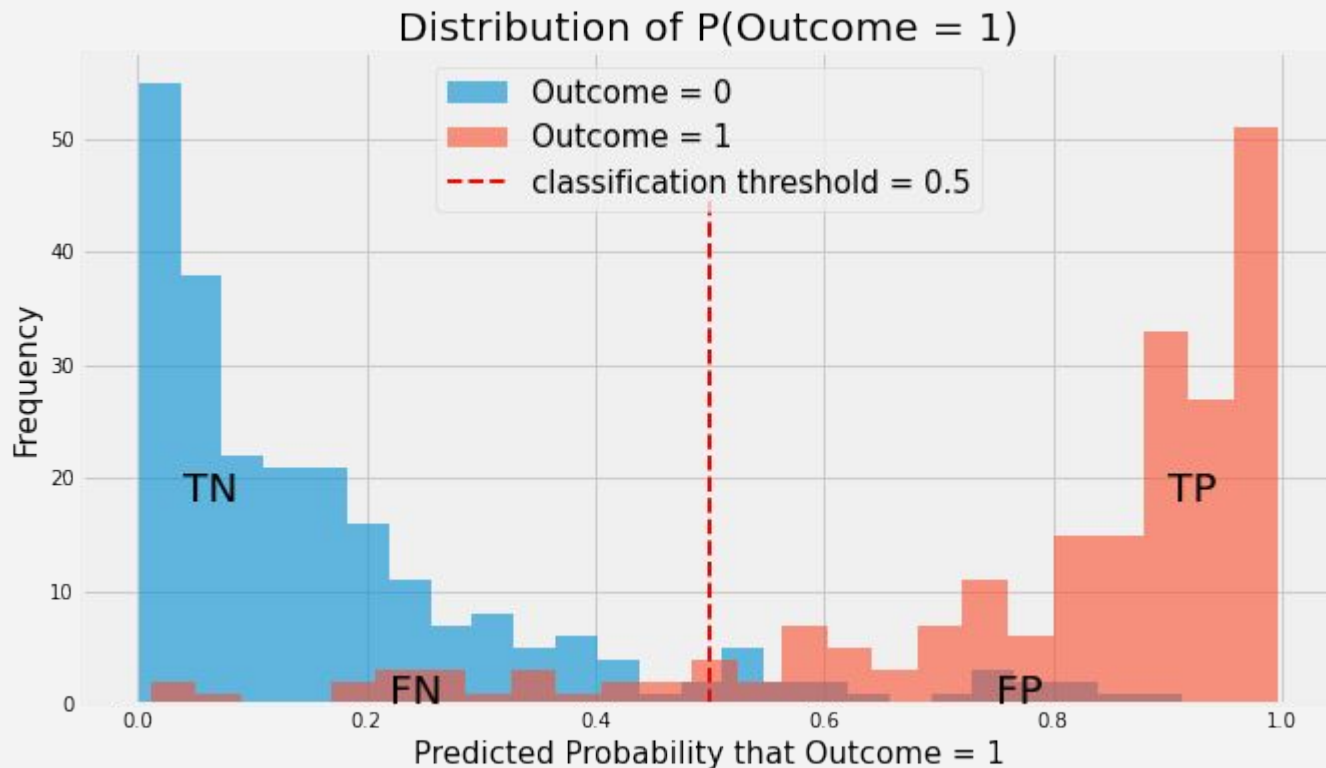
Naive
Bayes

89.8%

90.7%

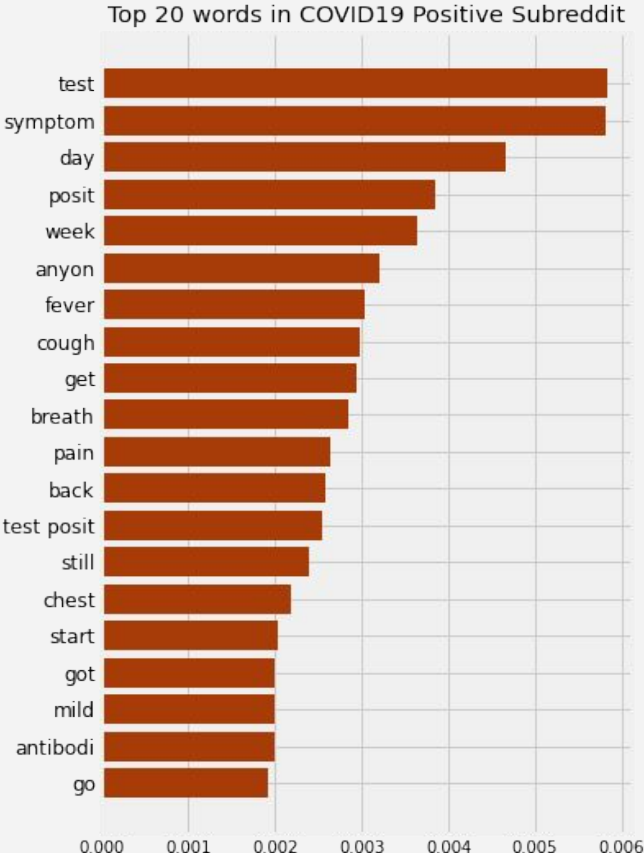


Model Performance

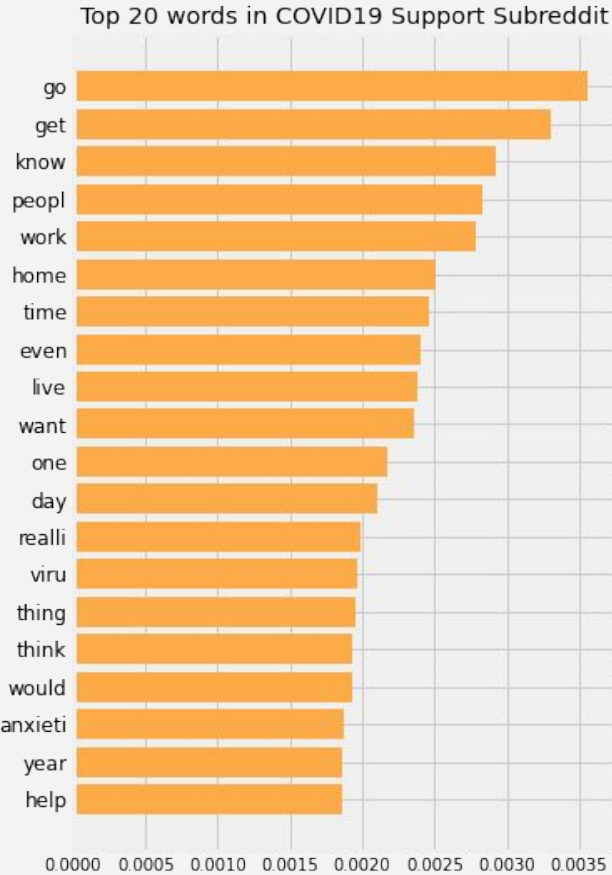


Conclusion

Physical condition



Emotions/Mental state



Recommendations

1. Provide more information in the “About Community” column

Redirect users to the subreddit that is more relevant

2. Deploy the model

Give automated subreddit suggestions to users before users publish their posts

3. Compile and share encouraging posts

Pick up encouraging posts using sentiment analysis techniques and create a tab within subreddit (eg. survivor stories (COVID19 Positive) or ways to make staying at home more bearable (COVID19 Support))



Thank you

r/COVID19_support
r/COVID19positive

Elise Nerissa