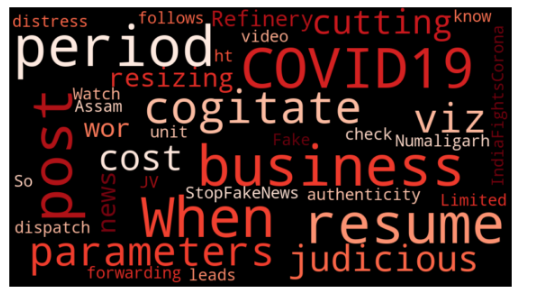
**Covido:Emotionify**

#Indiafightscorona

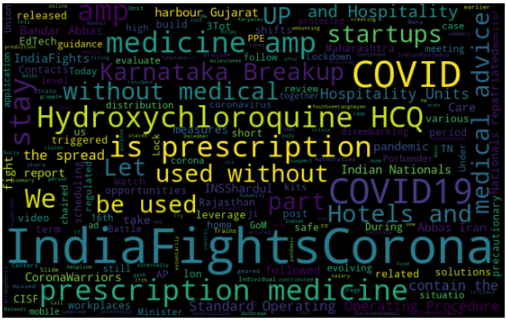
Negative Tweets:



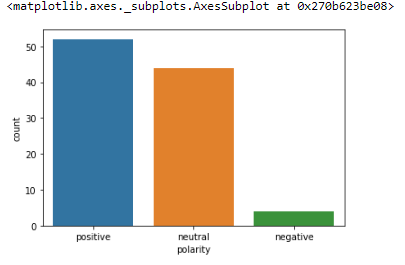
Positive Tweets:



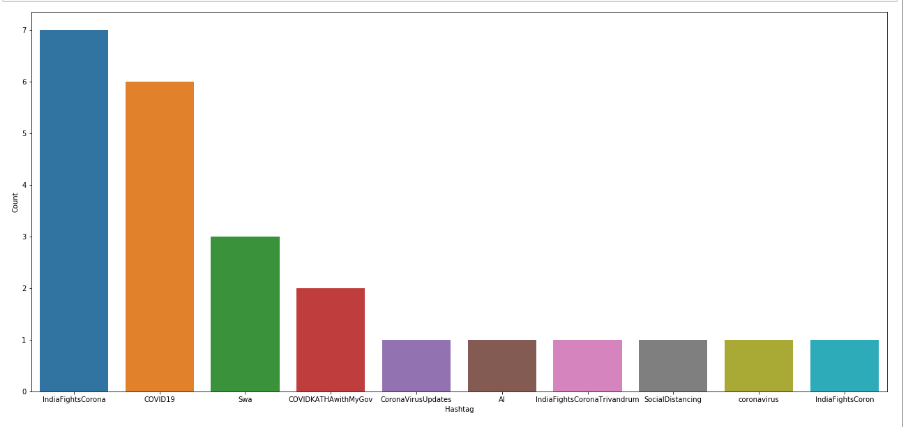
Neutral Tweets

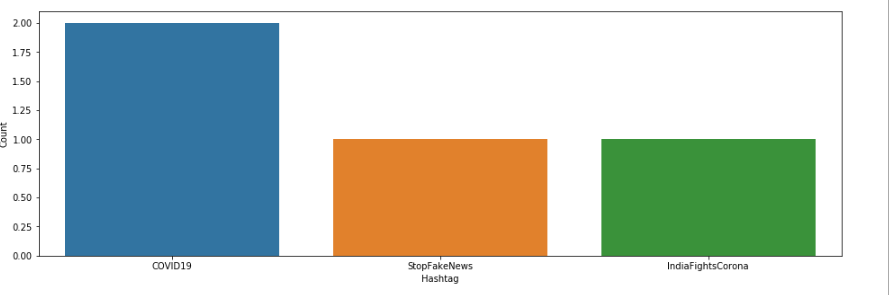


Comparing Number Of Tweets



Selecting Top 10 Positive and Negative Hashtags





A similar analysis has been performed for :

#jantacurfew:

These tweets are more on the positive and neutral side.

#tablighijamat #nizamuddin:

We find more hatred and negativity for these tweets.

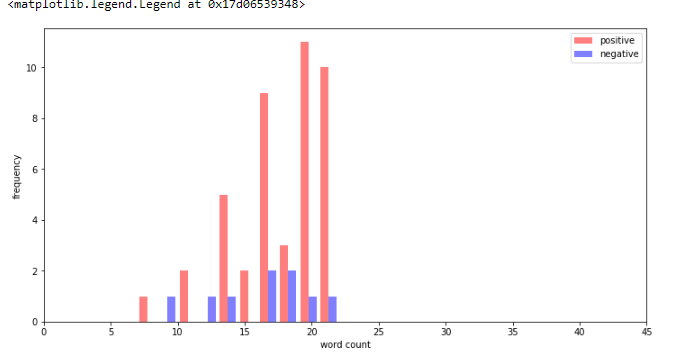
**Lockdown Analysis**

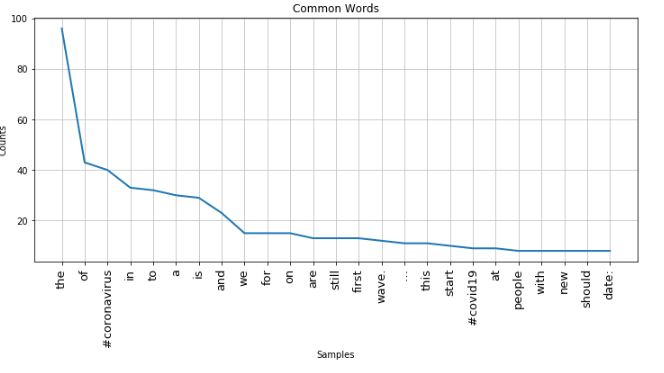
Further to know people’s emotions,I tried collecting data for different phases of lockdown from different countries.

Since it was the beginning (22nd March 2020 ) ,there was more curiosity and positivity among people to fight and defeat the virus.

Due to limitation of time and exams,I could only finish analysis of lockdown 1.

I visualised the lockdown by comparing which words were more frequently used,what were the popular hashtags, and compared the ranking of words.



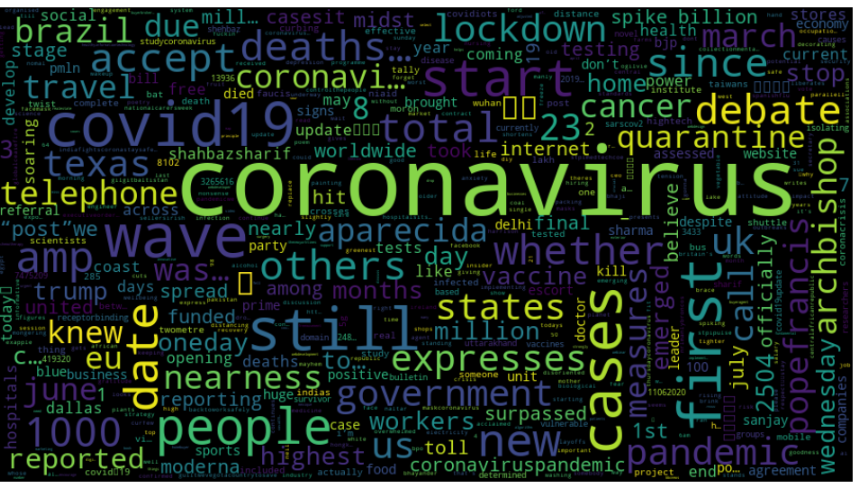


I noticed that sentences fall between 5 to 25 words.

A large proportion is positive.

Generating Wordclouds: For positive,negative and neutral tweets,I generated word clouds.

I also analysed the difference in the wordclouds after removing stop words ,tokenizing and so cleaner data getting displayed.



**Feature Extraction**

* Each message which is represented by a list of tokens is made into a vector that a machine learning model can understand.
* I used Bag of words model to do this.
* Each vector will have as many dimensions as there are unique words in the tweeter corpus.

There were a lot of zero counts. The result is displayed in the form of Sparse Matrix.

TF-IDF was used to find term frequency- inverse document frequency.

**Model Building and evaluation**

I have used a **Naive Bayes (NB)** classifier which is based on the Bayes Theorem.

* To cross-validate and select the best parameter configuration at the same time, I have used GridSearchCV
* It allows me to test out different hyperparameter configurations using for example the KFold strategy to split the model into random parts to find out if it's generalizing well or if it's overfitting.
* All combinations are tested and scored and the best model returned.

I got an accuracy score of 75% for the model tested.