INTRODUCTION

1.1 Overview

The covid-19 had taken the world by storm this year. The first case was recorded in China during December last year. Since then there is no stopping the spread, it went on to become a pandemic to the world. The first cases in india was recorded during March 2020. The government took immediate measures and announced country-wide lockdown to curb the spread of the noval corona virus. Therefore It would be Necessary to Analyze People Sentiments Towards the Government's Actions.

To study public sentiments, we chose Twitter as our target field. As one of the world's biggest social network platforms, Twitter hosts abundant user-generated posts, which closely reflect the public's reactions towards this pandemic with low latency. This is done using a Process called Sentimental Analysis.

Sentiment analysis (also known as opinion mining or emotion AI) refers to the use of natural language processing, text analysis, computational linguistics to systematically identify, extract, quantify, and study affective states and subjective information.

1.2 Purpose

In These times it becomes highly important for the government to know the Public opinions and reactions, So that it can be Considered Before formulating new laws or implementing Further Lockdowns . Hence It becomes neccessary to analyze the Public Sentiments. It allows us to gain an overview of the wider public opinion behind certain topics and also derive Understable patterns and Outcomes that are Useful for the Government or any Other Public/Private Organizations.

LITERATURE SURVEY

2.1 Existing Problem

The Main existing Problem Currently is the gap between the Government and the Public. There is no proper communication about people's opinions and often are misunderstood or doesnot reach the Government Authorities.Similiarly the government doesnot realize if the measures taken reach the people ,or to what extent those measures are implemented on Ground,or even if the public is satisfied with these actions . Therefore there arises a need to study people Sentiments to minimize these problems.

2.2 Proposed Solution

To Build a web application that performs sentimental analysis of Public responses ,on Twitter, to the Pandemic and Government Decisions on extending the Lockdown. The Tweets from the users in response to the Governments decision are scrapped based on Hashtags and Keywords and in the specific Time range.
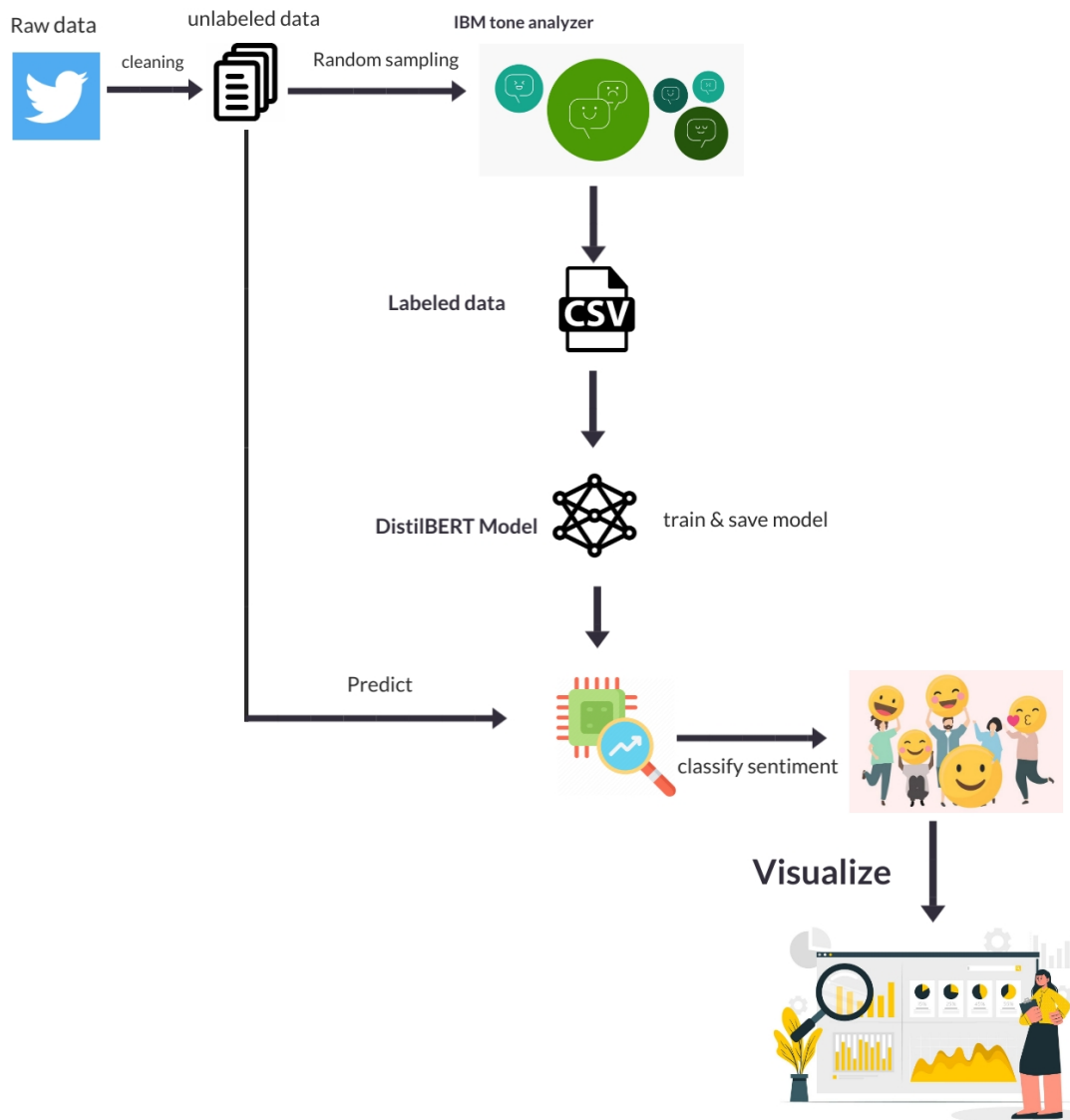Nature Language Processing is performed on the tweets to extract information and it is classified into its respective Sentiment.
Finally, the data is plotted and is visualized on the dashboard.
Thus Providing a Clear Statistical Report on the Public Response.

THEORETICAL ANALYSIS

3.1 Block Diagram

Raw data   unlabeled data   IBM tone analyzer

cleaning   Random sampling

Labeled data

DistilBERT Model   train & save model

Predict   classify sentiment

Visualize

3.2 Software Designing

1. Architecture

- Dash framework is built upon flask , which serves as web app base.
- IBM cloud is used as Web Server
- Files are saved in cloud for access.
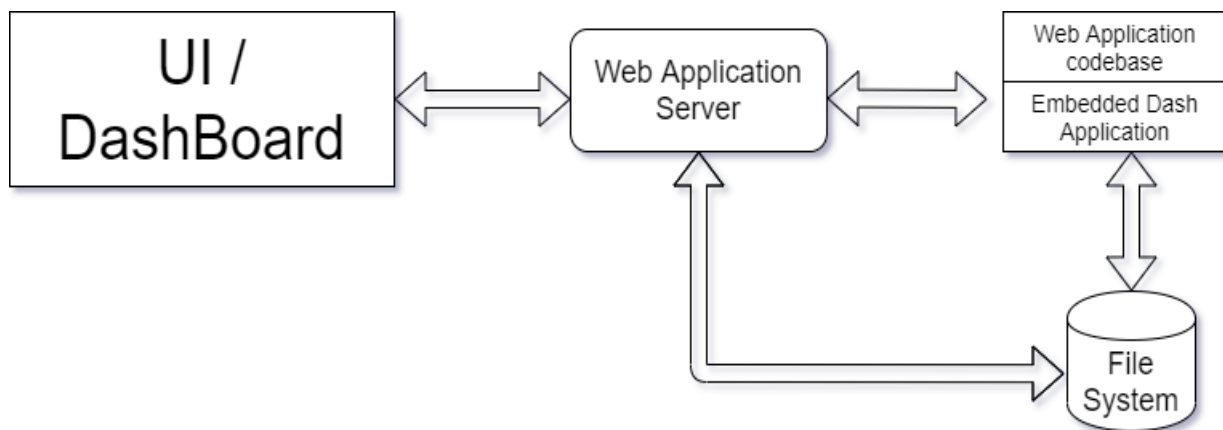
2. Modules

- Data Extraction
- Data Cleaning and pre-processing
- Model training
- Prediction
- Visualization

3. Data

- Tweets and respective Sentiment
- Topic-wise Tweets and Sentiment
- Most frequent Words

EXPERIMENTAL INVESTIGATIONS

1. Data Source

- Twitterscraper API used for scraping tweets based on given time period , using relevent hastags
- for each lockdown we got around 20k tweets,after filtering by english
- Hashtag used to extract relevant tweets are #covid19india, #covid19,#IndiafightsCoronavirus

Topic	No.of tweets
For training each sentiment	7501
Tweets extracted during each lockdown	12000-13000
Total tweets extracted on the whole	60000

Phase	Time Period	No.of Tweets
Lockdown 1	25th March - 14th April (21 days)	**13219**
Lockdown 2	15th April -3rd May (19days)	**13078**
Lockdown 3	4th May - 17th May (14days)	**12981**
Lockdown 4	18th May - 31st May (14days)	**12990**
Total	68 days	**52268**

2. Analytical Study

➤ For Labelling Raw Twitter Data , IBM Tone Analyzer Tool was Used .

➤ IBM tone Analyzer Labels Each tweet with Five different Sentiments ,i.e Fearful , Analytical , Joy , Sad , Anger.
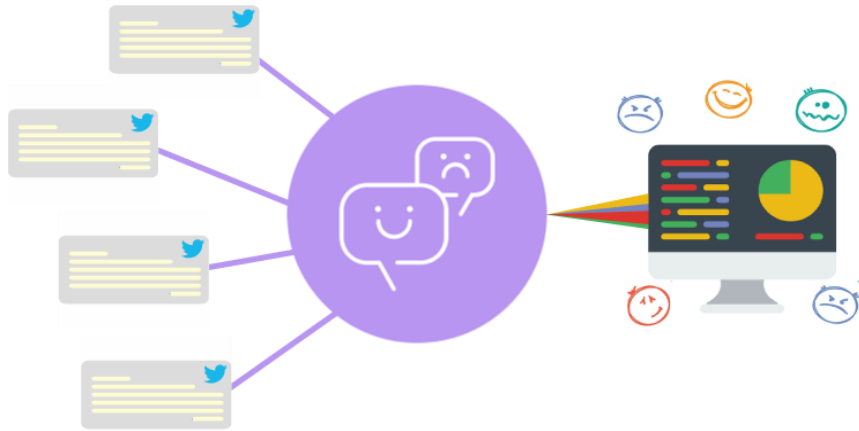
➤ This Labelled Data is Trained Using DistilBERT Model.

➤ DistilBERT is a small, fast, cheap and light Transformer model trained by distilling Bert base. It has 40% less parameters than bert-base-uncased, runs 60% faster while preserving over 95% of Bert's performances as measured on the GLUE language understanding benchmark.

➤ For Each Tweet we performed Binary Classification for Every Sentiment.

➤ This means that each tweet is being predicted for all the Sentiments .Thus We Obtain Multi-Dimensional Sentiments.

➤ Since we use a separate model to classify each Sentiment, Its Accuracy is enhanced.

➤ An Example Prediction is shown Below.

Analytics 🤔
user15 ✓
@user_15

If the number of death in India cross thousand or so like in Italy, Modi may have enough reasons to declare emergency like European nations and control public movement and gathering. #COVID19

5:45 PM · Mar 15, 2020

3 Retweets   233 Likes

Angry 😡
user7 ✓
@user_7

INDIA is worried about Maharashtra & Mumbai in particular, so much so, the Covid19 curve for INDIA depends on the performance of Maharashtra. News is depicting that it is perhaps out of control. Citizen feels that planning & organising of migrant movement is pathetic.

2:15 PM · May 27, 2020

46 Retweets   1.7K Likes

Confident 👍
user3 ✓
@user_3

#Covid19 is a part of our life till the vaccine arrives.It may take 12 to 18 months.We can't be lockdown for so long.We have to come out and combat the virus for our livelihood and adapt all the precautions.We are ready with all the equipment to serve our customers

11:23 AM · Apr 17, 2020

6 Retweets   275 Likes

Sadness 😢
user1 ✓
@user_1

#update: Despite our best efforts, the #COVID19 +ve Pt at MDU, #RajajiHospital, passed away few minutes back.He had medical history of prolonged illness with steroid dependent COPD, uncontrolled Diabetes with Hypertension

9:56 PM · Apr 09, 2020

Fear 😟
user2 ✓
@user_2

Dharavi Corona Community Spread Risk is worrisome to whole nation, Why?

15 Lac people stay here in 2.16 sq km.
Sanitation facilities are shared
Social Distancing is not just possible
Total Count Now at 13
How to deal with this?

9:56 PM · Apr 09, 2020

Joy 🙂
user5 ✓
@user_5

"Finally some good news after a long time
#Covid19 vaccine shows positive results in 8 people in early trials
#Covid_19

12:30 PM · May 15, 2020

< >

TOPIC ANALYSIS

➤ We have also Analyzed Trending Topics Related to the the Government Actions . All the Tweets are Categorized According to these Topics .

➤ All the Tweets of a particular Topic are Predicted for their Sentiment Using Vader Sentiment .Thus We Obtain a Topic-Based Sentimental Analysis.

➤ VADER is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media.

➤ It Provides Three Scores ,Namely Positive ,Negative and the Compound Score.

➤ The compound score is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive). This is the most useful metric if you want a single unidimensional measure of sentiment for a given sentence. Calling it a 'normalized, weighted composite score' is accurate.

3. METRICS

- To Analyze the Sentiment , Certain metrics were Designed which are defined below.

Metric Name	What does it mean?	How do we calculate
Sentiment Level	It Signifies the Intensity of a Sentiment.	(No.of Tweets of a certain Sentiment)/(Total No. of Tweets)
ΔSentiment Level	The Day-to-Day Difference in Sentimental Level	Sentiment level on Day N - Sentiment level on Day N-1
Sentiment Density	The Average Number Of Sentiments Possessed by a Tweet	(Total Count of all the Sentiments) /(Total No.of Tweets)
%Δ in Sentiment Density	The Percentage Change in Daily Sentiment Density.	(Sentiment Density on Day N /Sentiment Density on Day N-1) - 1
% Growth in Covid Cases	The Percentage Change in Daily Covid 19 Cases in India.	(Cases on Day N /Cases on Day N-1) - 1

4.TOOLS USED
- Tweepy- API for extracting Tweets
- Programming : Python3
- Deep Learning Framework - Tensorflow
- Deep learning Model -DistilBERT
- Ploting tool - Plotly
- Web framework- Dash (built upon Flask).
- Deployment Server - IBM Cloud

FLOWCHART
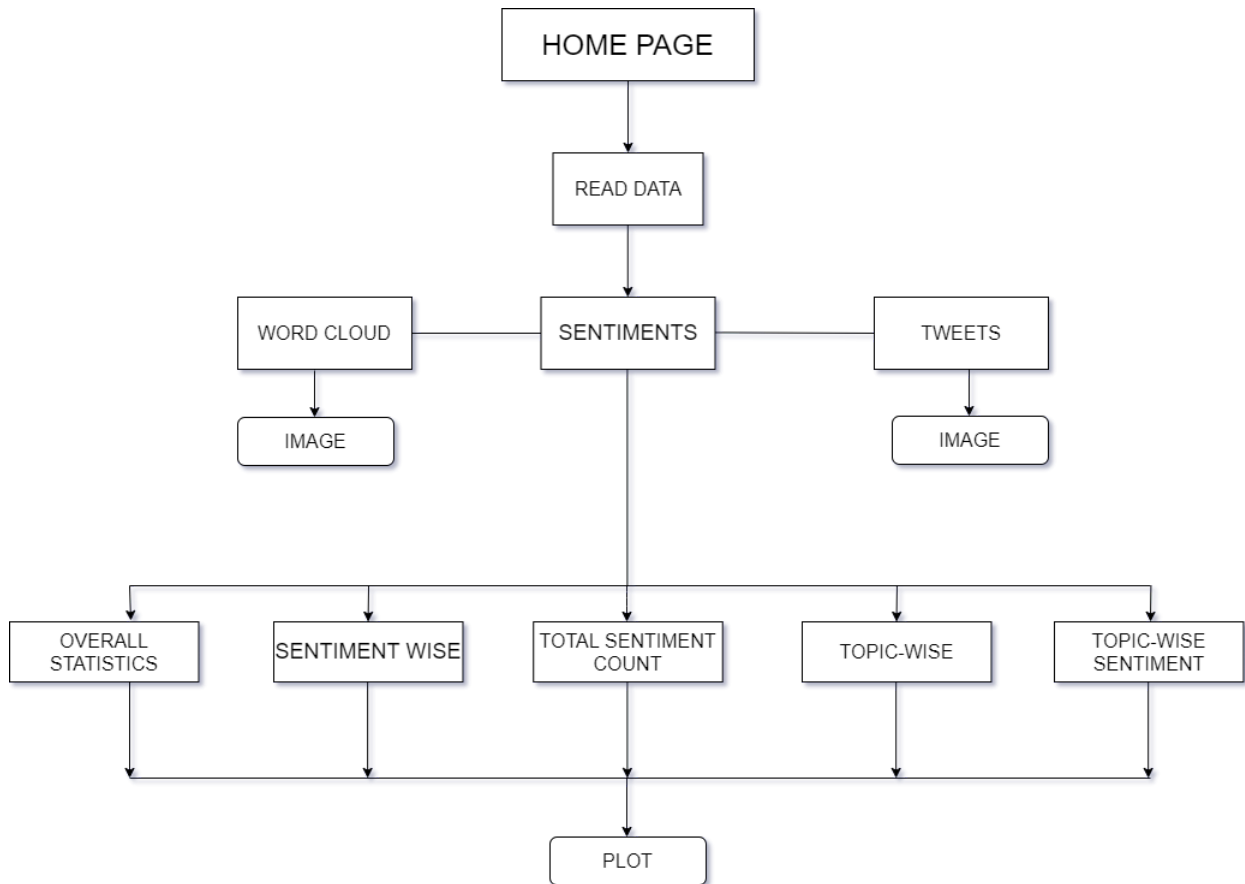
RESULT

After Analyzing the Results From Our Model , We Derived the Following Interesting Outcomes.

1. Sentiment-Density Trends.

- Growth Rate = ((No. of Cases on Present Day - No. of cases on previous day) /(No. of Cases on Previous Day)) * 100
- The above graph shows sentiment-wise variations with the percentage of rise in covid cases.
- Significant percentage of the Tweets are Analytical in Nature, An analytical tone indicates a persons reasoning and analytical attitude about things. An analytical tweet might be perceived as intellectual, rational, systematic, emotionless, or impersonal.

- We can see Serious Anger among the Public during the Lockdown, But we also see Notable Confidence with the Tweets . This Signifies the presence of People from Both the sides i.e those embracing the Lockdown and also presence of those against the governments actions.
- Sentiments such as Sadness and Fear show similar trends among Users,which has gradually increased Over time.
- We also Observe tweets with Joy during these times,it mostly indicates the response to Vaccine developments, appreciation for frontline workers ,government actions and also the Strong meme game of our country.

2. TOTAL SENTIMENT COUNT

Comparison of Total Tweets of Respective Sentiments

▶ We Observe That Most of the Tweets are Analytical in Nature i.e About 17000 Tweets are Analytical. It establishes that most of the Tweets are Factual and Rational In Nature .

▶ The Sentiments Anger and Confidence have almost the same Number of Tweets. It indicates the Representation of Users from Opposite Perspectives and Opinions. We conclude that as much as Lockdown was Received ,It also Received Equal amounts of anger and dissatisfaction from the Public.

⚑ There are about 10,000 Tweets from Users during the time Period with Sentiments expressing Fear and Sadness and Also About 6000 tweets expressing Joy . It shows that a Section of the Public have Been affected by the Lockdown Measures.

3. TRENDING TOPICS

Number of Tweets Topic-Wise
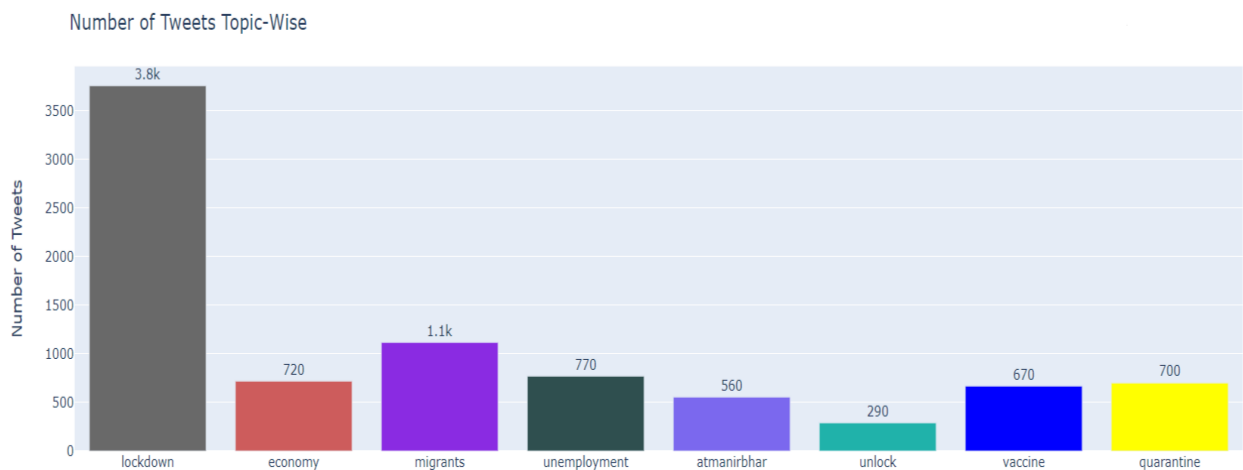
⚑ From the Extracted Twitter Data ,It is Seen That many have mentioned "Lockdown" in their Tweets. An Expected and Obvious Trend.

⚑ Also The Topics like "Migrants"," Unemployment" and "Economy" were highly Discussed About during the Lockdown as they were some of the important Issues.

⚑ Also Some Events and packages were disscussed including "Aatmanirbhar", "Vaccine", "Unlock" and Ofcourse "Quarantine".

⚑ These were some of the Important Trending Topics that were Evident By the Tweets. So we wanted to Dwell deep into the Topics and Analyze their Sentiments for further Understanding . Outcomes from This Topic-Analysis is Shown Below .

4. TOPIC ANALYSIS

Topic-Wise Sentiment

Range — Negative - [0,1] | Positive - [0,1] | Compound - [-1,1]
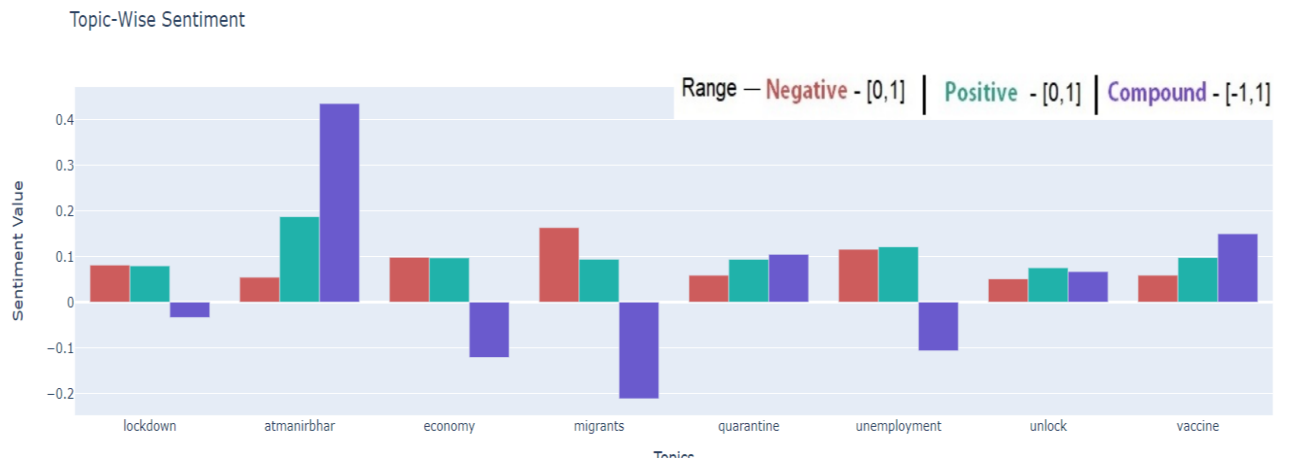
- ▶ The Above Graph Clearly Depicts the Overall Sentiment of the Users on These Topics.
- ▶ 'AatmaNirbhar' is the name given to the covid relief package announced by PMO India, which has highly appreciated and accepted positively on twitter.
- ▶ The people on twitter Raged on the 'Economy' , 'Unemployability' and 'Migrant' issues and expressed their Anger and Dissatisfaction on the Government.
- ▶ 'Vaccine' and 'Unlock' Topics have received Good response and is Welcomed by people on Twitter.

5.WORD CLOUD

▶ In the above picture we can observe few most frequently used words in twitter with respect to covid19 in india.

▶ We observe that many words are related to government actions, we can conclude that people have talked a lot about decisions or actions made by government. It Gives us a Clear Picture about What People Discussed Regarding the Government Measures.

6. TWEETS AND THEIR PREDICTIONS

▶ The above are some of tweets from the extracted data and their sentiment predicted by our Model.

▶ We can notice that the sentiments are predicted pretty accurately by the Model.

ADVANTAGES

▶ The Main Advantage is that the Government Gets a Clear Understanding of the Public Responses on Twitter.

▶ Instead of Classfying into just Negative or Positive Polarities, Each Tweet is sent for Prediction through Multiple Sentiment-Specific Models.This means that each tweet is being predicted for all the Sentiments.Thus We Obtain Multi-Dimensional Sentiments.

▶ Since we use a separate model to classify each Sentiment, Its Accuracy is enhanced.

▶ We Get to See the Variations in Sentiments during each Phase of Lockdown and their Announcements ,Thus obtaining a clear idea of How People Might React for Further Lockdowns in the Future.

▶ It can also be used during other Emergency Situations by the Government.

DISADVANTAGES

▶ The Extracted Data is English Language Specific, So the Regional Languages are Filtered Out.

APPLICATIONS

▶ It can be used by government or private organizations for knowing peoples opinions on certain actions and current dilemma.
▶ Private companies can also use this for understanding sentimental state,economic status of the people on social media , so that they can decide on actions on their release,marketing, production of the products.

CONCLUSION

▶ In this project, we analyzed the sentiments of COVID-19-related tweets in several ways. The overall trend shows that the public has been more optimistic over time. Digging into the multi-dimensional sentiment analysis, we found that the sentiments increase on announcements and attain peak during the beginning of lockdown. Besides, the Sentiment Density indicates that the public turned out to be less loaded with emotions. At last, the Topics behind the sentiments unfolded more details.
▶ To fight the coronavirus not only needs the guidance from the government but also a positive attitude from the public. Our analysis provides a potential approach to reveal the public's sentiment status and help institutions respond timely to it.

FURTHER SCOPE

Our analysis has shown some relationships between confirmed cases' growth and the trends of sentiments. With more granular data such as geographic data, demographic information, and so on, further insights can be generated, such as public sentiment monitoring the hardest-hit areas. With a more specific target, the analysis would be more valuable for institutions or governments to take action.

WEBLIOGRAPHY

🚩 https://dash.plotly.com/

🚩 https://plotly.com/python/

🚩 https://nbviewer.jupyter.org/github/amaiya/ktrain/blob/master/tutorials/tutorial-04-text-classification.ipynb

🚩 https://cloud.ibm.com/docs/tone-analyzer?topic=tone-analyzer-about

APPENDIX

A. Source code

▶ **Scraping twitter**

```python
from twitterscraper import query_tweets
import datetime as dt
import pandas as pd
bg_time = dt.date(2020,5,31)
ed_time = dt.date(2020,6,24)
# begindate=bg_time,enddate=ed_time
language = 'english'
tweets= query_tweets('covid19
india',begindate=bg_time,enddate=ed_time,lang=languag
e)
tweets= pd.DataFrame(t.__dict__ for t in tweets)
tweets.to_csv('tweets-5-june.csv',index=False)
```

▶ **Model training**

```python
from sklearn.model_selection import train_test_split

sadxtrain, sadxtest, sadytrain, sadytest =
train_test_split(df['clean_text'],
fxtrain, fxtest, fytrain, fytest =
train_test_split(df['clean_text'],
axtrain, axtest, aytrain, aytest =
train_test_split(df['clean_text'],
jxtrain, jxtest, jytrain, jytest =
train_test_split(df['clean_text'],
cxtrain, cxtest, cytrain, cytest =
train_test_split(df['clean_text'],
anlxtrain, anlxtest, anlytrain, anlytest =
```

```python
train_test_split(df['clean_text'],
df['analytical'],shuffle=True)


import ktrain
from ktrain import text
MODEL_NAME = 'distilbert-base-uncased'
t = text.Transformer(MODEL_NAME, maxlen=100,
class_names=['1','0'])
trn = t.preprocess_train(sadxtrain, sadytrain)
val = t.preprocess_test(sadxtest, sadytest)
model = t.get_classifier()
learner = ktrain.get_learner(model, train_data=trn,
val_data=val, batch_size=6)
learner.fit_onecycle(8e-5, 4)
learner.validate(class_names=t.get_classes())
predictor = ktrain.get_predictor(learner.model,
preproc=t)


#import json
model_json = learner.model.to_json()
with open("sadmodel.json", "w") as json_file:
    json_file.write(model_json)


learner.model.save_weights("model.h5")
print("Saved model to disk")
from keras.models import load_model
newmodel=load_model('sadmodel.h5')
```

▶ **Predict**

```python
import numpy as np
import pandas as pd
```

```python
import os,sys
import ktrain

analytical=ktrain.load_predictor('/content/drive/My
Drive/IBM data/analytical pred')
anger=ktrain.load_predictor('/content/drive/My
Drive/IBM data/anger pred')
confident=ktrain.load_predictor('/content/drive/My
Drive/IBM data/confident pred')
sadness=ktrain.load_predictor('/content/drive/My
Drive/IBM data/sad pred')
fear=ktrain.load_predictor('/content/drive/My
Drive/IBM data/fear pred')
joy=ktrain.load_predictor('/content/drive/My
Drive/IBM data/joy pred')
pred=pd.read_csv('/content/drive/My Drive/IBM
data/mar-april.csv')
import time
txt=pred['text'].tolist()
st=time.time()
pan=analytical.predict(txt)
panger=anger.predict(txt)
pconfi=confident.predict(txt)
psad=sadness.predict(txt)
pfear=fear.predict(txt)
pjoy=joy.predict(txt)
pred.head()
pred.to_csv('/content/drive/My Drive/IBM
data/mar-april-predictedres1.csv', sep=',',
encoding='utf-8')
```

▶ Web app

```python
import dash
import pandas as pd
import dash_core_components as dcc
import dash_html_components as html
from dash.dependencies import Input, Output
import plotly.express as px
from plotly.subplots import make_subplots
import plotly.graph_objects as go
import dash_bootstrap_components as dbc


@app.callback(Output("page-content", "children"),
[Input("url", "pathname")])
def render_page_content(pathname):
    if pathname in ["/", "/page-1"]:
        pag1 = html.Div([
            html.H2(children='Covid19 Twitter
Analysis India',
                    style={'text-align': 'center'}),
            dcc.Tabs(id="tabs",
                    value='tab-1',
                    children=[
                        dcc.Tab(label='Overall
Statistics', value='tab-1'),
                        dcc.Tab(label='Sentiment
Wise', value='tab-2'),
                        dcc.Tab(label='Total
Sentiment Count', value='tab-3'),
                        dcc.Tab(label='Topic Wise',
value='tab-4'),
                        dcc.Tab(label='Topic Wise -
```

```python
Sentiment', value='tab-5')
                    ]),
          html.Br(),
          #dcc.Graph(id='date-line'),
          html.Div(id='date-line')


    ])
    return pag1
```