SENTIMENT ANALYSIS OF COVID-19 TWEETS

TEAM NAME   :   CATALYST

TEAM MEMBERS :   UNNIMAYA M
                 ANANYA R
                 MANASA RADHAKRISHNAN

INTRODUCTION

OVERVIEW

The sentiments of Indians after the extension of lock down announcements is analyzed with the relevant tags on twitter and a predictive analytics model is built to understand the behavior of people if the lock down is further extended. A dashboard with visualization of people's reaction to the government announcements on lock down extension is also developed.

The twitter sentiment analysis model helps to understand the following:

- Get to know people's sentiment towards the epidemic.
- Understand the sentiments of people on government decision to extend the lock down

PURPOSE

The following are the purposes of our project -

- Our project will show the different sentiments in the form of histogram.
- After each live telecast about the status of lock down, the histogram will be updated according to the live tweets.
- A comparison of the old and updated histograms will be shown side by side.
- On clicking each part of the graph, more details about the tweets will be shown.

The latest updates about the corona virus and related topics will be displayed on a single website to make it easier to observe the changes related to this socially relevant topic.

LITERATURE SURVEY

EXISTING PROBLEM
- The existing systems have functions that allow the user to enter the hashtags and the amount of tweets related to that topic be displayed.
- They do not show a sentimental view of the situation like our project.
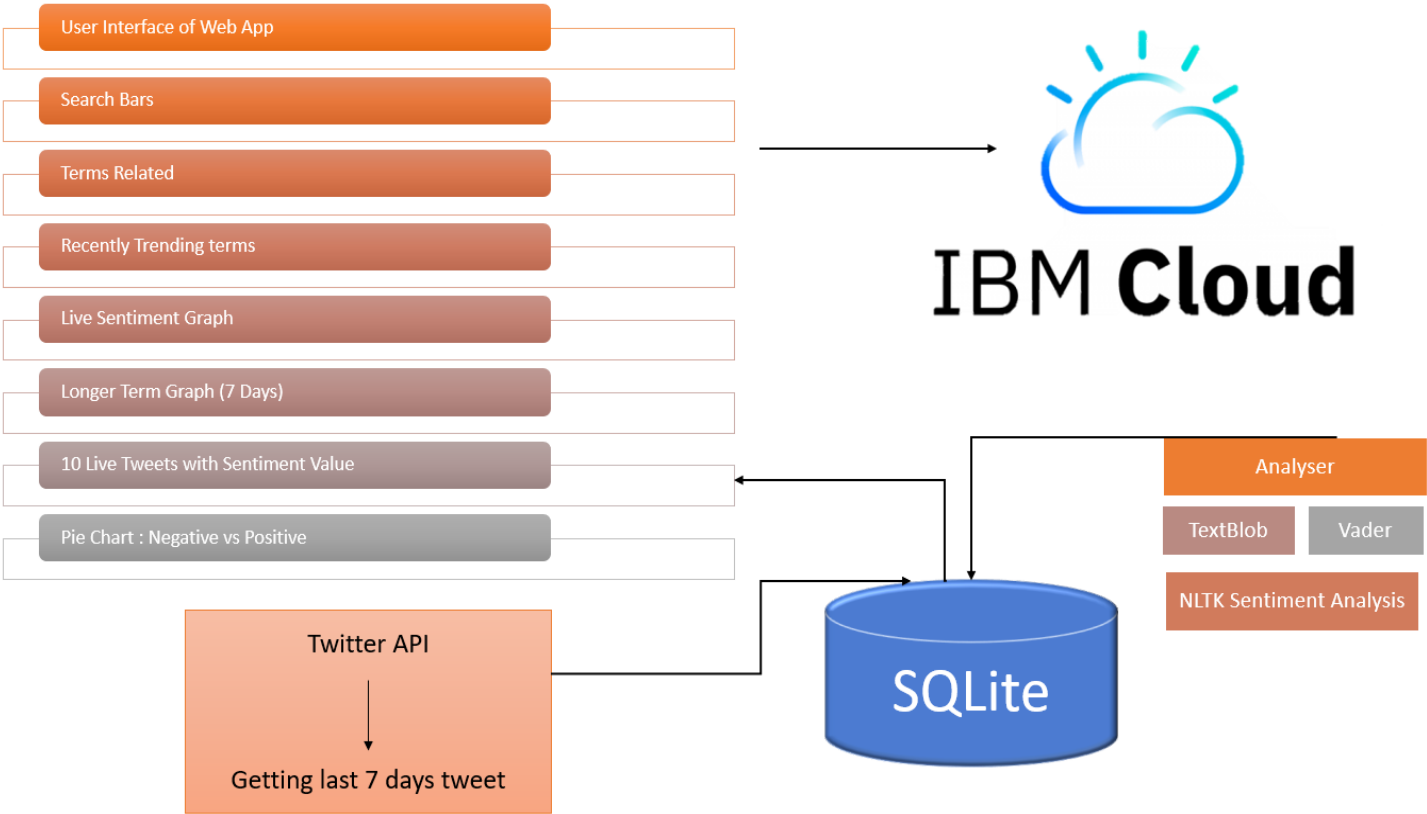- The number of COVID-19 cases in different locations are also displayed in these systems.

PROPOSED SOLUTION
There are websites that provide some functions similar to our solution. However, in our project:
- We display the latest updates as the live changes occur. According to the latest news, the histogram will display the changes with live tweets.
- In addition to the statistics in the histogram, our project also has the ability to show the overall sentiment of the general public.
- By analyzing the tweets, our project will be able to understand the sentiment of the public. This can prove to be helpful in the cases where it is important to understand what people generally feel about a motion/statement passed by the government or any other powerful body in the form of a pie chart.
- Moreover, a comparison of old and new statistics will be shown after every important statement is made.
- The live tweets will also be displayed with their sentiment values.

THEORITICAL ANALYSIS

BLOCK DIAGRAM:

| User Interface of Web App |
| Search Bars |
| Terms Related |
| Recently Trending terms |
| Live Sentiment Graph |
| Longer Term Graph (7 Days) |
| 10 Live Tweets with Sentiment Value |
| Pie Chart : Negative vs Positive |

IBM Cloud

Analyser

TextBlob   Vader

NLTK Sentiment Analysis

SQLite

Twitter API

Getting last 7 days tweet

SOFTWARE DESIGNING:

The software designing of our project is done using the framework called Dash along with coding using HTML and CSS. It uses various packages and libraries imported using python. We obtain the tweets using a python package and the data keeps changing according to the live tweets posted by the public. A search bar is provided for entering the keywords and using the tweets imported using the python package and the graphs of the sentiment analysis are being drawn by the system.

A search bar is provided on the screen that helps to enter the keywords needed to search the tweets. On entering the keyword, the imported packages function together to retrieve tweets in which that keyword is present. Here is an example :

Keyword : corona
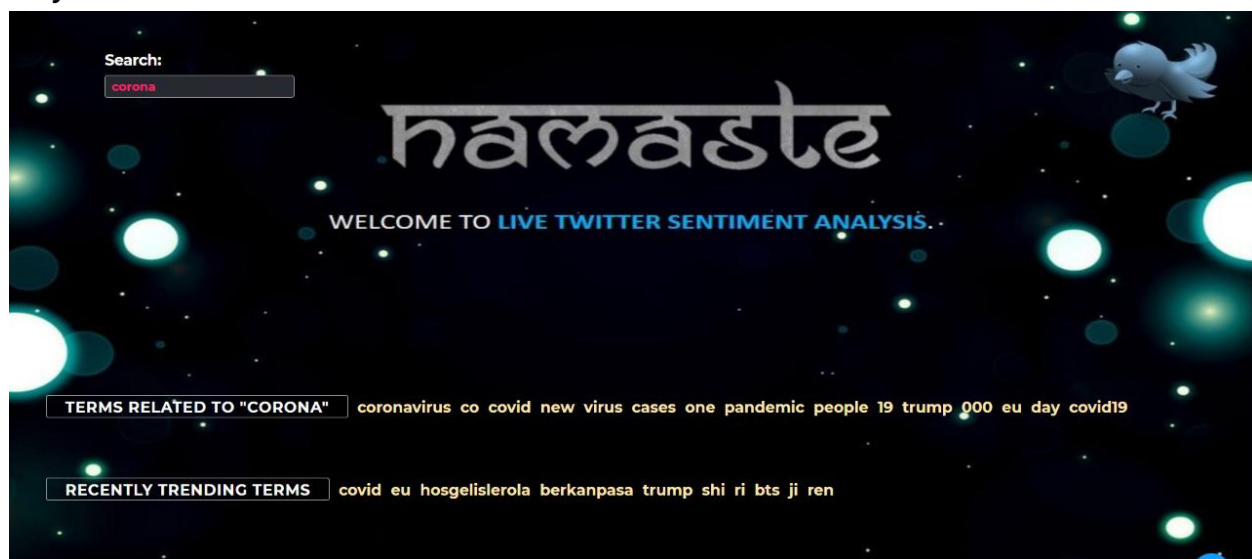
Here, the keyword "corona" is used to search for tweets. Correspondingly, two graphs are then displayed. The graphs being :
1. Live Sentiment
2. Long Term Sentiment

Live sentiment for: "corona"

Longer-term sentiment for: "corona"

Live Sentiment graph shows the sentiment or the feedback of the general public in a live manner while Long Term Sentiment graph shows people's feedback cumulatively. The feedback for a longer period of time is added up and shown in this case. This allows the user to see the difference in actual sentiment of the public over a period of time.

The tweets corresponding to the keyword "corona" are obtained and sentimental analysis is performed and the output is shown in the form of a pie chart. The chart has 2 side : positive and negative depicted by the

different colored sides of the chart as shown.

Date	Tweet	Sentiment
2020-07-14T14:30:56.655000	Dari manteman yang disana pun ngeiyain juga Intie semua negara sama U call it zero corona Yah karena gak di tes massal lagi Baru ketauan kalo ada korban Inget orang yg kena corona ini kalo gak parah gak bakalan ke rs deh	0
2020-07-14T14:30:53.659000	RT @kylegriffin1: Nearly a month after Mike Pence published an optimistic coronavirus op-ed, he has been proved wrong in nearly every way o...	-0.2023
2020-07-14T14:30:46.679000	I thought UK by 2027 trending was in reference to when it'd be Corona free	0.5106
2020-07-14T14:30:44.661000	@CoronaDoubter I think you will be exempt just like the other window lickers	0.4404
2020-07-14T14:30:41.514000	Corona ruiniert Berliner Clubs: Berlin lebt seit Jahren unter anderem vom Party-Tourismus. Doch wegen Corona bleiben viele Clubs bis Ende 2020 geschlossen, die Menschen feiern an anderen Orten. Dadurch droht der Stadt der wirtschaftliche Ruin.... https://t.co/IPXB8HcgCJ (via DW) https://t.co/yvQO5VfGW5	-0.8271
2020-07-14T14:30:41.641000	RT @VIM_Media: En el estado de #California, el numero de muertos por coronavirus en la prision estatal de #SanQuentin ha llegado a nueve, l...	0.34
2020-07-14T14:30:38.633000	RT @productivityfin: 5 things to consider before applying for a Bounce Back Loan https://t.co/j77LtLSRUG #cbils #bouncebackloan #coronavi...	0
2020-07-14T14:30:32.620000	Nao sei se na quarentena teve mais casos de corona ou de novos empreendedores/as.	0
2020-07-14T14:30:31.631000	RT @nowthisnews: Dr. Fauci remains 'cautiously optimistic' that a coronavirus vaccine will be available this winter https://t.co/TRRcF7N1pp	0.3182
2020-07-14T14:30:28.658000	RT @lindyli: WOW Trump has the gall to blame Obama and Biden for stopping testing The coronavirus hit the U.S. in 2020. This shitshow...	0.368

Each tweet is given a sentiment point with which the percentage of the pie chart shown below it is calculated and plotted.
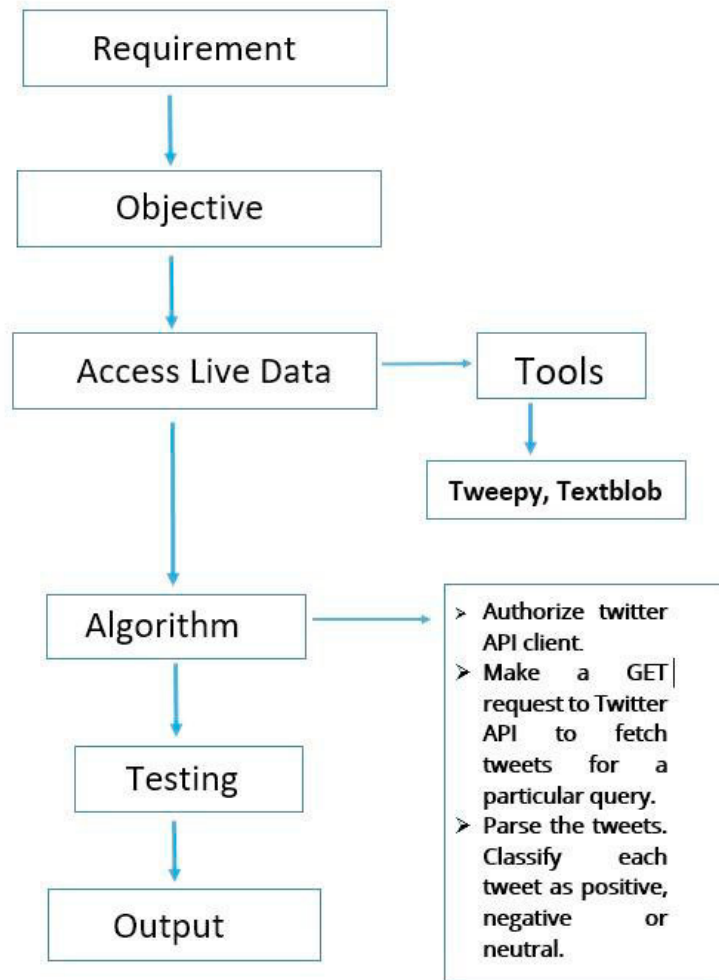
Experimental Investigation

➤ We found that there are many sentimental analysis web applications online. But only a few updates them every minute.
➤ The UI of many are not easy to use. They are complicated.
➤ Many include analysis as text.
➤ Only some has visual representation.

So our main goal is to develop a web app that looks attractice and is easy to use. It should focus on visual representation and shoul be easy for the viewers to analyze.

FLOWCHART

Requirement

↓

Objective

↓

Access Live Data	→	Tools

↓

Tweepy, Textblob

Algorithm	→	> Authorize twitter API client. > Make a GET request to Twitter API to fetch tweets for a particular query. > Parse the tweets. Classify each tweet as positive, negative or neutral.

↓

Testing

↓

Output

RESULT

The project aims to help workers of all kinds of sectors. It provides a number of positive enhancements to the lives of the beneficiaries. The statistics shown in the graphs and pie charts in this system can be very useful in the following cases:

- The overall sentiment of the public can be estimated very easily.
- This is useful for the government as they can use it to understand the public opinion after a particular change in policies that are being brought.
- This system can help in the efficient production of materials during the quarantine period. This can be done by searching using the keywords as the materials needed and so the needed materials can be produced by their respective sectors.
- The statistics can also act as a form of survey for any kind of purpose.
- It can also be used to get the general feedback of the public after a new product is launched. The graphs can show positive and negative reviews very easily.

ADVANTAGES AND DISADVANTAGES

Advantages of sentiment analysis :-

Sentiment analysis is a useful tool for any organization or group for which public sentiment or attitude towards them is important for their success - whichever way that success is defined.

On social media and online forums millions of people are busily discussing and reviewing businesses, companies, and organizations. And those opinions are being 'listened to' and analysed. Those being discussed are making use of this enormous amount of data by using computer programs that don't just locate all mentions of their products, services, or business, but also determine the emotions and attitudes behind the words being used.

- The results from sentiment analysis help businesses understand the conversations and discussions taking place about them, and helps them react and take action accordingly. They can quickly identify any negative sentiments being expressed, and turn poor customer experiences into very good ones.
- They can create better products and services, and they can formulate the marketing messages they send out according to the sentiments being expressed by their target audience or customers. All of which adds up to increased sales and revenue.
- By listening to and analysing comments on Facebook and Twitter, local government departments can gauge public sentiment towards their department and the services they provide, and use the results to improve services such as parking and leisure facilities, local policing, and the condition of roads.
- Universities can use sentiment analysis to analyze student feedback and comments garnered either from their own surveys, or from online sources such as social media. They can then use the results to identify and address any areas of student dissatisfaction, as well as identify and build on those areas where students are expressing positive sentiments.
- And by analysing the sentiment behind customer reviews on sites like TripAdvisor and Yelp, hotels and restaurants can not only manage their

reputations by improving the services offered, but can also gauge the general customer attitude to their business or brand.

- Businesses can compare their results with those of their competitors to better understand people's attitude to their business. They can identify where they may be excelling, or identify where there's room for improvement compared to the competition. They can also conduct market research into general sentiment around key issues, topics, products, and services, before developing and launching their own new services, products or features.

Disadvantages of Sentiment Analysis :-

- Sentiment analysis tools can identify and analyse many pieces of text automatically and quickly. But computer programs have problems recognizing things like sarcasm and irony, negations, jokes, and exaggerations - the sorts of things a person would have little trouble identifying. And failing to recognize these can skew the results.
- 'Disappointed' may be classified as a negative word for the purposes of sentiment analysis, but within the phrase "I wasn't disappointed", it should be classified as positive.
- With short sentences and pieces of text, for example like those you find on Twitter especially, and sometimes on Facebook, there might not be enough context for a reliable sentiment analysis. However, in general, Twitter has a reputation for being a good source of information for sentiment analysis, and with the new increased word count for tweets it's likely it will become even more useful.
- So, automated sentiment analysis tools do a really great job of analysing text for opinion and attitude, but they're not perfect.

APPLICATIONS

Nowadays, sentiment analysis or opinion mining is a hot topic in machine learning. We are still far to detect the sentiments of s corpus of texts very accurately because of the complexity in the English language and even more if we consider other languages such as Chinese.

While this pandemic has continued to affect the lives of millions, a number of countries have resorted to complete lockdown. During this lockdown, people have taken social networks to express their feelings and find a way to calm themselves down.
We are in a time of new desperate needs, new product requirements in the market are rising drastically. It is not easy to identify them, analysing people requirements through socialmedia would be one solution to tackle this. Futhermore each update from the Government regarding this pandemic and people's reaction and difficulties in going ahead with it can also be analysed effectively.

CONCLUSION

Our project enables the user to enter in the keyword that they want, to obtain a sentimental analysis of the tweets related to that specific topic. For our project, we have mainly focused on "COVID-19" and "corona" as keywords. In the search bar provided in our user interface, we can type in the said keywords. The system then scans the tweets on twitter and obtains the ones related to the keywords that we have provided. These tweets are then displayed along with two separate graphs showing the overall sentiment of the public. A pie chart is also depicted, showing "negative" and "positive" side of tweets, to help get an idea of the overall feedback of the people tweeting. It also has the ability to keep track of tweets live. So as the tweets keep coming in, the graphs and pie charts keep up with the changes.

Our project makes use of the framework called Dash. The backend development uses HTML and CSS for coding and designing the user interface. A number of packages have been imported using Python, for various functions in the project as well.

FUTURE SCOPE

From future perspective, this study can be used to analyze the changing emotions and sentiments of people from these countries and check whether there are major shifts in them over the period of time. It is expected that as the spread of this pandemic will increase, the sentiments and emotions in the tweets may change on the lines of what was seen in the case of China.

Moreover, we would like to make a web application for users to input keywords and get analyzed results. They can review needs and respnoses to a service.

BIBLIOGRAPHY

For learning more about Dash :
https://www.datacamp.com/community/tutorials/learn-build-dash-python
https://pythonprogramming.net/live-graph-twitter-sentiment-analysis-gui-dash-python/

For clarifications :
https://app.slack.com/client/T013B843QU9/C013NEFSC48

APPENDIX

SOURCE CODE:

Here are a few of the important parts of our code :

A few modules and libraries are imported first :

```python
1  import dash
2  from dash.dependencies import Output, Event
3  import dash_core_components as dcc
4  import dash_html_components as html
5  import plotly
6  import random
7  import plotly.graph_objs as go
8  from collections import deque
9  import sqlite3
10 import pandas as pd
```

Next, inside of our update_graph_scatter function, we'll add

```python
1  conn = sqlite3.connect('twitter.db')
2      c = conn.cursor()
3          df = pd.read_sql("SELECT * FROM sentiment WHERE tweet LIKE
   '%COVID-19%' ORDER BY unix DESC LIMIT 1000", conn)
4      df.sort_values('unix', inplace=True)
5      df['sentiment_smoothed'] = df['sentiment'].rolling(int(len(df)/5)).mean()
6      df.dropna(inplace=True)
```

Then we replace the X and Y variables here with our data and form the rest of the code around this.

```
1   X = df.unix.values[-100:]
2   Y = df.sentiment_smoothed.values[-100:]
```

For this to be streaming live, you will need to have the following program also running, with the API credentials filled out:

```
1   from tweepy import Stream
2   from tweepy import OAuthHandler
3   from tweepy.streaming import StreamListener
4   import json
5   import sqlite3
6   from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
7   from unidecode import unidecode
8   import time
9
10  analyzer = SentimentIntensityAnalyzer()
```

From the above code excerpt we can understand that they modules and libraries are imported like tweepy, json and sqlite in order to obtain the tweets that we need, to deal with json and to access the needed databases respectively.
The SentimentIntensityAnalyzer is also assigned to analyzer variable to make it easier to handle the rest of the project.

We now try to include elements that help in dynamically graphing terms for sentiment. For this we include the words to search by :

```
1   dcc.Input(id='sentiment_term',value='COVID-19', type='text'),
```

This is included inside :

```
1   app.layout = html.Div(
```

```
 2    [  html.H2('Live Twitter Sentiment'),
 3        dcc.Input(id='sentiment_term', value='COVID-19', type='text'),
 4        dcc.Graph(id='live-graph', animate=False),
 5        dcc.Interval(
 6           id='graph-update',
 7           interval=1*1000
 8        ),
 9    ]
10 )
```

Now we include the input:

```
 1  @app.callback(Output('live-graph', 'figure'),
 2            [Input(component_id='sentiment_term', component_property='value')],
 3            events=[Event('graph-update', 'interval')])
```

Next, we want to make our query use the term typed into the search box.

```
 1  df = pd.read_sql("SELECT * FROM sentiment WHERE tweet LIKE ? ORDER BY unix
    DESC LIMIT 1000", conn ,params=('%' + sentiment_term + '%',))
```

We will now include timestamps and combine to obtain :

```
 1      conn = sqlite3.connect('twitter.db')
 2      c = conn.cursor()
 3       df = pd.read_sql("SELECT * FROM sentiment WHERE tweet LIKE ? ORDER BY
    unix DESC LIMIT 200", conn ,params=('%' + sentiment_term + '%',))
 4      df.sort_values('unix', inplace=True)
 5      df['sentiment_smoothed'] = df['sentiment'].rolling(int(len(df)/2)).mean()
 6
 7      df['date'] = pd.to_datetime(df['unix'],unit='ms')
 8      df.set_index('date', inplace=True)
 9      df.dropna(inplace=True)
10      X = df.index
```

```
11      Y = df.sentiment_smoothed
```

_____

_