#### **Sentiment Analysis Of COVID-19 Tweet- Visualization Dashboard**

|  |  |
| --- | --- |
| Sr.No. | TOPIC |
| 1. | Introduction |
| 2. | Literature Survey |
| 3. | Theoretical Analysis |
| 4. | Experimental Investigation |
| 5. | Flow Chart |
| 6. | Result |
| 7. | Advantages & Disadvantages |
| 8. | Application |
| 9. | Conclusion |
| 10. | Future Scope |
| 11. | Bibliography |
| 12. | Appendix |

Team Name : Mind Bridges

Rakshita Iyer (Team Leader)

Sharvari Gaitonde

Lourd Bharathy Das

1. INTRODUCTION

1.1 Overview:

This project studies the sentiment of people towards the lockdown extension using tweets.

Extracting opinions from social networks and classifying emotion & sentiments using different machine learning algorithms,make this a valuable decision support tool for the Government.

In this covid-19 pandemic it is very important to know the people’s response to governments decisions.

After analyzing the tweets we will plot graphs on NodeRed which will give a clear idea to our government officials about the reaction of people on extension of lockdown and people’s sentiment towards the pandemic.

1.2 Purpose:

The purpose of the project is to make a web application that supports the Government to make decisions on lockdown and understand people’s emotion on pandemic.To understand the behavior of people on further lockdown extension and build a predictive analytics model .

A dashboard with visualization is created based on people’s reaction to the govt announcements on lockdown extension with Node-Red and Watson Studio.

2. LITERATURE SURVEY

2.1 Existing problem

Whole World is battling the Pandemic of Covid-19, Countries announced Lockdown to control the spread of this virus. People all over the world started sharing their views about this lockdown and Covid situation on many social media platforms. Governments across the world are trying to move closer to their citizens to achieve transparency and engagement. The explosion of social media is opening new opportunities to achieve it. Hence there is a need for a platform which can help the officials to study what people think about the decisions made by the Government.

2.2 Proposed Solution

Our project uses this social media Platform i.e. **Twitter** to study the **sentiment and emotion** of People regarding the extension of lockdown and the overall Covid crisis in **India**. This analysis is done by separating the tweets on the basis of country i.e. India and some #tags. #tags like #lockdown and #indiafightscorona are used to filter out tweets. This will store all the tweets that contains these hashtags and we can perform Sentiment and Emotional Analysis.

3. Theoretical Analysis

3.1 Block Diagram:

The project can be categorized into 3 major sections:

1. Data Preparation which includes Data Collection and Data Cleaning. This is done using Python (Jupyter Notebook)
2. Analysis of Tweets which consists of Emotion classification and Sentiment Classification of tweets. Algorithms like KNN and TF-IDF are used to increase the accuracy.
3. Deployment is done using Node-Red which helps us to build User Interface and Interactive Dashboards.



3.2 Software Designing :

Two major Software used in this project includes:

1. Jupyter Notebooks (IBM Watson Studio):

A Jupyter notebook is a web-based environment for interactive computing.

Notebooks include all of the building blocks you need to work with data:

* The data
* The code computations that process the data
* Visualizations of the results

Text and rich media to enhance understanding We have run code that process our data, and immediately view the results of the computation which we have saved in csv files.

1. Node-Red:

Node-RED is a programming tool for wiring together hardware devices, APIs and online services in new and interesting ways.It provides a browser-based editor that makes it easy to wire together flows using the wide range of nodes in the palette that can be deployed to its runtime in a single-click.

We have used various nodes like cloud object storage,graphs,function,csv nodes etc to implement our flow.Using chart nodes we have implemented different graphs.

4. Experimental Investigation:

4.1 Data Collection:

Data Collection is done using GetOldTweet3 Library. As our project is about the reaction of people on extended Lockdown, #lockdown is used to filter the tweets and the location is set to India. For now we have set the tweet count to 50 for fast working of the program. The tweets are stored in the form of List. The tweets were found to be in different language, so we used google translate module to translate the tweets into English. The List of tweets is converted into string and stored in a csv file which gets saved in Assets of Watson studio.

4.2 Data Cleaning:

For data cleaning, we used nltk library (Natural Language Tool Kit) From which we have used stopwords and punkt files. For cleaning the data we used the same csv fle where we saved our tweets. The saved tweets are converted into lowercase and by tokenization the words are seperated into list of words. After tokenization, we performed stemming where the words are broken down into its root word. After that, all the stopwords and special characters are identified and removed from the list. This list after all the operations is converted into string and then stored in the same csv file.

4.3 Emotional & Sentiment Analysis

Emotions are classified into fear, joy, trust, anger, anticipation, sadness, surprise and disgust.

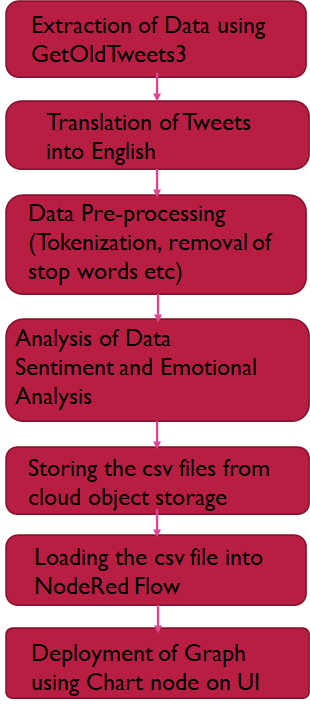
Sentiments are classified as positive, negative, neutral.

Analysis is implemented using Jupyter Notebooks of IBM Watson studio. Methods used in coding includes : NRClexicon for emotional analysis, vaderSentiment (SentimentIntensityAnalyzer) to find the polarity of tweets and classify them, nltk FreqDist get the final count the Emotions and sentiments, TF-IDF and K-nearest Neighbor Theorems for training, testing and clustering of data for better accuracy.

4.4 Deployment (Visualization Dashboard)

Node-Red nodes used are Inject node to trigger the flow, function nodes to implement some Java-Script codes to get the desired output, cos-get node to Upload the file from cloud object storage to the nodered flow, csv node to get the data of specific rows and chart node to plot the charts of Sentiment and Emotional Analysis.

5. Flowchart

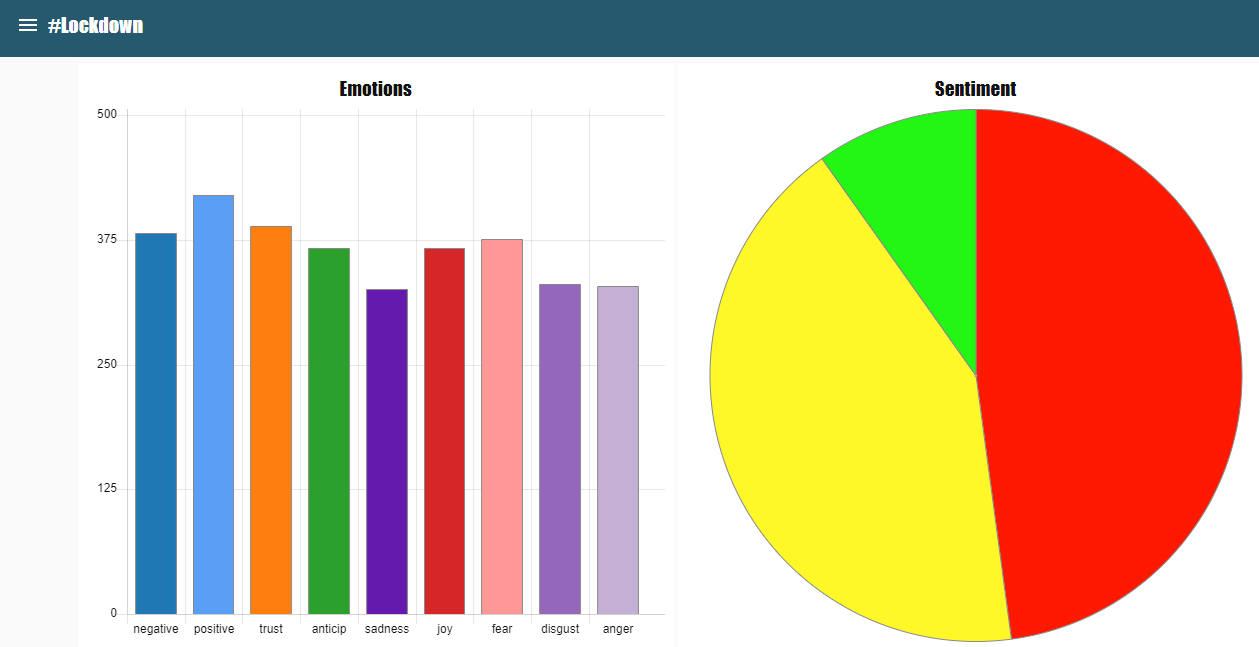


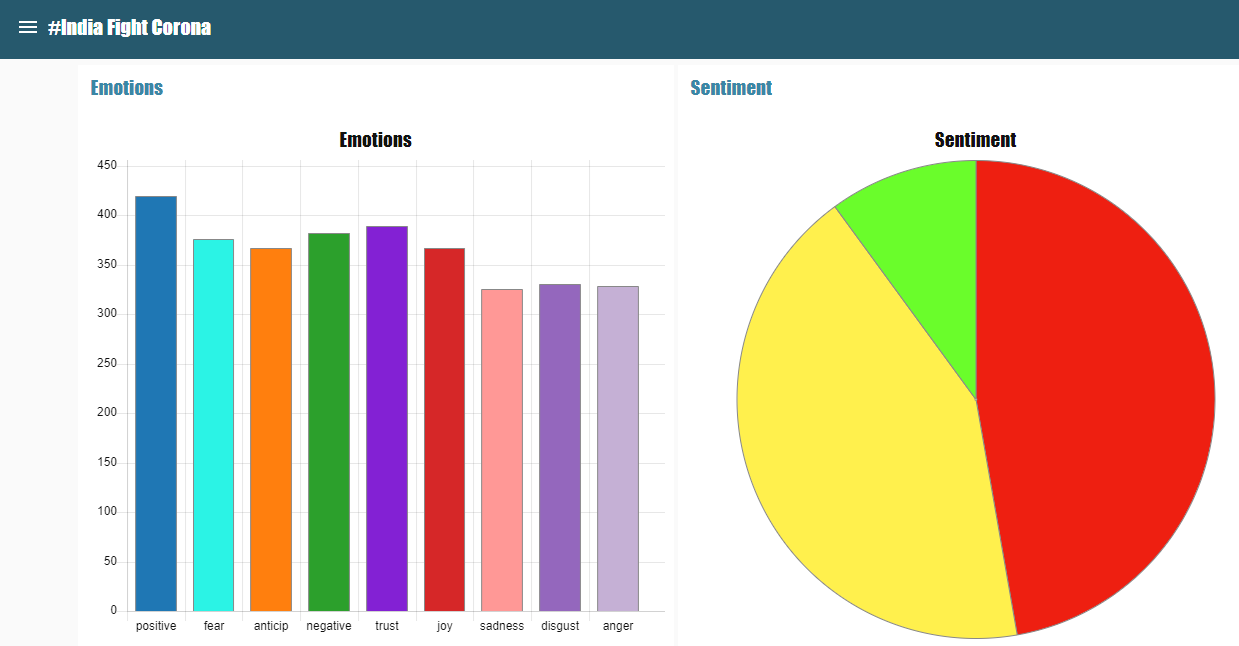
6. Result

An User Interface is Built using Node-Red. The chart nodes produce graphs. Sentiments are

represented using Pie-chart where as Emotions are represented using Bar-graph (X-axis represents name of emotion and Y-axis count)







7. Advantages & Disadvantages

7.1 Advantages:

* Tf-IDF is used for generating weightage values which will help to obtain a more accurate classification for our model.
* The clustering algorithm ie. KNN increases the predictive power of our model.
* Tweets of different languages across our country can also be analyzed to get the desired outputs from our model as we are using translator function.
* The project is completely based on software which is easily available hence is cost effective.

7.2 Disadvantage:

* It’s a tremendously difficult task even for a human being to understand a human emotion. Many times humans misinterpret sentiments of the text and when a machine does the job it gets way more tedious.

So there are chances where predictions may go wrong from time to time. But by using this technique you will get the opportunity to get it right about 70-80% of the time.

8. Applications

* The possibility to extract opinions from social networks and classify sentiment using different machine learning algorithms,make this a valuable decision support tool for the Government. The officials can predict how people will react to their certain decisions and plan future strategy.
* This project will help the Government to think on the rules and regulations implemented from a common man perspective and this will create a feeling of trust and security among the citizens. This project will act as a communication bridge between the government and its citizens.

9. Conclusion

Twitter is a platform where people share their personal opinions. In this project we analyzed the emotion of the people through their tweets and segregated them into positive, negative and neutral Sentiments. The aim of this project is to understand the behavior of people on further lockdown extensions and a predictive analytics model is built. A dashboard with visualization is created based on people’s reaction to the government announcements on lockdown extension and corona virus in India.

10. Future Scope

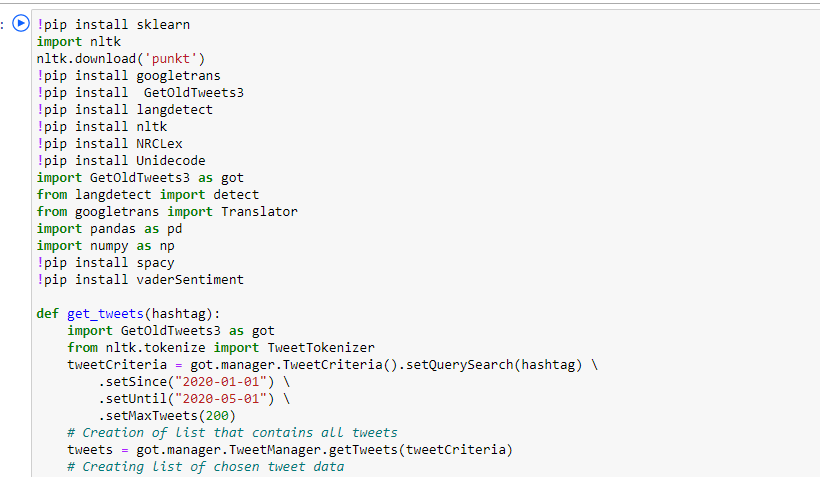
* Twitter Developer account is required to get Live tweets. We have applied for it but the access is not provided yet. Hence we have used GetOldTweets3 which can give us latest tweets which were posted a week before. Hence after getting access to developer account we can work with the Live Tweets.
* The data here is static till we go and re-run the program in Jupyter Notebook, therefore we are planning to implement a code which will re-run this program after certain interval of time.
* After implementation of above code to refresh the graph immediately as per data, instead of inject node in Node-Red a button can be used so by clicking, the user can trigger the flow and graph will get updated automatically. For now we have set inject node into repeat so it will trigger the flow after a specific interval of time.

11.Bibliography

* <https://medium.com/@AIY/getoldtweets3-830ebb8b2dab>
* <https://medium.com/@debarghyanandi/emotion-classification-in-machine-learning-4183f6c5d07a>
* <https://towardsdatascience.com/statistical-sentiment-analysis-for-survey-data-using-python-9c824ef0c9b0>
* <https://flows.nodered.org/>

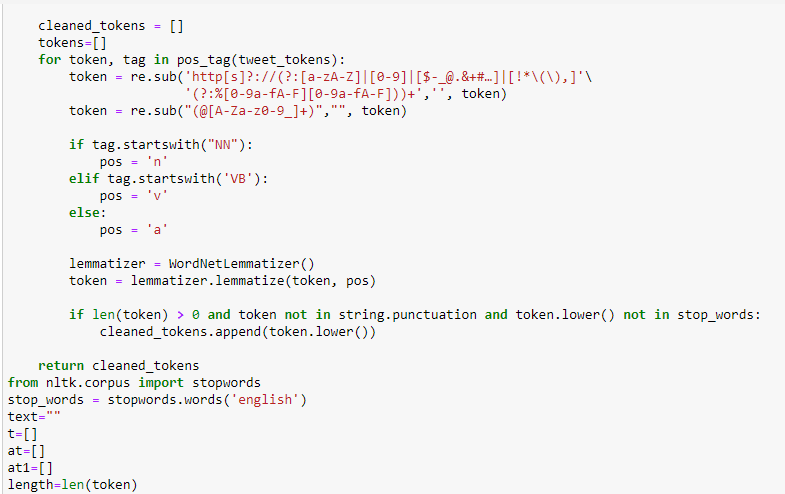
Appendix:

1. Source code (Jupyter Notebook)





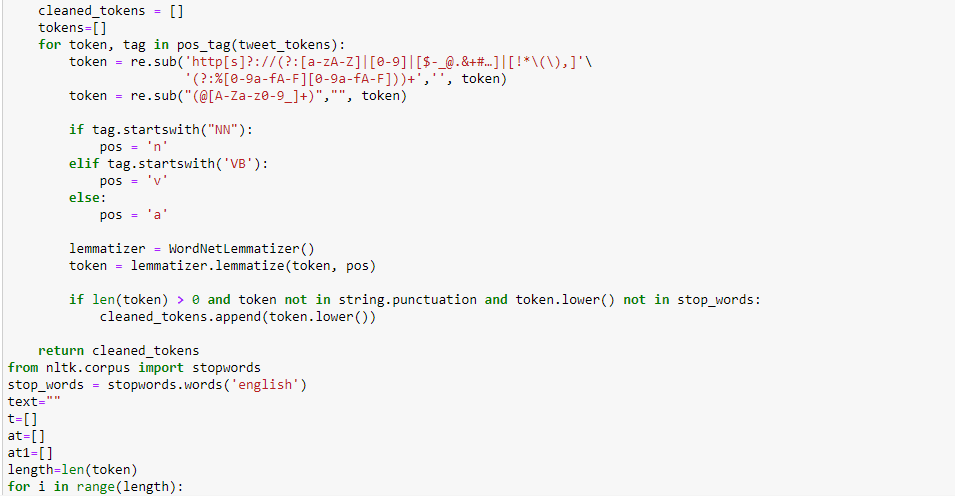














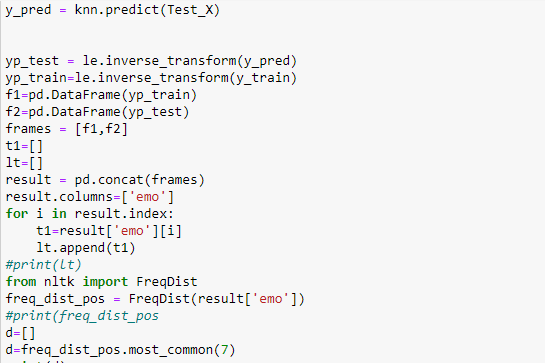




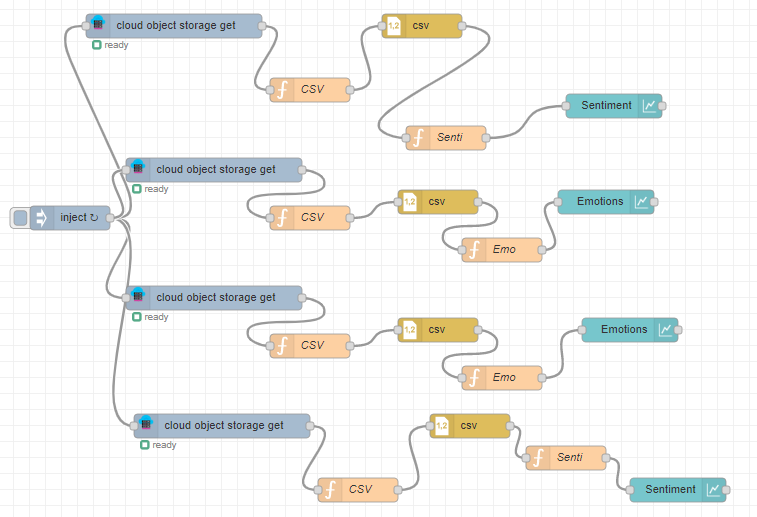




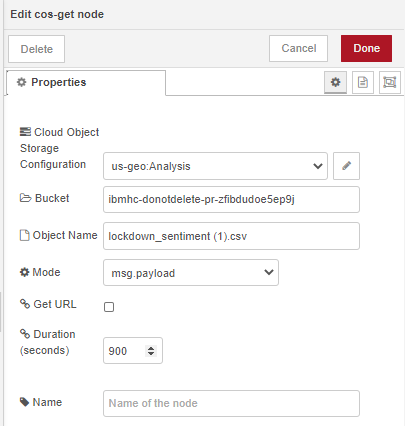




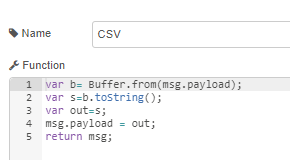
1. Node-Red Flow



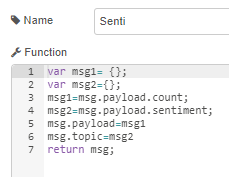
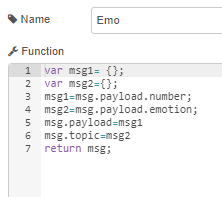
Cos-get node configuration:



CSV function node:



Senti/ Emo function node:

Dashboard Tabs and Groups:

