Sentiment Analysis of COVID-19 Tweets – Visualization Dashboard

Project Report

Team Rinj

**INDEX**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | |  | **INTRODUCTION** | | | |  | |  | |  | |
|  | |  | 1.1 Overview | | | |  | |  | |  | |
|  | |  | 1.2 Purpose | | | |  | |  | |  | |
| **2** | |  | **LITERATURE SURVEY** | | | |  | |  | |  | |
|  | |  | 2.1 Existing problem | | | | | |  | |  | |
|  | |  | 2.2 Proposed solution | | | | | |  | |  | |
| **3** | |  | **RESULT** | |  | |  | |  | |  | |
| **4** | |  | **ADVANTAGES & DISADVANTAGES** | | | | | |  | |  | |
| **5** | |  | **APPLICATIONS** | | | |  | |  | |  | |
| **6** | |  | **CONCLUSION** | | | |  | |  | |  | |
| **7** | |  | **FUTURE SCOPE** | | | | **`** | |  | |  | |
| **8** | |  | **BIBILOGRAPHY** | | | |  | |  | |  | |
|  | |  | **APPENDIX** | | | |  | |  | |  | |
|  |  | | | A. Source code | | | |  | |  | |  | |
|  |  | | |  | |  | |  | |  | |  | |

**INTRODUCTION**

**1.1 Overview**

* Sentiment Analysis refers to the use of Natural Language Processing (NLP) to determine the attitude, opinions and emotions of a speaker, writer, or other subject within an online mention.
* Twitter is a social media platform where we can use sentiment analysis to analyze various sentiments of various topics.
* COVID-19 also known as the novel coronavirus has had a substantial effect on the routine as well as the mental state of the people. In this situation people are turning towards social media platforms like twitter to express their sentiments towards the events taking place all across the country.

**1.2 Purpose**

* The problem statement that we choose to work on is to analyze sentiments of people during the epidemic so keeping this thought in our mind and making the web app according to it that is mainly focused to analyze sentiments of people from their texts extracted from twitter.

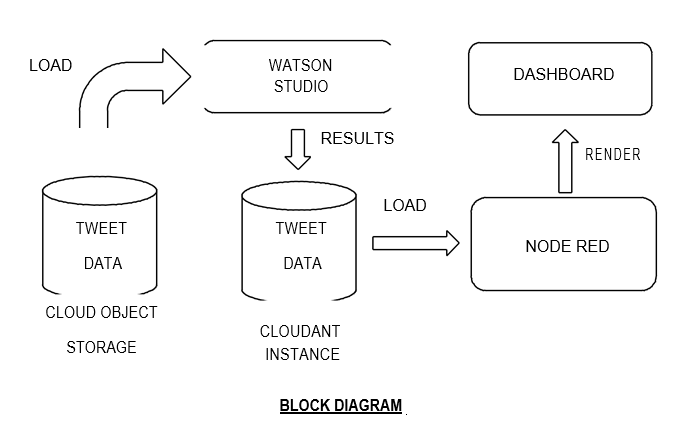
**LITERATURE SURVEY**

**2.1 Existing problem**

The objectives of the study are first, to analyze the twitter tweets of various people across the various regions of the Covid-19 topic(hashtag) and second, is to develop a platform where we can display the various sentiments that we have analyzed. Some of the problems are as follows:

1. Filtering the data  
2. Hydrating tweet\_id  
3. Storing and processing such a huge amount of data 4. Building an UI to display the results

**2.2 Proposed solution**



Twitter sentiment analysis is developed to analyse customers perspectives toward the critical to success in the marketplace. The program is using a machine-based learning approach which is more accurate for analysing a sentiment; together with natural language processing techniques will be used.

As a result, program will be categorized sentiment into positive and negative . Most of the problems faced by us can easily be solved by using IBM Cloud to build our project. Since IBM is providing us 5GB of Cloud Object Storage which can be connected to Watson Studio, where we can spawn notebooks with 8GB RAM which is more than sufficient to bear the load of the data.

This notebook would be responsible for processing the data and storing the results to a Cloudant instance.

We would then be using Node-RED for hosting the UI and retrieving data from the Cloudant instance.

**EXPERIMENTAL INVESTIGATIONS**

There were a substantial number of datasets which were considered for this project, one of which was the IEEE dataset named "CORONAVIRUS (COVID-19) TWEETS DATASET." It was a good dataset for this project, but because not all tweet\_ids in it were geotagged it had to be skipped as the twitter API provides only 900 hydrations of tweet\_id per 15 mins.

To avoid this problem I went through the "CORONAVIRUS (COVID-19) GEO-TAGGED TWEETS DATASET" it consisted of tweet\_id of only those tweets which are geo-tagged. This was a very generous gesture by the one who made this dataset (Rabindra Lamsal), but since we could only do 900 hydrations per 15 mins it was not used.

Later a dataset was used which was already hydrated, geo-tagged and made available for public use. From that dataset all the Indian tweets related to COVID-19 were then filtered. This was then used as the input for this project.

**RESULT**

A dashboard was created which consisted of interactive views (charts and tables) to get the abstract of sentiments of the people of India towards various government decisions along with some other interesting details like hashtags, geotags and user mentions.

**ADVANTAGES**

The Watson application includes following advantages:

* Easy to create.
* Less time required to develop.
* High performance with IBM cloud.
* The model provides with high accuracy sentiment analysis and also have undergone special training for covid19 tweets.
* The accuracy is around 96% which makes the model usable by everyone.

**DISADVANTAGES:**

The Watson news search application includes following disadvantages:

1. Network issues can also cause problems in plotting of graphs and lead to slowdown of the site.
2. No real time tweet data is being processed, since we did not have a premium or enterprise plan on twitter API, we could not run it on real time data without exhausting the API limit.

**APPLICATIONS**

1. The predictive analysis provided by our model can help recognize the sentiments and views of the masses on important news like extension of lockdown or easiness and sectors now allowed to function and these sentiments then further help the manufacturers/service providers predict about the demand/supply chain as per the positivity/negativity provided by the people and hence work on the same.
2. It also helps analyse the people’s sentiment towards the epidemic.
3. It helps understand the sentiments of people on government's decisions during the epidemic.
4. Our analysis helps the user to understand and comapre how people sentiments changed during different phase of lockdown.

**CONCLUSION**

This report focuses on sentiment analysis of covid-19 tweets and displaying results on visualization dashboard. Python libraries such as pandas, NumPy, nltk, sklearn, pickle, twitterscraper, matplotlib and many more are used for the implementation of the concept.

It predicts sentiment of statement entered by user, displays live tweets and its sentiment and also contains visualization dashboard containing graphs which updates daily.

**FUTURE SCOPE**

1. We can make an app of the same for mobile devices.
2. Extensions for browsers can also be made which will predict and show sentiments next to tweets on twitter page.
3. We can add real time tweet processing to it which will keep the data updated on the site. However, for it to happen we would have to acquire a premium or enterprise level API from twitter.

**BIBILOGRAPHY**

**Names**: Rishabh Jain , Jatin Agrawal ,

**College Name**: NMIMS MPSTME

**Work Title:** Sentiment Analysis of COVID-19 Tweets – Visualization

Dashboard

**References:**

1. News search app Documentation:

<https://developer.ibm.com/patterns/create-a-cognitive-news-search-app/>

1. Slack integration: <https://slack.com/intl/en-in/>
2. IBM Cloud: <https://www.ibm.com/cloud/get-started>
3. Node red tutorial:

<https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/>

1. About API:

<https://www.youtube.com/watch?v=s7wmiS2mSXY&feature=youtu.be>

1. Watson services and product: <https://www.ibm.com/watson/products-services>
2. IBM Watson Discovery:

<https://developer.ibm.com/articles/introduction-watson-discovery/>

1. Demo application: <https://discovery-news-demo.ng.bluemix.net/>

**APPENDIX**

**Link to Node Red work space:**

<https://news-search-network.eu-gb.mybluemix.net/red/#flow/b349b775.662b18>

**Link to Node Red UI:**

<https://news-search-network.eu-gb.mybluemix.net/ui/#!/0?socketid=f4KiGnLn9Dca9avgAAAB>