MULTI-PLATFORM SENTIMENT ANALYSIS

M. Madhusudhan, Assistant Professor, Dept. of CSE, CMR Technical Campus, Hyderabad.

Syed Yusuf Ali, UG student, Dept. of CSE, CMR Technical Campus, Hyderabad.

Nomula Dinakar, UG student, Dept. of CSE, CMR Technical Campus, Hyderabad.

Maskula Rahul Sai, UG student, Dept. of CSE, CMR Technical Campus, Hyderabad.

ABSTRACT- The information we aim to provide is mined from social media and analyzed, we provide an overview of public sentiment data from websites like YouTube, Twitter, and Reddit about any particular topic for a specific time frame. This is the era of social media where the internet has become a must for every teenager and adult. The lifeline of a normal human being during the lockdown period is knowing about the happenings around the world through social media sites like YouTube, Twitter, Reddit, and many more. People are heavily dependent on social media sites to know about anything. The masses are also showering their views regarding their topic of interest be it positive or negative and these views can be helpful to other people. So, our webapp gives the viewers an insight into these. The webapp retrieves through the comments on a particular topic (that the user has searched for) in various social media sites and returns a response on a scale of -1 to 1, about the sentiments of the people where -1 is negative sentiment i.e., people are unsatisfied with it and +1 is positive sentiment i.e., people are satisfied. This data is presented in a graphical format that provides higher readability and ease of consumption to the user. The programming language used is Python.

Key words: Sentiment analysis, Public sentiment, Comment mining, Social media, Python.

1. INTRODUCTION

Know what the world thinks!

It is very difficult to know what the world thinks of a particular topic, our webapp is focused on painting the true picture of the world for you. The information that is provided is mined from social media and analyzed by us using the sentiment analysis model. We provided an overview of analyzed data from websites like YouTube, Twitter, and Reddit about your topic for a relevant timeframe. This data is presented in a visual format that provides higher readability and ease of consumption.

So, what am I seeing?

- YouTube: 100 most relevant comments each from across 10 most popular videos, for instance, 100x10=1000 comments!
- Twitter: around 1000 most relevant tweets!
- Reddit: 150 most relevant comments each from across 5 most active subreddits, for instance, 150x5=750 comments!

What is the significance?

Most social networks project the views of the most vocal but a minority of users on their platforms, however, the majority of the users' opinion is not taken into consideration. We plan on providing a non-biased overview by mining each comment from various social media sites which returns a score on a scale of -1 to 1, which signifies the sentiments of people where '-1' being most negative sentiment i.e. people are unsatisfied with it and '+1' being most positive sentiment i.e. people are satisfied.

2. METHODOLOGY

For our project, we needed an accurate sentiment analysis model as well as a very fast one as we are mining a huge amount of data through the social media network and analyzing it at the runtime, therefore after research we found the best option was 'TextBlob'. We referred to many research papers on 'TextBlob' which is a python language module for performing analysis tasks on the text. And many YouTube lectures to obtain a very stable and fast model so that we can also use it for production purpose

For the user interface purpose, we made the web application on the dash a mini framework based on 'Flask' written in python language, we also used the SQLite database for the prototype, we referred to many websites and found that we can use either Django or Flask for the python web development so we chose Flask as it is easy to learn and have a flexible structure.

There were many options available for making the dashboard, we used plotly-dash to make the interactive dashboard.
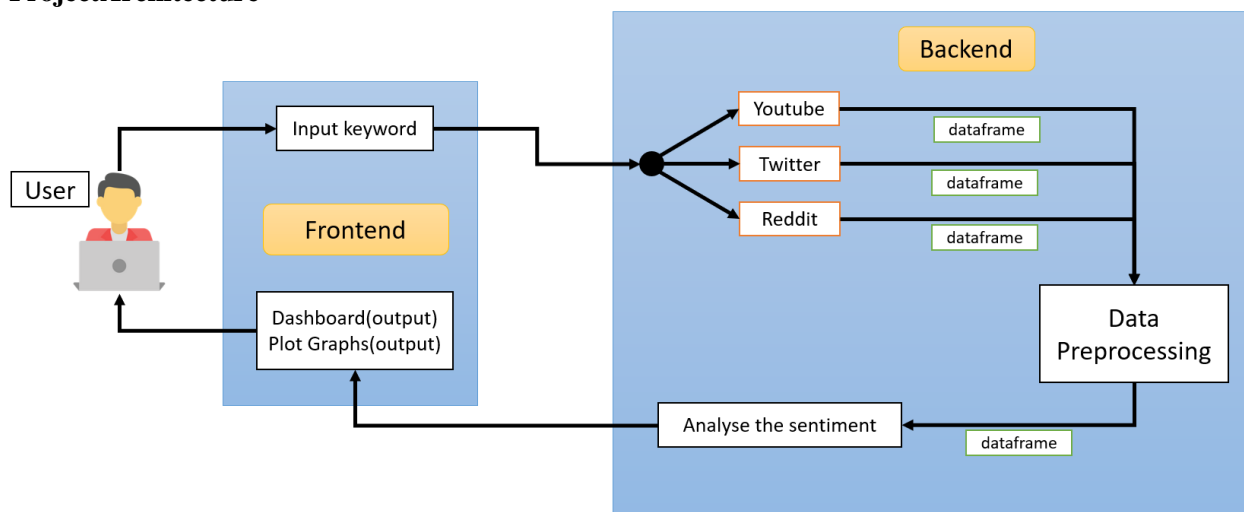
ProjectArchitecture

Figure 2.1: Architecture diagram.

The user gives an input keyword regarding the topic he/she wants to search. The keyword is searched by the backend program and top comments from YouTube, Twitter, and Reddit are gathered and put into a data frame. The data frame consists of the comments that the people have entered for a particular post. Then the data frame is passed through our TextBlob based sentiment analysis model which can find the sentiment of the comment and return a sentiment score

from a scale of -1 to +1. The whole process of sentiment analysis is parallelized with the help of 'asyncio' library for python. The graphs are then plotted against a particular axis and these graphs are presented in front of the user on the webapp. These graphs help the user to analyze whether a particular topic is appreciated by the public or not.

3. RESULTS AND DISCUSSION

Our solution

We are focused to paint the true picture of the world for the users. The information that is provided is mined from social media and analyzed by us, we provided an overview of data from websites like YouTube, Twitter, and Reddit about your topic for a relevant timeframe.

This data is presented in a visual format that provides higher readability and ease of consumption.

For making the sentiment analysis model, one approach is to make the model using natural language processing. If we have used that method then we need to tokenize the sentence and then we need to remove the stop words.

Stop words are those words that do not add any weight to the meaning of the sentence like the articles etc. But the drawback of using that approach is that is slow and, in our case, we are analyzing large volume of data at runtime so it is not feasible for us to use a slow model.

So, for analyzing the sentiment we can also use the premade module of the natural language toolkit NLTK which is called "Text blob" which does the job of tokenizing, removing stop words, and analyzing the text as mentioned above.

We also implemented parallelization for sentiment analysis models with the help of 'asyncio' which is a library for implementing parallelization in a python program.

Web-Application Design

The application has 4 tabs: Home, YouTube, Twitter, Reddit. You can click on any tab to mine the data from that social media network.

Figure 3.1: Search bar and various tabs.

After you give the input and press enter, our program will search for the input term in the given social media database and feeds its output as the input to the sentiment analysis model which will then analyze each entry in the data frame parallelly and corresponding to that row it will add a new column and assign the value whether the comment/tweet is positive or negative i.e., 1 or 0.

For mining we use two approaches, one is through the API of that social media network and the other approach we use is based on web scraping.

Then the output data frame is fed to the dashboard program as input, which then plots the interactive graphs using plotly-dash and a callback function to update the graph figure.

Dashboard design

The dashboard is designed on the 'Dash', we made 4 different interactive graphs for each of the social media networks.
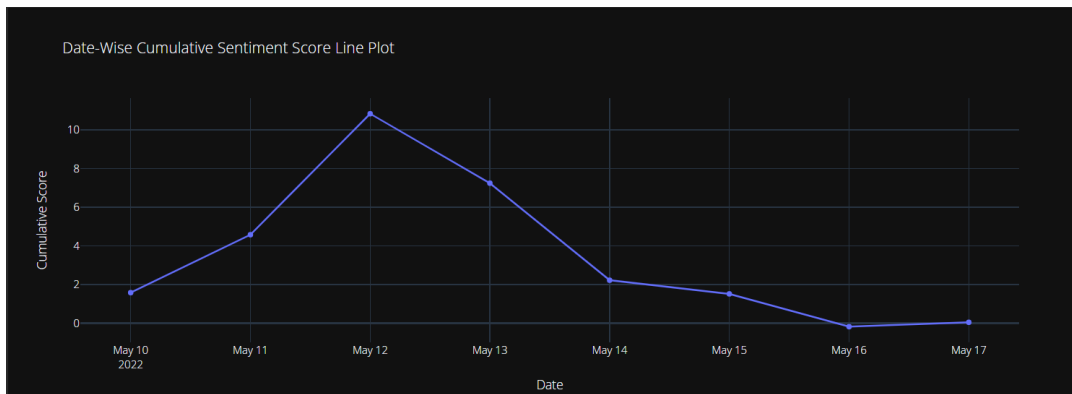

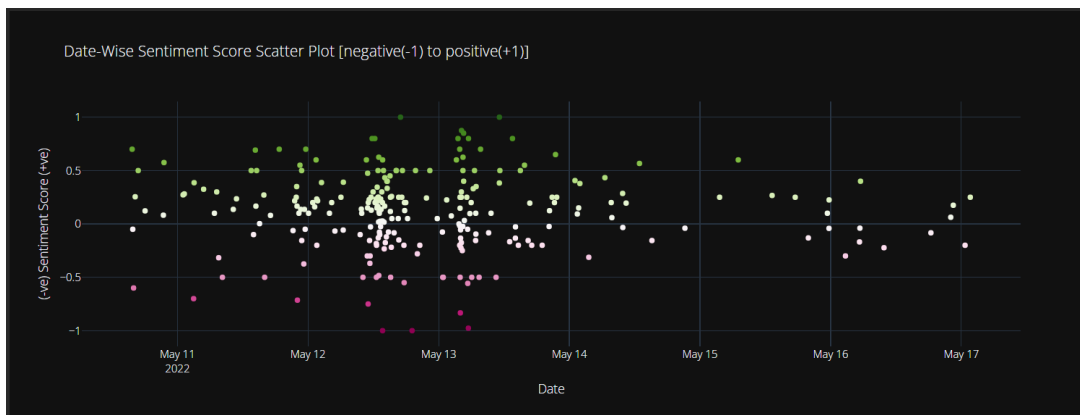Figure 3.2: Line plot (mean sentiment vs. date-time).

Figure 3.3: Scatter plot of each individual comment.

All the graphs and plots are interactive. If you select and drag it will automatically zoom the graph and if you hover over a point in the graph, it'll show the detailed view of that point like date, time, score, etc.

4. CONCLUSION AND FUTURE WORK

Hence, we have successfully implemented the Sentimental analysis application. Here we aimed to provide the information which is mined from social media and analyzed by us, we provided an overview of data from websites like YouTube, Twitter, and Reddit about any particular topic. We were successful in presenting the data in a visual format that provides higher readability and ease of consumption.

The scope of this project is extended to:

- Multinational companies and Firms
- International or National Hotel chains
- Mid-Scale businesses
- Government officials and political parties

Future Work:

- We need to make the application fast by shifting everything on API rather than using the scraping approach.
- We have to add more social media platforms like Facebook, Instagram, etc. to mine on.
- We have to make a mobile application for our project
- We will implement a better machine learning model using TensorFlow

5. ACKNOWLEDGEMENT

6. REFERENCES

[1] Sentiment analysis for Twitter data. URL: https://www.aclweb.org/anthology/W11-0705.pdf
[2] Shrija Madhu, "An approach to analyzing suicidal tendency in blogs and tweets using Sentiment Analysis," International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.4, pp.34-36, 2018.
[3] URL: https://www.youtube.com/watch?v=1gQ6uG5Ujiw
[4] Internet, Google, YouTube, StackOverflow.