A

Major Project

On

MULTI-PLATFORM SENTIMENT ANALYSIS

(Submitted in partial fulfillment of the requirements for the award of Degree)

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

By

SYED YUSUF ALI (187R1A0587)

NOMULA DINAKAR (187R1A05A4)

MASKULA RAHUL SAI (197R5A0504)

Under the Guidance of

M. Madhusudhan
(Assistant Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CMR TECHNICAL CAMPUS

UGC AUTONOMOUS

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi) Recognized Under Section 2(f) & 12(B) of the UGC Act.1956,
Kandlakoya (V), Medchal Road, Hyderabad-501401.

2018-2022

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project entitled "**MULTI-PLATFOM SENTIMENT ANALYSIS**" is being submitted by **SYED YUSUF ALI (187R1A0587), NOMULA DINAKAR (187R1A05A4) & MASKULA RAHUL SAI (197R5A0504)** in partial fulfillment of the requirements for the award of the degree of B. Tech in Computer Science and Engineering to the Jawaharlal Nehru Technological University Hyderabad, is a record of bonafide work carried out by him/her under our guidance and supervision during the year 2021-22.

The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Mr. M. Madhusudhan
Assistant Professor
Internal Guide

Dr. A. Raji Reddy
DIRECTOR

Dr. K. Srujan Raju
HOD

EXTERNAL EXAMINER

Submitted for viva voce Examination held on _____

ACKNOWLEGDEMENT

Apart from the efforts of us, the success of any project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project.

We take this opportunity to express our profound gratitude and deep regard to our guide, **Mr. M. Madhusudhan**, Assistant Professor for his exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing, help and guidance given by him shall carry us a long way in the journey of life on which we are about to embark.

We also take this opportunity to express a deep sense of gratitude to Project Review Committee (PRC) Coordinators: **Mr. A. Uday Kiran, Mr. J. Narasimha Rao, Dr. T. S. Mastan Rao, Mrs. G. Latha, Mr. A. Kiran Kumar,** for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

We are also thankful to **Dr. K. Srujan Raju**, Head, Department of Computer Science and Engineering for providing encouragement and support for completing this project successfully.

We are obliged to **Dr. A. Raji Reddy**, Director for being cooperative throughout the course of this project. We would like to express our sincere gratitude to Sri. **Ch. Gopal Reddy**, Chairman for providing excellent infrastructure and a nice atmosphere throughout the course of this project.

The guidance and support received from all the members of **CMR Technical Campus** who contributed to the completion of the project. We are grateful for their constant support and help.

Finally, we would like to take this opportunity to thank our family for their constant encouragement, without which this assignment would not be completed. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project.

SYED YUSUF ALI (187R1A0587)

NOMULA DINAKAR (187R1A05A4)

MASKULA RAHUL SAI (197R5A0504)

ABSTRACT

The information we aim to provide is mined from social media and analysed, we provide an overview of data from websites like YouTube, Twitter, and Reddit about any particular topic for a requested time frame. This data is presented in a visual format that provides higher readability and ease of consumption. 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 pandemic 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 website gives the viewers an insight into these. The website 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. The language used is Python for the back end.

LIST OF FIGURES

LIST OF SCREENSHOTS

TABLE OF CONTENTS

1. INTRODUCTION

1. INTRODUCTION

1.1 PROJECT SCOPE

This project is titled as “Multi-Platform Sentiment Analysis”. This webapp will provide the user with a facility to analyze the public sentiment regarding any topic of interest, the sentiment will be evaluated from mined comments across multiple social media platforms by the click of a button.

1.2 PROJECT PURPOSE

This project has been developed so that one can visualize the sentiment pertaining to a topic by mining the comments posted by people on posts related to that topic in various social media platforms.

1.3 PROJECT FEATURES

The main features of this project are that it gets the user to type in a keyword which can be any concept or event in the webapp and then comments will be mined from relevant posts about that topic and at the backend, the python code will analyze the sentiment and show it through the user dashboard in the form of graphs, charts and scatter plots.

2. SYSTEM ANALYSIS

2. SYSTEM ANALYSIS

SYSTEM ANALYSIS

System Analysis is the important phase in the system development process. The System is studied to the minute details and analyzed. The system analyst plays an important role of an interrogator and dwells deep into the working of the present system. In analysis, a detailed study of these operations performed by the system and their relationships within and outside the system is done. A key question considered here is, “what must be done to solve the problem?” The system is viewed as a whole and the inputs to the system are identified. Once analysis is completed the analyst has a firm understanding of what is to be done.

2.1 PROBLEM DEFINITION

The world today revolves around how the masses perceive a concept or event. It is of utmost importance to organizations who rely on public perception of their product or policies to know how the promotion of their product and policies is being received by the public and take measures, if needed, to correct their promotion or implementation attempts or to rollback the offered product or policies. In a way this webapp creates a feedback loop for taking corrective measures for such organizations. These organizations could be Multinational companies offering their products through ad campaigns or the government implementing their policies.

2.2 EXISTING SYSTEM

There are not many applications or programs that offer a view into the public sentiment analysis across multiple platforms through a single app with high levels of

readability into the data that has been harnessed and show it in reference to the timeframe from which the data has been collected.

2.2.1 LIMITATIONS OF EXISTING SYSTEM

- Relies on data mostly extracted from a single website.
- Lack of visual aids like graphs, charts and plots to represent the analyzed result.
- No representation of the time-frame of the data that is being extracted which is a major factor in sentiment analysis.
- No option to switch through various platforms to check how the sentiment of a topic differs across different platforms.

To avoid all these limitations and make the process more effortless, there needs to be an integrated webapp.

2.3 PROPOSED SYSTEM

We are proposing a system that extracts data from multiple social media websites, we will provide an overview of data from websites like YouTube, Twitter and Reddit about any particular topic. This data is presented in a visual format that provides higher readability and ease of consumption. The website retrieves through the comments on a particular topic in various social media sites and tells us about the sentiments of the people regarding the topic of interest.

2.3.1 ADVANTAGES OF THE PROPOSED SYSTEM

- Visualizes the true picture of the of the public sentiment regarding a concept or event.
- Shows the timeframe with respect to the comments that are mined for sentiment analysis.
- The data is presented in a visual format with the help of graphs and charts that provides higher readability and ease of consumption for the user.

- Great tool for MNCs, firms or Government officials and political parties to monitor their products and policies.

2.4 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. Three key considerations involved in the feasibility analysis are:

- Economic Feasibility
- Technical Feasibility
- Social Feasibility

2.4.1 ECONOMIC FEASIBILITY

The developing system must be justified by cost and benefit. Criteria to ensure that effort is concentrated on project, which will give best, return at the earliest. One of the factors, which affect the development of a new system, is the cost it would require. The following are some of the important financial questions asked during preliminary investigation:

- The costs conduct a full system investigation.
- The cost of the hardware and software.
- The benefits in the form of reduced costs or fewer costly errors.

Since the system is developed as part of project work, there is no manual cost to spend for the proposed system. Also all the resources are already available, it give an indication of the system is economically possible for development.

2.4.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the

available technical resources. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

2.4.3 BEHAVIORAL FEASIBILITY

This includes the following questions:

- Is there sufficient support for the users?
- Will the proposed system cause harm?

The project would be beneficial because it satisfies the objectives when developed and installed. All behavioural aspects are considered carefully and conclude that the project is behaviourally feasible.

2.5 HARDWARE & SOFTWARE REQUIREMENTS

2.5.1 HARDWARE REQUIREMENTS:

Hardware interfaces specifies the logical characteristics of each interface between the software product and the hardware components of the system. The following are some hardware requirements.

- CPU : i3 and above.
- Hard disk : 100 MB available space.
- RAM : 4GB and above.
- Input devices : Keyboard, Mouse.

2.5.2 SOFTWARE REQUIREMENTS:

Software Requirements specifies the logical characteristics of each interface and software components of the system. The following are some software requirements.

- Operating system : Windows 7, 8, 10.
- Languages : Python, HTML, CSS.
- Tools : Visual Studio Code.

3. ARCHITECTURE

3. ARCHITECTURE

3.1 PROJECT ARCITECTURE

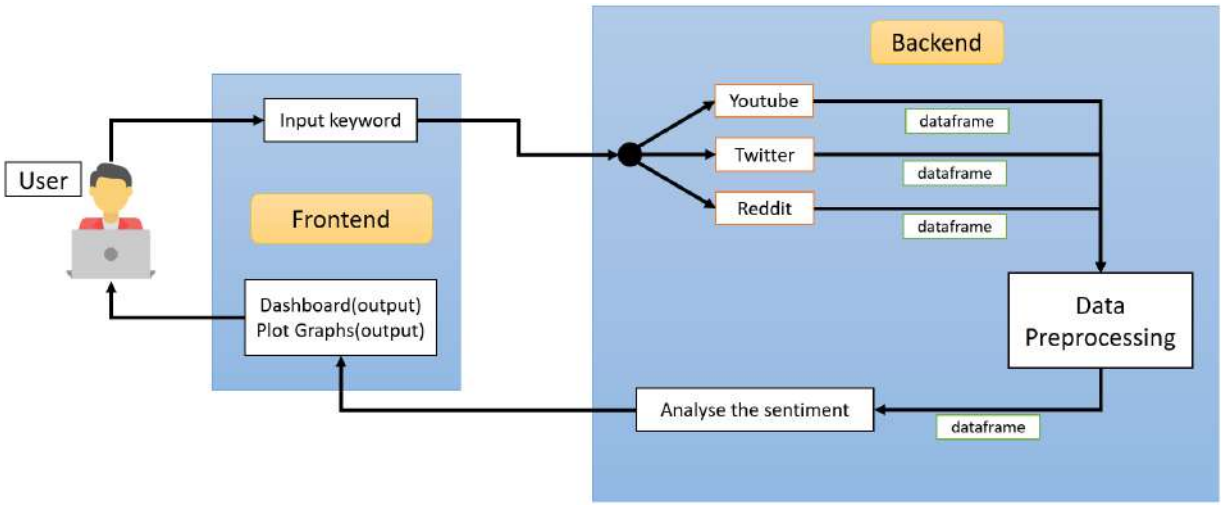


Figure 3.1 Project Architecture for Multi-Platform Sentiment Analysis

3.2 DESCRIPTION

The user is supposed to open the webapp and insert the keyword into the textbox provided at the frontend dashboard. This keyword is used to fetch relevant posts from various multimedia platforms such as YouTube, Reddit and Twitter by their respective APIs and comments from these relevant posts will be mined and preprocessed at the backend by the python code. Then these processed comment strings will be used to assign sentiment score using textblob module for the keyword and show the data through the user dashboard in the form of various graphs and scatter plots on different tabs for different platforms.

3.3 USE CASE DIAGRAM

In the use case diagram, we have basically two actors who are the user and the Analysis module. The user enters the keyword into the webapp, he can do so in one of the multiple tabs provided for mining data from various websites such as YouTube, Twitter and Reddit. The Analysis module assigns sentiment score to the comments that are mined from relevant posts through API, evaluates the cumulative sentiment and shows this data to the user. The user is able to observe this data in graphical format. The user can be a government entity or a multi-national firm.

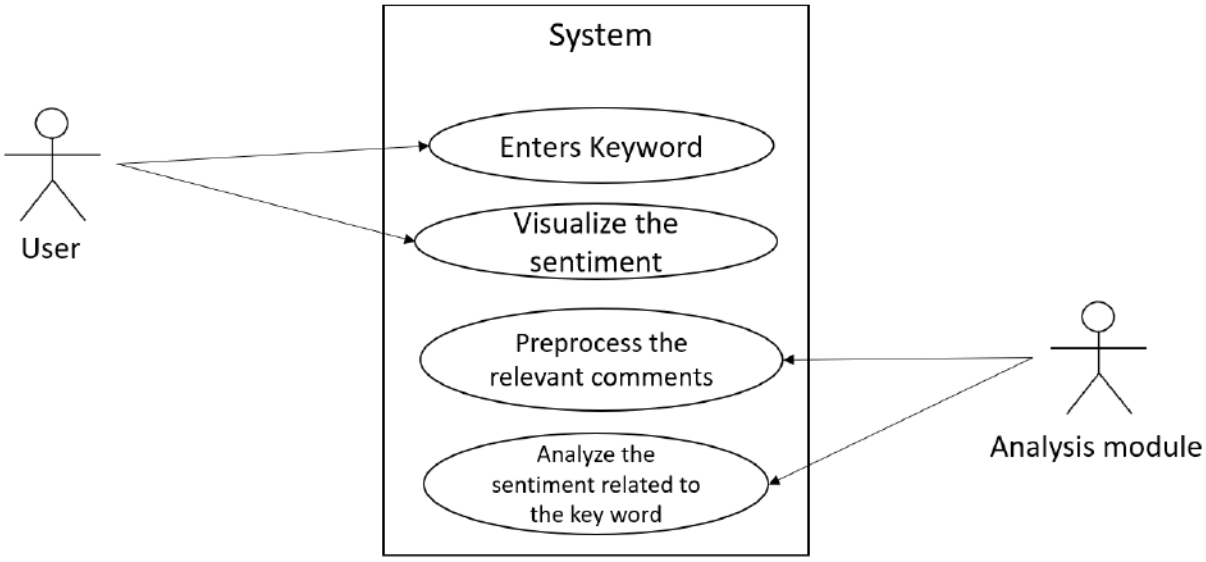


Figure 3.2 Use Case Diagram for Multi-Platform Sentiment Analysis

3.4 CLASS DIAGRAM

Class Diagram is a collection of classes and objects. We have two classes, namely, User and Analysis. The User class contains a variable named ‘keyword’ of type string and a member function, insertKeyword() to accept the user inserted keyword. The Analysis class has a member named ‘comment’ of type string which is relevant to the keyword of the User class, preprocessComment() is a member function to perform preprocessing tasks, analyzeSentiment() function analyzes the sentiment from the preprocessed comment and the showSentiment() function displays the analyzed sentiment to the dashboard in multiple graph format.

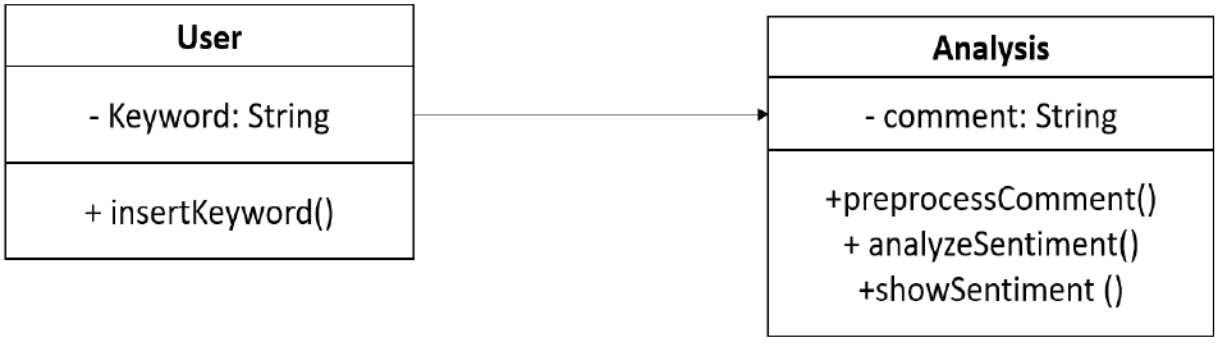


Figure 3.3 Class Diagram for Multi-Platform Sentiment Analysis

3.5 SEQUENCE DIAGRAM

The sequence diagram describes how—and in what order—a group of objects works together. First, the User enters the keyword, the API then proceeds to retrieve the posts relevant to the provided keyword, the comments are mined from these posts and are sent to the Analysis module. The Analysis module analyzes the comments and shows the sentiment evaluated from those comments back to the User in forms of different plots.

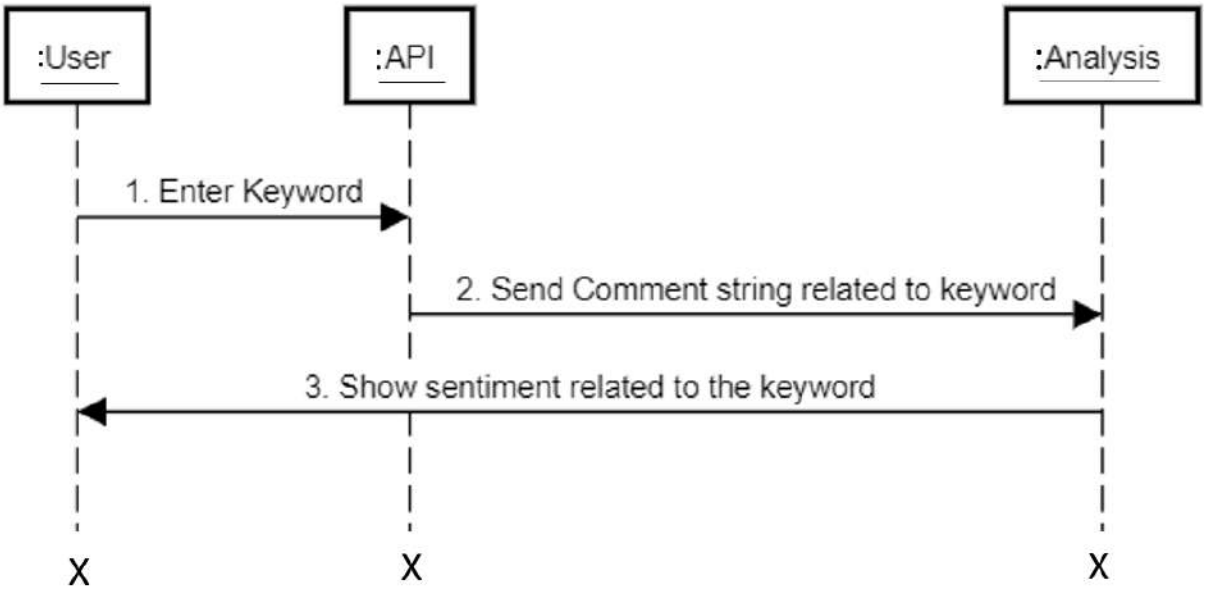


Figure 3.4 Sequence Diagram for Multi-Platform Sentiment Analysis

3.6 ACTIVITY DIAGRAM

Activity diagram describes about flow of activity states i.e. it depicts the behavior of a system. From the initial state, the first activity state comprises of the user entering the keyword. Then flow control goes to the API where comments relevant to the keyword are retrieved and control switches to Analysis, here, the sentiment from the comments is evaluated. After that, the control goes back to user where the evaluated sentiment is viewed and then it terminates at the final state.

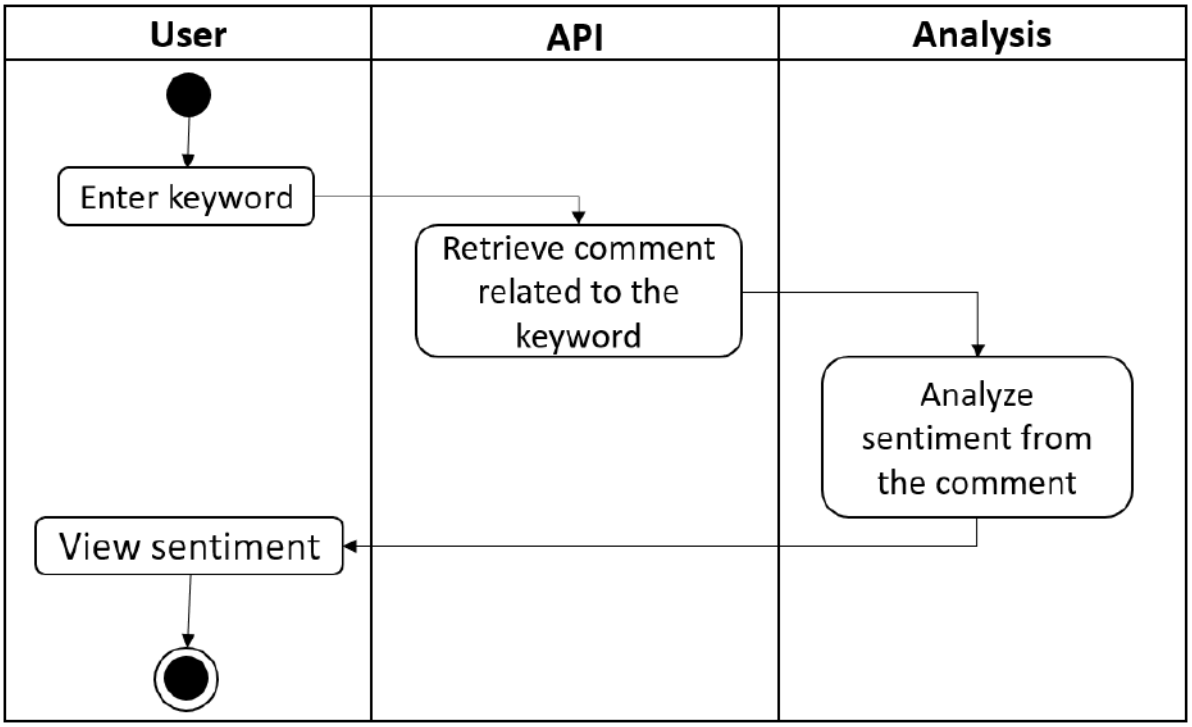


Figure 3.5 Activity Diagram for Multi-Platform Sentiment Analysis

4. IMPLEMENTATION

4. IMPLEMENTATION

4.1 Sample Code

```
from textblob import TextBlob



# ANALYSING SENTIMENT AND RATING THEM FROM -1 TO +1



def analyse_sentiment(df,term):
    f_list = df[term].to_list()
    pol = list()
    for l in f_list:
        analysis = TextBlob(l).sentiment
        t = analysis.polarity
        pol.append(t)
    df['sentiment'] = pol
    pol2 = list()
    for l in pol:
        if (l > 0):
            pol2.append("Positive")
        elif (l < 0):
            pol2.append("Negative")
        else:
            pol2.append("Neutral")
```

```
    df['roundoff'] = pol2
    df = df[df.sentiment != 0]
    return df


def pretty_txt(input_value):
    input_value = input_value.replace(" ", "")
    input_value = input_value.replace("-", "")
    input_value = input_value.replace("[^a-zA-Z#]", " ")
    return (input_value)


# To mine the required data from Reddit


import praw
import pandas as pd


#REDDIT MAIN FUNCTION FOR MINING


reddit = praw.Reddit(client_id='O8o5A', client_secret='4pCcZYQ2-o',
user_agent='Reddit WebScraping')


def top_posts(topic):
    posts=[]
    try:
        f_subreddit = reddit.subreddit(topic)
        for post in f_subreddit.hot(limit=5):
```

```
            posts.append([post.title, post.score, post.id, post.num_comments])
        posts = pd.DataFrame(posts,columns=['title', 'score', 'id', 'num_comments'])
        posts.set_index('title',inplace=True)
        return posts
    except:
        posts.append(["Null","0","0","0"])
        posts = pd.DataFrame(posts,columns=['title', 'score', 'id', 'num_comments'])
        return posts


def to_id_list(posts):
    id_list= posts["id"].tolist()
    return id_list


def mine_comments(id_list):
    comments=[]
    try:
        for i in id_list:
            submission = reddit.submission(id=i)
            submission.comments.replace_more(limit=None)
            for comment in submission.comments.list():

comments.append([submission.title,submission.score,submission.upvote_ratio,comment.
body,comment.score,comment.created_utc])


comments=pd.DataFrame(comments,columns=['title','s_score','upvote_ratio','comments','
c_score','c_date'])
            comments['c_date'] = pd.to_datetime(comments['c_date'],unit='s')
```

```
        return comments
    except:
        comments.append(["Null","0","0","Null","0","0"])

comments=pd.DataFrame(comments,columns=['title','s_score','upvote_ratio','comments','c_score','c_date'])
        return comments
from Reddit import top_posts
from Reddit import to_id_list
from Reddit import mine_comments

import plotly.graph_objs as go

from Senti import analyse_sentiment,pretty_txt
#FUNCTION TO PLOT REDDIT GRAPH


def REDDIT_graph(n_clicks,input_value):
    input_value = pretty_txt(input_value)
    df1=top_posts(input_value)
    list1=to_id_list(df1)
    comment=mine_comments(list1)
    comment=analyse_sentiment(comment,"comments")
    comment['Date'] = comment.apply(lambda row: str(row.c_date).split(" ", 1)[0], axis = 1)
    comment2 = comment.groupby('Date', as_index=False)[['sentiment']].sum()
```

```
data=[]
trace_close = go.Scatter(x = list(comment2.Date),
                y=list(comment2.sentiment),
                mode='lines',
                name="Close",
                line=dict(color="#ff3333"))
data.append(trace_close)
figure1 = go.Figure(data)
figure1.update_layout(
title="Date-Wise Cumulative Sentiment Score Line Plot",
xaxis_title="Date",
yaxis_title="Cumulative Score",
template='plotly_dark',
plot_bgcolor= 'rgba(0, 0, 0, 0)'
)

data=[]
trace_close = go.Scatter(x = list(comment.c_date),
                y=list(comment.sentiment),
                mode='markers',
                name="markers",
                marker_color=comment['sentiment'])
data.append(trace_close)
figure2 = go.Figure(data)
figure2.update_layout(
```

```
title="Date-Wise Sentiment Score Scatter Plot [negative(-1) to positive(+1)]",
xaxis_title="Date",
yaxis_title="(-ve) Sentiment Score (+ve)",
template='plotly_dark',
plot_bgcolor= 'rgba(0, 0, 0, 0)'
)

data=[]
trace_close = go.Box(x = list(comment.Date),
                     y=list(comment.sentiment),
                     name="Close",
                     line=dict(color="#ff3333"))
data.append(trace_close)
figure3 = go.Figure(data)
figure3.update_layout(
title="Date-Wise Sentiment Score Box Plot [negative(-1) to positive(+1)]",
xaxis_title="Date",
yaxis_title="(-ve) Sentiment Score (+ve)",
template='plotly_dark',
plot_bgcolor= 'rgba(0, 0, 0, 0)'
)

data=[]
trace_close = go.Pie(labels=list(comment.roundoff),
                     name="Close"
```

```
                    )
    data.append(trace_close)
    figure4 = go.Figure(data)
    figure4.update_layout(
    title="Pie-Chart",
    template='plotly_dark',
    plot_bgcolor= 'rgba(0, 0, 0, 0)'
    )
    figure = (figure1,figure2,figure3,figure4)

    return figure

import dash
import dash_core_components as dcc
import dash_html_components as html
from layout import html_layout

from twittergraph import TWT_graph
from youtubegraph import YT_graph
from redditgraph import REDDIT_graph


global yt_vid_comments
yt_vid_comments = []
```

```
app = dash.Dash(__name__)

app.index_string = html_layout

tabs_styles = {
    'height': '44px',

}
tab_style = {
    'border': '3px solid #111111',
    'padding': '6px',
    'fontWeight': 'bold',
    'background':'#111111'
}

tab_selected_style = {
    'borderTop': '3px solid #000000',
    'borderBottom': '3px solid #282828',
    'color': '#ffffff',
    'padding': '6px',
    'fontWeight': 'bold',
    'background':'#282828'

}
```

#LAYOUT OF THE MAIN DASH APP

```
app.layout = html.Div([
   dcc.Tabs(value='tab-0',children=
   [dcc.Tab(label='Home', value='tab-0', style=tab_style,
selected_style=tab_selected_style, children=[
        html.P(""),
        html.H3("Multi-platform Sentiment Analysis"),
        html.Div([html.H4("Know what the world thinks!"),
            html.P("We are focused to paint the true picture of the world for you. 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."),
            html.H4("So what am I seeing?"),
            html.P("-- Youtube : 100 most relevant comments each from across 10 most
popular videos, "),
            html.P("for instance, 100x10=1000 comments !"),
             html.P("-- Twitter: around 1000 most relevant tweets!"),
            html.P("-- Reddit : 150 most relevant comments each from across 5 most
active subreddits,  "),
            html.P("for instance, 150x5=750 comments !"),
            html.H4("What is the significance?"),
            html.P("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."),

html.P("Note: This data is changing every second and hence the results take time to analyze and convert this data into a consumable format so a little Patience will be appreciated")

])

]),

dcc.Tab(label='YouTube', value='tab-1', style=tab_style, selected_style=tab_selected_style, children=[

html.Div([

#html.Div(html.H1(children="Team Zion")),

html.P(""),

html.H3("Enter the term you want to analyse"),

html.Div([

dcc.Input(

id = "yquery-input",

placeholder = "Enter the query you want to search",

type = "text",

value = "Covid19",

style={"margin-right": "15px"}

),

html.Button('Submit', id='ysubmit-val', n_clicks=0),

html.P(""),

]),

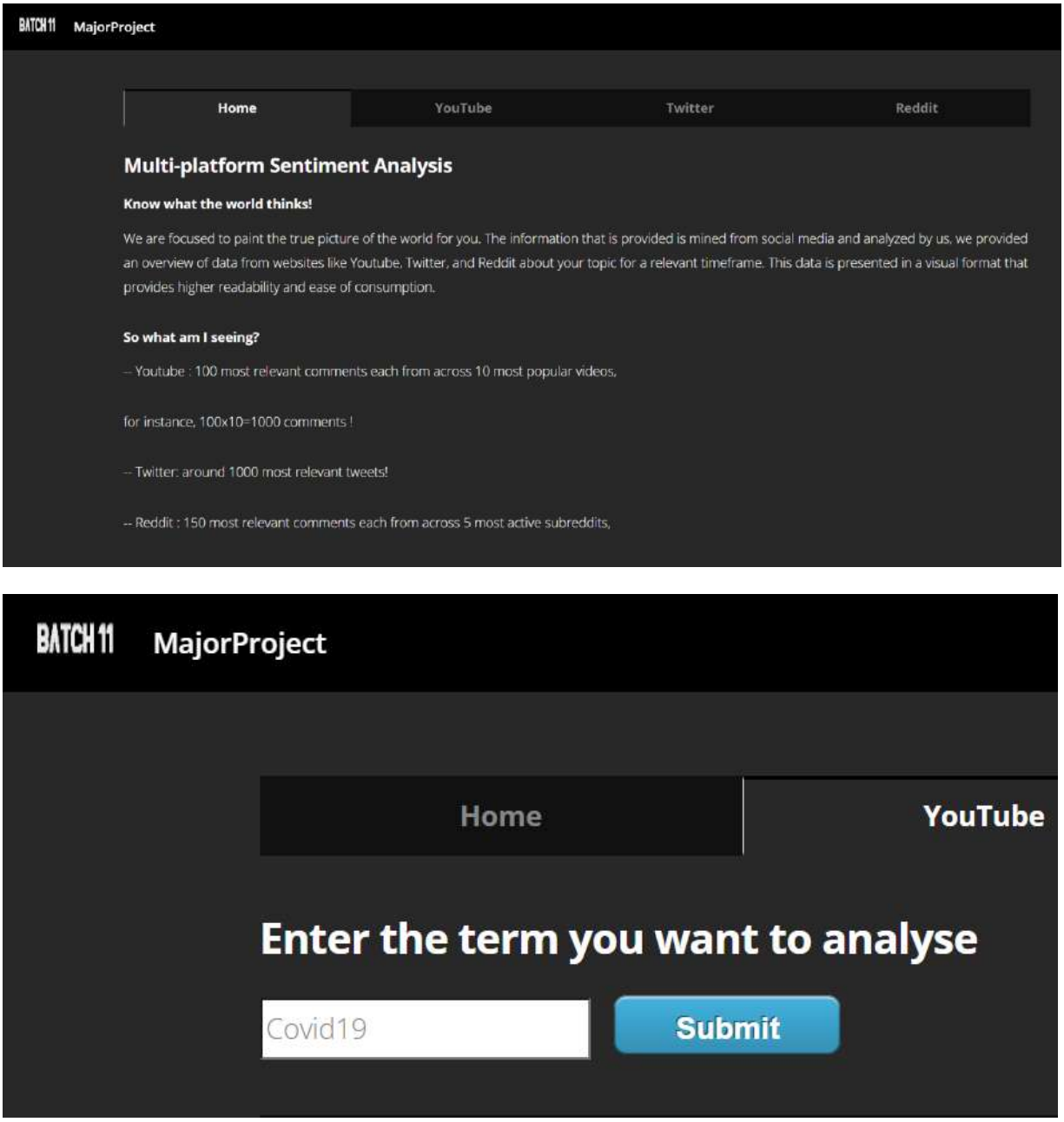
```
        html.Div(
            dcc.Graph(id="y-graph1", config={'displayModeBar': False})
            ),
        html.P(""),
        html.Div(
            dcc.Graph(id="y-graph2",)
            ),
        html.P(""),
        html.Div(
            dcc.Graph(id="y-graph3",)
            ),
        html.P(""),
        html.Div(
            dcc.Graph(id="y-graph4",)
            )

    ])
])
```

5. SCREENSHOTS

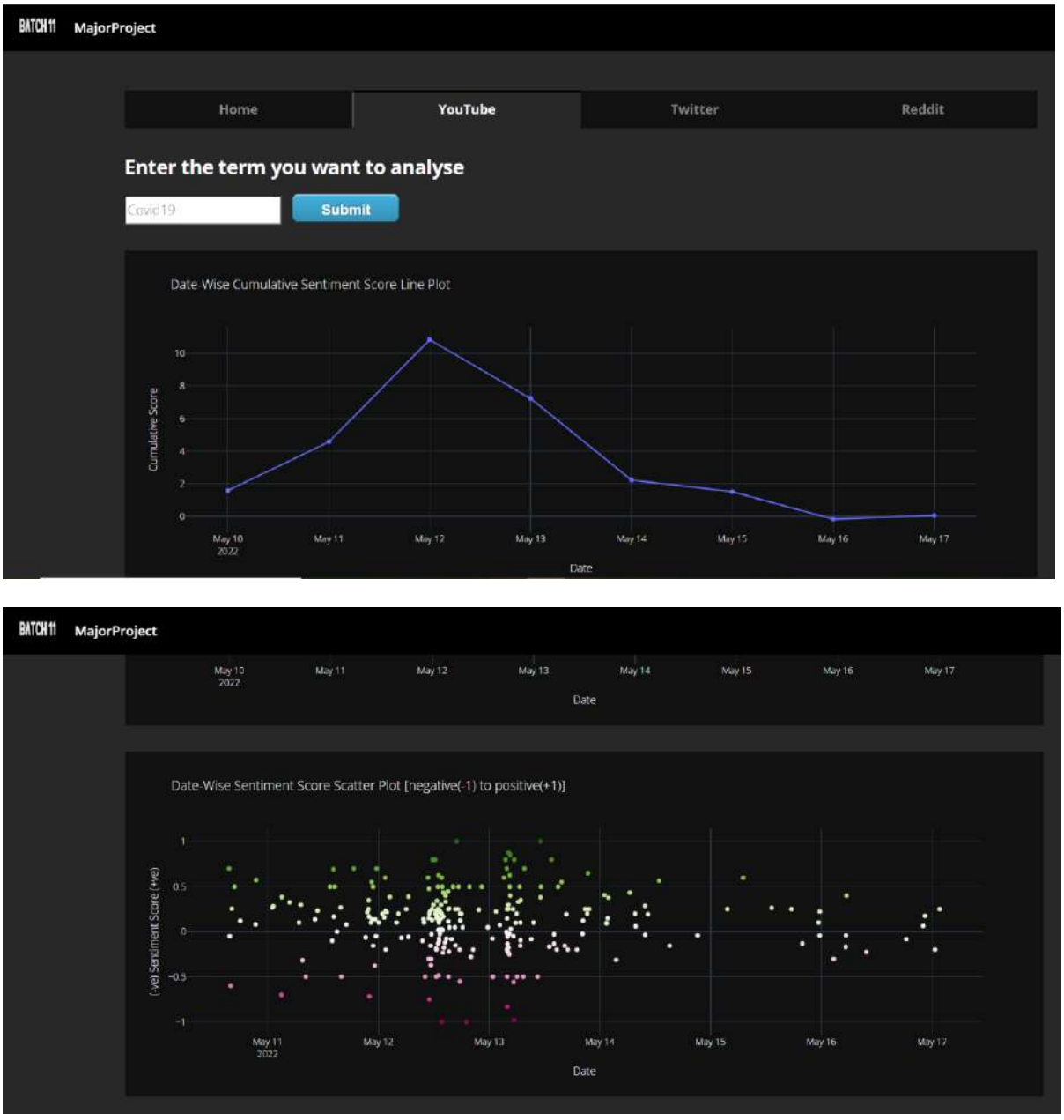
5. SCREENSHOTS

5.1 HOME TAB AND TEXTBOX



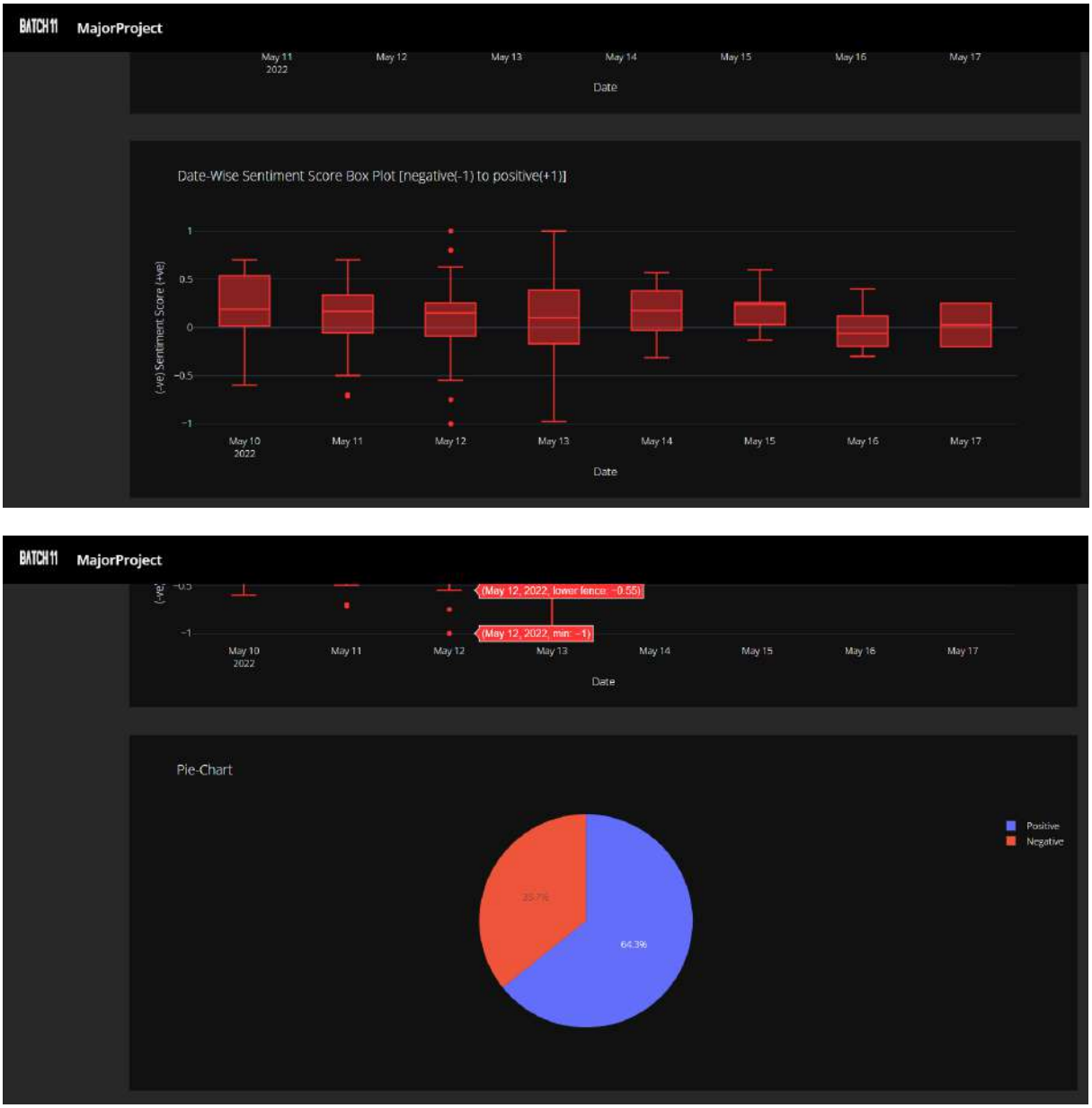
Screenshot 5.1 Home Tab and Textbox

5.2 LINE PLOT AND SCATTER PLOT



Screenshot 5.2 Line plot and Scatter plot

5.3 BOX PLOT AND PIE CHART



Screenshot 5.3 Box plot and Pie chart

6. TESTING

6. TESTING

6.1 INTRODUCTION TO TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of tests. each test type addresses a specific testing requirement.

6.2 TYPES OF TESTING

6.2.1 UNIT TESTING

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

6.2.2 FUNCTIONAL TESTING

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases.

6.3 TEST CASES

6.3.1 USER DASHBOARD

Test case ID	Test case name	Purpose	Input	Output
1	User types a meaningful topic in the textbox.	To check if results are returned for correct word.	“Covid19”.	Sentiment graphs for the term “Covid 19”.
2	User types in a meaningless combination of letters.	To check if results are returned for invalid term.	Invalid term (random combination of letters).	No change in user dashboard plots.
3	User types in multiple valid terms (space separated) in the textbox.	To check if results are returned for multiple terms which are space separated.	“Covid Russia War”	Sentiment graphs for the cumulative score of the terms is returned.
4	User types in same term on different dates.	To check if results are reflected across the timeframe correctly.	“Covid19”	Different sentiment values are observed on different dates.

7. CONCLUSION

7. CONCLUSION & FUTURE SCOPE

7.1 PROJECT CONCLUSION

This project is an interactive platform that tells you what people are thinking about a given input. 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. This web app is focused on painting the true picture of the public sentiment on a concept or event.

7.2 FUTURE SCOPE

In future, this project can be enhanced by:

- Giving user the flexibility to choose the specific timeframe from which the data needs to be extracted.
- Allowing the user to select the number of comments that are to be mined.
- Adding more platforms (e.g. Instagram, etc.).

8. BIBLIOGRAPHY

8. BIBILOGRAPHY

8.1 WEBSITES

[1] https://flask.palletsprojects.com/en/2.1.x/

[2] https://dash.plotly.com/introduction l

[3] https://docs.python.org/3/tutorial/

[4] https://textblob.readthedocs.io/en/dev/

[5] https://developers.google.com/youtube/v3/getting-started

[6] https://developer.twitter.com/en/docs

8.2 GITHUB LINK

https://github.com/trexyuzi/Multi-Platform-Sentiment-Analysis

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

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

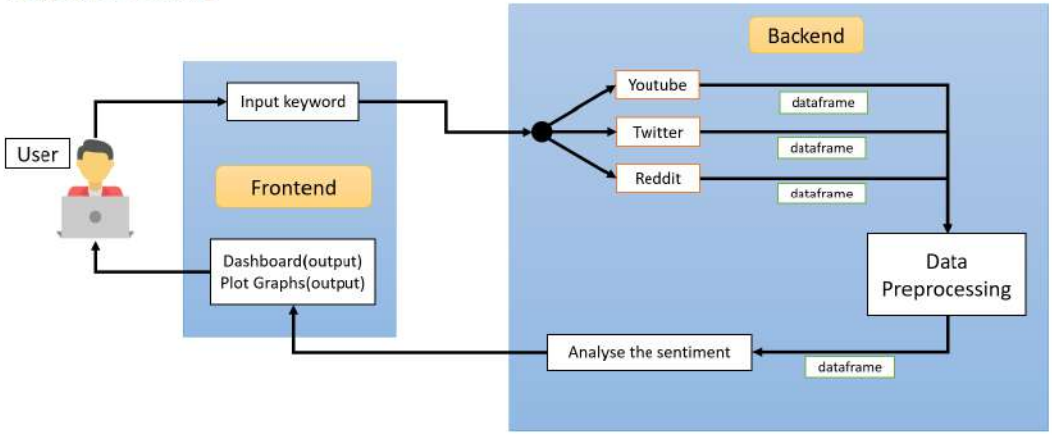


Figure 2.1: Architecture diagram.

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.

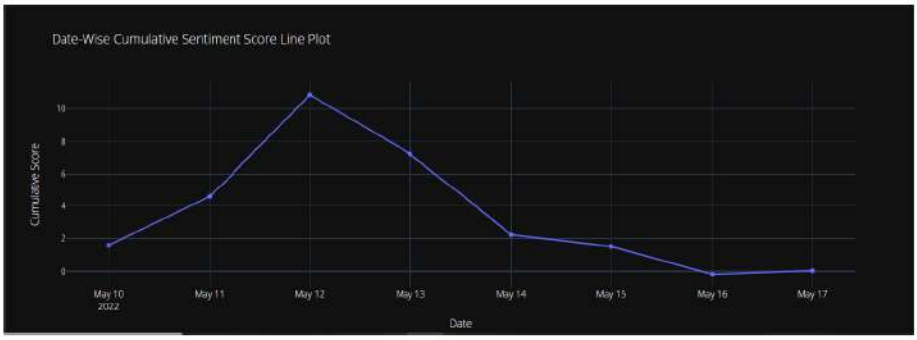


Figure 3.2: Line plot (mean sentiment vs. date-time).

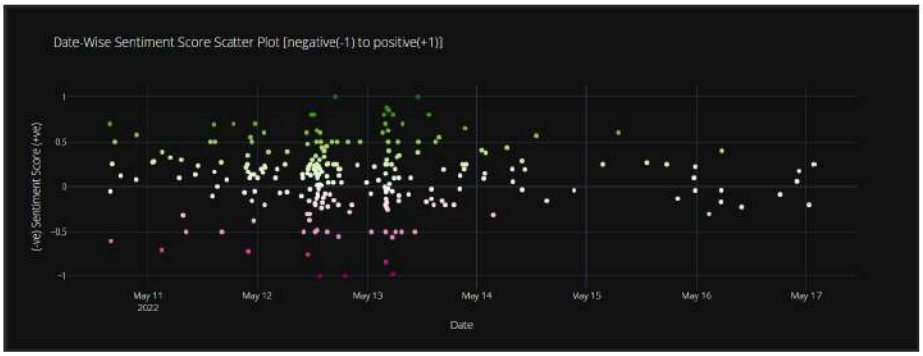


Figure 3.3: Scatter plot of each individual comment.

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

The success of this paper includes help from our guide as well. We are grateful to our guide, Mr. M. Madhusudhan, Assistant professor, CMR Technical Campus, for his expertise that assisted us in our research.

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.

ISSN: 2582-3930

International Journal of Scientific Research in Engineering and Management

is hereby awarding this certificate to

Syed Yusuf Ali

in recognition the publication of manuscript entitled

MULTI-PLATFORM SENTIMENT ANALYSIS

published in Ijsrem Journal Volume 06 Issue 06 June 2022

www.ijsrem.com

Editor in Chief
E-mail: editor@ijsrem.com

ISSN: 2582-3930

International Journal of Scientific Research in Engineering and Management

is hereby awarding this certificate to

Nomula Dinakar

in recognition the publication of manuscript entitled

MULTI-PLATFORM SENTIMENT ANALYSIS

published in Ijsrem Journal Volume 06 Issue 06 June 2022

www.ijsrem.com

Editor in Chief
E-mail: editor@ijsrem.com

ISSN: 2582-3930

International Journal of Scientific Research in Engineering and Management

is hereby awarding this certificate to

Maskula Rahul Sai

in recognition the publication of manuscript entitled

MULTI-PLATFORM SENTIMENT ANALYSIS

published in Ijsrem Journal Volume 06 Issue 06 June 2022

www.ijsrem.com

Editor in Chief
E-mail: editor@ijsrem.com