

Social Sentiment Profiling on Public Health and Safety using AI/ML

Tanisha Bhola

B.E student, Electronics and Telecommunication, Vivekanand Education Society's Institute Of Technology, Mumbai, India

*

Abstract – In today's rapidly growing landscape of social data, researchers and organizations are increasingly tapping into online conversations for valuable insights. This includes understanding sentiments around crucial topics like the global COVID-19 outbreak, crime, depression, and other societal issues. By analyzing data from platforms like Twitter, Reddit, and Facebook using Python and machine learning models, sentiment analysis helps classify user posts as positive, negative, or neutral. This approach aids organizations in making informed decisions without offending public sentiment. Leveraging Python's versatility, sentiment analysis is performed on Twitter data using libraries like Tweepy and TextBlob. Graphical representations provide insights into sentiment trends around hashtags such as COVID-19, crimes, depression, and others.

Key Words: Crime-related Keywords: Crime, Theft, Assault, Violence, Fraud
COVID-19-related Keywords: COVID-19, Lockdown, Pandemic

1. INTRODUCTION

Microblogging emerged as a vital tool for rapid communication, especially as mobile devices became the dominant platform for accessing online content, with Twitter being a prominent player in this space. With its real-time updates and concise format, Twitter quickly became a global phenomenon, enabling users to share thoughts, news, and queries effectively. Its impact extends beyond casual updates, with users often breaking news and mobilizing movements during major events. The platform's demographics skew towards younger age groups, but it maintains a significant presence across various demographics. Twitter's influence on society is profound, from shaping public discourse to influencing decision-making in both public and private sectors. Leveraging social media data for sentiment analysis provides invaluable insights for organizations, helping them make informed decisions, improve services, and foster better relationships with citizens. This data-driven approach not only enhances organizational

efficiency but also contributes to the overall betterment of society by addressing its needs and aspirations.

1.1 LITERATURE REVIEW

1. Pokharel, Bishwo Prakash (2020):

- This study concentrates on Twitter sentiment analysis during the COVID-19 outbreak in Nepal. By collecting and pre-processing tweets using regular expressions in Python, the research aimed to understand the sentiments prevalent among Nepali Twitter users.

- The findings highlighted a mix of emotions among the population, with a predominantly positive and hopeful outlook, alongside instances of fear, sadness, and disgust. Such insights are invaluable for understanding public perception during crises and informing decision-making processes.

2. Pak, Alexander and Paroubek, Patrick:

- The research conducted by Pak and Paroubek revolves around the use of the Twitter API to gather a corpus of text posts. This corpus was then categorized into three classes: positive sentiments, negative sentiments, and objective texts.

- The ultimate goal was to train a classifier capable of distinguishing between positive and negative sentiments expressed in tweets. By leveraging machine learning techniques, the study aimed to develop a model that could accurately classify the sentiment conveyed in Twitter posts.

3. Chandra, Y. and Jana, A.:

- This paper provides insights into sentiment analysis techniques beyond traditional machine learning approaches. It emphasizes the diverse range of methodologies available for sentiment analysis, including machine learning and deep learning techniques.

- Additionally, the study highlights the potential of using Twitter data obtained through Python scripts, such as Tweepy, for sentiment analysis. By extracting sentiments from Twitter data, organizations can gain valuable insights into public opinion on various topics of interest.

4. Yadav, Nikhil et al.:

- The research conducted by Yadav and colleagues focuses on utilizing Twitter sentiment analysis for product evaluation. Companies increasingly rely on microblogs like Twitter to gauge public sentiment towards their products.
- By analyzing user reactions on Twitter, organizations can gather valuable feedback and insights to improve their products and services. The study underscores the importance of excluding retweets from analysis to focus solely on original user-generated content for more accurate sentiment analysis.

5. Jayasanka, R. A. S. C. et al.:

- The research from the University of Moratuwa in Sri Lanka emphasizes the significance of thorough sentiment analysis to obtain accurate insights. It introduces a novel approach to sentiment analysis using Adverb-Adjective Combinations (AACs) to gauge the strength of subjective expressions in sentences.
- Unlike traditional methods that rely on specific parts of speech, this approach offers a nuanced understanding of sentiment intensity. By considering AACs, the study aims to provide more accurate sentiment analysis results, thereby enhancing the reliability of sentiment analysis methodologies.

2. PROBLEM STATEMENT

Title: Twitter Opinion Mining and Sentiment Analysis

The complexity of human language presents challenges in teaching machines to accurately analyze sentiments in textual data, including grammatical nuances, cultural variations, slang, and misspellings. Sentiment analysis, also known as opinion mining, is a crucial natural language processing technique used to interpret and classify emotions in subjective data, such as emails, survey responses, and social media posts. This process is especially useful in social media monitoring, offering organizations valuable insights into public opinion on various topics.

Steps Involved

1. Scraping Data from Social Platforms:

- Data from social platforms, particularly Twitter, is scraped using the Scala library "Twitter4j" and

the Twitter API. This involves registering a developer account with Twitter and providing necessary parameters such as API Key, Access token key, Consumer Secret key, and Access Secret Token Key.

- Filters are applied to retrieve tweets related to specific keywords, enabling targeted data collection.

2. Cleaning of Scraped Data:

- Data processing involves tokenization, where tweets are split into individual words or tokens, typically using whitespace or punctuation characters. This facilitates further analysis using techniques like the bag-of-words model.

- Data filtering involves removing stop words, numbers, and punctuations, as well as stemming to reduce derived words to their roots. This enhances the quality of the dataset by eliminating irrelevant information.

3. Feature Extraction:

- Feature extraction techniques such as TF-IDF (Term Frequency-Inverse Document Frequency) are employed to assess the importance of terms in the corpus. This facilitates the identification of trending topics and the creation of word clouds, aiding in data visualization and analysis.

4. Sentiment Analysis:

- Sentiment analysis is performed using a custom algorithm to determine the polarity of tweets.

- Polarity is determined by counting positive and negative words in each tweet and comparing them against predefined lists. Positive sentiment leads to higher scores, while negative sentiment decreases scores.

Tools and Technologies

- Python: Used for its versatility and libraries suitable for data science projects.
- Flask: A Python web framework facilitating the development of web applications.
- Bootstrap: A HTML, CSS & JS Library for simplifying web page development.
- Tailwind CSS: A utility-first CSS framework for building custom designs.
- HTML: Standard markup language for creating structured web pages.

This sentiment analysis project comprises four main modules: scrapper, sentiment analysis tool, algorithm, and dashboard. The scrapper gathers

social media data, the sentiment analysis tool analyzes sentiments using machine learning techniques, and the dashboard presents the analyzed data for visualization and interpretation.

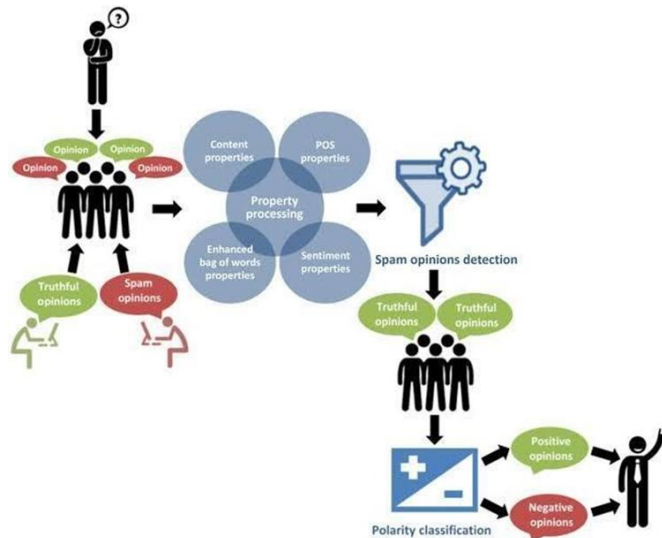


Fig: Public opinion classification

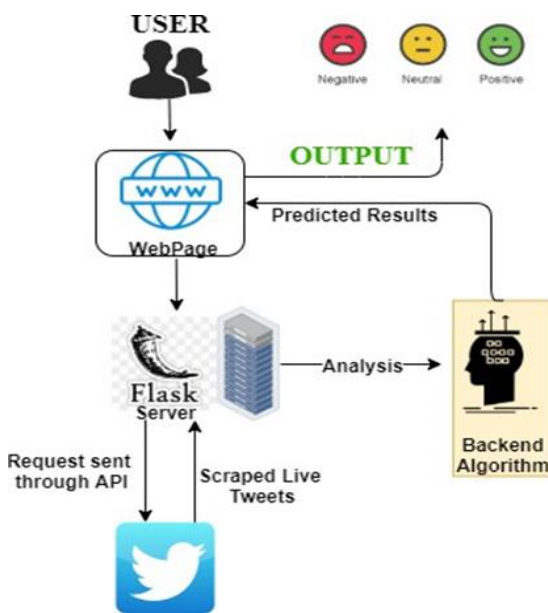


Fig: Block Diagram for Sentiment Analysis

The ability to extract insights from social data through sentiment analysis is invaluable for organizations worldwide, enabling informed decision-making and a deeper understanding of public sentiment.

Moving to the algorithm part, at the beginning we have generated

- API Key,
- Consumer Secret key,
- Access token key,
- Access Secret Token Key.

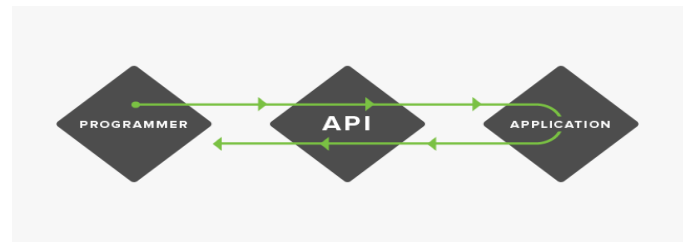
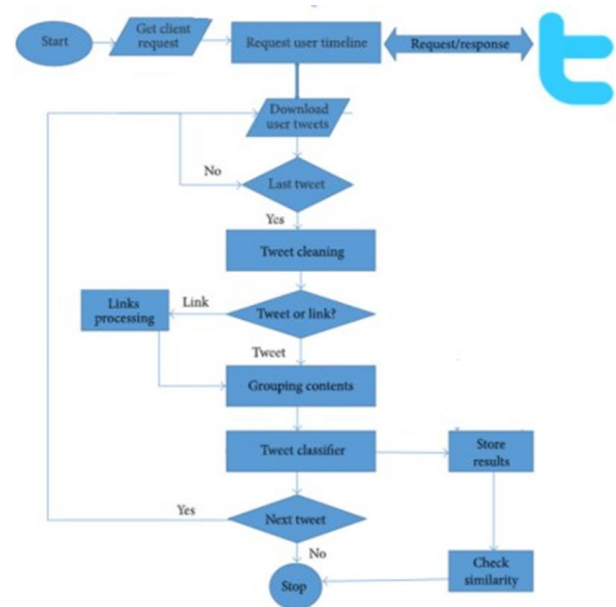


Fig: Twitter API

The Twitter API serves as a powerful tool for accessing and interacting with Twitter data, allowing developers to compose tweets, read profiles, and access high volumes of tweets on specific topics or locations. By leveraging various APIs, developers can create fully automated systems to interact with Twitter.

1. Backend Algorithm Flowchart



The backend algorithm for sentiment analysis involves four main steps: data scrapping, data cleaning, segregation, and graphical representation. Data is first scraped from Twitter using the Tweepy library in Python, followed by cleaning operations to remove noise such as punctuation, digits, and links. The cleaned data is then processed for sentiment analysis, and results are continuously stored and processed until a defined count is reached.

The integration of backend and frontend functionalities is depicted in the block diagram. User input, such as hashtags, is taken and integrated with the Flask server, which houses the backend algorithms. The server interacts with Twitter using

generated API and consumer keys, scrapes tweets, performs sentiment analysis, and returns results to the webpage for display.

2. Dashboard

The dashboard presents graphical representations of sentiments extracted from social media platforms like Twitter and Instagram. It allows users to search for sentiments related to trending topics, with the current project focusing on COVID-19 and crimes.

Through this system, users gain insights into public sentiments, categorized as positive, negative, or neutral, facilitating informed decision-making and understanding of public opinion trends.

3. Implementation

3.1 Importing Libraries

The implementation begins by importing necessary libraries such as numpy, pandas, tweepy, TextBlob, and Matplotlib. These libraries provide functionalities for array operations, data manipulation, accessing Twitter API, text processing, and data visualization.

A. Scrapping of Data from Social Platform
Twitter data is scraped using Tweepy library, establishing authentication and connection with Twitter API. Tweepy OAuthHandler authenticates the app using access tokens. Tweets are collected based on specified keywords or filters, with properties like tweet text, ID, and time saved for analysis.

Create an application

Application Details

Name *

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description *

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website *

Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens. (If you don't have a URL yet, just put a placeholder here but remember to change it later)

Callback URL

Where should we return after successfully authenticating? OAuth 1.0a applications should explicitly specify their oauth_callback URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

Developer Agreement
☐ Yes, I have read and agree to the [Twitter Developer Agreement](#).

Fig: Twitter application development

Your Access Token
This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access Token

Access Token Secret

Access Level Read and write

Owner

Owner ID

Fig: Twitter API key generation

B. Number of Tweets to be Scraped

The number of tweets to be scraped is determined using Tweepy API, specifying the count parameter in the search function. Tweets are filtered based on language, and a JSON response is received from Twitter, containing tweet properties.

C. Tweet Cleaning

Scraped tweets undergo cleaning using the cleantext function, removing noise like punctuation, digits, and links. Word embeddings reduce the need for extensive text cleaning, capturing relative meaning within the training text.

D. Classification of Data

TextBlob library is used for sentiment analysis, classifying tweets into positive, negative, or neutral categories based on polarity and subjectivity scores. Results are stored and displayed, allowing insights into public sentiment.

3.2 Integration

3.2.1 Interfacing Role of Flask

Flask is used for interfacing between the backend and frontend. Various Flask tools like render_template, redirect, and request facilitate HTML generation, redirection, and request handling, simplifying website development.

3.2.2 Frontend

HTML, CSS, and Bootstrap are utilized for frontend development, comprising pages like Home, About, and Team. The Home page enables users to input hashtags or Twitter handles for analysis, while the About page provides project details, and the Team page lists project members.

3.2.3 Integration of Frontend with Backend and Website Link Generation

Flask integrates frontend with backend functionalities, generating website links using URL routing. Flask's route decorator and add_url_rule function define URL rules, passing variable parts as

keyword arguments to view functions. Trailing slashes ensure URL uniqueness, while relative link targets enhance website navigation.

Through this implementation, users can efficiently analyze Twitter data, extract sentiments, and visualize trends, fostering informed decision-making and understanding of public opinions.

4. RESULT

4.1 Output of backend part

```

Tweets ... Time
0 If we can't get a Covid-19 relief package thro... 2020-12-04 06:51:32
1 Joe Ageyo\nMurathe\nRuto\n\nMy boss,covid-19 m... 2020-12-04 06:51:32
2 RT @TheMikeOliver: What have you learnt from ... 2020-12-04 06:51:32
3 RT @KannadaGrahaka: Industry experts say the d... 2020-12-04 06:51:31
4 RT @narangmedical: In our endeavor to contain ... 2020-12-04 06:51:31
...
94 @LisaMarieBoothe The worst thing anyone can do... 2020-12-04 06:50:53
95 RT @stanleywidianto: Dr. Sardjono, a radiologi... 2020-12-04 06:50:53
96 Investigation launched as 44 cases of Covid-19... 2020-12-04 06:50:53
97 Yo: Mullen Sports Medicine's Mahmood and Garci... 2020-12-04 06:50:52
98 RT @POPSUGARLatina: After recovering from COVI... 2020-12-04 06:50:52

[99 rows x 3 columns]
Positive: 39.0 %
Negative: 17.0 %
Neutral: 44.0 %
2020-12-04 06:51:32

```

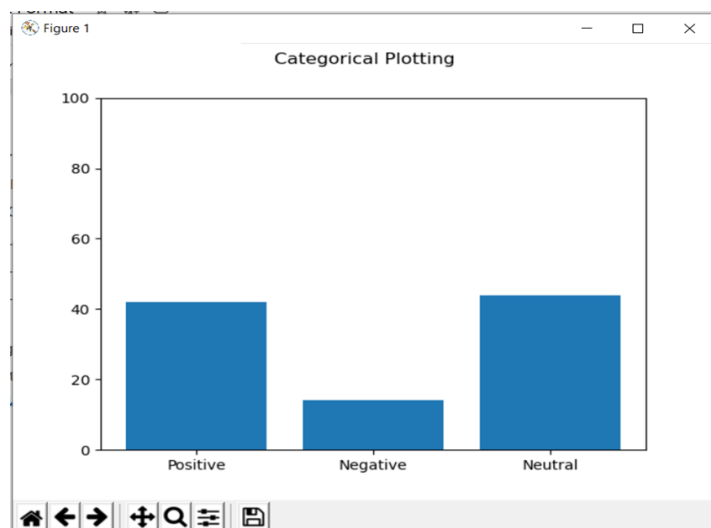


Fig: Output of backend part

4.2 Output of frontend part

4.2.1 Layout of created web

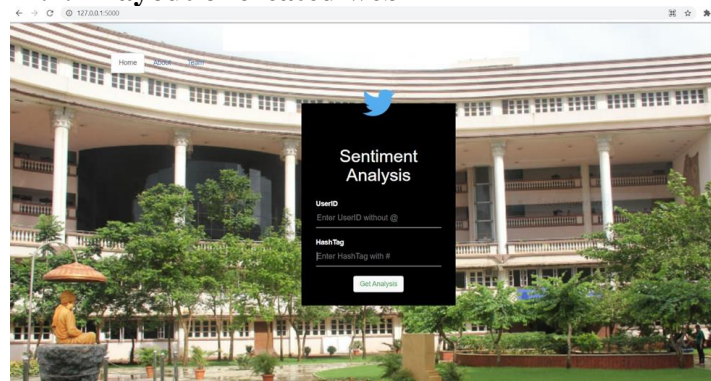
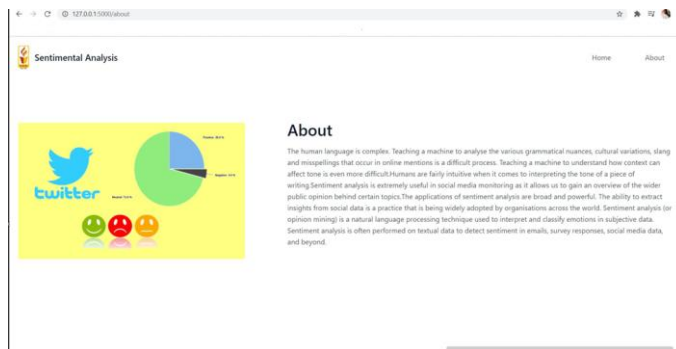


Fig: Frontend Home page



4.2.2 Sentiment analysis on keyword lockdown

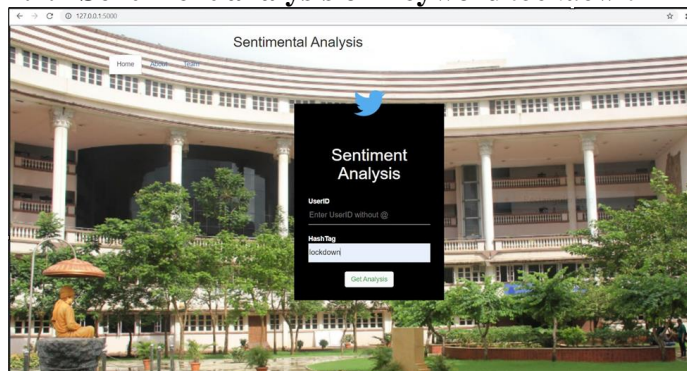
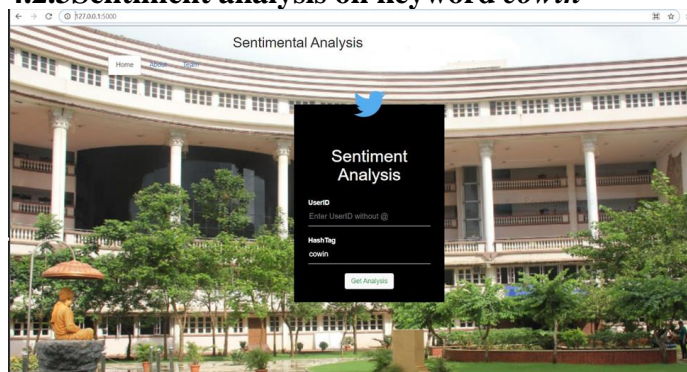


Fig: Sentiment analysis on keyword lockdown

4.2.3 Sentiment analysis on keyword cowin



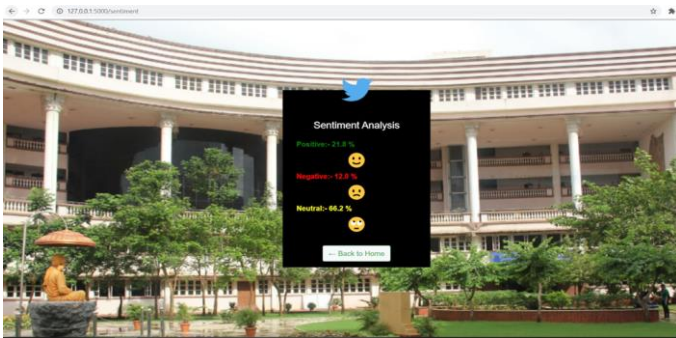


Fig: Sentimental analysis on keyword cowin

4.2.4 Sentiment analysis on keyword *Remdesevir*

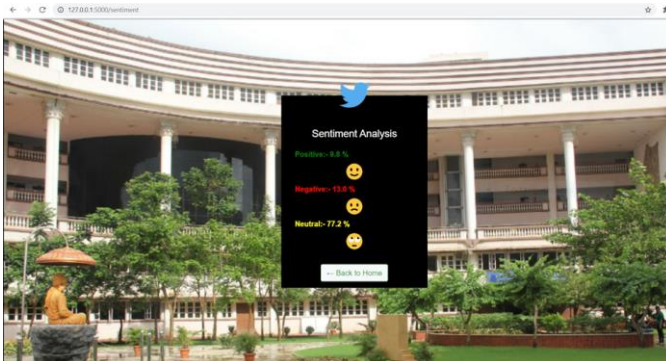
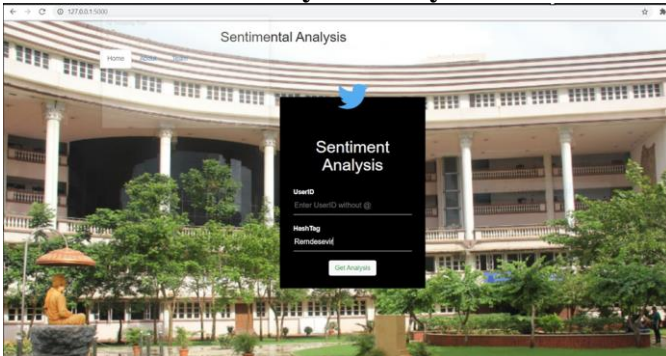


Fig: Sentimental Analysis on keyword Remdesevir

4.2.5 Sentiment analysis on UserID *Narendramodi*

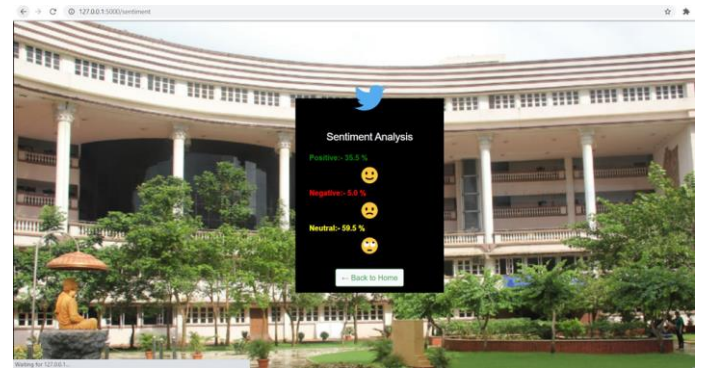
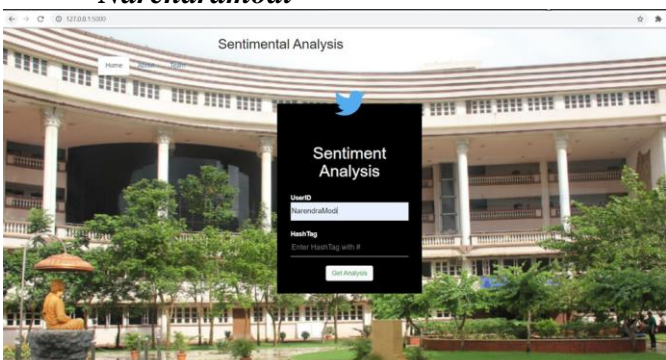


Fig: Sentiment Analysis on UserID Narendramodi

After plotting numeric information from Twitter data, word clouds are generated for three categories: neutral, positive, and negative sentiments. The word clouds depict words commonly found in each category, reflecting people's opinions and reactions during the outbreak.

- Neutral Category: Words like "flu," "coronavirus," and "novel" are common but don't lean towards positivity or negativity.
- Positive Category: Terms such as "new cases," "treatment," and "vaccine" signify positive actions and efforts during the pandemic.
- Negative Category: Words like "pandemic," "death," and "fear" highlight the challenges and negative impact of the crisis.
- Other categories include expressions of anger or frustration towards China's handling of the pandemic and emotions related to increasing death rates globally.

These word clouds provide insights into public sentiment and help classify text data into relevant categories based on the prevailing emotions expressed.

5. CONCLUSIONS

Sentiment analysis is widely applicable across various domains, including blog articles, product reviews, social media platforms, and movie reviews. Implementing sentiment analysis using approaches like SemiWandNet is often easier than traditional classifier methods. The commercial interest in sentiment analysis is strong, as companies seek to understand public perception of **their products and** consumers rely on reviews to make informed decisions.

The agenda-setting theory of media highlights how media influences public opinion by selectively presenting information. Social networking sites play a significant role when direct experience or alternative sources of information are limited. However, media's role in shaping public opinion can be both positive and negative, with examples ranging from raising awareness of social issues to potentially distorting facts.

In democratic societies, public opinion is crucial, and the media plays a vital role in informing it. However, media's impact can vary, sometimes amplifying social movements or controversies, while other times, it may unintentionally distort facts or influence judicial decisions. Overall, understanding the influence of mass media on public opinion is essential for ensuring a well-informed and democratic society.

6. REFERENCES

1. Pokharel, Bishwo Prakash. (2020). Twitter Sentiment Analysis During Covid-19 Outbreak in Nepal.
2. Twitter as a Corpus for Sentiment Analysis and Opinion Mining Alexander Pak, Patrick Paroubek.
3. Y. Chandra and A. Jana, "Sentiment Analysis using Machine Learning and Deep Learning," 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2020, pp. 1-4.
4. Nikhil Yadav, Omkar Kudale, Srishti Gupta, Aditi Rao, Ajitkumar Shitole, "Twitter Sentiment Analysis Using Machine Learning For Product Evaluation", Inventive Computation Technologies (ICICT) 2020 International Conferenceon, pp. 181-185, 2020.
5. R. A. S. C. Jayasanka, M. D. T. Madhushani, E. R. Marcus, I. A. A. U. Aberathne and S.C. Premaratne Faculty of Information Technology, University of Moratuwa, Sri Lanka.
6. <https://www.w3schools.com/html/>