A

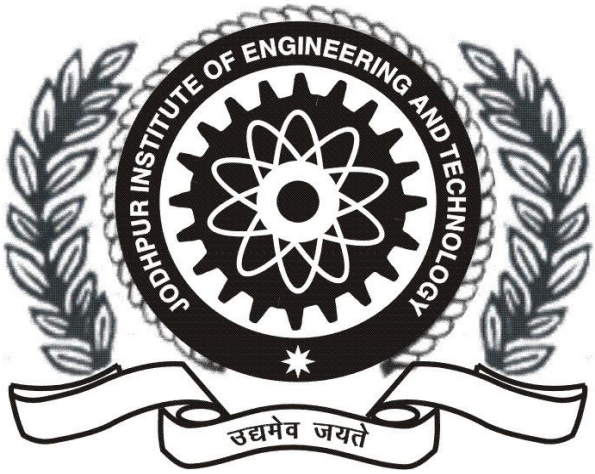
PROJECT REPORT

ON

(REAL – TIME SOCIAL MEDIA SENTIMENT ANALYSIS)

In partial fulfillment of

B.Tech. IV yr **(Computer Science & Engineering.)**



Submitted To:

Mr. Rajendra Purohit

(Associate Professor)

Submitted by:

Pranav Chouhan (20EJICS106)
Radha (20EJICS118)
Section – B

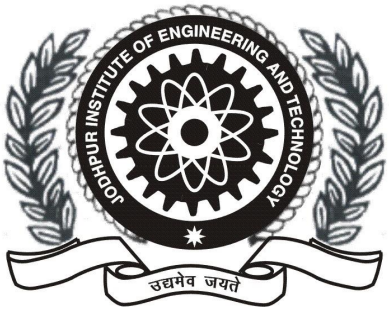
Jodhpur Institute of Engineering and Technology, Jodhpur
Department of Computer Science and Engineering Session 2023-24

PREFACE

An excellent practical work and knowledge makes an engineer successful in every field and these qualities are very important for an Engineer, in our opinion and it is also outcome of our project work also. Not only theoretical knowledge but also practical knowledge is essential for a technical student and thus the project work is beneficial for him/her to learn more and more about practical works.

This report is the culmination of extensive research, analysis, and collaboration, aimed at providing valuable insights into Social Media Sentiment Analysis.

Throughout the course of this study, we have endeavored to gather relevant information, employ rigorous methodologies, and engage in thoughtful discussions to ensure the quality and accuracy of the content presented here. This report gives good knowledge about project work. We will never say that it is complete in itself out. We have tried to do our best in preparing this report. We think this report will be taken as reference by coming students.



Jodhpur Institute of Engineering and Technology, Jodhpur
Department of Computer Science and Engineering
Session 2023-24

CERTIFACTE

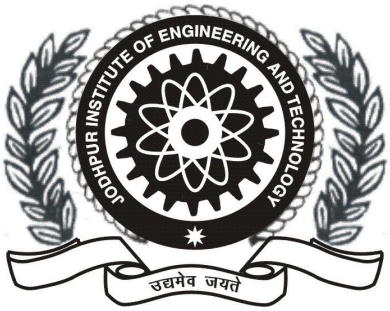
This is to certify that **Mr. Pranav Chouhan S/o Prakash Chouhan** pursuing, B.tech(Bachelor of Technology), 4th year, from Jodhpur Institute of Engineering and Technology, Jodhpur has successfully completed the project entitled **"Real- Time Social Media Sentiment Analysis"** has been carried out under the guidance and supervision.

The conducted performance of the student remained excellent during the period of Project Work.

Date: 16 December, 2023

(Ms. Harshita Khangarot)
Associate Professor, CSE
Project Guide

(Mr. Rajendra Purohit)
Associate Professor, CSE
Project Head



Jodhpur Institute of Engineering and Technology, Jodhpur
Department of Computer Science and Engineering
Session 2023-24

CERTIFACTE

This is to certify that **Ms. Radha D/o Shiv Nandan** pursuing, B.tech(Bachelor of Technology), 4th year, from Jodhpur Institute of Engineering and Technology, Jodhpur has successfully completed the project entitled **"Real- Time Social Media Sentiment Analysis"** has been carried out under the guidance and supervision.

The conducted performance of the student remained excellent during the period of Project Work.

Date: 16 December, 2023

(Ms. Harshita Khangarot)
Associate Professor, CSE
Project Guide

(Mr. Rajendra Purohit)
Associate Professor, CSE
Project Head

ACKNOWLEDGEMENT

The project report on Real – Time Social Media Sentiment Analysis is an outcome of guidance, moral support & devotion bestowed on us through our work. For this let me acknowledge & express our performance sense of gratitude & thanks to everybody who have been source of inspiration during the project preparation.

First & foremost, I offer my sincere phrases of thanks with innate humility to **"Ms. Harshita Khangarot Ma'am"** to guide our project for providing help whenever needed. So, we are very thankful to acknowledge for key role played by her in providing us with her precious ideas, suggestion & help that enabled in shaping project work.

I would like to recognize the invaluable assistance of **"Mrs. Mamta Garg Ma'am"** and **"Anju Jangid Ma'am"** for advising us and introducing the project phase to us in an easy way to understand, which has helped me to complete my report easily and effectively on time. The activity was also helpful for our future work and I would also like to express my sincere gratitude towards JIET college who gave us this golden opportunity.

TABLE OF CONTENTS

1. INTRODUCTION

Businesses must be quick to respond potential crises or market trends in today's fast – changing landscape. Marketers rely on Sentiment analysis software to learn what customers feel about the company's brand, products, and services in real time and take immediate actions. Based on their findings. They can configure the software to send alerts when negative sentiments are detected for specific keywords.

Objective:

The goal of this project is to develop a real-time Social Media Sentiment Analysis Tool that enables companies and organizations to monitor and analyze sentiment in social media posts. This tool will help them gain insights into how their brand or products are perceived and allow for timely responses to maintain a positive reputation. The problem addressed by this project is the lack of an efficient and real-time tool for monitoring and analyzing sentiment on social media platforms. Existing tools may not provide real-time insights, lack customization options, or have limited accuracy in sentiment analysis.

Scope of the Project:

Customer support teams use sentiment analysis tools to personalize responses based on the mood of the conversation. Business use sentiment analysis to derive intelligence and form actionable plans in different areas.

- Improve customer service
- Brand monitoring
- Market research
- Track campaign performance

Features:

Real-time Monitoring: Companies and organizations need to protect their brand reputation by addressing negative sentiment promptly and amplifying positive sentiment. Continuously monitor social media feeds across major platforms such as Twitter, Facebook, and Instagram.

Sentiment Analysis: Social media provides a wealth of customer feedback that can be valuable for product improvement and customer satisfaction. Employ advanced Natural Language Processing (NLP) techniques to analyze and classify social media posts into three categories: positive, negative, or neutral. Track sentiment trends over time to provide insights into how sentiment is evolving.

Customization: Understanding the sentiment of potential customers and competitors can inform strategic decisions and marketing campaigns. Allow users to customize sentiment categories and keywords for monitoring, ensuring relevance to specific contexts.

Visualization: Present sentiment trends and insights through interactive dashboards and visualizations for easy interpretation. Rapid identification of negative sentiment and potential crises allows for immediate response and damage control.

Outcome:

Social media sentiment analysis unlocks the essence of what people express about your brand or business across social platform. By analayzing social sentiment, we gain valuable understanding to strengthen our social media monitoring plan, enabling us to grasp how people perceive our products, services and brand.

2. REVIEW/LITERATURE SURVEY

[1] This particular research paper will build on Chakraborty et al.'s research paper: "Predicting stock movement using sentiment analysis of the Twitter feed". In their article, the researchers have found that Twitter data could predict stock prices very well on stable days in the stock market. However, the researchers used a boosted regression tree model to predict the stock price difference for the next day with the current day's stock market Sentiment. This paper will implement neural networks to see if they produce better results than the boosted tree model. Specifically, a Multilayer Perceptron Neural Network (MLP) model will be employed. This paper aims to improve the previous writing using MLP and analyze the effectiveness of using Twitter data to predict stock market trends and prices. In our work, we predict the future movement of the United States' stock market by analyzing the sentiment of Twitter posts related to the Stock market. To do this, we collected stock-related tweets and obtained their average sentiment value by using SVM. After that, we prepared the training set with those tweets and with corresponding DJIA or Apple Inc. closing stock index differences between the present-day and next day. Then we tested on similar stock related tweets on a different timeline to see how much we can predict the stock index. We used a Boosted Regression Tree model and a Multilayer Perceptron Neural Network model to do this. We were able to derive answers for both of our hypotheses. From the results of our work, it is seen that tweets do play a role in the prediction of stock market movement. Furthermore, it is implied that Neural Networks perform better than the Boosted Regression Tree. For all three sets of data with the keywords: "stock market", "stocktwits", "AAPL", the Multilayer Perceptron Neural Network model has a lower MAE and RMSE than the Boosted Regression Tree model. From our results, it is also clear that too high and too low differences in Stock Indexes are challenging to predict with Boosted Regression Tree. However, except for those days, our models predicted very well on the given data set. Furthermore, using a data range of more than one year may provide more accurate results. Additionally, analyzing the models in different economic situations such as booms

or recession may allow us to better see the productivity of the models. Besides, the use of a neural network for classifying the sentimental analysis tweets may offer better results.

[2] In this paper, real time monitoring of public sentiment, trend tracking, and understanding public opinion, with implications for diverse fields such as marketing, politics, and customer service. The study underscores how sentiment analysis on Twitter can be a powerful tool for gaining insights into public sentiment and harnessing this data for decision-making. In summary,this research paper goes beyond merely acknowledging the challenges of sentiment analysis on Twitter. It provides a deep and rich exploration of the intricacies involved, offering a roadmap for researchers and practitioners to navigate this challenging terrain successfully. By addressing the unique characteristics of Twitter data, adopting specialized preprocessing, and considering the role of features, the paper contributes significantly to the advancement of sentiment analysis techniques, particularly within the context of real-time social media platforms like Twitter.

[3] According to the paper Sentiment analysis has proven to be a valuable tool to gauge public opinion in different disciplines. It has been successfully employed in financial market prediction, health issues, customer analytics, commercial valuation assessment, brand marketing, politics, crime prediction, and emergency management. Many of the published studies have focused on sentiment analysis of Twitter messages, mainly because a large and diverse population expresses opinions about almost any topic daily on this platform. This paper proposes a comprehensive review of the multifaceted reality of sentiment analysis in social networks. We not only review the existing methods for sentiment analysis in social networks from an academic perspective, but also explore new aspects such as temporal dynamics, causal relationships, and applications in industry. We also study domains where these techniques have been applied, and discuss the practical applicability of emerging Artificial Intelligence methods. This paper emphasizes the importance of temporal characterization and causal effects in sentiment analysis in social networks, and explores their applications in different contexts such as stock market value, politics, and cyberbullying in educational

centers. A strong interest from industry in this discipline can be inferred by the intense activity we observe in the field of intellectual protection, with more than 8,000 patents issued on the topic in only five years. This interest compares positively with the effort from academia, with more than 2,300 articles published in 15 years. But these papers are unevenly split across domains: there is a strong presence in marketing, politics, economics, and health, but less activity in other domains such as emergencies. Regarding the techniques employed, traditional techniques such as dictionaries, neural networks, or Support Vector Machines are widely represented. In contrast, we could still not find a comparable representation of advanced state-of-the-art techniques such as Transformers-based systems like BERT, T5, T0++, or GPT-
2/3. This reality is consistent with the results found by the authors of this work, where computationally expensive tools such as GPT-3 are challenging to apply to achieve competitive results compared to those from simpler, lighter and more conventional techniques. These results, together with the interest shown by industry and academia, suggest that there is still ample room for research opportunities on domains, techniques and practical applications, and we expect to keep observing a sustained cadence in the number of published papers, patents and commercial tools made available.

[4] Microblog data like Twitter, on which users post real time reactions to and opinions about “everything”, poses newer and different challenges. Some of the early and recent results on sentiment analysis of Twitter data are by Go et al. (2009), (Bermingham and Smeaton, 2010) and Pak and Paroubek (2010). Go et al. (2009) use distant learning to acquire sentiment data. They use tweets ending in positive emoticons like “:)” “:-)” as positive and negative emoticons like “:(” “:-(” as negative. They build models using Naive Bayes, MaxEnt and Support Vector Machines (SVM), and they report SVM outperforms other classifiers. In terms of feature space, they try a Unigram, Bigram model in conjunction with parts-ofspeech (POS) features. They note that the unigram model outperforms all other models. Specifically, bigrams and POS features do not help. Pak and Paroubek (2010) collect data following a similar distant learning paradigm. They perform a different classification task though: subjective versus objective. For subjective data

they collect the tweets ending with emoticons in the same manner as Go et al. (2009). For objective data they crawl twitter accounts of popular newspapers like “New York Times”, “Washington Posts” etc. In addition we explore a different method of data representation and report significant improvement over the unigram models. Another contribution of this paper is that we report results on manually annotated data that does not suffer from any known biases. Our data is a random sample of streaming tweets unlike data collected by using specific queries. The size of our hand-labeled data allows us to perform crossvalidation experiments and check for the variance in performance of the classifier across folds.

[5] The primary objective of this paper is to investigate the role of Twitter in political campaigning and how candidates use the platform to communicate with voters and constituents. The study conducted an exploratory analysis of Twitter usage by political candidates during an election campaign. It involved collecting and analyzing tweets from a sample of political candidates. The paper focused on several aspects, including tweet frequency, content, interaction with followers, and sentiment expressed in the tweets. The underscores that Twitter is not only a medium for broadcasting campaign-related information but also a space where candidates express personal opinions, engage in issue discussions, and share their stance on various topics. These findings emphasize the adaptability of Twitter as a versatile tool for politicians to address a wide range of objectives. Ultimately, this study has far-reaching implications for the realm of political communication, offering a deeper understanding of how social media, in this case, Twitter, can influence electoral outcomes and shape the relationships between candidates and their constituents. It underscores the need for politicians to navigate the digital landscape strategically, balancing campaign messages, engagement with followers, and the management of sentiment to effectively connect with the modern, tech-savvy electorate. As the digital sphere continues to evolve, future research may further elucidate the intricate dynamics of social media in political campaigns and its implications for democratic processes.

[6] Agarwal et al. recognize the practical applications of sentiment analysis on Twitter data. These applications extend to real-time monitoring of public sentiment, trend tracking, and understanding public opinion, with implications for diverse fields such as marketing, politics, and customer service. The study underscores how sentiment analysis on Twitter can be a powerful tool for gaining insights into public sentiment and harnessing this data for decision-making. In summary, this research paper goes beyond merely acknowledging the challenges of sentiment analysis on Twitter. It provides a deep and rich exploration of the intricacies involved, offering a roadmap for researchers and practitioners to navigate this challenging terrain successfully. By addressing the unique characteristics of Twitter data, adopting specialized preprocessing, and considering the role of features, the paper contributes significantly to the advancement of sentiment analysis techniques, particularly within the context of real-time social media platforms like Twitter. This re search underscores the complexity of sentiment analysis in the Twitter verse and offers valuable guidance for researchers and practitioners. It emphasizes the importance of tailored approaches, specialized preprocessing, and the consideration of unique Twitter features. By addressing these challenges, the paper contributes to the advancement of sentiment analysis techniques, particularly in the context of realtime social media platforms like Twitter.

Sentiment analysis has garnered significant attention in recent years. Researchers have been exploring methods to extract and analyze sentiments expressed in Instagram content, which mainly consists of images, captions, and comments. The primary objective of sentiment analysis for Instagram is to understand the emotional tone and opinions expressed in visual and textual content shared on the platform. Researchers aim to uncover insights related to user sentiment, preferences, and trends. Sentiment analysis for Instagram is a dynamic field, driven by the unique combination of visual and textual content. The research in this area focuses on developing innovative methods to extract and interpret sentiment in diverse contexts, from personal posts to brand-related content. With the continued growth of Instagram, sentiment analysis will likely remain a significant area of study in the years to come.

3. REQUIREMENT SPECIFICATION

Requirement Specifications for the Career Catalyst – Real-time Social Media Sentiment Analysis:

4.1 Software Development Life Cycle (SDLC) Model:

We have employed the Agile Software Development Life Cycle (SDLC) model for this project. Agile is an iterative and incremental model that is well-suited for dynamic projects such as the development of the Sentiment Analysis Tool. This model emphasizes flexibility, collaboration, and customer satisfaction, allowing for the adaptation of requirements throughout the development process.

The key features of the Agile model for the Resume Analysis Tool are:

- **Requirements Gathering and Analysis:** Continuous collaboration with stakeholders to gather and refine project requirements, ensuring a clear understanding of user needs for resume parsing, analytics, and reporting. The primary stakeholders include end-users, social media administrators, and developers.
- **Design:** Iterative design processes, adapting the system architecture and interface based on ongoing feedback and evolving requirements.
- **Implementation:** Regularly delivering incremental software updates, integrating NLP models, database components, and user interface elements in small, manageable iterations.
- **Testing:** Conducting continuous testing throughout the development process, with a focus on rapid identification and resolution of issues. Emphasis on automated testing for efficiency.
- **Deployment:** Regularly deploying functional increments of the system, ensuring that users can access and benefit from the latest features.
- **Maintenance:** Ongoing maintenance and updates after each iteration, with a commitment to responsiveness and adaptability in addressing user feedback and emerging requirements. Purpose: The purpose of this document is to outline the requirements for the development of a real-time social media sentiment analysis system.

- **Scope:** The system will focus on monitoring and analyzing sentiments expressed on various social media platforms in real-time.

4.2 Interfaces:

Software Interfaces:

- Operating System: Cross-platform compatibility(Windows, Linux, macOS).
- Web Browser: Compatible with major browsers (Chrome, Firefox, Safari).

Hardware Interfaces:

- Server Hardware: Standard server with sufficient processing power and storage.
- User Hardware: Desktops, laptops, or tablets with internet connectivity.

Communication Interfaces:

- HTTP/HTTPS for communication between the user's device and the server.

4.3 Hardware Requirements:

Server:

- Processor: Dual-core or higher.
- RAM: 8 GB or higher.
- Storage: 100 GB SSD or higher.

Internet Connectivity:

- High-speed internet for cloud services.

User Devices:

- Desktop, laptop, or tablet with modern browser support.
- Internet connectivity for accessing the web-based interface.

4.4 Software Requirement:

Server-side:

- Python: Backend scripting.
- spaCy, NLTK: Natural Language Processing.
- MySQL: Database management.
- Streamlit: Web interface development. Client-side:
- Web Browser: Chrome, Firefox, Safari.

- NLP Models: spaCy's en_core_web_sm model and pyresparser for resume parsing.

Client-side:

- Web Browser: Chrome, Firefox, Safari.
- NLP Models: spaCy's en_core_web_sm model and pyresparser for resume parsing.

4.5 Functional Requirements:

Data Collection:

- The system shall be capable of collecting real-time data from various social media platforms (e.g., Twitter, Facebook, Instagram).
- The system should support multiple languages for sentiment analysis. **Sentiment Analysis:**
- The system must employ natural language processing (NLP) techniques to analyze and categorize social media content into positive, negative, or neutral sentiments.
- Sentiment analysis should consider context, sarcasm, and colloquial language.
- The accuracy of sentiment analysis should be adjustable and configurable.

Real-time Monitoring:

- The system must provide real-time monitoring of social media feeds.
- Monitoring should include user comments, posts, and mentions.
- The system should be capable of handling a high volume of data in realtime.

User Interface:

- The system must have a user-friendly interface for end-users and administrators.
- The interface should display sentiment trends, charts, and detailed sentiment analysis reports.
- Users should be able to filter and customize the displayed data based on time, platform, and sentiment.

4.6 Non-Functional Requirements: Performance:

- The system should handle a minimum of 10,000 social media posts per minute.
- Response time for sentiment analysis should be less than 2 seconds.

Reliability:

- The system should have a backup and recovery mechanism for data in case of a failure.
- The system should be available 99.9% of the time.

Scalability:

- The system should be scalable to accommodate an increasing volume of social media data.
- It should support the addition of new social media platforms without significant modifications.

Security:

- Data encryption should be implemented for data in transit and at rest.
- Access control mechanisms should restrict unauthorized access to sensitive data.

4. TECHNICAL DETAILS OF PROJECT/STUDY

5.1 Objective

The goal of this project is to develop a real-time Social Media Sentiment Analysis Tool that enables companies and organizations to monitor and analyze sentiment in social media posts. This tool will help them gain insights into how their brand or products are perceived and allow for timely responses to maintain a positive reputation. Real-time social media sentiment analysis on YouTube involves continuously monitoring and analyzing the sentiments expressed in comments, discussions, and interactions on the YouTube platform. This process utilizes natural language processing (NLP) and machine learning techniques to determine whether the sentiment behind user-generated content is positive, negative, or neutral. Real-time social media sentiment analysis on YouTube serves as a valuable tool for content creators, marketers, and platform administrators to understand audience reactions, identify trends, and respond promptly to emerging sentiments.

5. DESIGN DOCUMENT

A design document, often referred to as a design specification or technical design document, is a comprehensive and detailed written description of the design and architecture of a software system or a project. It serves as a blueprint for developers, engineers, and other stakeholders, providing a structured outline of how the software or project is intended to be built, including its components, interactions, and functionalities. A well-structured design document is crucial for ensuring a clear and consistent implementation of the project and is often created during the planning and early development stages.

6.1 Functional Description:

The project is used to develop a real-time Social Media Sentiment Analysis Tool that enables companies and organizations to monitor and analyze sentiment in social media posts. This tool will help them gain insights into how their brand or products are perceived and allow for timely responses to maintain a positive reputation.

The primary objective of this project is to gain inside into public opinions, emotions and attitude towards specific topics, brands, events or products as they unfold on social media. Firstly, data is collected through Instagram platform and majorly based on specific keywords, hashtags, or user defined criteria to focus on relevant content. In preprocessing it involves tasks like text normalization, removing noise (e.g., special characters, URLs, and emojis), and language detection.

The core of the project involves sentiment analysis, which classifies each piece of social media content into different sentiment categories such as positive, negative, or neutral. It can be performed by using techniques like machine learning algorithm or deep learning models. The sentiment analysis is done in real time allowing the system to process and categorize new social media content as it is generated. A real-time social-media sentiment analysis project provides valuable insights for business, organisations and individuals to understand public perception, track brand reputation, and respond to emerging trends or issues promptly. It has

applications in marketing, customer service, market research, and public opinion monitoring.

6.2 Functional Partitions:

To efficiently manage the development of a real-time social media sentiment analysis project, you can break it down into functional partitions or components. These partitions help in organizing the project, assigning responsibilities, and ensuring that each component performs its specific role.

Data Collection:

- Implement data collection from various social media platforms and APIs. Handle real-time data streaming and data retrieval.
- Normalize text, remove noise, and perform language detection. Tokenize text for further analysis.

Sentiment Analysis:

- Develop sentiment analysis models or integrate pre-built sentiment analysis libraries. Classify social media content into sentiment categories (positive, negative, neutral).
- Continuously process incoming social media data. Update sentiment statistics and trends in real-time. Provide language-specific text processing and sentiment analysis when necessary.

Visualization and Reporting:

- Create interactive visualizations, charts, and graphs to represent sentiment data. Send notifications or alerts to users or administrators.

Data Storage and Archiving:

- Store historical social media data for future analysis. Implement data archiving and retrieval mechanisms ensure data security and compliance.

6.3 Data Description:

This data description provides an overview of the types of data that are collected, processed, and analyzed in a real-time social media sentiment analysis project. It

highlights the importance of data quality, real-time processing, and compliance with privacy regulations when handling social media data. For Instagram it used data like public photos and captions, comments, and likes.

6.4 User Interface Design:

Designing a user interface for a real-time social media sentiment analysis project is essential to provide users with an intuitive and interactive experience for monitoring sentiment trends and insights.

Main Dashboards:

- The main dashboard serves as the central hub where users can access all key features and insights. At the top, display the project name and logo for branding.
- Include a navigation menu on the left for easy access to various sections of the application. Ensure that the user interface is responsive and compatible with various devices, including desktops, tablets, and mobile phones.

Real-Time Sentiment Trends:

- A prominent section of the dashboard should display real-time sentiment trends with charts and graphs.
- Use line charts, bar graphs, or heatmaps to show how sentiments are changing over time. Allow users to filter trends by social media platform, sentiment type, and date range.

Custom Search and Monitoring:

- Include a search bar that enables users to enter keywords or hashtags to monitor in real-time. Show live updates for the monitored streams with sentiment breakdowns.
- Display a section for alerts and notifications, which can be customized by the user. Alerts can be triggered based on significant sentiment changes, keyword mentions, or specific events.

Data Analysis:

- Provide an option to access historical sentiment data and analytics.
- Users can select a date range and view sentiment trends, popular keywords, and sentiment breakdowns for specific periods. Include interactive charts and data visualization for historical analysis.

6.5 Module Description:

A real-time social media sentiment analysis project can be organized into several modules, each responsible for specific functionalities. These modules help break down the project into manageable components. And to create a comprehensive real-time social media sentiment analysis system that provides valuable insights to users, monitors public sentiment, and helps organizations make data-driven decisions based on the sentiment trends observed in social media data.

Data Collection Module:

- Responsible for collecting real-time social media data from Youtube. Cleans and prepares raw social media data for analysis.
- Normalizes text, removes noise (e.g., URLs, special characters), and identifies the language of the content. Tokenizes text for further analysis.

Sentiment Analysis Module:

- Implements the core sentiment analysis functionality. Utilizes sentiment analysis models, which could be machine learning, or deep learning models.
- Classifies social media content into sentiment categories such as positive, negative, or neutral.

Real-Time Processing Module:

- Continuously processes incoming social media data and triggers sentiment analysis as new data arrives.
- Updates sentiment statistics and trends in real-time. Manages the realtime data stream and ensures data consistency.

Visualization and Reporting Module:

- Creates interactive visualizations and reports for users to understand sentiment insights.

- Allows users to explore sentiment data easily. Defines and manages criteria for triggering alerts based on sentiment changes, trending topics, or other user-defined conditions.

Data Storage Module

- Archives historical social media data for future analysis and research. Manages the storage of data in databases or data lakes.
- Ensures data security, privacy, and compliance with regulations.

6.6 UML Diagrams –

UML, which stands for Unified Modeling Language, is a way to visually represent the architecture, design, and implementation of complex software systems. When you're writing code, there are thousands of lines in an application, and it's difficult to keep track of the relationships and hierarchies within a software system. UML diagrams divide that software system into components and subcomponents.

UML is a standardized modeling language that can be used across different programming languages and development processes, so the majority of software developers will understand it and be able to apply it to their work.

Though many engineers dread diagrams, they're useful in an Agile development environment: they keep development productive and focused. Instead of thinking them as just a "nice to have," treat your UML diagrams as core aspects of documentation. UML diagrams can help engineering teams:

- Bring new team members or developers switching teams up to speed quickly.

- Navigate source code.
- Plan out new features before any programming takes place.
- Communicate with technical and non-technical audiences more easily.

Sequence Diagram –

A sequence diagram is a type of interaction diagram because it describes how—and in what order—a group of objects works together. These diagrams are used by software developers and business professionals to understand requirements for a new system or to document an existing process. Sequence diagrams are sometimes known as event diagrams or event scenarios.

Benefits –

Sequence diagrams can be useful references for businesses and other organizations. Try drawing a sequence diagram to:

- Represent the details of a UML use case.
- Model the logic of a sophisticated procedure, function, or operation.
- See how objects and components interact with each other to complete a process.
- Plan and understand the detailed functionality of an existing or future scenario.

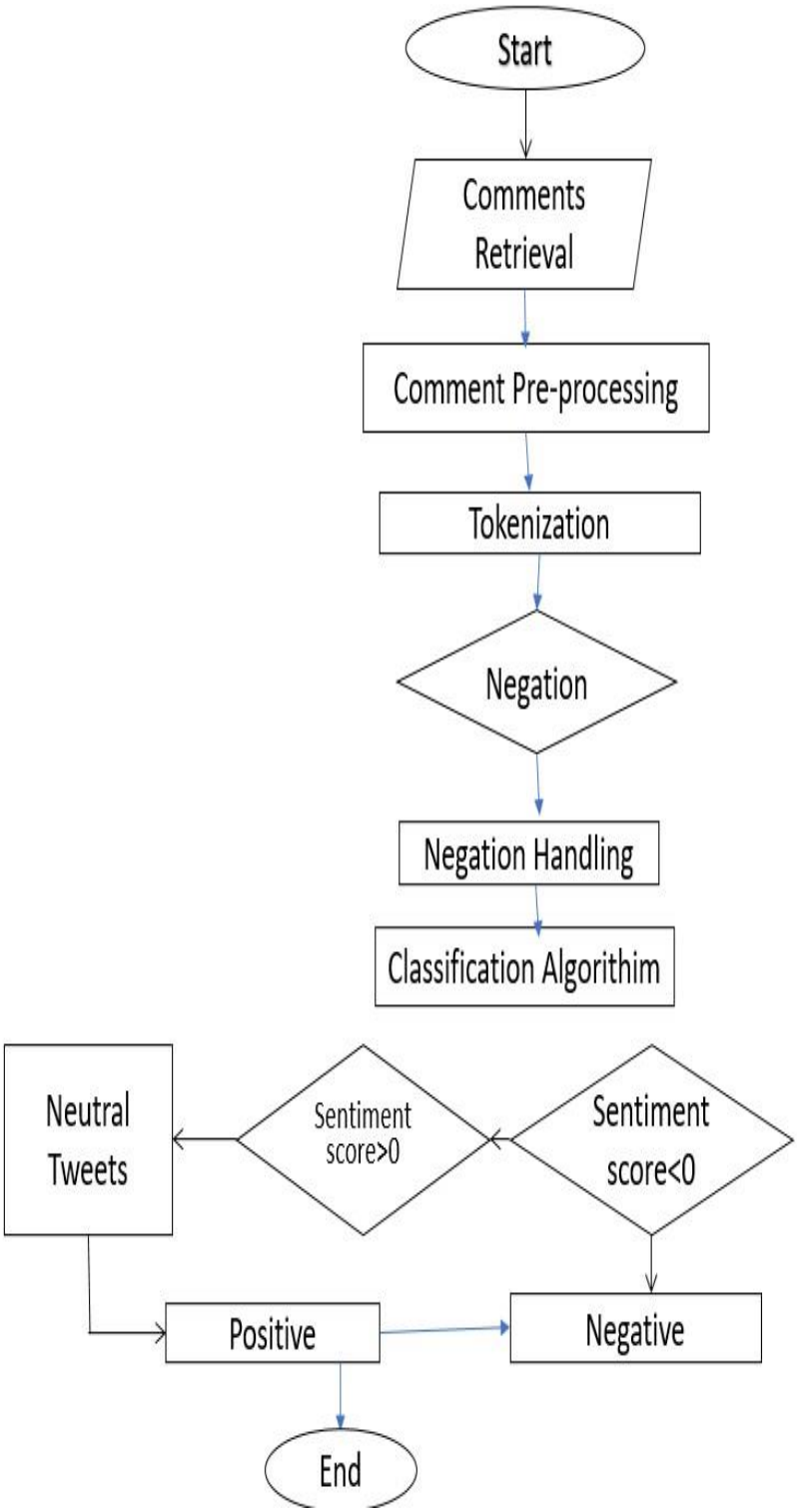

Start
Comments Retrieval
Comment Pre-processing
Tokenization
Negation
Negation Handling
Classification Algorithim
Neutral Tweets
Sentiment score>0
Sentiment score<0
Positive
Negative
End

Fig 5(a)

Data Flow Diagram -

A Data Flow Diagram (DFD) is a graphical representation of how data flows within a system. It illustrates the processes, data stores, data sources, and data destinations involved in a system and how they interact with each other. DFDs are commonly used in systems engineering and software development to model the flow of data through a process or a system.

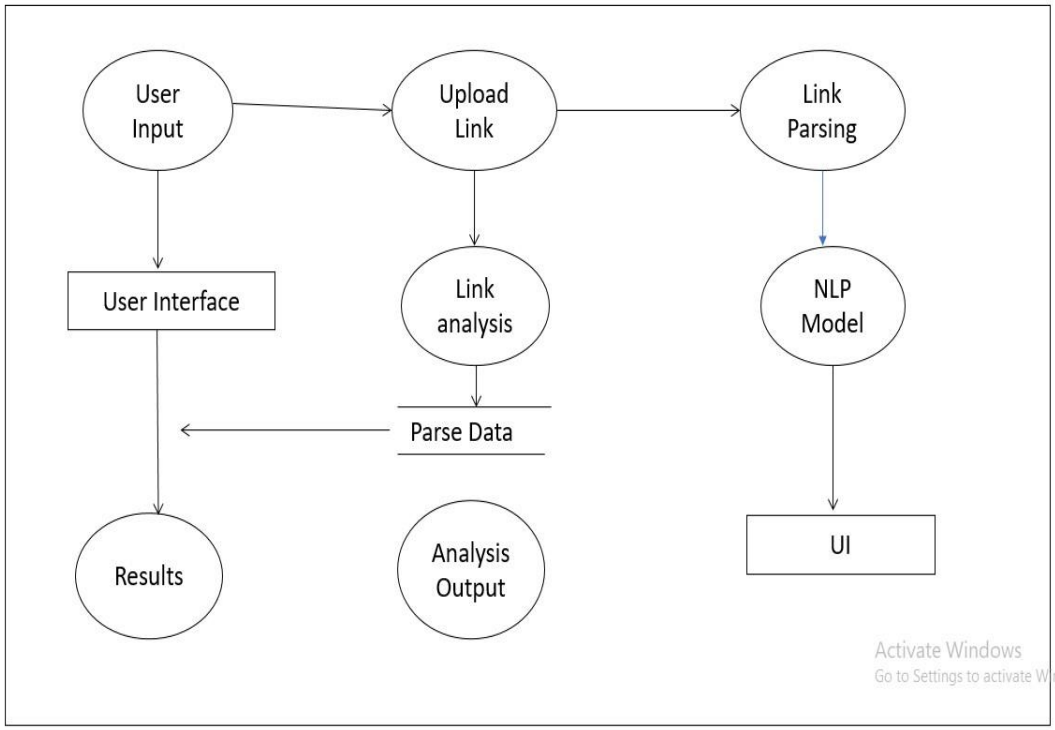


Fig 5(b)

6. RESULTS

Screenshots:

Loading Data

In [3]:
```
df = pd.read_csv('/kaggle/input/youtube-statistics/comments.csv')
df.head()
```

Out[3]:

	Unnamed: 0	Video ID	Comment	Likes	Sentiment
0	0	wAZZ-UWGVHI	Let's not forget that Apple Pay in 2014 requir...	95.0	1.0
1	1	wAZZ-UWGVHI	Here in NZ 50% of retailers don't even have co...	19.0	0.0
2	2	wAZZ-UWGVHI	I will forever acknowledge this channel with t...	161.0	2.0
3	3	wAZZ-UWGVHI	Whenever I go to a place that doesn't take App...	8.0	0.0
4	4	wAZZ-UWGVHI	Apple Pay is so convenient, secure, and easy t...	34.0	2.0

In [4]:
```
df.shape
```

Out[4]: (18409, 5)

Fig 6(a)

In [15]:
```
sns.histplot(df['Sentiment'], kde=True)
plt.show()
```

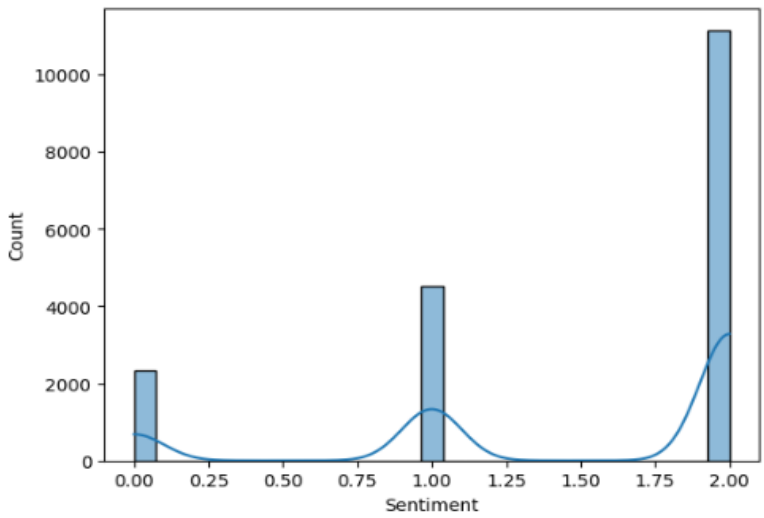


Fig 6(b)

Support Vector Classifier

In [35]:
```
model = SVC()
model.fit(X_train, y_train)
```

Out[35]: SVC()

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [36]:
```
y_pred = model.predict(X_test)
```

In [37]:
```
def eval(name, y_test, y_pred):
    cm = confusion_matrix(y_test, y_pred)
    t1 = ConfusionMatrixDisplay(cm)
    print('Classification Report for SVC: \n')
    print(classification_report(y_test, y_pred))
    t1.plot()
```

In [38]:
```
eval('Classification Report', y_test, y_pred)
```

```
Classification Report for SVC:

              precision    recall  f1-score   support

         0.0       0.74      0.15      0.25       479
         1.0       0.62      0.57      0.60       954
         2.0       0.76      0.92      0.83      2166

    accuracy                           0.72      3599
   macro avg       0.71      0.55      0.56      3599
weighted avg       0.72      0.72      0.69      3599
```

Fig 6(c)

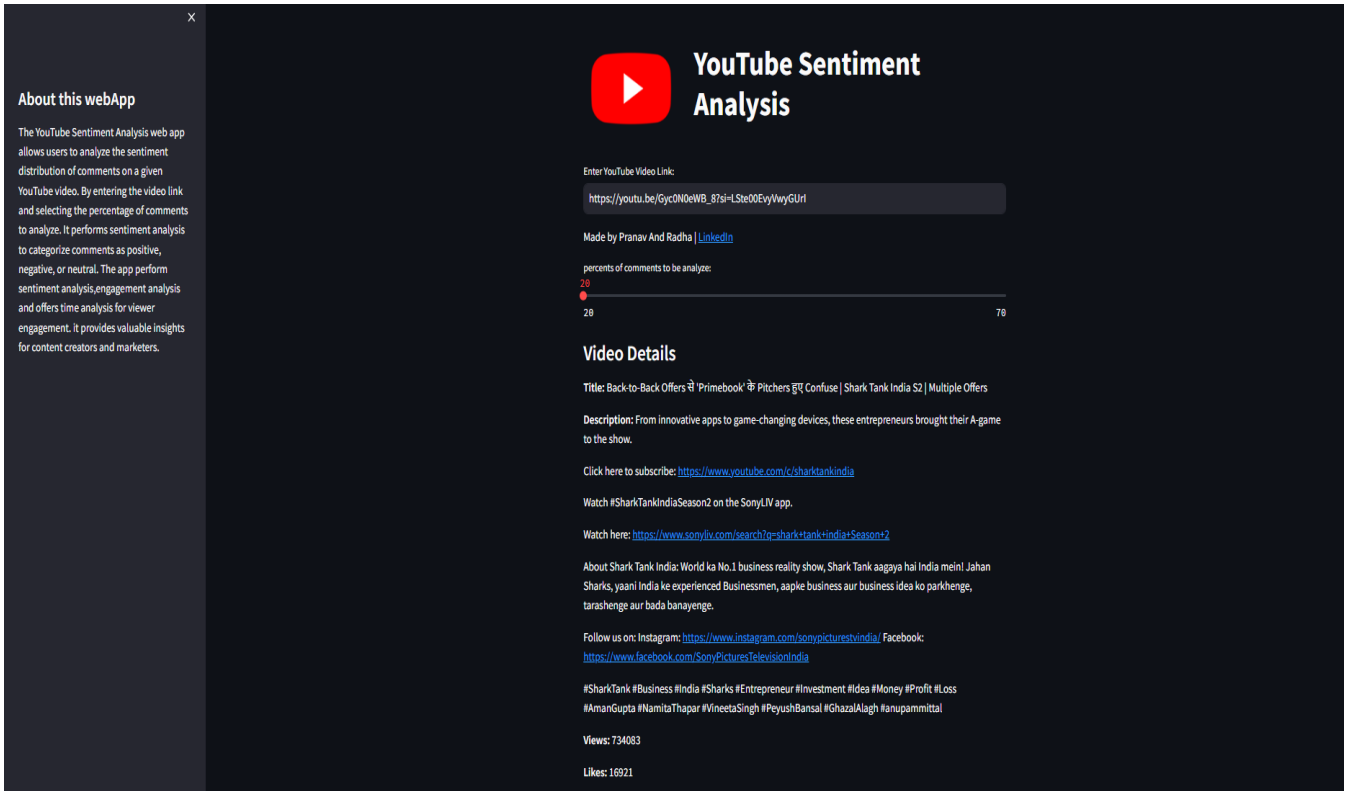


Fig 6(d)

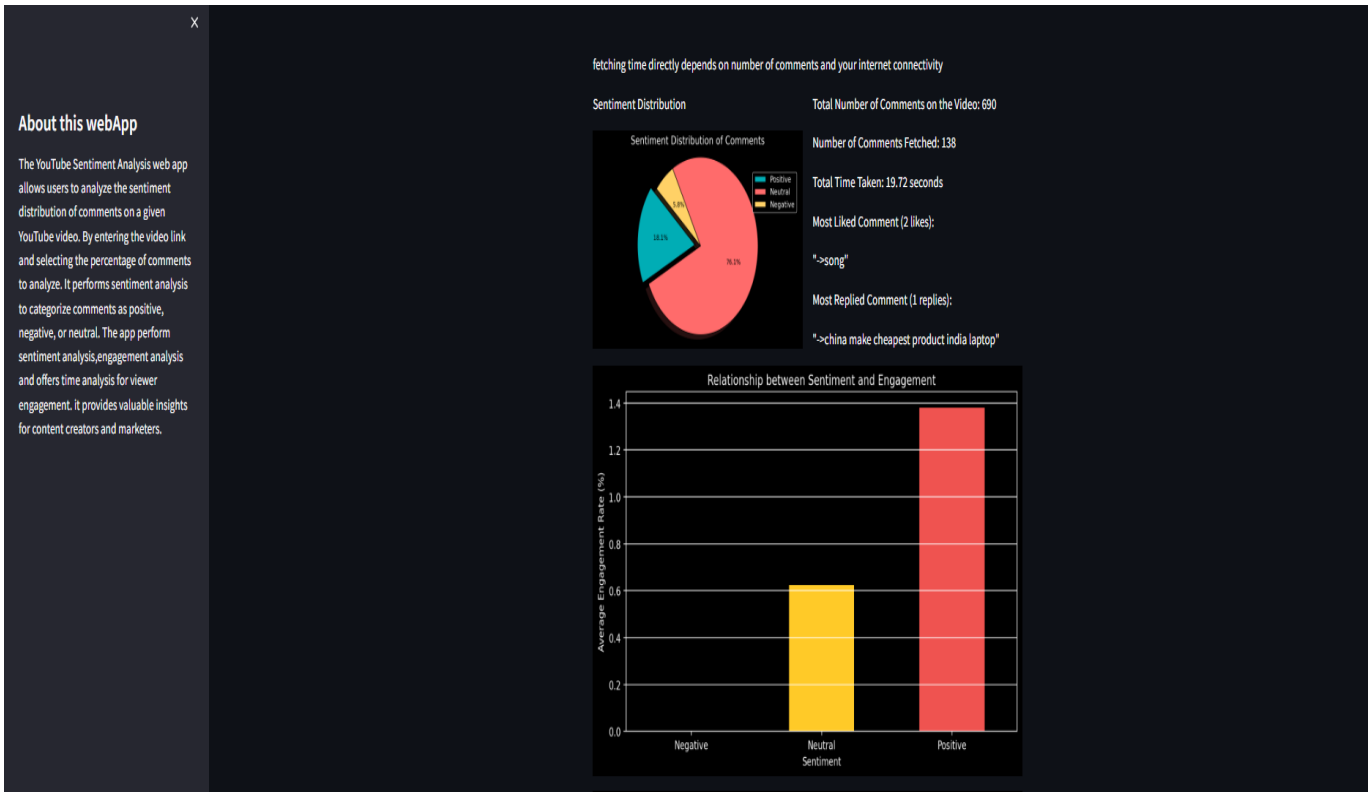

X
About this webApp
The YouTube Sentiment Analysis web app allows users to analyze the sentiment distribution of comments on a given YouTube video. By entering the video link and selecting the percentage of comments to analyze. It performs sentiment analysis to categorize comments as positive, negative, or neutral. The app perform sentiment analysis,engagement analysis and offers time analysis for viewer engagement. it provides valuable insights for content creators and marketers.
fetching time directly depends on number of comments and your internet connectivity
Sentiment Distribution
Sentiment Distribution of Comments
Positive
Neutral
Negative
Total Number of Comments on the Video: 690
Number of Comments Fetched: 138
Total Time Taken: 19.72 seconds
Most Liked Comment (2 likes):
"->song"
Most Replied Comment (1 replies):
"->china make cheapest product india laptop"
Relationship between Sentiment and Engagement
Average Engagement Rate (%)
Negative
Neutral
Positive
Sentiment

Fig 6(e)

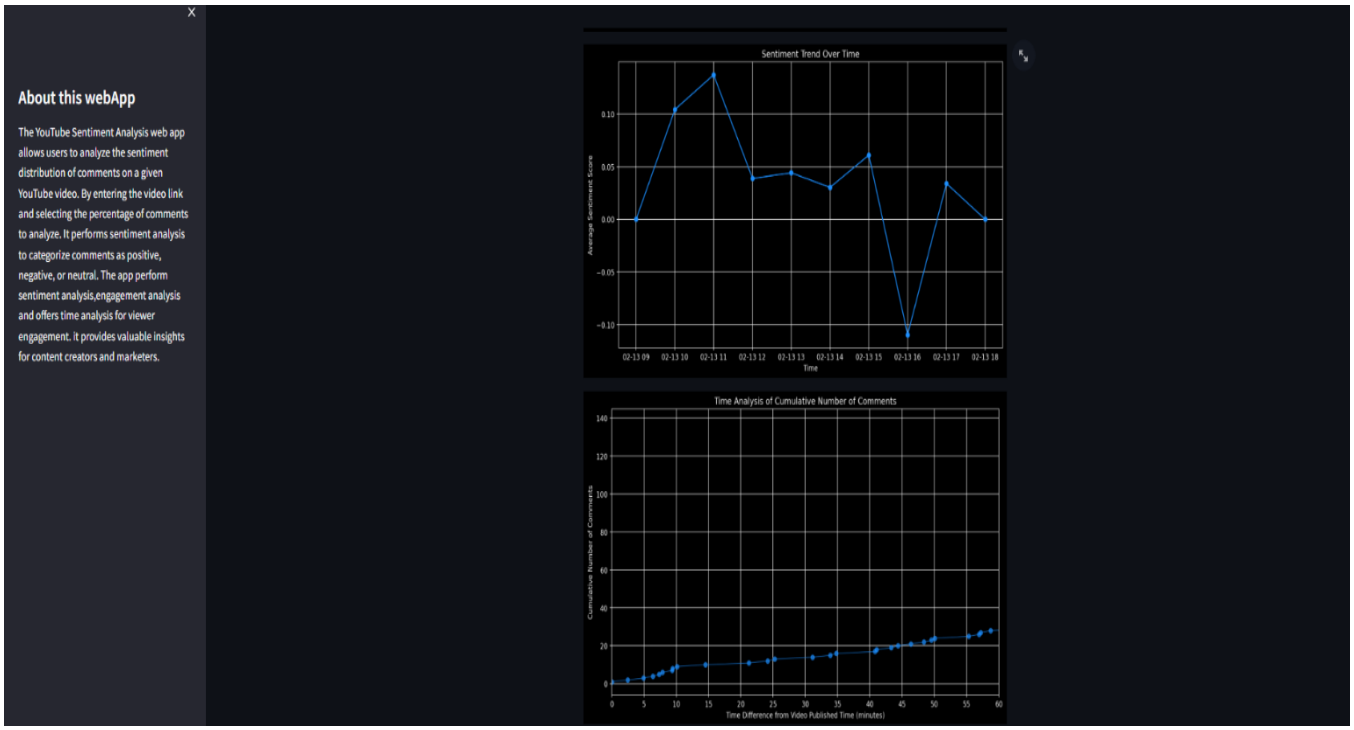

X
About this webApp
The YouTube Sentiment Analysis web app allows users to analyze the sentiment distribution of comments on a given YouTube video. By entering the video link and selecting the percentage of comments to analyze. It performs sentiment analysis to categorize comments as positive, negative, or neutral. The app perform sentiment analysis,engagement analysis and offers time analysis for viewer engagement. it provides valuable insights for content creators and marketers.
Sentiment Trend Over Time
Average Sentiment Score
Time
Time Analysis of Cumulative Number of Comments
Cumulative Number of Comments
Time Difference from Video Published Time (minutes)

Fig 6(f)

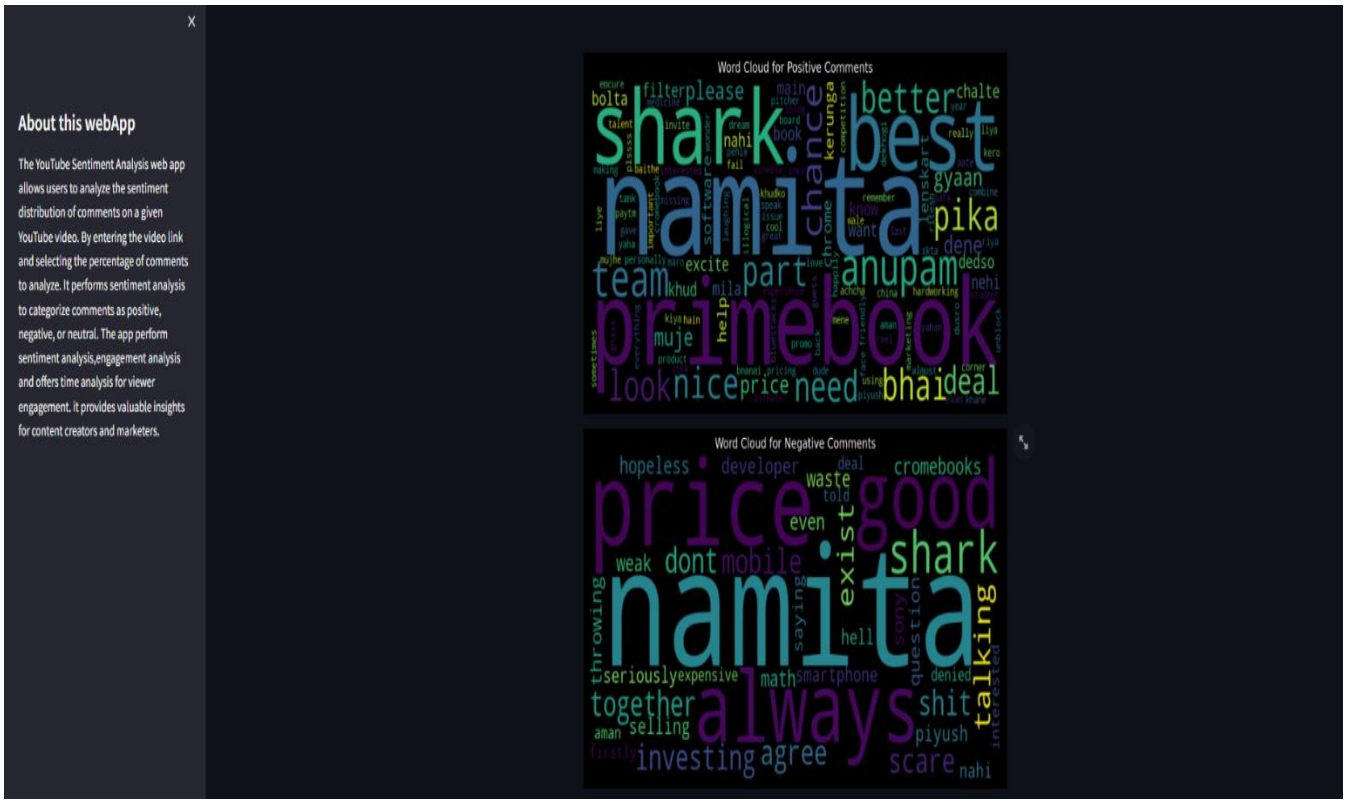

About this webApp
The YouTube Sentiment Analysis web app allows users to analyze the sentiment distribution of comments on a given YouTube video. By entering the video link and selecting the percentage of comments to analyze. It performs sentiment analysis to categorize comments as positive, negative, or neutral. The app perform sentiment analysis,engagement analysis and offers time analysis for viewer engagement. it provides valuable insights for content creators and marketers.
Word Cloud for Positive Comments
Word Cloud for Negative Comments

Fig 6(g)

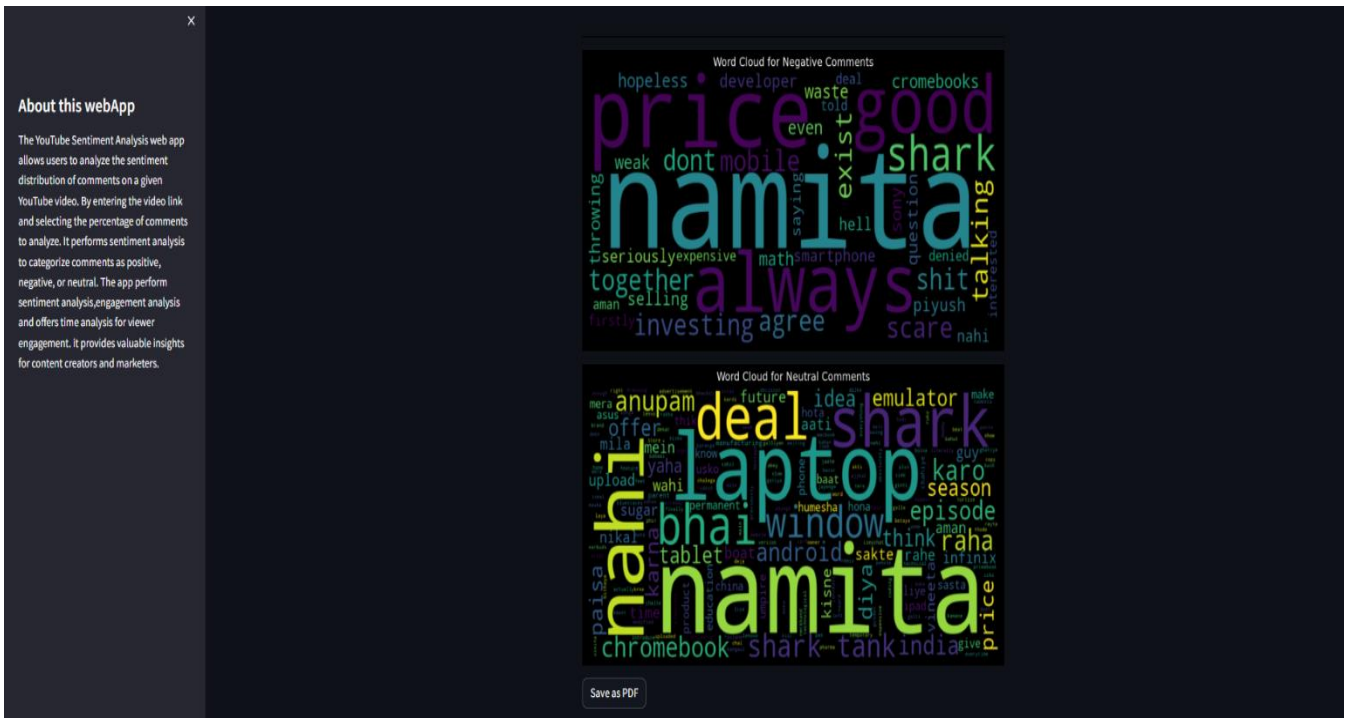

About this webApp
The YouTube Sentiment Analysis web app allows users to analyze the sentiment distribution of comments on a given YouTube video. By entering the video link and selecting the percentage of comments to analyze. It performs sentiment analysis to categorize comments as positive, negative, or neutral. The app perform sentiment analysis,engagement analysis and offers time analysis for viewer engagement. it provides valuable insights for content creators and marketers.
Word Cloud for Negative Comments
Word Cloud for Neutral Comments
Save as PDF

Fig 6(h)

7. WORK DISTRIBUTION

S. No.	Name of function/module	Start Date	End Date	Responsible Person
1.	Data Collection and Preprocessing	2023-9-15	2023-10-11	Radha
2.	NLP Model	2023-10-01	2023-10-30	Pranav
3.	Sentiment Analysis	2023-10-20	2023-11-12	Pranav Chouhan and Radha
4.	User Interface Design	2023-11-10	2023-11-21	Pranav Chouhan
5.	Maintenance and Updates	2023-11-23	2023-11-25	Pranav Chouhan
6.	Other Works	2023-11-25	2023-12-04	Pranav Chouhan and Radha
7.	Report Work	2023-12-05	2023-12-10	Pranav Chouhan and Radha

8. CONCLUSION

In conclusion, real-time social media sentiment analysis is a crucial and dynamic field with far-reaching implications for businesses, individuals, and organizations. The ability to monitor, analyze, and respond to sentiments expressed on social media platforms in real-time provides valuable insights into public opinions, brand perception, and emerging trends.

The implementation of real-time sentiment analysis involves the integration of natural language processing (NLP), machine learning, and data processing technologies. These systems must be capable of handling large volumes of social media data, providing accurate and context-aware sentiment categorization, and delivering timely insights to end-users.

The user interface of real-time sentiment analysis tools plays a pivotal role, offering an intuitive and customizable platform for users to visualize sentiment trends, access detailed reports, and make informed decisions. The responsiveness and scalability of these systems are crucial, ensuring that they can adapt to the ever-changing landscape of social media interactions and accommodate increasing data volumes.

Security and ethical considerations are paramount in the development and deployment of real-time sentiment analysis systems. Adherence to privacy regulations, data encryption, and measures to mitigate biases are essential to ensure responsible and lawful use of user-generated content.

As technology continues to evolve, real-time social media sentiment analysis remains at the forefront of enhancing user engagement, brand management, and crisis response strategies. The insights derived from these systems empower businesses to make data-driven decisions, respond promptly to customer.

9. FUTURE WORK

The future of real-time social media sentiment analysis holds exciting prospects, and ongoing research and development efforts aim to address current challenges and explore new opportunities. Continued collaboration between researchers, data scientists, and industry practitioners will play a crucial role in shaping the future of real-time social media sentiment analysis, ensuring that these systems remain effective, ethical, and aligned with the evolving landscape of online communication.

Here are some potential avenues for future work in this field:

- **Enhanced Contextual Understanding:** Future work could focus on improving the contextual understanding of sentiment analysis systems. This includes better handling of sarcasm, nuanced language, and cultural variations to enhance the accuracy of sentiment predictions.
- **Multimodal Analysis:** Integrating multimodal analysis that considers not only text but also images, videos, and audio content could provide a more comprehensive understanding of sentiments expressed on social media. This would require the development of sophisticated models capable of analyzing diverse types of media.
- **Real – time Emotion Recognition:** Going beyond simple positive, negative, or neutral categorizations, future sentiment analysis systems might delve deeper into recognizing specific emotions expressed in social media content. This could involve the development of emotion-specific models and algorithms.
- **Improved Handling of Biases:** Addressing and mitigating biases in sentiment analysis models is an ongoing challenge. Future work could focus on developing methods to identify and reduce biases, ensuring fair and unbiased sentiment analysis across diverse user demographics.
- **Personalized Sentiment Analysis:** Customizing sentiment analysis for individual users or specific user groups could provide more personalized insights. This could involve tailoring models based on user preferences, historical interactions, and feedback.
- **Explainable AI in Sentiment Analysis:** Enhancing the interpretability and explainability of sentiment analysis models is crucial. Future research might focus on developing models that provide clear explanations for their predictions, increasing user trust and understanding.

- **Privacy – Preserving Techniques:** Given the increasing concerns about privacy, research could explore techniques that enable sentiment analysis while preserving user privacy. This may involve decentralized or federated learning approaches.

10. REFERENCES

Literature Survey References -

[1] Sai Vikram Kolasani (2010), Rida Assaf Trumbull High School, Trumbull, USA, University of Chicago/ Computer Science, Chicago, USA.

[2] Agrawal et al. recognize the practical applications of sentiment analysis on Twitter data.

[3] Abbasi, A., Hassan, A., & Milan, D. (2014). Benchmarking twitter sentiment analysis tools. In Proceedings of the Ninth International Conference on Language Resources and Evaluation in European.

[4] Federico Neri, Carlo Aliprandi and Federico Capeci in August 2012, Conference: International Conference on Advances in Social Networks Analysis and Mining.

[5] Li, X., & Lam, W. (2014), How candidates use Twitter during an election campaign: an exploratory study.

[6] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment Analysis of Twitter Data.

Website references –

- https://www.repustate.com/blog/real-time-sentiment-analysis/
- https://www.supportlogic.com/resources/blog/the-escalation-matrixbest-practices-and-going-beyond/?utm_campaign=23.09-WW-PSCHPROS---Google-Net-New---TOF-Campaigns&utm_medium=PaidAdvertising&utm_source=Google&utm_term=sentiment%20analysis &utm_content=NN-TOF-Escalation-Matrix&qgad=676274975433&qgterm=sentiment%20analysis&qgad=676274975433&qgterm=sentiment%20analysis&gclid=Cj0KCQiAj_CrBhD-

ARlsAliMxT_yuAZiAcBp5tshSLKqzhai313Ot7BTxJd27ZEsA_Cd7pSoOQKoFgaAjdlEALw_wcB

- https://blog.hootsuite.com/social-media-sentiment-analysis-tools/

- https://monkeylearn.com/blog/sentiment-analysis-youtubecomments/
- https://aws.amazon.com/what-is/sentimentanalysis/#:~:text=Sentiment%20analysis%20is%20the%20process,s ocial%20media%20comments%2C%20and%20reviews.