ECHO INSIGHTS - YouTube Sentiment Analysis Tool

A PROJECT REPORT

Submitted by

SANJEEV U (2116210701232)

SENTHILNAATHAN K (2116210701238)

SHARAN ADHITYA C D (2116210701241)

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

ANNA UNIVERSITY, CHENNAI

MAY 2024

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

BONAFIDE CERTIFICATE

Certified that this project titled **"ECHO INSIGHTS, YOUTUBE SENTIMENT ANALYSIS APP"** is the bonafide work of **"SANJEEV U (2116210701232)", "SENTHILNAATHAN K (2116210701238)", "SHARAN ADHITYA C D (2116210701241)"**who carried out the work under my supervision. Certified further that to the best of my knowledge, the work reported herein does not form part of any other thesis or dissertation based on which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Mr.K.Vijay B.Tech., M.E.,

PROJECT COORDINATOR

Assistant Professor (SG),

Department of Computer Science and Engineering
Rajalakshmi Engineering College, Chennai - 602 105
Submitted to Project Viva-Voce Examination held on _____

Internal Examiner **External Examiner**

ABSTRACT

In an era where social media is omnipresent, ECHO INSIGHTS emerges as a streamlined solution for analyzing YouTube video sentiment. This project focuses on developing a robust social media analytical tool designed specifically for YouTube sentiment analysis. By leveraging YouTube API V3, ECHO INSIGHTS extracts comments from any given video link and organizes them into a data frame using Pandas. To ensure accessibility for a global audience, these comments are translated into English with the Google Translate module in Python. Subsequently, sentiment analysis is performed using pre-trained Vader models, known for their effectiveness in interpreting social media text. The analysis results are showcased in an interactive dashboard created with Streamlit, enabling users to visualize and interpret sentiment patterns effortlessly. The dashboard features a range of visualizations, including the best and worst comments, providing a clear view of the most positive and negative feedback. Additionally, an overall sentiment score is calculated and displayed, summarizing the general tone of the video's reception. One of the standout features of ECHO INSIGHTS is the word cloud, which highlights the most frequently used words in the comments, offering a quick overview of prevalent topics and trends within the audience's feedback. ECHO INSIGHTS aims to empower content creators, marketers, and analysts by providing a comprehensive understanding of audience sentiment and engagement on YouTube videos. By offering detailed insights into viewer opinions and reactions, users can make more informed decisions regarding content creation and strategy development. This tool not only helps in understanding the current sentiment but also in tracking changes over time, thereby enabling continuous improvement and optimization of content. Overall, ECHO INSIGHTS is designed to be an invaluable asset for anyone looking to delve deeper into the sentiment behind YouTube videos. By transforming raw comment data into meaningful insights, it facilitates a deeper connection with the audience and aids in crafting content that resonates more effectively.

ACKNOWLEDGEMENT

First, we thank the almighty god for the successful completion of the project. Our sincere thanks to our chairman **Mr. S. Meganathan B.E., F.I.E.,** for his sincere endeavor in educating us in his premier institution. We would like to express our deep gratitude to our beloved Chairperson **Dr. Thangam Meganathan Ph.D.,** for her enthusiastic motivation which inspired us a lot in completing this project and Vice Chairman **Mr. Abhay Shankar Meganathan B.E., M.S.,** for providing us with the requisite infrastructure. We also express our sincere gratitude to our college Principal, **Dr. S. N. Murugesan M.E., PhD.,** and **Dr. P. KUMAR M.E., PhD, Director computing and information science , and Head Of Department of Computer Science and Engineering** and our project coordinator **Mr.K.Vijay B.Tech., M.E.,,** for his encouragement and guiding us throughout the project towards successful completion of this project and to our parents, friends, all faculty members and supporting staffs for their direct and indirect involvement in successful completion of the project for their encouragement and support.

SANJEEV U

SENTHILNAATHAN K

SHARAN ADHITYA C D

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 INTRODUCTION

In today's fast-paced world, Social media sites like YouTube are effective means for content producers to interact with their viewers in the current digital era. To properly customize their content strategies, marketers and content creators must have a thorough understanding of audience sentiment and engagement dynamics. The proposed project, ECHO INSIGHTS, provides a thorough social media analytical tool targeted at YouTube videos in order to meet this demand.

To give consumers practical insights into audience interactions, ECHO INSIGHTS makes use of cutting-edge technology including the YouTube API, sentiment analysis algorithms, and interactive data visualization tools. Through the process of extracting, translating, and analyzing comments from YouTube videos, ECHO INSIGHTS allows users to identify important topics and sentiment patterns that emerge from audience interactions.

Streamlit was used to create an intuitive dashboard interface that allows users to view word clouds representing popular themes, analyze sentiment trends, and access comprehensive sentiment analysis results. To help with additional analysis and documentation, ECHO INSIGHTS also provides easy ways to print or save analysis findings as PDFs.

Content producers, marketers, and researchers can obtain more profound understanding of the attitudes and behaviors of their YouTube audience using ECHO INSIGHTS. This will enable them to make well-informed decisions and maximize their content strategies for maximum effect and success

1.2 PROBLEM STATEMENT

To Create an AI-powered sentiment analysis solution specialized in interpreting emotions within social media content, empowering individuals and organizations to manage their online reputation and perception effectively.

How might we create an AI-powered sentiment analysis solution specialized in interpreting emotions within social media content? Empowering individuals and organizations to manage online reputation and perception effectively.

1.3 SCOPE OF WORK

The scope of this project is to develop a comprehensive social media analytical tool centered on YouTube videos, ECHO INSIGHTS entails integrating multiple technologies and functionalities." The tool will translate comments into English and retrieve them from videos using a smooth connection with the YouTube API. This will assure accessibility.

The emotional tone and sentiment of comments will be analyzed by sophisticated sentiment analysis algorithms, including Vader bots that have been taught, to provide insightful information about audience reactions. The main component of the service will be an interactive dashboard created with Streamlit, which enables users to examine word clouds that represent hot subjects in comments and see sentiment trends over time.

1.4 AIM AND OBJECTIVES OF THE PROJECT

The ECHO INSIGHTS project aims to provide an approachable social media analysis tool with a YouTube video focus. Through the use of interactive data visualization tools, sentiment analysis algorithms, and the YouTube API, the project seeks to offer practical insights into the dynamics of audience interaction and sentiment.

The goal of this project is to develop an extensive tool for YouTube comment analysis. In order to effectively extract comments from videos while respecting API rate limitations, it will make use of the YouTube API. The program will translate comments into English to help with accessibility and analysis in the event of a language barrier. Sentiment analysis will be used to classify comments according to emotions, providing insights into the responses of the audience. Users will be able to observe sentiment patterns and actively explore the data using an interactive dashboard that was created with Streamlit.

Finally, a well-defined project management plan with clear milestones and timelines will ensure efficient collaboration and timely completion

1.5 RESOURCES

This project has been developed through widespread secondary research of accredited manuscripts, standard papers, business journals, white papers, analysts' information, and conference reviews. Significant resources are required to achieve an efficacious completion of this project.

The following prospectus details a list of resources that will play a primary role in the successful execution of our project:

- A properly functioning workstation (PC, laptop, net-books etc.) to carry out desired research and collect relevant content.
- Stable internet connection throughout the process is required for utilizing all the features without any interruption or delays, since this project utilizes certain modules like YouTube API, which requires a stable internet connection in order to retrieve all the comments from the given YouTube Video link provided as the input.

CHAPTER 2
LITERATURE SURVEY

2.1. Sentiment Analysis on Twitter with Python's Natural Language Toolkit and VADER Sentiment Analyzer (Elbagir et al., 2019):

This paper talks about figuring out how people feel from stuff they write on Twitter (social media platform). It uses a tool called VADER, which might be useful for our project, even though it focuses on Twitter. Still a good reference for understanding sentiment analysis.

2.2. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython (McKinney, 2012):

Data manipulation It taught me how to organize and play around with data (data manipulation) using this library called pandas. Super helpful because we'll need to do that with all the comments I collect. Definitely keeping this one around.

2.3. Python Data Analysis with pandas (Bernard, 2016):

Practical It shows how to actually use the library in real-world situations, which is perfect for putting what we learned from the first paper (McKinney, 2012) into practice.

2.4. The Power of Social Media Analytics: Text Analytics Based on Sentiment Analysis and Word Clouds on R (Kabir et al., 2018):

This paper goes beyond just figuring out how people feel (sentiment analysis). It talks about using those feelings and what people say (text analytics) to understand social media better. Plus, it shows how to make cool pictures (word clouds) that show what people are talking about the most. This could be a cool way to present the results of my project.

2.5. Sentiment Analysis Using Machine Learning Techniques on Python (Rathee et al., 2018):

Machine learning This one dives into super advanced stuff using machines (machine learning) to understand feelings (sentiment analysis). Even though I might not use this in my project right now, it's cool to see how far this technology can go.

2.6. Sentiment Analysis for Youtube Videos with User Comments: Review (Alhujaili and Yafooz, 2021):

It's all about figuring out how people feel (sentiment analysis) from YouTube videos and the comments people leave. This is exactly what we want to do in my project, so this paper is super important.

2.7. Web Application Development with Streamlit (Khorasani et al., 2020):

This paper is like a guide for building the actual program (web application) people will use to interact with my project. Streamlit seems like a good tool to use, so this will be helpful when I get to that stage.

2.8 Language Translation Papers (Manish & Mohammad, 2019; Kolhar & Alameen, 2021):

If I ever want to add a feature where the program translates comments into different languages (machine translation), these papers would be useful. They explain how translation works and what to consider, like if it might not be perfect or could cost money.

2.9. Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques (Dashtipour et al., 2016):

Similar to the translation papers, this one focuses on figuring out how people feel (sentiment analysis) in comments written in different languages (multilingual). This could be helpful if I ever want to analyze comments from all over the world.

2.10. Semantic Sentiment Analysis of Twitter (Saif et al., 2012):

This paper goes a bit deeper than just positive or negative feelings (sentiment analysis). It talks about understanding the actual meaning (semantic analysis) behind what people write. This might be more than I need for my project now, but it's interesting to know there are ways to get a more nuanced understanding of what people are saying.

2.11. Scikit-learn: Machine Learning in Python (Pedregosa et al., 2011):

This one's another advanced topic about machines and learning (machine learning). While we might not use it now, it could be a good starting point if we ever want to explore using machines in my project in the future.

2.12. K. Sahoo, et al "Exploratory data analysis using Python"

These papers might offer broader background knowledge on data analysis principles, which can be helpful when working with comment dataframes. These papers seem like general guides for working with data (data analysis). They might come in handy when we need to understand the information I collect from the comments (dataframes).

CHAPTER 3

SYSTEM DESIGN

3.1 GENERAL

In this section, we would like to show how the general outline of how all the components end up working when organized and arranged together. It is further represented in the form of a flow chart below.

3.2 SYSTEM ARCHITECTURE DIAGRAM

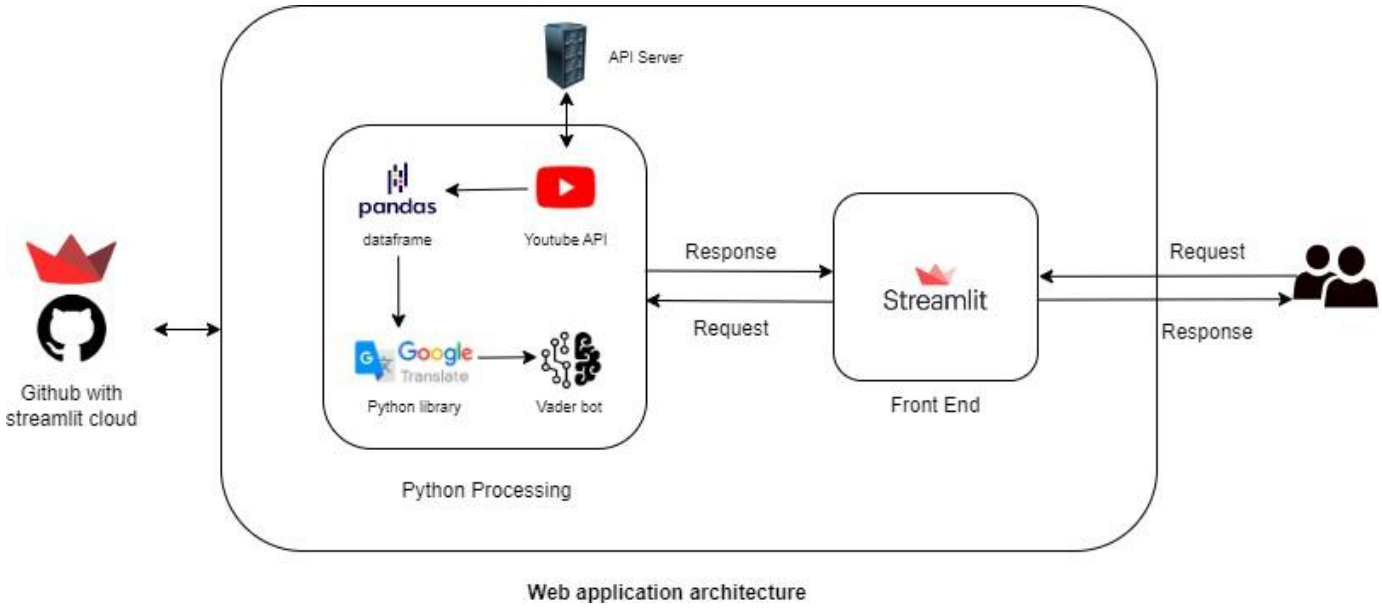


Fig 3.1: System Architecture

The Fig 3.1 is the architecture diagram of ECHO INSIGHTS, with user request and response models connected to streamlit, a front-end framework and the comments are fetched from YouTube API V3 and a data frame is formed using

out of it and then translated using google trans module of library of python and processed using Vader bots for sentiment. Then the results are sent as response to the streamlit dashboard with an intuitive UI and graphs. Later the code is pushed to GitHub and can be hosted using streamlit cloud.

3.3 DEVELOPMENT ENVIRONMENT

3.3.1 HARDWARE REQUIREMENTS

The hardware requirements may serve as the basis for a contract for the system's implementation. It should therefore be a complete and consistent specification of the entire system. It is generally used by software engineers as the starting point for the system design.

Processor	Core i3 or more
RAM	4GB or more
Browser	Chrome 64+ or Edge 79+
OS	Any OS with supported browsers

TABLE 3.1 HARDWARE REQUIREMENTS

3.3.2 SOFTWARE REQUIREMENTS

The software requirements for "ECHO INSIGHTS" are as follows:

1) Any mobile or computing devices with a web browser
2) All the libraries installed or via req.txt (Streamlit, Vader, Google trans)

CHAPTER 4

PROJECT DESCRIPTION

4.1 METHODOLOGY

With the help of your YouTube sentiment analysis app, you may gain a deeper understanding of your audience. Through a user-friendly Streamlit interface, you may examine comments related to videos you provide by utilizing the YouTube Data API. The sentiment analysis program VADER then analyzes each comment's underlying emotions and assigns a score to classify it as positive, negative, or neutral. Even comments written in foreign languages can be easily translated for thorough examination with Google Translate.

The software does more than just provide emotion scores. It offers a basic sentiment that encapsulates the atmosphere of the comments area. Additionally, it highlights the most encouraging and unfavorable remarks, which can help you figure out what appeals to your audience and where you might make improvements. An eye-catching word cloud illustrates the most frequently used words and insights.

With an interactive dashboard that lets you easily study trends, filter comments, and dig deeper into the data, Streamlit excels once more. This user-centered strategy guarantees a seamless and enlightening encounter. After deployment and testing, the app is prepared to provide insightful information about the mood of your audience with a few clicks.

4.2 MODULE DESCRIPTION

The development process for "ECHO INSIGHTS – Youtube Sentiment analysis tool encompasses four key modules.

4.2.1 DASHBOARD

The main dashboard, uses streamlit, a python-based web framework that ensures easy integration with py models and other components much easier. Here, the Input, i.e., the YouTube video which is to be analyzed, is taken as URL from user and manipulates it with the processes and returns the result for user in intuitive dashboard.

4.2.2 FETCH COMMENTS

Then initial module is “fetch_comments”. This module utilizes googleclient’s YouTube API V3 to fetch the comments the given youtube video link. Here, we set the max counts to 100. This API will take the given URL of the YouTube video, get the video ID from it, and returns the Author, published time, Updated time, Like count and text of each comment. This data will be stored in a dataframe (a Data table).

4.2.3 TRANSLATE

The "Translate_cmt" translates YouTube comments. It uses Google Translate's Python library (googletrans) to turn comments into a chosen language. This module gets the dataframe from the fetch_comments module as an input, translates it using the googletrans library and finally outputs the translated comments as another dataframe for further analysis. Beware, translation speed can vary depending on your device's processing power.

4.2.4 SENTIMENT

Next comes “vader_sentiment”. This module uses the vader bot, which is a pre trained model used to alter the sentiments for the given texts and updates the sentiment value of each comment as a new column to dataframe alongside the previous data of the comments.

Vader bot provides a better result, in terms of finding the sentiment of the given text, when compared to other pre trained models that can be used for sentimental analysis.

4.2.5 INSIGHTS

The “Insights” module is used to provide the analysis of the comments after all the processing stages like translating and sentiment assigning. This module displays the top 10 comments based on the sentiment as a table, worst10 comments based on the sentiment as a table. Also, this module shows the overall sentiment of the given YouTube video’s comments, thereby giving an overview of the video’s overall sentiment, by aggregating the sentiments of each comment. Then, this module shows the graph stating the distribution of the sentiment in the given YouTube video. This graph is typically like a bell-curve. If it is leaned towards right side, then the video is positive and if it is leaned towards left side, then the video is negative. Then, another graph is also displayed showing the comments over time, by plotting the published date against the comment count using plotly library of python.

4.2.6 WORDCLOUD

“Word cloud”, a simple python library that uses matplot lib to plot most frequently used words based on sizes for a better visual glance and return to key page. This module utilizes this Wordcloud python library, to visualize the most frequently used words in the given YouTube video’s comments. The size of the words displayed is correlated to the frequency of the word. The bigger the word is, the frequent it is present in the comments of the given YouTube video.

CHAPTER 5
RESULTS AND DISCUSSIONS

5.1 OUTPUT

The following contains the images of the working application

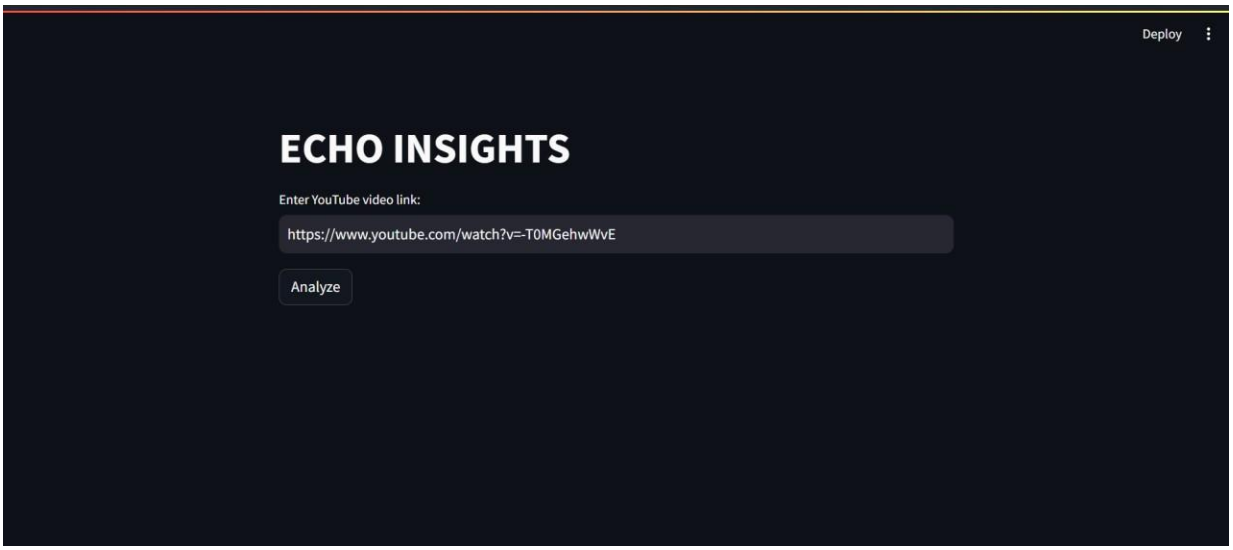


Fig 5.1.1 – Welcome page

Above fig 5.1.1 is the home page of the ECHO INSIGHTS web application. This page prompts the user to input the URL of the YouTube video which is to be analyzed. On clicking the analyses button, the YouTube API, provided by Google, fetches YouTube video's comments.

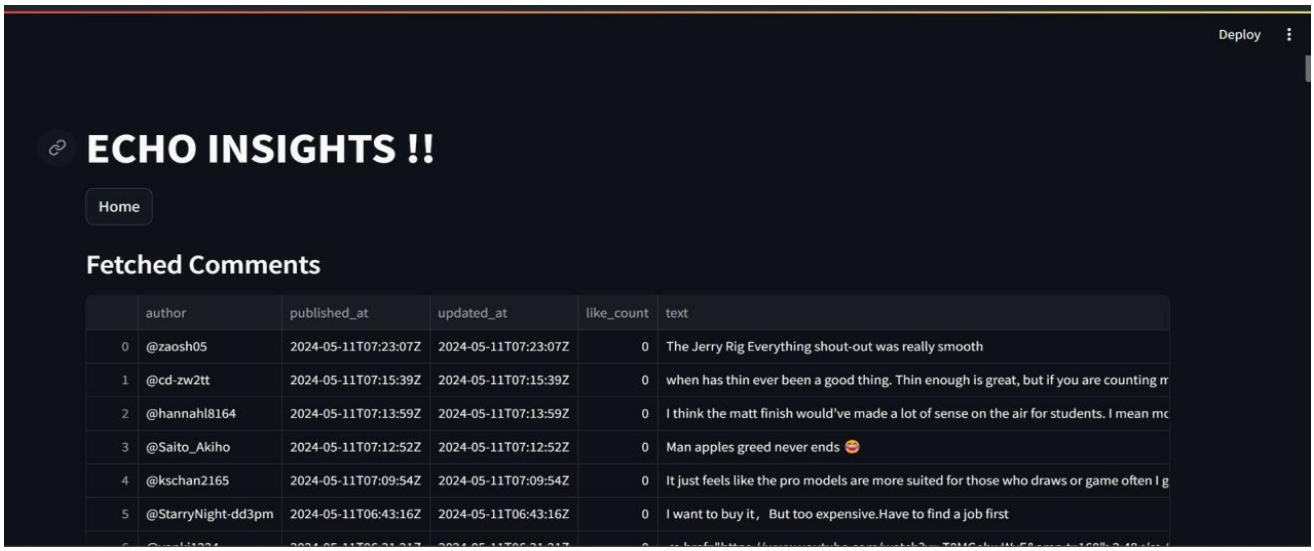


Fig 5.1.2 – Displays fetched comments

Fig 5.1.2 is a Secondary page that shows the comments of the YouTube video, that is fetched by the YouTube API. Here, it displays the fetched data, which has the columns like author, published at, updated at, like count and text of each comment, in the form of a table.

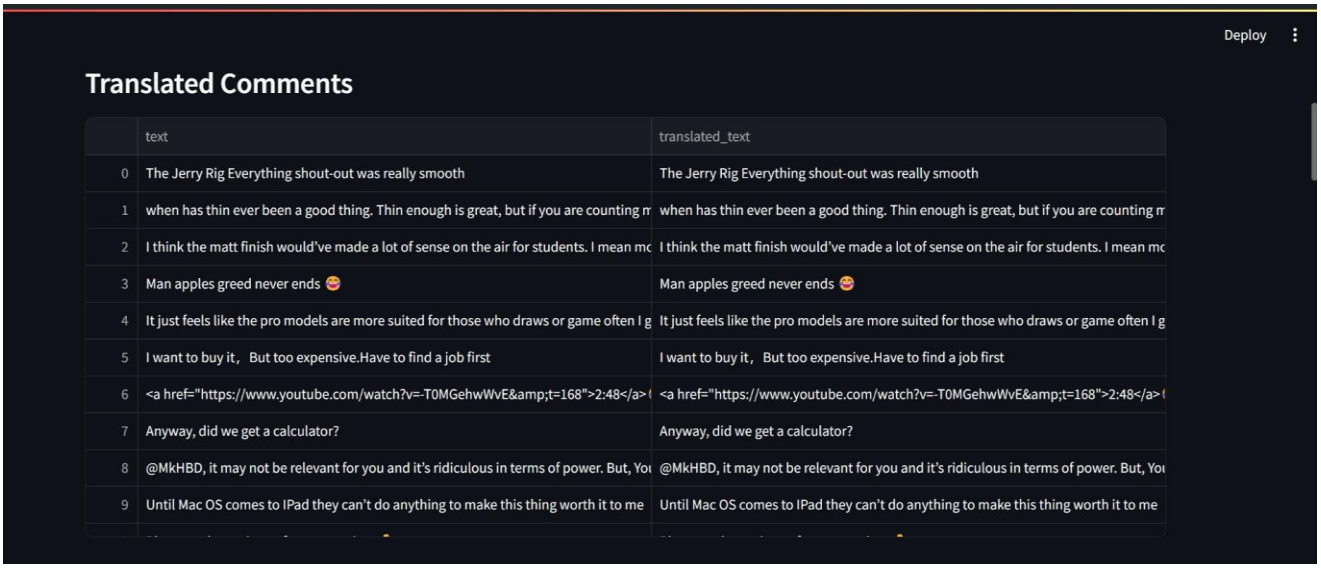
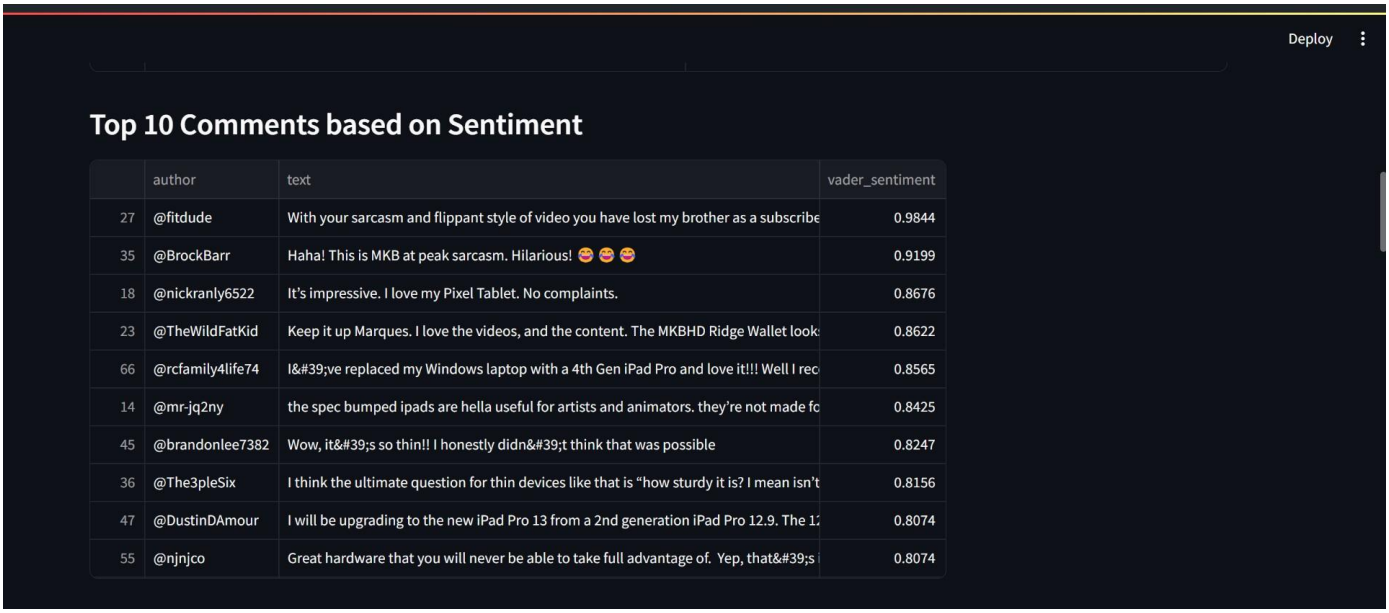


Fig 5.1.3 – Display translated comments

As shown in fig 5.1.3, the google trans library from python to translate all the comments in

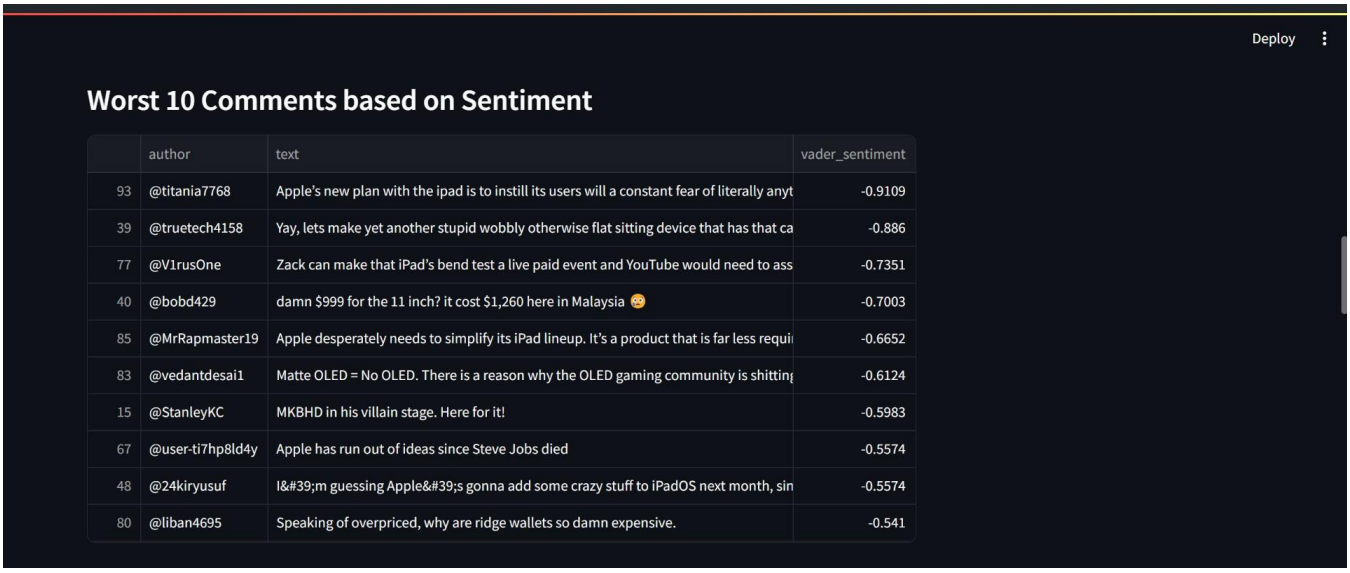
the YouTube video into English (to an extent) and appends the translated comments as another column named “Translated text”, alongside each comment, in the dataframe. But only the text and translated text are displayed here.



	author	text	vader_sentiment
27	@fitdude	With your sarcasm and flippant style of video you have lost my brother as a subscribe	0.9844
35	@BrockBarr	Haha! This is MKB at peak sarcasm. Hilarious! 😂 😂 😂	0.9199
18	@nickranly6522	It’s impressive. I love my Pixel Tablet. No complaints.	0.8676
23	@TheWildFatKid	Keep it up Marques. I love the videos, and the content. The MKBHD Ridge Wallet look	0.8622
66	@rcfamily4life74	I've replaced my Windows laptop with a 4th Gen iPad Pro and love it!!! Well I rec	0.8565
14	@mr-jq2ny	the spec bumped ipads are hella useful for artists and animators. they’re not made fo	0.8425
45	@brandonlee7382	Wow, it's so thin!! I honestly didn't think that was possible	0.8247
36	@The3pleSix	I think the ultimate question for thin devices like that is “how sturdy it is? I mean isn’t	0.8156
47	@DustinDAmour	I will be upgrading to the new iPad Pro 13 from a 2nd generation iPad Pro 12.9. The 1	0.8074
55	@njnjco	Great hardware that you will never be able to take full advantage of. Yep, that's	0.8074

Fig 5.1.4 – Sorted comments based on sentiment – increasing

As seen in fig in fig 5.1.4, After adding sentiments to the results using the Vader bot, here, the top 10 comments are displayed to the user in a data frame table format with author, comment text the sentiment value of each comment based on the sentiment values that is assigned using the bot. This provides an insight of the best comments present in the video.



	author	text	vader_sentiment
93	@titania7768	Apple's new plan with the ipad is to instill its users will a constant fear of literally anyt	-0.9109
39	@truetech4158	Yay, lets make yet another stupid wobbly otherwise flat sitting device that has that ca	-0.886
77	@V1rusOne	Zack can make that iPad's bend test a live paid event and YouTube would need to ass	-0.7351
40	@bobd429	damn $999 for the 11 inch? it cost $1,260 here in Malaysia 😳	-0.7003
85	@MrRapmaster19	Apple desperately needs to simplify its iPad lineup. It's a product that is far less requi	-0.6652
83	@vedantdesai1	Matte OLED = No OLED. There is a reason why the OLED gaming community is shittin	-0.6124
15	@StanleyKC	MKBHD in his villain stage. Here for it!	-0.5983
67	@user-ti7hp8ld4y	Apple has run out of ideas since Steve Jobs died	-0.5574
48	@24kiryusuf	I'm guessing Apple's gonna add some crazy stuff to iPadOS next month, sin	-0.5574
80	@liban4695	Speaking of overpriced, why are ridge wallets so damn expensive.	-0.541

Fig 5.1.5 – Sorted comments based on sentiment

As shown in fig 5.1.5, After adding sentiments to the results using the Vader bot, here, the worst 10 comments are displayed to the user in a data frame table format with author, comment text the sentiment value of each comment based on the sentiment values that is assigned using the bot. This provides an insight of the best comments present in the video.

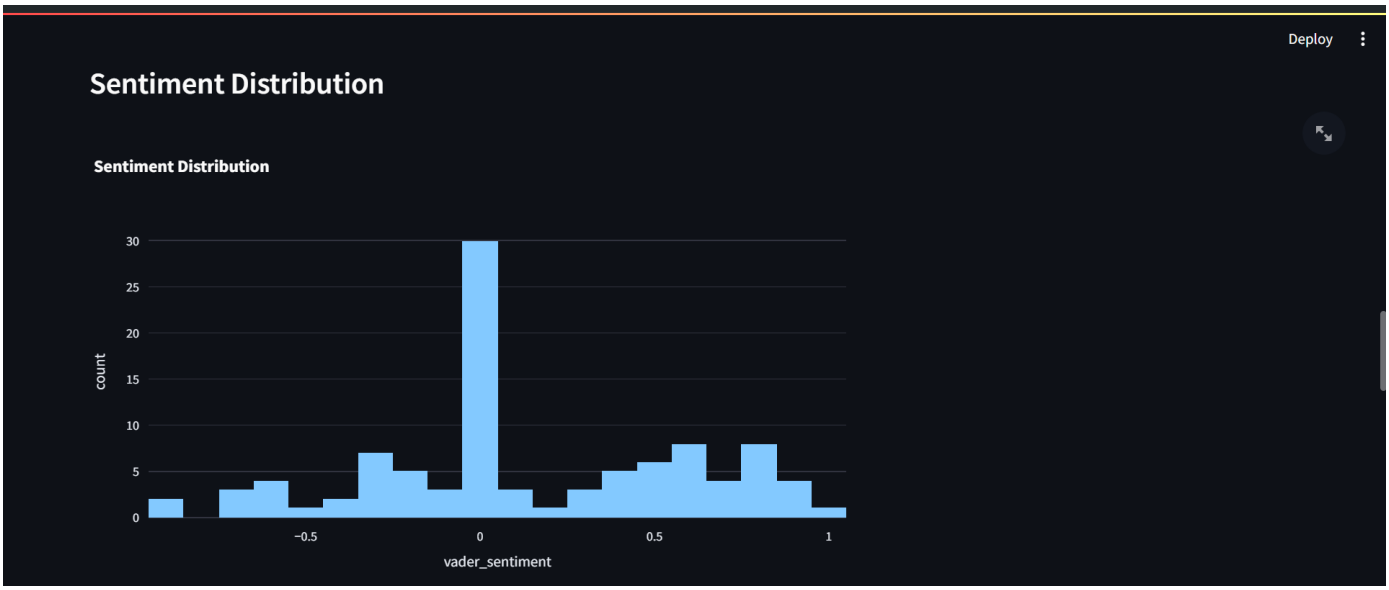


Fig 5.1.6 – Sentiment distribution graph

As seen in fig 5.1.6, On processing the data, the sentiment distribution graph is plotted using

plotly library of python. this shows the sentiment distribution over -1 to 1 with count (as of 100 results fetched). This graph is typically like a bell-curve. If it is leaned towards right side, then the video is positive and if it is leaned towards left side, then the video is negative

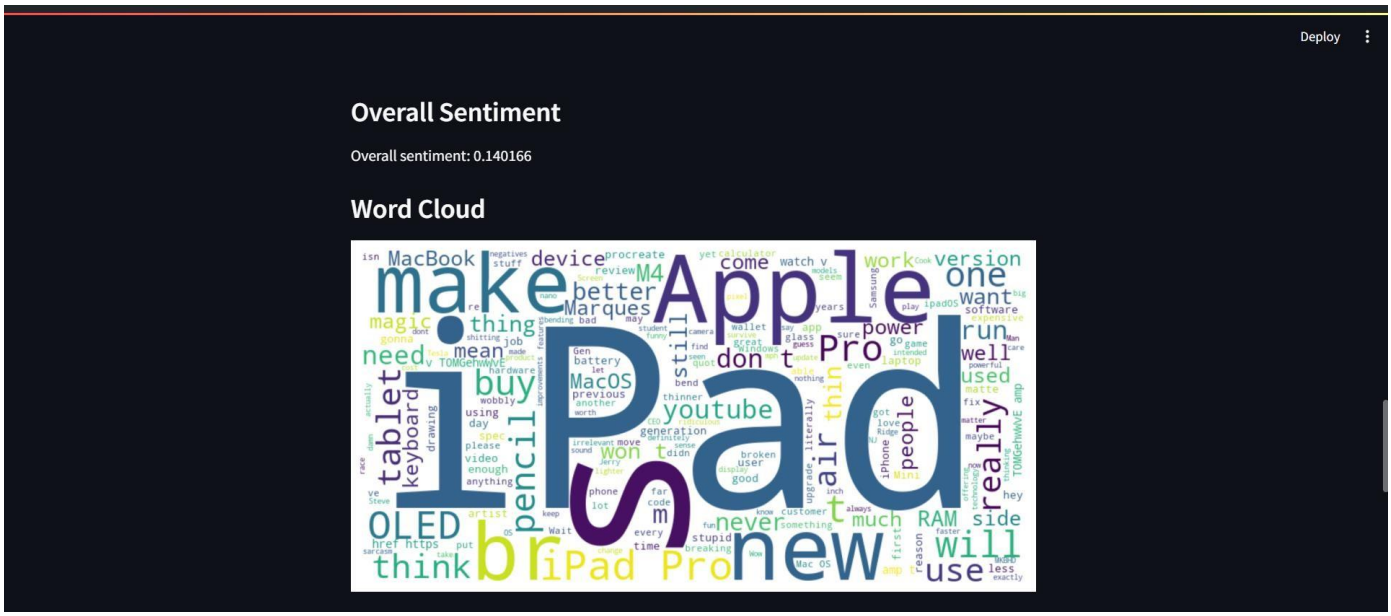


Fig 5.1.7 – word cloud of comments

As shown in fig 5.1.7 here, Wordcloud python library, is used visualize the most frequently used words in the given YouTube video's comments. The size of the words displayed is correlated to the frequency of the word. The bigger the word is, the frequent it is present in the comments of the given YouTube video.

Fig 5.1.8 – comments over time graph

This is the final module as seen in fig 5.1.8 here we plot the comments over time as a graph, where the comment date is plotted against the comment count using plotly library of python. This is used to find out the reach of the video over the time period.

5.2 RESULT

With our in-depth study of YouTube comments, you can unleash the power of audience insights! This analysis reveals the general sentiment—positive, negative, or neutral—by unlocking the emotional undertones of the comments area of your video. We go above and above by highlighting the most encouraging and critical remarks, serving as valuable nuggets of insight into what appeals to your audience and possible areas for development. However, the examination doesn't end there! The most often used words are graphically represented in a stunning word cloud that takes center stage. This offers an overview of the main ideas and subjects that generated the greatest conversation in the comments.

We care about the user experience more than just statistics. With Streamlit, we created an interactive dashboard that gives you control. Sort comments according to sentiment. For an analysis that is laser-focused, filter comments according to sentiment or keywords. Examine sentiment patterns over time to observe changes in audience perception. Investigate specific comments and their sentiment scores to learn more about your audience's responses.

You may adjust your material to the tastes of your audience by utilizing the power of this study. By cultivating a supportive and active community around your channel, you can eventually soar to new heights with your YouTube presence.

CHAPTER 6

CONCLUSION AND FUTURE ENHANCEMENT

6.1 CONCLUSION

To sum up, ECHO INSIGHTS is an essential resource for researchers, marketers, and content creators that want to understand the nuances of audience emotion and interaction on YouTube. Through the utilization of cutting-edge technology such as the YouTube API and sentiment analysis algorithms, customers are enabled to obtain practical insights regarding audience interactions. Comprehensive sentiment analysis results and interactive data visualization tools are easily accessible through the dashboard's user-friendly design, which is powered by Streamlit. Users may maximize effect and refine content strategy by using ECHO INSIGHTS to uncover sentiment trends, emerging topics, and patterns within audience interactions.

6.2 FUTURE ENHANCEMENT

To give consumers a complete picture of their digital footprint, extend ECHO INSIGHTS to evaluate sentiment and interaction patterns across a variety of social media sites, including Facebook, Instagram, and Twitter. Introducing configurable reporting options will enable customers to adapt analysis results to their own requirements. These options include the capacity to produce comprehensive reports in a variety of formats, including interactive dashboards, CSV files, and PDFs. Allow users to proactively modify their content strategies for optimal effect and success by integrating predictive analytics tools to foresee future audience behavior and trends based on historical data.

APPENDIX

SOURCE CODE:

Fetch.py

```
import googleapiclient.discovery
from googleapiclient.discovery import build import
pandas as pd
def fetch_comments(video_id):

    api_service_name = "youtube"
    api_version = "v3"
    DEVELOPER_KEY = "AIzaSyB47XVqTefMFpRGzMavn1KMc9W1Jf8BrCg"

    youtube = googleapiclient.discovery.build(api_service_name, api_version,
developerKey=DEVELOPER_KEY)

    request = youtube.commentThreads().list(
    part="snippet",
    videoId=video_id,
    maxResults=100
    )
    response = request.execute()
    comments = []
    for item in response['items']:
      comment = item['snippet']['topLevelComment']['snippet']
      comments.append([
        comment['authorDisplayName'],
```

```
            comment['publishedAt'],
            comment['updatedAt'],
            comment['likeCount'],
            comment['textDisplay']
        ])

    df = pd.DataFrame(comments, columns=['author', 'published_at', 'updated_at',
'like_count', 'text'])

    return df
```

Pdfp.py

```
# import pandas as pd
# from pandas_profiling import ProfileReport

# def generate_profile(df):
#     # Generate the profile report
#     profile = ProfileReport(dataframe, explorative=True) #
        return profile
import pandas as pd
from pandas_profiling import ProfileReport
# from pydantic import BaseModel, BaseSettings, Field, PrivateAttr



def generate_profiling_report(df):
```

```
    profile = ProfileReport(df, explorative=True) return
    profile.to_html()
```

Index.py

```
import streamlit as st
import pandas as pd
import plotly.express as px
from fetch import fetch_comments
from trans import translate_comments
from vader import analyze_sentiment
from wordcld import generate_wordcloud
from pdpf import generate_profiling_report

def input_page(): st.title("ECHO
    INSIGHTS")

    # Get YouTube video link from user
    youtube_link = st.text_input("Enter YouTube video link:") if
    st.button("Analyze") and youtube_link:
        # Extract video ID from the link video_id
        = youtube_link.split("=")[-1]
        # Set query parameters to navigate to the next page st.session_state["query_params"]
        = {"youtube_link": youtube_link, "video_id":
video_id}
```

```
def result_page(): st.title("ECHO
    INSIGHTS !!")

    if st.button("Home"):
        # Clear session state
        st.session_state.clear() #
        Redirect to input page
        input_page()

    # Get video ID from query parameters
    query_params = st.session_state.get("query_params", {})
    youtube_link = query_params.get("youtube_link", "")
    video_id = query_params.get("video_id", "")
    # Fetch comments
    df = fetch_comments(video_id)
    st.subheader("Fetched Comments")
    # Display selected columns' heads in Streamlit
    st.write(df.head(15))

    # Translate comments
    translated_comments = translate_comments(df)
    selected_columns = ['text', 'translated_text'] selected_df =
    df[selected_columns] st.subheader("Translated
    Comments") st.write(selected_df.head(15))
```

```
# Perform sentiment analysis
sentiment_analysis = analyze_sentiment(translated_comments)
selected_columns = ['author','text', 'translated_text','vader_sentiment']
selected_df = df[selected_columns]


# Combine sentiment analysis with original data df['vader_sentiment'] =
sentiment_analysis['vader_sentiment']

# Sort comments based on sentiment score
top_comments = df.sort_values(by='vader_sentiment', ascending=False).head(10)

# Display top 10 comments with sentiments
st.subheader("Top 10 Comments based on Sentiment")
st.write(top_comments[['author', 'text', 'vader_sentiment']])
worst_comments = df.sort_values(by='vader_sentiment', ascending=True).head(10)
# Display worst 10 comments with sentiments
st.subheader("Worst 10 Comments based on Sentiment")
st.write(worst_comments[['author', 'text', 'vader_sentiment']])

# Display sentiment distribution plot
st.subheader("Sentiment Distribution")
fig = px.histogram(df, x='vader_sentiment', nbins=30, title="Sentiment Distribution")

st.plotly_chart(fig)

# Display overall sentiment
overall_sentiment = sentiment_analysis['vader_sentiment'].mean() st.subheader("Overall
```

```
    Sentiment")
    st.write(f"Overall sentiment: {overall_sentiment}")

    st.subheader("Word Cloud")
    generate_wordcloud(translated_comments['text'])

    # st.subheader("Pandas Profiling Report")
    # profiling_report = generate_profiling_report(df) #
    st.write(profiling_report)
    # Comments over time
    df['comment_published'] = pd.to_datetime(df['published_at'])
    df['comment_date'] = df['comment_published'].dt.date
    comments_over_time =
df.groupby('comment_date').size().reset_index(name='comment_count')
    st.subheader("Comments Over Time")
    fig_comments_over_time = px.line(comments_over_time, x='comment_date',
y='comment_count', title='Comments Over Time')
    st.plotly_chart(fig_comments_over_time)



def main():
    # Render the appropriate page based on query parameters if
    st.session_state.get("query_params"):
        result_page() else:
        input_page()
```

```
if _name__== "_main_":
    main()
```

Translate.py

```
def translate_comments(df):
    # Your code to translate comments here
    from googletrans import Translator
    def translate_text(text, src='auto', dest='en'): translator =
      Translator()

      translated_text = translator.translate(text, src=src, dest=dest) return
      translated_text.text

    # Apply translation to the 'text' column
    df['translated_text'] = df['text'].apply(lambda x: translate_text(x)) return df
```

Vader.py

```
def analyze_sentiment(df):
    from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
```

```
    # Function to perform sentiment analysis using VADER
    def get_vader_sentiment(text):
        analyzer = SentimentIntensityAnalyzer()
        sentiment_score = analyzer.polarity_scores(text)
        # We extract the compound score which represents the
        overall sentiment return sentiment_score['compound']

    # Apply sentiment analysis to the 'translated_text' column df['vader_sentiment']
    = df['translated_text'].apply(get_vader_sentiment) return df
```

Wordcloud.py

```
import streamlit as st
from wordcloud import
WordCloud import
matplotlib.pyplot as plt


def generate_wordcloud(df_column):
    comment_text = ' '.join(df_column)
    wordcloud = WordCloud(width=800, height=400,
background_color='white').generate(comment_text)
    fig, ax = plt.subplots(figsize=(10, 5))


    ax.imshow(wordcloud, interpolation='bilinear')
    ax.axis('off')
    st.pyplot(fig)
```

REFERENCES

[1] R. Singh and A. Tiwari, "YOUTUBE COMMENTS SENTIMENT ANALYSIS," *Int. J. Sci. Res. Eng. Manag.(IJSREM*, *no. May*, 2021, [Online]. Available: https://www.researchgate.net/profile/Ritika-Singh-38/publication/351351202_YOUTUBE_COMMENTS_SENTIMENT_ANALYSIS/links/6092e8b2299bf1ad8d7d79a7/YOUTUBE-COMMENTS-SENTIMENT-ANALYSIS.pdf

[2] R. Manish, A. Mohammad, Department of Computer Science & Engineering, Sant Gadge Baba Amravati University, Amravati, India, and Department of Computer Science & Engineering, Sant Gadge Baba Amravati University, Amravati, India, "Language translation: Enhancing bi-lingual machine translation approach using python," *I-manag. S J. Comput. Sci.*, vol. 7, no. 2, p. 36, 2019.

[3] R. Pradhan, "Extracting Sentiments from YouTube Comments," in 2021 Sixth International Conference on Image Information Processing (ICIIP), IEEE, Nov. 2021, pp. 1–4.

[4] K. Sahoo, A. K. Samal, and J. Pramanik, "Exploratory data analysis using Python," *Technology*, 2019, [Online]. Available: https://www.researchgate.net/profile/Dr-Subhendu-Pani/publication/337146539_IJITEE/links/5dc70b124585151435fb427e/IJITEE.pdf

[5] M. Khorasani, M. Abdou, and J. H. Fernández, *Web Application Development with Streamlit*. Apress.

[6] S. Elbagir and J. Yang, "Sentiment Analysis on Twitter with Python?s Natural Language Toolkit and VADER Sentiment Analyzer," in *IAENG Transactions on Engineering Sciences*, WORLD SCIENTIFIC, 2019, pp. 63–80.

[7] W. McKinney, *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. "O'Reilly Media, Inc.," 2012.

[8] J. Bernard, "Python Data Analysis with pandas," in *Python Recipes Handbook: A Problem-Solution Approach*, J. Bernard, Ed., Berkeley, CA: Apress, 2016, pp. 37–48.

[9] A. I. Kabir, R. Karim, S. Newaz, and M. I. Hossain, “The power of social media analytics: Text analytics based on sentiment analysis and word clouds on R,” *Inform. Econ.*, vol. 22, no. 1/2018, pp. 25–38, Mar. 2018.

[10] N. Rathee, N. Joshi, and J. Kaur, “Sentiment Analysis Using Machine Learning Techniques on Python,” in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, IEEE, Jun. 2018, pp. 779–785.

[11] R. F. Alhujaili and W. M. S. Yafooz, “Sentiment Analysis for Youtube Videos with User Comments: Review,” in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, IEEE, Mar. 2021, pp. 814–820.

[12] M. Kolhar and A. Alameen, “Artificial intelligence based language translation platform,” *Intell. Autom. Soft Comput.*, vol. 28, no. 1, pp. 1–9, 2021.

[13] K. Dashtipour *et al.*, “Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques,” *Cognit. Comput.*, vol. 8, pp. 757–771, Jun. 2016.

[14] H. Saif, Y. He, and H. Alani, “Semantic Sentiment Analysis of Twitter,” in *The Semantic Web – ISWC 2012*, Springer Berlin Heidelberg, 2012, pp. 508–524.

[15] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.