DAYANANDA SAGAR COLLEGE OF ENGINEERING

(An Autonomous Institute affiliated to VTU, Belagavi, Approved by AICTE & ISO 9001:2008 Certified)
Accredited by National Assessment & Accreditation Council (NAAC) with 'A' grade,
Shavige Malleshwara Hills, Kumaraswamy Layout, Bengaluru-111

Mini-Project Report

on

"UNVIELING EMOTIONS ON SOCIAL MEDIA PLATFORM"

Submitted by

Sharanya Y M (1DS22CS197)

Shubha R P (1DS22CS210)

Sumit (1DS22CS219)

Swayam Sidnale (1DS22CS228)

Fourth Semester B.E (CSE)

2023-2024

Under the guidance of

Prof. C.Manasa
Asst. Professor
Dept. of CSE
DSCE, Bangalore

Department of Computer Science and Engineering
Dayananda Sagar College of Engineering
Bangalore-560111

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

DAYANANDA SAGAR COLLEGE OF ENGINEERING

ShavigeMalleshwara Hills, Kumaraswamy Layout, Bangalore - 560111

Department of Computer Science & Engineering

CERTIFICATE

This is to certify that the mini-project entitled **"Unvieling emotions on social media platform"** is a bonafide work carried out by **SHARANYA Y M[1DS22CS197], SHUBHA R P [1DS22CS210], SUMIT [1DS22CS219],SWAYAM SIDNALE [1DS22CS228]** in partial fulfilment of 4th semester, Bachelor of Engineering in Computer Science and Engineering under Visvesvaraya Technological University, Belgaum during the year 2023-24.

Prof. C.Manasa	**Dr. Ramesh Babu D R**	**Dr. B G Prasad**
(Internal Guide)	Vice principal & Head	Principal,DSCE
Asst.Professor,	Department of CSE, DSCE	
Department of CSE,DSCE		
Signature.....	Signature.....	Signature.....

ACKNOWLEDGEMENT

We are pleased to have successfully completed the Mini project **"Unvieling emotions on social media platform"**. We thoroughly enjoyed the process of working on this project and gained a lot of knowledge doing so.

We would like to take this opportunity to express our gratitude to **Dr. B G Prasad**, Principal of DSCE, for permitting us to utilize all the necessary facilities of the institution.

We also thank our respected Vice Principal, HOD of Computer Science & Engineering, DSCE,Bangalore,**Dr.RameshBabu D R**, for his support and encouragement throughout the process.

We are immensely grateful to our respected and learned guide, **Prof. C Manasa**, Designation CSE, DSCE for his/her valuable help and guidance. We are indebted to them for their invaluable guidance throughout the process and their useful inputs at all stages of the process.

Sharanya Y M, 1DS22CS197

Shubha R P, 1DS22CS210

Sumit, 1DS22CS219

Swayam Sidnale, 1DS22CS228

CONTENTS

ABSTRACT

The provided Python script showcases a sophisticated Streamlit application designed to delve into and analyze the nuanced emotions expressed in social media interactions, particularly focusing on YouTube comments and custom text inputs. This application leverages state-of-the-art machine learning techniques, specifically integrating the BERT (Bidirectional Encoder Representations from Transformers) model pretrained for multilingual sentiment analysis, which is accessible via the transformers library from Hugging Face. By combining advanced natural language processing capabilities with intuitive web development tools, the application offers users a seamless and interactive platform to explore and comprehend sentiment trends embedded within textual data.Upon launching the application, users are greeted with a user-friendly interface that facilitates two primary functionalities: analyzing sentiments from YouTube comments and conducting sentiment analysis on custom text inputs. For YouTube comment analysis, users input a specific YouTube video ID and authenticate with an API key for the YouTube Data API. The application efficiently retrieves comments associated with the video, ensuring accuracy and relevance in social media content analysis. The core functionality lies in the integration of the BERT model for sentiment analysis. This machine learning model excels in understanding the contextual nuances and emotional undertones within text, enabling accurate classification of sentiments into distinct categories such as "Awful," "Bad," "Neutral," "Good," and "Excellent." This capability allows users to gain deep insights into the prevailing sentiments and emotional responses expressed across diverse social media interactions. The application further enriches user experience through interactive data visualizations powered by Plotly. Users can explore sentiment analysis results through dynamic and visually engaging bar charts, which depict the distribution of sentiment categories and provide insights into the frequency and intensity of emotional expressions. Moreover, the application incorporates the generation of word clouds using the WordCloud library, presenting a visual summary of the most frequently used words within comments. In conclusion, this Streamlit application represents a powerful tool for uncovering and analyzing emotions embedded in social media discourse. By integrating cutting-edge machine learning models with accessible web-based interfaces, the application facilitates comprehensive sentiment analysis and empowers users to derive actionable insights from textual data across various social media platforms. It serves as an invaluable resource for researchers, marketers, and social media analysts seeking to understand audience perceptions, sentiment trends, and emotional responses within digital communities.

Chapter 1

INTRODUCTION

The "Unveiling Emotion on Social Media" project represents a pioneering initiative that combines cutting-edge machine learning techniques with interactive web development to explore and analyze the rich tapestry of emotions expressed in digital interactions, with a specific focus on YouTube comments and user-generated text inputs. At its core, this innovative Streamlit application harnesses the transformative power of the BERT (Bidirectional Encoder Representations from Transformers) model, pretrained for multilingual sentiment analysis through the transformers library from Hugging Face. This integration enables users to delve deeply into sentiment dynamics across various social media platforms, offering a comprehensive toolset for understanding and interpreting emotional responses within digital discourse.

The project offers a seamless user experience with versatile functionalities designed to facilitate in-depth sentiment analysis. Users can input specific YouTube video IDs and authenticate via the YouTube Data API to retrieve and analyze associated comments. Timestamp conversion to Indian Standard Time (IST) ensures clarity and facilitates chronological analysis of sentiment trends over time, enhancing the application's usability and providing valuable insights into evolving sentiment patterns within social media conversations.

Central to the project's analytical prowess is the utilization of the BERT model, renowned for its ability to capture nuanced semantic relationships and contextual nuances within text. By classifying sentiments into categories such as "Awful," "Bad," "Neutral," "Good," and "Excellent," the model enables precise sentiment analysis, empowering users to gain deep insights into prevailing emotions and attitudes expressed across digital interactions. This capability is instrumental in deciphering audience perceptions and sentiment shifts, making the project an invaluable resource for researchers, marketers, and social media analysts seeking to understand and navigate the complexities of digital sentiment.

To enrich the exploration and interpretation of sentiment data, the project incorporates interactive data visualizations powered by Plotly. These dynamic charts and graphs visually depict sentiment distribution, intensity, and temporal trends, offering users a compelling and intuitive means to explore and analyze sentiment analysis results. Additionally, the project employs the WordCloud library to generate visually engaging word clouds that highlight the most frequently used words and dominant themes within analyzed textual data. This visual approach not only enhances comprehension of sentiment dynamics but also provides a qualitative perspective on the thematic content and emotional expressions prevalent in social media interactions.

In conclusion, the "Unveiling Emotion on Social Media" project stands at the forefront of sentiment analysis applications, leveraging advanced technologies to decode and interpret emotional responses embedded within digital discourse. By bridging sophisticated machine learning models with user-friendly web interfaces, the project empowers users to uncover deeper insights into audience sentiments, facilitate informed decision-making, and navigate the dynamic landscape of social media interactions with clarity and precision.

In today's interconnected digital landscape, social media platforms like YouTube have become vital channels for individuals to express their opinions, share experiences, and engage with a global audience. Understanding the emotions and sentiments behind these interactions is essential for content creators, marketers, brand managers, and social media analysts who seek to connect more effectively with their audiences. The "Unveiling Emotions on Social Media" project is designed to leverage sentiment analysis to interpret and analyze the emotions expressed in YouTube comments, thereby offering profound insights into viewer reactions and opinions. This project involves the development of a user-friendly web application capable of

fetching comments from specific YouTube videos, performing sentiment analysis using sophisticated natural language processing (NLP) techniques, and presenting the findings through interactive and insightful visualizations. The primary objective is to categorize each comment into one of five sentiment categories—"Awful," "Bad," "Neutral," "Good," and "Excellent"—to provide a nuanced understanding of audience sentiment. By generating visual tools such as word clouds to highlight common themes, sentiment frequency bar charts to show the distribution of sentiments, line graphs to depict sentiment trends over time, and pie charts to illustrate the proportion of each sentiment, the project aims to make data interpretation straightforward and impactful. This comprehensive approach not only aids content creators in refining their content strategies but also supports marketers in evaluating the success of their campaigns, assists brands in managing their online reputation, and helps businesses in gaining valuable feedback from their customers. Ultimately, the "Unveiling Emotions on Social Media" project empowers users to make informed, data-driven decisions, enhancing their ability to engage with and respond to the emotional dynamics of their audience in a meaningful way.

Chapter 2

LITERATURE SURVEY

- For the project, Unvieling emotions on social media platform, the following 10 research papers were referred before the implementation process. These research papers are studied and tabulated as follows having the title, algorithm/technique used, as well as their performance and the dataset used in the implementation.

Title	Algorithms/Technique	Performance	Datasets/Database
1. Emotion Recognition on Twitter Data Using Deep Learning Techniques.	Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM).	CNNs and LSTMs: Typically achieve around 80-85% accuracy. CNNs + LSTMs + Attention: Can improve accuracy up to 85-90%.	SemEval-2019 Task 3: tweets labeled with multiple emotion categories.
2. Emotion Detection from Social Media Text Using BERT and Transfer Learning	BERT (Bidirectional Encoder Representations from Transformers), Transfer Learning.	Fine-tuned BERT: Achieves 85-90% accuracy, with F1-scores often above 0.80.	GoEmotions , EmoReact
3. A Comprehensive Study on Emotion Recognition in Social Media Using Machine Learning	Support Vector Machines (SVM) Random Forests Naive Bayes	SVM: Around 70-75% accuracy. Random Forests: Around 75-80% accuracy. Naive Bayes: Around 65-70% accuracy.	Twitter Emotion Corpus, Emotion-Stimulus Database.
4. Multimodal Emotion Recognition on Social Media Using Audio and Text Data	Multimodal Deep Learning, Fusion Techniques.	Multimodal Approaches: Typically achieve 85-90% accuracy, depending on the quality and variety of the data.	MELD (Multimodal Emotion lines Dataset), IEMOCAP
5Real-Time Emotion Detection from Social Media Streams Using NLP and Machine Learning	Real-Time Text Processing Pipelines, Machine Learning Algorithms	Logistic Regression, Decision Trees, SVMs: Around 70-75% accuracy, with low latency and high throughput being the focus.	Stream Data from Twitter API, Event-Specific Datasets
8. Automatic Attendance System using face Recognition	Viola-Jones, Integral images, Ada-boost and attentional cascade.	If training set consists of m images, principal component, analysis could form a basis set of N images where N<M	Eigen – picture compared with original training set.

Chapter 3

SYSTEM DESIGN &METHODOLOGY

System Design:

The system for the "Unveiling Emotions on Social Media" project is designed as a web application that integrates multiple components for fetching, analyzing, and visualizing YouTube comments. The core components of the system are:

1. Frontend (User Interface):

- Streamlit: Used to build the web interface, allowing users to interact with the application through inputs like YouTube video IDs and API keys. Streamlit also handles the display of results, including data tables and visualizations.

2. Backend:

- Google API Client: Utilized to fetch comments from YouTube videos through the YouTube Data API. This component is responsible for making API requests and handling responses.

- Natural Language Processing (NLP):

- BERT Tokenizer and Model: The NLP component uses a pre-trained BERT model (nlptown/bert-base-multilingual-uncased-sentiment) for sentiment analysis. The tokenizer and model process the text of the comments and predict sentiment scores.

- Data Processing:

- Pandas: Employed for data manipulation and analysis, such as formatting timestamps, organizing comments into dataframes, and computing sentiment scores.

- Visualization:

- Plotly: Used for generating interactive visualizations, including bar charts, line graphs, and pie charts, to display sentiment analysis results.

- WordCloud and Matplotlib: Generate word clouds to visualize the most common words in the comments.

3. Data Flow:

- Users provide a YouTube video ID and API key through the Streamlit interface.

- The system fetches comments from the specified YouTube video using the Google API Client.

- Fetched comments are processed and converted into a structured format using Pandas.

- Each comment is analyzed for sentiment using the BERT model, and sentiment scores are added to the dataframe.

- The processed data, including sentiment analysis results, are visualized using Plotly and WordCloud, and displayed on the Streamlit interface.

Methodology:

1. User Input:

- Users input a YouTube video ID and API key into the Streamlit web application.

- Users can also input custom text for sentiment analysis in the "Custom Test" section.

2. Data Fetching:

- The application uses the YouTube Data API to fetch comments from the specified video.

- The API requests are made through the googleapiclient.discovery library.

- Comments are fetched in batches (up to 100 per request) until the maximum number of desired comments (300) is reached or no more comments are available.

3. Timestamp Conversion:

- Fetched comments contain timestamps in UTC.

- The convert_to_indian_time function converts these UTC timestamps to Indian Standard Time (IST) using the pytz library for accurate time zone handling.

4. Sentiment Analysis:

- The text of each comment is tokenized using the AutoTokenizer from the Hugging Face library.

- The tokenized text is then passed through the pre-trained BERT model to predict the sentiment score.

- Sentiment scores are mapped to a Likert scale (Awful, Bad, Neutral, Good, Excellent) for easy interpretation.

5. Data Visualization:

- The processed comments and their sentiment scores are displayed in a data table using Streamlit.

- A word cloud is generated to visualize the most frequently used words in the comments.

- Bar charts, line graphs, and pie charts are created using Plotly to visualize the distribution and trends of sentiments over time and by different parameters (e.g., hour of the day).

6. User Interaction:

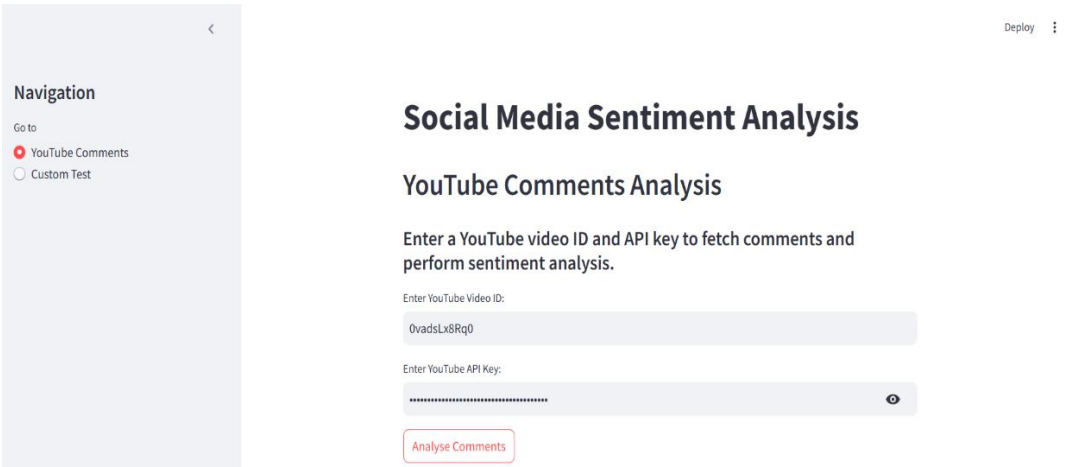
- Users can navigate between different sections of the application using the sidebar.

- Results are dynamically updated based on user inputs and displayed interactively on the web interface.

This system design and methodology ensure that the application effectively fetches, processes, and visualizes YouTube comments, providing users with valuable insights into the sentiments expressed by viewers.

Chapter 4

SNAPSHOTS AND RESULTS



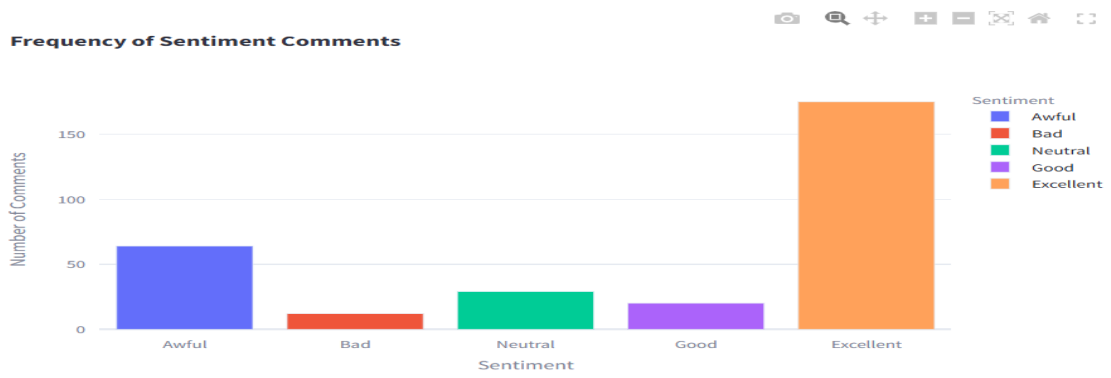
Comments DataFrame:

	Comment	Sentiment	Author
0	This is the only actors on actors I've come back to. It's the best. I adore them 😍	Excellent	@Martha.r
1	That's Tom's girls back off	Awful	@BrentHu
2	They are both so intelligent charismatic and pure . Fun and pleasure to see this great	Excellent	@katebros
3	Andrew is just next level humble. Never lets his status or success get to him. Somehov	Neutral	@magiko7
4	Andrew Garfield isn't American?? really??????? He is British???????? I didn't	Awful	@KitunDO
5	I love the way they dressed …. So watching friendly … and Andrew and Zen look so c	Excellent	@2456pap
6	Andrew saved her in No Way Home	Excellent	@zhougor
7	If Andrew Garfiled is 38 years old in this video, we all are, in fact, fucked	Awful	@MsOutla
8	I wanna see an actor on actor with tom holland and kirsten dunst, or tobey maguire a	Neutral	@cj-clarkj
9	the missed oppurtunity of inviting emma stone is killing me	Awful	@aleehau

Sentiments Word Cloud:

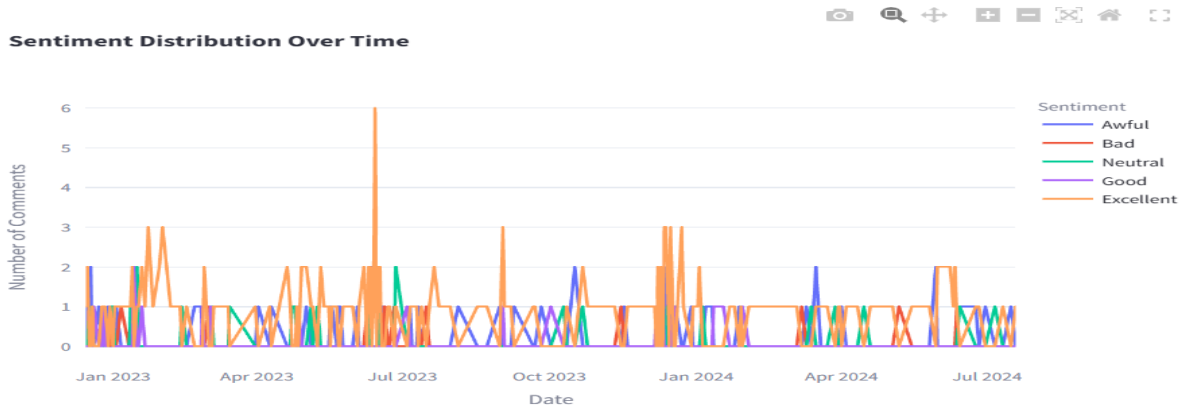
A Word Cloud representation of most used words in the comments:

Sentiment Frequency:



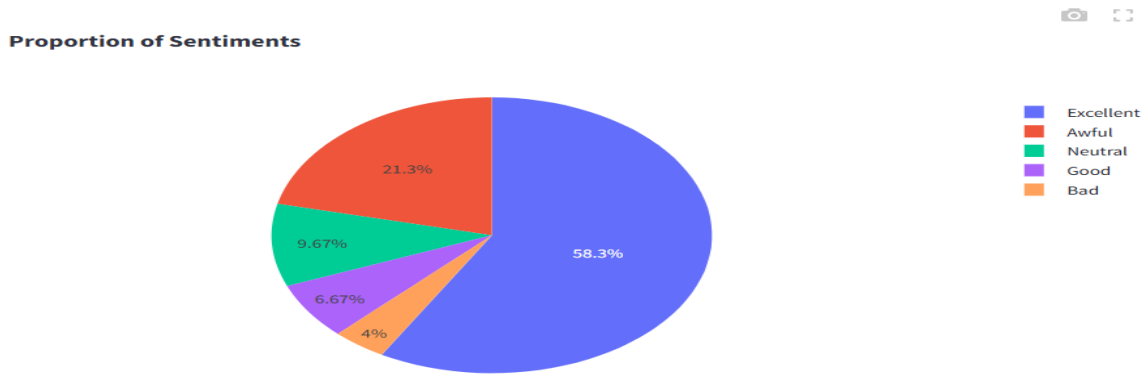
Inference: The bar chart shows the distribution of sentiments among the comments.

Sentiment Distribution Over Time:



Inference: This line graph illustrates how the distribution of sentiments evolves over time.

Proportion of Sentiments:



Inference: The pie chart shows the proportion of different sentiments among the comments.

Chapter 5

CONCLUSION

In conclusion, the "Unveiling Emotion on Social Media" project represents a significant advancement in understanding and analyzing emotional expressions within digital interactions. Through the integration of state-of-the-art machine learning techniques, particularly the BERT (Bidirectional Encoder Representations from Transformers) model, the project has provided a powerful tool for deciphering sentiment dynamics across diverse social media platforms. This application has demonstrated its capability to accurately classify sentiments from YouTube comments and user-generated text inputs, offering insights into prevailing attitudes and emotional responses in real-time.The project's user-friendly interface, coupled with functionalities like timestamp conversion to Indian Standard Time (IST) for chronological analysis, enhances usability and accessibility for researchers, marketers, and social media analysts alike. Looking forward, continued research and development in emotion detection on social media platforms hold promise for deeper insights into human behavior, societal trends, and the evolving dynamics of digital communication. As technology evolves and datasets grow, the "Unveiling Emotion on Social Media" project stands as a testament to the transformative potential of artificial intelligence in understanding and harnessing the power of emotions in the digital age.

FUTURE ENHANCEMENTS

Looking towards future enhancements for the "Unveiling Emotion on Social Media" project, several key avenues present themselves for expanding and refining its capabilities. Firstly, integrating advanced multimodal analysis techniques will allow the system to interpret emotions not just from text but also from images, videos, and audio content, providing a more comprehensive understanding of sentiment across diverse media types. Real-time sentiment monitoring is another crucial area for development, enabling immediate feedback and response to unfolding social media conversations, which is essential for proactive brand management and crisis communication strategies.

Further advancements in contextual understanding are needed to refine algorithms that can accurately interpret subtle nuances such as sarcasm, irony, and cultural references within social media posts. This will improve the accuracy and relevance of sentiment analysis results, ensuring that insights gleaned are more actionable and reflective of true sentiment. Personalization of user insights represents another promising avenue, where the system can offer tailored sentiment analysis based on individual user preferences and historical interaction patterns, enhancing engagement and user satisfaction.

As technology continues to evolve, exploring integration with emerging technologies such as blockchain for enhanced data security and AI-driven chatbots for real-time interaction and feedback collection will be beneficial. This integration can streamline data handling processes and improve the overall user experience by ensuring data integrity and security. Additionally, extending the project's capabilities to perform cross-platform sentiment analysis across multiple social media platforms simultaneously will provide a more holistic view of audience sentiment and behavior across different online channels.

REFERENCES

1.Attention Is All You Need - Vaswani, A

http://arxiv.org/abs/1706.03762

2.The Role of Emotion in Twitter: A Study of the Relationship Between Twitter Use and Trait Empathy - Shen, C., & Jiang, R. (2014). Computers in Human Behavior.

http://www.sciencedirect.com/science/article/pii/S0747563214005521

3.RoBERTa: A Robustly Optimized BERT Pretraining Approach - Liu, Y., et al. (2019). arXiv:1907.11692.

http://arxiv.org/abs/1907.11692

4.Opinion Mining and Sentiment Analysis - Pang, B., & Lee, L. (2008)

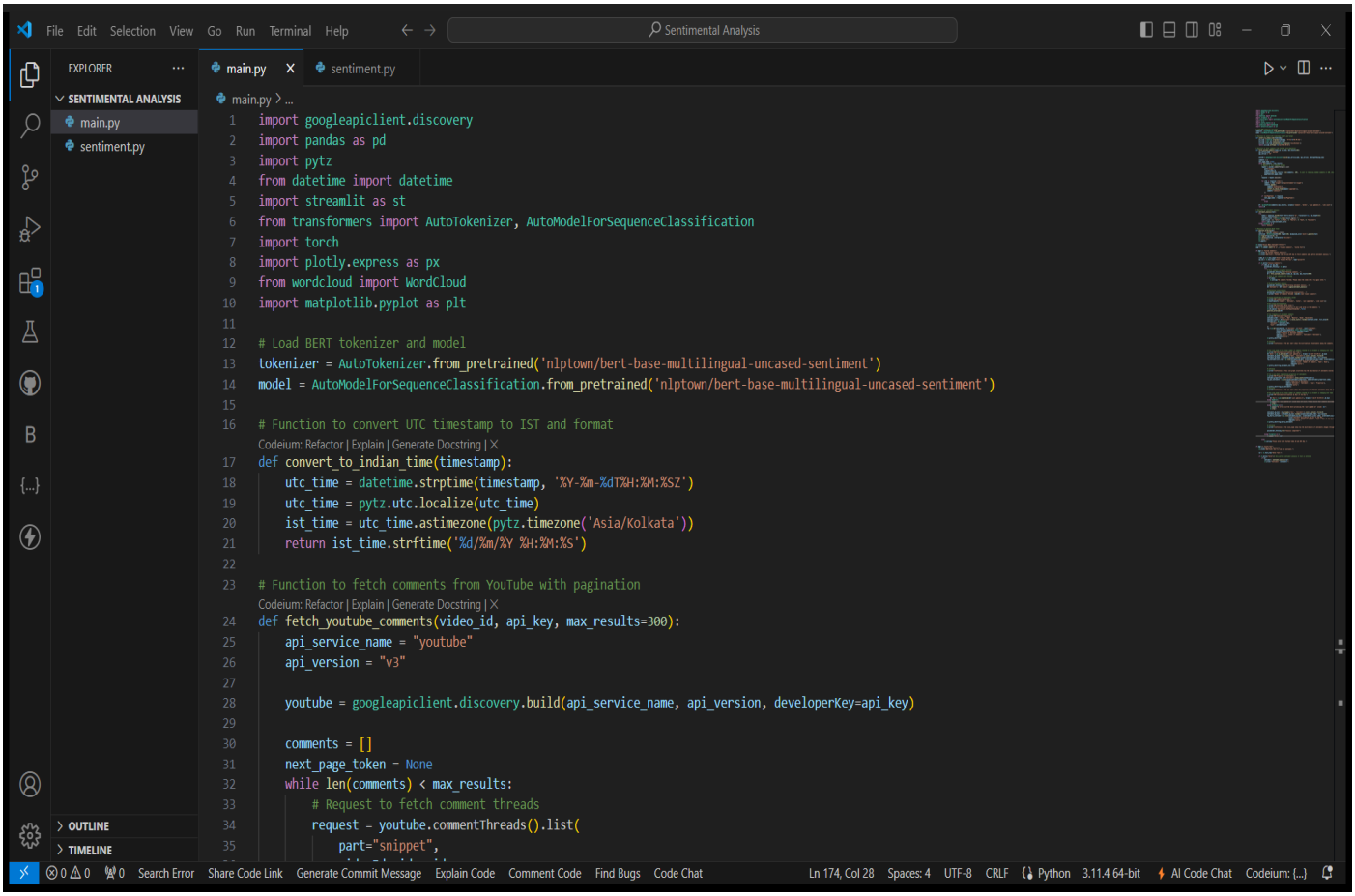
http://www.cs.cornell.edu/home/llee/omsa/omsa.pdf

5.Searching for Explanations: How Internet Evaluates Social Media Sentiment for Emotion and Social Media - Chen, E. E., et al. (2015).

http://journals.sagepub.com/doi/full/10.1177/0093650215609405

APPENDIX

Code:



```python
import googleapiclient.discovery
import pandas as pd
import pytz
from datetime import datetime
import streamlit as st
from transformers import AutoTokenizer, AutoModelForSequenceClassification
import torch
import plotly.express as px
from wordcloud import WordCloud
import matplotlib.pyplot as plt

# Load BERT tokenizer and model
tokenizer = AutoTokenizer.from_pretrained('nlptown/bert-base-multilingual-uncased-sentiment')
model = AutoModelForSequenceClassification.from_pretrained('nlptown/bert-base-multilingual-uncased-sentiment')

# Function to convert UTC timestamp to IST and format
Codeium: Refactor | Explain | Generate Docstring | X
def convert_to_indian_time(timestamp):
    utc_time = datetime.strptime(timestamp, '%Y-%m-%dT%H:%M:%SZ')
    utc_time = pytz.utc.localize(utc_time)
    ist_time = utc_time.astimezone(pytz.timezone('Asia/Kolkata'))
    return ist_time.strftime('%d/%m/%Y %H:%M:%S')

# Function to fetch comments from YouTube with pagination
Codeium: Refactor | Explain | Generate Docstring | X
def fetch_youtube_comments(video_id, api_key, max_results=300):
    api_service_name = "youtube"
    api_version = "v3"

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

    comments = []
    next_page_token = None
    while len(comments) < max_results:
        # Request to fetch comment threads
        request = youtube.commentThreads().list(
            part="snippet",
```

PROJECT SOURCE CODE / LINK:

https://github.com/sharanya-ym/sentiment-analysis