MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

Project Report

on

YouTube Comment Analyzer

Submitted By:

Sagar Patidar(0901AD211046)

Raghuveer Mewada(0901AD211040)

Faculty Mentor:

Prof. Geetika Hazra

Assistant Professor

CENTRE FOR ARTIFICIAL INTELLIGENCE

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE
GWALIOR - 474005 (MP) est. 1957

JULY-DEC.

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

CERTIFICATE

This is certified that **Sagar Patidar**(0901AD211046) , **Raghuveer Mewada** (0901AD211040) has submitted the project report titled **YouTube Comment Analyzer** under the mentorship of **Prof. Geetika Hazra** in partial fulfilment of the requirement for the award of degree of Bachelor of Technology in **Artificial Intelligence And Data Science** from Madhav Institute of Technology and Science, Gwalior.

Geetika 23/11/23

Prof. Geetika Hazra
Faculty Mentor
Assistant Professor
Centre for Artificial Intelligence

28/11/23

Dr. R. R. Singh
Coordinator
Centre for Artificial Intelligence

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

DECLARATION

I hereby declare that the work being presented in this project report, for the partial fulfilment of requirement for the award of the degree of Bachelor of Technology in **Artificial Intelligence And Data Science** at Madhav Institute of Technology & Science, Gwalior is an authenticated and original record of my work under the mentorship of **Prof. Geetika Hazra , Assistant Professor ,Centre for Artificial Intelligence**

I declare that I have not submitted the matter embodied in this report for the award of any degree or diploma anywhere else.

Sagar Patidar(0901AD211046)

Raghuveer Mewada(0901AD211040)

III Year, Centre for Artificial Intelligence

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

ACKNOWLEDGEMENT

The full semester project has proved to be pivotal to my career. I am thankful to my institute, **Madhav Institute of Technology and Science** to allow me to continue my disciplinary/interdisciplinary project as a curriculum requirement, under the provisions of the Flexible Curriculum Scheme (based on the AICTE Model Curriculum 2018), approved by the Academic Council of the institute. I extend my gratitude to the Director of the institute, **Dr. R. K. Pandit** and Dean Academics, **Dr. Manjaree Pandit** for this.

I would sincerely like to thank my department, **Centre for Artificial Intelligence,** for allowing me to explore this project. I humbly thank **Dr. R. R. Singh**, Coordinator, Centre for Artificial Intelligence, for his continued support during the course of this engagement, which eased the process and formalities involved.

I am sincerely thankful to my faculty mentors. I am grateful to the guidance of **Prof. Geetika Hazra , Assistant Professor ,Centre for Artificial Intelligence**, for his continued support and guidance throughout the project. I am also very thankful to the faculty and staff of the department.

Sagar Patidar(0901AD211046)
Raghuveer Mewada(0901AD211040)

III Year, Centre for Artificial Intelligence

ABSTRACT

In the digital era, online platforms like YouTube host an immense volume of user-generated content, with comments playing a crucial role in audience engagement. The YouTube Comment Analyzer model presented in this project aims to leverage machine learning to extract valuable insights from the vast comment sections of YouTube videos.

This model employs natural language processing (NLP) techniques to analyze sentiments, identify key topics, and evaluate the overall sentiment polarity of comments. it categorizes comments into positive, negative, or neutral sentiments, providing content creators with a deeper understanding of audience reactions.

We looked at when comments were made to see if people talk more right after a video is uploaded. We also made a picture of the most used words to know what people talk about the most. Another tool helped us know if comments were positive, negative, or neutral.

Our findings showed that people engage differently over time, especially in the first few days after a video goes live. We also learned what topics or feelings show up the most in comments. This could help people who make videos better understand what their viewers like or don't like. But, we should remember that these tools might not catch everything, so it's still important to think about what people say carefully.

In short, this project helps us understand what people say about YouTube videos. It could help video creators and others make better decisions about what content to create, keeping viewers' interests in mind.

Keyword:
YouTube Comment Analysis
Sentiment Analysis
User Engagement Insights
YouTube Comment Feedback
Comment Sentiment Classifier
Sentiment Polarity Analyzer
YouTube Analytics Model
YouTube Data API v3

सार:

डिजिटल युग में, YouTube जैसे ऑनलाइन प्लेटफ़ॉर्म में उपयोगकर्ताओं द्वारा बनाए गए सामग्री की एक अत्यधिक वॉल्यूम होती है, जिसमें टिप्पणियाँ दर्शक संवाद में महत्त्वपूर्ण भूमिका निभाती हैं। यह प्रस्तुत परियोजना में दिखाया गया YouTube Comment Analyzer मॉडल मशीन लर्निंग का उपयोग करने का उद्देश्य है, YouTube वीडियोज़ की विशाल टिप्पणी अनुभागों से मूल्यवान अंशों को निकालना।

यह मॉडल प्राकृतिक भाषा प्रसंस्करण (NLP) तकनीकों का उपयोग करता है ताकि टिप्पणियों का भावनात्मक विश्लेषण किया जा सके, महत्त्वपूर्ण विषयों की पहचान की जा सके, और टिप्पणियों की समग्र भावना को मूल्यांकित किया जा सके। यह टिप्पणियों को सकारात्मक, नकारात्मक या न्यूट्रल भावनाओं में वर्गीकृत करता है, जिससे सामग्री निर्माताओं को दर्शकों के प्रतिक्रियाओं की गहरी समझ मिल सके।

हमने देखा कि टिप्पणियाँ कब की गई थीं, इससे हमें यह जानने में मदद मिली कि क्या लोग वीडियो अपलोड होने के तुरंत बाद अधिक बात करते हैं। हमने सबसे अधिक प्रयुक्त शब्दों की एक तस्वीर बनाई ताकि हमें पता चल सके कि लोग सबसे अधिक किस बारे में बात करते हैं। एक और टूल ने हमें यह बताया कि क्या टिप्पणियाँ सकारात्मक, नकारात्मक या न्यूट्रल हैं।

हमारे अनुसंधान ने दिखाया कि लोग समय के साथ अलग-अलग ढंग से संबंधित होते हैं, खासकर वीडियो लाइव होने के पहले के कुछ दिनों में। हमने यह भी सीखा कि कौन-कौन से विषय या भावनाएँ टिप्पणियों में सबसे ज्यादा प्रकट होती हैं। यह उन लोगों की मदद कर सकता है जो वीडियो बनाते हैं ताकि वे अच्छी तरह समझ सकें कि उनके दर्शक को क्या पसंद है या क्या नहीं। लेकिन हमें याद रखना चाहिए कि ये टूल सब कुछ नहीं पकड़ सकते, इसलिए लोगों की बातों पर सोचना महत्त्वपूर्ण है।

संक्षेप में, यह परियोजना हमें यह समझने में मदद करती है कि लोग YouTube वीडियों के बारे में क्या कहते हैं। यह वीडियो निर्माताओं और अन्यों को बेहतर निर्णय लेने में मदद कर सकती है कि किस प्रकार की सामग्री बनानी चाहिए, दर्शकों की रुचियों को ध्यान में रखते हुए।

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1: Introduction

Our project, the YouTube Comment Analyzer, is like a clever friend for Creators. We are creating a smart tool that looks at what people say in the comments. We want to understand if they're feeling happy, sad, or just okay and discover what topics they like talking about the most. It's like having a helpful friend who guides video creators on what their viewers like. Our big goal is to make YouTube a more positive and enjoyable place for everyone. By understanding people's feelings and interests, we hope to bring more smiles and good vibes to the world of YouTube, making it a happier space for all of us!

1.1 Project Aim

Our project aims to improve the YouTube experience for everyone by creating a clever tool. This tool will look at the comments people leave on videos to understand if they feel happy, sad, or just okay. We want to know what topics they enjoy talking about the most. It's like having a friendly helper for video creators, guiding them on what viewers like. Our goal is to use this information to make YouTube a more positive and enjoyable community. By understanding what makes people happy, we hope to bring smiles and good vibes to the world of YouTube and make it a better place for everyone.

1.2 Objective

Our project's main objective is to create a super-smart tool for Creator. We want this tool to understand how people feel when they leave comments on videos, whether they're happy, sad, or just okay. The goal is to figure out what topics people like talking about the most. With this information, our aim is to help video creators make even better videos that everyone enjoys. We also want to assist moderators in keeping the online community a friendly and positive place. Our objective is to use smart technology to bring more joy, understanding, and good vibes to YouTube for everyone who watches and creates content.

Literature Review

Previous Work

Previous work in sentiment analysis, language translation, and YouTube Data API integration has laid the foundation for the development of a comprehensive YouTube comment analyzer. Notable studies by Choudhury and Breslin (2010) emphasized sentiment detection challenges on YouTube, using lexicons like SentiWordNet. Agrawal (2009) explored lexicon-based sentiment analysis techniques, showcasing the application of sentiment lexicons for user comment analysis. Asghar et al.'s work (2014) emphasized language-specific sentiment lexicons. In the realm of YouTube Data API, research by Siersdorfer et al. (2010) and Ni et al. (2011) demonstrated the significance of API integration for collecting and analyzing large volumes of user comments and event classification on YouTube.

Current Technology

In the current technological landscape, sentiment analysis benefits from advanced Natural Language Processing (NLP) techniques, providing more accurate classification. Custom sentiment analysis APIs from cloud providers, such as Google Cloud Natural Language API and Microsoft Azure Text Analytics, offer pre-trained models for seamless integration. Neural Machine Translation (NMT) dominates the language translation space, with models like those used by Google Translate providing contextually relevant translations. Cloud-based translation APIs, such as Google Cloud Translation API and Microsoft Azure Translator, simplify language translation integration. For YouTube Data API integration, the latest YouTube Data API v3 and YouTube Analytics API provide comprehensive features for accessing comments, conducting searches, and gaining insights into user interactions. Combining these technologies ensures the development of a robust YouTube comment analyzer, offering insights into sentiment, multilingual understanding, and real-time YouTube data analysis for content creators and analysts.

CHAPTER 2: SYSTEM SPECIFICATION

2.1 Hardware Specification

Since the hardware is an important part while developing a web project, it's necessary to find hardware requirements.

Processor–

The minimum level of the required processor for this platform is Pentium IV with 800 MHz processing speed.

Since the time taken for processing the instruction depends on the processor's power, it is very important to choose the required processor

RAM–

For a higher speed of processing, it also depends on the memory. Therefore, for better performance minimum RAM should be 2 GB.

Hard disk–

In modern days, a vast amount of data is generated daily from internet platforms. So, a good size hard disk is required for the storage of the processed data.

Cache Memory–

The access time of the operation tasks mainly depends on the cache memory. Therefore, the recommended cache memory is 1000 KB

2.2 Software Specification

Our system should meet the following minimum specifications

OS — Ubuntu, Windows, Mac OS

We will be using HTML, CSS, JAVASCRIPT, and Python and we also need an application Visual Studio Code and PyCharm as a code editor.

CHAPTER 3: METHODOLOGY

3.1 Data Collection:

For this project, data collection was done with the help of API provided by YouTube (YouTube Data API v3). The API provides access to YouTube's vast repository of videos, channels, playlists, and comments, enabling the retrieval of valuable information necessary for analyzing user comments on the platform.

- **Data Acquisition Process**
  - **API Key Generation:** Access to the YouTube Data API v3 necessitated the generation of an API key through the Google Developers Console. This unique key served as the authentication mechanism for making requests to the API endpoints.

  - **Endpoint Queries:** The API endpoints were utilized to retrieve information about videos, including their titles, descriptions, upload dates, view counts, and most importantly, the comments posted by users. Requests were made to obtain comment threads associated with selected videos based on search criteria and video identifiers.

  - **Data Filtering and Storage:** Upon receiving the API responses, the data underwent preliminary filtering to exclude irrelevant or duplicate comments. The collected data, including comment text, user details, timestamps, and engagement metrics, was stored in a structured format for further analysis.

	Author	Comment	Published At	Like Count	Reply Count
0	World Affairs by Unacademy	Use Code PD10 TO SAVE MONEY on Unacadem...	2023-08-13T08:06:45Z	144	10
1	Brajesh Pandey	Modi ji ko bana do republican Candiate	2023-08-22T14:50:15Z	0	0
2	Gangapuram Bhavani	President koi bhi ho America ko unka deep stat...	2023-08-21T17:33:29Z	0	0
3	vishalsagu3	Sir.... thumbnail ke..liye Trump ki .. kya pho...	2023-08-21T11:46:39Z	0	0
4	Arjun Raj	Vivek becoming President is something we Keral...	2023-08-21T11:15:02Z	0	0
5	Kaushik Baur	Indians in abroad hi India ka dushman hai, kam...	2023-08-21T09:04:59Z	0	0
6	Gigi mathew	Right wing my behind. If it wasn't for Mar...	2023-08-20T18:32:47Z	0	0
7	Piyush Maurya	Jai shree Krishna	2023-08-20T14:51:02Z	0	0
8	davenash	one of his policy is anti-immigrant 😂😂😂	2023-08-19T19:10:42Z	0	0
9	ALEX PRICE	Biden again will be president....	2023-08-19T13:47:25Z	0	0
10	Anjali Sharma	Literally 'gadha'😂😂😂😂	2023-08-19T09:29:14Z	0	0
11	Aditya Pradhan	Convert ka meaning kya hai😅... Sab convert ho ...	2023-08-19T09:01:22Z	0	0
12	50thAnniversary	Democrats gadhe hi hain yahan. 😅	2023-08-18T23:08:52Z	0	0
13	Clueless Lass-i Joins Internet	Only Jaby Said That In His Reaction Video, ...	2023-08-18T16:34:37Z	0	0
14	ashutosh singh	South Indian going to rule the world's every a...	2023-08-18T12:00:55Z	0	0
15	dassunder	Dont support ramaswamy	2023-08-18T11:01:59Z	0	0

Figure 3.1

3.2 Data Pre-processing:

The obtained data from the YouTube Data API v3 arrived in a formatted structure, containing relevant attributes such as comment text, user details, timestamps, and etc. As a part of the preprocessing pipeline, language translation to English was implemented to ensure uniformity and enhance the accuracy of subsequent analysis and modelling tasks.

Azure Translation Services were directly employed to translate all comments, regardless of their original language, into English. Azure Translation for Language Standardization: Leveraging Azure's translation capabilities, all comments, irrespective of their original language, were translated into English using the available APIs.

```
body = []
for i in range(len(df)):
    body.append({'text': df['comment'][i]})

request = requests.post(constructed_url, params=params, headers=headers, json=body)
response = request.json()

data = (json.dumps(response, sort_keys=True, ensure_ascii=False, indent=4, separators=(',', ': ')))
comment_translated = []
for i in range(len(response)):
    comment_translated.append(response[i]['translations'][0]['text'])

return comment_translated
```

Figure 3.2

3.3 Data Visualisation:

Visualization plays a pivotal role in comprehending patterns, trends, and insights within datasets. In this YouTube comment analysis project, visual representations were generated to elucidate temporal trends and the linguistic landscape of comments through graphical plots and word cloud representations.

3.3.1 Overall Comments Over Time

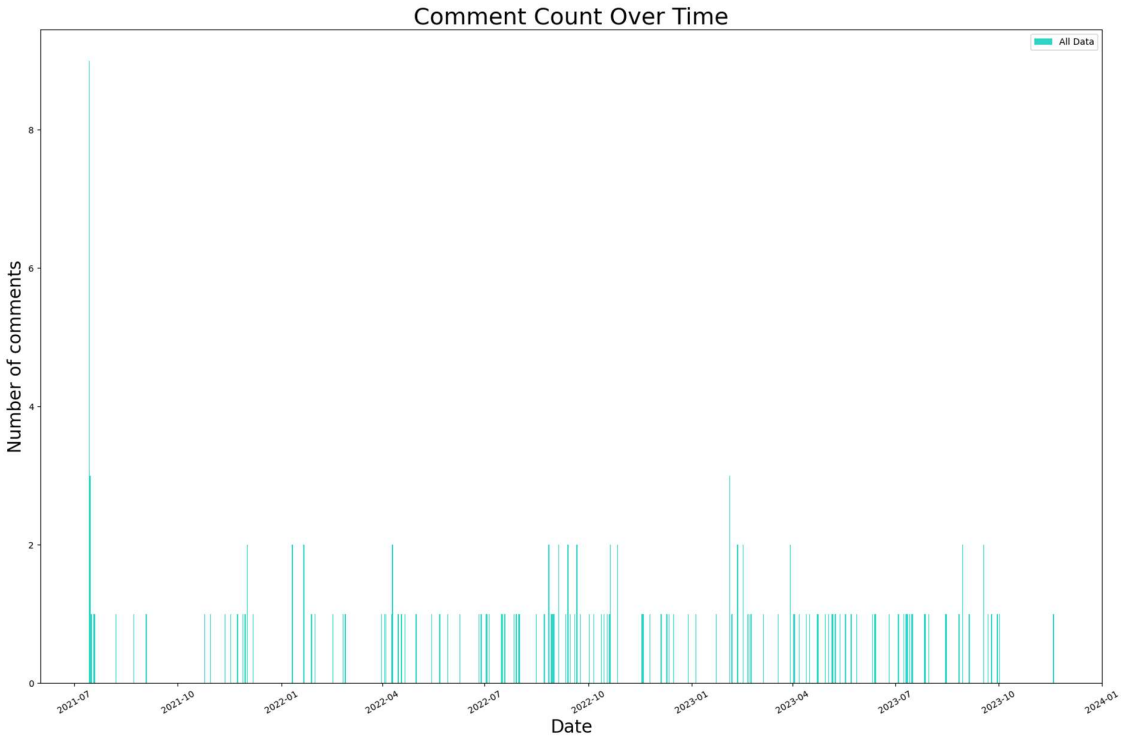

Figure 3.3.1

The first visualization illustrates the distribution of comments over time, encompassing the entire dataset duration. The x-axis represents time, segmented into relevant intervals (daily, weekly, or monthly), while the y-axis denotes the count of comments posted during those intervals. This graph provides an overview of the comment influx and potential spikes or trends observed across the entire period.

3.3.2 First 7 Days' Comments Over Time

The second visualization specifically focuses on comments within the initial seven days following video upload. Similar to the first graph, this visualization showcases the comment frequency but specifically emphasizes the comments' pattern and engagement within the crucial early period after video publication. It helps identify trends in user engagement shortly after the video's release.

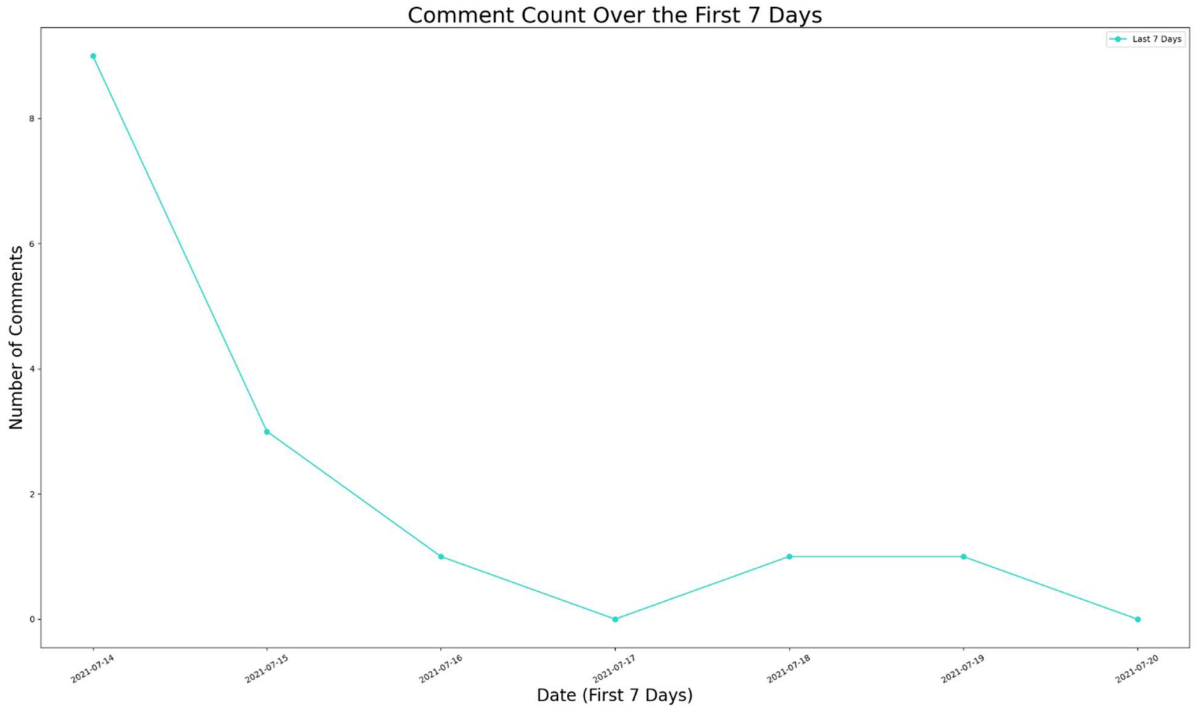

Figure 3.3.2

3.3.3 Word Cloud Representation

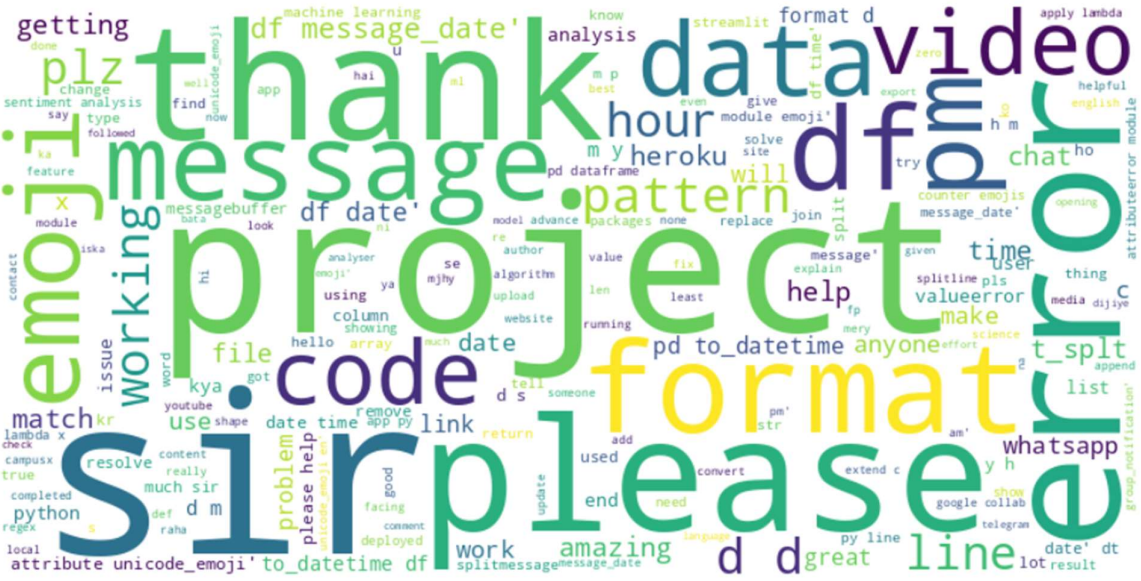

Figure 3.3.3

a word cloud representation was generated based on the comments collected. The word cloud visually emphasizes frequently occurring words in the comment corpus. Word size corresponds to the frequency of appearance in the comments, providing a quick visual snapshot of the most common topics or sentiments expressed by viewers.

3.4 Server Building:

Flask is a popular web framework that can be used to build a server for a machine-learning model. Flask is a lightweight and flexible framework that allows you to quickly create web applications in Python. Flask provides a simple way to handle HTTP requests and responses, making it easy to integrate with machine learning models.

app.py

```python
import pandas as pd
import matplotlib.pyplot as plt
from flask import Flask, render_template,request

app = Flask(__name__)

@app.route("/")
def home():
    return render_template("index.html", graph=None)

@app.route('/', methods=['POST'])
def analyze_video():
    youtube_link = request.form.get('youtube-link')
    is_valid, video_id = get_id.validate_and_extract_video_id(youtube_link)
    if is_valid:
        df = comment.get_comments(video_id)
        statistics = comment.get_statistics(video_id)
        translate_df = HtoHE.translate(df)
        positive, negative, neutral, p_score, n_score, ne_score = textAnalysis.analysis(translate_df)
        textData = {
            "positive": positive,
            "negative": negative,
            "neutral": neutral,
            "p_score": p_score,
            "n_score": n_score,
            "ne_score": ne_score
        }
        # print(textData)

        graph_all, graph_last_7_days , cloud = generate_graph(df)

        return render_template('index.html' ,
                               graph=graph_all,graph_7 =graph_last_7_days,
                               statistics=statistics, video_id = video_id,
                               cloud=cloud,textData = textData ,error=False)
```

Figure 3.4

3.5 Model Used: Azure Text Analytics for Comment Analysis

For the analysis of comments gathered from YouTube using the YouTube Data API v3, an AI-based model from Azure Text Analytics was utilized. Azure Text Analytics provides a suite of natural language processing tools that enable sentiment analysis, key phrase extraction, language detection, and other text analytics functionalities, which were instrumental in deriving insights from the comment dataset.

```
from azure.ai.textanalytics import TextAnalyticsClient
from azure.core.credentials import AzureKeyCredential

def analysis(df):

    endpoint = "https://sentimentcomment.cognitiveservices.azure.com/"
    text_analytics_client = TextAnalyticsClient(endpoint=endpoint, credential=AzureKeyCredential(key))

    documents = df[0:10]
    response = text_analytics_client.analyze_sentiment(documents=documents)

    for i in range(len(df)//10):
        documents = df[i*10:(i+1)*10]
        response = text_analytics_client.analyze_sentiment(documents=documents)

        for document in response:
            # print("Document Sentiment: {}".format(document.sentiment))
            if document.sentiment == "positive":
                positive += 1
            elif document.sentiment == "negative":
                negative += 1
            else:
                neutral += 1

            p_score += document.confidence_scores.positive
            n_score += document.confidence_scores.negative
            ne_score += document.confidence_scores.neutral

    return positive, negative, neutral, p_score/len(df), n_score/len(df), ne_score/len(df)
```

Figure 3.5

3.6 User Interface:

HTML, CSS, and JavaScript can be used to build a user interface for a machine-learning model. HTML provides the structure of the page, CSS provides the styling and layout, and JavaScript provides the interactivity.

HTML:

```
<body>
  <div id="left-section">
    <h3>YouTube Comment Analyzer</h3>
    <form action="/" method="post">
      <input type="text" name="youtube-link" placeholder="YouTube Video Link" id="youtube-link" />
      <button type="submit">Analyse</button>
    </form>
    {% if statistics %}
    <iframe id="video_embed" width="300" height="200" src="https://www.youtube.com/embed/{{ video_id }}"></iframe>
    <p class="video_name">{{ statistics.video_name }}</p>
    <h4 id="channel_name">{{ statistics.channel_name }}</h4>
    {% elif error %}
    <p class="error">Please enter valid YouTube Video URL</p>
    {% endif %}
  </div>
  <div id="right-section">
    <div id="graph-container">
      {% if graph %}
      <h1 class="heading">Statistics</h1>
      <div id="statistics">
        <div class="count view">
          <span class="stat_text">View</span>
          <span>{{ statistics.view_count }}</span>
        </div>
        <div class="count like">
          <span class="stat_text">Like</span>
          <span>{{ statistics.like_count }}</span>
        </div>
        <div class="count comment">
          <span class="stat_text">Comment</span>
          <span>{{ statistics.comment_count }}</span>
        </div>
        <div class="count published_at">
          <span class="stat_text">Published At</span>
          <span>{{ statistics.published_at|formatdatetime }}</span>
```

3.6.1

CSS:

```
body {
    font-family: Arial, sans-serif;
    margin: 0;
    padding: 0;
    display: flex;
    height: 100vh;
}

#left-section {
    width: 20%;
    background-color: #323341;
    color: white;
    height: 100%;
    padding: 20px;
    position: fixed;
}

#right-section {
    flex: 1;
    padding: 20px;
    background: #101118;
    overflow-y: auto;
    /* Add scroll to the right section */
    min-height: 100vh;
    height: 100%;
    margin-left: 22%;
}
```

Figure 3.6.2

JavaScript:

```
<script>
  function changeGraph() {
    var selectedValue = document.getElementById('graph-type').value

    if (selectedValue === 'complete') {
      document.querySelector('.graph.graph2').classList.remove('none')
      document.querySelector('.graph.graph1').classList.add('none')
    } else if (selectedValue === 'seven-day') {
      document.querySelector('.graph.graph1').classList.remove('none')
      document.querySelector('.graph.graph2').classList.add('none')
    }
  }
</script>
```

Figure 3.6.3

Final UI:

Figure 3.6.4

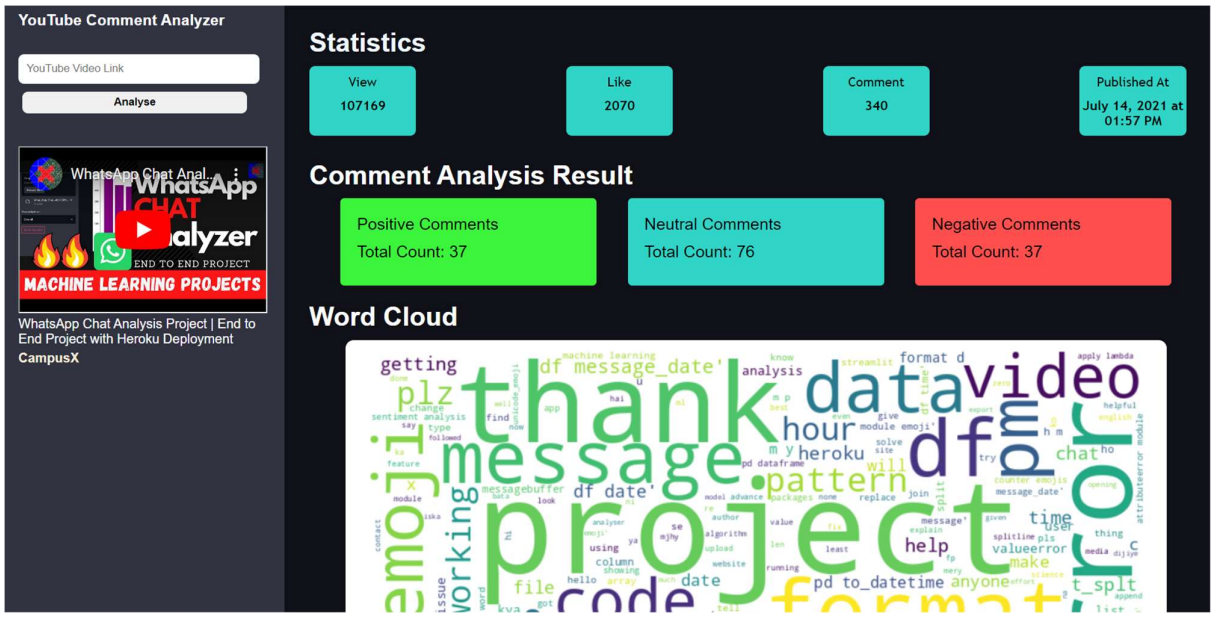

Figure 3.65

3.7 Application of the YouTube Comment Analyzer Model:

- Content Optimization
- Community Engagement
- Feedback Analysis
- Trend Identification
- User Experience Enhancement

3.8 Problems Faced during Model Development

- **Manage Data Quantity**

  The huge volume of comments available for analysis give a significant challenge due to resource limitations, including computational power, storage capacity, and processing time. Analyzing a massive number of comments could overwhelm available resources and hinder timely analysis and model development**.**

Limiting the scope of comment analysis to a specific subset or a set number of comments becomes a pragmatic strategy. By strategically sampling or setting a cap on the number of comments analyzed, the project can focus on a representative subset rather than attempting to process the entire vast volume.

3.9 Limitations of the YouTube Comment Analyzer Model

- **API Limitations and Financial Constraints:**

  One limitation encountered during the project was the restriction imposed by the usage limits of the free access tier of the API service. As the project progressed, the free tier's limitations, including the restricted number of API requests or access to premium features, hindered the ability to scale or access advanced functionalities. Due to budget constraints, upgrading to a paid API subscription for expanded access or higher usage limits became unfeasible for the project's scope or financial resources. This constraint affected the frequency of data retrieval, analysis capabilities, and access to enhanced features, ultimately impacting the depth and breadth of insights derived from the YouTube comment dataset.

- **Handling Large-Scale Data:**

  The time required for processing extensive volumes of data posed a significant limitation during the project. Analysing the substantial dataset obtained from YouTube comments using various natural language processing (NLP) techniques demanded considerable computational resources and time.

  As the dataset size increased, the processing time grew significantly, impacting tasks such as sentiment analysis, language translation, and key phrase extraction. The scalability of the computational infrastructure became a limiting factor in handling larger datasets effectively within reasonable time frames.

CHAPTER 5: TECHNIQUES AND TOOLS

Technologies and tools refer to the various software, hardware, and other resources that are used to accomplish a specific task or achieve a particular goal. They are often used interchangeably, but generally, technologies refer to the broader range of resources and methods used to achieve a specific outcome, while tools refer to specific software applications, hardware devices, or other resources used to carry out specific tasks.

4.1 Technologies used

- **Python**

  Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy-to learn syntax emphasizes readability and therefore reduces the cost of program maintenance.

- **Libraries**

  - **Pandas**

    Panda is an open-source library designed primarily for working quickly and logically with relational or labeled data. It offers a range of data structures and procedures for working with time series and numerical data. The NumPy library serves as the foundation for this library. Pandas is quick and offers its users exceptional performance & productivity.

  - **NumPy**

    NumPy is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete

Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more.

- **Matplotlib**

  Matplotlib is a widely used Python library for creating visualizations, including charts, graphs, and plots. It provides a wide range of tools for data visualization, making it a popular choice in the field of machine learning (ML). Matplotlib supports a wide range of plots, including line plots, scatter plots, bar plots, histograms, and more.

- **Backend or Server**

  Flask is a popular Python web framework that allows developers to build web applications easily and quickly. It is a lightweight and flexible framework that provides essential tools and features for building web applications. Flask is based on the WSGI (Web Server Gateway Interface) standard and is compatible with most web servers and platforms.

  Flask provides a range of features for web application development, including routing, templating, session management, and security. It also supports the use of extensions, which allows developers to add additional functionality to their web applications, such as database integration, authentication, and more.

- **API**
  - **YouTube Data API V3**

    The YouTube Data API v3 is a powerful tool provided by YouTube to access and interact with their vast repository of video content. It allows developers to retrieve information such as video details, playlists, channels, and Comment. This API enables the integration of YouTube data into applications, websites.

  - **Azure Text Analytics**

    Azure Text Analytics is a service provided by Microsoft Azure that leverages natural language processing (NLP) capabilities to extract valuable insights from unstructured text data. It is designed to analyze and understand the sentiment, key phrases, and language of text documents, making it a powerful tool for text-based data processing.

Sentiment Analysis: Azure Text Analytics can determine the sentiment expressed in a piece of text, classifying it as positive, negative, or neutral. This feature is useful for understanding the emotional tone of user reviews, social media posts, and other textual content.

- **Azure Text Translation**

  Azure Text Translation is a service provided by Microsoft Azure that facilitates language translation within applications, websites, and other solutions. This service leverages advanced machine learning and neural machine translation technologies to provide accurate and context-aware translations for various languages.

- **Frontend or User Interface**

  - **HTML**

    HTML (Hypertext Markup Language) is a markup language used for creating web pages and other information that can be displayed in a web browser. HTML is the backbone of the World Wide Web and is used to structure and display content such as text, images, and multimedia on websites.HTML consists of a series of elements that define the structure of a web page.

  - **CSS**

    CSS (Cascading Style Sheets) is a style sheet language used for describing the presentation of a document written in HTML (Hypertext Markup Language). CSS is used to define the visual appearance of a web page, including its layout, colours, fonts, and other styling options.CSS works by separating the content of a web page from its presentation

  - **JavaScript**

    JavaScript (JS) is a programming language used to create dynamic and interactive web pages. It is widely used in web development for creating client-side scripts that run in web browsers, and server-side scripts that run on web servers. JavaScript is a high-level, object-oriented programming language that supports eventdriven, functional, and imperative programming styles.

4.2 Tools used

- **PyCharm**

  PyCharm is an integrated development environment (IDE) used for Python programming. PyCharm provides a range of features designed to help developers write and debug Python code more efficiently. Some of its key features include code completion, code analysis, debugging, and version control integration. PyCharm also provides a range of tools for web development, including support for popular web frameworks such as Django and Flask

- **VS code**

  Visual Studio Code, often abbreviated as VS Code, is a free, open-source code editor developed by Microsoft. It is designed for building and debugging modern web and cloud applications and supports a wide range of programming languages, including Python, JavaScript, and TypeScript. VS Code provides a range of features, including syntax highlighting, code completion, and debugging tools

CHAPTER 5 : CONCLUSION

In conclusion, our YouTube Comment Analyzer project has been a journey to make YouTube a friendlier and more enjoyable space. We've learned to understand people's feelings in the comments and discovered the hot topics they love to talk about. With this information, video creators can now make content that everyone loves. Our aim is to be a helpful tool, like a wise friend, guiding creators and keeping the YouTube community positive and happy.

As our project wraps up, it's not just about computers and data – it's about creating a place where everyone on YouTube feels good, smiles more, and enjoys the experience. The adventure to make YouTube even better doesn't stop here; it continues with the hope of spreading joy and good vibes throughout the platform.

Overall, our YouTube Comment Analyzer is like a helpful friend for YouTube. We've learned to understand if people are happy or not in the comments and discovered the favorite topics they like to talk about. This means video creators can now make better videos that everyone enjoys.

References

- https://flask.palletsprojects.com/en/3.0.x/
- https://docs.python.org/3/
- https://matplotlib.org/stable/index.html
- https://developers.google.com/youtube/v3/docs
- https://learn.microsoft.com/en-us/python/api/azure-ai-textanalytics/?view=azure-python
- https://learn.microsoft.com/enus/python/api/azurecore/azure.core.credentials?view=azure-python
- https://learn.microsoft.com/en-us/azure/ai-services/translator/
- https://stackoverflow.com/
- https://pypi.org/project/wordcloud/