**Fr. Conceicao Rodrigues College of Engineering** Father Agnel Ashram, Bandstand, Bandra –west, Mumbai-50 

**Department of Computer Engineering**

**SOCIAL MEDIA ANALYTICS LAB**

**Experiment No: 2**

**Aim:** Data Collection-Select the social media platforms of your choice (Twitter, Facebook, LinkedIn, Youtube, Web blogs etc), connect to and capture social media data for business (scraping, crawling,parsing).

**Objective:** To capture social media data and perform sentiment analysis from business perspective.

**Lab outcomes:**

*At the end of this lab session, students will be able to…*

1. Acquire hands on skills needed to work with social media data.

**Theory:**

● Crawling: Involves going though specific websites and related links, more like going through a collection of things and inspecting them. This is the first stage of scraping and Parsing. basically, visiting and going through a site could be termed crawling that is if being done by a bot; web crawler.

● Scraping: It is a form of copying, in which specific data is gathered and copied from the *web*, typically into a specified storage location.

● Parsing: This involves breaking down the above scraped data into smaller bits of it, this is to aid understanding of the scraped data. This stage is employed in various fields of data extraction and mining.

**Steps**

● **Step1:**

○ Get API Key - Go to the Google Cloud console.

(https://console.cloud.google.com/) and sign in with your Google Account ○ Click the project drop-down menu in the top bar and select or create the project you want to use. ○ Click the hamburger menu in the top left and select APIs & Services >Select Youtube Data V3 API> Credentials.

Academic Year- 23-24

**Fr. Conceicao Rodrigues College of Engineering** Father Agnel Ashram, Bandstand, Bandra –west, Mumbai-50 

**Department of Computer Engineering**

**SOCIAL MEDIA ANALYTICS LAB**

○ Click the Create credentials button and select API key.

○ The API key will be displayed in a pop-up window. You can click the RESTRICT KEY button to restrict the API keys usage, such as by IP address or referrer.

○ Click the COPY button to copy the API key to your clipboard.

■ You can use the API key in your application to access the Google Cloud APIs. Be sure to keep the API key confidential, as it can be used to access your Google Cloud resources. ● Step2

○ Search Youtube Video for scraping comments.

○ Copy the Youtube Video Id

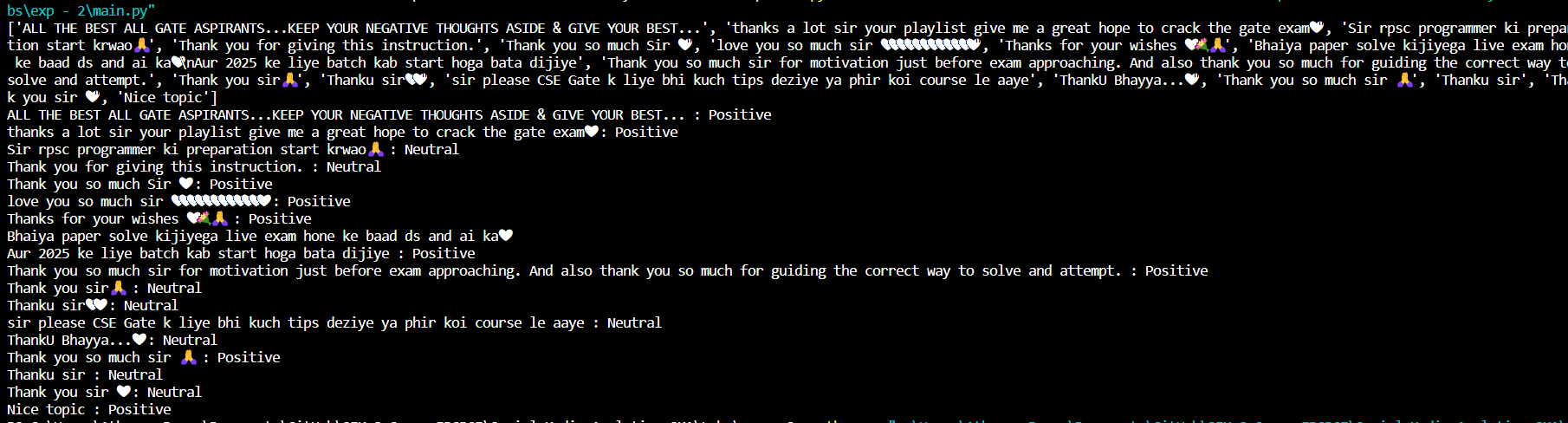
**Student’s Task**

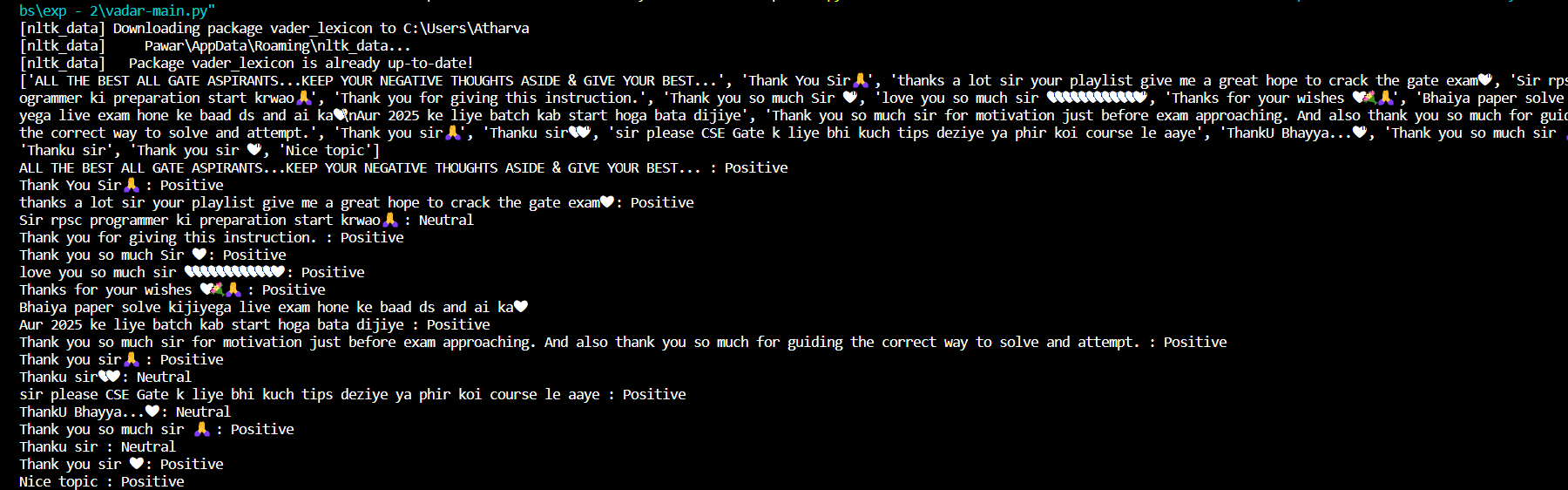
1. Select youtube video link of your choice and write a code to print comments of that video.



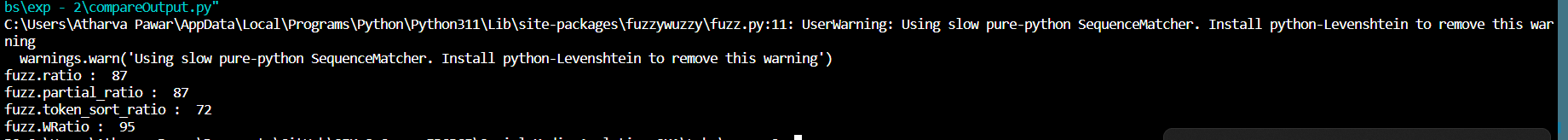


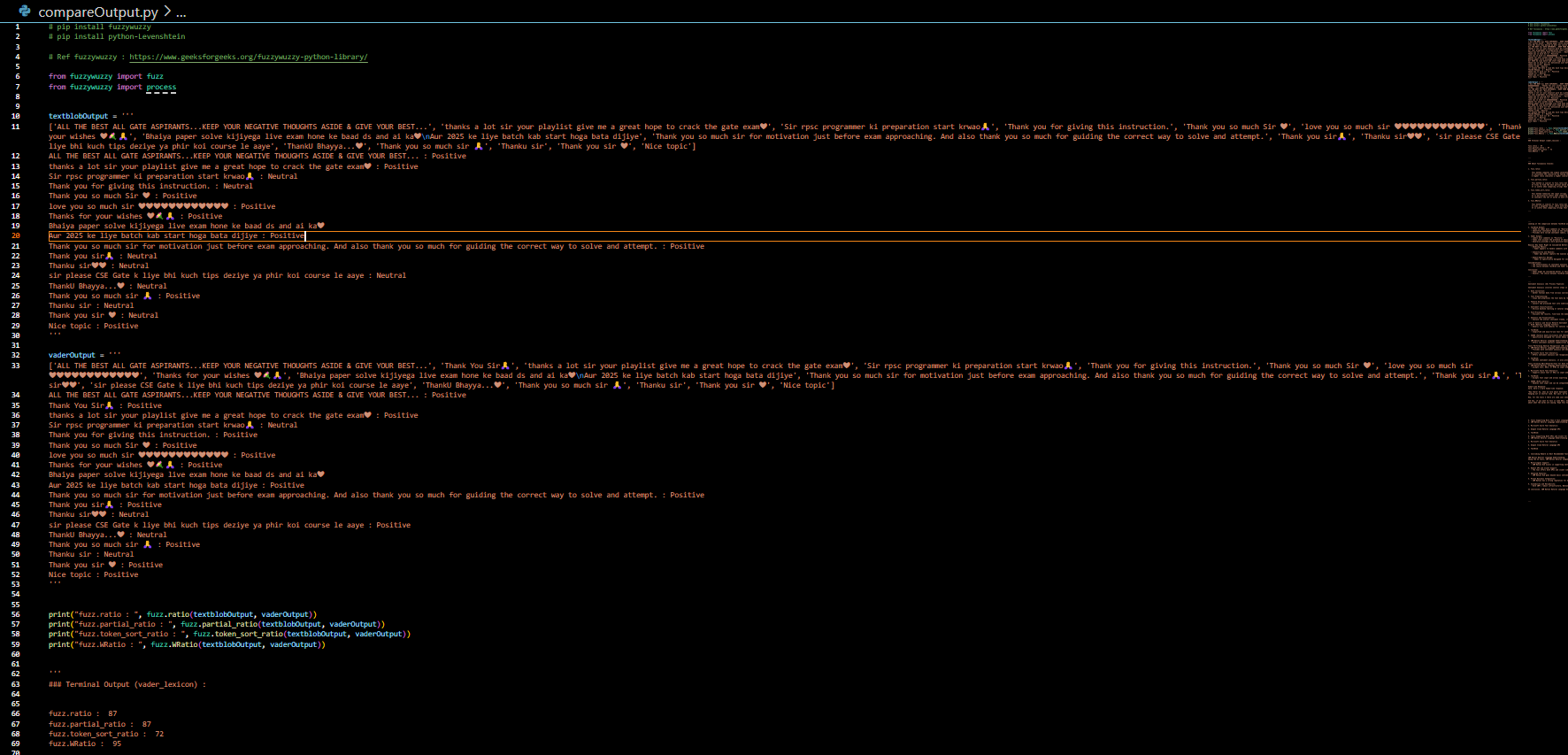
1. Write a code to perform the sentiment analysis of the comments of the selected video using Textblob and Vader.





1. Write your remarks on the results of analysis done in qno 2.





1. TextBlob Output:

- Tends to label more comments as "Neutral."

- Identifies sentiments based on the overall tone of the text without considering emojis explicitly.

- Provides more varied sentiment labels, including "Neutral," "Positive," and "Negative."

2. Vader Output:

- Labels most comments as "Positive."

- Takes into account the presence of emojis and assigns a positive sentiment if positive emojis are present.

- Generally provides a more positive-leaning sentiment classification.

Reasons Why Vader is Better:

Handling Emojis:

- Vader appears to handle comments with emojis better. It recognizes positive sentiments associated with emojis like ❤ and 💐 and assigns an overall positive sentiment.

Subjectivity and Nuances:

- Vader may better capture the nuances and subjectivity in comments, especially in mixed sentiment expressions. For example, comments like "sir please CSE Gate k liye bhi kuch tips deziye ya phir koi course le aaye" are labeled as "Positive" by Vader, possibly acknowledging the positive intent.

Domain-Specific Design:

- Vader is specifically designed for social media text, where sentiments can be conveyed through unconventional language, emoticons, and slangs. This makes it well-suited for sentiment analysis in platforms like YouTube.

**4. Article discussion**

**Title: Sentiment analysis for mining texts and social networks data: Methods and tools.** Zucco, Chiara, et al. "Sentiment analysis for mining texts and social networks data: Methods and tools." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10.1 (2020): e1333.

**Answer the following questions based on above article.**

1. **Explain the Sentiment Analysis (SA) process pipeline**.

1. Data Collection:

- Gather textual data from various sources, such as social media posts, comments, reviews, or other text-based content.

2. Text Preprocessing:

- Clean and preprocess the text data by removing noise, stopwords, and irrelevant information. Tokenization and stemming may also be applied.

3. Feature Extraction:

- Convert the processed text into numerical features using techniques like TF-IDF (Term Frequency-Inverse Document Frequency) or word embeddings.

4. Sentiment Classification:

- Utilize machine learning or natural language processing models to classify the sentiment of each piece of text into categories like positive, negative, or neutral.

5. Post-Processing:

- Evaluate the results, fine-tune the model, and handle any misclassifications. Post-processing may include adjusting for context or considering sentiment intensity.

6. Analysis and Visualization:

- Analyze the overall sentiment trends, create visualizations, and extract meaningful insights from the sentiment-labeled data.

1. **List the generic and Social Network Sentiment Analysis Tools which are free to use.**

1. NLTK (Natural Language Toolkit)

2. TextBlob

3. VADER (Valence Aware Dictionary and sEntiment Reasoner)

4. IBM Watson Natural Language Understanding

1. **List the tools that perform ER (Emotion Recognition) along with Sentiment Analysis in English**.

1. IBM Watson Natural Language Understanding

2. Microsoft Azure Text Analytics

3. TextBlob

**4. List the tools that allows URLs/ keywords/text/ as option for data entering for analysis and allows Jason/ csv as data export .**

1. IBM Watson Natural Language Understanding

2. Microsoft Azure Text Analytics

3. TextBlob

4. VADER (NLTK Library)

5. **List the tools that support more than 5 text languages for the analysis.**

1. IBM Watson Natural Language Understanding

2. Microsoft Azure Text Analytics

3. Google Cloud Natural Language API

4. TextBlob

**6. List the tools that support both APIs and client for the most used programming languages**.

1. IBM Watson Natural Language Understanding

2. Microsoft Azure Text Analytics

3. Google Cloud Natural Language API

4. TextBlob

**7. Write the concluding remark on the tool (among 24) that is best recommended for business analysis. Justify the same.**

IBM Watson Natural Language Understanding

1. Multilingual Support:

- IBM Watson NLU excels in supporting sentiment analysis for a diverse range of languages. This is crucial for businesses operating globally, ensuring a comprehensive understanding of customer sentiments across different regions.

2. Robust API and Client Support:

- The tool offers both APIs and client libraries for popular programming languages, making it highly versatile and easily integrable into existing business applications. This flexibility caters to diverse development environments and the preferences of developers.

3. Advanced Features:

- IBM Watson NLU goes beyond basic sentiment analysis. It provides additional features like emotion recognition, entity recognition, and concept tagging, offering a more nuanced understanding of textual data.

4. Proven Business Integration:

- IBM Watson has a strong reputation for business applications and has been successfully integrated into various industries, including finance, healthcare, and customer service. This track record makes it a reliable choice for businesses seeking actionable insights from textual data.

5. Scalability and Reliability:

- With IBM's robust infrastructure, Watson NLU ensures scalability and reliability, crucial factors for businesses dealing with large volumes of textual data.