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.

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

○ 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.

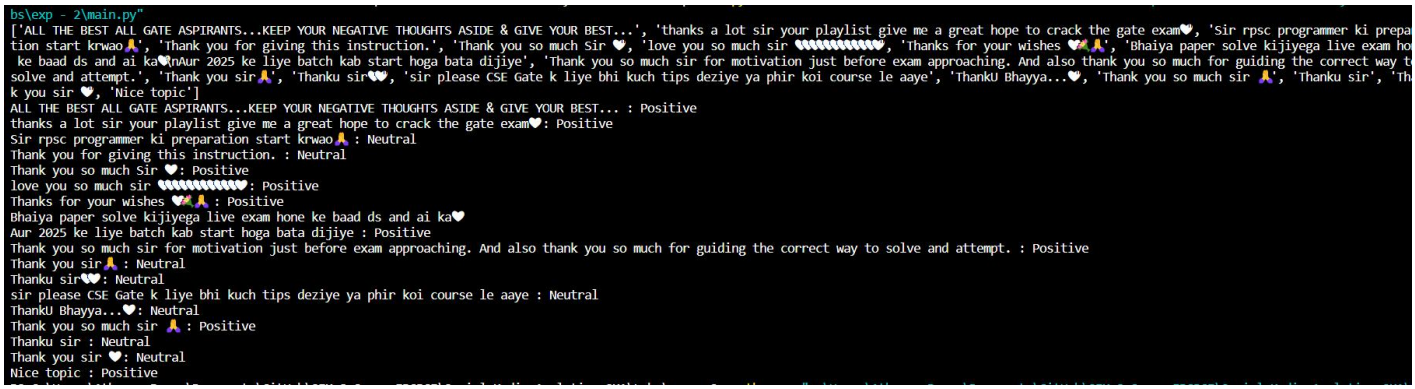
xt-blobmain.py > ...

```python
import requests
from textblob import TextBlob
import pandas as pd

video_id = '9NQSnAOtrbw'
api_key = "apikey"

# Retrieve video information
video_info_url = f"https://www.googleapis.com/youtube/v3/videos?part=snippet&id={video_id}&key={api_key}"
video_info_response = requests.get(video_info_url)
video_info_data = video_info_response.json()
video_info_data

# Retrieve video comments
comments_url = f"https://www.googleapis.com/youtube/v3/commentThreads?part=snippet&videoId={video_id}&key={api_key}"
comments_response = requests.get(comments_url)
comments_data = comments_response.json()

# Extract the comments
comments = [item["snippet"]["topLevelComment"]["snippet"]["textOriginal"] for item in comments_data['items']]
print(comments)

def get_comment_sentiment(comment):
    analysis = TextBlob(comment)
    if analysis.sentiment.polarity > 0:
        return "Positive"
    elif analysis.sentiment.polarity == 0:
        return "Neutral"
    else:
        return "Negative"

comment_list = []
sentiment_list = []

for comment in comments:
    sentiment = get_comment_sentiment(comment)
    comment_list.append(comment)
    sentiment_list.append(sentiment)
    print(f"{comment} : {sentiment}")

sentiment_df = pd.DataFrame({"Comments": comment_list, "Sentiment": sentiment_list})
sentiment_df.head()
sentiment_df.to_csv("YouTube_Comments_Sentiment.csv")
```

vader-main.py > ...

```python
import requests
import pandas as pd
from nltk.sentiment.vader import SentimentIntensityAnalyzer

# Download the Vader lexicon for sentiment analysis
import nltk
nltk.download('vader_lexicon')

video_id = '9NQSnAOtrbw'
api_key = "apikey"

# Retrieve video comments
comments_url = f"https://www.googleapis.com/youtube/v3/commentThreads?part=snippet&videoId={video_id}&key={api_key}"
comments_response = requests.get(comments_url)
comments_data = comments_response.json()

# Extract the comments
comments = [item["snippet"]["topLevelComment"]["snippet"]["textOriginal"] for item in comments_data['items']]
print(comments)

# Sentiment analysis using Vader
analyzer = SentimentIntensityAnalyzer()

comment_list = []
compound_scores = []

for comment in comments:
    sentiment_scores = analyzer.polarity_scores(comment)
    compound_score = sentiment_scores['compound']

    comment_list.append(comment)
    compound_scores.append(compound_score)

    sentiment = 'Positive' if compound_score >= 0.05 else ('Neutral' if -0.05 < compound_score < 0.05 else 'Negative')
    print(f"{comment} : {sentiment}")

sentiment_df = pd.DataFrame({"Comments": comment_list, "Compound Score": compound_scores})
sentiment_df.head()
sentiment_df.to_csv("YouTube_Comments_Sentiment_Vader.csv")
```

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



```
bs\exp - 2\main.py"
['ALL THE BEST ALL GATE ASPIRANTS...KEEP YOUR NEGATIVE THOUGHTS ASIDE & GIVE YOUR BEST...', 'thanks a lot sir your playlist give me a great hope to crack the gate exam❤', 'Sir rpsc programmer ki prepa
tion start krwao🙏', 'Thank you for giving this instruction.', 'Thank you so much Sir ❤', 'love you so much sir ❤❤❤❤❤❤❤❤❤❤', 'Thanks for your wishes 💞🙏', 'Bhaiya paper solve kijiyega live exam ho
 ke baad ds and ai ka❤\nAur 2025 ke liye batch kab start hoga bata dijiye', 'Thank you so much sir for motivation just before exam approaching. And also thank you so much for guiding the correct way t
solve and attempt.', 'Thank you sir🙏', 'Thanku sir❤❤', 'sir please CSE Gate k liye bhi kuch tips deziye ya phir koi course le aaye', 'ThankU Bhayya...❤', 'Thank you so much sir 🙏', 'Thanku sir', 'Th
k you sir ❤', 'Nice topic']
ALL THE BEST ALL GATE ASPIRANTS...KEEP YOUR NEGATIVE THOUGHTS ASIDE & GIVE YOUR BEST... : Positive
thanks a lot sir your playlist give me a great hope to crack the gate exam❤: Positive
Sir rpsc programmer ki preparation start krwao🙏 : Neutral
Thank you for giving this instruction. : Neutral
Thank you so much Sir ❤: Positive
love you so much sir ❤❤❤❤❤❤❤❤❤❤: Positive
Thanks for your wishes 💞🙏 : Positive
Bhaiya paper solve kijiyega live exam hone ke baad ds and ai ka❤
Aur 2025 ke liye batch kab start hoga bata dijiye : Positive
Thank you so much sir for motivation just before exam approaching. And also thank you so much for guiding the correct way to solve and attempt. : Positive
Thank you sir🙏 : Neutral
Thanku sir❤❤: Neutral
sir please CSE Gate k liye bhi kuch tips deziye ya phir koi course le aaye : Neutral
ThankU Bhayya...❤: Neutral
Thank you so much sir 🙏 : Positive
Thanku sir : Neutral
Thank you sir ❤: Neutral
Nice topic : Positive
```

```
bs\exp - 2\vadar-main.py"
[nltk_data] Downloading package vader_lexicon to C:\Users\Atharva
[nltk_data]     Pawar\AppData\Roaming\nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
['ALL THE BEST ALL GATE ASPIRANTS...KEEP YOUR NEGATIVE THOUGHTS ASIDE & GIVE YOUR BEST...', 'Thank You Sir🙏', 'thanks a lot sir your playlist give me a great hope to crack the gate exam🖤', 'Sir rps
ogrammer ki preparation start krwao🙏', 'Thank you for giving this instruction.', 'Thank you so much Sir 🖤', 'love you so much sir 🖤🖤🖤🖤🖤🖤🖤🖤🖤🖤', 'Thanks for your wishes 💖🙏', 'Bhaiya paper solve
yega live exam hone ke baad ds and ai ka🖤\nAur 2025 ke liye batch kab start hoga bata dijiye', 'Thank you so much sir for motivation just before exam approaching. And also thank you so much for guid
the correct way to solve and attempt.', 'Thank you sir🙏', 'Thanku sir🖤', 'sir please CSE Gate k liye bhi kuch tips deziye ya phir koi course le aaye', 'ThankU Bhayya...🖤', 'Thank you so much sir
'Thanku sir', 'Thank you sir 🖤', 'Nice topic']
ALL THE BEST ALL GATE ASPIRANTS...KEEP YOUR NEGATIVE THOUGHTS ASIDE & GIVE YOUR BEST... : Positive
Thank You Sir🙏 : Positive
thanks a lot sir your playlist give me a great hope to crack the gate exam🖤: Positive
Sir rpsc programmer ki preparation start krwao🙏 : Neutral
Thank you for giving this instruction. : Positive
Thank you so much Sir 🖤: Positive
love you so much sir 🖤🖤🖤🖤🖤🖤🖤🖤🖤🖤: Positive
Thanks for your wishes 💖🙏 : Positive
Bhaiya paper solve kijiyega live exam hone ke baad ds and ai ka🖤
Aur 2025 ke liye batch kab start hoga bata dijiye : Positive
Thank you so much sir for motivation just before exam approaching. And also thank you so much for guiding the correct way to solve and attempt. : Positive
Thank you sir🙏 : Positive
Thanku sir🖤: Neutral
sir please CSE Gate k liye bhi kuch tips deziye ya phir koi course le aaye : Positive
ThankU Bhayya...🖤: Neutral
Thank you so much sir 🙏 : Positive
Thanku sir : Neutral
Thank you sir 🖤: Positive
Nice topic : Positive
```

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

```
bs\exp - 2\compareOutput.py"
C:\Users\Atharva Pawar\AppData\Local\Programs\Python\Python311\Lib\site-packages\fuzzywuzzy\fuzz.py:11: UserWarning: Using slow pure-python SequenceMatcher. Install python-Levenshtein to remove this warning
  warnings.warn('Using slow pure-python SequenceMatcher. Install python-Levenshtein to remove this warning')
fuzz.ratio :  87
fuzz.partial_ratio :  87
fuzz.token_sort_ratio :  72
fuzz.WRatio :  95
```

compareOutput.py > ...

```
# pip install fuzzywuzzy
# pip install python-Levenshtein

# Ref fuzzywuzzy : https://www.geeksforgeeks.org/fuzzywuzzy-python-library/

from fuzzywuzzy import fuzz
from fuzzywuzzy import process


textblobOutput = '''
['ALL THE BEST ALL GATE ASPIRANTS...KEEP YOUR NEGATIVE THOUGHTS ASIDE & GIVE YOUR BEST...', 'thanks a lot sir your playlist give me a great hope to crack the gate exam❤', 'Sir rpsc programmer ki preparation start krwao🙏', 'Thank you for giving this instruction.', 'Thank you so much Sir ❤', 'love you so much sir ❤❤❤❤❤❤❤❤❤❤❤', 'Thank your wishes ❤🙏🙏', 'Bhaiya paper solve kijiyega live exam hone ke baad ds and ai ka❤\nAur 2025 ke liye batch kab start hoga bata dijiye', 'Thank you so much sir for motivation just before exam approaching. And also thank you so much for guiding the correct way to solve and attempt.', 'Thank you sir🙏', 'Thanku sir❤❤', 'sir please CSE Gate kliye bhi kuch tips deziye ya phir koi course le aaye', 'ThankU Bhayya...❤', 'Thank you so much sir 🙏', 'Thanku sir', 'Thank you sir ❤', 'Nice topic']
ALL THE BEST ALL GATE ASPIRANTS...KEEP YOUR NEGATIVE THOUGHTS ASIDE & GIVE YOUR BEST... : Positive
thanks a lot sir your playlist give me a great hope to crack the gate exam❤ : Positive
Sir rpsc programmer ki preparation start krwao🙏 : Neutral
Thank you for giving this instruction. : Neutral
Thank you so much Sir ❤ : Positive
love you so much sir ❤❤❤❤❤❤❤❤❤❤❤ : Positive
Thanks for your wishes ❤🙏🙏 : Positive
Bhaiya paper solve kijiyega live exam hone ke baad ds and ai ka❤
Aur 2025 ke liye batch kab start hoga bata dijiye : Positive
Thank you so much sir for motivation just before exam approaching. And also thank you so much for guiding the correct way to solve and attempt. : Positive
Thank you sir🙏 : Neutral
Thanku sir❤❤ : Neutral
sir please CSE Gate k liye bhi kuch tips deziye ya phir koi course le aaye : Neutral
ThankU Bhayya...❤ : Neutral
Thank you so much sir 🙏 : Positive
Thanku sir : Neutral
Thank you sir ❤ : Neutral
Nice topic : Positive
'''

vaderOutput = '''
['ALL THE BEST ALL GATE ASPIRANTS...KEEP YOUR NEGATIVE THOUGHTS ASIDE & GIVE YOUR BEST...', 'Thank You Sir🙏', 'thanks a lot sir your playlist give me a great hope to crack the gate exam❤', 'Sir rpsc programmer ki preparation start krwao🙏', 'Thank you for giving this instruction.', 'Thank you so much Sir ❤', 'love you so much sir ❤❤❤❤❤❤❤❤❤❤❤', 'Thanks for your wishes ❤🙏🙏', 'Bhaiya paper solve kijiyega live exam hone ke baad ds and ai ka❤\nAur 2025 ke liye batch kab start hoga bata dijiye', 'Thank you so much sir for motivation just before exam approaching. And also thank you so much for guiding the correct way to solve and attempt.', 'Thank you sir🙏', 'Thanku sir❤❤', 'sir please CSE Gate k liye bhi kuch tips deziye ya phir koi course le aaye', 'ThankU Bhayya...❤', 'Thank you so much sir 🙏', 'Thanku sir', 'Thank you sir ❤', 'Nice topic']
ALL THE BEST ALL GATE ASPIRANTS...KEEP YOUR NEGATIVE THOUGHTS ASIDE & GIVE YOUR BEST... : Positive
Thank You Sir🙏 : Positive
thanks a lot sir your playlist give me a great hope to crack the gate exam❤ : Positive
Sir rpsc programmer ki preparation start krwao🙏 : Neutral
Thank you for giving this instruction. : Positive
Thank you so much Sir ❤ : Positive
love you so much sir ❤❤❤❤❤❤❤❤❤❤❤ : Positive
Thanks for your wishes ❤🙏🙏 : Positive
Bhaiya paper solve kijiyega live exam hone ke baad ds and ai ka❤
Aur 2025 ke liye batch kab start hoga bata dijiye : Positive
Thank you so much sir for motivation just before exam approaching. And also thank you so much for guiding the correct way to solve and attempt. : Positive
Thank you sir🙏 : Positive
Thanku sir❤❤ : Neutral
sir please CSE Gate k liye bhi kuch tips deziye ya phir koi course le aaye : Positive
ThankU Bhayya...❤ : Neutral
Thank you so much sir 🙏 : Positive
Thanku sir : Neutral
Thank you sir ❤ : Positive
Nice topic : Positive
'''


print("fuzz.ratio : ", fuzz.ratio(textblobOutput, vaderOutput))
print("fuzz.partial_ratio : ", fuzz.partial_ratio(textblobOutput, vaderOutput))
print("fuzz.token_sort_ratio : ", fuzz.token_sort_ratio(textblobOutput, vaderOutput))
print("fuzz.WRatio : ", fuzz.WRatio(textblobOutput, vaderOutput))


'''
### Terminal Output (vader_lexicon) :


fuzz.ratio :  87
fuzz.partial_ratio :  87
fuzz.token_sort_ratio :  72
fuzz.WRatio :  95
```

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.

2. 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

3. **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.