

▼ IT 350

Assignment 2 - Data Analytics and Visualization of Live Stream Data

Name: Niraj Nandish

Roll no.: 191IT234

Batch no.: 7

Link to Colab notebook - <https://colab.research.google.com/drive/1ub5QHB3Dicz7rVMxKYNrrWn1QsSB74zz?usp=sharing>

```
!pip install google-api-python-client google-auth-oauthlib google-auth-httplib2 google-api-python-client hatesonar
```

```
Requirement already satisfied: google-api-python-client in /usr/local/lib/python3.7/dist-packages (1.12.10)
Requirement already satisfied: google-auth-oauthlib in /usr/local/lib/python3.7/dist-packages (0.4.6)
Requirement already satisfied: google-auth-httplib2 in /usr/local/lib/python3.7/dist-packages (0.0.4)
Requirement already satisfied: hatesonar in /usr/local/lib/python3.7/dist-packages (0.0.7)
Requirement already satisfied: scikit-learn==0.22.2.post1 in /usr/local/lib/python3.7/dist-packages (0.22.2.p
Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (3.2.2)
Requirement already satisfied: scipy>=0.17.0 in /usr/local/lib/python3.7/dist-packages (from scikit-learn==0.
Requirement already satisfied: joblib>=0.11 in /usr/local/lib/python3.7/dist-packages (from scikit-learn==0.2
Requirement already satisfied: numpy>=1.11.0 in /usr/local/lib/python3.7/dist-packages (from scikit-learn==0.
Requirement already satisfied: httplib2<1dev,>=0.15.0 in /usr/local/lib/python3.7/dist-packages (from google-
Requirement already satisfied: six<2dev,>=1.13.0 in /usr/local/lib/python3.7/dist-packages (from google-api-p
Requirement already satisfied: google-auth<3dev,>=1.16.0 in /usr/local/lib/python3.7/dist-packages (from goog
Requirement already satisfied: uritemplate<4dev,>=3.0.0 in /usr/local/lib/python3.7/dist-packages (from googl
Requirement already satisfied: google-api-core<3dev,>=1.21.0 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: packaging>=14.3 in /usr/local/lib/python3.7/dist-packages (from google-api-cor
Requirement already satisfied: protobuf>=3.12.0 in /usr/local/lib/python3.7/dist-packages (from google-api-co
Requirement already satisfied: pytz in /usr/local/lib/python3.7/dist-packages (from google-api-core<3dev,>=1.
Requirement already satisfied: requests<3.0.0dev,>=2.18.0 in /usr/local/lib/python3.7/dist-packages (from goc
Requirement already satisfied: googleapis-common-protos<2.0dev,>=1.6.0 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: setuptools>=40.3.0 in /usr/local/lib/python3.7/dist-packages (from google-api-
Requirement already satisfied: cachetools<5.0,>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from google-
Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.7/dist-packages (from google-a
```

```

Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.7/dist-packages (from google-auth<3dev
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from packa
Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in /usr/local/lib/python3.7/dist-packages (from pyasn1-mo
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from requests<3.
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from requests<3.0
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests<3.0.0dev
Requirement already satisfied: requests-oauthlib>=0.7.0 in /usr/local/lib/python3.7/dist-packages (from googl
Requirement already satisfied: oauthlib>=3.0.0 in /usr/local/lib/python3.7/dist-packages (from requests-oauth
Requirement already satisfied: pandas>=0.22.0 in /usr/local/lib/python3.7/dist-packages (from hatesonar) (1.3
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dist-packages (from pandas>
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib)
Requirement already satisfied: cyclor>=0.10 in /usr/local/lib/python3.7/dist-packages (from matplotlib) (0.11

```

1. Use the allocated data source (as per your batch) and build a utility to extract and curate
- data for the analytics tasks specified in part 2. Ensure that this module provides a Live Stream of data.

```

import googleapiclient.discovery

api_service_name = "youtube"
api_version = "v3"
DEVELOPER_KEY = "AIzaSyAiJYSZn_aVFt9Cplqnx54O7LuHD2MqZR4"
videoID = 'IsBc_Vbt1Uw'
youtube = googleapiclient.discovery.build(api_service_name, api_version, developerKey=DEVELOPER_KEY)
request = youtube.commentThreads().list(part="snippet,replies", maxResults=100, videoId=videoID, textFormat="plain
response = request.execute()
comment_list={}
next_page_token = response['nextPageToken']

for item in response['items']:
    comment_list[item['snippet']['topLevelComment']['id']] = item['snippet']['topLevelComment']['snippet']['textDi
    if 'replies' in item:
        for reply in item['replies']['comments']:
            comment_list[reply['id']] = reply['snippet']['textDisplay']

```

```

while next_page_token:
    request = youtube.commentThreads().list(
        part="snippet,replies",
        maxResults=100,
        videoId=videoID,
        textFormat="plainText",
        pageToken=next_page_token
    )
    response = request.execute()
    next_page_token = ''
    if "nextPageToken" in response:
        next_page_token = response['nextPageToken']
    for item in response['items']:
        comment_list[item['snippet']['topLevelComment']['id']] = item['snippet']['topLevelComment']['snippet']['textDisplay']
        if 'replies' in item:
            for reply in item['replies']['comments']:
                comment_list[reply['id']] = reply['snippet']['textDisplay']

```

2. Using the data collected as part of question 1, perform a data analytics task as per your batch allotment (Use an existing library if available for this part)

```

from hatesonar import Sonar
from pprint import pprint

sonar = Sonar()
hate_count = {
    "hate_speech": 0,
    "offensive_language": 0,
    "neither": 0,
}
hate_arr = []
offensive_arr = []
for commentID in comment_list:
    sonar_res = sonar.ping(text=comment_list[commentID])
    hate_count[sonar_res["top_class"]] += 1

```

```

hate_arr.append(sonar_res["classes"][0]["confidence"])
offensive_arr.append(sonar_res["classes"][1]["confidence"])

pprint(hate_count)

/usr/local/lib/python3.7/dist-packages/sklearn/utils/deprecation.py:144: FutureWarning: The sklearn.linear_mo
warnings.warn(message, FutureWarning)
/usr/local/lib/python3.7/dist-packages/sklearn/base.py:318: UserWarning: Trying to unpickle estimator Logisti
UserWarning)
/usr/local/lib/python3.7/dist-packages/sklearn/base.py:318: UserWarning: Trying to unpickle estimator TfidfTr
UserWarning)
/usr/local/lib/python3.7/dist-packages/sklearn/base.py:318: UserWarning: Trying to unpickle estimator TfidfVe
UserWarning)
{'hate_speech': 33, 'neither': 12562, 'offensive_language': 666}

```

3. Finally, build a visualization module for data obtained from the task carried out in part 2 so that any changes in the stream are reflected in the visualization

```

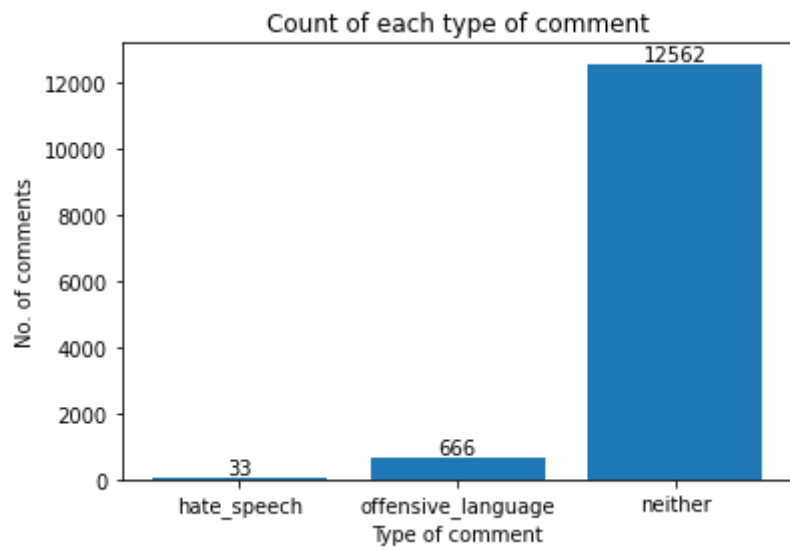
import matplotlib.pyplot as plt

plt.bar(list(hate_count.keys()), list(hate_count.values()))
for (i, v) in enumerate(hate_count.values()):
    plt.text(i, v, str(v), ha='center', va='bottom')
plt.title('Count of each type of comment')
plt.xlabel("Type of comment")
plt.ylabel("No. of comments")
plt.show()

plt.plot([i for i in range(100)], hate_arr[:100], label='Hate Speech')
plt.plot([i for i in range(100)], offensive_arr[:100], label='Offensive language')
plt.title('Hate speech and Offensive language confidence score of first 100 comments')
plt.ylabel("Confidence Score")
plt.legend(bbox_to_anchor=(1, 1))
plt.show()

```





Hate speech and Offensive language confidence score of first 100 comments

