AML-3204 Social Media Analytics

Assignment (20%)

Analyzing YouTube Data

In this assignment, you will write a Python script to download user interaction data from YouTube for each "youtubeId" provided in the "vdoLinks.csv" file. For each of the YouTube video id, you will extract the following information:

1. 100 comments
2. Description of the video
3. View Count
4. Like Count
5. Dislike Count
6. Comment Count
7. Duration of the video
8. Favorite Count

While extracting the data, if any of the "youtubeId" does not work, please ignore that ID and proceed with the following "youtubeId".

After the data extraction is done, then please find out the following:

1. A list of the top-10 videos based on the total views
2. A list of the bottom-10 videos based on the total views
3. The most liked video
4. The least liked video
5. The video with the highest duration
6. Apply sentiment analysis on the downloaded comments for each of the videos. For sentiment analysis, you can use VADER or any other tool of your choice. Your program should list the sentiments score for each of the movies

Data from YouTube can be extracted in many ways. You can use any of the available approaches. However, we have practiced one method for the task in our Lab session that might be handy.

Submission:

To submit, please do the following:

Tasks (70 Marks):

1. Write a short report on the assignment.
   The report should have the following:
       a. Data collection procedure
       b. Data cleansing steps
       c. Analysis Steps
       d. Using a bar diagram, show the top-10 videos based on total views

e. Using a bar diagram, show the bottom-10 videos based on total views (NB. In the bottom-10 list, do not include the videos that have zero views)
f. The title of the video that has the most likes
g. The title of the video that has the least likes
h. The title of the video that has the highest duration
i. Using a bar diagram, show the top-10 videos that have the highest positive sentiment scores, which you calculated from the comments
j. Using a bar diagram, show the bottom-10 videos that have the highest negative sentiment scores, which you calculated from the comments

Note: Use appropreate additional visualization techniques to present your findings if required.

Please upload the report and the Python file

Presentation (30 Marks)

Presentation time: 15 minutes

Presentation is a two-step procedure:

Step-1: Present your slides describing your work

Step-2: Demonstrate your python solution