Assignment 4

Course: ISA 414

Instructor: Dr. Arthur Carvalho

Points:100

Due date: October 25th, 2022, before 11:59 pm

Submission instructions: this assignment is to be done individually. All your answers should be in a single Jupyter notebook or python file. Your code must be well formulated (*i.e.*, no errors) and sound (*i.e.*, it does what the question asks it to do). In particular, the grader must be able to open your submission using VS Code and run the code without running into errors. Code with errors may receive zero points. Submit the final document on Canvas before the due date.

Description: suppose you got a new job (you are on fire!) and now work for a social-media analytics company. Your first task is to develop a script that monitors users' posts on Twitter. Specifically, your script must: 1) collect tweets every 10 seconds associated with a specific hashtag; and 2) estimate in real-time the overall sentiment behind all the never-before-collected tweets. A company using this script will then be able to react quickly in case users are posting too many negative tweets about the company. Using the template on the following page, solve these tasks:

a) Connect to the Twitter API and create a MongoDB database called "*Asst4*" and a collection called "*SocialMedia*." You are highly encouraged to obtain your own Twitter API keys. **[5 points]**

b) Write some code that continuously collects five tweets associated with a hashtag. For testing purposes, it is advisable to use a unique hashtag, *e.g.*, #ISA414_*YourMiamiID*. **[5 points]**

c) For each of the five collected tweets, determine whether that tweet was collected before. This step may require some research. [**20 points**]

d) For each tweet never before seen, do the following:

1. Apply the VADER sentiment-analysis model (see Lecture 15) and retrieve only the "*compound*" score [**20 points**];
2. Print the words "*New Tweet*: " followed by the tweet's text, followed by the word "*Score*: " followed by the calculated sentiment score, followed by the word *"Time*: *"* followed by the time the tweet was posted/created **[20 points]**;
3. Add the sentiment score to the variable cumulative_score [**5 points**];

4. For the sake of future auditings, store the tweet's text, score, and when the tweet was posted as three key-value pairs inside a MongoDB document. [**20 points**]

e) After all the above, print a message to inform the user about the current cumulative sentiment score and put your code to sleep for 10 seconds to avoid using the API too often. **[5 points]**

Solution Template

```
# import the required modules
from time import sleep
…

# connect to the Twitter API and create your MongoDB connection
…

# define relevant variables
cumulative_score = 0
…

# start an infinite loop to collect tweets continuously
while True:

    # retrieve five tweets
    search_results  = …

    # for each tweet in your search
    for tweet in search_results:

        # verify whether tweet was not collected previously
        if … :

            # calculate the sentiment score associated with tweet
            score = …

            # print the text in tweet, the score, and the time tweet was posted
            …

            # update cumulative_score
            …

            # store collected data inside MongoDB
            …

    # print current cumulative_score
    …

    # sleeping for 10 seconds to avoid too many Twitter queries
    sleep(10)
```