MIDASS@IIITD Summer Internship Task

GENERAL INSTRUCTIONS

Failing to follow any of the instructions below will lead to rejection of your submission.

1. This task has 3 problems
    a. A python problem
    b. A Natural Language Processing (NLP) problem
    c. A Computer Vision (CV) problem
2. The **Python problem is mandatory** for everyone. Out of **NLP and CV tasks, you have to submit any one** of them.
3. You can submit solutions to all 3 problems if you want. However, no extra time or points will be rewarded.
4. You have to submit the **link of a GitHub repository and solution CSV file** via https://goo.gl/forms/9wm6IhKrmvi4tEfV2. (Specific detail given in task.)
5. The Github Repository should follow the following structure. Instructions about what should be in these directories are given in problem statements.

```
/root
|
|_/Pyhton_problem
|
|_/(CV/NLP)_Problem
```

6. We recommend committing regularly on Github while developing.
7. You have **10 days** to submit the solutions. No extensions will be provided.
8. Code of each problem should be in a Jupyter Notebook.
9. Make sure your codes are properly documented. We recommend the following:-
    a. Before each code block have a markdown block which mentions the following
        i. What the code block is doing.
        ii. What is your intuition behind doing this? Why do you think it is useful?

    b. For bigger code blocks also add comments in the block.
    c. Keep everything in this notebook things which worked, things which did not work. This notebook should be a snapshot of the process you follow to solve a Data problem.
10. You are not allowed to add any new data samples from an online source or self-labeled to the provided dataset (Just to make it clear, data augmentation techniques while learning is allowed, adding completely new samples from any external sources is not allowed).
11. If you have any doubts feel free to email at midas@iiitd.ac.in or hitkuli@iiitd.ac.in

Python Problem

You have to write a python script which can fetch all the tweets(as many as allowed by Twitter API) done by midas@IIITD twitter handle and dump the responses into JSONlines file.

The other part of your script should be able to parse these JSONline files to display the following for every tweet in a tabular format.

- The text of the tweet.
- Date and time of the tweet.
- The number of favorites/likes.
- The number of retweets.
- Number of Images present in Tweet. If no image returns None.

Submission

Submission of this task should just include the **code in Jupyter notebook**. Make sure you **remove Twitter API keys** from the solution while uploading them to GitHub. We would not be responsible for any misuse in case you accidentally make them public or share them with us.

Computer Vision Problem

You can download train and test dataset from here. This dataset has 4 classes. Labels for training data are provided, you have to submit labels of test data. Feel free to use any Machine learning or Deep learning technique.

Submission

In the form, you would have to submit the CSV file of test labels. CSV file should be named as *your_full_name.csv* each row of this file should be the following:-

- *Test_image_index, predicted class*

Dataset folder has a sample submission file. Strictly follow that format.

Your GitHub repo should have code in Jupyter notebook.

NLP Problem

Please refer to https://competitions.codalab.org/competitions/19955#learn_the_details-data-specifications for the problem statement. The dataset for the task is available at https://github.com/Semeval2019Task9/Subtask-A. You have to follow all the rules as mentioned in the problem.

Submission

In the form, you would have to submit the CSV file of test labels. CSV file should be named as *your_full_name.csv*. Submission CSV should follow the format given on the dataset GitHub link (https://github.com/Semeval2019Task9/Subtask-A/blob/master/SubtaskA_Trial_Test.csv)

NOTE - Submission CSV file format has to be the same as defined in the dataset repository. However, submission filename should still be *your_full_name.csv*, and not *submission.csv* as suggested in the repository.

Your GitHub repo should have code in Jupyter notebook.

We advise you to follow our Twitter page https://twitter.com/midasIIITD and Like our Facebook page https://www.facebook.com/midasIIITD to get all updates including status of your results, timelines, and etc.