COMP-9318 Final Project

Instructions:

1. This note book contains instructions for **COMP9318 Final-Project**.
2. You are required to complete your implementation in a file `submission.py` provided along with this notebook.
3. You are not allowed to print out unnecessary stuff. We will not consider any output printed out on the screen. All results should be returned in appropriate data structures returned by corresponding functions.
4. This notebook encompasses all the requisite details regarding the project. Detailed instructions including **CONSTRAINTS**, **FEEDBACK** and **EVALUATION** are provided in respective sections. In case of additional problem, you can post your query @ Piazza.
5. This project is **time-consuming**, so it is highly advised that you start working on this as early as possible.
6. You are allowed to use only the permitted libraries and modules (as mentioned in the **CONSTRAINTS** section). You should not import unnecessary modules/libraries, failing to import such modules at test time will lead to errors.
7. You are **NOT ALLOWED** to use dictionaries and/or external data resources for this project.
8. We will provide you **LIMITED FEEDBACK** for your submission (only **15** attempts allowed to each group). Instructions for the **FEEDBACK** and final submission are given in the **SUBMISSION** section.
9. For **Final Evaluation** we will be using a different dataset, so your final scores may vary.
10. Submission deadline for this assignment is **23:59:59 on 27-May, 2018**.
11. **Late Penalty: 10-% on day-1 and 20% on each subsequent day.**

Introduction:

In this Project, you are required to devise an algorithm/technique to fool a binary classifier named `target-classifier`. In this regard, you only have access to following information:

1. The `target-classifier` is a binary classifier classifying data to two categories, *i.e.*, **class-1** and **class-0**.
2. You have access to part of classifiers' training data, *i.e.*, a sample of 540 paragraphs. 180 for **class-1**, and 360 for **class-0**, provided in the files: `class-1.txt` and `class-0.txt` respectively.
3. The `target-classifier` belong to the SVM family.
4. The `target-classifier` allows **EXACTLY 20 DISTINCT** modifications in each test sample.
5. You are provided with a test sample of **200** paragraphs from **class-1** (in the file: `test_data.txt`). You can use these test samples to get feedback from the target classifier (**only 15 attempts** allowed to each group.).
6. **NOTE: You are not allowed to use the data `test_data.txt` for your model training (if any). VIOLATIONS in this regard will get ZERO score**.

-to-do:

- You are required to come up with an algorithm named `fool_classifier()` that makes best use of the above-mentioned information (**point 1-4**) to fool the `target-classifier`. By fooling the classifier we mean that your algorithm can help mis-classify a bunch of test instances (**point-5**) with minimal possible modifications (**EXACTLY 20 DISTINCT** modifications allowed to each test sample).
- **NOTE:: We put a harsh limit** on the number of modifications allowed for each test instance. You are only allowed to modify each test sample by **EXACTLY 20 DISTINCT tokens (NO MORE NO LESS)**.
- **NOTE:: ADDING** or **DELETING** one word at a time is **ONE** modification. Replacement will be considered as **TWO** modifications (*i.e.,* **Deletion** followed by **Insertion**).

Constraints

Your implementation `submission.py` should comply with following constraints.

1. You should implement your methodology using `Python3`.
2. You should implement your code in the function `fool_classifier()` in the file `submission.py`.
3. You are only allowed to use pre-defined class `strategy()` defined in the file: `helper.py` in order to train your models (if any).
4. You **should not** do any pre-processing on the data. We have already pre-processed the data for you.
5. You are supposed to implement your algorithm using **scikit-learn (version=0.19.1)**. We will **NOT** accept implementations using other Libraries.
6. You are **not supposed to augment** the data using external/additional resources. You are only allowed to use the partial training data provided to you (*i.e.*, `class-1.txt` and `class-0.txt`).
7. You are **not** allowed to use the test samples (*i.e.*, `test_data.txt`) for model training and/or inference building. You can only use this data for testing, *i.e.*, calculating success %-age (as described in the **EVALUATION** section.). **VIOLATIONS IN THIS REGARD WILL GET ZERO SCORE**.
8. You are **not** allowed to hard code the ground truth and any other information into your implementation `submission.py`.
9. Considering the **RUNNING TIME**, your implementation is supposed to read the test data file (*i.e.*, `test_data.txt` with 200 test samples), process it and write the modified file (`modified_data.txt`) within **12 Minutes**.
10. Each modified test sample in the modified file (`modified_data.txt`) should not differ from the original test sample corresponding to the file (`test_data.txt`) by more than 20 tokens.
11. **NOTE::** Inserting or Deleting a word is **ONE** modification. Replacement will be considered as **TWO** modifications (*i.e.*, deletion followed by insertion).

Implementation Details

1. In the file `submission.py`, you are required to implement a function named: `fool_classifier()` that reads a text file named: `test_data.txt` from Present Working Directory(PWD), and writes out the modified text file: `modified_data.txt` in the same directory.
2. We have provided the implementation of **strategy** class in a seperate file `helper.py`. You are supposed to use this class for your model training (if any) and inference building.
3. **Detailed description of input and/or output parts is given below:**

Input:

- The function `fool_classifier()` reads a text files named `test_data.txt` having almost (500-1500) test samples. Each line in the input file corresponds to a single test sample.
- **Note:** We will also provide the partial training data ((i) `class-0.txt` and (ii) `class-1.txt`) in the test environment. You can access this data using the class: `strategy()`.

Output:

- You are supposed to write down the modified file named `modified_data.txt` in the same directory, and in the same format as that of the `test_data.txt`. In addition, your program is supposed to return the instance of the `strategy` class defined in `helper.py`.

- **Note:** Please make sure that the file: `modified_data.txt` is generated by your code, and it follows the **MODIFICATION RESTRICTIONS (ADD** and/or **DELETE EXACTLY 20 DISTINCT TOKENS)**. In case of **ERRORS**, we will **NOT** allow more feedback attempts.

EVALUATION:

1. For evaluation, we will consider a bunch of test paragraphs having:
   - Approximately 500-1500 test samples for class-1, with each line corresponding to a distinct test sample.The input test file will follow the same format as that of `test_data.txt`.
   - We will consider the success rate of your algorithm for final evaluation. By success rate we mean %-age of samples miss-classified by the `target-classifier` (*i.e.,* instances of `class-1`, classified as `class-0` after `20` distinct modifications).

Example: ¶

1. Consider 200 test-samples (classified as **class-1** by the `target-classifier`).
2. For-Example, after modifying each test sample by (**20 DISTINCT TOKENS**) the `target-classifier` mis-classifies **100** test samples (*i.e.,* 100 test samples are classified as **class-0** then your **success %-age** is:
3. **success %-age** = (100) x 100/200 = **50%**