

Bonus Assignment #1

General Instructions:

1. **There is NO PARTIAL credit for the bonus assignment submission that has partial/incomplete code.**
2. All three parts (**Part I**, **Part II**, and **Part III**) of this Assignment must be completed successfully to earn FULL credit for this Assignment.
3. You must collect data of the past year from **StackOverflow** for **Part II** of this Assignment using any of the libraries listed for **StackExchangeAPI** (<https://api.stackexchange.com/>). See the list of common **StackExchangeAPI** libraries here:
 - <https://stackapps.com/questions/3881/seapi-a-lightweight-python-wrapper-for-the-stack-exchange-api>
 - <https://stackapps.com/questions/6841/stackapi-another-python-wrapper>
 - <https://stackapps.com/questions/198/py-stackexchange-an-api-wrapper-for-python>
 - <https://stackapps.com/questions/3417/stack-py-a-python-module-for-accessing-the-stack-exchange-2-1-api>
4. All of your source code must be clearly documented and functional; **ZERO credit will be given to the submission that has nonfunctional code.**
5. Submit your comparative analysis report for the results you obtained for all experiments you executed.
6. **ZERO credit will be given to the submission that has NO comparative analysis report.**

7. Submit your **source-code in three separate folders ((Part I, Part II, and Part III))** and live video of your run that has your code and your output.
8. The **Stackoverflow** dataset for **Part I** is available from different sources :
 - <https://www.ics.uci.edu/~dubois/stackoverflow/>
 - <https://archive.org/download/stackexchange>
 - <https://www.kaggle.com/stackoverflow/stackoverflow?select=votes>
 - <https://cloud.google.com/bigquery/public-data>

Requirements:

Part I:

Use Anaconda **Python** 3.7, and **Tensorflow/Keras** to create the StackOverFlow_Recommender ipynb script, and the provided **Stackoverflow** dataset to implement the following:

- 1) Use the provided Stackoverflow dataset (**answers.csv**)
- 2) Use Google Colab (<https://colab.research.google.com>) or your personal computer CPU and GPU
- 3) The intent is to make recommendations for a user who posted a question and got answered, and find other questions that you recommend to the same user based on the provided tags and their scores. Basically, users working on specific domain will ask similar questions and answers. If someone interested in python related questions, we will recommend similar/related questions in Python but not in Java for example.
- 4) The provided dataset needs some preprocessing and cleaning for the special characters.
- 5) Execute 3 experiments for the using the following packages/algorithms:
 1. Surprise/SVD/SVD++
 2. TensorFlow/Keras/Collaborative Filtering
 3. Restricted Boltzmann Machine
- 6) Provide a comparative analysis report discussing the results you obtain from the 3 experiments you executed.

Part II:

Redo **Part I** by using **StackExchangeAPI** or any wrapper libraries (listed above) for **StackExchangeAPI**, to pull data of the past year from **StackOverFlow**.

Part III:

Reuse the architecture, docker microservices, and Google cloud discussed in **Assignment #5**, redo **Assignment #5 Forecasting** for **Stackoverflow questions/answers** for the **Repos** used in **Assignment #5** by using **StackExchangeAPI** or any wrapper libraries (listed above) for **StackExchangeAPI**, to pull data of the past year from **StackOverFlow**.

Similar to **Assignment #5 Forecasting** for the different **Repos**, but rather than looking for how many **GitHub issues opened** or **closed** for every **Repo**, in this part of the bonus assignment you will be looking for how many **questions** and **answers** for every **Repo** name (use the **Tags** to search/index **Repo names**); you need to use the answers/question **Tags** in **answers.csv** as an example and create a dictionary/HashMap for every repo name (tag).

Assignment Deliverables:

You are required to submit a SINGLE Zip file that has the following deliverables are:

1. Your IPYNB script
2. All of your source code and output
3. Output report that has your assignment run saved in OUTPUT.pdf
4. Video recording of 10 minutes as a demo for the run of your assignment using <https://screencast-o-matic.com/>

Post your assignment as a SINGLE ZIP file on Blackboard.

Dr. Atef Bader