Bonus Assignment #1

General Instructions:

1. **Due Date is 5/8/21 by 11:59pm**.

2. **There is NO PARTIAL credit for the bonus assignment submission that has partial/not complete code.**

3. All of your source code must be clearly documented and functional; **ZERO credit will be given to the submission that has nonfunctional code**.

4. Submit your comparative analysis report for the results you obtained for all experiments you executed.

5. **ZERO credit will be given to the submission that has NO comparative analysis report.**

6. Submit your **IPYNB** script and live video of your run that has your code and your output.

7. The dataset for **Stackoverflow** available from different sources :
   - https://www.ics.uci.edu/~duboisc/stackoverflow/
   - https://archive.org/download/stackexchange
   - https://www.kaggle.com/stackoverflow/stackoverflow?select=votes
   - https://cloud.google.com/bigquery/public-data

8. Before you start working on this assignment you must read/review the following **RAPIDS** libraries :
   - https://rapids.ai/start.html
   - https://docs.rapids.ai/api
   - https://docs.rapids.ai/api/cudf/stable/
   - https://docs.rapids.ai/api/cuml/stable/
   - https://docs.rapids.ai/api/cugraph/stable/

Requirements:

Use Anaconda **Python** 3.7, **Tensorflow/Keras**, **RAPIDS** (https://rapids.ai/about.html), the Recommender ipynb script discussed in the class tutorial, and the provided Stackoverflow dataset to implement the following features for **Stackoverflow** dataset:

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) Reuse and modify the ipynb script discussed in the lecture and implement the 4 experiments provided but using the Stackoverflow dataset.
6) Choose a class for any machine learning algorithm from cuML library and call that Experiment #5 to make recommendations.
7) Provide a comparative analysis report discussing the results you obtain from the 5 experiments you executed.

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 ALL captured screen-shots of 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