Natural Language Processing - Assignment 04

Deadline: 21/11/2021 11:59:59 PM **Maximum Marks: 50**

Instructions:

- The assignment is to be attempted in a group of max 2.
- Language allowed: Python
- You are allowed to use libraries such as NLTK for data preprocessing.
- For Plagiarism, institute policy will be followed.
- You need to submit README.pdf, Code files (it should include both .py files/.ipynb files), and Output.pdf.
- Mention methodology, preprocessing steps and assumptions you may have in README.pdf.
- Mention your sample outputs in the output.pdf.
- You are advised to prepare a well-documented code file.
- Submit code, readme and output files in ZIP format with the following name: A4_<roll_no1>_<roll_no2>.zip
- Use classroom discussion for any doubt.

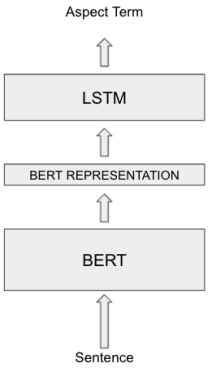
Dataset: Download here

- You can find the details of the dataset here

Task: Aspect Term Extraction and Sentiment Classification

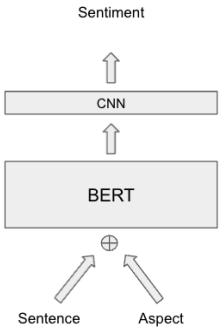
1. Implement the following architecture to perform the task of aspect term extraction.

   Note: You are free to choose the size and layers of LSTM.



2. Implement the following architecture to perform the task of sentiment classification.

   Note:
   a. You are free to choose the CNN architecture (filter size, numbers, etc.).
   b. You can adopt any technique to model the instance. However, the default technique should be contextual, i.e., for each aspect take +/- 5 tokens as input. Two examples for context size +/- 2 are shown below.
      - Context window size: ±2
        - *Although the display is poor*
        - *find its battery_life amazing.*



Note:

- *Use HuggingFace implementation for BERT.*
- *BERT finetuning is not mandatory but it is recommended to finetune to improve the model's performance.*
- *If there is no aspect term in a sentence, skip those. This assignment does not require aspect category detection.*

Evaluation:

1. Aspect-term Extraction: Precision, Recall and F1-score.
2. Aspect Sentiment Classification: Accuracy.