IISER Bhopal

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH BHOPAL

Course instructor: **Dr. Jasabanta Patro** Assignment number: **4**

Course: **DSE 407/607: NLP** Date: **October 24, 2024**

Marks: **10** Date of submission: **November 15, 2024**

Regulations:

- Each student is required to submit solutions based on the specified task.
- Multiple submissions are not allowed.
- Plagiarism: Strictly prohibited. All work should be original. The code will be checked for plague (as well as AI detector) and appropriate action will taken if found guilty of copying.

Submission Guidelines:

- **Deliverables:** public URL of **(i) code** (with proper comments).
  - The Colab notebook should only contain the inference part of the model and load the pre-trained weights. The training part should be commented out.
  - The model should be able to load weights from your public GitHub repository (create a repo with the trained model and download weights from it).
- File naming convention:

  rollno_name_nlpassignment1.ipynb
- Students need to submit only the URL of the Colab notebook (with public access) with clear instructions for running the code. The runtime of the code should not be more than 10 minutes.
- Deadline: All assignments must be submitted by the deadline. Late submissions will be penalized.

Marking:

- Marking will be done based on only one criteria, (i) **model performance**.
- The performance of each submission will be evaluated using average macro F1-score based on the predicted labels and the gold ones.
- All submitted code should be reproducible with public access. If the results cannot be reproduced, the submission will be considered incomplete and the submission will not be marked.

Text Classification Using Transformer and Pretrained Language Model:

Given a target text snippet, predict the perceived emotion(s) of the speaker. Specifically, select whether each of the following emotions apply: joy, sadness, fear, anger, surprise, or disgust. In other words, label the text snippet with: joy (1) or no joy (0), sadness (1) or no sadness (0), fear (1) or no fear (0), anger (1) or no anger (0), surprise (1) or no surprise (0).

Further details on the task (Track-A) is provided in the following URL:

https://github.com/emotion-analysis-project/SemEval2025-task11

The sample dataset is provided in the above URL. Students can refer it to understand the task and the dataset format. The train dataset is available on the URL : https://github.com/debajyotimaz/nlp_assignment. Students are supposed to submit the code of the best performing algorithm and that algorithm will be evaluated on our custom test set (which will be kept private). Use this colab notebook to get started: https://colab.research.google.com/drive/13yNxUnB866IqHF8H1mnXIRQYT-t4COBA?usp=sharing.

Task overview:

1. **Base model:**
   - Use any embedding and model, and get the best performance.
   - Usage of pre-processing techniques is optional.
2. **Note:** Comment out the training part of the code (we can undo comments to check the training part also). The model should already be trained, and the deliverable will be focused on inference only.