In this assignment you will fine-tune a BERT model for text classification.

**Set up:**

1. Go through and understand the slides A4.pptx. Install the libraries and run its code .

2. Choose and obtain a text classification dataset. You can select from Hugging Face using the datasets library as shown in the slides. Please make sure the dataset has texts and labels (nominal). There should be at least 1000 training and at least 1000 test examples. In case the dataset is too large to train with on your computer, you can use only its portion (see the slides). *Do not use use the imdb dataset from the slides.*

3. Select any BERT model or one of its variation with a suitable size and case/uncased. Obtain it and make sure you can work with it as in the slides. *Do not use a BERT model that is already fine-tuned on your dataset.*

**Tasks:**

1. Fine-tune the BERT model on the training dataset. You can use any network architecture on top of the BERT, the final layer must be softmax for classification. Choose a suitable number of epochs and a suitable batch size. Test the fine-tuned model on the test dataset. Report test accuracy.

2. Look at the actual classification predictions of the fine-tuned model on the test dataset. Take at least 10 examples the model predicts correctly and at least 10 examples which the model predicts incorrectly, from these make at least 3 general observations (an observation should not be specific to any one example).

3. Come up with at least 5 examples demonstrating BERT model's contextual embeddings analogous to the "dog" and "cat" example towards the end of the slides *(but do not include that example)*. Include your examples in the report along with the cosine similarity scores.

**Submission:**

1. Report (pdf file):

Write a report in which you clearly include the following:

* (1 point) A brief description of the dataset you used for fine-tuning, including its genre of text, classes and size (and if you used only a portion).
* (1 point) A brief description of the BERT model you chose to use.
* (1 point) A brief description of your network and your training setting.
* (2 points) Result of Task 1.
* (1 point) Your comments on the Task 1 results.
* (3 points) The 3 observations from Task 2 (include the correct and incorrect examples and their predictions).
* (3 points) The 5 examples of Task 3 and the results along with your comments.

2. Code (2 points): Submit all the code you wrote for doing this assignment as an executable .py file. The grader should be able to import it and replicate your results.

3. (1 point) Your text classification dataset for Task 1, if it cannot be downloaded through your submitted code.