1. Submission will happen over Moodle.
2. Your submission will be run on HPC. We have set up a clone conda environment over HPC for you to verify the dependencies for your submission. The conda environment is present in the following directory

   /scratch/cse/phd/csz208845/COL772/A2/

   Note that conda activate function in HPC requires loading and unloading of anaconda modules. The full set of commands needed for activating the environment are:

   module load apps/anaconda/3
   conda activate /scratch/cse/phd/csz208845/COL772/A2/
   module unload apps/anaconda/3

   We have included following packages in the environment

   pandas==1.1.3
   numpy==1.19.5
   scikit-learn==0.24.2
   scipy==1.5.3
   nltk==3.5

   Nltk data has been downloaded in the environment. You have to verify that your submission works with the above libraries. In case you have any additional dependencies, raise a query over piazza.

3. Your submission should include a run-train.sh script which will be run as follows

   bash run-train.sh <data_directory> <model_directory>

   Your script should read training data from <data_directory>/training.csv. Your script should save the model after training at <model_directory>.

4. Your submission should include a run-test.sh script which will be ran as follows

   bash run-test.sh <model_directory> <input_file_path> <output_file_path>

   Your script should load the trained model from <model_directory> score text in <input_file_path> and write prediction to <output_file_path>. Input and output formats are as specified in the assignment document. Please find a sample input and output file attached.