ReadMe

DLFA - Cohort 3 Module 3 – Deep Learning for Natural Language Processing

Kaggle Competition - 28th May 2023

Natural Language Processing - Irrelevant/Inappropriate Questions Classification

This Kaggle competition on Natural Language Processing (https://www.kaggle.com/t/bde6f23028154933a99e4b4ca8a3dff2) is a team activity. This activity carries 10 marks towards the DLFA program. It is mainly to test your understanding of the topics that have been covered so far in the Deep Learning program and specifically on Natural Language Processing. This competition will start on Sunday, May 28 and end on Saturday, June 17, 2023 11:59 PM. We have helped you in forming teams during the entire duration of the competition. Each team can have a maximum of four to five members. The details of all the teams have been uploaded on the LMS. You are welcome to rename your team while making the submission on Kaggle. So, you are encouraged to hold team meetings as per your convenience to discuss your progress and submission. We also require that as a team, you do not share your code with any other team. The submission has to be made on Kaggle only. However, we will be evaluating your code as well. Please join the competition and accept the rules by clicking the link given above. After that, you can form a team and make submission using the format given on Kaggle. We will provide support on Saturday, June 3, 2023 with the mentors for a live session in place of the usual assignment session, in case you have any queries or are facing any issues. The mentors will be there to help you in resolving any issues you face, but will NOT help you in coding.

Please download the test and train data using the google Colab Notebook which is being provided to you. The train data, the test data and the sample_submission file (all in .csv format) are provided on Kaggle, and can also be downloaded using the colab notebook which we have provided.

The Problem Statement, Objective, Background, Dataset, and Steps for implementation are provided within the Instruction Notebook (Colab).

As this competition aims to classify 'irrelevant/inappropriate' questions and 'relevant/appropriate' questions, you are required to build an RNN (Long Short Term Memory/Gated Recurrent Unit) or a Convolutional Neural Network (CNN) or BERT (with/without 'attention') using either Keras or PyTorch deep learning libraries on train and test dataset samples and perform classification of approximately 261221 questions asked on a well known public form as 'irrelevant/inappropriate' questions or 'relevant/appropriate' questions. Please predict the results to sync with sample_submissions.csv. You are also welcome to use pre-trained models such as BERT (with 'attention' or without 'attention') to train and evaluate the test data.

During the entire duration of the competition (Tuesday, May 28 - Friday, June 17), the mentor(s) can be approached for guidance when there are any difficulties in implementing specific steps. On Saturday, June 3, there will be a mentoring session on the Kaggle Competition in place of the usual assignment session. However, it is the team's responsibility to work towards the solution in the given time. Your team's solution notebook should be presented to us, latest, by 11:59 PM on 17th June 2023. Please make the submission on Kaggle exactly in the format given in the 'sample_submission.csv' file. Any deviation from the specified format will result in an invalid submission. You can make up to 20 submissions (as an individual or as a team) in a day. After June 17, 2023 11:59 PM, you will not be able to make any submissions. Failure in making a single submission will result in the entire team getting zero marks.

Note: Submit the notebook to dlfa.support@talentsprint.com after the completion of kaggle competition to ensure the 'proof of code' for your final submission on kaggle before Saturday, June 17, 2023 11:59 PM.
● In colab notebook, goto *File* and *save and pin revision* to save the changes.

- Make sure to give the edit access and share only one file which resembles the final predictions that are submitted on kaggle.