Coursework specification: 490 - NLP (2020)

Your task will be to develop a regression model to predict the quality of a machine translated sentence. We provide two language pairs: English->German and English->Chinese. You can select any of these language pairs to submit a model for.
Details on the task are outlined here: https://competitions.codalab.org/competitions/22831.

Teams: You should self-organise to form groups of up to 3 students. Submissions with fewer students will *not* be assessed differently. Please choose a team leader to create an account on codalab, download the data and make the submissions. Specify on the profile the team name and team members (names of students).

Data and evaluation: This website will provide a training set to build models, a development set for model selection or hyperparameter tuning, and a blind test set. Feel free to re-split the training to create a separate test set for preliminary experiments, but for the final submission we recommend using the provided splits. The evaluation of your systems will be done automatically by the online platform. You should be given a performance score right after submission.

Submission: Each team is allowed to submit up to 10 systems for a given language pair. The leaderboard will only show your best system, but you will be able to see the results for other submissions and report on them. Make sure to submit your systems to the CodaLab platform early enough to make sure your format is correct and you can obtain results to enter on your written report before the submission deadline.

Report: In addition to submitting one (or more) systems you will need to submit to CATe a written report (pdf format) by February 28th (midnight) where you include:

1) Details on the design of model(s) built: architecture, hyperparameters (and their selection process), algorithms used for learning, optimisation, etc., deep learning framework used (PyTorch, Keras+?, TensorFlow, etc)
2) Type of input data: word embeddings (which type? Pre-trained? Contextualised?) vs bag of words vs pre-extracted features, etc.
3) Any pre-processing done to the data
4) Performance obtained for the blind test set (up to 10 submissions) and any validation experiments (either on the given dev set or a subset of the training data)
5) Analysis/insights/discussion on the results and challenges encountered in designing your models.

Your report should not have more than 4 pages of content and 1 additional for references. It should follow the following latex style: `https://www.overleaf.com/read/xhxbhtgjgbxv`
Make sure you download or make a copy of the overleaf project as this one is shared with all of you.

Code: Finally, you will need to submit your code: please add the url to a colab file or git repository in your pdf report.

Marking: Your mark (20% of the course mark) will be based on the diversity and creativity of ideas used to devise your models (including pre-processing), not on the performance of the models on the task. The code will be checked and marked according to whether it follows basic coding good practices, such as organization, readability and documentation (at least a readme file) and comments. You are allowed to use any libraries for NLP/ML, but make sure to add some of your own code.