NLP with Deep Learning - Course work

Your task will be to develop models to identify and categorize offensive language. The task is outlined here: https://competitions.codalab.org/competitions/20011. More specifically, the course work will require systems for three sub-tasks:

- Sub-task A - Offensive language identification: given a text, your system should categorise it into OFF (offensive) and NOT (not offensive).
- Sub-task B - Automatic categorization of offense types: given a text, your system should categorise it into TIN (targeted at an individual) and UNT (untargeted).
- Sub-task C – Offense target identification: IND (individual), GRP (group), and OTH (other).

Each subtasks can be dealt with by building classifiers (2-3 classes).

This website will provide your training set to build models, test sets, as well as a platform to submit your predictions on the test set. The evaluation of your systems will be done automatically by the online platform (https://competitions.codalab.org/competitions/20011#participate). You should be given a performance score right after submission, if not you will be given the test set to test locally. You will have to create an account on the task website to obtain the data and participate.

You can self-organise to form groups of up to 3 students.

In addition to submitting one (or more) systems for each task online you will need to submit to CATe a written report by March 1st (midnight) where you include:

1) Details on the design of classifier(s) built for each task: 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 in the submissions and any validation experiments
5) Discussion on the results and challenges encountered in designing your classifiers

Finally, you will need to submit your code as a colab file, or upload it to a git repository and share the link in your report.

Your report should not have more than 5 pages. Your mark will be based on the diversity and creativity of ideas used to devise your classifiers, not on the performance of the classifiers on the task, nor the actual code.

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.