

CS410 Text Information Systems(Fall 2020)
Project Proposal

1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

Edward Ma - kcma2(Captain)

Oran Chan - wlchan2

2. Which competition do you plan to join?

We are planning to work on the competition of Text Classification

3. If you choose the IR competition, are you prepared to learn state-of-the-art IR methods like query expansion, feedback, rank fusion, learning to rank, etc.? Name some more concrete methods or tools that you may have heard of. If you choose the classification competition, are you prepared to learn state-of-the-art neural network classifiers? Name some neural classifiers and deep learning frameworks that you may have heard of. Describe any relevant prior experience with such methods.

We are interested in SVM, XGBoost and especially the BERT model which has proven performance on downstream tasks. Hence, we decided to use PyTorch to fine-tune this state-of-the-art method to tackle various classification tasks. Toxic comment classification is one of my experiences working on BERT.

4. Which programming language do you plan to use?

Python