Pick one of the following two text classification tasks from the GLUE Links to an external site.benchmark set.
Stanford Sentiment Treebank (SST-2)Links to an external site.
Multi-genre natural language inference (MNLI)Links to an external site.

Download the data, design a classifier of your choice, and report results on the test set.

The sentiment task involves predicting a binary label for a phrase (0/1=negative/positive sentiment). SST-2 is a derivative of an earlier task Links to an external site.that labeled all phrases in a syntactic tree and included neutral labels. The tree-based labeling is more complex -- please just stick with the 0/1 task.

The MNLI task involves 3 labels for the relation between sentence pairs: neutral, contradiction and entailment. For MNLI, the test set is privately held, so use the development set as your test set and randomly choose 10% to the training data for your development set.

You may use either a neural bag of words model, a CNN, or an RNN or transformer if you are more ambitious or more experienced. You should submit a short but comprehensive description of the approach you used with details of the implementation and variations that you assessed with the development data, together the accuracy of your best development configuration on the test set. For example, if you take a BOW approach, specify the pre-trained word embeddings you used and any word-based weighting you used. Be sure to only include one result on the test data. Your grade will be based on the quality of your write-up and achieving results in the ballpark of what others have achieved for that method.

You can develop your own models with pytorch / tensorflow or keras. If you want to play around with pretrained models, you can use Huggingface Transformers to do that. AI2NLP could also be useful if you are familiar with that.