Homework 4: Text Classification 2

Objectives:

- Gain experience with Keras/TensorFlow
- Gain experience with text classification
- Gain experience with deep learning model variations and embeddings

Turn in:

- This program should be created in a notebook (Jupyter, Google, or Kaggle)
- Print to pdf and upload your pdf to eLearning

Instructions: Use Keras/TensorFlow

1. Go to Kaggle.com. Find a text classification data set that interests you. Divide into train/test. Create a graph showing the distribution of the target classes. Describe the data set and what the model should be able to predict.
2. Create a sequential model and evaluate on the test data
3. Try a different architecture like RNN, CNN, LSTM, etc and evaluate on the test data
4. Try different embedding approaches for your best performing model in Step 3, and evaluate on the test data
5. In a text block in your notebook, write up your analysis of the performance of various approaches

Grading Rubric:

- Each part is worth 0 to 20 points
- Your grade is not determined by the accuracy achieved, but by how much work and thought you put into it

Bonus 10 points:

- Use an advanced architecture such as a transformer or a pretrained model (from TensorFlow Hub, Hugging Face, or elsewhere), and evaluate its results compared to your best model above.
- Create a TensorBoard graphic and list two useful items of information you gleaned from the interface. Include screen shots.

Caution: All course work is run through plagiarism detection software comparing students' work as well as work from previous semesters and other sources.