Challenge Overview

For this challenge, you will work on the topic of Natural Language Processing. More specifically the task of [sentiment analysis](#).

The dataset we will work with is the IMBD [Large Movie Review Dataset](#). It contains 50k movie reviews made of text representing the viewer's comments along with a label "positive" or "negative".The goal of this challenge is for you to build and compare different approaches to tackle the problem which ideally improves upon the provided baselines.

Starting point

To help you get started, we have attached a sample [notebook](#) that shows you the gist of a basic approach. As in real life, there may be issues with the baseline notebook provided and it's up to you to take initiative to scour the literature, forums, papers and tutorials to find out helpful implementations that will help you improve upon the baseline.

Deliverables

You are to deliver:
- A well-commented Jupyter/Google Colab notebook or GitHub repository showcasing **at least one** additional NLP models for sentiment analysis improving over the baseline
- At least one deployed ML model that can be interacted with via an API or web interface (live web demo)
- A written report outlining your choices of model architecture and deployment architecture

Some questions to answer in your report:
- Beyond the provided baseline, what other approaches did you decide to pick?
- How did you develop a training/validation and test set? Did you use cross-validation?
- What performance metric did you pick to measure how well your model did? Why ?
- What services did you use to deploy the trained ML model? What are its modules/components?
- For the different models you picked, what were their relative strengths and weaknesses?

Deadline and Submission

You can put your notebook on either a GitHub repository and invite mhkane to it or send it as a link on a [Google Colab Notebook](#) along with your report to kanemh@mit.edu

The deadline for submitting your notebook and report is **Monday, June 20th 2022** at 11:59pm GMT+1. No late submissions will be accepted.

Main Assessment Criteria

- At least one additional model developed and investigated which improves over the baseline.
  - Ideally, model coming from a family of architectures such as:
    - LSTM
    - RNN
    - CNN
    - Transformers
- At least 1 model we can interact with either via API or via a web interface
  - For deploying a model, you can just pick one model (including the provided baseline) and deploy it. Some example of tools
    - [Gradio](#)
    - [Streamlit](#)
    - [Spaces - Hugging Face](#)
- Written report including details on chosen models, their performance and services used to deploy them. Nice to have:
  - visualizations, graphs, no typos
  - Explanation of your approach, choices and reasoning
  - Length guideline: 4-6 pages. Can be more if you wish.

Bonus Points (by order of importance):

- If you find yourself having additional time or wanting to exceed expectations, you can:
  - Deploy a transformer model that can be interfaced with via a web interface
  - Have more than 1 ML model developed and investigated which improves over the baseline.

If you have any additional questions during this challenge phase, do not hesitate to reach out to kanemh@mit.edu