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 notebook or GitHub repository along with a written report outlining the additional approaches you have picked to solve this problem. In the repository, we will look for evidence of well-documented and written code.

For the report, we will look for a well-motivated choice of approaches, clear analysis of strengths and weaknesses of each approach along with evidence of detailed understanding of the concepts involved for this challenge.

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?
- What different models did you try? What were the observed results?
- What performance metric did you pick to measure how well your model did. Why ?
- 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 **Wednesday, May 25th 2022** at 11:59pm GMT+1. No late submissions will be accepted.

Main Assessment Criteria

- 2-3 additional models developed and investigated which improve over the baseline.
  - Ideally, models coming from different family of architectures such as LSTM, CNN and Transformers
- Written analysis of model results mixing statistical approach (performance metrics) and detailed example investigation showing understanding of machine learning and NLP concepts.
- Easy to read and well-explained technical report in English outlining thought process and learnings behind solution.
  - Nice to have: visualizations, graphs, no typos
  - Length guideline: 4-6 pages. Can be more if you wish.

Bonus Points:

- First, make sure you complete the core assessment.
- If you find yourself having additional time or wanting to exceed expectations, you can use recent MLOps tools to deploy a model of your choice. Some examples of tools:
  - [Gradio](#)
  - [Streamlit](#)
  - [Spaces - Hugging Face](#)

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