Assignment 4

Submission deadline: 30th September 2023

# Problem:

Make a model to detect speech. Preferably online.

# Restrictions:

* Your model should be trained only on Google colab (free version). It should be deployable/testable there too.
* Write your codes yourself. Allowed libraries: numpy, scipy, pandas, tensorflow, pytorch, librosa, matplotlib, pdb.

# Running the code:

$ python main.py -i abc.wav -o abc.csv

-i: input wav file

-o: output csv file

* It has two columns: <start time in s>, <end time in s>
* Each row has the start and end times of a speech segment. Multiple rows for multiple speech segments

# Baseline:

* <https://github.com/snakers4/silero-vad>

# Evaluation Metrics:

* Segment wise: precision, recall, F1
* Event wise: precision, recall, F1
* Use sedeval library <https://tut-arg.github.io/sed_eval/>

# Baseline Results:

# Useful Datasets:

A sample audio and transcriptions are shared in the folder. You may use this for development. The test files will be similar to this.

* <https://www.kaggle.com/datasets/lazyrac00n/speech-activity-detection-datasets>
* <https://pixabay.com/sound-effects/search/ambient/>
* <http://www.openslr.org/12>

# Report:

Write the report as a research paper. Page limit 2 pages + references. It should contain the following sections:

* Introduction (it can be very short)
* Method description
* Results
* Discussion (on salient points of your method that others may not have thought about)