This READMe is for the lstm.ipynb model. However, the gru jupyter notebook is exactly the same except that it uses a different model name and that the LSTM layer has been replaced by a GRU layer.

Note: The current output of the jupyter notebooks may be from a different preprocessed dataset than what actually perfomed the best. If you want to have the “correct” output for our best-performing model, you may have to rerun/retrain the entire model.

lstm.ipynb - READMe

Description: This notebooks trains a LSTM model

-----------------------

Prerequisites:

- Logged in with a Google account, with access to Google Colab, Google Colab's CPUs and GPUs (subscription needed to have prioritised access GPU's for a significant longer time) (for this notebook, you only need CPUs)

- Granting permission to run each of our Google Colab notebooks and python files.

- Mount drive at /content/drive/ [This is done when running the notebook, see below]

- The folders/files regarding Preprocessed training and validation datasets are at the described location.

-----------------------

How to run:

Before running, the user can specify

- which preprocessed datasets they want to work with

- which embedding they want to work with

In the second section "Choices", there is a variable

PREPROCESSING\_CHOICE defined by PREPROCESSING\_OPTIONS[INDEX], where INDEX is an integer [0;8]. The user can choose INDEX from that range, which corresponds to the preprocessed dataset.

In the second section, there is also a variable "embedding\_choice", an integer from [1;5], where each integer encodes an embedding (as described in the comments)

After selecting INDEX and embedding\_choice:

Run all sections, top to bottom

-----------------------

Output:

In the directory /content/drive/CIL22022, the trained models are stored as file (if chosen to save the mode) and the performance of the chosen model is logged.