# ENDEMIC

## Folder Contents

* data.zip - Contains the ECTF data according to the format expected by the pipeline. PolitiFact and GossipCon to be downloaded from their original source: https://github.com/KaiDMML/FakeNewsNet/tree/master/dataset
* covid19.sh - Shell config file which ensures the end-to-end working.
* extract\_features.py - Prepares the data by extracting various inputs from the tweet objects containing json files.
* preprocess.py - Prepares all the data according to the embedding vector inputs.
* model3.py - Contains all the model component classes
* training.py - Runs the training epochs and evaluation steps
* main.py - Loads all the data and initialises training, followed by evaluation
* config.py - Contains all args which are parsed after running main.py, along with default values.

## Compute Requirements:

- 5GB GPU for training

- 1GB for inference

## How to run?

- Unzip data.zip

OR

- Add/Update data in data/ECTF/ in json format, as fake.json, genuine.json, unlabelled.json. (sample data given in CTF are given here in json format)

- Install the requirements

```

pip install -r requirements.txt

```

- Change the project and directory name in covid19.sh (optional)

- Run the pipeline (changing the paths appropriately, if any of the above steps are done differently)

```

sh ./covid19.sh

```