

1 Description of Submitted Files

There are six folders containing the implemented models:

- **logistic_regression**: Run *log_reg.py* to perform logistic regression
- **BERT**: A finished copy is named as *BERT.ipynb* and exported as pdf with name *BERT.pdf*
- **bow, lstm, bi_lstm, att_lstm**: Each contains a *_model.py, _sentiment.py and run_lstm.pbs. *_model.py contain the model structure and _sentiment.py contain the training and testing process.

Following files are related to data preprocessing:

- *preprocess.py* would prepare the data, transform the tokens to ids and save for later use.
- *save_data.py* would pad the data prepared by *preprocess.py* according to the desired sequence_length and save them to numpy array.

2 Execution Instructions

- *log_reg.py* could be run directly on Blue Waters.
- *BERT.ipynb* could be run on Google Colab.
- To run the rest models, One may login to computer node on Blue Waters and enter module load python/2.0.1. Then, one need to run *preprocess.py* and *save_data.py* to prepare the data. Data will be saved to the folder **pre-processed_data**. Finally, one can either submit jobs using run_lstm.pbs or directly run *_sentiment.py after you cd to the folder of different models.