# \_UIT-VSFC folder

Contains raw data (sentences in text). More info can be found in the README.txt of that folder.

# segmented\_data folder

Contains raw data but the words are segmented using RDRSegmenter from VnCoreNLP. These files are in csv format.

# checkpoint folder

Contains all the best trained models (.pth) to load.

# models folder

Contains our notebooks for the implementation of RNN, LSTM, Transformer, BERT models and an Ensemble notebook.

Note: The code to preprocess data for RNN and LSTM is in the RNN.ipynb notebook.

# preprocessed folder

Contains preprocessed data represented in One-hot or Tokenizer Encoding. More info about these preprocessed data can be found in the README.txt file inside the folder preprocessed/data\_preprocessed.

# output\_for\_ensemble folder

Contains all the output vectors of RNN, LSTM, Transformers to be used to test Ensemble performance and the notebooks to generate them.