Homework 3

Due: 4/23/2025

Task: Develop a Named Entity Recognition (NER) tool based on BERT model.

Sample code: You can find an open source at https://colab.research.google.com/drive/1sSVjRSUf5FzgSbe4FWw--2vanLOiqN8m?fbclid=IwAR2mQw_H-W8ZEow-51mR1geICfnEy2Q22KwdP3BiI_VmUOMW5bkeCEqUiTM#scrollTo=eTPw5A7fFeMC

With document at https://bc165870081.medium.com/ner%E6%95%99%E5%AD%B8%E7%AF%84%E4%BE%8B-ea4216047728

Your task for this homework is:

1. Use the DNRTI dataset for training, validation, and testing. The dataset is available from the E3 platform.
2. Train your model using train.txt.
3. Validate your model using valid.txt.
4. Evaluate your model using test.txt.
5. Bonus: Usually, SecBERT with LSTM/BiLSTM and/or CRF could provide better performance; you can try it.
6. Upload the report (.pdf) , the prediction file, and all program files (.ipynb) to Github.
7. Use your Github URL as the answer to the homework.

Example code: LSTM+CNN+CRF (you should change its embedding to SecBERT for better performance):
https://github.com/TheAnig/NER-LSTM-CNN-Pytorch/blob/master/Named_Entity_Recognition-LSTM-CNN-CRF-Tutorial.ipynb