Malay-speech fine-tune guidance

* In pre-trained model (The model on which we do fine-tuning) package should include six file:

Pytorch\_model.bin

Config.json

Preprocessor\_config.json

Special\_tokens\_map.json

Vocab.json

Tokenizer\_config.json

* In loading\_data transcript: “uob\_dataloading.py”

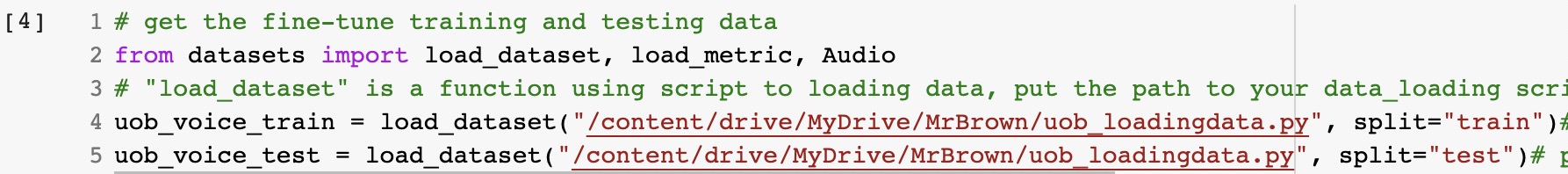
Here put the path to our prepared zip file (including one train folder, one test folder, one train.csv and one test.csv)

Here put the unzipped file name (usually same with the zip file name without extension)



* In fine-tuning code: “Malaya-speech\_fine-tune\_demo.ipynb”

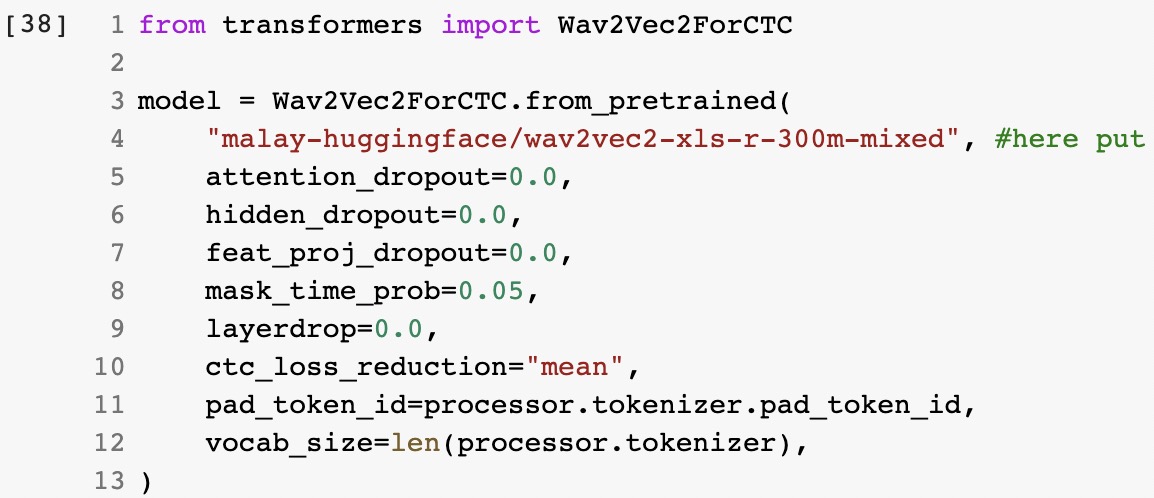
Here put the path to the “uob\_dataloading.py”



Here put the path to the place you want to put the loaded data, or it will store the dataset to cloud (as we cannot connect the cloud)



Here put the path to pretrained model (the malaya-speech STT model pacakge)



Here put the path to the package you want to store the fine-tuned model (also means those six files that can consist the model: Pytorch\_model.bin, Config.json, Preprocessor\_config.json, Special\_tokens\_map.json, Vocab.json, Tokenizer\_config.json)

