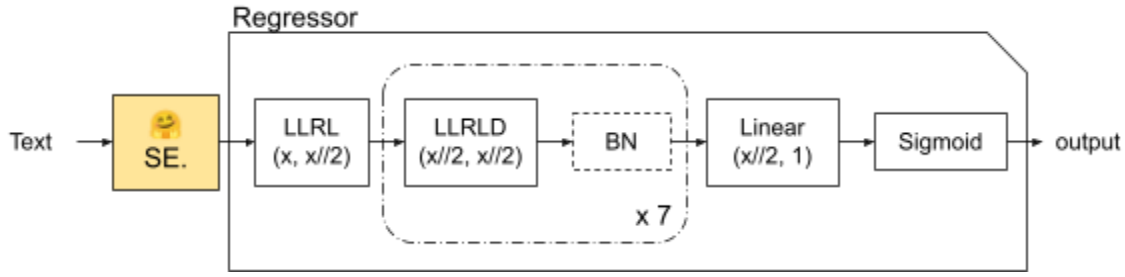
My model

The sentence is encoded by a sentence encoder hosted on a hugging face and I have created my own architecture for the regressor model.



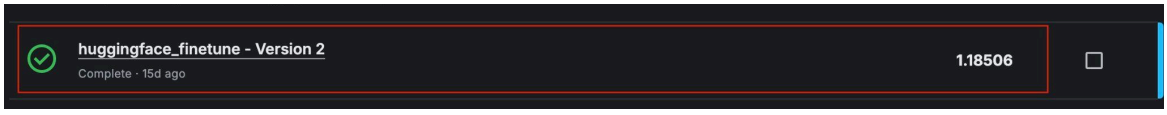
Breakdown

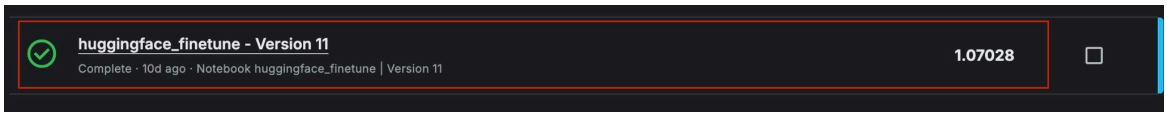
1. 🤗Sentence encoder (**SE.**)
   I used `sentence-transformers/paraphrase-multilingual-mpnet-base-v2.`
2. **LLRL** is group of Linear + LeakyReLU
3. **LLRLD** is for Linear + LeakyReLU + Dropout
   a. Dropout rate is 20% (0.2)
4. **BN** is BatchNorm1d and applies to every 3 layers to prevent gradient vanishing.
5. The model's output will be `Sigmoid(logit) * 5`

Methodology

1. Due to lack of training data, I decided to split data into 10 folds and train 10 models separately.
2. Each model will train with `batch_size=32`, `lr=0.01`, `weight_decay=0.01`. Additionally, The model applied `ReduceLROnPlateau` learning rate scheduler with `patience=10` and `factor=0.5`. Every model will fine-tune the transformer for their own and optimize `Mean Square Error` loss function.
3. The final answer will be arithmetic mean among the 10 models

Results





Kaggle name: Jirayuwat B. / Student ID: 6432023321