

Checkpoint 2.2:

boolq validation set:

distilRB-rand: overall acc: 0.622, f1: 0.383

distilroberta: overall acc: 0.664, f1: 0.637

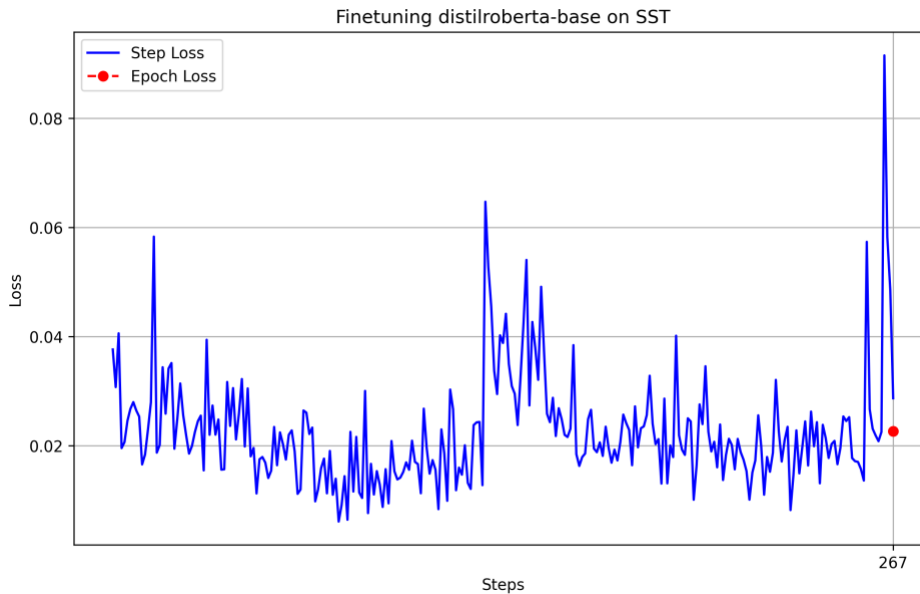
distilRB-KQV: overall acc: 0.637, f1: 0.448

distilRB-nores: overall acc: 0.654, f1: 0.527

Checkpoint 2.3:

Training plot saved to

results/finetuning\_loss\_plot\_distilroberta\_base\_regression.png



Validation: mae: 0.129, r: 0.781

Test: mae: 0.129, r: 0.794

SST test set:

distilRB-rand: mae: 0.484, r: -0.005

distilRB-KQV: mae: 0.225, r: 0.373

distilRB-nores: mae: 0.139, r: 0.780

Checkpoint 2.4: Extra Credit

boolq validation set:

distilroberta: overall acc: 0.664, f1: 0.637

distilRB-improved: overall acc: 0.668, f1: 0.590

SST test set:

distilroberta: mae: 0.129, r: 0.794

distilRB-improved: mae: 0.185, r: 0.709