

F1 on Test Set with 20% of Training Set  
Used for Supervised Finetuning

