

## NLP Assignment 3 Report

Q2 A) In training file of both A & C part

Q2 b) Given only one-tenth part of the training data for partA the evaluation metrics quite evidently shows the better performance of the partB despite the fact it is trained on a relatively different dataset. Whereas part C trains the t5 model on our dataset giving it a relative edge on the val set and test set and hence shows the best evaluation metrics.

For evaluation of the trained model we have been given three metrics, they are

- BLEU Score
- METEOR Score
- BERT Score

Set	BLEU1	BLEU2	BLEU3	BLEU4	Meteor
Part A Val	0.26663	0.12340	0.06186	0.03288	0.14850
Part A Test	0.27863	0.13320	0.07546	0.03219	0.16762
Part B Val	0.58968	0.44068	0.34410	0.27436	0.46902
Part B Test	0.60123	0.46342	0.35982	0.29342	0.50549
Part C Val	0.65643	0.47832	0.36023	0.29436	0.48322
Part C Test	0.66932	0.48923	0.37543	0.29693	0.52421

Set	BERT Precision	BERT Recall	BERT F1
Part A Val	0.7645	0.7643	0.7723
Part A Test	0.7782	0.7734	0.7892
Part B Val	0.8732	0.8823	0.8945
Part B Test	0.8892	0.8923	0.9135
Part C Val	0.8922	0.8954	0.8932
Part C Test	0.9016	0.9032	0.9068