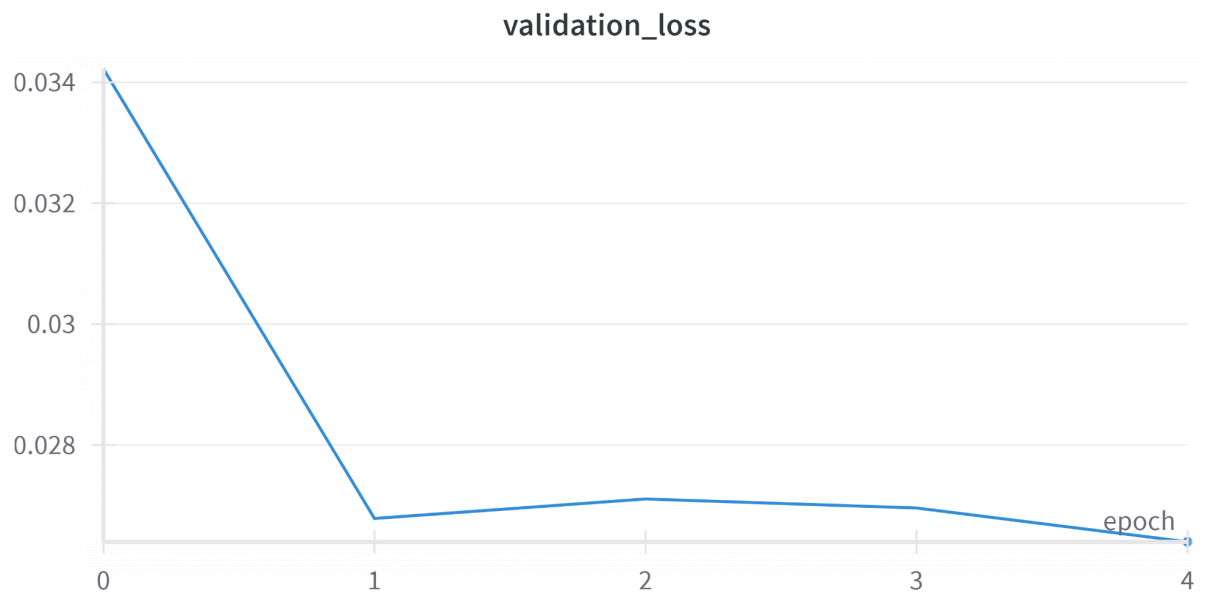
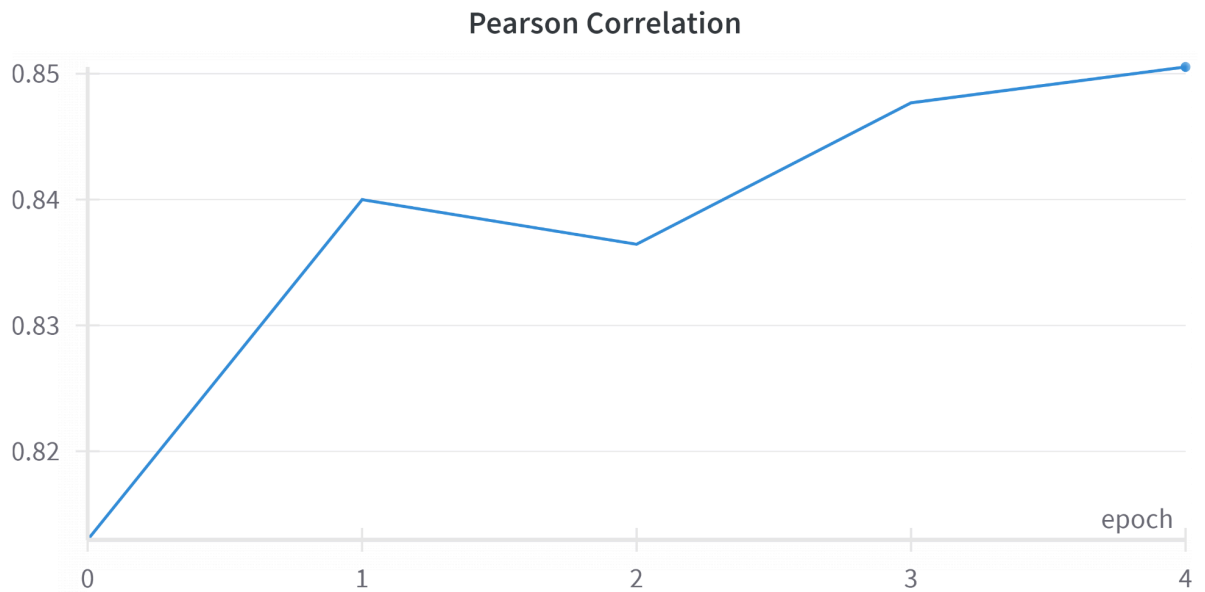


## Task1

A.





For the train and validation losses, the rate of change of loss is very high in the first 2 epochs and gradually reduces.

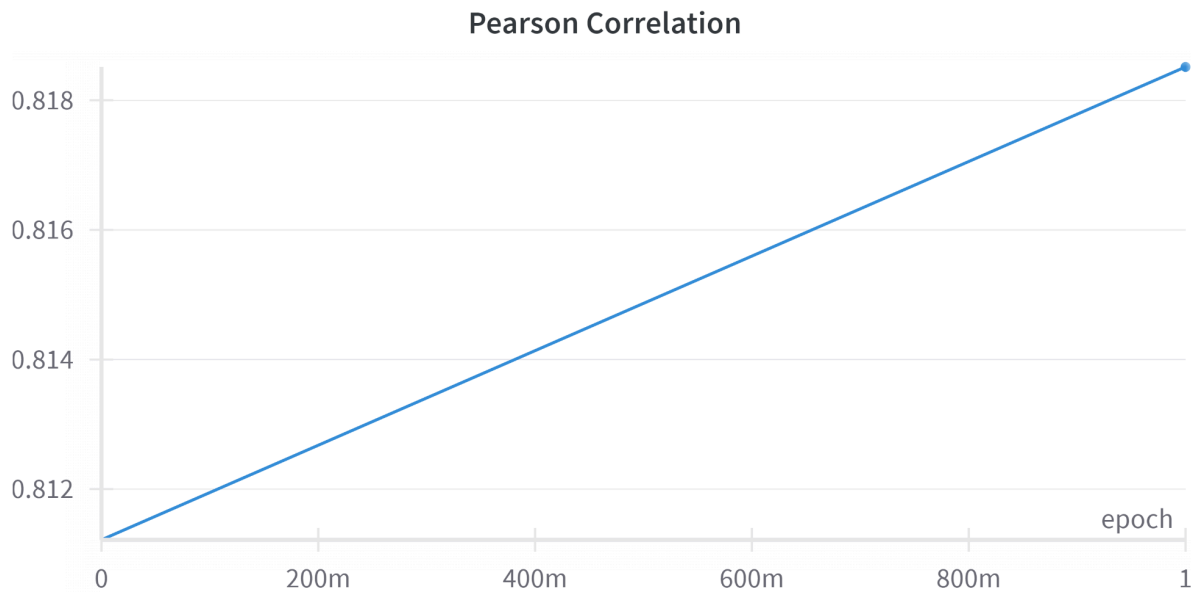
The evaluation metric used was the Pearson correlation. At the end of training the model for 5 epochs, we achieved a Pearson score of 0.8505 on the validation dataset.

The Pearson score decreased at the third epoch since the validation loss went up.

- B. The evaluation metric used was the Pearson correlation. Without finetuning, the Pearson correlation was 0.792 on the validation dataset.

C.





We ran the 1C for 2 epochs. Training loss and validation loss decreased over the 2 epochs. When we increased the number of epochs, we observed overfitting in the model, therefore we stuck to 2 epochs.

The evaluation metric used was the Pearson correlation. At the end of training the model for 5 epochs, we achieved a Pearson score of 0.8185 on the validation dataset, which is better than 1B.

### **Comparison Between Models:**

1A performs better than 1B and 1C. 1B was just a pre-trained model, which was not fine-tuned at all. 1A was trained for a greater number of epochs than 1C, which could be one of the reasons why it is performing better.

Task 2:

A.

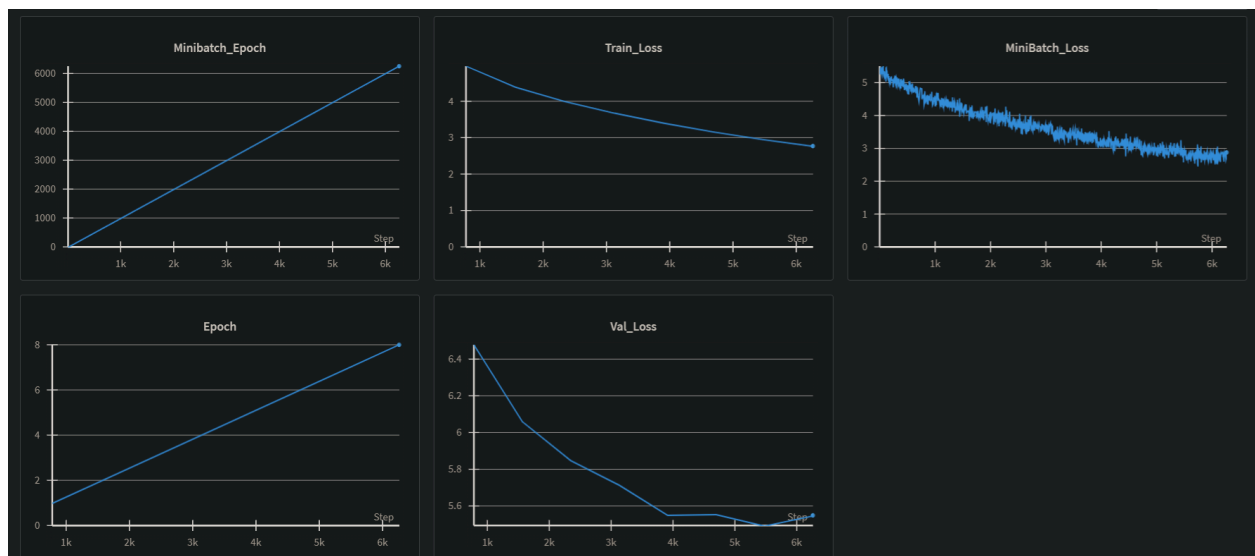
Both validation and training graphs are steep and not converging. This shows that the model can be trained for more epochs for better results.

### **Validation dataset results:**

**BLEU:**

BLEU-1 score: 0.718031894381054

BLEU-2 score: 0.5519852998864307  
BLEU-3 score: 0.41669291365055505  
BLEU-4 score: 0.32828448914569114  
**Meteor:** 0.23495087207185508  
**BERT Score:**  
Precision: 0.841719959371  
Recall: 0.8504983473340725  
F1: 0.84591299649631



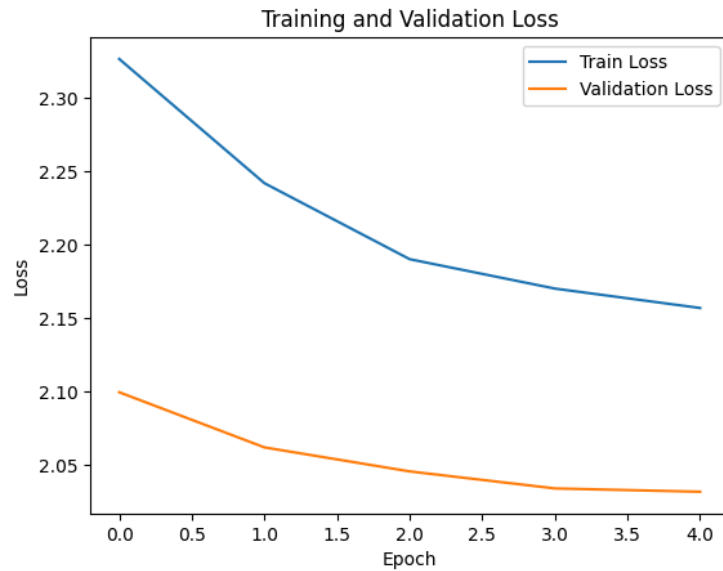
B.

BLEU-1 score: 0.8693316903269345  
BLEU-2 score: 0.7955778532875366  
BLEU-3 score: 0.7345549424282534  
BLEU-4 score: 0.6863329636040714

Meteor score: 0.5742548079549654

BERT score: 0.9165949821472168

C.



We ran it for 5 epochs. Training and validation loss both are decreasing at a very good pace. Model is not overfitting.

BLEU-1 score: 0.8279538607142208

BLEU-2 score: 0.728682425223939

BLEU-3 score: 0.6487978033927436

BLEU-4 score: 0.589032636540817

Meteor score: 0.5064402745439278

Drive link for trained models:

<https://drive.google.com/drive/u/2/folders/1efnWFsBPrmR1c7kznAwhlkkZVero5QvN>

### Contributions:

Task 1 - Richa and Sahil

Task 2- Vartika and Prakhar