

NLP ASSIGNMENT-3 REPORT

PART-1: TEXT SIMILARITY

a) BERT Model for Text Similarity Task:

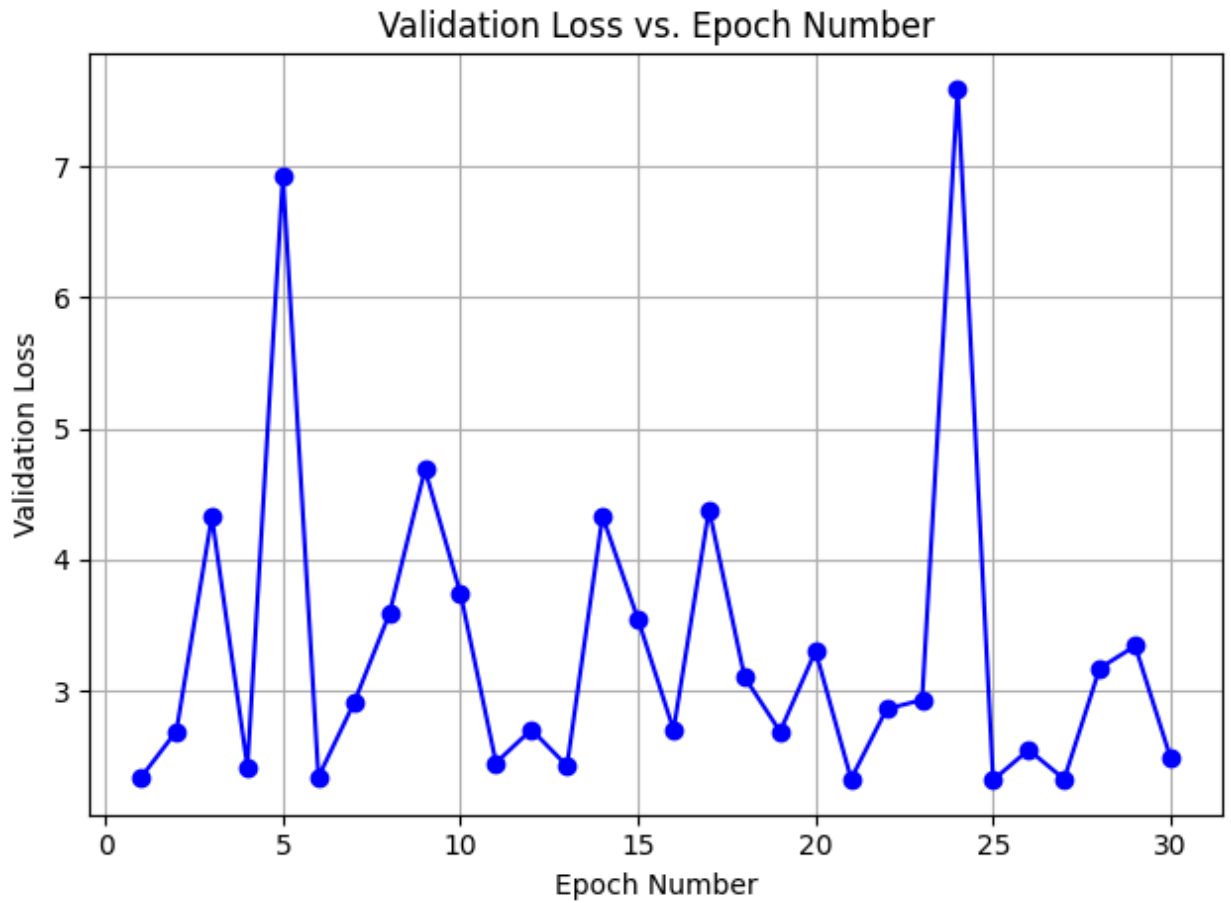
The Google drive link to the saved models are:

https://drive.google.com/drive/folders/1RXJ5WTc-eaTpK8e3WolQ_v5Ac_-cSp5p?usp=s_haring

Training loss vs epoch graph:



Validation Loss vs Epoch:



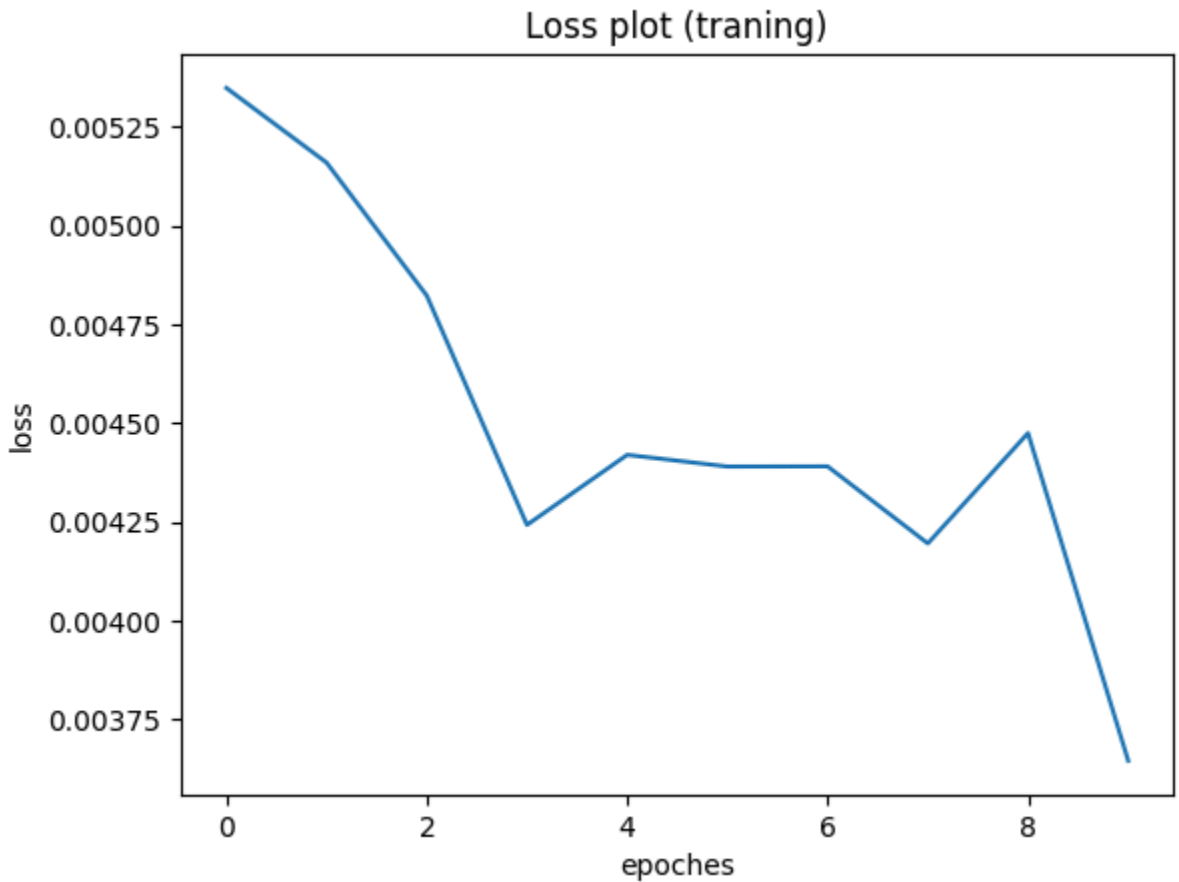
The value of the Pearson correlation metric is:

b) Sentence-BERT Model and SentenceTransformers Framework:

Pearson correlation coefficient:

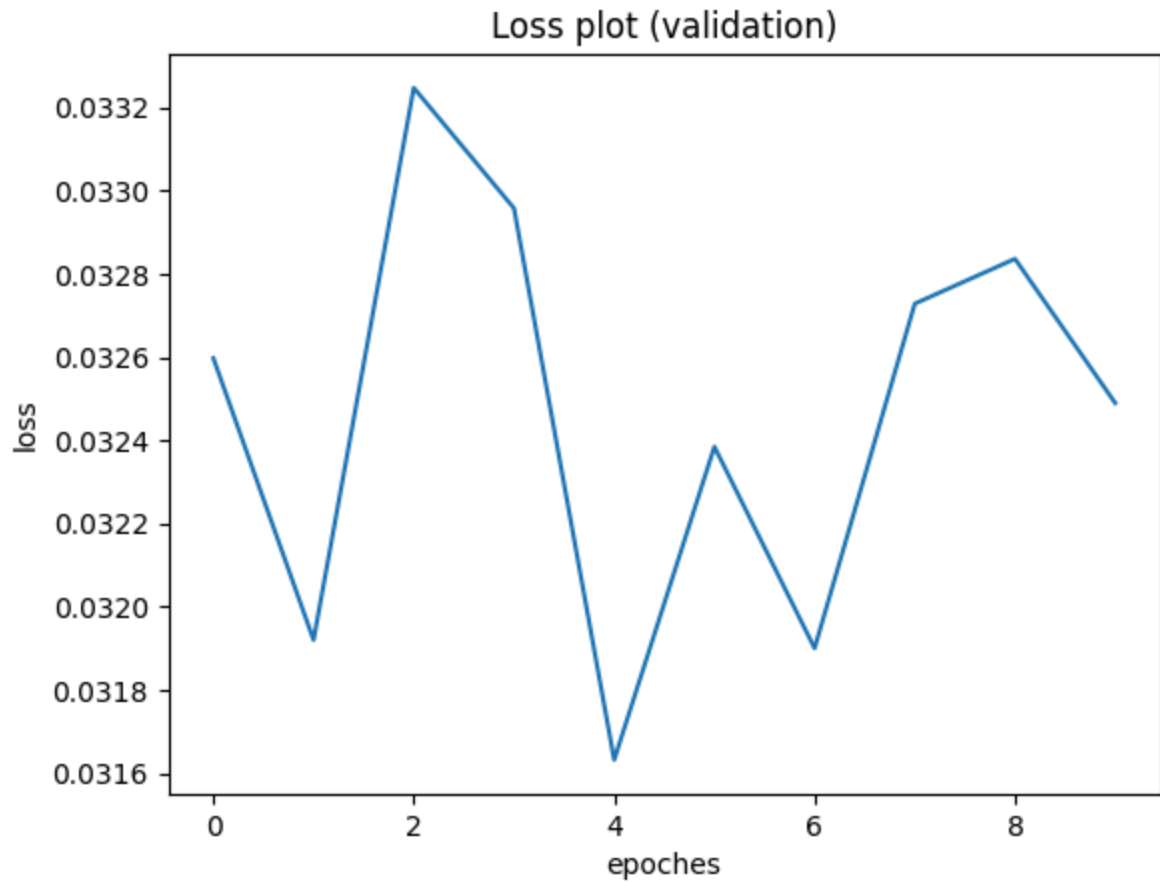
statistic=0.8155824666904162, pvalue=4.339986148814607e-259

c) Fine-tuning SentenceBERT for STS Task:



as the epoch increase the loss decrease first and then increase and then decrease again

At the beginning of training, the model is learning patterns and features from the data. After some epochs, the loss may start to increase again. This could happen due to overfitting, where the model starts to memorize the training data instead of generalizing well to unseen data. At last decrease could be a result of the model learning more generalizable features or patterns in the data.



After the epoch model overfits the training data

Pearson correlation coefficient: statistic=0.8303416793059837,
pvalue=8.902318663885784e-277

Google Drive Link for Saved Model Checkpoints:

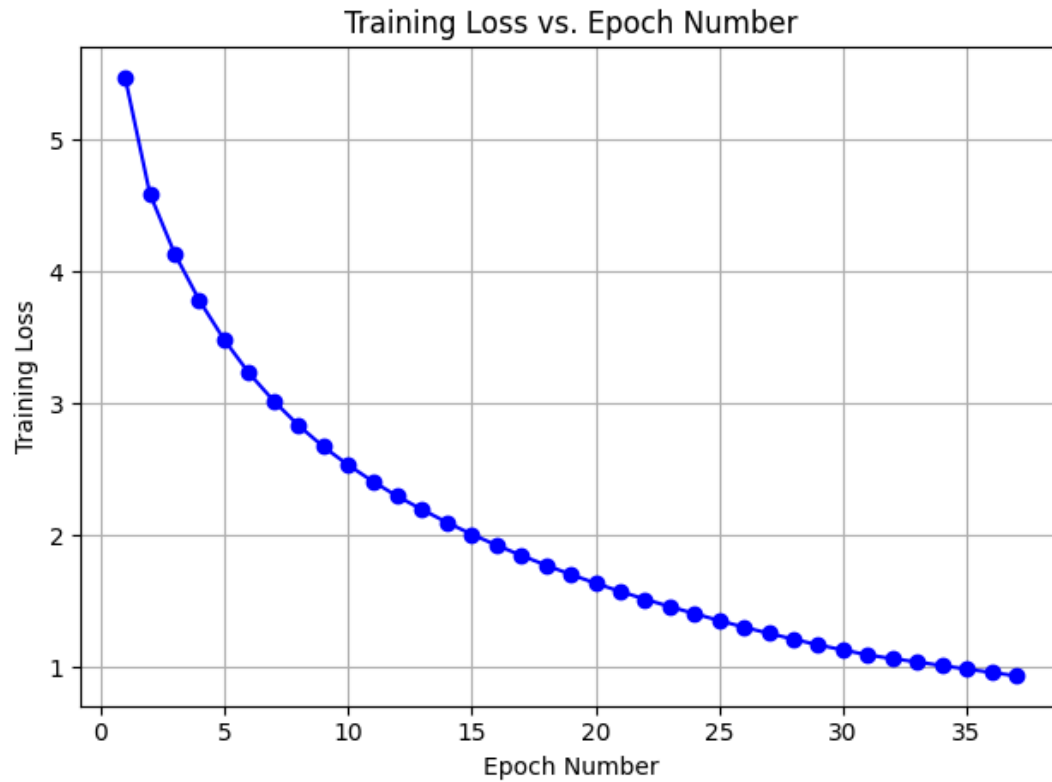
[NLPassign3 - Google Drive](#)

https://drive.google.com/drive/folders/1UiSB0hcMrIeU3O1PItgpMq6jcZvUnQNi?usp=drive_link

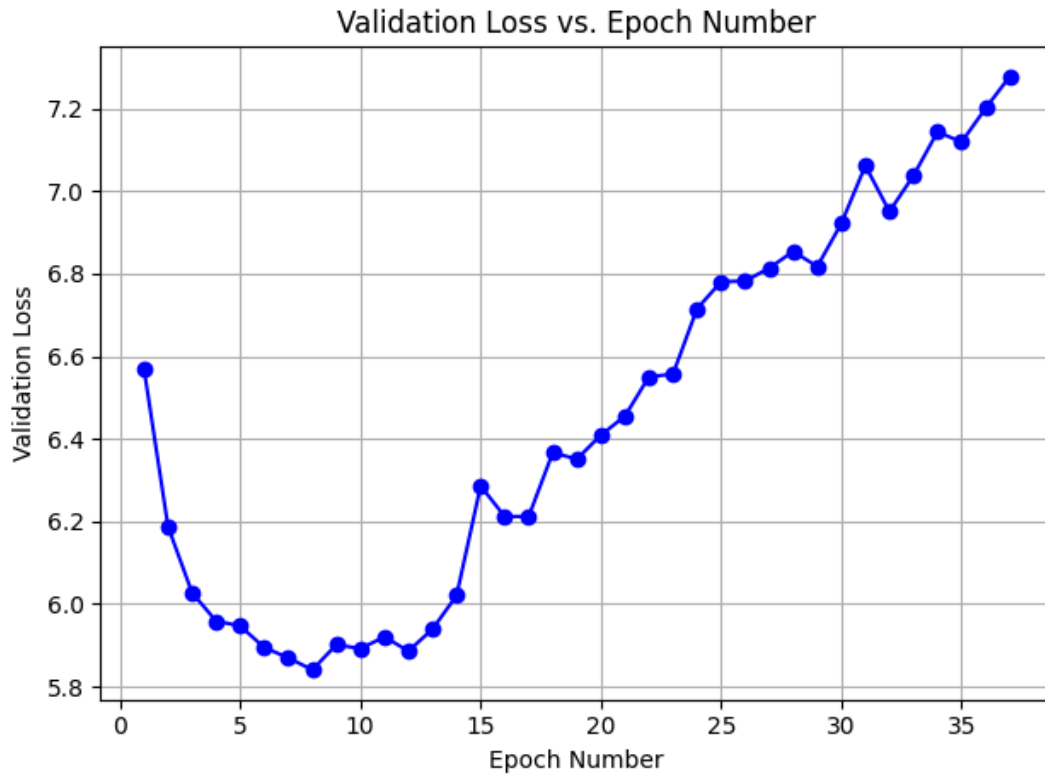
PART-2: MACHINE TRANSLATION

a) Encoder-Decoder Transformer:

Graphs



Nothing Special in the training loss, it decreases with each epoch as expected.



Here the validation loss first decreases indicating that the model is generalizing well but after epoch number 8 the validation loss increases indicating that after epoch 8 the model starts to overfit the training data. Hence Model at epoch 8 is the best one.

Evaluation Metrics

- a) METEOR Score: 0.22941659673628328
- b) Bleu-1 Score: 0.2574172392254841
- c) Bleu-2 Score: 0.1306106188497208
- d) Bleu-3 Score: 0.07239764582103247
- e) Bleu-4 Score: 0.04232843806620156

Link of saved model checkpoints:

https://drive.google.com/drive/folders/10Bq0EnupqOU7GV6sv47voR8yWF_kwTDf?usp=sharing

b) Zero-Shot Evaluation of T5-small model:

Using the pretrained 't5-small' model, we obtained translations from English to German on the WMT16 validation and test datasets. These have been stored in a pickle file, and some samples of the translations are as follows.

Screenshot for sample translations on validation dataset:

Source (English): Mr Modi is on a five-day trip to Japan to strengthen economic ties with the third largest economy in the world.
 Translation (German): Herr Modi ist auf einer fünftägigen Reise nach Japan, um die wirtschaftlichen Beziehungen zur dritten größten Wirtschaftsmacht der Welt zu stärken.
 Source (English): High on the agenda are plans for greater nuclear co-operation.
 Translation (German): Auf der Tagesordnung stehen Pläne für eine verstärkte nukleare Zusammenarbeit.
 Source (English): India is also reportedly hoping for a deal on defence collaboration between the two nations.
 Translation (German): Indien hofft ebenfalls auf ein Abkommen über die Verteidigungszusammenarbeit zwischen den beiden Nationen.
 Source (English): Karratha police arrest 20-year-old after high speed motorcycle chase
 Translation (German): Die Polizei in Karratha verhaftet 20-jährige nach der Hochgeschwindigkeits-Motorradjagd

Screenshot for sample translations on test dataset:

Translation (German): Gleichzeitig sehen 55 Prozent der registrierten Wähler in der ganzen Welt Trump auch nicht als einen guten Führer an.
 Source (English): Both Clinton and Trump, however, get good marks on leadership.
 Translation (German): Sowohl Clinton als auch Trump zeichnen sich jedoch für die Führung aus.
 Source (English): Of the six candidates tested, Biden, Carson and Bush do the best on honesty.
 Translation (German): Von den sechs getesteten Kandidaten tun Biden, Carson und Bush das Beste in Bezug auf Ehrlichkeit.
 Source (English): More see Sanders as honest than not, but like Carson, many don't have an opinion of him.
 Translation (German): Mehr sehen Sanders als ehrlich als nicht, aber wie Carson haben viele keine Meinung von ihm.
 Source (English): The margin of error for the sample of 351 Democratic primary voters is 6 percentage points.
 Translation (German): Die Fehlermarge für die Stichprobe von 351 demokratischen Wählern beträgt 6 Prozentpunkte.
 Source (English): Ben Carson draws close behind Donald Trump in national poll
 Translation (German): Ben Carson steht hinter Donald Trump in einer nationalen Umfrage
 Source (English): Ben Carson is running close behind Donald Trump in the newest CBS News/New York Times poll
 Translation (German): Ben Carson läuft in der neuesten CBS News/New York Times Umfrage knapp hinter Donald Trump zurück

The sample translations are shown above. Now, we computed the Meteor score, along with Bleu-1, Bleu-2, Bleu-3 and Bleu-4 scores. The values were as follows:

- f) METEOR Score: 0.5904659873087972
- g) Bleu-1 Score: 0.6326
- h) Bleu-2 Score: 0.49457
- i) Bleu-3 Score: 0.39911156
- j) Bleu-4 Score: 0.32721

```
[nick_data] Package omw-1.4 is already up-to-date.
Meteor Score is: {'meteor': 0.5904659873087972}
Bleu-1 Score is: {'bleu': 0.6326077952348977, 'precisions': [0.6340460920882179], 'brevity_penalty': 0.997731557892608, 'length_ratio': 0.997734}
Bleu-2 Score is: {'bleu': 0.49457010679249214, 'precisions': [0.6340460920882179, 0.3875319177529902], 'brevity_penalty': 0.997731557892608, 'length_ratio': 0.997734}
Bleu-3 Score is: {'bleu': 0.3991115621266788, 'precisions': [0.6340460920882179, 0.3875319177529902, 0.2605034584018822], 'brevity_penalty': 0.997731557892608, 'length_ratio': 0.997734}
Bleu-4 Score is: {'bleu': 0.32721372144381805, 'precisions': [0.6340460920882179, 0.3875319177529902, 0.2605034584018822, 0.1807296433975305], 'length_ratio': 0.997734}
```

The pipelines for inference have been constructed in code in Google Colab as instructed for demonstration purposes.

- c) Fine-tuning T5-small for German to English Translation:

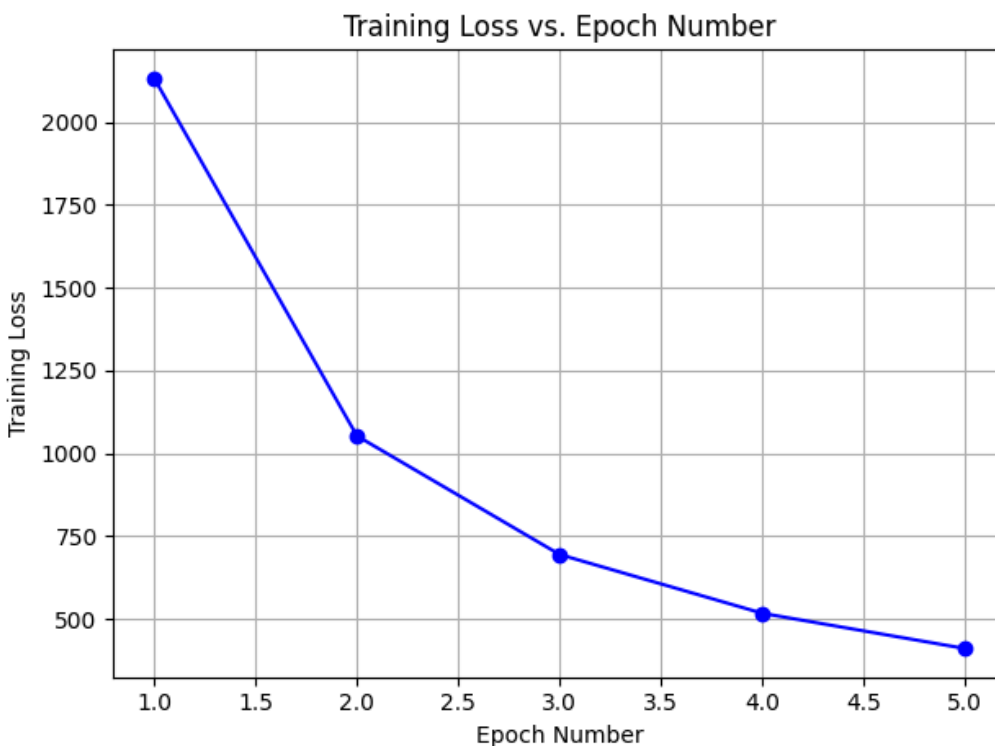
Links to Saved Model Checkpoints:

- 1) https://drive.google.com/file/d/11ToMIZ1X4SQ2isug9_hlb2xosvLZRcl4/view?usp=drive_link

- 2) https://drive.google.com/file/d/1-1PNwdGzcNn59NCrW2hfJPmycJg7PJj9/view?usp=drive_link
- 3) https://drive.google.com/file/d/1-7KVOgiLHc8qxs4tdN4R9AY23zLLgRUN/view?usp=drive_link
- 4) https://drive.google.com/file/d/1-BaVa1qi2iF8JejwbLIS4GyxI13QZuhe/view?usp=drive_link
- 5) https://drive.google.com/file/d/1-CDVUdzXCbCfW-vjqTG90R_f0Rl84Ini/view?usp=drive_link

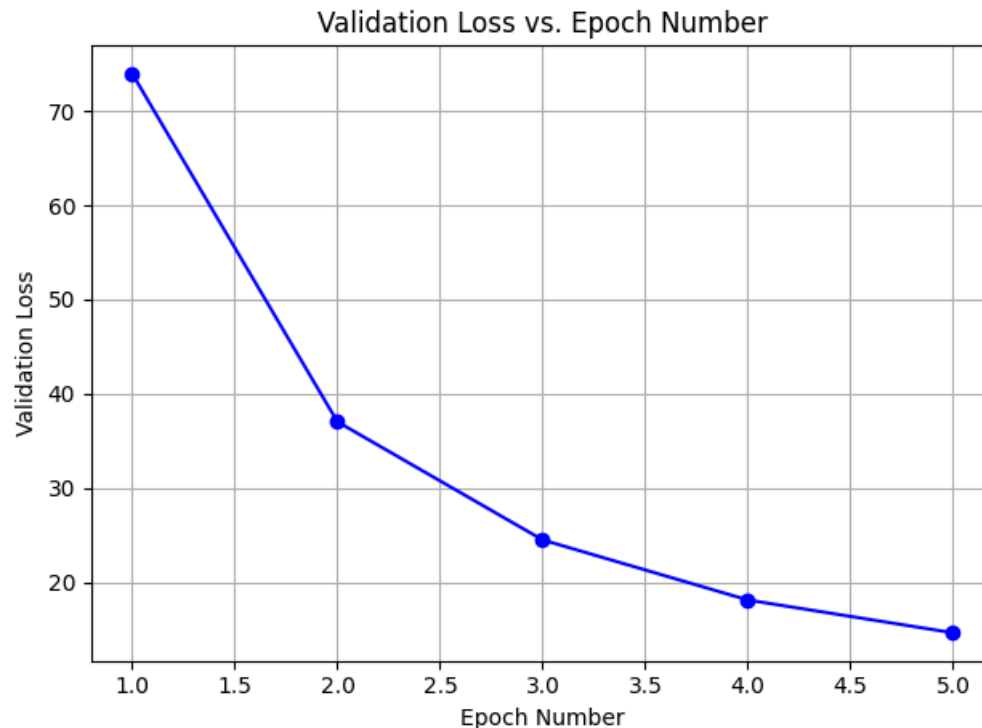
Now, we will be adding the training and validation loss graphs versus number of epochs. We trained this model for 5 epochs, and trained the first encoder (Self attention) layer of the t5-small model.

Training Loss vs Epochs Graph:



In this graph, we see the expected behavior of the training loss decreasing at a decreasing rate. This means with more epochs, the loss will decline, and the amount of decrease will be lesser with every epoch.

Validation Loss vs Epoch:



In this graph also, we see the validation loss decreasing at a decreasing rate. As the model trains, we run it on the validation set to ensure that the model does not overfit on the training data. As the loss is decreasing, we can say that overfitting is not happening so far. As the number of epochs will increase, the data will overfit, and the validation loss would increase. However, here since we are only running 5 epochs, we do not see overfitting come into play (training more epochs was a bit difficult as training was taking a lot of time).

Sample Translations from German to English on Validation Dataset:

```

Source (German): George Webster, 28, wurde die Anklage bei einer Anhörung von dem Obersten Gericht in Glasgow verlesen.
Translation (English): George Webster, 28, a été slée à l'occasion d'une a hearing of the Ob First Court in Glasgow.
Source (German): Er wird beschuldigt, am 7. Juni 2013 eine Frau im Scotland's Hotel in Pitlochry in Perthshire vergewaltigt zu haben.
Translation (English): Er wird schuldigt er, daß am 7. Juni 2013 eine Frau in Scotland's Hotel in Pitlochry in Perthshire walte hat.
Source (German): Die Anklage lautet, dass Webster sie angriff, während sie "bewusstlos war, schlief, und kein Einverständnis signalisieren konnte."
Translation (English): Die Anklage lautet, dass Webster sie griff, während sie 'wiss, schlief, and no agreement signaled'.
Source (German): Webster wird darüber hinaus vorgeworfen, am 4. Mai 2014 eine zweite Frau im Golf View Hotel in Naim im schottischen Hochland vergewaltigt zu haben.
Translation (English): Webster wird darüber hinaus vorworfen, am 4. Mai 2014, eine zweite Frau in Golf View Hotel in Naim, in der Schottischen Hochland, zu haben.
Source (German): Richterin Lady Rae setzt den Verhandlungstermin für den 17. November am Obersten Gericht in Edinburgh an.
Translation (English): Richterin Lady Rae setzt die Verhandlungstermin für 17 November am Obersten Gericht in Edinburgh.
Source (German): Rückbesinnung auf das sehr amerikanische Ideal der Arbeitsrechte als Menschenrechte
Translation (English): Restitution based on the very American ideal of working rights as a human rights based on the very American ideal of human rights.
Source (German): Die Kongressabgeordneten Keith Ellison und John Lewis haben einen Gesetzesvorschlag eingebracht, um die Organisation von Gewerkschaften als Bürgern
Translation (English): Keith Ellison and John Lewis have introduced a legislative proposal to establish the organisation of unions as citizens rights.
Source (German): "So wie Gewerkschaften sterben, sterben auch die Mittelklassejobs," sagte Ellison, ein Demokrat aus Minnesota und stellvertretender Vorsitzender de
Translation (English): "So as unionssterben,sterben also the Mittelklassejobs," said Ellison, a Demokrat aus Minnesota and a Vice President of Progressive Caucus in
Source (German): Daher stelle ich stolz gemeinsam mit der Bürgerrechtsikone John Lewis das Mitarbeiterermächtigungsgesetz vor.

```

Sample Translations on test dataset:

Translation (English): Es sind hauptsächlich Republikaner, wenn als sechs von 10, die sagen, es ist schlechter geworden, ihre Ansichten Clinton.
Source (German): Obwohl Clinton von den Wählern ihrer Partei als ehrlich und vertrauenswürdig angesehen wird, ist dies weiterhin ein Problem für sie unter eingetrag
Translation (English): Obwohl Clinton von den Wählern ihrer Partei als ehrlich und vertrauenswürdig betrachtet wird, ist dies weiterhin ein Problem für die Wähler, die
Source (German): 62 Prozent der Wahlberechtigten halten sie nicht für eine ehrliche und vertrauenswürdige Person.
Translation (English): 62 % of the electables think that they are not a genuine and trustworthy person.
Source (German): Zur gleichen Zeit sehen 55 Prozent der Wahlberechtigten landesweit auch Trump als nicht ehrlich und nicht vertrauenswürdig an.
Translation (English): In the same time, 55 % of the electorates landeswide also Trump as not honest and unworthy.
Source (German): Clinton und Trump erhalten jedoch gute Noten bei Führung.
Translation (English): Clinton et Trump obtain gute Notens in Führung.
Source (German): Von den sechs getesteten Kandidaten schneiden Biden, Carson und Bush am besten bei Ehrlichkeit ab.
Translation (English): Von the six tested candidates schneiden Biden, Carson and Bush best in Ehrlichkeit.
Source (German): Eine Mehrheit sieht Sanders eher als ehrlich an, aber wie bei Carson haben viele keine Meinung von ihm.
Translation (English): a majority sees Sanders eher als ehrlich, but as Carson has many no opinion of him.
Source (German): Die Fehlerquote für die Stichprobe von 351 Demokratischen Wählern bei den Vorwahlen beträgt 6 Prozentpunkte.
Translation (English): Die Fehlerquote für die Stichprobe von 351 Demokrat Wähler bei den Vorwahlen beträgt 6 %.
Source (German): Ben Carson rückt bei landesweiter Umfrage nahe hinter Donald Trump auf
Translation (English): Ben Carson rückrückt in landesweiter Umfrage near hinter Donald Trump in a position in a position in a position in a position in a position i
Source (German): Ben Carson befindet sich bei der neuesten Umfrage von CBS News/New York Times dicht hinter Donald Trump
Translation (English): Ben Carson is behind Donald Trump in the new question of CBS News/New York Times dicht hinter Donald Trump.
Source (German): Auf die beiden Kandidaten entfällt inzwischen die Hälfte der Unterstützung der Wähler in der neuen Umfrage, und anderen verlieren dramatisch an Hal
Translation (English): The two candidates entfällt inzwischen half of the support of the elector in the new question - and the other loses dramatically Hal

The computed metrics are as follows:

- a) METEOR Score: 0.322533
- b) Bleu-1 Score: 0.375224
- c) Bleu-2 Score: 0.24312
- d) Bleu-3 Score: 0.16746
- e) Bleu-4 Score: 0.11874

Comparing Performance of 2 models

As we can see Performance of the “T5-small” model, which has been fine-tuned for our task, is much better in comparison to the encoder-decoder transformer model that's been trained from scratch. The reason for this is as follows:

- 1) The amount of data that's used for training the encoder-decoder transformer model from scratch is very less in comparison to the amount of data T5 was pre trained on. As we know deep networks require large amounts of data to provide good results, the encoder-decoder transformer model performs poorly.
- 2) Also T5 model was initially pretrained so it understands how the language works, whereas there was no such training in case of encoder-decoder transformer model.
- 3) Also the size of the T5 model is larger than the encoder-decoder transformer model.

REFERENCES:

- 1) [BERT \(huggingface.co\)](https://huggingface.co/BERT)
- 2) [SentenceTransformers Documentation — Sentence-Transformers documentation \(sbert.net\)](https://huggingface.co/SentenceTransformers)
- 3) <https://www.google.com/url?q=https://www.sbert.net/examples/training/sts/README.html&sa=D&source=apps-viewer-frontend&ust=1712592807806006&usg=AOvVaw11pPhPgbTacU1I6Krucj4p&hl=en>
- 4) [Load \(huggingface.co\)](https://huggingface.co/Load)

- 5) [Language Translation with nn.Transformer and torchtext — PyTorch Tutorials 2.2.1+cu121 documentation](#)
- 6) [T5 \(huggingface.co\)](#)
- 7) [Translation \(huggingface.co\)](#)
- 8) [BLEU - a Hugging Face Space by evaluate-metric](#)
- 9) [METEOR - a Hugging Face Space by evaluate-metric](#)
- 10) [BERT Score - a Hugging Face Space by evaluate-metric](#)
- 11)

CONTRIBUTIONS:

Part 1-a) Jay Saraf, Vatsal Chaudhary

Part 1-b,c) Harsh

Part 2-a) Vatsal Chaudhary

Part 2-b,c) Rohan Gupta

Members:

- 1) Harsh (2020061)
- 2) Jay Saraf (2020438)
- 3) Rohan Gupta (2020113)
- 4) Vatsal Chaudhary (2020549)