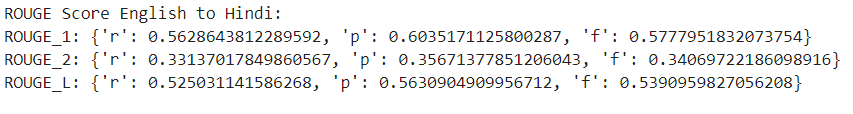
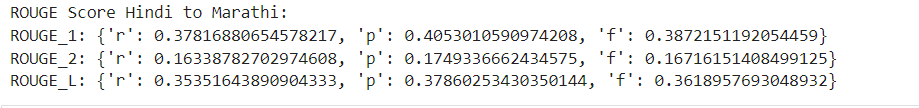
NLLB Scores:

|  |  |  |
| --- | --- | --- |
| Translation | Corpus level bleu Score | Sentence level bleu Score |
| English to Hindi | 0.62 | 0.31 |
| Hindi to English | 0.66 | 0.38 |
| Hindi to Marathi | 0.5501 | 0.20 |
| Marathi to Hindi | 0.5505 | 0.23 |



A number on a white background

Description automatically generated



A number on a white background

Description automatically generated

IndicTrans Scores:

|  |  |  |
| --- | --- | --- |
| Translation | Corpus level bleu Score | Sentence level bleu Score |
| English to Hindi | 0.69 | 0.35 |
| Hindi to English | 0.75 | 0.44 |
| Hindi to Marathi | 0.605 | 0.21 |
| Marathi to Hindi | 0.608 | 0.27 |

A number on a white background

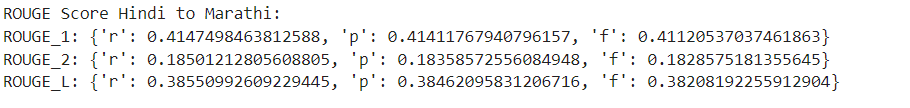
Description automatically generated

A number on a white background

Description automatically generated

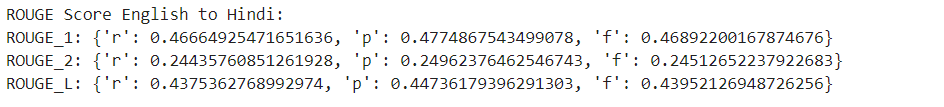
A number on a white background

Description automatically generated



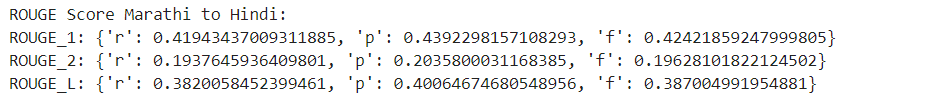
ChatGPT Scores:

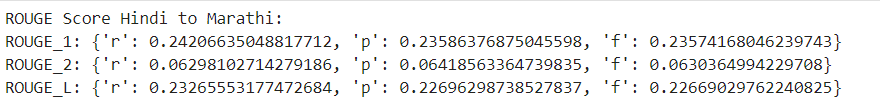
|  |  |  |
| --- | --- | --- |
| Translation | Corpus level bleu Score | Sentence level bleu Score |
| English to Hindi | 0.56 | 0.24 |
| Hindi to English | 0.67 | 0.38 |
| Hindi to Marathi | 0.45 | 0.12 |
| Marathi to Hindi | 0.50 | 0.18 |



A number on a white background

Description automatically generated





By doing this assignment I explore various Translation model and learn how to use them. I learn that we can BLEU Scores is a metric which is used to check the quality of translation, it is of two type one is corpus level, and another is sentence level.When we want to check the quality of each and every sentence in that case we should use sentence level bleu score because it calculates score for each sentence, whereas corpus level focus on whole translated corpus.

As we can see that generally corpus level score is better than sentence level because I explained sentence level is more rigorous.

IndicTrans outperform both NNLB and ChatGPT in terms of blue scores and reason is obvious that IndicTrans is pre-trained on Indian languages so it have better understanding of Indian languages , but because we use only subset of dataset it is not sure that IndicTrans is better than other two models.

In terms of Rouge scores all models are competing with each other but still IndicTrans result is better than other two , because of the same reason as I already mentioned.

Here NLLB is giving better results than ChatGPT, one of the possibilities can be because ChatGPT results are based on only 50 sentences so result is lower than NLLB which was calculated for 1000 sentences.