**NLP Project Documentation**

1. Summarization. :-

We used the Text rank summary model which had an accuracy of 70% and the TF-IDF model which had an accuracy of approximately 60%

Dataset:

https://www.kaggle.com/code/lusfernandotorres/text-summarization-with-large-language-models/input

1. Translation. :-

We used the LSTM deep learning model which had an accuracy of 75%.

Dataset: ara.txt

https://www.manythings.org/anki/

1. Named Entity Recognition (NER). :-

We used the CRF (Conditional Random Fields) model which had a 92% accuracy.

Dataset used: NER\_dataset.csv

<https://www.kaggle.com/datasets/namanj27/ner-dataset>

1. Sentiment Analysis:

We used the naïve Bayes algorithm which had an accuracy of 84% and SVM which had an accuracy of 82%.

Dataset used: <https://drive.google.com/file/d/1PbhF3Njd0nFj8JlQUdBI21ocTpZROmp5/view?usp=drive_link>