

## NLP Assignment-4

Firstly, I have done the preprocessing which includes lowering the words and then removing the space and punctuation and later removing the stop words and then do the tokenization.

Then I have make the vocab by removing the words which are not significant and have also give each words unique index and vice-versa. Then split the data into train, test and validation.

Finding the embeddings of the dataset.

Finally we have calculated the Confusion matrix and other parameters like F1 score, recall value & precision score.

Precision is the ratio of true positives to the total number of predicted positives, while recall is the ratio of true positives to the total number of actual positives. The F1 score is the harmonic mean of precision and recall, and ranges between 0 and 1, with a higher score indicating better performance.

For the sentiment analysis dataset:

F1 score is :0.5357737104825291

Recall Score is :0.5796579657965797

Accuracy is:0.49502262443438916

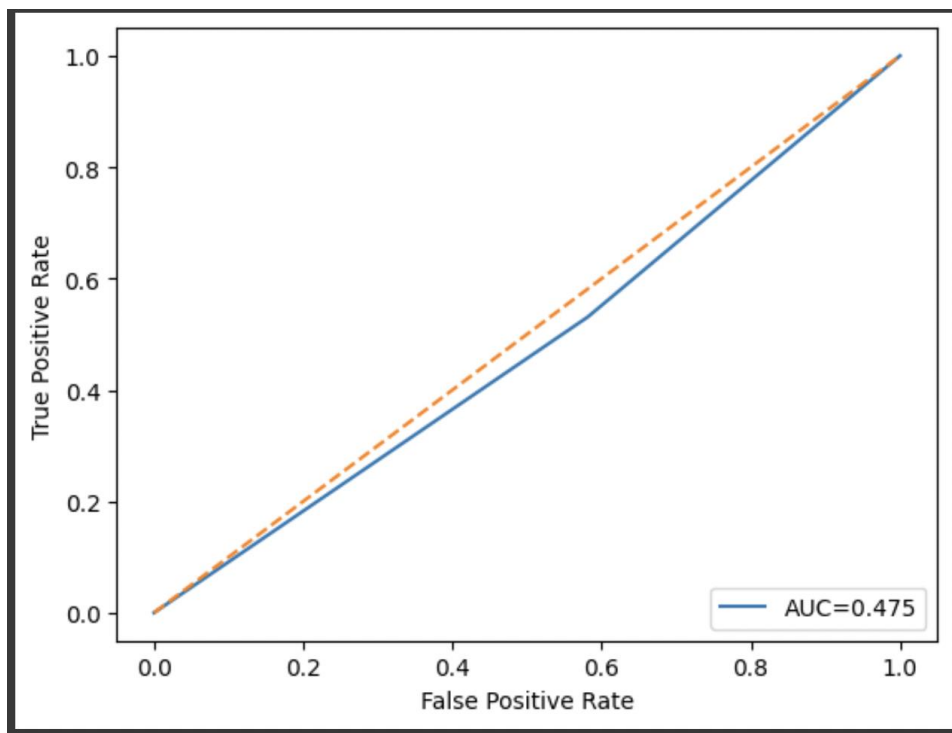
Precision is :0.4980665119876257

A confusion matrix is a table used to evaluate the performance of a machine learning model for a classification problem. It summarizes the predictions made by a model on a set of test data, comparing the predicted class labels with the true class labels.

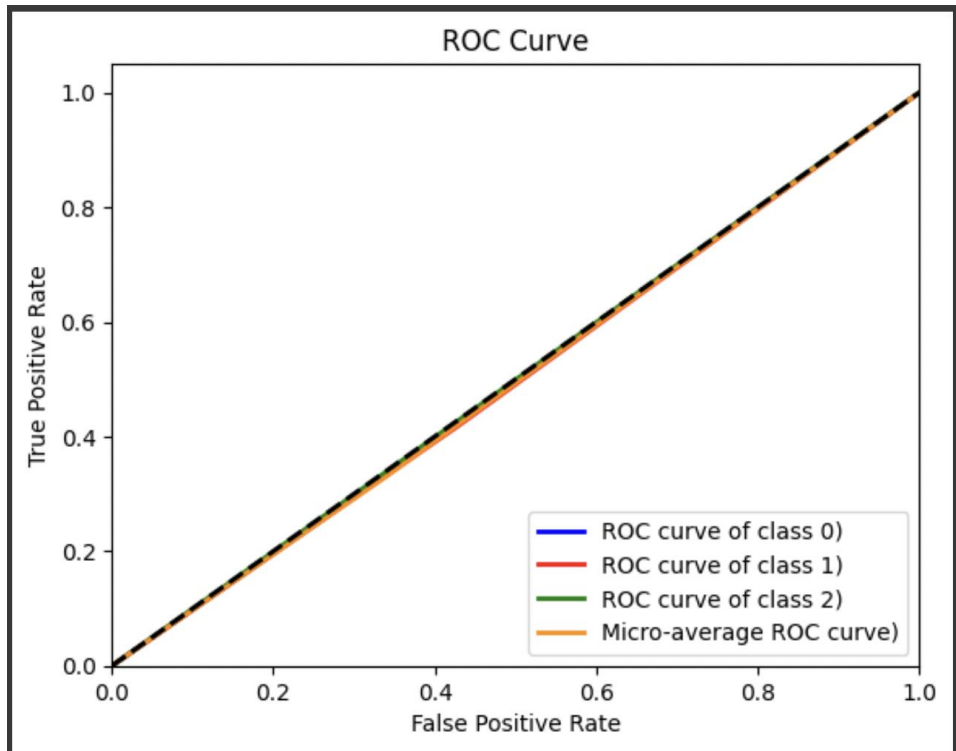
Evaluated Confusion matrix is :

[[450 649]

[467 644]]



For the second dataset



Model & embedding links of google drive:-

Model- [https://drive.google.com/file/d/1--0jk8YQppjd6vd3IRBQzdsQEysGGjIB/view?usp=share\\_link](https://drive.google.com/file/d/1--0jk8YQppjd6vd3IRBQzdsQEysGGjIB/view?usp=share_link)

Model:- [https://drive.google.com/file/d/18FGGiGdFh0g7tjFSd-zY8KMqistblyHI/view?usp=share\\_link](https://drive.google.com/file/d/18FGGiGdFh0g7tjFSd-zY8KMqistblyHI/view?usp=share_link)

Embeddings:

[https://drive.google.com/file/d/1xmtOxPiIJ0eNrrdCk4Cpzw1MT5wqyv2T/view?usp=share\\_link](https://drive.google.com/file/d/1xmtOxPiIJ0eNrrdCk4Cpzw1MT5wqyv2T/view?usp=share_link)