

Pipeline:

First I loaded the IMDB dataset from the “datasets” library. This dataset contains reviews about movies and whether they are positive reviews or negative reviews. Then for preprocessing, I just removed the HTML tags present in the reviews(if any special character is saved as HTML tags). After that I loaded the BERT model and tokenizer to FineTune the dataset. After training was completed, I calculated the Accuracy and F1 score of the Test dataset and then saved the model. Then loaded the saved model for predicting sentiment on an independent review.

This was the basic pipeline I created using the resources provided. The major problem was that since the training set was very large with big reviews, the training took a lot of time. I chose GPU runtime and only 3 epochs but still it took over 2 hours to train