I have used the DistilBERT model , which is a smaller version of BERT model which is trained using the Knowledge Distillation method. I haven’t experimented much by trying different models,since this is already a good model. The Evaluation metrics is just calculating the accuracy of the model and I got an accuracy of 0.932 for the training dataset and 0.942 for the test dataset.

I couldn’t tune much and experiment with different parameters due to time constraints, i just trained with different learning rate and max length of tweet.