Problem:
Implement a deep neural network for a sentiment classification task. The model should achieve a minimum 80% F-score or higher.

Dataset:
We use IMDB Review Sentiment Analysis Dataset which contains movie reviews and their corresponding sentiment labels (i.e. 'positive' and 'negative'). The dataset contains 49582 reviews and we divide it into train (39723), test (4995), and development (4998) sets.

Model Architecture:
The classifier model uses a pre-trained BERT model from huggingface library for getting the word embeddings. A dropout layer is added to avoid overfitting followed by a linear output layer. The linear layer outputs a probability distribution over the two classes, namely 'positive' and 'negative'. We consider the class with the higher probability as the prediction of the model. To train the model we use cross entropy loss and Adam optimizer.

Hyperparameters:
We use the following values for the hyperparameters during training:

Adam Learning Rate	2e-5
Training batch size	16
Training epochs	4
dropout	0.3

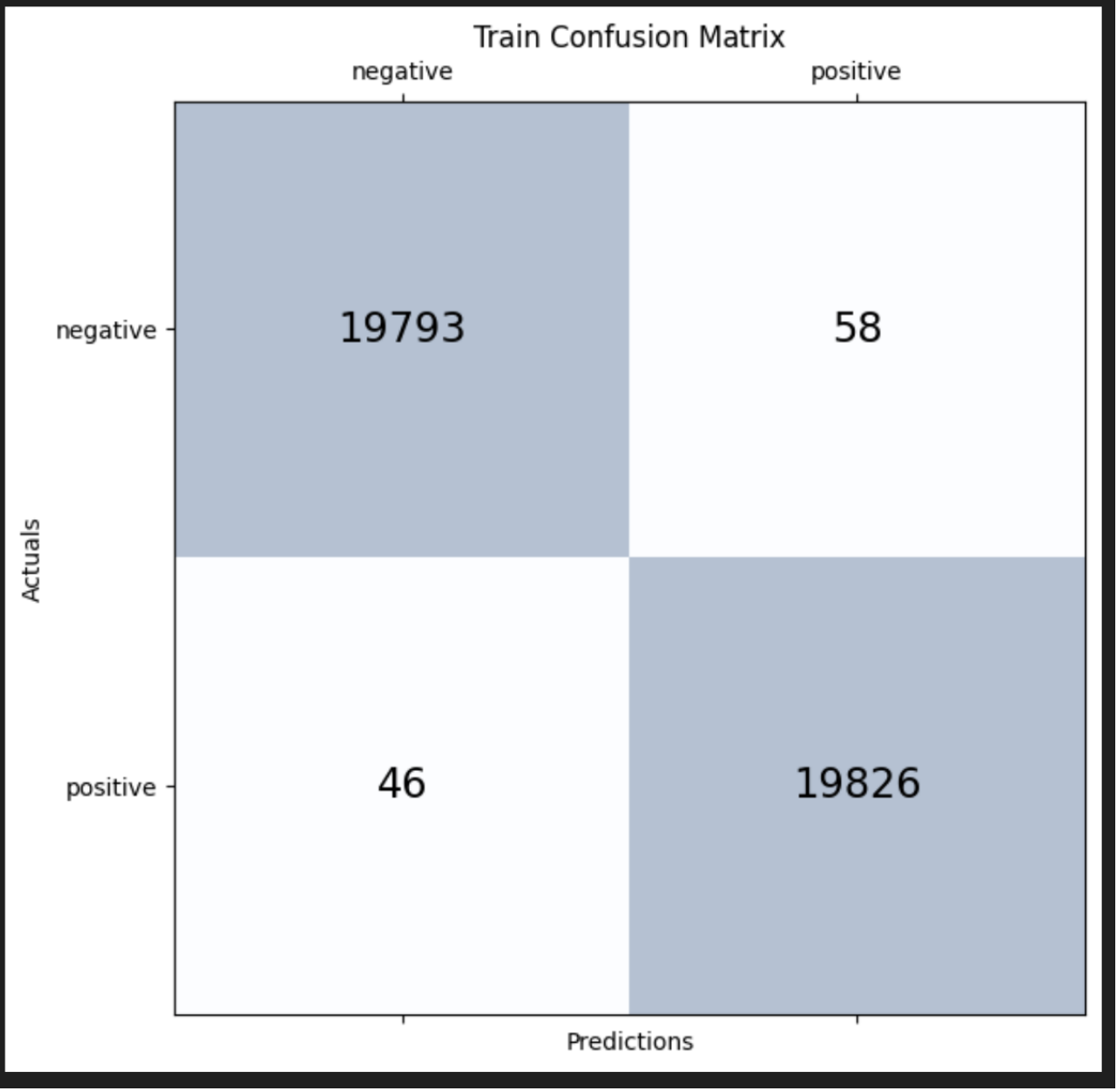
Performance:
The trained model is evaluated on the train, test, and dev sets. The results are reported here

Dataset	**F-score**
train	1
test	0.93
dev	0.93

Train set F-score

F-Score

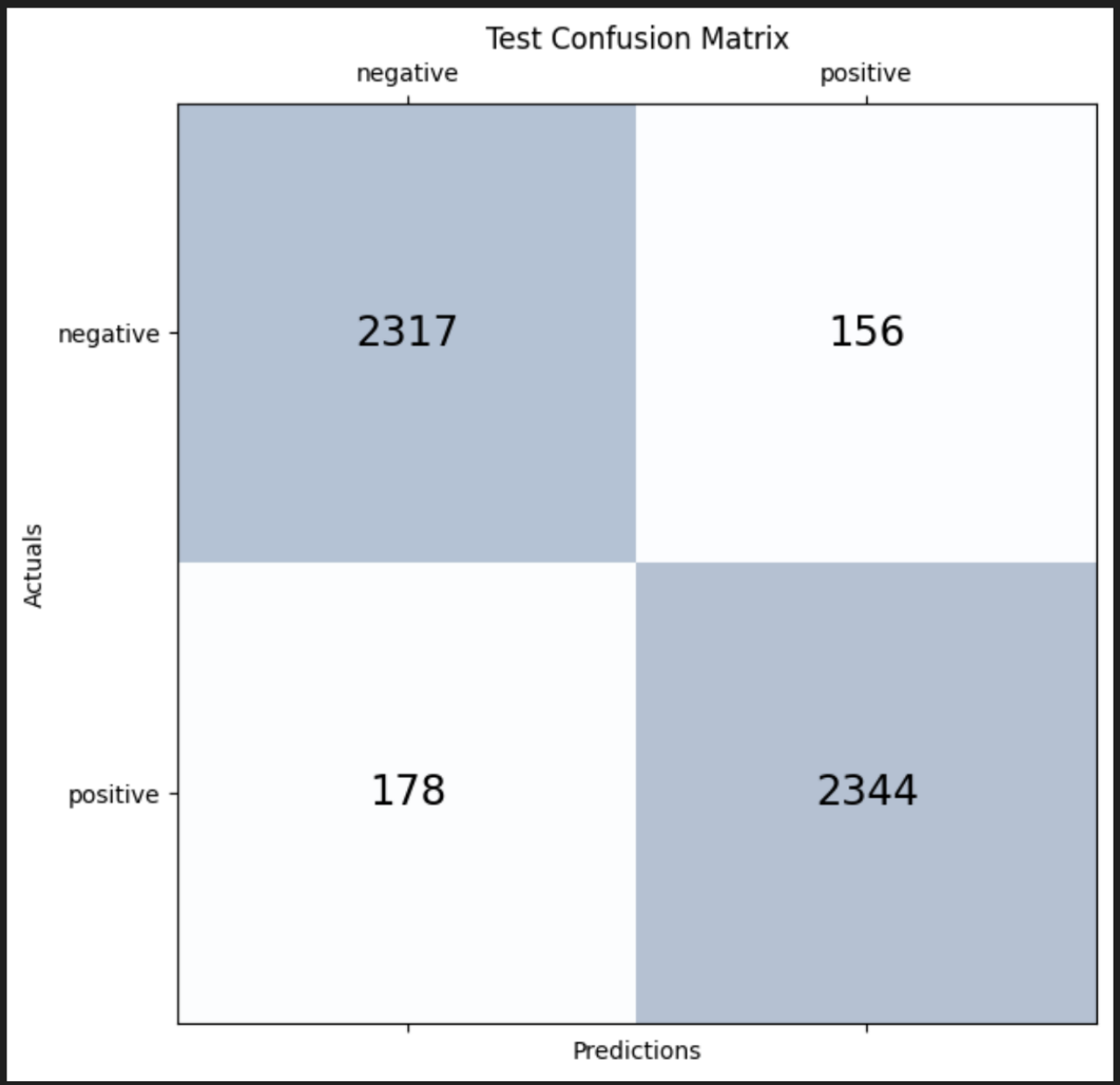
	precision	recall	f1-score	support
negative	1.00	1.00	1.00	19851
positive	1.00	1.00	1.00	19872
accuracy			1.00	39723
macro avg	1.00	1.00	1.00	39723
weighted avg	1.00	1.00	1.00	39723



Test set F-score

F-Score

	precision	recall	f1-score	support
negative	0.93	0.94	0.93	2473
positive	0.94	0.93	0.93	2522
accuracy			0.93	4995
macro avg	0.93	0.93	0.93	4995
weighted avg	0.93	0.93	0.93	4995



Development set F-score

F-Score

	precision	recall	f1-score	support
negative	0.93	0.94	0.94	2522
positive	0.94	0.93	0.93	2476
accuracy			0.93	4998
macro avg	0.93	0.93	0.93	4998
weighted avg	0.93	0.93	0.93	4998