Natural Language Processing (CS60075)

Assignment 2: News Topic Classification

Name: Suthar Utsav Kaushikkumar

Roll No.: 23CS60R49

TASK 2: News Topic Classification using BERT Model

Challanges Faced During Implementation:

1) Using only the computer's regular power (CPU), it took about 10 minutes to teach the model. But when we added a special graphics card (GPU), it trained in less than a minute.
2) I tried making the transformer model myself to learn, but it took a lot of time and research. Unfortunately, it didn't work well, so I switched to using a prebuilt model.
3) When we split the text into smaller parts (tokenizing), the model performed differently depending on whether we used an attention mask or not. Finally, we chose the model with an attention mask because it's important for tasks like classification.
4) We tried out different maximum sequence lengths to see what worked best through trial and error.

Hyperparameters used for the Task:

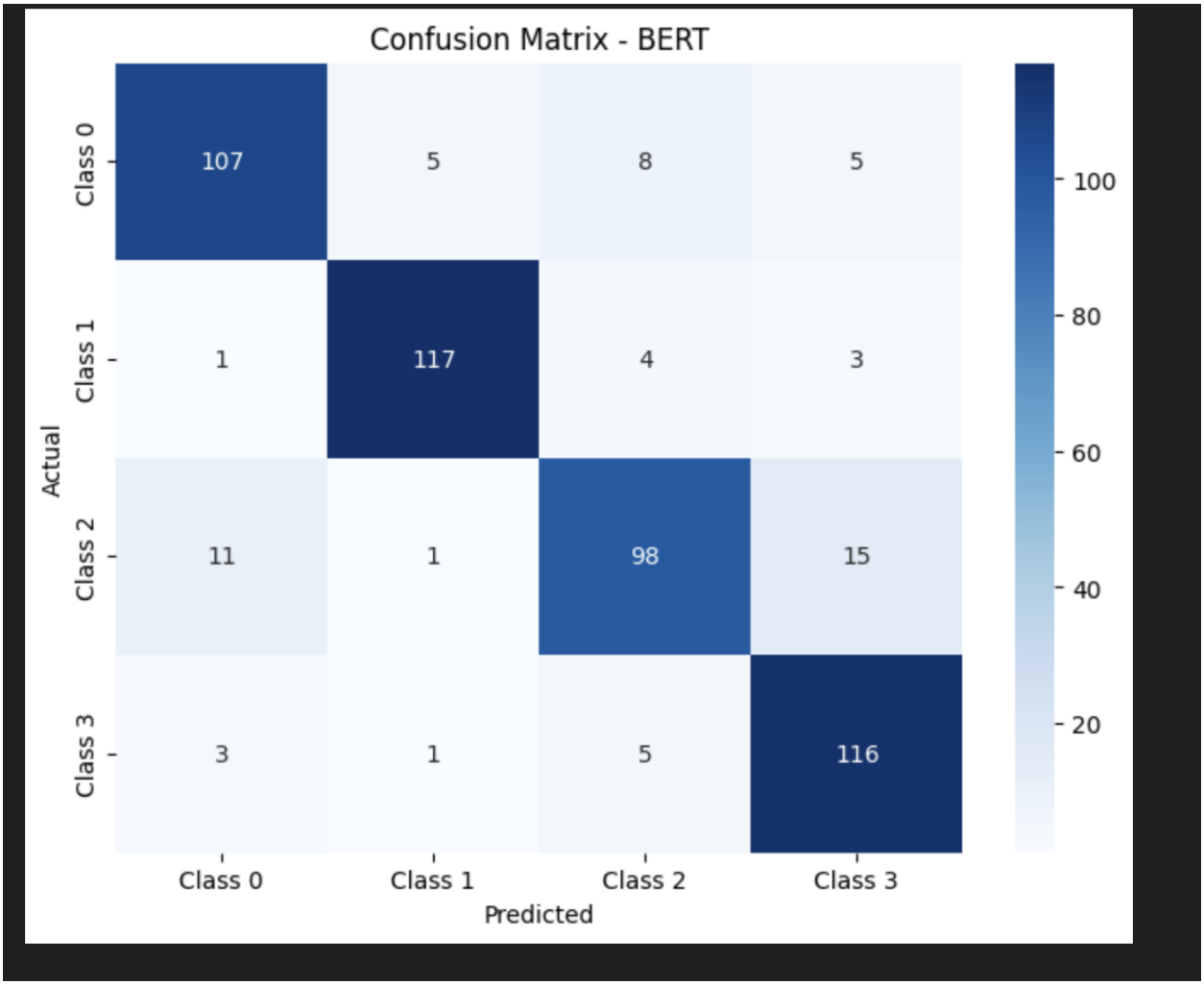
1) Max Sequence Length: 64
2) Model: roberta-base
3) Batch Size: 64
4) Truncation = True
5) Loss Function: Cross Entropy Loss
6) Optimizer: Adam
7) Learning Rate: 0.00002
8) Number of epochs = 5

Performance of the BERT Model:

1) Training Accuracy BERT: 0.9633
2) validation Accuracy BERT: 0.9150
3) Testing Accuracy BERT: 0.8760

4) Test F1 Score BERT: 0.8754

Confusion Matrix Bert



	precision	recall	f1-score	support
0	0.88	0.86	0.87	125
1	0.94	0.94	0.94	125
2	0.85	0.78	0.82	125
3	0.83	0.93	0.88	125
accuracy			0.88	500
macro avg	0.88	0.88	0.88	500
weighted avg	0.88	0.88	0.88	500