Training BERTimbau

Sentiment Classification Model

BERTimbau

BERT-BASE-PORTUGUESE-CASED

- pre trained in brazilian portuguese

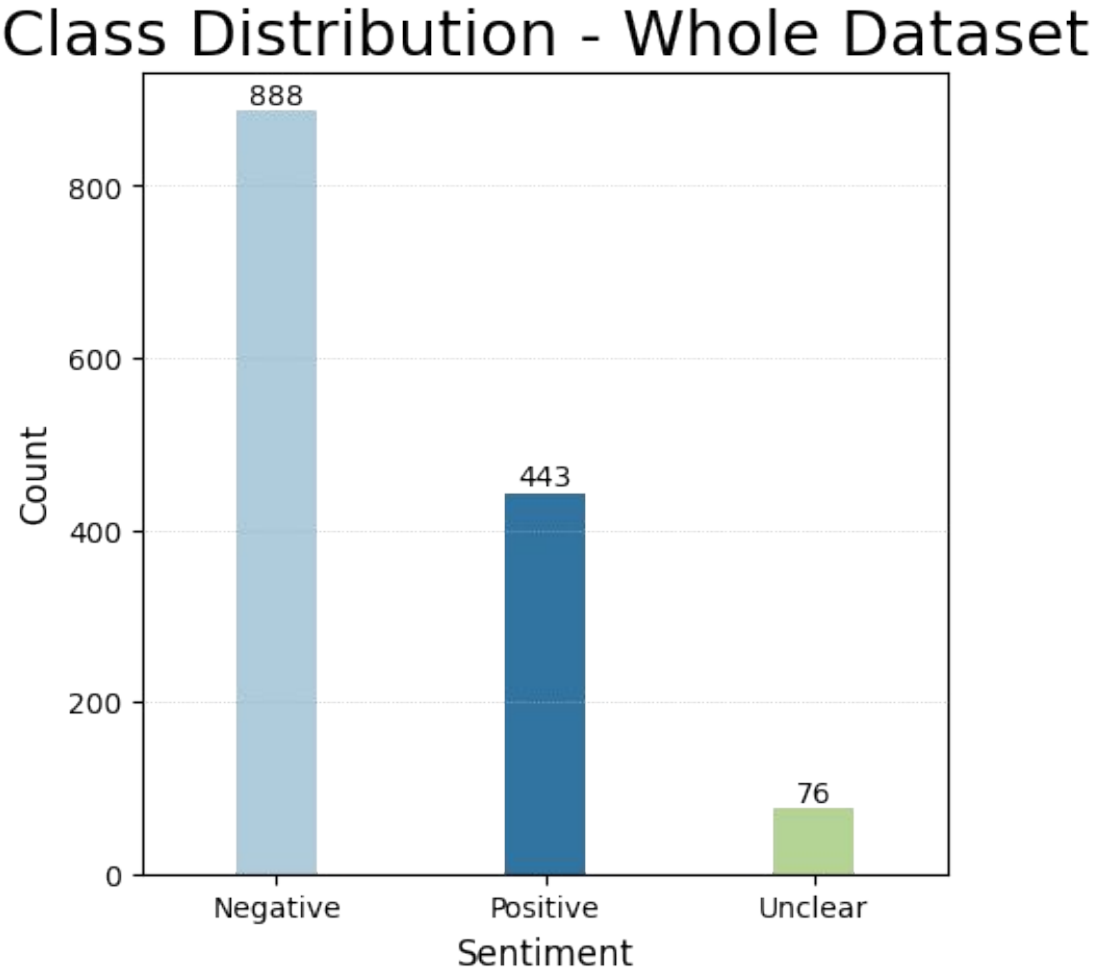
- sentence textual similarity

- recognizing textual entailment

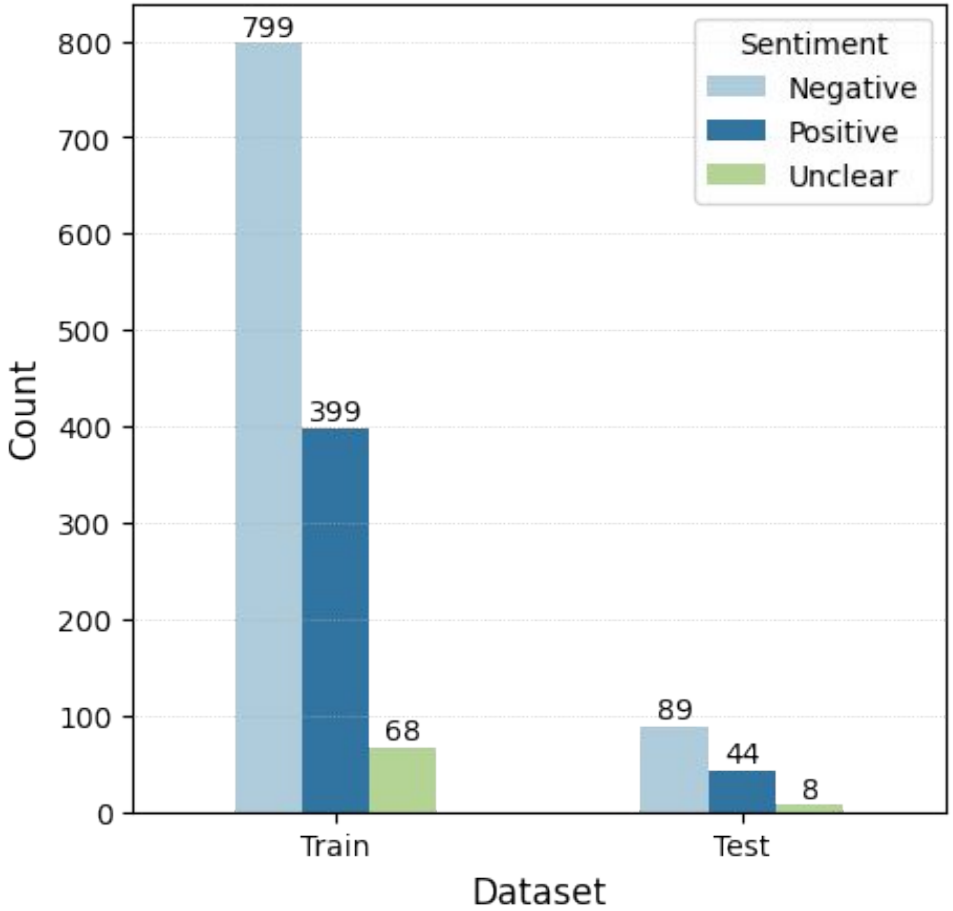
- name entity recognition

– 9:1 TRAIN/TEST SPLIT –

PREPARING THE DATA…


Class Distribution - Whole Dataset
888
443
76
800
600
400
200
0
Count
Negative
Positive
Unclear
Sentiment

Class Distribution - Train/Test Sets

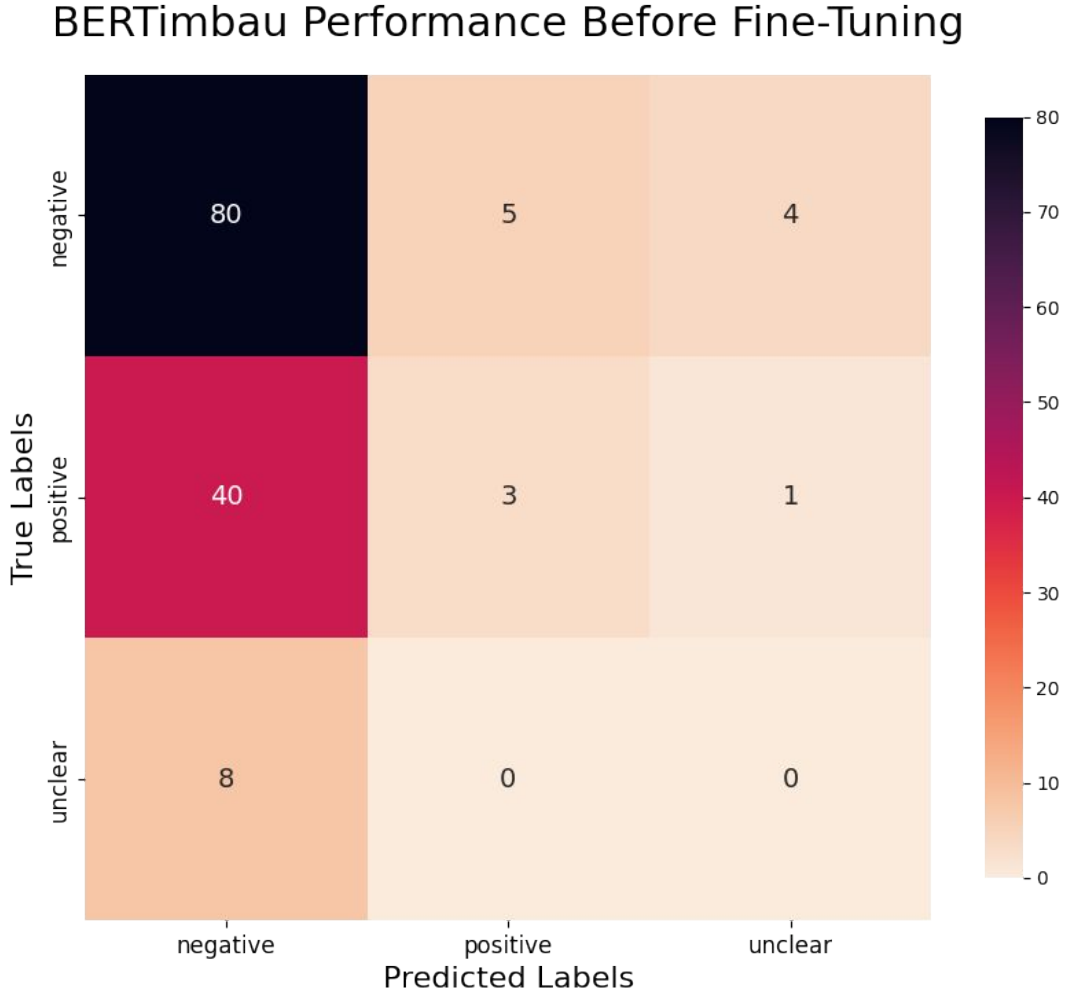


TEST BERTimbau...

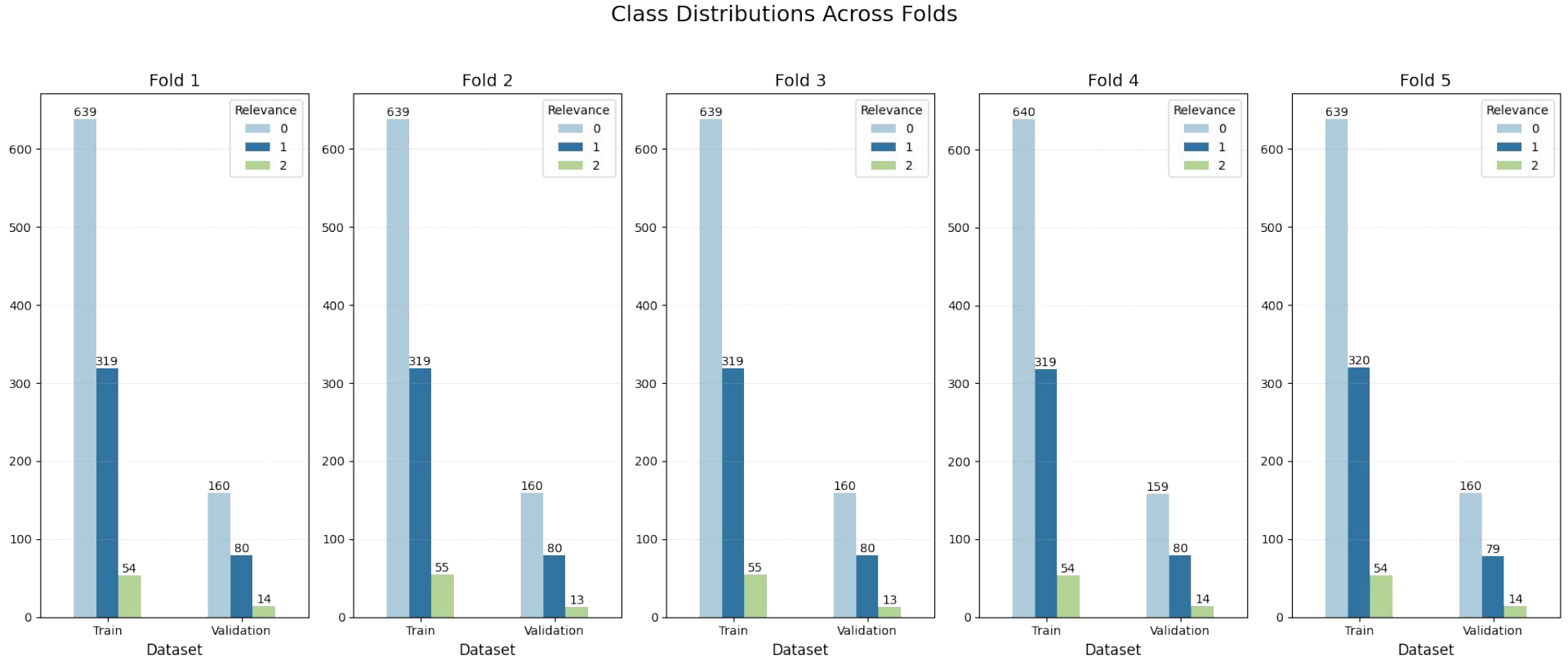
BERTimbau Performance Before Fine-Tuning

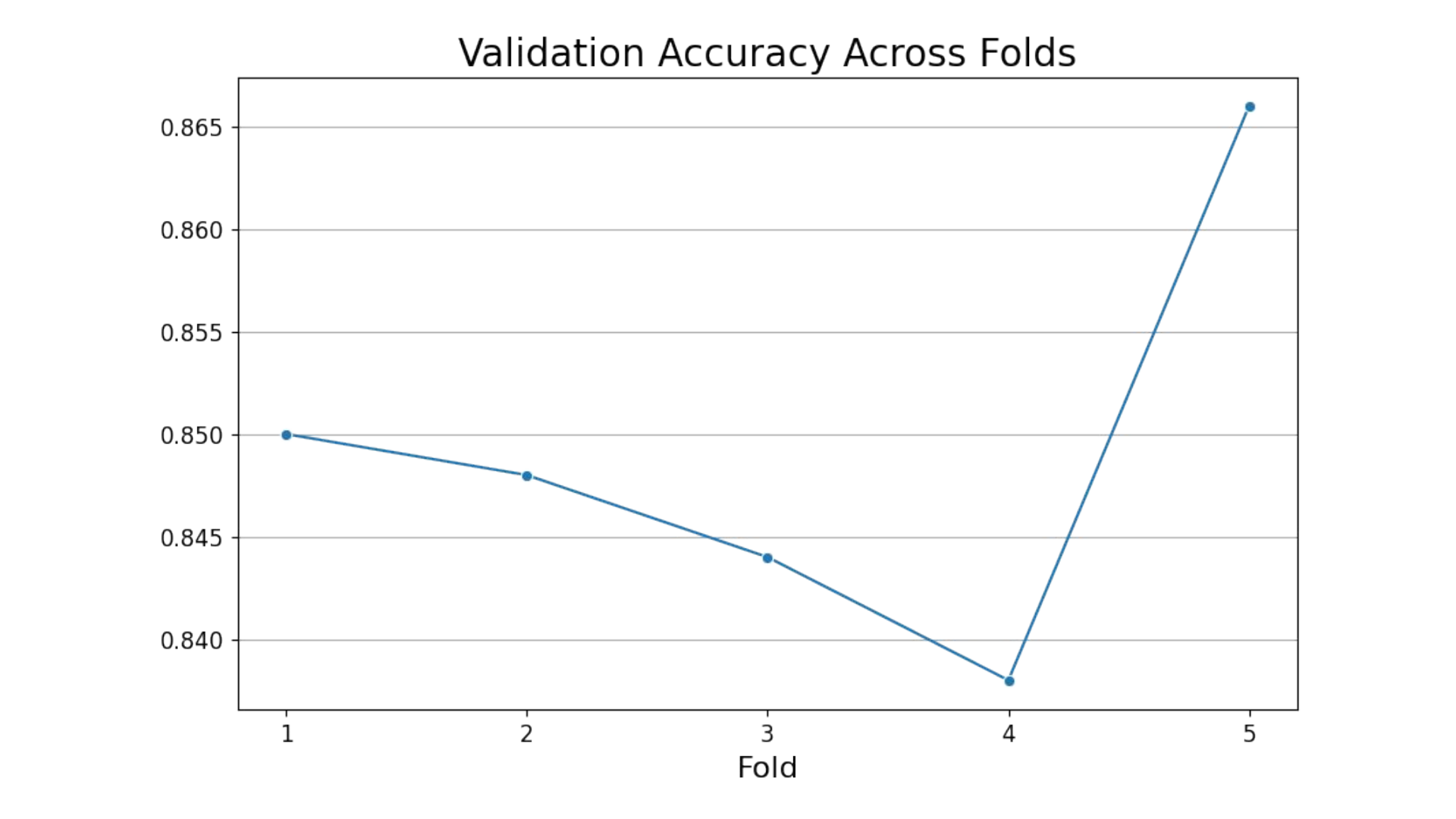
	precision	recall	f1-score	support
0	0.62	0.9	0.74	89.0
1	0.38	0.07	0.12	44.0
2	0.0	0.0	0.0	8.0
macro avg	0.33	0.32	0.28	141.0
weighted avg	0.51	0.59	0.5	141.0

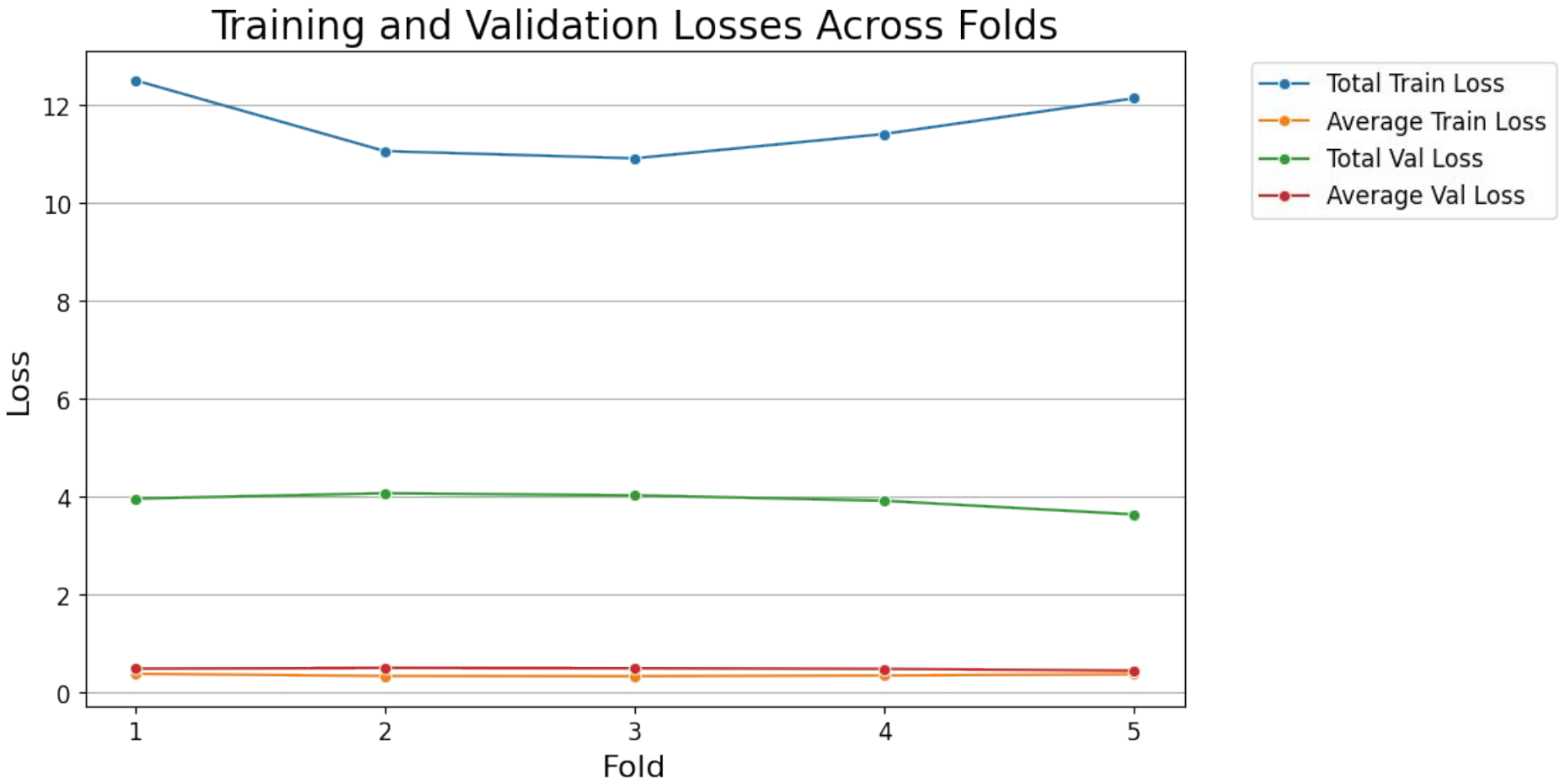
Overall Accuracy: 0.59

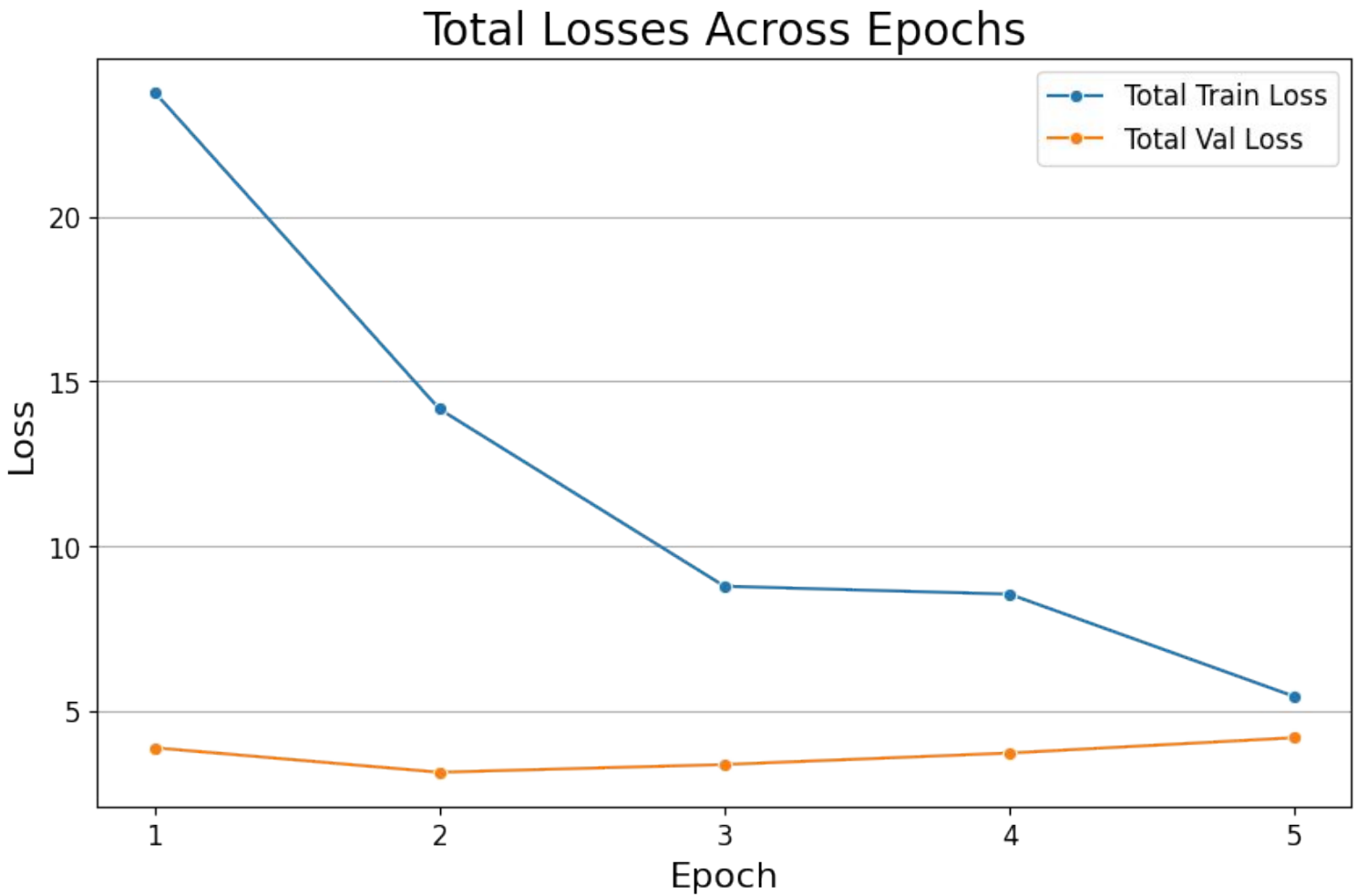

BERTimbau Performance Before Fine-Tuning
True Labels
negative
positive
unclear
80
5
4
40
3
1
8
0
0
negative
positive
unclear
Predicted Labels
80
70
60
50
40
30
20
10
0

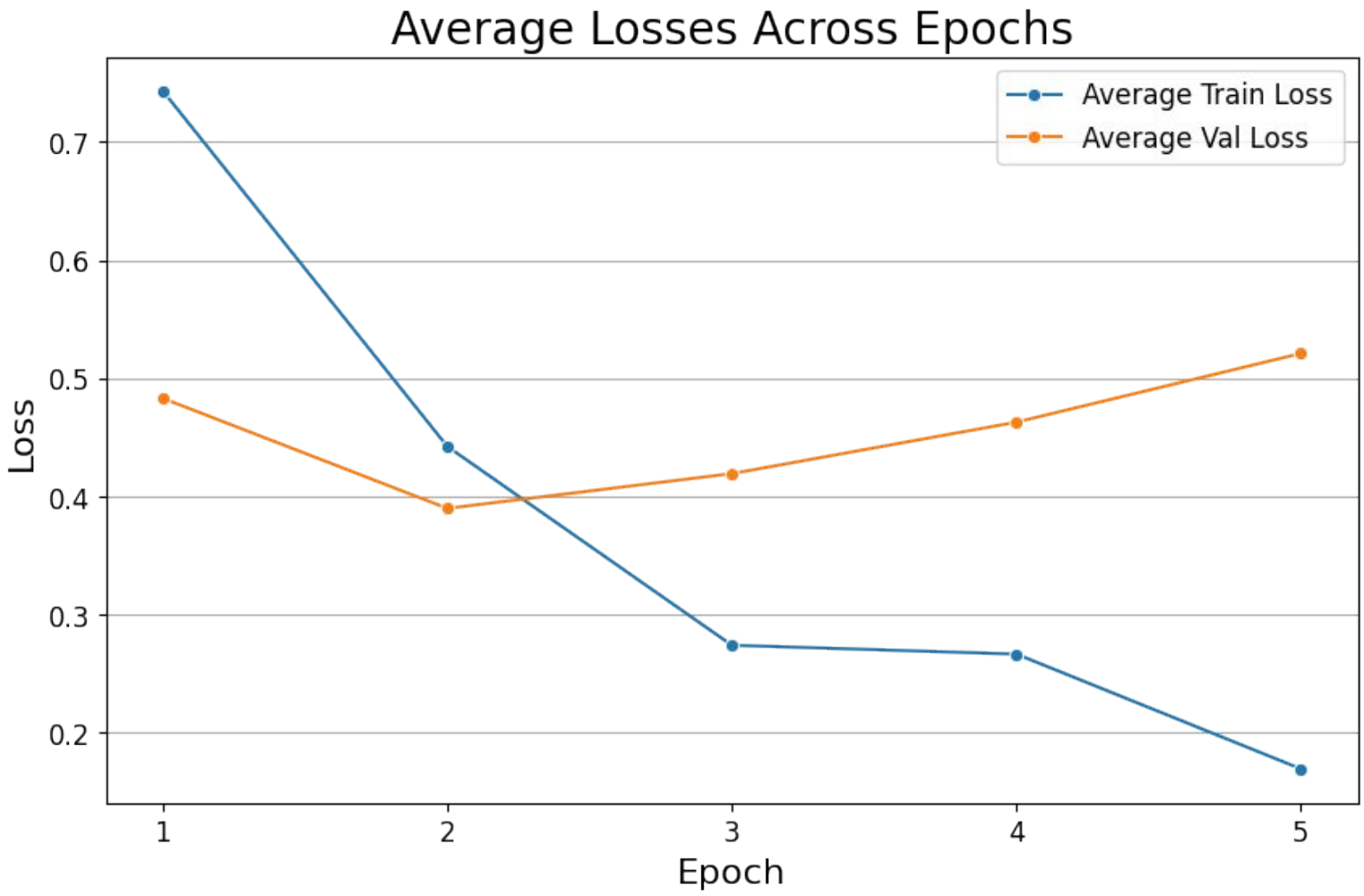
FINE TUNING...


Class Distributions Across Folds
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5
Relevance
0
1
2
639
319
54
160
80
14
639
319
55
160
80
13
639
319
55
160
80
13
640
319
54
159
80
14
639
320
54
160
79
14
0
100
200
300
400
500
600
Train
Validation
Dataset


Validation Accuracy Across Folds
0.865
0.860
0.855
0.850
0.845
0.840
1
2
3
4
5
Fold


Training and Validation Losses Across Folds
Total Train Loss
Average Train Loss
Total Val Loss
Average Val Loss
Loss
Fold
0
2
4
6
8
10
12
1
2
3
4
5


Total Losses Across Epochs
Total Train Loss
Total Val Loss
Loss
20
15
10
5
1
2
3
4
5
Epoch


Average Losses Across Epochs
Average Train Loss
Average Val Loss
Loss
0.7
0.6
0.5
0.4
0.3
0.2
1
2
3
4
5
Epoch

TEST NEW MODEL…

BERTimbau Performance After Fine-Tuning

	precision	recall	f1-score	support
0	0.8	0.94	0.87	89.0
1	0.78	0.64	0.7	44.0
2	0.0	0.0	0.0	8.0
macro avg	0.53	0.53	0.52	141.0
weighted avg	0.75	0.79	0.77	141.0

Overall Accuracy: 0.79

BERTimbau Performance Before Fine-Tuning

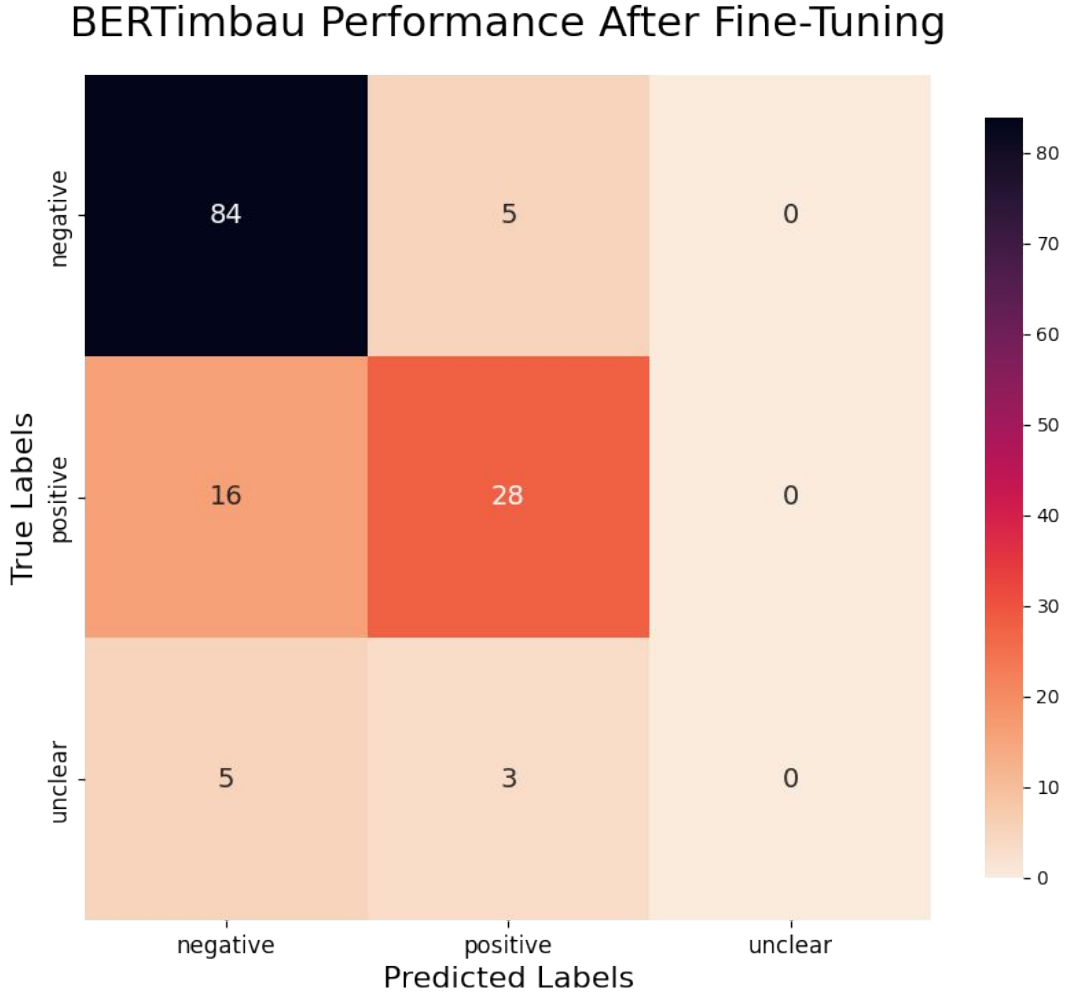
	precision	recall	f1-score	support
0	0.62	0.9	0.74	89.0
1	0.38	0.07	0.12	44.0
2	0.0	0.0	0.0	8.0
macro avg	0.33	0.32	0.28	141.0
weighted avg	0.51	0.59	0.5	141.0

Overall Accuracy: 0.59

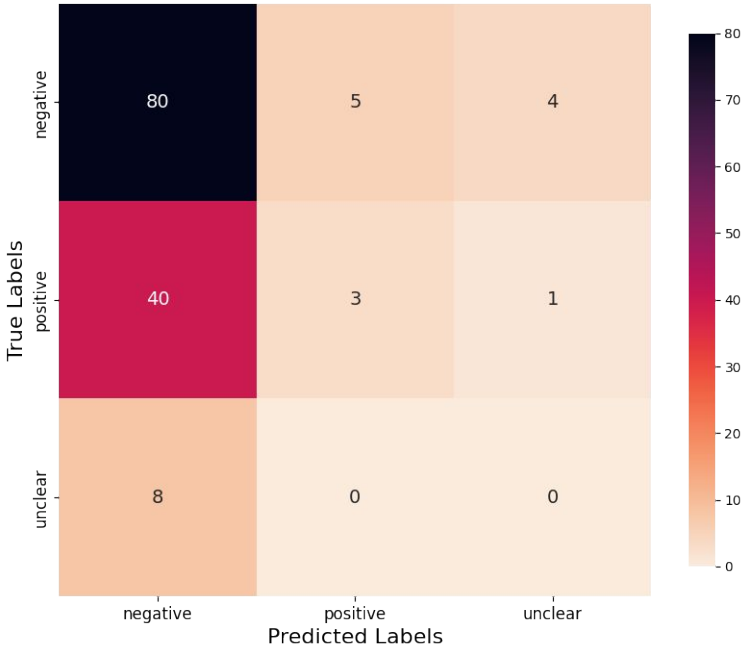
BERTimbau Performance After Fine-Tuning

	precision	recall	f1-score	support
0	0.8	0.94	0.87	89.0
1	0.78	0.64	0.7	44.0
2	0.0	0.0	0.0	8.0
macro avg	0.53	0.53	0.52	141.0
weighted avg	0.75	0.79	0.77	141.0

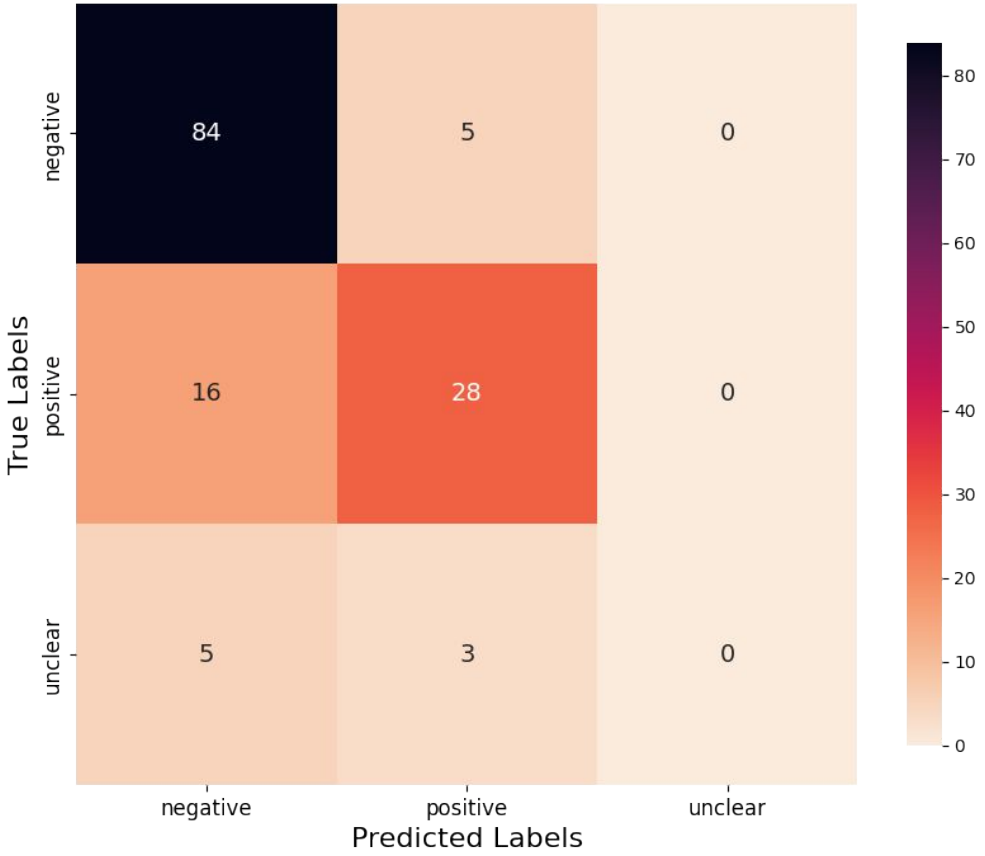
Overall Accuracy: 0.79


BERTimbau Performance After Fine-Tuning
True Labels
negative
positive
unclear
84
5
0
16
28
0
5
3
0
negative
positive
unclear
Predicted Labels
80
70
60
50
40
30
20
10
0

BERTimbau Performance Before Fine-Tuning



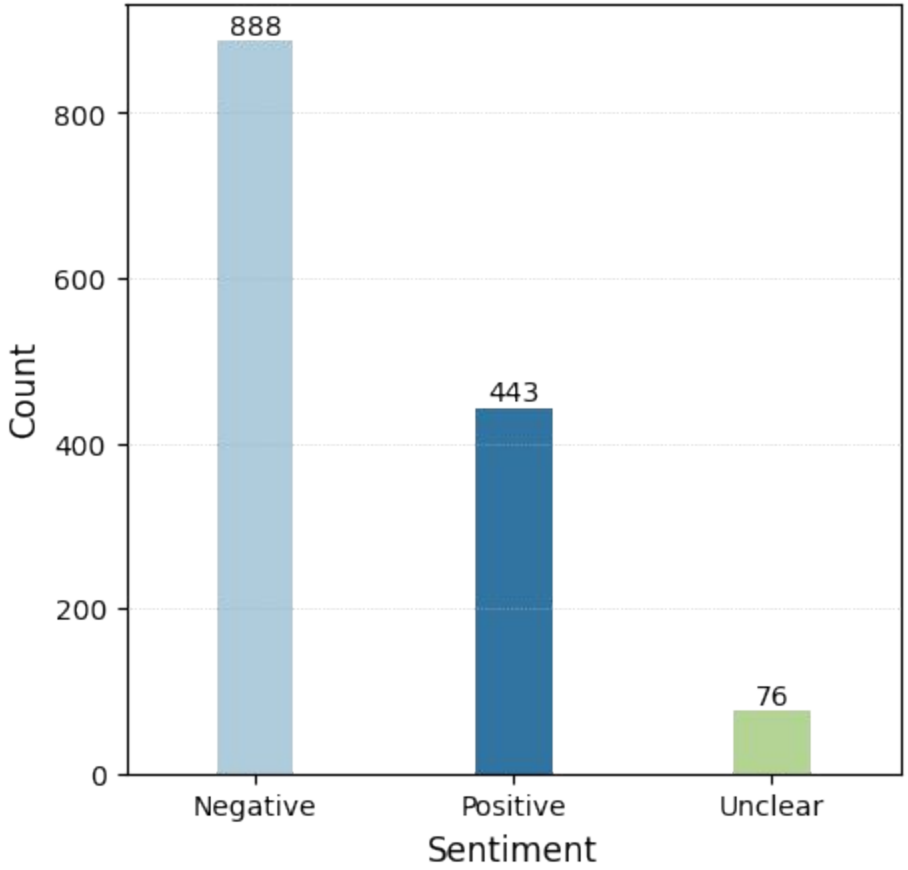
BERTimbau Performance After Fine-Tuning



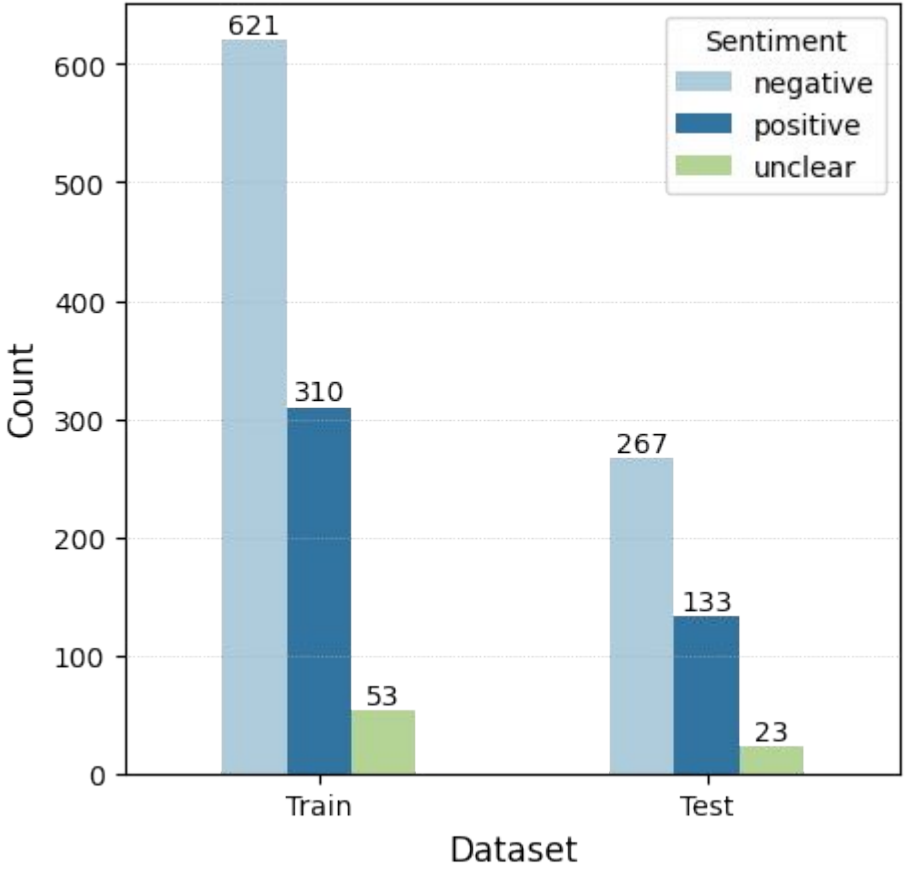
– 7:3 TRAIN/TEST SPLIT –

PREPARING THE DATA…

Sentiment Distribution - Whole Dataset



Sentiment Distribution - Train/Test Sets

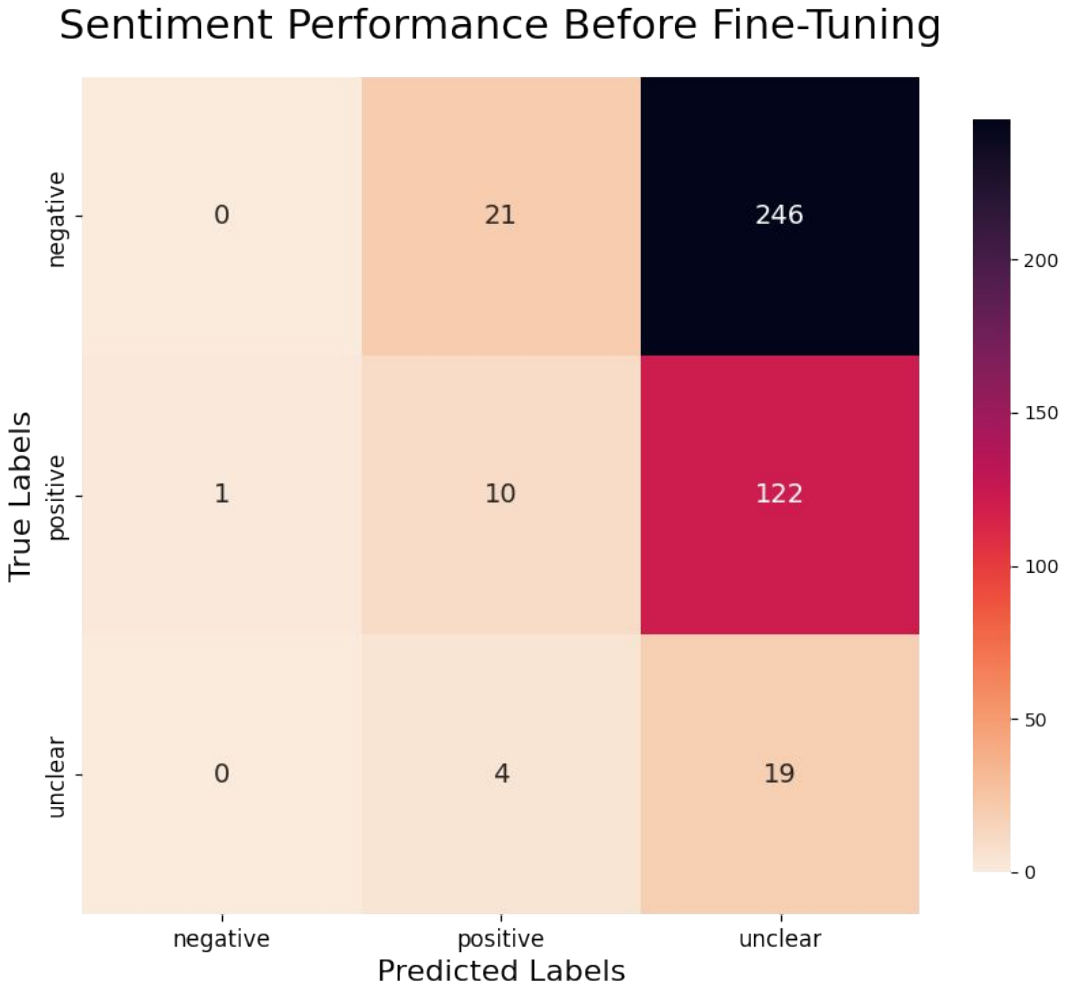


TEST BERTimbau...

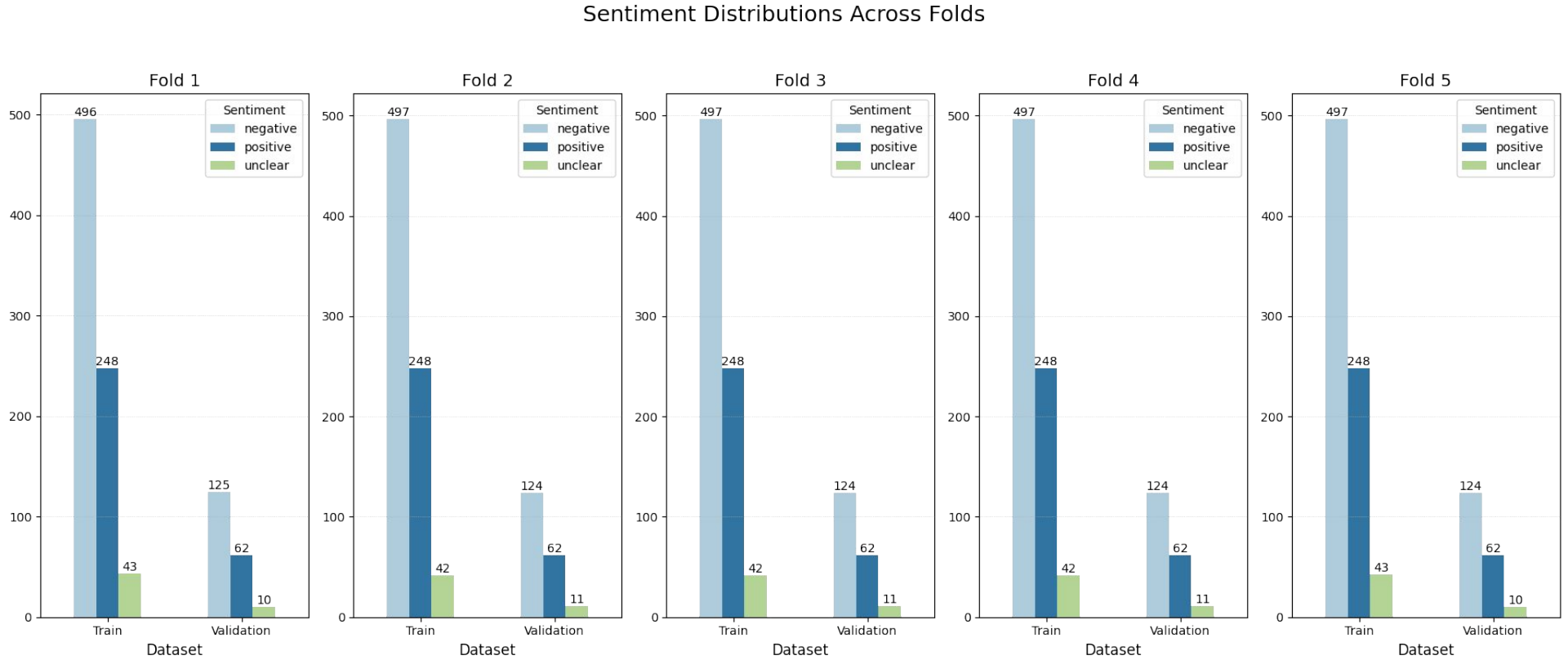
Sentiment Performance Before Fine-Tuning

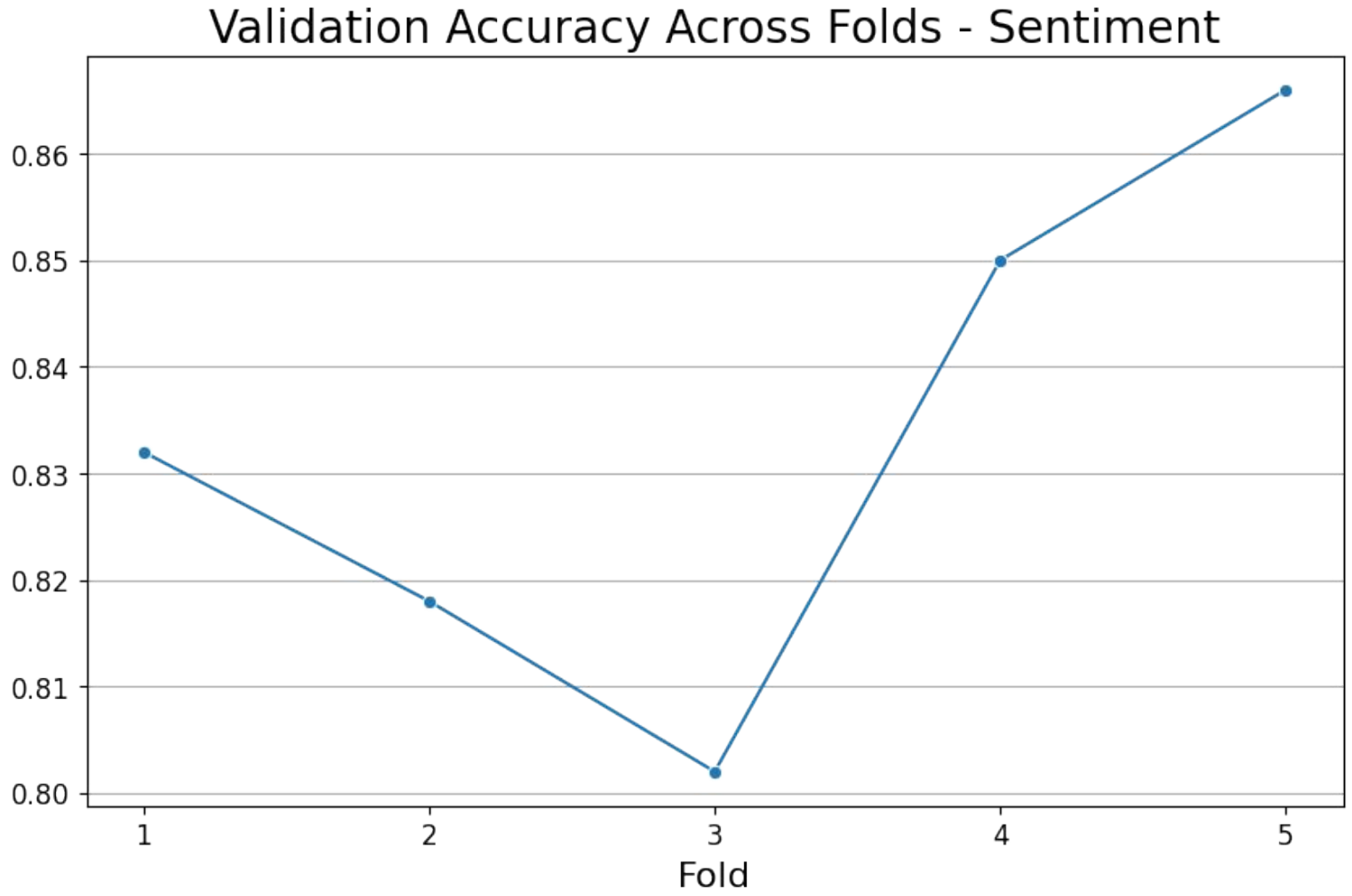
	precision	recall	f1-score	support
0	0.0	0.0	0.0	267.0
1	0.29	0.08	0.12	133.0
2	0.05	0.83	0.09	23.0
macro avg	0.11	0.3	0.07	423.0
weighted avg	0.09	0.07	0.04	423.0

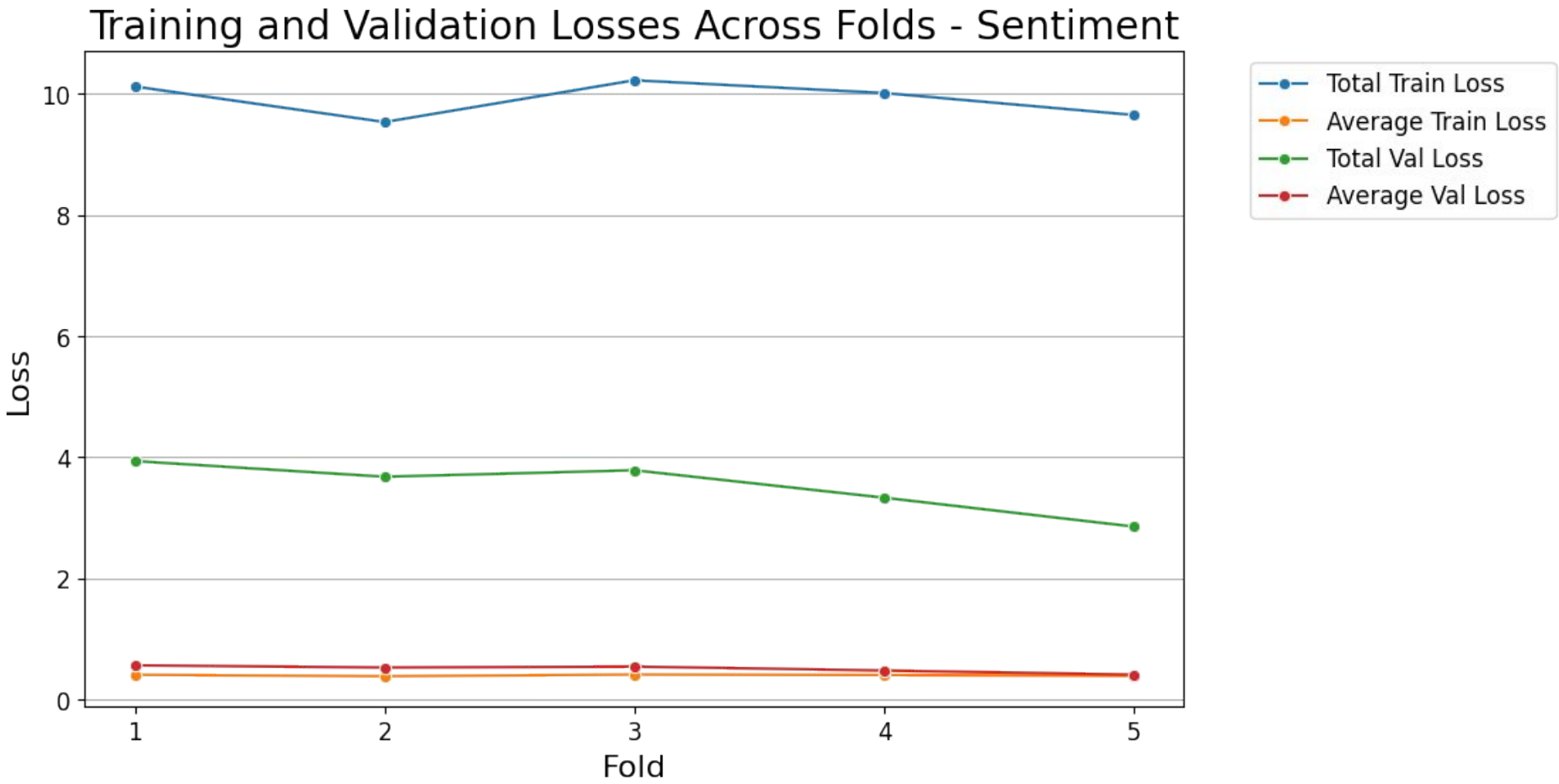
Overall Accuracy: 0.07

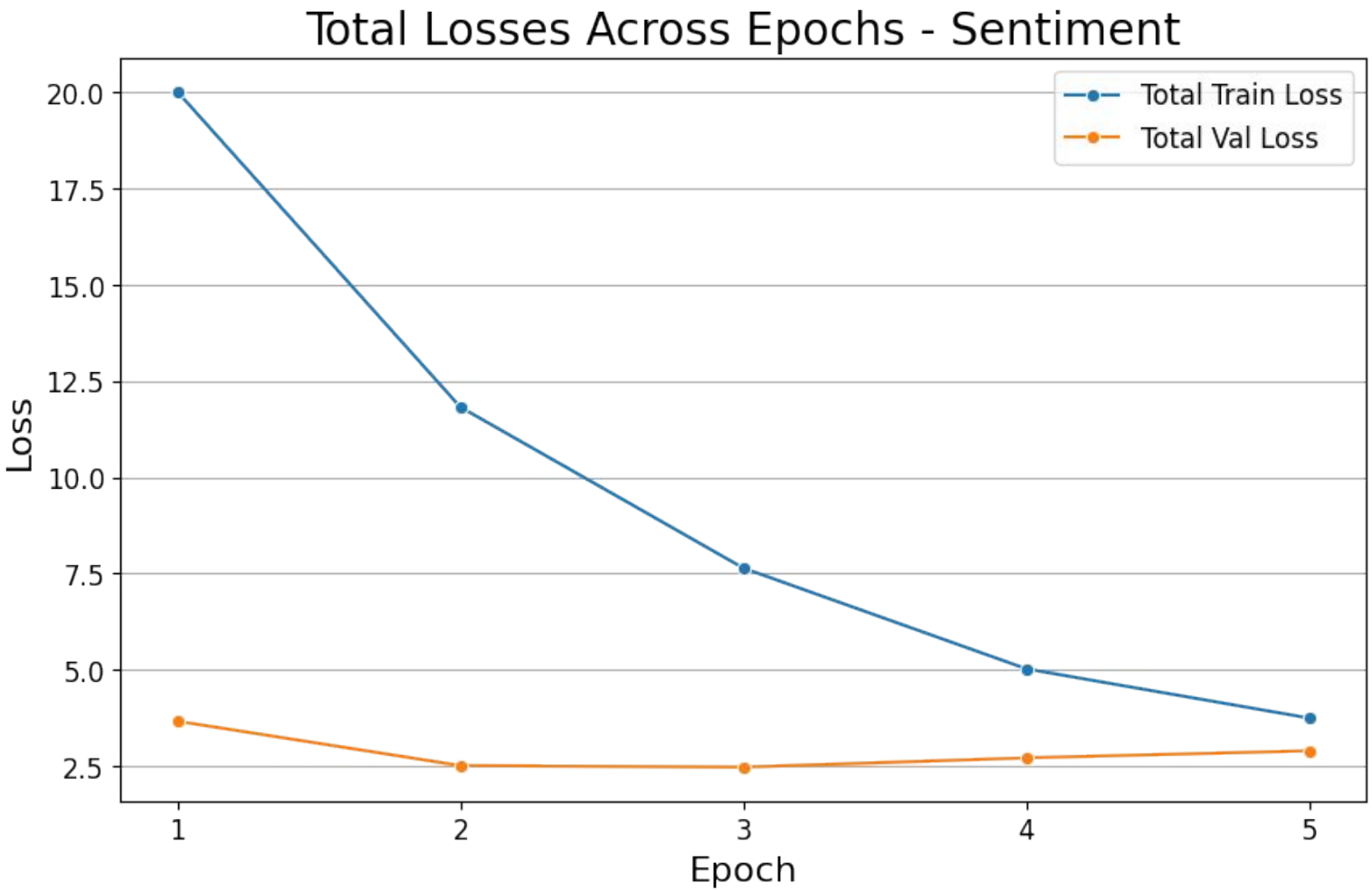

Sentiment Performance Before Fine-Tuning
True Labels
negative
positive
unclear
0
21
246
1
10
122
0
4
19
negative
positive
unclear
Predicted Labels
200
150
100
50
0

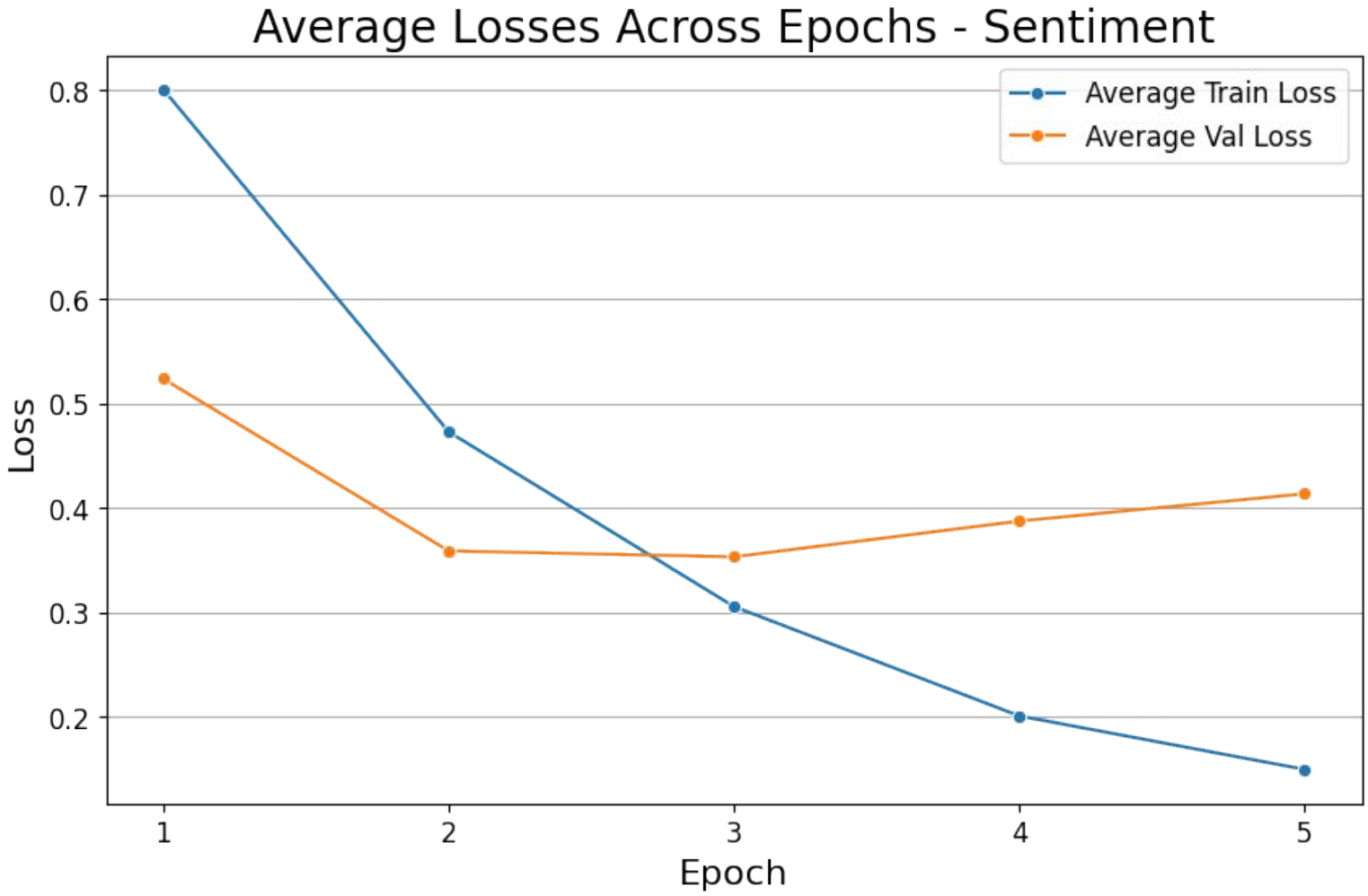
FINE TUNING...


Sentiment Distributions Across Folds
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5
Sentiment
negative
positive
unclear
496
248
43
125
62
10
497
248
42
124
62
11
497
248
42
124
62
11
497
248
42
124
62
11
497
248
43
124
62
10
0
100
200
300
400
500
Train
Validation
Dataset


Validation Accuracy Across Folds - Sentiment
0.86
0.85
0.84
0.83
0.82
0.81
0.80
1
2
3
4
5
Fold


Training and Validation Losses Across Folds - Sentiment
Loss
Fold
0
2
4
6
8
10
1
2
3
4
5
Total Train Loss
Average Train Loss
Total Val Loss
Average Val Loss


Total Losses Across Epochs - Sentiment
Total Train Loss
Total Val Loss
Loss
Epoch
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
1
2
3
4
5


Average Losses Across Epochs - Sentiment
Average Train Loss
Average Val Loss
Loss
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1
2
3
4
5
Epoch

TEST NEW MODEL…

Best Sentiment Performance After Fine-Tuning

	precision	recall	f1-score	support
0	0.89	0.88	0.89	267.0
1	0.75	0.84	0.79	133.0
2	0.36	0.17	0.24	23.0
macro avg	0.67	0.63	0.64	423.0
weighted avg	0.82	0.83	0.82	423.0

Overall Accuracy: 0.83

Sentiment Performance Before Fine-Tuning

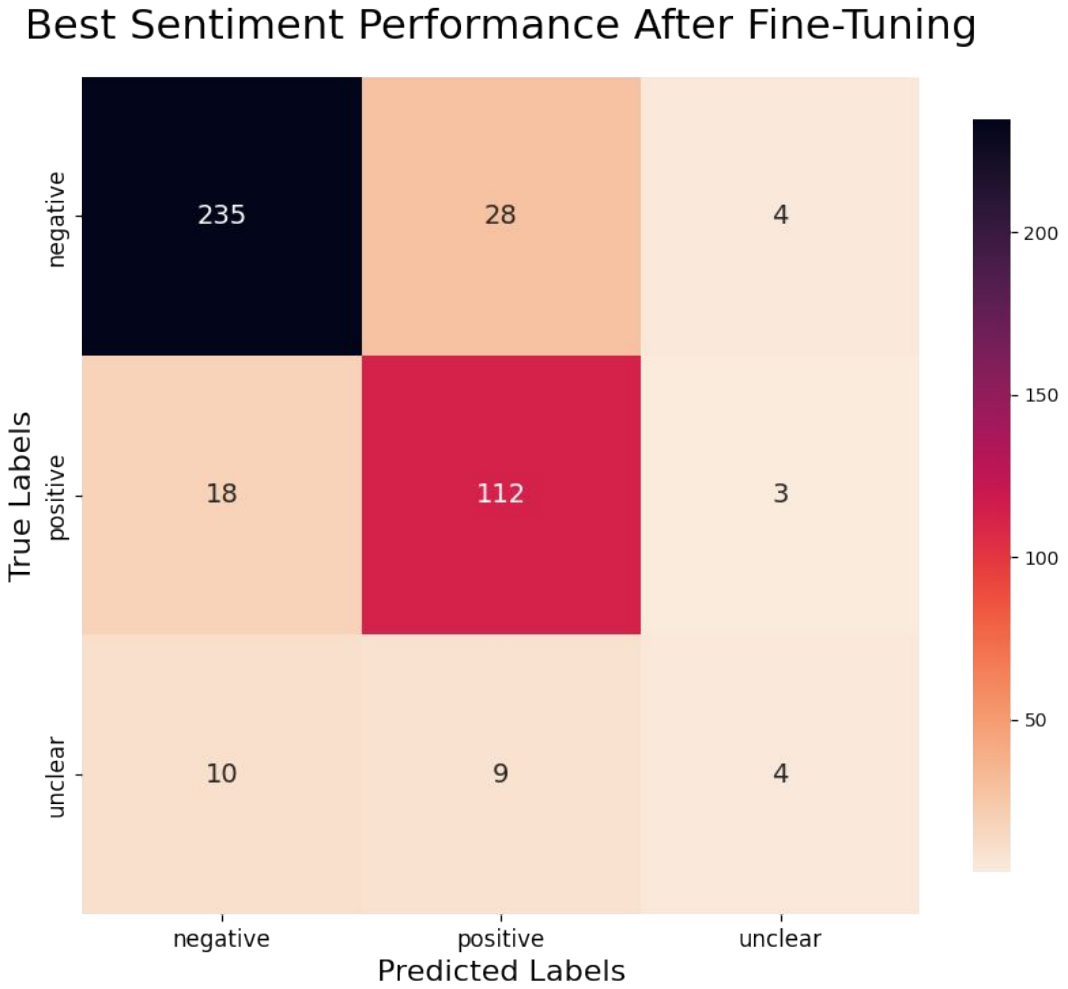
	precision	recall	f1-score	support
0	0.0	0.0	0.0	267.0
1	0.29	0.08	0.12	133.0
2	0.05	0.83	0.09	23.0
macro avg	0.11	0.3	0.07	423.0
weighted avg	0.09	0.07	0.04	423.0

Overall Accuracy: 0.07

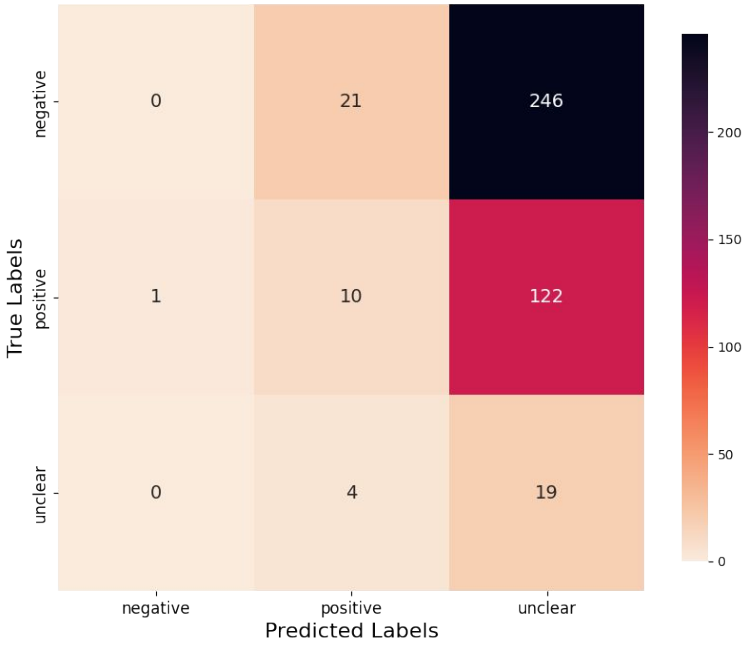
Best Sentiment Performance After Fine-Tuning

	precision	recall	f1-score	support
0	0.89	0.88	0.89	267.0
1	0.75	0.84	0.79	133.0
2	0.36	0.17	0.24	23.0
macro avg	0.67	0.63	0.64	423.0
weighted avg	0.82	0.83	0.82	423.0

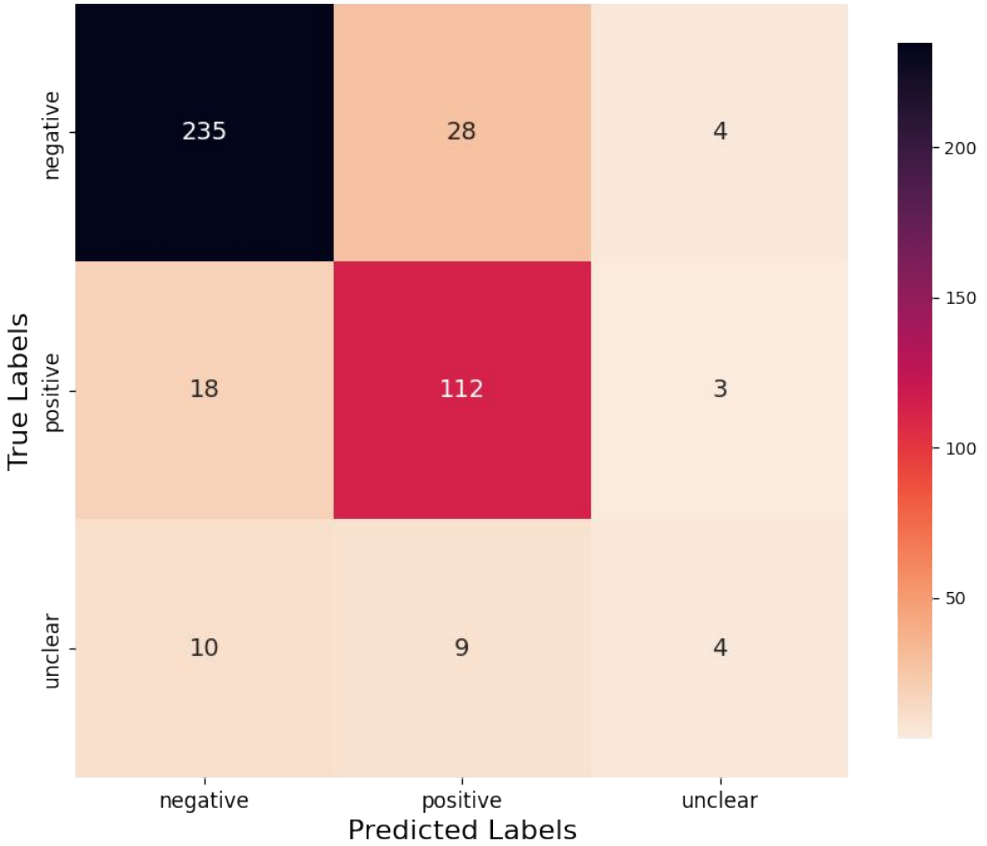
Overall Accuracy: 0.83


Best Sentiment Performance After Fine-Tuning
True Labels
negative
positive
unclear
235
28
4
18
112
3
10
9
4
200
150
100
50
negative
positive
unclear
Predicted Labels

Sentiment Performance Before Fine-Tuning



Best Sentiment Performance After Fine-Tuning



EVALUATION ACROSS FOLDS…

Sentiment Performance After Fine-Tuning: Fold 1

	precision	recall	f1-score	support
0	0.9	0.87	0.88	267.0
1	0.74	0.9	0.81	133.0
2	0.0	0.0	0.0	23.0
macro avg	0.54	0.59	0.57	423.0
weighted avg	0.8	0.83	0.81	423.0

Overall Accuracy: 0.83

Sentiment Performance After Fine-Tuning: Fold 2

	precision	recall	f1-score	support
0	0.87	0.88	0.88	267.0
1	0.78	0.79	0.78	133.0
2	0.0	0.0	0.0	23.0
macro avg	0.55	0.56	0.55	423.0
weighted avg	0.79	0.81	0.8	423.0

Overall Accuracy: 0.81

Sentiment Performance After Fine-Tuning: Fold 3

	precision	recall	f1-score	support
0	0.88	0.89	0.89	267.0
1	0.75	0.86	0.8	133.0
2	0.0	0.0	0.0	23.0
macro avg	0.54	0.59	0.56	423.0
weighted avg	0.79	0.83	0.81	423.0

Overall Accuracy: 0.83

Sentiment Performance After Fine-Tuning: Fold 4

	precision	recall	f1-score	support
0	0.87	0.93	0.9	267.0
1	0.8	0.83	0.81	133.0
2	0.5	0.04	0.08	23.0
macro avg	0.72	0.6	0.6	423.0
weighted avg	0.83	0.85	0.83	423.0

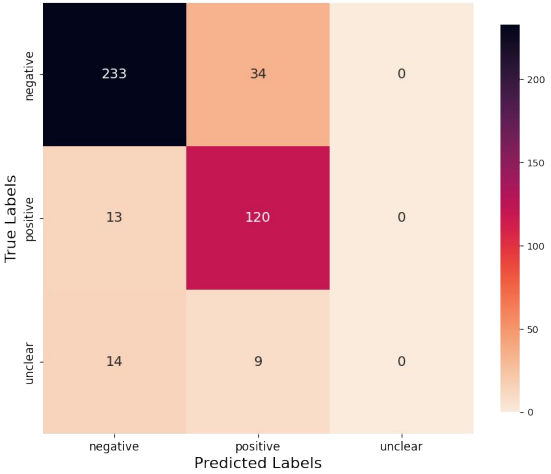
Overall Accuracy: 0.85

Sentiment Performance After Fine-Tuning: Fold 5

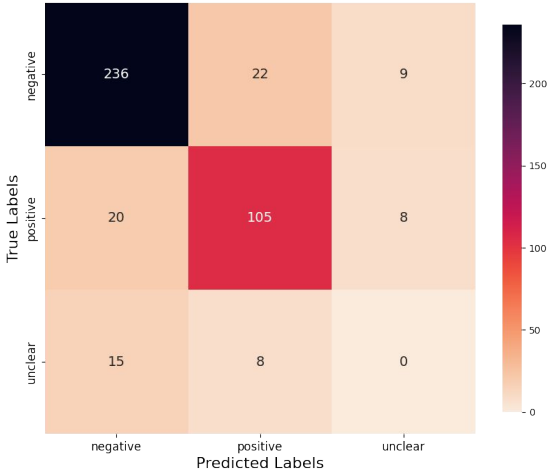
	precision	recall	f1-score	support
0	0.89	0.88	0.89	267.0
1	0.75	0.84	0.79	133.0
2	0.36	0.17	0.24	23.0
macro avg	0.67	0.63	0.64	423.0
weighted avg	0.82	0.83	0.82	423.0

Overall Accuracy: 0.83

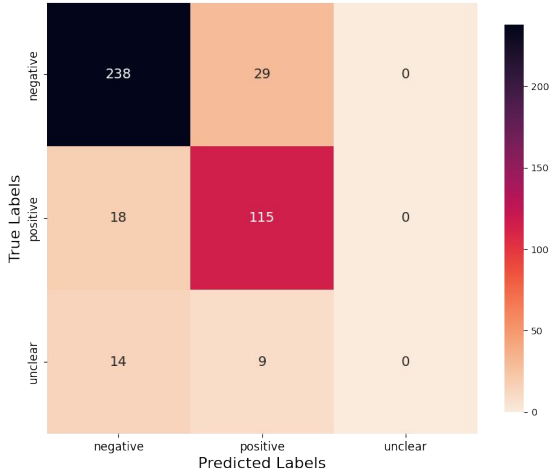
Sentiment Performance After Fine-Tuning: Fold 1



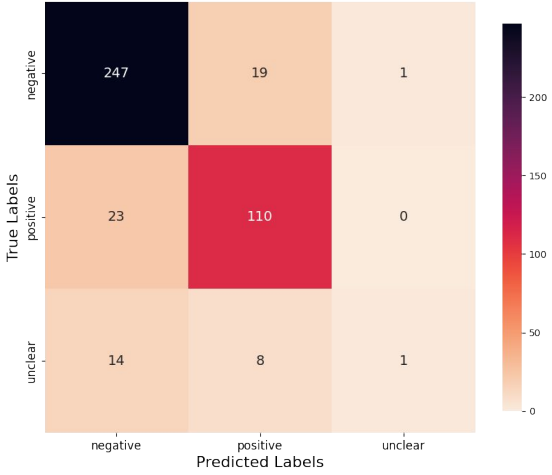
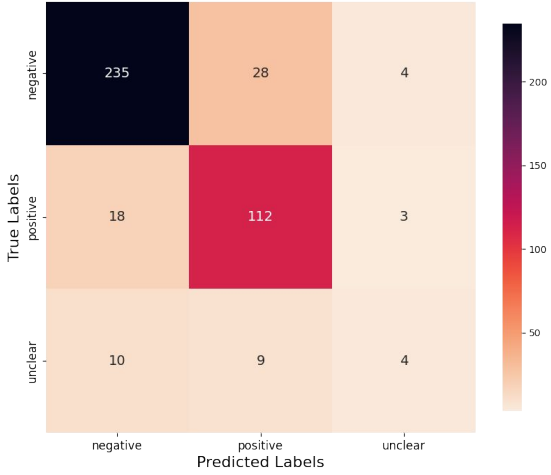
Sentiment Performance After Fine-Tuning: Fold 2



Sentiment Performance After Fine-Tuning: Fold 3



Sentiment Performance After Fine-Tuning: Fold 4



Sentiment Performance After Fine-Tuning: Fold 5