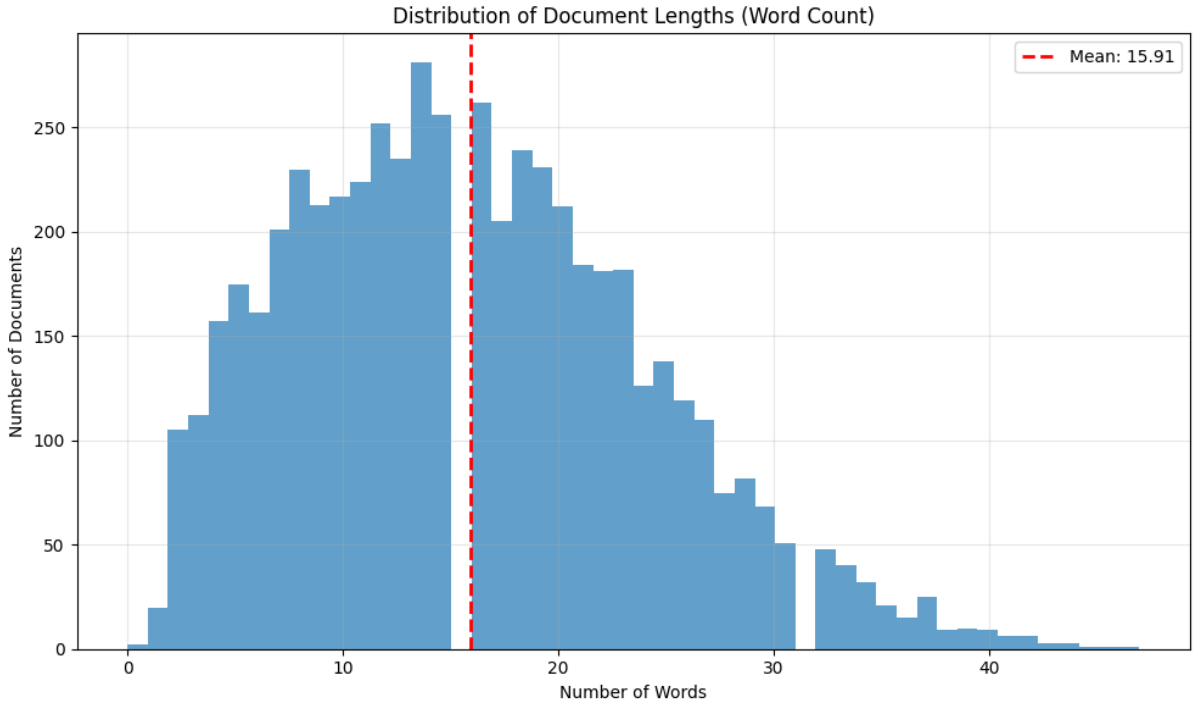
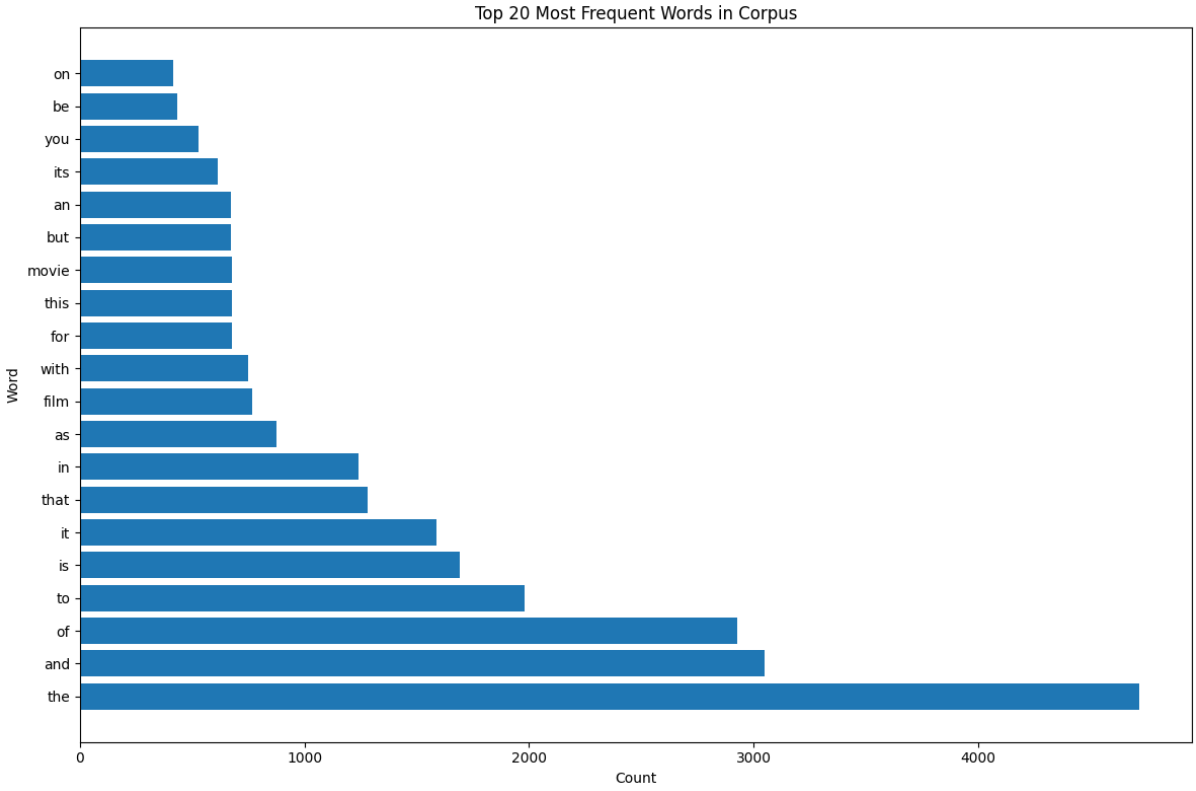
```
Model Summary:
==========================================================
Input Size: 10000
Layer 1: Linear(10000, 512)
    Parameters: 5,120,512
    Activation: ReLU + Dropout(0.3)
Layer 2: Linear(512, 256)
    Parameters: 131,328
    Activation: ReLU + Dropout(0.3)
Layer 3: Linear(256, 128)
    Parameters: 32,896
    Activation: ReLU + Dropout(0.3)
Layer 4: Linear(128, 64)
    Parameters: 8,256
    Activation: ReLU + Dropout(0.3)
Layer 5: Linear(64, 2)
    Parameters: 130
==========================================================
Total Trainable Parameters: 5,293,122

PyTorch Model Structure:
SentimentMLP(
  (fc1): Linear(in_features=10000, out_features=512, bias=True)
  (fc2): Linear(in_features=512, out_features=256, bias=True)
  (fc3): Linear(in_features=256, out_features=128, bias=True)
  (fc4): Linear(in_features=128, out_features=64, bias=True)
  (fc5): Linear(in_features=64, out_features=2, bias=True)
  (dropout): Dropout(p=0.3, inplace=False)
)

Total trainable parameters (using PyTorch's function): 5,293,122
```

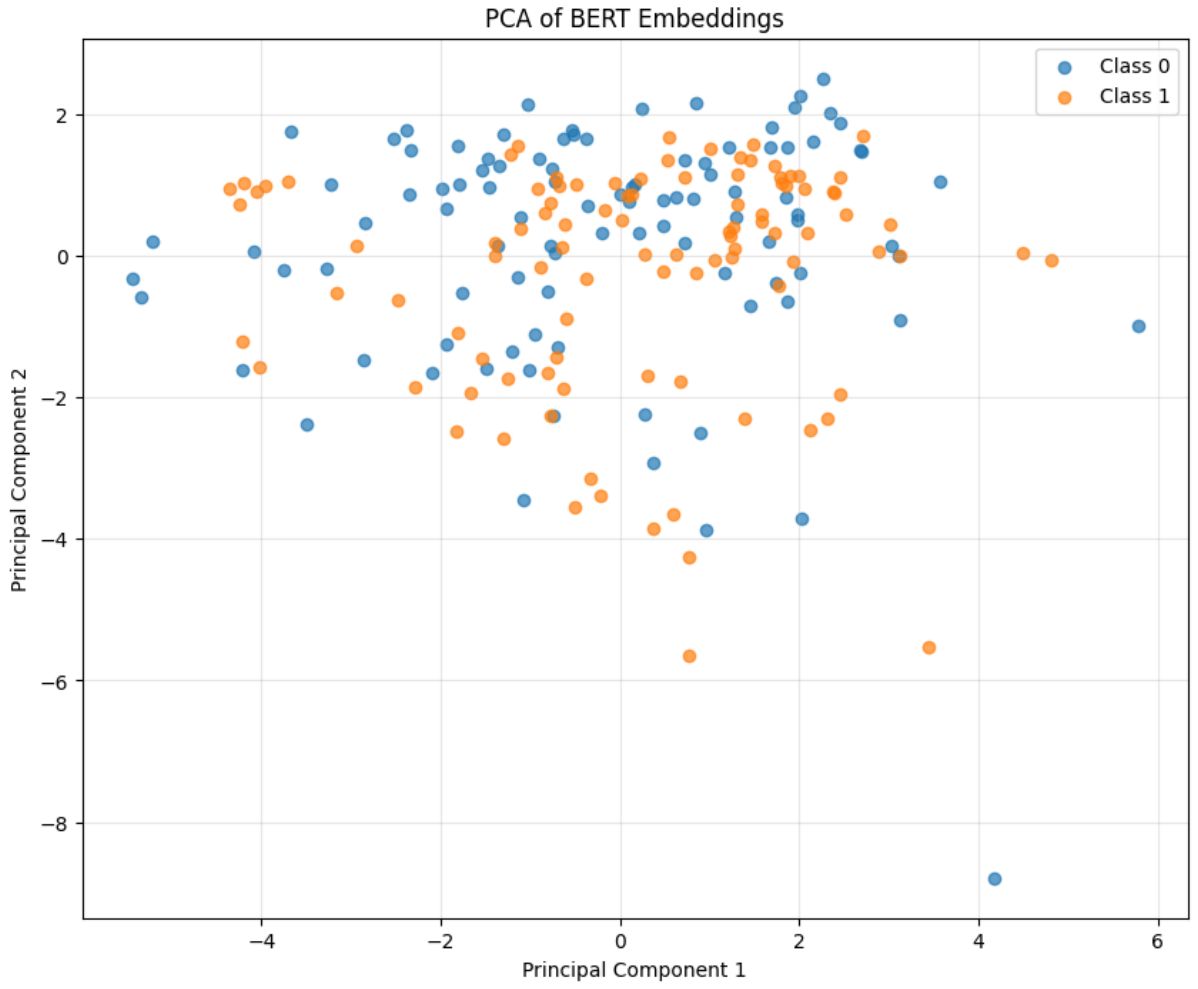
```
Training set shape: (5536, 2)
Validation set shape: (1385, 2)
Test set shape: (1821, 2)
Training features shape: (5536, 10000)
Validation features shape: (1385, 10000)
Test features shape: (1821, 10000)
Vocabulary size: 10000
Sample of vocabulary (first 20 words): ['10' '100' '101' '11' '112' '12' '120' '127' '129' '13' '14' '140' '146'
 '15' '15th' '16' '163' '170' '1790' '18']
Number of batches in training set: 87
Number of batches in validation set: 22
Number of batches in test set: 29
Batch features shape: torch.Size([64, 10000])
Batch labels shape: torch.Size([64])
Sum of first sample features: 18.0
Label counts in batch: tensor([31, 33])
Total trainable parameters in BoW model: 5,293,122
BoW representation sparsity: 0.0015
Average words per document: 15.91
```

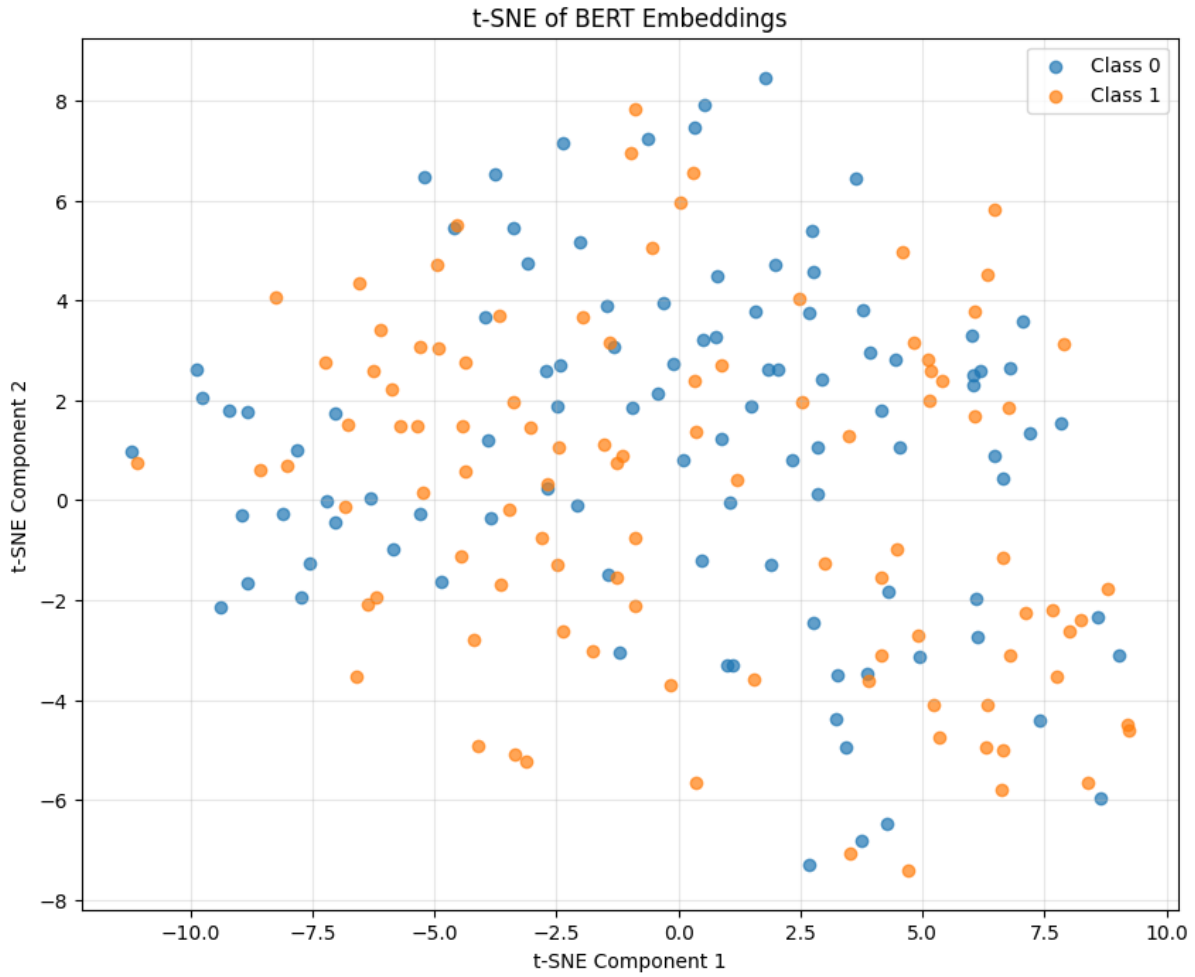





```
Using device: cpu
Training set shape: (5536, 2)
Validation set shape: (1385, 2)
Test set shape: (1821, 2)
Loading tokenizer and model for bert-base-uncased...
/usr/local/lib/python3.11/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens), set it as secret in your Google Colab and restart your
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or datasets.
  warnings.warn(
tokenizer_config.json: 100% 48.0/48.0 [00:00<00:00, 5.27kB/s]
config.json: 100% 570/570 [00:00<00:00, 58.6kB/s]
vocab.txt: 100% 232k/232k [00:00<00:00, 5.26MB/s]
tokenizer.json: 100% 466k/466k [00:00<00:00, 10.6MB/s]
model.safetensors: 100% 440M/440M [00:02<00:00, 188MB/s]
Model loaded with embedding size: 768
Processing training data...
Processing chunks:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|       | 1/4 [00:03<00:09,  3.02s/it]
Extracting embeddings:  50%|     | 2/4 [00:04<00:04,  2.36s/it]
Extracting embeddings:  75%|   | 3/4 [00:07<00:02,  2.38s/it]
Extracting embeddings: 100%| | 4/4 [00:08<00:00,  2.05s/it]
Processing chunks:  25%|       | 1/4 [00:08<00:24,  8.22s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|       | 1/4 [00:02<00:06,  2.00s/it]
Extracting embeddings:  50%|     | 2/4 [00:03<00:03,  1.94s/it]
Extracting embeddings:  75%|   | 3/4 [00:06<00:02,  2.11s/it]
Extracting embeddings: 100%| | 4/4 [00:07<00:00,  1.78s/it]
Processing chunks:  50%|     | 2/4 [00:15<00:15,  7.62s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|       | 1/4 [00:02<00:05,  2.00s/it]
Extracting embeddings:  50%|     | 2/4 [00:03<00:03,  2.00s/it]
Extracting embeddings:  75%|   | 3/4 [00:05<00:01,  2.00s/it]
Extracting embeddings: 100%| | 4/4 [00:06<00:00,  1.65s/it]
Processing chunks:  75%|   | 3/4 [00:22<00:07,  7.20s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|       | 1/4 [00:01<00:05,  1.70s/it]
Extracting embeddings:  50%|     | 2/4 [00:03<00:03,  1.82s/it]
Extracting embeddings:  75%|   | 3/4 [00:05<00:01,  1.76s/it]
Extracting embeddings: 100%| | 4/4 [00:05<00:00,  1.48s/it]
Processing chunks: 100%| | 4/4 [00:28<00:00,  7.05s/it]
Processing validation data...
Processing chunks:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|       | 1/4 [00:01<00:04,  1.60s/it]
Extracting embeddings:  50%|     | 2/4 [00:03<00:04,  2.01s/it]
Extracting embeddings:  75%|   | 3/4 [00:05<00:01,  1.96s/it]
Extracting embeddings: 100%| | 4/4 [00:06<00:00,  1.63s/it]
Processing chunks:  25%|       | 1/4 [00:06<00:19,  6.60s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|       | 1/4 [00:02<00:06,  2.10s/it]
Extracting embeddings:  50%|     | 2/4 [00:04<00:04,  2.10s/it]
Extracting embeddings:  75%|   | 3/4 [00:06<00:02,  2.05s/it]
Extracting embeddings: 100%| | 4/4 [00:06<00:00,  1.72s/it]
Processing chunks:  50%|     | 2/4 [00:13<00:13,  6.77s/it]
```

```
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:02<00:06,  2.10s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:03<00:03,  1.98s/it]
Extracting embeddings:  75%|███████▌  | 3/4 [00:06<00:02,  2.08s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:06<00:00,  1.75s/it]
Processing chunks:  75%|███████▌  | 3/4 [00:20<00:06,  6.88s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:01<00:05,  1.80s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:04<00:04,  2.04s/it]
Extracting embeddings:  75%|███████▌  | 3/4 [00:06<00:02,  2.02s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:06<00:00,  1.70s/it]
Processing chunks: 100%|██████████| 4/4 [00:27<00:00,  6.85s/it]
Processing test data...
Processing chunks:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:02<00:07,  2.50s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:04<00:04,  2.15s/it]
Extracting embeddings:  75%|███████▌  | 3/4 [00:06<00:01,  1.99s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:07<00:00,  1.77s/it]
Processing chunks:  25%|██▌       | 1/4 [00:07<00:21,  7.10s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:01<00:05,  1.99s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:04<00:04,  2.06s/it]
Extracting embeddings:  75%|███████▌  | 3/4 [00:06<00:02,  2.03s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:06<00:00,  1.73s/it]
Processing chunks:  50%|█████     | 2/4 [00:14<00:13,  6.99s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:02<00:06,  2.10s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:04<00:03,  1.98s/it]
Extracting embeddings:  75%|███████▌  | 3/4 [00:05<00:01,  1.99s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:06<00:00,  1.68s/it]
Processing chunks:  75%|███████▌  | 3/4 [00:20<00:06,  6.90s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:01<00:05,  1.90s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:04<00:04,  2.08s/it]
Extracting embeddings:  75%|███████▌  | 3/4 [00:06<00:02,  2.04s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:07<00:00,  1.75s/it]
Processing chunks: 100%|██████████| 4/4 [00:27<00:00,  6.98s/it]
Embeddings saved to disk.
Embedding size: 768
Total trainable parameters in embedding model: 566,338
```





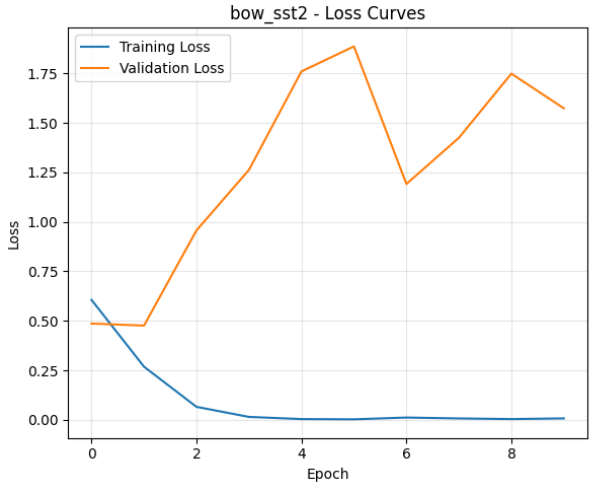

```
Extracting embeddings: 100%|██████████| 1/1 [00:00<00:00,  1.97it/s]Example embeddings shape: (3, 768)
Cosine similarity between example sentences:
Similarity between "I loved this movie, it was amazing!" and "This film was terrible and boring.": 0.8485
Similarity between "I loved this movie, it was amazing!" and "The plot was interesting but the acting was poor.": 0.7988
Similarity between "This film was terrible and boring." and "The plot was interesting but the acting was poor.": 0.8938
BERT embeddings implementation completed successfully.
```

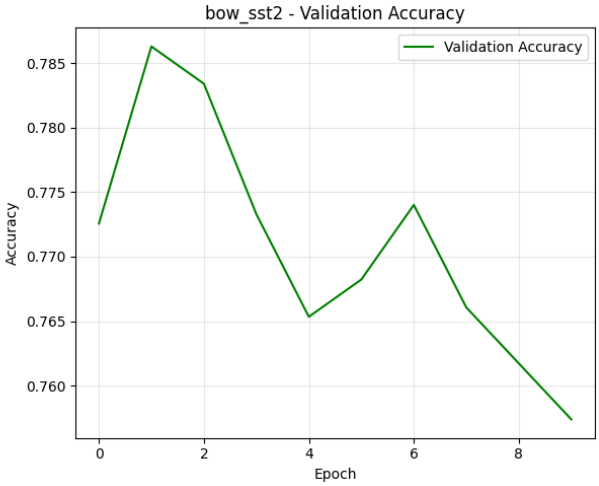
```
Loading tokenizer and model for bert-base-uncased...
Model loaded with embedding size: 768
Processing chunks:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:02<00:07,  2.66s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:04<00:04,  2.45s/it]
Extracting embeddings:  75%|███████▌  | 3/4 [00:07<00:02,  2.38s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:08<00:00,  2.04s/it]
Processing chunks:  25%|██▌       | 1/4 [00:08<00:24,  8.17s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:02<00:07,  2.50s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:04<00:04,  2.20s/it]
Extracting embeddings:  75%|███████▌  | 3/4 [00:06<00:02,  2.30s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:07<00:00,  1.95s/it]
Processing chunks:  50%|█████     | 2/4 [00:16<00:16,  8.07s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:02<00:07,  2.40s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:04<00:04,  2.17s/it]
Extracting embeddings:  75%|███████▌  | 3/4 [00:06<00:02,  2.09s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:07<00:00,  1.82s/it]
Processing chunks:  75%|███████▌  | 3/4 [00:23<00:07,  7.72s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:01<00:05,  2.00s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:03<00:03,  1.94s/it]
Extracting embeddings:  75%|███████▌  | 3/4 [00:05<00:01,  1.97s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:06<00:00,  1.65s/it]
Processing chunks: 100%|██████████| 4/4 [00:30<00:00,  7.52s/it]
Processing chunks:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:01<00:05,  2.00s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:04<00:04,  2.06s/it]
```

```
Extracting embeddings:  75%|          | 3/4 [00:06<00:01,  1.99s/it]
Extracting embeddings: 100%|          | 4/4 [00:06<00:00,  1.73s/it]
Processing chunks:  25%|          | 1/4 [00:06<00:20,  6.90s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|          | 1/4 [00:02<00:06,  2.10s/it]
Extracting embeddings:  50%|          | 2/4 [00:04<00:04,  2.04s/it]
Extracting embeddings:  75%|          | 3/4 [00:06<00:02,  2.21s/it]
Extracting embeddings: 100%|          | 4/4 [00:07<00:00,  1.83s/it]
Processing chunks:  50%|          | 2/4 [00:14<00:14,  7.20s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|          | 1/4 [00:02<00:06,  2.00s/it]
Extracting embeddings:  50%|          | 2/4 [00:04<00:04,  2.24s/it]
Extracting embeddings:  75%|          | 3/4 [00:06<00:02,  2.22s/it]
Extracting embeddings: 100%|          | 4/4 [00:07<00:00,  1.80s/it]
Processing chunks:  75%|          | 3/4 [00:21<00:07,  7.24s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|          | 1/4 [00:02<00:06,  2.00s/it]
Extracting embeddings:  50%|          | 2/4 [00:04<00:04,  2.06s/it]
Extracting embeddings:  75%|          | 3/4 [00:06<00:01,  1.99s/it]
Extracting embeddings: 100%|          | 4/4 [00:06<00:00,  1.72s/it]
Processing chunks: 100%|          | 4/4 [00:28<00:00,  7.15s/it]
Processing chunks:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|          | 1/4 [00:01<00:05,  1.80s/it]
Extracting embeddings:  50%|          | 2/4 [00:03<00:03,  1.92s/it]
Extracting embeddings:  75%|          | 3/4 [00:05<00:01,  1.96s/it]
Extracting embeddings: 100%|          | 4/4 [00:06<00:00,  1.68s/it]
Processing chunks:  25%|          | 1/4 [00:06<00:20,  6.89s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|          | 1/4 [00:01<00:05,  2.00s/it]
Extracting embeddings:  50%|          | 2/4 [00:03<00:03,  2.00s/it]
Extracting embeddings:  75%|          | 3/4 [00:06<00:02,  2.05s/it]
```

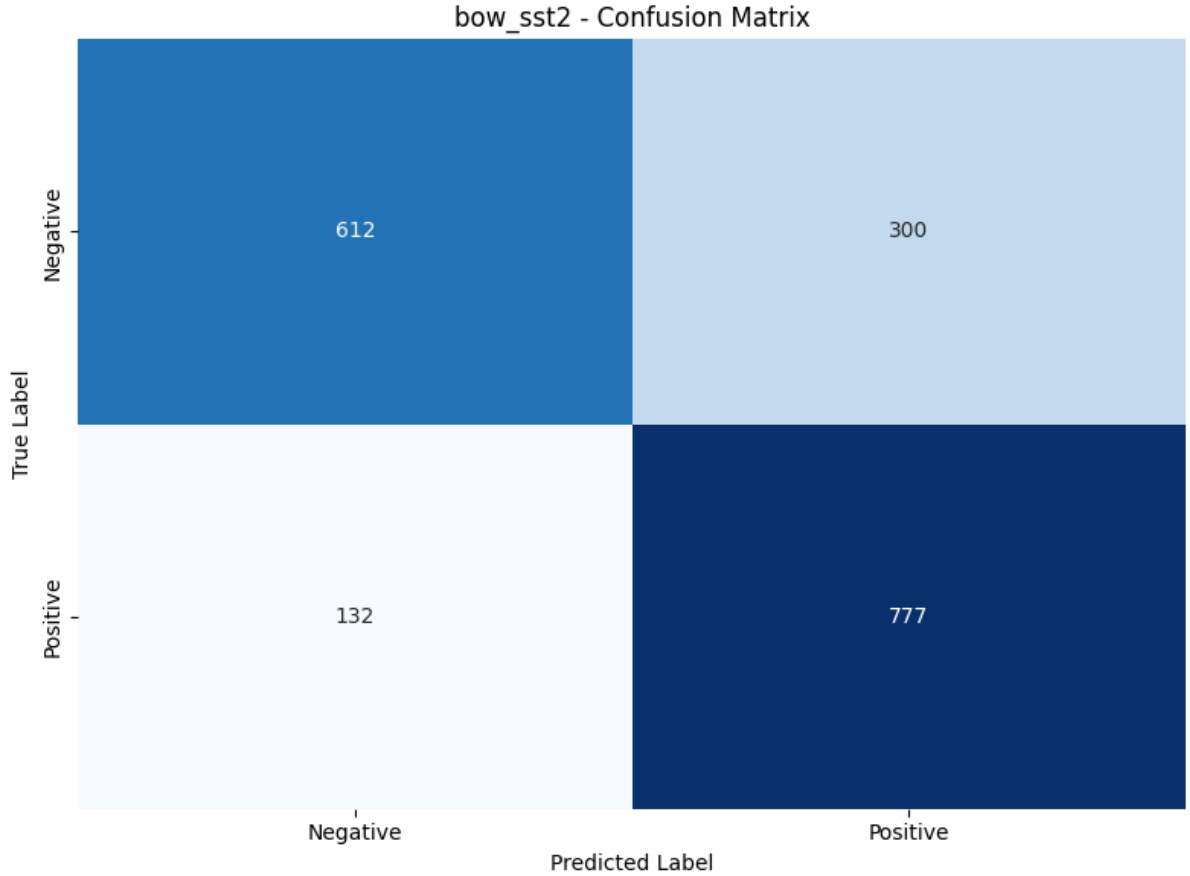
```
Extracting embeddings: 100%|          | 4/4 [00:06<00:00,  1.73s/it]
Processing chunks:  50%|          | 2/4 [00:13<00:13,  6.90s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|          | 1/4 [00:02<00:06,  2.30s/it]
Extracting embeddings:  50%|          | 2/4 [00:04<00:04,  2.18s/it]
Extracting embeddings:  75%|          | 3/4 [00:06<00:02,  2.24s/it]
Extracting embeddings: 100%|          | 4/4 [00:07<00:00,  1.88s/it]
Processing chunks:  75%|          | 3/4 [00:21<00:07,  7.26s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|          | 1/4 [00:02<00:07,  2.40s/it]
Extracting embeddings:  50%|          | 2/4 [00:04<00:04,  2.05s/it]
Extracting embeddings:  75%|          | 3/4 [00:06<00:02,  2.03s/it]
Extracting embeddings: 100%|          | 4/4 [00:06<00:00,  1.73s/it]
Processing chunks: 100%|          | 4/4 [00:28<00:00,  7.12s/it]
```

```
Using device: cpu

==================================================
TRAINING BAG-OF-WORDS MODEL ON SST2 DATASET
==================================================
Epoch 1/10 [Train]: 100%|          | 87/87 [00:22<00:00,  3.82it/s]
Epoch 1/10 [Val]: 100%|          | 22/22 [00:01<00:00, 19.99it/s]
Epoch 1/10
Train Loss: 0.6059, Val Loss: 0.4862, Val Acc: 0.7726
Checkpoint saved at checkpoints/bow_sst2_best.pt with val_acc: 0.7726
Epoch 2/10 [Train]: 100%|          | 87/87 [00:24<00:00,  3.60it/s]
Epoch 2/10 [Val]: 100%|          | 22/22 [00:01<00:00, 18.37it/s]
Epoch 2/10
Train Loss: 0.2682, Val Loss: 0.4760, Val Acc: 0.7863
Checkpoint saved at checkpoints/bow_sst2_best.pt with val_acc: 0.7863
Epoch 3/10 [Train]: 100%|          | 87/87 [00:22<00:00,  3.80it/s]
Epoch 3/10 [Val]: 100%|          | 22/22 [00:00<00:00, 22.01it/s]
Epoch 3/10
Train Loss: 0.0652, Val Loss: 0.9570, Val Acc: 0.7834
Epoch 4/10 [Train]: 100%|          | 87/87 [00:23<00:00,  3.69it/s]
Epoch 4/10 [Val]: 100%|          | 22/22 [00:00<00:00, 22.04it/s]
Epoch 4/10
Train Loss: 0.0145, Val Loss: 1.2627, Val Acc: 0.7733
Epoch 5/10 [Train]: 100%|          | 87/87 [00:24<00:00,  3.59it/s]
Epoch 5/10 [Val]: 100%|          | 22/22 [00:01<00:00, 21.95it/s]
Epoch 5/10
Train Loss: 0.0035, Val Loss: 1.7603, Val Acc: 0.7653
Epoch 6/10 [Train]: 100%|          | 87/87 [00:22<00:00,  3.80it/s]
Epoch 6/10 [Val]: 100%|          | 22/22 [00:00<00:00, 22.05it/s]
Epoch 6/10
Train Loss: 0.0023, Val Loss: 1.8864, Val Acc: 0.7682
Epoch 7/10 [Train]: 100%|          | 87/87 [00:22<00:00,  3.78it/s]
Epoch 7/10 [Val]: 100%|          | 22/22 [00:01<00:00, 18.31it/s]
Epoch 7/10
Train Loss: 0.0112, Val Loss: 1.1914, Val Acc: 0.7740
Epoch 8/10 [Train]: 100%|          | 87/87 [00:24<00:00,  3.49it/s]
Epoch 8/10 [Val]: 100%|          | 22/22 [00:01<00:00, 19.94it/s]
Epoch 8/10
Train Loss: 0.0067, Val Loss: 1.4248, Val Acc: 0.7661
Epoch 9/10 [Train]: 100%|          | 87/87 [00:25<00:00,  3.45it/s]
Epoch 9/10 [Val]: 100%|          | 22/22 [00:01<00:00, 18.36it/s]
Epoch 9/10
Train Loss: 0.0036, Val Loss: 1.7490, Val Acc: 0.7617
Epoch 10/10 [Train]: 100%|          | 87/87 [00:25<00:00,  3.47it/s]
Epoch 10/10 [Val]: 100%|          | 22/22 [00:01<00:00, 18.37it/s]
Epoch 10/10
Train Loss: 0.0069, Val Loss: 1.5738, Val Acc: 0.7574
Final model saved at checkpoints/bow_sst2_final.pt
```

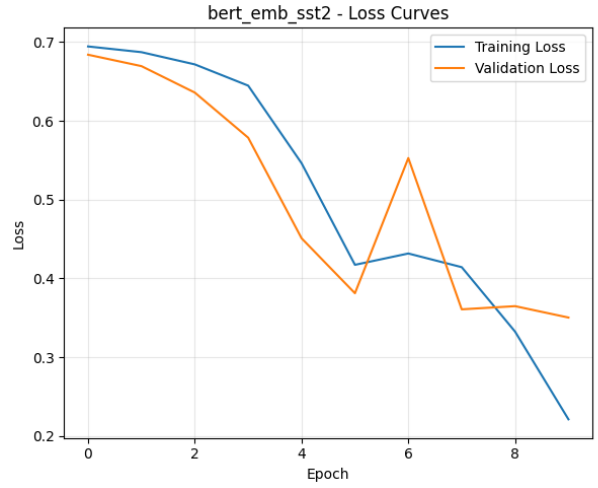


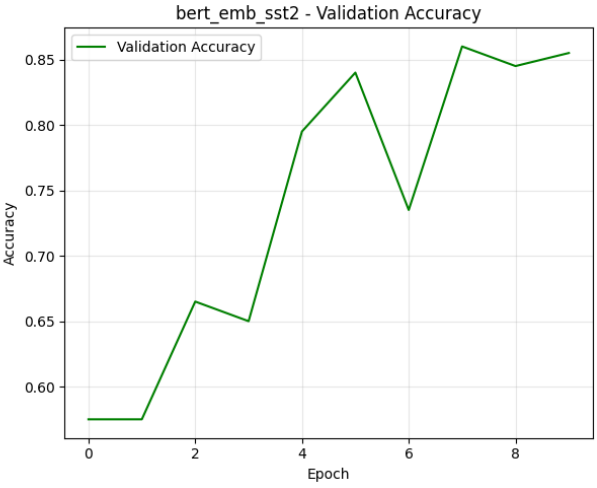



```
Evaluating: 100%|██████████| 29/29 [00:01<00:00, 21.88it/s]

BoW Model Test Loss: 1.5643, Test Accuracy: 0.7628

Classification Report:
              precision    recall  f1-score   support

    Negative       0.82      0.67      0.74       912
    Positive       0.72      0.85      0.78       909

    accuracy                           0.76      1821
   macro avg       0.77      0.76      0.76      1821
weighted avg       0.77      0.76      0.76      1821
```

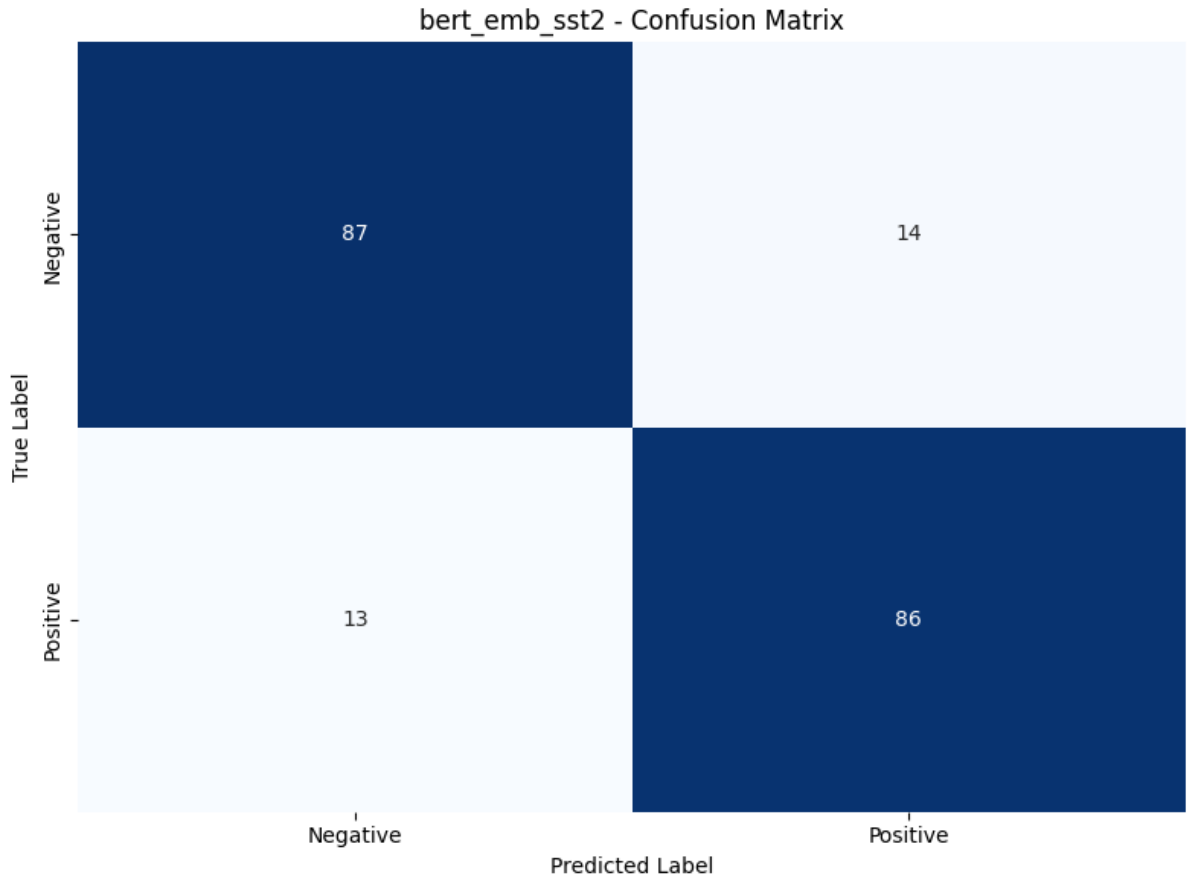


```
TRAINING BERT EMBEDDINGS MODEL ON SST2 DATASET
==================================================
Epoch 1/10 [Train]: 100%|██████████| 7/7 [00:00<00:00,  9.74it/s]
Epoch 1/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.46it/s]
Epoch 1/10
Train Loss: 0.6943, Val Loss: 0.6839, Val Acc: 0.5750
Checkpoint saved at checkpoints/bert_emb_sst2_best.pt with val_acc: 0.5750
Epoch 2/10 [Train]: 100%|██████████| 7/7 [00:00<00:00, 11.92it/s]
Epoch 2/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.76it/s]
Epoch 2/10
Train Loss: 0.6871, Val Loss: 0.6694, Val Acc: 0.5750
Epoch 3/10 [Train]: 100%|██████████| 7/7 [00:00<00:00, 10.03it/s]
Epoch 3/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.76it/s]
Epoch 3/10
Train Loss: 0.6716, Val Loss: 0.6358, Val Acc: 0.6650
Checkpoint saved at checkpoints/bert_emb_sst2_best.pt with val_acc: 0.6650
Epoch 4/10 [Train]: 100%|██████████| 7/7 [00:00<00:00, 10.07it/s]
Epoch 4/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.32it/s]
Epoch 4/10
Train Loss: 0.6446, Val Loss: 0.5786, Val Acc: 0.6500
Epoch 5/10 [Train]: 100%|██████████| 7/7 [00:00<00:00,  8.79it/s]
Epoch 5/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.62it/s]
Epoch 5/10
Train Loss: 0.5462, Val Loss: 0.4510, Val Acc: 0.7950
Checkpoint saved at checkpoints/bert_emb_sst2_best.pt with val_acc: 0.7950
Epoch 6/10 [Train]: 100%|██████████| 7/7 [00:00<00:00,  7.91it/s]
Epoch 6/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.65it/s]
Epoch 6/10
Train Loss: 0.4173, Val Loss: 0.3813, Val Acc: 0.8400
Checkpoint saved at checkpoints/bert_emb_sst2_best.pt with val_acc: 0.8400
Epoch 7/10 [Train]: 100%|██████████| 7/7 [00:00<00:00,  8.84it/s]
Epoch 7/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.44it/s]
Epoch 7/10
Train Loss: 0.4317, Val Loss: 0.5528, Val Acc: 0.7350
Epoch 8/10 [Train]: 100%|██████████| 7/7 [00:00<00:00, 10.02it/s]
Epoch 8/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.46it/s]
Epoch 8/10
Train Loss: 0.4143, Val Loss: 0.3608, Val Acc: 0.8600
Checkpoint saved at checkpoints/bert_emb_sst2_best.pt with val_acc: 0.8600
Epoch 9/10 [Train]: 100%|██████████| 7/7 [00:00<00:00,  8.72it/s]
Epoch 9/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 23.50it/s]
Epoch 9/10
Train Loss: 0.3326, Val Loss: 0.3649, Val Acc: 0.8450
Epoch 10/10 [Train]: 100%|██████████| 7/7 [00:00<00:00,  8.80it/s]
Epoch 10/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.15it/s]
Epoch 10/10
Train Loss: 0.2213, Val Loss: 0.3504, Val Acc: 0.8550
Final model saved at checkpoints/bert_emb_sst2_final.pt
```






```
Evaluating: 100%|██████████| 7/7 [00:00<00:00, 714.48it/s]

BERT Embeddings Model Test Loss: 0.3649, Test Accuracy: 0.8650

Classification Report:
              precision    recall  f1-score   support

    Negative       0.87      0.86      0.87       101
    Positive       0.86      0.87      0.86        99

    accuracy                           0.86       200
   macro avg       0.86      0.87      0.86       200
weighted avg       0.87      0.86      0.87       200
```

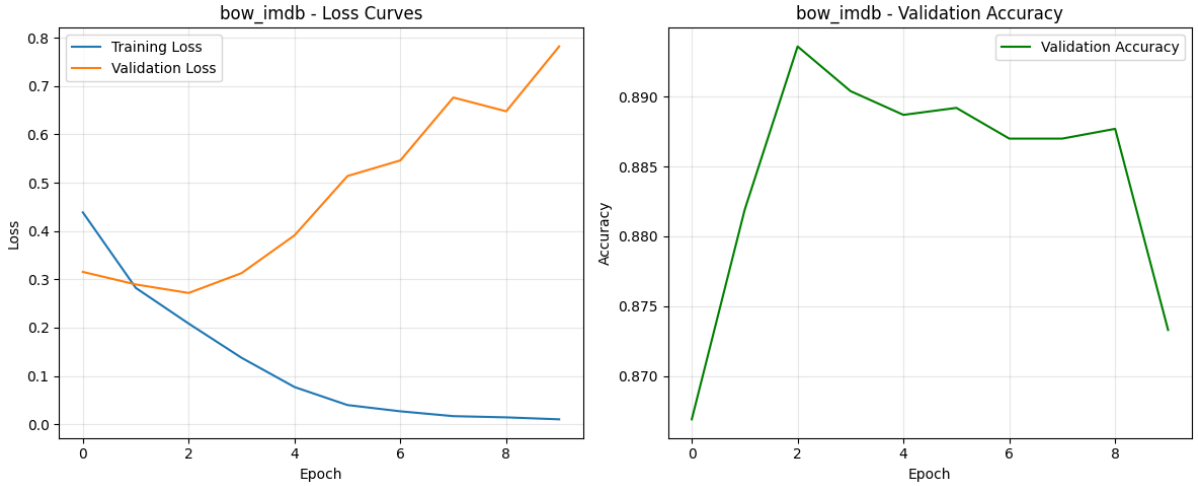



```
==================================================
COMPARING MODELS ON SST2 DATASET
==================================================
BoW Model Best Validation Accuracy: 0.7863
BoW Model Test Accuracy: 0.7628
BERT Embeddings Model Best Validation Accuracy: 0.8600
BERT Embeddings Model Test Accuracy: 0.8650
```

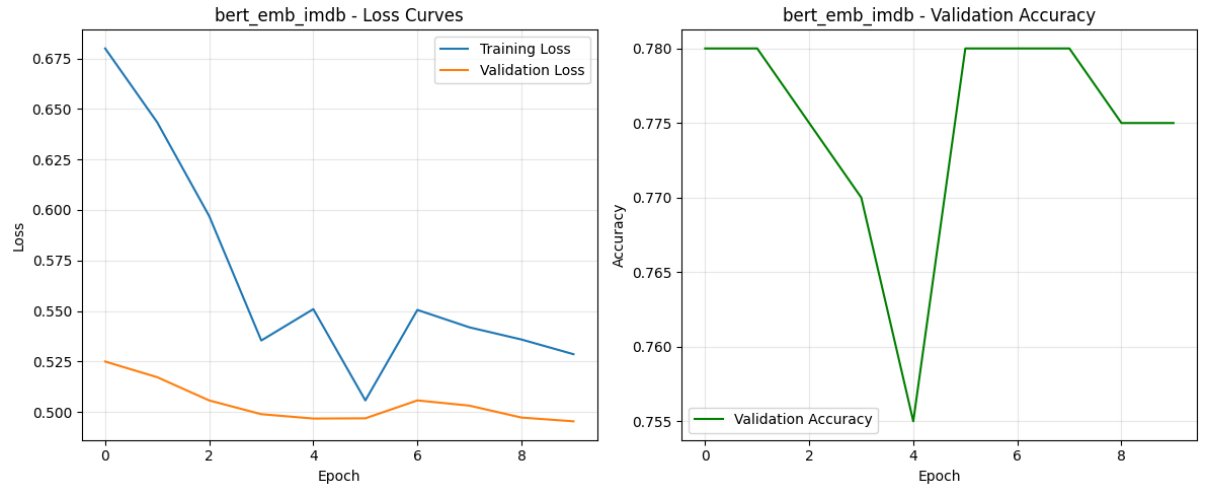
```
Processing IMDB training data for embeddings...
Processing chunks:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:05<00:15,  5.27s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:09<00:09,  4.87s/it]
Extracting embeddings:  75%|███████▌  | 3/4 [00:14<00:04,  4.89s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:16<00:00,  4.02s/it]
Processing chunks:  25%|██▌       | 1/4 [00:16<00:48, 16.07s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:03<00:11,  3.90s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:08<00:08,  4.02s/it]
Extracting embeddings:  75%|███████▌  | 3/4 [00:12<00:04,  4.06s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:13<00:00,  3.38s/it]
Processing chunks:  50%|█████     | 2/4 [00:29<00:29, 14.62s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:04<00:12,  4.10s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:08<00:08,  4.16s/it]
Extracting embeddings:  75%|███████▌  | 3/4 [00:12<00:04,  4.13s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:13<00:00,  3.40s/it]
Processing chunks:  75%|███████▌  | 3/4 [00:43<00:14, 14.24s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:04<00:12,  4.00s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:08<00:08,  4.06s/it]
Extracting embeddings:  75%|███████▌  | 3/4 [00:12<00:04,  4.08s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:13<00:00,  3.28s/it]
Processing chunks: 100%|██████████| 4/4 [00:56<00:00, 14.17s/it]
Processing IMDB validation data for embeddings...
Processing chunks:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:03<00:11,  3.99s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:07<00:07,  3.94s/it]
Extracting embeddings:  75%|███████▌  | 3/4 [00:12<00:04,  4.06s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:13<00:00,  3.35s/it]
Processing chunks:  25%|██▌       | 1/4 [00:13<00:40, 13.40s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:04<00:12,  4.10s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:08<00:08,  4.16s/it]
Extracting embeddings:  75%|███████▌  | 3/4 [00:12<00:04,  4.23s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:13<00:00,  3.50s/it]
Processing chunks:  50%|█████     | 2/4 [00:27<00:27, 13.81s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:03<00:11,  3.90s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:07<00:07,  3.96s/it]
Extracting embeddings:  75%|███████▌  | 3/4 [00:12<00:04,  4.07s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:13<00:00,  3.35s/it]
Processing chunks:  75%|███████▌  | 3/4 [00:40<00:13, 13.62s/it]
Extracting embeddings:   0%|          | 0/4 [00:00<?, ?it/s]
Extracting embeddings:  25%|██▌       | 1/4 [00:03<00:11,  3.80s/it]
Extracting embeddings:  50%|█████     | 2/4 [00:07<00:07,  3.86s/it]
```

```
Extracting embeddings:  75%|███████▌  | 3/4 [00:11<00:03,  3.97s/it]
Extracting embeddings: 100%|██████████| 4/4 [00:13<00:00,  3.27s/it]
Processing chunks: 100%|██████████| 4/4 [00:54<00:00, 13.52s/it]

================================================
CONTINUING TRAINING OF BAG-OF-WORDS MODEL ON IMDB DATASET
================================================
Loaded BoW model checkpoint from checkpoints/bow_sst2_best.pt
Checkpoint info - Epoch: 1, Val Acc: 0.7863
Epoch 1/10 [Train]: 100%|██████████| 625/625 [01:59<00:00,  5.21it/s]
Epoch 1/10 [Val]: 100%|██████████| 157/157 [00:08<00:00, 19.39it/s]
Epoch 1/10
Train Loss: 0.4386, Val Loss: 0.3153, Val Acc: 0.8669
Checkpoint saved at checkpoints/bow_imdb_best.pt with val_acc: 0.8669
Epoch 2/10 [Train]: 100%|██████████| 625/625 [02:06<00:00,  4.94it/s]
Epoch 2/10 [Val]: 100%|██████████| 157/157 [00:07<00:00, 20.12it/s]
Epoch 2/10
Train Loss: 0.2822, Val Loss: 0.2893, Val Acc: 0.8819
Checkpoint saved at checkpoints/bow_imdb_best.pt with val_acc: 0.8819
Epoch 3/10 [Train]: 100%|██████████| 625/625 [02:03<00:00,  5.07it/s]
Epoch 3/10 [Val]: 100%|██████████| 157/157 [00:08<00:00, 18.26it/s]
Epoch 3/10
Train Loss: 0.2083, Val Loss: 0.2717, Val Acc: 0.8936
Checkpoint saved at checkpoints/bow_imdb_best.pt with val_acc: 0.8936
Epoch 4/10 [Train]: 100%|██████████| 625/625 [02:04<00:00,  5.01it/s]
Epoch 4/10 [Val]: 100%|██████████| 157/157 [00:07<00:00, 19.88it/s]
Epoch 4/10
Train Loss: 0.1374, Val Loss: 0.3129, Val Acc: 0.8904
Epoch 5/10 [Train]: 100%|██████████| 625/625 [02:07<00:00,  4.90it/s]
Epoch 5/10 [Val]: 100%|██████████| 157/157 [00:07<00:00, 20.15it/s]
Epoch 5/10
Train Loss: 0.0768, Val Loss: 0.3911, Val Acc: 0.8887
Epoch 6/10 [Train]: 100%|██████████| 625/625 [02:10<00:00,  4.80it/s]
Epoch 6/10 [Val]: 100%|██████████| 157/157 [00:07<00:00, 20.67it/s]
Epoch 6/10
Train Loss: 0.0395, Val Loss: 0.5138, Val Acc: 0.8892
Epoch 7/10 [Train]: 100%|██████████| 625/625 [02:11<00:00,  4.77it/s]
Epoch 7/10 [Val]: 100%|██████████| 157/157 [00:08<00:00, 19.16it/s]
Epoch 7/10
Train Loss: 0.0266, Val Loss: 0.5463, Val Acc: 0.8870
Epoch 8/10 [Train]: 100%|██████████| 625/625 [02:13<00:00,  4.67it/s]
Epoch 8/10 [Val]: 100%|██████████| 157/157 [00:08<00:00, 18.92it/s]
Epoch 8/10
Train Loss: 0.0166, Val Loss: 0.6765, Val Acc: 0.8870
Epoch 9/10 [Train]: 100%|██████████| 625/625 [02:11<00:00,  4.76it/s]
Epoch 9/10 [Val]: 100%|██████████| 157/157 [00:08<00:00, 19.15it/s]
```

```
Epoch 9/10
Train Loss: 0.0141, Val Loss: 0.6479, Val Acc: 0.8877
Epoch 10/10 [Train]: 100%|██████████| 625/625 [02:13<00:00,  4.67it/s]
Epoch 10/10 [Val]: 100%|██████████| 157/157 [00:08<00:00, 18.47it/s]
Epoch 10/10
Train Loss: 0.0100, Val Loss: 0.7823, Val Acc: 0.8733
Final model saved at checkpoints/bow_imdb_final.pt
```



```
CONTINUING TRAINING OF BERT EMBEDDINGS MODEL ON IMDB DATASET
================================================
Loaded BERT Embeddings model checkpoint from checkpoints/bert_emb_sst2_best.pt
Checkpoint info - Epoch: 7, Val Acc: 0.8600
Epoch 1/10 [Train]: 100%|██████████| 7/7 [00:00<00:00, 10.55it/s]
Epoch 1/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.43it/s]
Epoch 1/10
Train Loss: 0.6800, Val Loss: 0.5249, Val Acc: 0.7800
Checkpoint saved at checkpoints/bert_emb_imdb_best.pt with val_acc: 0.7800
Epoch 2/10 [Train]: 100%|██████████| 7/7 [00:00<00:00,  8.87it/s]
Epoch 2/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.49it/s]
Epoch 2/10
Train Loss: 0.6434, Val Loss: 0.5172, Val Acc: 0.7800
Epoch 3/10 [Train]: 100%|██████████| 7/7 [00:00<00:00,  8.80it/s]
Epoch 3/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.44it/s]
Epoch 3/10
Train Loss: 0.5969, Val Loss: 0.5057, Val Acc: 0.7750
Epoch 4/10 [Train]: 100%|██████████| 7/7 [00:00<00:00,  8.68it/s]
Epoch 4/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.08it/s]
Epoch 4/10
Train Loss: 0.5354, Val Loss: 0.4989, Val Acc: 0.7700
Epoch 5/10 [Train]: 100%|██████████| 7/7 [00:00<00:00,  8.76it/s]
Epoch 5/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.41it/s]
Epoch 5/10
Train Loss: 0.5509, Val Loss: 0.4967, Val Acc: 0.7550
Epoch 6/10 [Train]: 100%|██████████| 7/7 [00:00<00:00,  8.68it/s]
Epoch 6/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 23.43it/s]
Epoch 6/10
Train Loss: 0.5057, Val Loss: 0.4969, Val Acc: 0.7800
Epoch 7/10 [Train]: 100%|██████████| 7/7 [00:00<00:00,  7.79it/s]
Epoch 7/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.23it/s]
Epoch 7/10
Train Loss: 0.5505, Val Loss: 0.5057, Val Acc: 0.7800
Epoch 8/10 [Train]: 100%|██████████| 7/7 [00:00<00:00,  8.73it/s]
Epoch 8/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.62it/s]
Epoch 8/10
Train Loss: 0.5419, Val Loss: 0.5031, Val Acc: 0.7800
Epoch 9/10 [Train]: 100%|██████████| 7/7 [00:00<00:00,  9.98it/s]
Epoch 9/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.93it/s]
Epoch 9/10
Train Loss: 0.5359, Val Loss: 0.4972, Val Acc: 0.7750
Epoch 10/10 [Train]: 100%|██████████| 7/7 [00:00<00:00,  9.98it/s]
Epoch 10/10 [Val]: 100%|██████████| 7/7 [00:00<00:00, 35.47it/s]
Epoch 10/10
Train Loss: 0.5286, Val Loss: 0.4954, Val Acc: 0.7750
Final model saved at checkpoints/bert_emb_imdb_final.pt
```




```
================================================
COMPARING MODELS ON IMDB DATASET
================================================
BoW Model Best Validation Accuracy: 0.8936
BERT Embeddings Model Best Validation Accuracy: 0.7800

================================================
FINAL COMPARISON ACROSS DATASETS
================================================
Bag-of-Words Model:
  SST2 Best Val Accuracy: 0.7863
  IMDB Best Val Accuracy: 0.8936

BERT Embeddings Model:
  SST2 Best Val Accuracy: 0.8600
  IMDB Best Val Accuracy: 0.7800
```