News Headline Classifier – Summary Report

1. Project Overview

This project fine-tunes a BERT model (bert-base-uncased) to classify BBC news headlines into five categories:

- Business
- Entertainment
- Politics
- Sport
- Tech

The model uses a frozen BERT backbone with a linear classifier on top, enabling fast and efficient fine-tuning on a relatively small dataset.

2. Hyperparameters

Hyperparameter	-	**Value**
Model Name	-	bert-base-uncased
Number of Labels	-	5
Batch Size	-	8
Learning Rate	-	5e-5
Epochs	-	10
Maximum Sequence Length	-	128
Optimizer	-	AdamW
Loss Function	-	CrossEntropyLoss

3. Training Details

- Frozen BERT Layers: Only the linear classifier was trained, reducing computation and minimizing overfitting.
- Learning Rate: 5e-5 provided stable convergence.
- Batch Size: 8 was optimal for GPU memory usage.
- Data Splits: Training (80%), Validation (10%), Test (10%) using train_test_split.
- Epochs: 10 epochs allowed sufficient learning without overfitting.
- Evaluation Metrics: Accuracy, Precision, Recall, and F1-score (macro) were used to ensure balanced performance across all classes.

4. Dataset Class Distribution

Class	Count
Politics	275
Entertainment	286
Sport	210
Business	212
Tech	242

Minor class imbalance exists but does not significantly impact macro metrics.

5. Model Performance

Epoch 1/10

Train Loss: 1.6089 | Train Acc: 0.2683

Val Loss: 1.5347 | Val Acc: 0.3828

Epoch 2/10

Train Loss: 1.4938 | Train Acc: 0.4411

Val Loss: 1.4373 | Val Acc: 0.5469

Epoch 3/10

Train Loss: 1.3973 | Train Acc: 0.5874

Val Loss: 1.3503 | Val Acc: 0.6172

Epoch 4/10

Train Loss: 1.3148 | Train Acc: 0.6778

Val Loss: 1.2679 | Val Acc: 0.6875

Epoch 5/10

Train Loss: 1.2419 | Train Acc: 0.7114

Val Loss: 1.1915 | Val Acc: 0.7422

Epoch 6/10

Train Loss: 1.1710 | Train Acc: 0.7734

Val Loss: 1.1209 | Val Acc: 0.8125

Epoch 7/10

Train Loss: 1.1095 | Train Acc: 0.8028

Val Loss: 1.0549 | Val Acc: 0.8359

Epoch 8/10

Train Loss: 1.0484 | Train Acc: 0.8354

Val Loss: 0.9930 | Val Acc: 0.8828

Epoch 9/10

Train Loss: 0.9922 | Train Acc: 0.8709

Val Loss: 0.9364 | Val Acc: 0.8984

Epoch 10/10

Train Loss: 0.9404 | Train Acc: 0.8923

Val Loss: 0.8841 | Val Acc: 0.9141

For Test data ,

Loss : 0.6674

Accuracy : 0.9187

Precision : 0.9122

Recall : 0.9267

F1 Score : 0.9145

6. Inference

Users can input any news headline, and the model outputs class probabilities.

Label Mapping:

0 → Business

1 → Entertainment

2 → Politics

3 → Sport

4 → Tech

Example:

- Input: "Stock markets hit record highs"
- Predicted Class: Business

7. Key Insights & Takeaways

1. Freezing BERT embeddings reduces training time while maintaining accuracy.
2. Small learning rate and AdamW optimizer ensure stable convergence.
3. Tokenizing headlines to a max length of 128 covers most inputs.
4. Saving test datasets and checkpoints ensures reproducible evaluation.
5. Future improvements:
    - Unfreeze last few BERT layers for deeper fine-tuning.
    - Experiment with other Transformer architectures (e.g., RoBERTa, DistilBERT).
    - Use data augmentation to increase dataset size and diversity.

Conclusion:

The fine-tuned BERT classifier performs effectively for BBC news headline categorization. It provides reliable predictions, efficient inference, and a strong baseline for further experimentation and deployment.