Name: Muaaz Hassan Ansari

Id: DHC-107

News Topic Classifier Using BERT

# Objective

The objective of this project is to fine-tune a transformer-based model (BERT) to classify news headlines into topic categories using the AG News dataset. The project demonstrates preprocessing, fine-tuning, evaluation, and deployment of a text classification model.

# Dataset

The AG News dataset is used, which contains news headlines categorized into four classes:  
1. World  
2. Sports  
3. Business  
4. Sci/Tech

# Methodology

1. Load and preprocess the AG News dataset.  
2. Tokenize the text using BERT tokenizer.  
3. Fine-tune the pre-trained BERT model (bert-base-uncased) for classification.  
4. Evaluate the model using accuracy and F1-score.  
5. Deploy the model using Gradio for interactive testing.

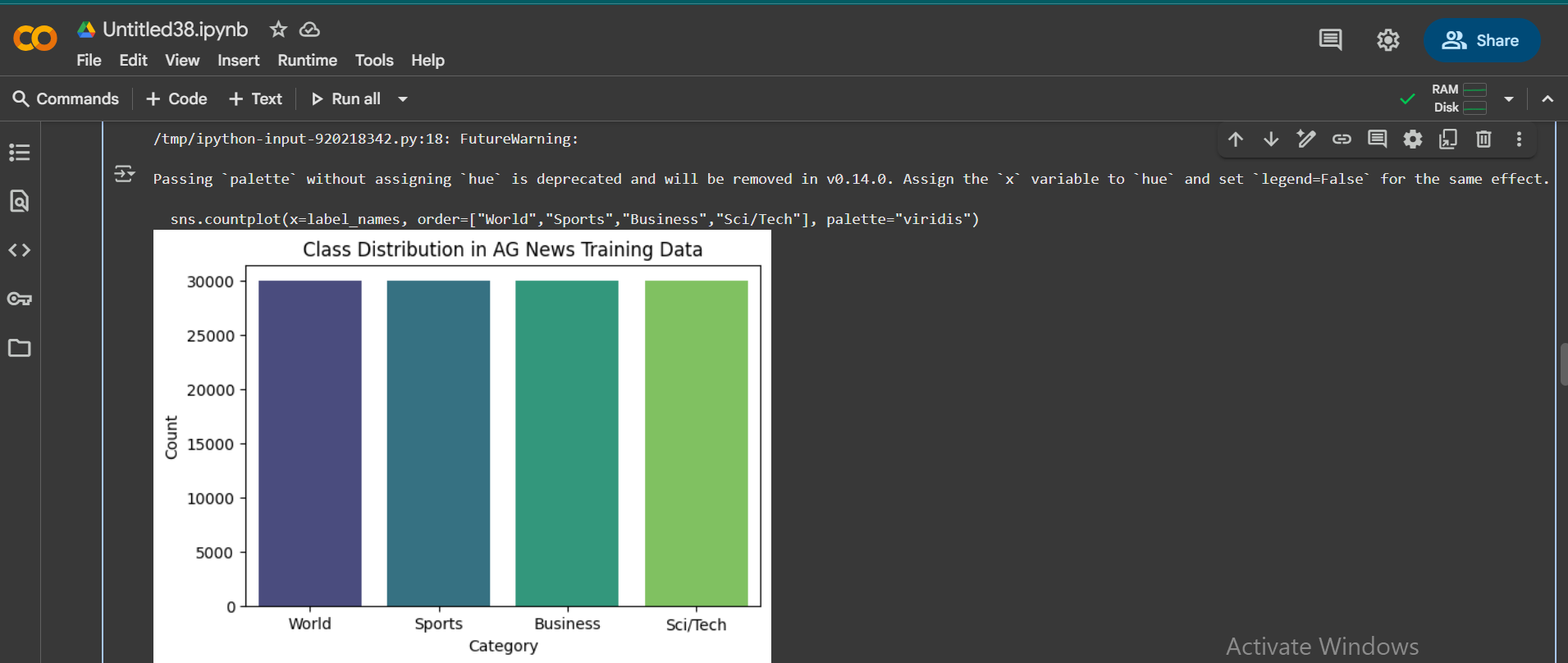
# Training Details

The model was trained using Hugging Face Transformers with the following configuration:  
- Pre-trained model: bert-base-uncased  
- Batch size: 16  
- Optimizer: AdamW  
- Epochs: 2-3  
- Evaluation metrics: Accuracy, F1-score

# Results

The fine-tuned BERT model achieved the following performance on the validation set:  
  
Epoch 1: Validation Loss = 0.228, Accuracy = 92.6%, F1-macro = 0.923  
Epoch 2: Validation Loss = 0.285, Accuracy = 92.0%, F1-macro = 0.917  
  
These results demonstrate that the model performs well on the AG News dataset.

# Visualizations

Several visualizations were created to better understand the dataset and model performance:  


# Deployment

The model was deployed using Gradio, allowing users to input custom news headlines and receive predicted categories. This provides an interactive way to test the classifier.

# Results

