README: Student Feedback Sentiment Classifier

🎓 Student Feedback Sentiment Classifier (Fine-Tuned BERT)  
  
This project fine-tunes a pre-trained BERT model to classify student course feedback into three sentiment categories: Positive, Neutral, and Negative.  
  
It was built as part of a final assignment for a course on fine-tuning large language models, demonstrating domain adaptation, model evaluation, error analysis, and deployment via an inference pipeline.  
  
📌 Project Structure  
  
student-feedback-finetune/  
│  
├── train\_model.py # Fine-tunes the BERT model  
├── evaluate\_model.py # Evaluates the fine-tuned model  
├── evaluate\_baseline.py # Evaluates pre-trained (non-fine-tuned) BERT  
├── analyze\_errors.py # Extracts misclassified test samples  
├── inference\_pipeline.py # Runs live predictions with confidence scores  
├── tokenize\_dataset.py # Tokenizes and splits dataset  
├── misclassified\_examples.csv # Output of error analysis  
├── requirements.txt # Dependencies for running the code  
├── README.md # Project overview and usage instructions  
└── student\_feedback\_finetuned\_model/ # Saved model and tokenizer  
  
🧠 Task Overview  
  
- Model: bert-base-uncased  
- Task: Multi-class sentiment classification  
- Labels: 0 = Negative, 1 = Neutral, 2 = Positive  
  
🚀 How to Run  
  
1. Install dependencies:  
 pip install -r requirements.txt  
  
2. Tokenize and Prepare Dataset:  
 python tokenize\_dataset.py  
  
3. Train the Model:  
 python train\_model.py  
  
4. Evaluate the Fine-Tuned Model:  
 python evaluate\_model.py  
  
5. Compare with Baseline:  
 python evaluate\_baseline.py  
  
6. Run Error Analysis:  
 python analyze\_errors.py  
  
7. Run Inference:  
 python inference\_pipeline.py  
  
📊 Model Performance  
  
| Metric | Fine-Tuned | Baseline (Untrained BERT) |  
|--------------|------------|----------------------------|  
| Accuracy | 98.18% | 43.63% |  
| F1 Score | 98.19% | 38.52% |  
  
🧪 Example Predictions  
  
Input: "The teaching is terrible and unfair." → Prediction: Negative (Confidence: 98.7%)  
Input: "Library is okay but can improve." → Prediction: Neutral (Confidence: 98.1%)  
Input: "The course was good." → Prediction: Positive (Confidence: 79.1%)  
  
📂 Dataset  
  
Custom CSV file containing student feedback and labeled sentiment. The dataset is tokenized and saved using Hugging Face Datasets format (load\_from\_disk()).  
  
🛠 Dependencies  
  
See requirements.txt for details.  
  
👨‍💻 Author  
  
Anurag K. Vadapalli   
Fine-Tuning LLMs – Final Project (2025)   
Northeastern University