Fine Tuning DistilBERT for Emotion Recognition

1. Executive Summary

This report details the successful implementation and fine tuning of a Large Language Model (LLM) for the purpose of emotion recognition. Using the open source DistilBERT architecture and the dair ai/emotion dataset i achieved a classification accuracy of 93.85% on the validation set. The training process was executed on a local environment equipped with an NVIDIA RTX 3070 Ti demonstrating the feasibility of high performance NLP tasks on consumer grade hardware without reliance on SaaS APIs.

2. Experimental Setup

2.1 Hardware & Environment

The experiment was conducted in a local Jupyter Notebook environment using the PyTorch framework with CUDA acceleration.

- GPU: NVIDIA GeForce RTX 3070 Ti Laptop GPU
- Compute Platform: CUDA 11.8
- Key Libraries: transformers (4.57.3), datasets (4.4.1), evaluate, scikit-learn.

2.2 Dataset

I utilized the publicly available dair-ai/emotion dataset from the Hugging Face Hub.

- Task: Multi-class Text Classification.
- Labels (6 classes): Sadness, Joy, Love, Anger, Fear, Surprise.
- Data Split:
  Training Set: 16,000 samples
  Validation Set: 2,000 samples

2.3 Model Architecture

- Base Model: distilbert-base-uncased (a distilled version of BERT, roughly 40% smaller and 60% faster while retaining 97% of performance).
- Modification: A sequence classification head was added with 6 output neurons corresponding to the emotion labels.
- Tokenizer: Standard DistilBERT tokenizer with padding and truncation set to a maximum length of 512 tokens.

3. Training Methodology

The model was fine tuned using the Trainer API from the Hugging Face transformers library. Mixed precision training (FP16) was enabled to optimize memory usage and training speed on the RTX 3070 Ti.

Hyperparameters:

- Epochs: 3
- Batch Size: 16 (Train) / 32 (Eval)
- Learning Rate: 2e-5
- Weight Decay: 0.01
- Optimization Strategy: Load best model at end (metric: Accuracy).

4. Results & Analysis

4.1 Training Progression

The model showed rapid convergence reaching over 92 accuracy within the first epoch. Training took approximately 9 minutes and 13 seconds (553 seconds total).

Epoch	Training Loss	Validation Loss	Accuracy
1	0.2604	0.2016	**92.80%**
2	0.1477	0.1847	**93.30%**
3	0.0970	0.1653	**93.85%**

4.2 Qualitative Evaluation (Inference)

Following training, the model was tested on unseen user generated text to verify real world applicability. The model correctly classified all test inputs:

1. *"I passed my exam!"* -> Joy (Correct)
2. *"I am so happy today!"* -> Joy (Correct)
3. *"I am furious about this."* -> Anger (Correct)

5. Conclusion

The experiment successfully demonstrated that open source LLMs can be effectively fine tuned for specialized NLP tasks like emotion recognition with minimal computational resources. Achieving 93.85 accuracy in under 10 minutes of training time validates the efficiency of DistilBERT for production grade text classification. The model and tokenizer were successfully saved (my_emotion_model) and are ready for deployment or inference in downstream applications.