**Tweet Emotion Detection using PEFT and Prompt-Based Fine-Tuning**

**Overview**

This project extends tweet sentiment classification using **Parameter-Efficient Fine-Tuning (PEFT)** and **pre-trained Transformer models** with a focus on **prompt-based learning**. The work is divided into three parts:

* **Part 1**: Data preprocessing and prompt generation.
* **Part 2**: Model fine-tuning using PEFT techniques.
* **Part 3**: Evaluation, logging, and inference.

**What is PEFT (Parameter-Efficient Fine-Tuning)?**

PEFT allows fine-tuning large-scale models efficiently by modifying only **small, trainable layers** instead of updating the entire model. This is useful for **memory-constrained environments** and speeds up training while preserving model accuracy.

**What is Prompt-Based Tuning?**

Prompt-Based Tuning is a fine-tuning method where models learn from **structured prompts** instead of updating all their parameters. This allows efficient adaptation for sentiment classification and reduces the need for extensive labeled data.

**Why Use Prompt-Based Tuning?**

* **Improves adaptability**: Guides the model’s reasoning with structured prompts instead of direct classification.
* **Reduces memory usage**: Works well with **PEFT techniques** like LoRA.
* **Enhances generalization**: Prompts allow better transfer learning to unseen examples.

**Types of Prompt-Based Tuning Used in HW8**

1. **Prefix-Tuning**: Adds trainable embeddings before input tokens without modifying the main model.
2. **Instruction-Tuning**: Uses explicit instructions in prompts (e.g., "Classify the emotion in this tweet: ...").
3. **Soft Prompt Tuning**: Uses learnable token embeddings instead of fixed text prompts, optimizing fine-tuning efficiency.

**How Prompt-Based Tuning Was Used in HW8**

* **Part 1**: Structured emotion classification prompts.
* **Part 2**: Injected prompts into **PEFT fine-tuned models** like LLaMA.
* **Part 3**: Evaluated models trained with prompt-based learning against traditional approaches. PEFT allows fine-tuning large-scale models efficiently by modifying only **small, trainable layers** instead of updating the entire model. This is useful for **memory-constrained environments** and speeds up training while preserving model accuracy.

**Why Use PEFT?**

* **Reduces memory usage**: Instead of updating all parameters, only small layers are trained.
* **Works with Large Models**: PEFT makes fine-tuning scalable even for large-scale transformers.
* **Supports Prompt-Based Learning**: Can be combined with **prompt tuning** for better adaptation.

**Models Used**

This project fine-tunes models using PEFT:

* **Qwen2.5-0.5B (Qwen2.5-0.5B (Base Model))** (Standard Transformer Baseline)
  + Serves as a benchmark for fine-tuning performance.
* **Meta LLaMA-3.2-1B-Instruct (Experiment 2)** (Optimized PEFT Model)
  + Uses PEFT fine-tuning with **prompt-based learning**.
  + Shows improved performance on key emotions like optimism, anger, and disgust.
* **Meta LLaMA-3.1-8B-Instruct (Meta LLaMA-3.1-8B-Instruct (Experiment 3))** (Alternative PEFT Approach)
  + Slightly different tuning approach, aiming for better fear and sadness detection.

**Model Comparison**

To evaluate the effectiveness of different fine-tuned transformer models, we compared their performance on tweet sentiment classification:

* **Qwen2.5-0.5B (Qwen2.5-0.5B (Base Model))**
  + Weak performance across all emotions, struggled with **anger, disgust, and fear detection**.
  + Low **trust detection (0.0037), love (0.0086), and optimism (0.0718)**.
* **Meta LLaMA-3.2-1B-Instruct (Meta LLaMA-3.2-1B-Instruct (Experiment 2) - Best Model)**
  + **Best in optimism detection (0.5158), anger (0.4283), and disgust (0.4262)**.
  + Improved performance compared to the base model.
* **Meta LLaMA-3.1-8B-Instruct (Meta LLaMA-3.1-8B-Instruct (Experiment 3))**
  + More balanced but slightly weaker than Meta LLaMA-3.2-1B-Instruct (Experiment 2).
  + Better in **fear detection (0.1469), sadness (0.3062)**, but still struggled with trust (0.0052).

**Conclusion**

After evaluating the models on tweet sentiment classification, **Meta LLaMA-3.2-1B-Instruct (Experiment 2) emerged as the best-performing model**. It achieved superior accuracy in capturing **optimism, anger, and disgust**, making it the most effective model for this task. **Meta LLaMA-3.1-8B-Instruct (Experiment 3) was a close second**, showing balanced performance across emotions.

The use of **PEFT and prompt-based fine-tuning** allowed efficient model adaptation, demonstrating that transformer-based approaches can be effectively tuned for tweet emotion detection while optimizing memory usage.