Summary

In this assignment, I fine-tuned the `bert-base-uncased` model for sentiment classification using the IMDb dataset via Hugging Face's `transformers` and `datasets` libraries. The project was implemented in **Google Colab**, using `transformers v4.52.4`, `datasets v3.6.0`, and PyTorch with **GPU acceleration**. To speed up training and save memory, **mixed-precision (`fp16=True`)** was enabled.

Due to runtime and memory limits, I initially trained on **5,000 samples**, which led to the model always predicting **"Positive"**. After investigating class imbalance and limited learning, I expanded the dataset to **20,000 training samples**, trained for **3 epochs**, and enabled evaluation using **accuracy and F1-score**. The fine-tuned model achieved an **accuracy of ~94%** and an **F1 score of ~96%** on a 1,000-sample test subset.

The model and tokenizer Ire saved locally and successfully reloaded for inference on sample texts. This exercise demonstrated the full NLP pipeline — from data loading and tokenization to fine-tuning, evaluation, and deployment — while highlighting how performance improves with thoughtful training and evaluation strategies, even under limited hardware.