Assignment 3: Sentiment Analysis with Transformer Models

Overview: Build a sentiment■classification pipeline on customer reviews using a transformer (e.g., DistilBERT). Deploy the model as a simple Streamlit app.

Objectives: • Fine■tune a transformer on a labeled text dataset. • Compare baseline TF■IDF■+■LogReg vs. transformer. • Deploy an interactive web UI for real■time inference.

Dataset: Use the “Amazon Reviews Polarity” subset (1.8■M examples) or any dataset with ≥■100■k labeled reviews.

Tasks: 1. Subsample balanced training set (max 200■k rows) to fit within hardware limits. 2. Implement baseline model; report precision/recall/F1. 3. Fine■tune DistilBERT (or equivalent) with Hugging Face, add early stopping. 4. Target F1 ≥■0.92 on validation. 5. Wrap the trained model in a Streamlit UI that lets users type/paste text and see sentiment + confidence. 6. Containerize the app with Docker.

Deliverables: • Code repo with notebooks and app. • Deployed URL or Docker image. • Short video (≤■2■min) showing the app (optional bonus).