Project Report

Project: YouTubeInsightAI

Author: Tanveer Ahmad

Date: May 04, 2025

1. Project Overview

YouTubeInsightAI is a sentiment analysis pipeline that classifies YouTube comments into positive, neutral, and negative sentiments using a fine-tuned BERT model. The system is designed to support content moderation, credibility analysis, and user feedback aggregation by automating sentiment classification from video comment sections.

2. Project Architecture

The architecture involves the following key stages:

- Scraping YouTube comments using `script.py`
- Cleaning and preprocessing comment data
- Fine-tuning BERT on labeled sentiment data
- Logging and tracking experiments using MLflow and DagsHub
- Deploying the model and loading it during inference
- GitHub Actions for CI/CD automation

3. Model Training

The dataset includes labeled YouTube comments categorized into three sentiment classes: negative (-1), neutral (0), and positive (1).

The BERT base model (`bert-base-uncased`) was fine-tuned using PyTorch, and performance metrics were logged with MLflow.

Training was tracked on DagsHub using MLflow's tracking URI.

4. Model Deployment

After training, the model was registered on MLflow and hosted through DagsHub. The tokenizer (`bert-base-uncased`) is loaded for preprocessing.

`loadmodel.py` loads the model and tokenizer to classify new comments passed from `script.py`.

5. Automation

A GitHub Actions workflow automates the training and evaluation process:

- Triggered on push or manual dispatch
- Runs training pipeline
- Uploads artifacts such as model weights and logs
- Uses `actions/setup-python@v4` and `actions/upload-artifact@v4`

6. Inference

The `script.py` file collects YouTube comments, which are then classified by `loadmodel.py`. The model outputs sentiment predictions which are counted and returned as a summary of sentiment categories.

7. Improvements & Future Work

- Enhance model with more YouTube-specific sentiment nuances
- Add credibility scoring system
- Integrate with a Chrome extension or web UI for live analysis
- Support multi-language sentiment detection