Social Media Trend Predictor

Mantripragada Venkata Karthikeya

Overview of the Project

- ▶ The goal is to predict trending hashtags on social media platforms like Twitter and Reddit.
- ▶ By using Natural Language Processing (NLP) techniques, we extract meaningful patterns from user-generated content.
- ▶ The solution leverages:
  - ▶ **LSTM (Long Short-Term Memory)** networks for sequential pattern recognition.
  - ▶ **Prophet** for time-series forecasting of hashtag trends.
- ▶ The solution is deployed using **FastAPI** for scalable API interactions and **Streamlit** for an easy-to-use UI.

Solution Approach

- **Data Collection**:
  - Scrape real-time trending hashtags from **Twitter** and **Reddit** using dedicated APIs.
  - This data includes the hashtag and the number of occurrences (or trend count) over time.
- **Data Preprocessing**:
  - Clean the scraped data: removing stop words, special characters, and normalizing text (lowercasing, lemmatization).
  - Encode categorical data, like hashtags, using Label Encoding.
  - Scale the target variable (count of occurrences) for better model performance.
- **Modeling**:
  - **LSTM** is used to capture temporal dependencies in hashtag trends over time.
  - **Prophet** forecasts future trends based on historical data, handling seasonality and holidays effectively.
- **Deployment**:
  - **FastAPI** is used to deploy the backend API, making it scalable and fast.
  - **Streamlit** provides a simple UI for users to interact with the system.

FastAPI Backend Implementation

- **Backend Responsibilities**:
  - The FastAPI server loads the trained models for prediction.
  - Exposes an API endpoint to accept hashtag input from users.
  - The API predicts the popularity (count) of a given hashtag.
  - **Example Endpoint**:
    - URL: 'http://127.0.0.1:8000/predict?hashtag=Example'
    - The server processes the hashtag, applies the trained model, and returns the predicted count.

- **Training**:
  - The LSTM model is trained on the preprocessed dataset with features like encoded hashtags and trend counts.
  - The trained model is saved as 'trend_predictor.h5' and can be used for real-time predictions.

- **API Details**:
  - FastAPI's asynchronous capabilities ensure the backend can handle multiple requests concurrently.
  - The model is served in real-time, predicting trends as soon as the user inputs a hashtag.

Streamlit UI for Interaction

- ▶ **User Experience**:
  - ▶ Streamlit provides an interactive UI where users can input any hashtag.
  - ▶ The input hashtag is sent to the FastAPI backend for prediction.
- ▶ **UI Features**:
  - ▶ Simple interface that displays the predicted trend count for a hashtag.
  - ▶ Real-time interaction with FastAPI to visualize trends as soon as the user submits the hashtag.
- ▶ **Advantages**:
  - ▶ **Non-technical Users** can easily interact with the system.
  - ▶ **Real-time Predictions**: Users can see predicted trends for hashtags in real time.
  - ▶ **Visual Feedback**: Trends can be shown graphically, giving users insights into how hashtag popularity evolves.