Offline Chat Reply Recommendation System

Objective

The objective of this project is to build an offline chat-reply recommendation system using transformer-based models. The system predicts User A's next possible reply when User B sends a message, using User A's previous conversation history as context.

Dataset

The dataset consists of two-person conversations provided in the shared Google Sheet. It contains long chat exchanges between User A and User B, which were used to create contextual input-response pairs.

Approach

1. Preprocessing: Cleaned and formatted the conversation data, combining messages into context-response pairs.
2. Tokenization: Used GPT-2 tokenizer to prepare the dataset for training.
3. Model: Fine-tuned GPT-2 (a Transformer-based language model) on the chat data to generate coherent replies.
4. Evaluation: Used BLEU score to measure the quality of generated responses.
5. Deployment: Saved the trained model as a joblib file for offline use.

Model and Tools

Model: GPT-2 (Transformer-based model)
Libraries: Transformers, Torch, Scikit-learn, Pandas, Numpy, NLTK, Joblib
Language: Python 3.10+

Evaluation

The model was evaluated using BLEU score to assess how well the generated responses matched the actual ones. The average BLEU score achieved demonstrates that the model can generate contextually appropriate replies.

Results

The model successfully generates coherent replies based on previous chat context. It works efficiently offline and can be deployed for chat-based recommendation or automation tasks.

Conclusion

This project demonstrates the capability of fine-tuned Transformer models in understanding and generating conversational text. Future improvements can include larger datasets, multi-turn dialogue handling, and integration with emotion detection.