CORPUS WINES CHAT BOT

Overall Approach

The chatbot is designed to answer user queries about Corpus Wine Company using a pre-defined corpus of questions and answers. The main steps include:

1. **Preprocessing:** Tokenizing and removing stopwords from the questions in the corpus.
2. **Vectorization:** Converting the preprocessed questions into TF-IDF vectors.
3. **Similarity Matching:** Calculating cosine similarity between the user's query and the corpus questions to find the most relevant answer.
4. **Session Management:** Maintaining conversation history using unique session IDs.

Frameworks/Libraries/Tools Used

1. **NLTK:** Used for text preprocessing (tokenization and stopword removal).
2. **scikit-learn:** Used for TF-IDF vectorization and cosine similarity calculation.
3. **Streamlit:** Used to create the web interface for the chatbot.
4. **uuid:** Used to generate unique session IDs for maintaining conversation history.

Problems Faced and Solutions

1. **Text Preprocessing:** Initial challenges with correctly tokenizing and removing stopwords from the text were resolved by fine-tuning the preprocessing function.
2. **Similarity Threshold:** Determining an appropriate threshold for cosine similarity to filter out irrelevant answers was critical. Experimentation with different threshold values helped in finding an optimal value.
3. **Session Management:** Ensuring the chatbot maintained accurate session history required implementing a local dictionary to store sessions and retrieving them effectively.

Future Scope

1. **Expanding the Corpus:** Adding more questions and answers to cover a broader range of queries.
2. **Improved NLP Techniques:** Incorporating advanced NLP models like BERT for better understanding and matching user queries.
3. **User Feedback Loop:** Allowing users to provide feedback on the answers to continuously improve the chatbot's performance.
4. **Integration with Backend Systems:** Connecting the chatbot to backend inventory systems to provide real-time information about wine availability and promotions.
5. **Multi-language Support:** Adding support for multiple languages to cater to a diverse customer base.