GOVERNMENT COLLEGE OF TECHNOLOGY COIMBATORE

ARTIFICIAL INTELLIGENCE 18SPC702

Name : Poornasri P , Keerthana R

Reg no : 71772117131,71772117306

Branch : Computer Science and Engineering

Course : Artificial Intelligence

Course code : 18SPC702

Assignment : 2

Project : Excel based chatbot

Project Description:

The Excel Question-Answer System is a cutting-edge AI-driven tool that simplifies querying Excel documents. This application empowers users to ask natural language questions about data within Excel files, making it especially valuable for analyzing financial, historical, or time-series datasets. By seamlessly merging data extraction with an advanced AI model, the system provides accurate, contextually relevant answers without the need for manual searches or complicated queries.

Methodology

1. **Data Handling and Preparation:**

   o **Uploading Files:** Users can easily upload a CSV file using Streamlit's file_uploader component. Once uploaded, the file is stored locally and read with the pandas library for further processing.

   o **Transforming Data:** The CSV is loaded into a pandas DataFrame, allowing for effective data manipulation. The DataFrame is then converted to a string format with to_string() to provide the AI model with a comprehensive view of the dataset, covering all rows and columns.

2. **User Interaction and Query Submission:**

   o **Entering Questions:** Using Streamlit's text_input, users can submit questions in plain language. The application can handle diverse types of queries, from targeted questions (e.g., "What was the stock price in 1928?") to more general inquiries (e.g., "What is the average annual sales growth?").

   o **Prompt Creation:** The user's query is combined with the dataset context to form a complete prompt. This ensures the AI model has a clear understanding of both the data and the question, facilitating accurate answers.

3. **Integration with AI Model:**

   o **Model Setup:** The system establishes a session with Google's gemini-1.5-pro model through the GenerativeModel API, allowing the AI to engage with the data context in a conversational format.

   o **Submitting Queries and Handling Responses:** The full prompt, including data context and user query, is sent to the AI model using the send_message function. The model interprets the input and generates a response based on the dataset.

- o **Error Management:** If the model request encounters an error (e.g., network or API issues), the system displays an error message and includes a retry mechanism to handle temporary issues, ensuring a stable user experience.

4. **Presenting Answers:**

- o **Processing AI Output:** The model's response is parsed, extracting the relevant answer from candidates[0].content.parts[0].text. This response is formatted in natural language, providing a clear and direct answer to the user's question.

- o **Displaying Results:** The answer is shown on the Streamlit interface under an "Answer" section, allowing users to receive immediate feedback and enabling an interactive querying experience.

- o **Session Continuity:** Users can ask multiple questions in a single session, with answers displayed in real time. A "Quit" button is available for users to exit the interface when finished.

AI Model Information

The project leverages Google's gemini-1.5-pro model, recognized for its strengths in natural language comprehension and generation, particularly in scenarios involving structured data queries.

Key Features and Capabilities

1. **Advanced Language Understanding:**

- o **Comprehensive Context Awareness:** The gemini-1.5-pro model is trained on a vast dataset, enabling it to interpret complex questions and provide responses that are accurate and contextually grounded. This makes it effective for answering nuanced queries and summarizing data trends or comparisons.

- o **Flexibility with Data Formats:** By transforming data into a readable text string, the model can analyze structured information without requiring a fixed schema, making it adaptable to diverse datasets.

2. **Contextual Answer Generation:**

- o **Insightful Data Analysis:** Utilizing advanced language modeling techniques, the model can generate targeted, context-specific answers. For example, it can reference particular years or identify trends in

chronological data like stock prices, providing valuable insights into financial and historical records.

- o **Ideal for Numerical Data:** The gemini-1.5-pro model excels at interpreting quantitative data, making it particularly effective for financial analysis and time-series datasets.

3. **Conversational Capabilities:**

- o **Interactive Data Exploration:** The chat-based format allows the model to handle a series of related questions in the same session, making it easy for users to seek clarification or delve deeper based on previous answers.

- o **Dynamic Response Handling:** Unlike static models, this system's interactive nature adjusts responses based on the evolving context, delivering a conversational and user-friendly analysis experience.

4. **User-Friendly Interface with Streamlit:**

- o **Simplified Interaction:** Although Streamlit is not a model, its web interface plays a crucial role in presenting the AI model's capabilities. It provides a straightforward and engaging platform for real-time querying and feedback, making complex data interactions accessible to users of all skill levels.

Output:

Excel Question-Answer System

Upload an Excel file and ask questions about the data.

Upload an Excel file

	Drag and drop file here Limit 200MB per file • XLSX	Browse files
	Untitled spreadsheet.xlsx 14.1KB	✕

Excel Data Preview

	Register number	Name	NM Course
0	71772117101	ABHILASH G	Cloud Application Development
1	71772117102	ABINAYA V	Cloud Application Development
2	71772117103	ANUMITHA R D	Artificial Intelligence
3	71772117104	ARIVUMATHI P K	Data analytics with cognos
4	71772117105	BALAJI R A	Applied Data Science
5	71772117106	BARATHKUMAR.K	Cloud Application Development
6	71772117108	CATHLYN JEBA GOLDY. T	Cloud Application Development

Enter your question:

list the no of students opted for artificial intelligence course also give their names

Ask

Answer

There are 6 students opted for Artificial Intelligence. Their names are:

- ANUMITHA R D
- MARIESWARAN P
- POORNASRI.P
- PRIYANKA S
- RAJESWARI P
- RUTHRA.L

Challenges Faced:

1. Handling Large Data Contexts

- **Challenge**: Since generative AI models typically have input size limits, sending large Excel datasets as prompts can be challenging. Including an entire dataset in the context can exceed the model's input capacity, leading to incomplete data analysis or truncated responses.
- **Solution**: To address this, the data context was carefully formatted to provide relevant portions or summaries of the dataset instead of the entire content. Selecting specific rows or columns based on the user's question could be a future improvement to further optimize data input.

2. Ensuring Accurate Responses to Varied Queries

- **Challenge**: Natural language questions can vary widely in phrasing, specificity, and intent. The generative model may struggle to accurately answer questions without a clear and well-defined context.
- **Solution**: To improve response relevance, the approach involved combining the dataset context and question in a structured prompt format, helping the model focus on the specific aspects of the data requested by the user. Testing the prompt format and tweaking question framing further optimized accuracy.

3. Error Handling and API Reliability

- **Challenge**: Network errors, API rate limits, or timeout issues could disrupt model responses, resulting in incomplete answers or failed requests.
- **Solution**: A retry mechanism was incorporated with error handling to capture any exceptions, allowing the system to wait and reattempt the request when necessary. This provided resilience and improved reliability, even under unstable network conditions.

4. Adapting to Natural Language Complexity

- **Challenge**: Generative AI models can sometimes misinterpret nuanced or complex questions, especially when the question requires multi-step reasoning (e.g., analyzing growth trends over multiple years).
- **Solution**: Streamlit's input mechanism allows users to clarify or rephrase questions, while testing different prompt formulations improved the model's interpretive ability. Additionally, users were encouraged to ask clear, direct questions for optimal results.

5. Integration of Streamlit and API-based Responses

- **Challenge**: Coordinating data flow between Streamlit, Google's API, and pandas for real-time processing and display required careful handling to ensure smooth interaction and data visualization.
- **Solution**: Streamlit components were implemented for interactive querying and response display, with conditional checks and interface feedback (like error messages) to enhance user experience. This made it easier to handle dynamic inputs and provide clear feedback to users.