Medical Assistant Bot Assignment

Problem Statement:
Develop a medical question-answering system utilizing the provided dataset containing medical information. The goal is to create a model that can effectively answer user queries related to medical diseases. You are free to augment the provided dataset with other medical datasets if you feel it is necessary.

Tasks:

Data Preprocessing:

Ensure the data is structured appropriately for model training, validation, and testing.

Model Training:

Develop a machine learning model using natural language processing (NLP) techniques to understand and respond to user queries. Choose an appropriate architecture for the model, considering factors such as the complexity of the dataset and the computational resources available.

Model Evaluation:

Evaluate the model's performance using appropriate metrics. State your reasoning for using the metric.

Example Interaction:

Include at least three example interactions where the user asks a question (something similar to the provided dataset), and the model provides an answer.

Documentation:

Provide a README file with a concise description of the approach used to tackle the problem. Include:

a. details about any assumptions made during model development and training.
b. model's performance, highlighting its strengths and weaknesses.
c. potential improvements or extensions to the solution, such as refining the model architecture, etc.

Submission Instructions:

Ensure the code is well-documented and organized, following best practices for readability and maintainability. Include comments to explain the purpose and functionality of each code segment. Upload the code to your GitHub account and send an email (before the below-mentioned deadline) to "datascience@supportiv.com" stating the completion of the assignment along with the GitHub link for the code base.

Grading Rubric:

- Approach to the Problem: Evaluation of the method used, including any assumptions made during model development.
- Code Documentation: Assessment of the clarity and completeness of the code documentation, including comments and explanations.
- Model Performance: Analysis of the model's performance using appropriate metrics and examples of predictions.
- Potential Improvements: Evaluation of potential enhancements or extensions to the solution and their feasibility.

We look forward to reviewing your solutions and understanding your approach to this task. Good luck!

Dataset link:
https://drive.google.com/file/d/1vXyLOFRc98f097x4CrK9gOOb3JsKvPmN/view?usp=sharing

Deadline: 2 days after receiving the challenge

Queries/clarifications: please reach out to zara@supportiv.com

Important Note: You must solve this problem independently without the assistance of third-party AI systems (such as OpenAI, Claude, or similar tools) in any part of your solution. However, consulting code documentation is permitted. In your submission, include a statement confirming that you did not use AI assistance.