DataSage: AI-Powered Data Science Tutor

Gokul G , B.Tech, IIIT Dharwad

22BDS026

Abstract

In the era of data-driven decision-making, learning Data Science effectively is essential. However, learners often struggle with scattered resources, lack of guidance, and limited real-time feedback. DataSage is an AI-powered virtual tutor developed using Gemini 2.0 Flash and LangChain, aimed at offering contextual, personalized learning experiences for aspiring Data Scientists. This report presents the design, implementation, and potential impact of the tool, developed using Streamlit and integrated with APIs for enhanced visuals and UI customization.

Keywords—Data Science Education, Conversational AI, Gemini, LangChain, Streamlit, Virtual Tutor, Personalized Learning

I. INTRODUCTION

The rise of online education has amplified the need for smart tutoring systems. While MOOCs and courses provide content, they lack personal interaction and adaptability. DataSage bridges this gap by enabling interactive learning sessions with an AI tutor that adapts to the user's expertise and responds only to Data Science-specific queries. The app aims to create an engaging environment with a beautiful UI, practical examples, and responsive Q&A mechanisms.

II. SYSTEM OVERVIEW

A. Objectives

- To provide a personalized AI tutor for Data Science learners.
- To support contextual conversations limited to Data Science topics.
- To offer a user-friendly interface with visual elements and memory

B. Scope

- Supports users ranging from beginners to advanced learners.
- Covers domains such as Python, Statistics, Machine Learning, and Data Visualization.

III. SYSTEM ARCHITECTURE

The system is built using a modular architecture consisting of:
- Frontend (Streamlit): For interactive UI with sidebar filters, chat interface, and 3D image integration.
- Backend (LangChain + Gemini 2.0 Flash): Handles prompt templates and LLM-based responses.
- Image API (Unsplash): Generates 3D AI assistant images dynamically.
- User Profiling: Users select their experience level and skill sliders for customized responses.

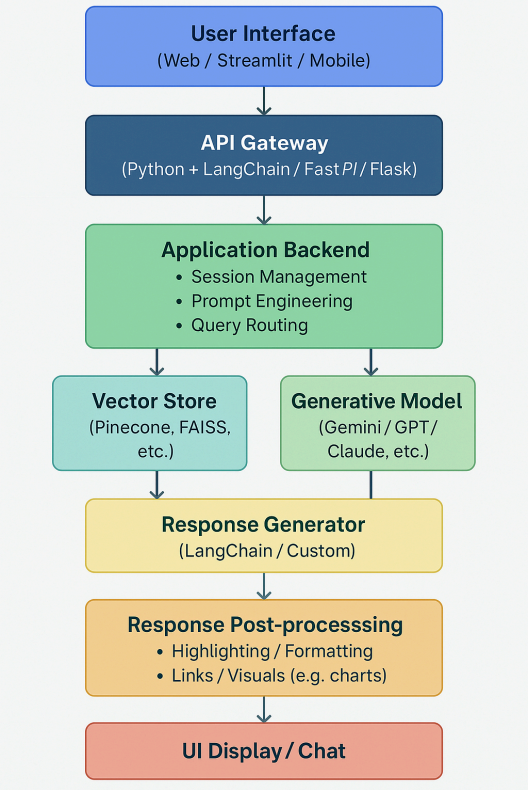

Figure 1: Architecture Diagram

IV. KEY FEATURES

• User Expertise Profiling (Beginner, Intermediate, Advanced)
• Theme Customization (Glassmorphism, Gradients)
• 3D Image Assistant (via Unsplash API)
• Chat Memory (LangChain chains with conversation awareness)
• Domain-Restricted Answering (Only Data Science)

V. TECHNOLOGIES USED

Component	**Technology**
UI	Streamlit
LLM	Gemini 2.0 Flash
Chain Management	LangChain
Image Integration	Unsplash API
Styling	**CSS (Glassmorphism)**

VI. RESULTS AND SCREENSHOTS

The app has been successfully deployed with the following features:
- Real-time Data Science Q&A
- Personalized suggestions
- Visually engaging layout

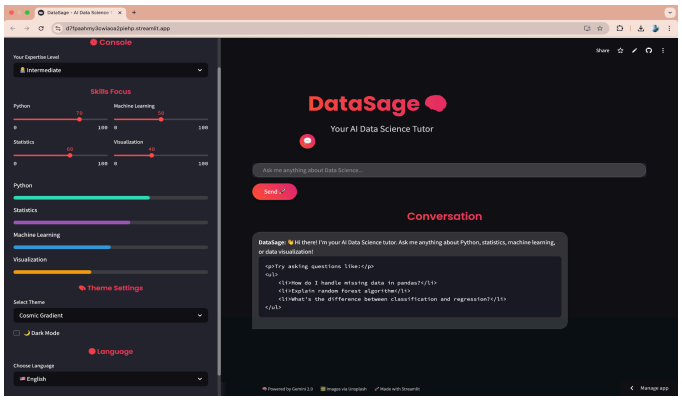

Figure 2: Screenshot of Home UI

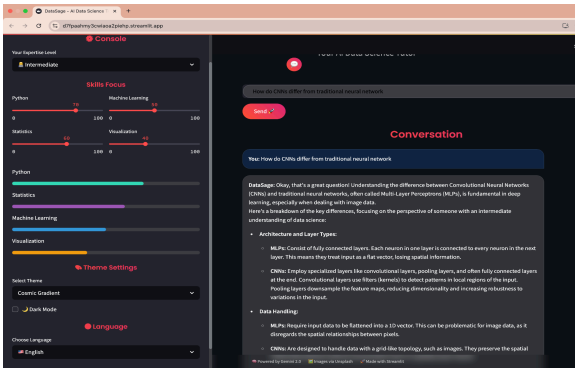

Figure 3: Screenshot of Chat Interface with AI responses

VII. FUTURE WORK

- Persistent Memory using Vector Database
- Multilingual Support
- Learning Analytics Dashboard
- Tutor Mode with custom lesson generation and quizzes

VIII. CONCLUSION

DataSage demonstrates the potential of conversational AI in personalized education. By combining the power of LLMs with real-time interaction and domain restriction, it enables users to learn Data Science more efficiently and enjoyably. This project serves as a foundation for future educational AI tools with increased interactivity and deeper learning customization.

REFERENCES

[1] LangChain Documentation – https://docs.langchain.com
[2] Gemini API – https://ai.google.dev
[3] Streamlit – https://streamlit.io
[4] Unsplash API – https://unsplash.com/developers
[5] IEEE Citation Guide – https://ieeeauthorcenter.ieee.org