

PROJECT REPORT

CSV-Based Retrieval-Augmented Generation (RAG) Chatbot System

Internship Organization: Ospyn

Technology Private Limited

Intern Name: Muhammad Rasin M

Internship Domain: Artificial Intelligence,

Data Science & Analytics

1. Introduction

This project focuses on building a CSV-based chatbot using modern AI tools. The chatbot allows users to upload a CSV file and ask questions related only to the uploaded data. The system retrieves relevant information from the CSV and generates accurate responses without hallucination.

2. Objective

The main objective of this project is to design a beginner-friendly RAG (Retrieval Augmented Generation) chatbot that can answer user queries strictly based on a CSV dataset.

3. Tools and Technologies Used

- Python
- Streamlit (for web interface)
- LangChain (for document loading and text splitting)
- FAISS (for vector storage and similarity search)
- Ollama (for embeddings and LLM response generation)
- CSV dataset (indian_food.csv)

4. System Architecture

The system works in four main steps:

1. Upload CSV file
2. Convert CSV content into text chunks
3. Generate embeddings and store them in FAISS
4. Retrieve relevant content and generate AI responses

5. Working of the Chatbot

Once a CSV file is uploaded, the chatbot processes the file and creates a vector database. When a user asks a question, the system searches for relevant rows in the CSV and provides the context to the language model, which then generates an accurate response.

6. File Type Limitation

The current implementation supports only CSV files. This design choice was made to keep the project simple and beginner-friendly.

7. Output and Results

The chatbot successfully answered questions related to the Indian food dataset, such as identifying vegetarian dishes. The responses were accurate and strictly based on the uploaded CSV file.

8. Advantages

- Prevents AI hallucination
- Simple and user-friendly interface
- Fast retrieval using FAISS
- Local LLM usage using Ollama

9. Limitations

- Supports only CSV files
- Answers depend on retrieved chunks
- Requires Ollama to be running locally

10. Conclusion

This project demonstrates a practical implementation of a CSV-based RAG chatbot. It is suitable for beginners and serves as a strong foundation for building more advanced document-based AI applications.