Enhancing Gift Recommendations with RAG

By,

Amulya Murahari - 002738903

Namitha J C - 0027195461

Sinchana Kumara - 002780971

Prompt Engineering and AI

16/07/2024

Introduction

The objective of a gift recommendation system is to streamline the gift-giving process by suggesting personalized options that align with user preferences and historical data, enhancing satisfaction and efficiency

Vision

To build an accurate platform for personalized gift recommendations, enhancing the joy of giving by ensuring every gift is meaningful and cherished.

Goal

To achieve a 90% satisfaction rate among users by providing accurate and personalized gift recommendations.

Strategy

Leverage advanced AI algorithms to analyze user preferences and behavior, continuously refining our recommendations to ensure highly personalized and accurate gift suggestions.

Importance and Relevance to the Course and Industry

1. This project integrates advanced AI and data analytics, aligning with course objectives of practical application of machine learning and software development

2. It addresses industry needs for personalized user experiences and enhances e-commerce by simplifying the gift-giving process, making it highly relevant and impactful.

Project Description:

The Gift Recommendation System uses machine learning to suggest personalized gifts based on user input, analyzing historical data and preferences. It features a user-friendly interface for specifying criteria like budget and interests.

Specific Problem Aimed to Solve:

The system simplifies the stressful process of finding suitable gifts by providing tailored suggestions that match the recipient's tastes and preferences. This reduces uncertainty and decision fatigue, enhancing the gift-giving experience.

Scope of the Project:

1

User Interface Development

Design an intuitive, responsive application for users to specify gift criteria.

2

Data Collection and Analysis

Gather and update data on user preferences and purchasing patterns.

3

Algorithm Implementation

Develop machine learning algorithms to predict and recommend gifts

4

Testing and Iteration

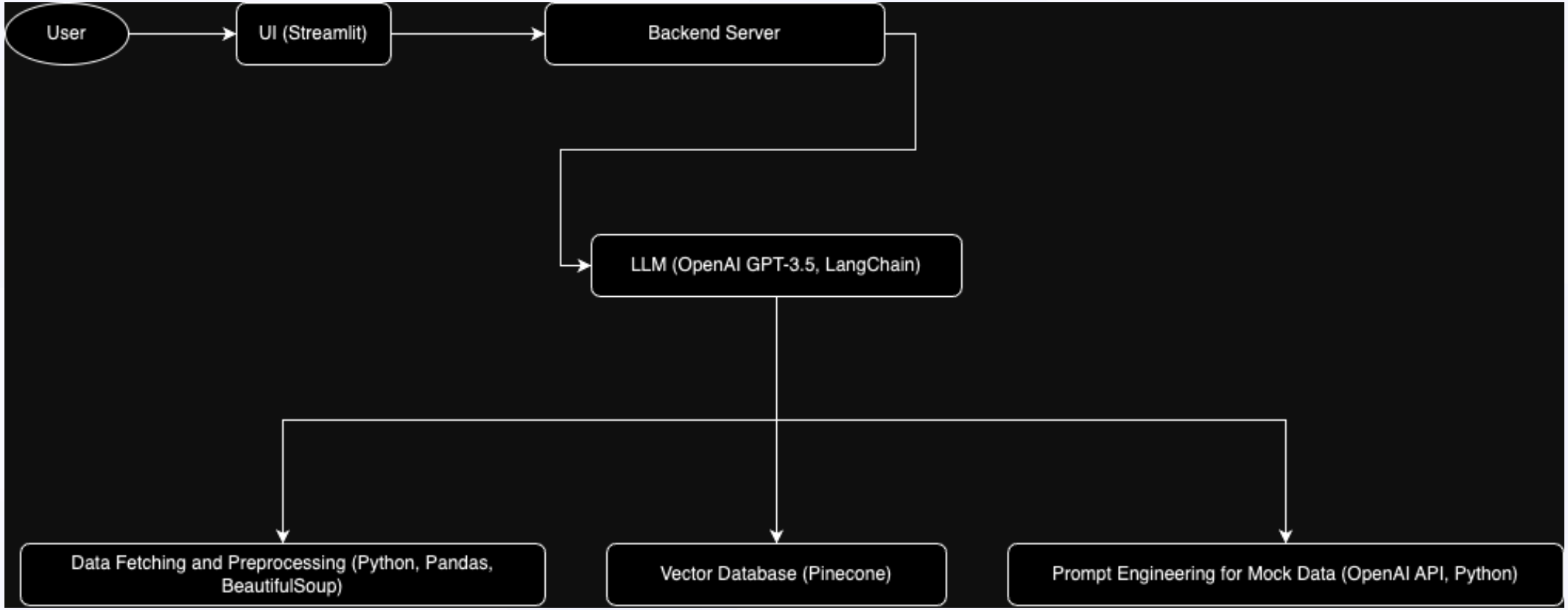
Conduct rigorous testing and gather user feedback

5

Integration and Deployment

Integrate with e-commerce platforms or operate as a standalone app

Project Architecture



Project architecture

Technologies and Tools Used:

- **Streamlit**: Interactive UI
- **OpenAI GPT-3.5 & LangChain**: Natural language processing
- **Pinecone/Weaviate**: Vector databases
- **Python, Pandas, BeautifulSoup**: Data fetching and preprocessing
- **OpenAI API**: Prompt engineering

Project Architecture

Component	Description
User Interface	A responsive web application built with Streamlit that allows users to specify their gift-giving criteria, such as budget, recipient's interests, and occasion.
Data Collection and Preprocessing	A system that gathers data on user preferences, gift items, and purchasing patterns, and preprocesses the data for input into the recommendation model.
Recommendation Algorithm	Machine learning models, such as those based on OpenAI GPT-3.5 and LangChain, that analyze the user input and past data to generate personalized gift recommendations.
Data Storage	A vector database, such as Pinecone or Weaviate, to efficiently store and query the data used by the recommendation algorithm.
Integration and Deployment	The system will be designed to be easily integrated into existing e-commerce platforms or to operate as a standalone application, with a focus on scalability and reliability.

Data Collection and Preprocessing

- **Source and Nature:** Collected data from Kaggle.com and includes categorical data , numerical data and textual data.

Steps Taken for Data Collection:

- **Kaggle Dataset Utilization:** A pre-compiled dataset from Kaggle containing comprehensive gift recommendation data.

Data Preprocessing Techniques Used:

- **Cleaning:** Remove duplicates, handle missing values, correct inconsistencies in the dataset.
- **Categorization:** Classify gifts into categories based on type, occasion, recipient demographics.
- **Text Preprocessing:** Utilize NLP techniques to process textual data.

RAG Pipeline Implementation

Overview of the RAG Pipeline: Combine the power of dense retrieval with a sequence-to-sequence model for generating recommendations

Steps

1

2

Retrieval Setup:

- **Indexing:** Create a searchable index of gift data using embeddings
- **Query Processing:** Transform user queries into embeddings using the same model.

Generation:

- **Context Integration:** Fuse retrieved documents into the decoder of a generative model.
- **Output Generation:** Generate personalized gift recommendations

Challenges and Solutions :

1

- **Scalability:** Handling large volumes of data and queries.
  - **Solution:** Use efficient indexing and query techniques such as FAISS.

2

Accuracy: Ensuring the relevance of retrieved documents.

- **Solution:** Continuously train and refine the retrieval model with user feedback.

Performance Metrics

1

Key Metrics to Evaluate the Project

- **Precision and Recall:** Measure the relevance of the recommendations.
- **Latency:** Time taken to generate recommendations.

2

Methods Used to Calculate These Metrics

- **Precision/Recall Calculation:** True positives, false positives, false negatives from test cases.
- **User Surveys:** Surveys to capture direct user feedback on recommendation relevance.
- **Latency Measurements:** Monitor backend logs to assess response times.

3

Initial Results and Observations

- **Effectiveness:** Early tests show promising precision and recall in controlled settings.
- **User Feedback:** Positive feedback, highlighting the personalization of the recommendations.
- **Performance Issues:** Identified latency spikes during high query volumes

Methods to Improve Metrics

Identifying Weak Areas:

By analyzing the calculated metrics to we will identify weak areas.

Improvements Proposed:

According to the calculated metrics, we will work on modifying the model. For example:

Improving Context Precision:

Use more advanced algorithms like BM25 to improve context retrieval.

Improving Faithfulness:

Fine-tune the generative model with more accurate data.

Noise Injection During Training:

Train the model with noisy data to make it more resilient to real-world noisy inputs.

Deployment Plan

1 Prepare The Application

Ensure Streamlit application code is clean, well-documented, and tested locally.

2 Set Up a GitHub Repository

Push your Streamlit application code to this repository.

3 Deploy to Streamlit Cloud

- Log in to Streamlit Cloud and connect your GitHub repo.
- Specify branch and main Python file.
- Deploy and configure settings.

4 Tools and Platforms for Deployment

- Streamlit Cloud
- GitHub

5 Plan for User Testing and Feedback

- Beta Testing - Share with select users.
- Iteration - Update based on feedback.

Future Work

Improvements for the Project

Enhanced Retrieval Techniques:

Exploring and implementing more advanced retrieval techniques (e.g., hybrid models combining keyword and neural retrieval).

Advanced AI Models:

Incorporating the latest advancements in AI and NLP for better performance.

Long-term Vision

Scalability:

Ensure the application can handle increasing user loads.

User-Centric Improvements:

Focus on enhancing user experience through continuous feedback and iteration.

Conclusion

The Gift Recommendation System leverages machine learning to provide personalized gift suggestions based on user input, including budget, recipient's interests, and occasion. By analyzing historical data, user preferences, and demographics, it generates tailored recommendations. The system's user-friendly interface ensures optimized and relevant gift options for enhanced customer satisfaction.

Key takeaways

1

Introduction to RAG and Gift Recommendation

Leveraging diverse data sources and sophisticated preprocessing ensures a robust foundation for personalized gift recommendations.

2

RAG Pipeline Implementation

The RAG pipeline effectively combines retrieval and generation techniques to deliver personalized and timely gift suggestions.

3

Performance Metrics

Precision, recall, and user satisfaction rate are critical metrics demonstrating the effectiveness and efficiency of the gift recommendation system.

4

Metric Improvement

Specific strategies and their impacts.

5

Deployment

Effective steps and tools for deploying the application.

6

Future Work

Vision for extending and enhancing the project.

Thank You!

We're open to Q&A now.