Enhanced Property Recommendation System Using Retrieval-Augmented Generation for Airbnb Listings

Authors:

Itay Bachar- 206948218

Sewar Hino- 323016485

Adir Toledano- 313535379

Abstract:

This project develops a Retrieval-Augmented Generation (RAG) system to improve property recommendations for Airbnb users by leveraging synthetic question-answer (QA) pairs for evaluation. Given the lack of ground truth data in the Airbnb dataset, we generated synthetic QA pairs based on Airbnb property descriptions. This dataset enhances the reliability of the RAG system in retrieving relevant properties. Our system, demonstrated through a prototype application, enables users to query property features like accessibility, infrastructure, and amenities, providing tailored property recommendations. The results confirm the RAG system's efficacy in returning relevant listings, showing promise for application across similar recommendation domains.

Introduction:

Personalized recommendations are crucial for platforms like Airbnb, where users have unique preferences. Traditional recommendation systems lack the ability to cater effectively to specific user needs, such as accessibility or amenity preferences. This project addresses this gap by implementing a RAG pipeline tailored for property recommendations that considers user-specific requirements. By introducing synthetic QA pairs as a ground truth, our approach evaluates the RAG system's ability to provide relevant, accessible, and infrastructure-specific listings, thus enhancing Airbnb's recommendation capabilities.

Methodology:

<u>Dataset and Preprocessing:</u>

We used an Airbnb dataset hosted on Hugging Face, containing property descriptions, amenities, and numerical attributes such as price and guest capacity. The data preprocessing was led by Adir, focusing on the following key tasks:

- Data Cleaning**: Irrelevant columns like `weekly_price` and `monthly_price` were removed to streamline the dataset.

- Handling Missing Values**: Missing numerical values, such as in `bedrooms` and `beds`, were filled with the mean value. Text fields with missing values were replaced with placeholders to avoid skewing model training.
- Data Type Validation**: Data types were standardized (e.g., ensuring `first_review` and `last_review` were in datetime format).
- Exploratory Data Analysis (EDA)**: Using histograms and KDE plots, Adir conducted EDA to identify outliers and understand feature distributions. This step provided insight into data patterns, such as the distribution of `price` and `number_of_reviews`.

Refer to Appendix for visualization samples.

<u>QA Dataset Generation:</u>

Since ground truth values for QA were absent, Itay generated a synthetic QA dataset using Cohere's large language model (LLM) API. This dataset creation process included:

1. **Sample Selection**: 500 Airbnb listings were sampled to capture a broad spectrum of property features.

2. **Topic-Based Prompting**: The QA pairs covered topics like accessibility, specific infrastructure (e.g., elevator), and basic amenities (e.g., internet, kitchen). Iterative prompts were crafted to ensure topic relevance.

3. **Validation and Refinement**: LLM outputs were parsed and validated for format consistency and topic alignment. Challenges with LLaMA 3 led to a transition to Cohere's LLM, which provided better control over topic specificity and output coherence.

Itay's final QA dataset, stored in a CSV, serves as a ground truth reference, essential for evaluating the RAG model's response quality.

<u>RAG System Development:</u>

The RAG system consists of embedding-based retrieval and a text generation model:

1. **Embedding Generation**: Property descriptions were embedded using the `all-MiniLM-L6-v2` model from the SentenceTransformers library. The embeddings were indexed with FAISS for fast similarity-based retrieval.

2. **Text Generation**: A language model (`distilgpt2`) was used to generate responses, summarizing retrieved properties in natural language.

3. **QA-Enhanced Response Generation**: Siwar's model utilized Itay's QA pairs in the response generation process, integrating relevant QA pairs to improve the response context.

4. **Evaluation**: The system was evaluated using precision, recall, and F1-score. Synthetic queries were matched against the ground truth QA pairs to measure response relevance.

Experiments:

Experimental Setup:

Our experiments assessed retrieval accuracy and response generation using the following approach:

1. **Synthetic Queries**: Queries focused on specific needs (e.g., "wheelchair-accessible property with an elevator") to test the system's ability to identify relevant properties.

2. **Evaluation Metrics**: Precision, recall, and F1-scores measured the accuracy of the system's retrieval. Evaluation also involved qualitative assessments of response relevance.

Results:

The RAG system demonstrated high retrieval accuracy, achieving a precision of 1.00, recall of 0.67, and F1-score of 0.80, indicating effective retrieval of relevant properties. Additionally, responses were contextually aligned with user needs, especially for accessibility-related queries.

Discussion:

This project successfully addressed the challenges of personalized recommendations by creating a synthetic QA dataset and implementing a RAG system. The Cohere-based QA generation improved ground truth quality, and FAISS-based retrieval proved efficient for large datasets. However, limitations such as API rate limits and cost constraints in the Cohere LLM remain. Future work could explore fine-tuning the retrieval model with domain-specific embeddings and increasing QA dataset coverage.

Appendix:

1. **Visualizations**: EDA plots (histograms, KDE, scatter plots) of features like `price`, `bedrooms`, and `number_of_reviews`.

2. **References**: Key papers, blogs, and resources that informed each stage of the project.

GitHub Repository: [Link to Repository with Full Code and Data]