**Brief Overview**

**Project: Personalized Product Recommendations for Unilever on Bigbasket**

**Background**: Bigbasket is a leading Indian e-commerce platform for groceries and household items, and a potential client for Unilever. Challenge: Bigbasket offers a vast product selection, making it difficult for customers to discover new products or find exactly what they're looking for. Traditional recommendation systems might suggest "frequently bought together" items, which can be limiting.

**Sample data**:

|  |  |  |
| --- | --- | --- |
| **Column**  **Name** | **Description** | **Example** |
| Product ID | Unique identifier for each product | 12345 |
| Product  Name | Name of the product | Knorr Instant Noodles– Masala |
| Description Textual description A delicious and convenient pack of instant noodles with of the product a flavorful masala taste.  Ingredients List of ingredients Wheat flour, palm oil, iodized salt, spices (including coriander, turmeric, chili pepper), dehydrated | | | | |
| vegetables (carrot, peas)  Category Category the Instant Noodles  product belongs to  Image URL URL of the product (link to image) image  **Sample Data:**   * + **Bigbasket Product Data (collected):**     1. Product ID: 1234   ○ Description: "Dove Shampoo, Nourishing Care, 650ml"  ○ Ingredients: "Water, Sodium Laureth Sulfate, Dimethicone, ..." (truncated for brevity)  ○ Category: "Hair Care"  ○ Brand: "Dove"   * + **Customer Interaction Data (potential, if available):**     1. Customer ID: 5678   ○ Past Purchases: [Dove Shampoo (ID: 1234), Conditioner (Brand X), Cooking Oil (Brand Y)]   * + **Langchain Training Data (example):**     1. Question: "What shampoo is good for dry hair?"   ○ Answer: "Dove shampoos are known to be moisturizing and suitable for dry hair."  ○ Question: "What can I cook with vegetable oil?"  ○ Answer: "Vegetable oil is a versatile ingredient for stir-fries, baking, and salad dressings."  **Technical Approach:**  1. **Data Preprocessing:**  ○ **Product Data:** We cleaned and standardized product descriptions in the Bigbasket data. This involved removing punctuation, converting text to lowercase, and potentially stemming or lemmatization.  ○ **Public Product Database:** If available, information from LLAMAIndEx (or similar source) was integrated into the product data. For example, a field for "Hair Type:  Normal to Dry" was added based on the LLAMAIndEx entry for Dove Shampoo.  ○ **Customer Interaction Data (if available):** Customer purchase history was anonymized and translated into a format suitable for the recommendation system. This could have involved creating a binary matrix indicating which products a customer had purchased in the past.  2. **Embedding Generation:**  ○ OpenAI Embeddings were used to convert textual data (product descriptions and ingredients) into numerical vectors. These embeddings captured the semantic meaning of the text, allowing for comparisons between products.  ○ For example, the embeddings for "Dove Shampoo" and "Hair Conditioner" would likely be closer than the embeddings for "Dove Shampoo" and "Cooking Oil" due to the semantic similarity.  3. **Retrieval Augmented Generation (RAG) with Langchain:**  ○ Langchain was trained on a prepared question-answer dataset related to grocery and household products. This dataset included questions about product attributes, uses, and potential pairings between products.  ○ The Langchain model learned to answer these questions based on the information it was trained on.  4. **Personalized Recommendations:**  ○ **Scenario:** A customer (ID: 5678) had purchased Dove Shampoo (ID: 1234) in the past.  ○ **Retrieval:**  ■ The embedding for Dove Shampoo (ID: 1234) was looked up in the vector database (Chroma).  ■ Other products in Chroma with similar embeddings, indicating similar product characteristics, were found. This could include other shampoos or hair care products.  ○ **Langchain for Personalization:**  ■ The retrieved similar products and the customer's basket information (Dove Shampoo) were used as context for Langchain.  ■ Langchain was prompted with questions like "What conditioner would complement Dove shampoo?" or "What hair care routine is recommended for dry hair?" based on the additional information from LLAMAIndEx  ○ **Recommendation Generation:**  ■ Langchain leveraged its knowledge base and the powerful LLM (potentially not GPT-3.5-tubo) to generate personalized recommendations. This might have included suggesting a specific conditioner brand compatible with Dove shampoo or recommending a hair care routine for dry hair that included additional products.  By combining these steps, the system moved beyond recommending "frequently bought together" items. It considered the semantic meaning of products (through embeddings), the customer's past purchases, and additional product information to provide tailored recommendations that addressed the customer's specific needs. | | | | | |