FOOD RECOMMENDER SYSTEM

Executive Summary

By Gifty Singh

<u>Food case:</u>

The vast majority of people are not fully aware of the rich array of recipes and food pairings that are available, thus potentially missing out on exciting culinary experiences. The aim of this project is to develop a recommender system to encourage users to try new recipes based on their unique preferences. We believe this also presents a lucrative business opportunity. Grocers could use this system to create novel food pairings and stimulate interest around emerging food trends, which could significantly boost profits and drive social change.

<u>Background:</u>

While food recommender systems exist, their primary focus is to suggest recipes that align with users' existing preferences. However, this approach only addresses part of the problem, as users may not know about new cuisines or whether they might enjoy them. This project aims to transcend this limitation by using a more robust model that includes real user feedback, thus increasing the novelty of suggested recipes. The aim is to improve the "hit rate" for both familiar and novel recipes, bringing the human element into the recommendations.

Dataset Details:
The dataset used in this project was obtained from food.com and combines both recipes and user interactions, totaling 1 million data points. The "recipe" table contains information such as the name, id, preparation time, ingredients, and nutrition details of each recipe. The "interactions" table contains data on user IDs, recipe IDs, ratings, and reviews.

Recipe Table:

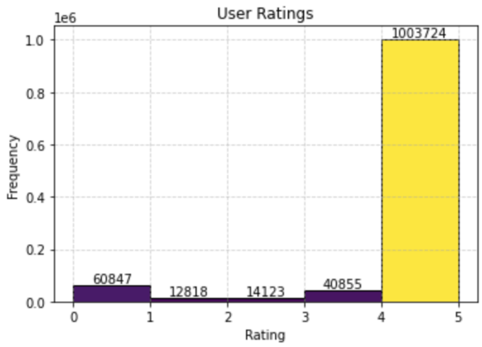
Column Name	Data Type	Description
name	String	The name of the recipe.
id	Integer	A unique identifier assigned to each recipe.
minutes	Integer	The amount of time required to prepare the recipe, in minutes.
contributor_id	Integer	A unique identifier assigned to the contributor of the recipe.
submitted	Date	The date the recipe was submitted.

tags	List of Strings	A list of tags associated with the recipe.
nutrition	List of Floats	Nutritional information about the recipe.
n_steps	Integer	The number of steps required to prepare the recipe.
steps	List of Strings	A list detailing each step required to prepare the recipe.
description	String	A textual description of the recipe.
ingredients	List of Strings	A list of the ingredients required for the recipe.
n_ingredients	Integer	The number of ingredients required for the recipe.

User Interactions Table:

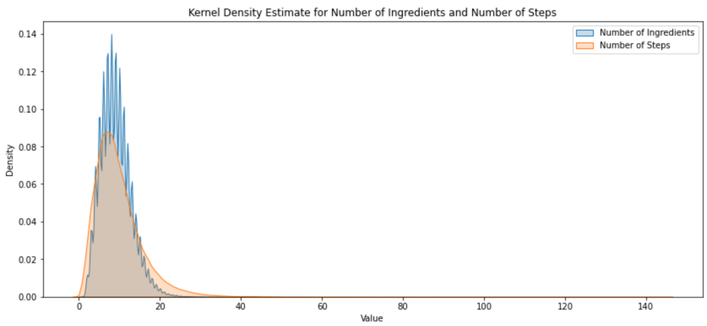
Column Name	Data Type	Description
user_id	Integer	A unique identifier assigned to each user.
recipe_id	Integer	A unique identifier assigned to each recipe.
rating	Integer or Float	The rating given by the user to the recipe.
review	String	The textual review provided by the user for the recipe.

Data Cleaning and Preprocessing:



- Exploring the rating of the users. 5 is the most used value.

During the initial cleaning process, we categorised data points as relevant or non-relevant to the goal of making recommendations. *'Names'* and *'ingredients'* emerged as the most significant features, creating a matrix to navigate through similar items using the term frequency–inverse document frequency (TF-IDF) tokenizer.

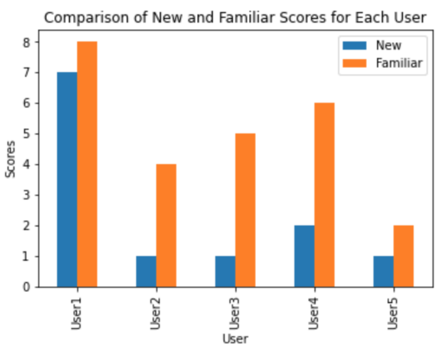


- Distribution of ingredients and steps using the kernel density estimate.

The cosine similarity function was used to find the degree of similarity between different items based on their voting frequency, a more discretely measure of rank we can control diversity.

Modeling and Results:

Our model's results were promising, with the generated recommendations aligning well with the target keywords. We conducted preliminary validation with five users, who demonstrated a hit rate of 60% - indicating that the majority of the recommended dishes were both familiar to and appreciated by the users. However, the users were also intrigued by some of the less familiar suggestions, especially those lower in the ranking.



User	New	Familiar
User1	3.5	4.0
User2	0.5	2.0
User3	0.5	2.5
User4	1.0	3.0
User5	0.5	1.0

- Shows user feedback to model 1 and model 2 combined.

Conclusions:

Our findings indicate that the cosine similarity function is an effective method for making recommendations based on similar items. However, real user feedback is crucial for gauging actual user preferences and refining the recommendation system. While our preliminary testing has shown promising results, we recognize that a more extensive evaluation is needed. In the next phase of the project, we aim to enhance our model by incorporating more user behaviour data and diversifying the dataset, which should allow for the inclusion of more "non-similar" items higher in the rankings, thus adding context and personalization to each user's recommendations.

This project has shown that there is enormous potential for recommender systems that can guide users towards new and exciting culinary experiences. As we continue to refine our approach, we believe our system can bring benefits not only to individual users, but also to businesses looking to capitalise on emerging food trends.