Project Report

1. Lumaa Recommendation Engine

Motivation behind my recommendation.

I was searching for a desert in youtube and I want to avoid sugars and diary in the desert. I was checking for the recipe and wanted to try building a recommendation engine.

2. Dataset

https://www.kaggle.com/datasets/thedevastator/better-recipes-for-a-better-life?select=recipes.csv

Features:

- recipe_name: The name of the recipe. (String)
- recipe_name: The name of the recipe. (String)
- prep_time: The amount of time required to prepare the recipe. (Integer)
- cook_time: The amount of time required to cook the recipe. (Integer)
- total_time: The total amount of time required to prepare and cook the recipe. (Integer)
- servings: The number of servings the recipe yields. (Integer)
- servings: The number of servings the recipe yields. (Integer)
- ingredients: A list of ingredients required to make the recipe. (List)
- directions: A list of directions for preparing and cooking the recipe. (List)
- rating: The recipe rating. (Float)
- url: The recipe URL. (String)
- cuisine_path: The recipe cuisine path. (String)
- nutrition: The recipe nutrition information. (Dictionary)
- timing: The recipe timing information. (Dictionary)

3. Preparation:

1. Removed the duplicates
2. Removed null values
3. Removed columns 'img_src', 'cuisine_path', 'prep_time', 'cook_time'
4. Feature engineered time and converted to minutes

5. Missing values of time are replaced with ‘missing’
6. Combined ingredients and directions together, removed duplicate words

Sample data of combined_ingredients_directions
Ingredients:
3 tablespoons butter, 2 pounds Granny Smith apples (or other firm, crisp apples), peeled, quartered, cored and sliced 1/4-inch thick, 1 pound Macintosh apples (or other soft-textured apples that fall apart when cooked), peeled, quartered, cored, and sliced 1/4-inch thick, ½ cup sugar, ½ cup dried cranberries, 1 sheet frozen puff pastry, thawed but still cold (follow package directions), 1 egg white, lightly beaten, 1 tablespoon sugar, 1 ……..

Directions:
Heat butter in a large skillet over medium-high heat. Add apples, 1/2 cup sugar and cranberries; cover and cook until apples release their liquid, about 5 minutes. Remove lid and continue to cook, stirring frequently, ……

Combined Ingredients and Directions:
unfold form smith keep allows temperature cranberries; onto at juices quartered, position butter, adjust 10 still minutes. airtight macintosh cookie brown, 2 corners sprinkle warm this thick, with consistency, serve ½ cream work parchment peeled, lightly (follow until (i lining tablespoons inches. be apples, package cooked), liquid, granny temperature. directions), ice apples fall ensures optional dough 400 container or to and pan floured pastry,.....

7. Removed the duplicates in ingredients
8. Used NLP techniques to preprocess “Ingredients”
9. Process the text with SpaCy, NLTK and applied following techniques
- Lowercasing
- Removing Punctuation
- Pronoun removal
- Removing Stop Words
- Tokenization and rejoined them to a single string

10. I applied the above functions to the 'Ingredients' column and named it 'processed_ingredients'. Similarly, I combined the 'Ingredients' and

'Directions' columns and named the resulting column 'processed_ingredients_directions'.

How are we going to recommend?

If the user wants to give query with ingredients we can recommend using “processed ingredients column”
If the user wants to give query with ingredients and directions we can recommend using 'processed_ingredients_directions' column
If the user wants to give query with ingredients and quantity we can recommend using ‘ingredients’ column

4. Overview of Recipe Recommendation System

This document explains the implementation of a recipe recommendation system built with Streamlit. The system leverages various Python libraries to filter and recommend recipes based on user preferences, such as desired ingredients, dietary restrictions, and other cooking parameters. Here’s a breakdown of how each component contributes to the application.

Filtering Recipes
The filter_recipes function applies multiple user-specified filters to the dataset:

Ingredient Exclusions: It removes recipes containing any undesired ingredients (donts) listed by the user. This includes specific items and broader categories like meat or dairy, depending on user preferences.
Rating Filter: It filters out recipes with ratings below a specified threshold.
Cooking Time Filter: It excludes recipes exceeding the maximum cooking time set by the user.
Servings Filter: It ensures that recipes meet the required serving size range, from minimum to maximum servings.

Text Processing with TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) is used to convert the text data of recipes into a numerical format that machine learning models can understand. This technique:

- **Term Frequency (TF)**: Measures how frequently a term appears in a document. In this context, it helps highlight the most prominent ingredients and directions in a recipe.
- **Inverse Document Frequency (IDF)**: Downweighs terms that appear frequently across documents, reducing their significance. This helps to diminish the weight of common cooking terms and emphasize unique ingredients.

The tfidf_vectorizer is initialized and fitted to the processed_ingredients of the filtered recipes, transforming the text into a TF-IDF matrix.

Cosine Similarity for Recommendation

Cosine similarity measures the cosine of the angle between two vectors. In this application, it is used to:

- Compare the TF-IDF vector of the user's query against each recipe's vector in the dataset.
- Identify recipes that are most similar to the user's preferences by calculating the cosine similarity between vectors.

The cosine similarity values are used to rank the recipes, with higher values indicating a closer match to the user's query.

Cut-off Score

The cut-off score is a threshold for determining which recipes are sufficiently similar to the user's query to be recommended. Recipes with cosine similarities below this score are not included in the final recommendations. This score can be adjusted to change the strictness of the match requirement:

- **Increasing the cut-off score**: Leads to fewer, more relevant recommendations.
- **Decreasing the cut-off score**: Results in more recommendations but potentially less relevance.

Streamlit User Interface

The user interface, created with Streamlit, allows users to input their preferences and view recommendations. It includes input fields for ingredients to include, ingredients to avoid, and sliders for rating, cooking time, and servings. Users can specify whether to include meat and dairy in the recipes.

Conclusion

This recipe recommendation system provides an interactive, user-friendly way to filter and find recipes based on detailed personal preferences. By leveraging advanced text processing and similarity measures, it ensures that the recommendations are tailored to the user's specific culinary interests and dietary needs.

Recipe Recommendation System

What ingredients you want to have in your recipe?

honey and banana

Ingredients to avoid (comma-separated)

peanut

Minimum rating

4.00

0.00 5.00

Maximum cooking time (minutes)

90

0 180

Minimum servings

2

1 10

Maximum servings

10

1 30

Is it OK to recommend with meat?

- (•) Yes
- ( ) No

Diary preferance?

- (•) Yes
- ( ) No

Recommend Recipes

	recipe_name	rating	total_time	servings	ingredients
700	Simple Avocado Milkshake	4.7	10	3	1 cup milk, 1 ripe avocado, peele
695	Banana, Avocado, and Spinach Smoothie	4.7	10	2	1 banana, sliced, ½ avocado, pe
73	Simple Summer Smoothie	4.5	10	4	1 banana, 1 cup frozen strawberr
519	Easy Mango Banana Smoothie	4.5	10	8	2 mangos - peeled, seeded, and s
525	Quick Banana Milkshake (Ice Cream Free)	4.6	5	2	3 ice cubes, 1 cup milk, 1 banana