Recipe Recommendation System

Aamna Ahsan

Overview

Cooking is a passion for many people. It is a skill necessary in everyday life. However, in the initial phase of learning, it is difficult to decide what to cook, and how to cook it. A recipe recommendation can help solve this difficulty. It can be tailored towards user preferences, nutrition/dietary restrictions, and favorite kinds of cuisines. This system can also help discover new dishes based on new cuisines, and assist in meal planning for someone in a busy routine or busy household.

Method and Data Processing:

A recipes.csv file was extracted, cleaned up and prepared:

1. Outliers were detected using box plots, these outliers were then removed, and revisualized.
2. NaN values were removed for unnecessary items, and necessary items were replaced with 1.
3. The time entries in the cooking, preparation and total time was switched to time in seconds.
4. The data was checked for duplicate values:
    - duplicated rows in recipes dataset: 0
    - duplicated rows in user interactions dataset: 0
5. Reviews and Ratings were analyzed and visualized as well.

Pre-Processing and Exploring the Data

Categorizing

This process began by categorizing the recipes into Quick, moderate, and long. This was done by manipulating the Cooking Time column. The recipe was categorized as ‘quick’ if the cooking time was less than 1 hour. The ‘moderate’ category was between 60 min-120 min. All recipes longer than 120 minutes were considered “long”. This shows that most of the recipes in this data consisted of ‘long’ cooking time.

Ingredient popularity was also measured, and the top 20 ingredients were identified.

Category: Dietary Restriction

This was created based on: ‘dairy-free’, ‘gluten-free’, ‘low-carb’, ‘vegetarian’ keywords. A new column was created to indicate if these were present with a “0” or “1”.

Category: “MealType”

A third category called, “MealType”, was created. This consisted of vegetarian, seafood, and meat items:

- meat_keywords **=** ['meat', 'chicken', 'beef', 'pork', 'lamb', 'turkey','duck','goat','rabbit','sausage','mutton','veal','liver','brain']
- seafood_keywords **=** ['seafood','fish', 'shrimp', 'crab', 'lobster', 'salmon','cod','crayfish','mussles','scallops','Anchovies','Oyester','squid', 'octopus','prawns']
- All other items were considered ‘vegetarian’.

Category: Level of difficulty

- Easy: Total time (which consisted of prep time and cooking time) = 30 minutes or less
- Medium: Total time = 30 minutes- 2 hours
- Difficult= 2 hours +

Category: Cuisines

This category was intended to specify different cuisines based on keywords:

This included: Italian, Mexican, Chinese, Indian, Japanese, Thai, Asian, Persian, American, Middle East as keywords.

Models

The following models were created based on specific features:

- K-Nearest Neighbor: Similarity based on features.
- Ingredient Based Feature
- And lastly, a model based on Reviews.

Future Improvements

1. A feature of “likes” can be added, to gain user reviews through the recommendation system.
2. A recommendation model can be created based on cooking time, and based on level of difficulty can be created. More “quick” and “moderate” recipes are needed after seeing the distribution in the pre-processing notebook.
3. A recommendation model can be created based on favorite ingredients.
4. More entries into cuisine types can also be created, features like seasonal recommendations can also be explored.