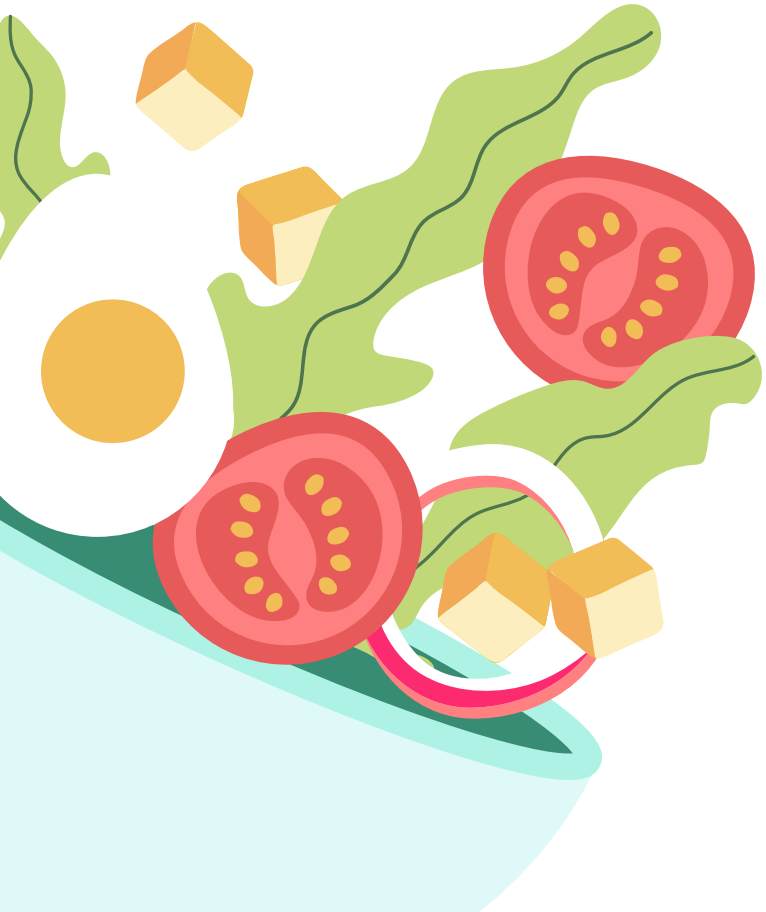


Diet Recommendation System

Team member: Zujun Gu, Yung-chieh Yang, Yijie Huang, Tengkai Gong, Zhiyi Weng

contents

Dataset

Motivation

Data
Exploration
& Visualization

Recommendatio
n
System

Dataset

Where is the dataset?

Food.com

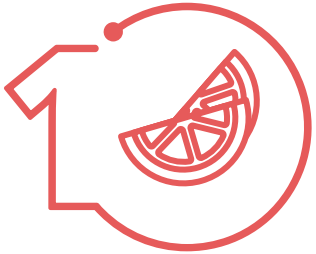
What's inside the dataset?

Basic Information, Timing Details, Instructions, Ingredients, User Reviews, Nutritional Content

How many?

522,517 recipes across
312 distinct categories

Motivation

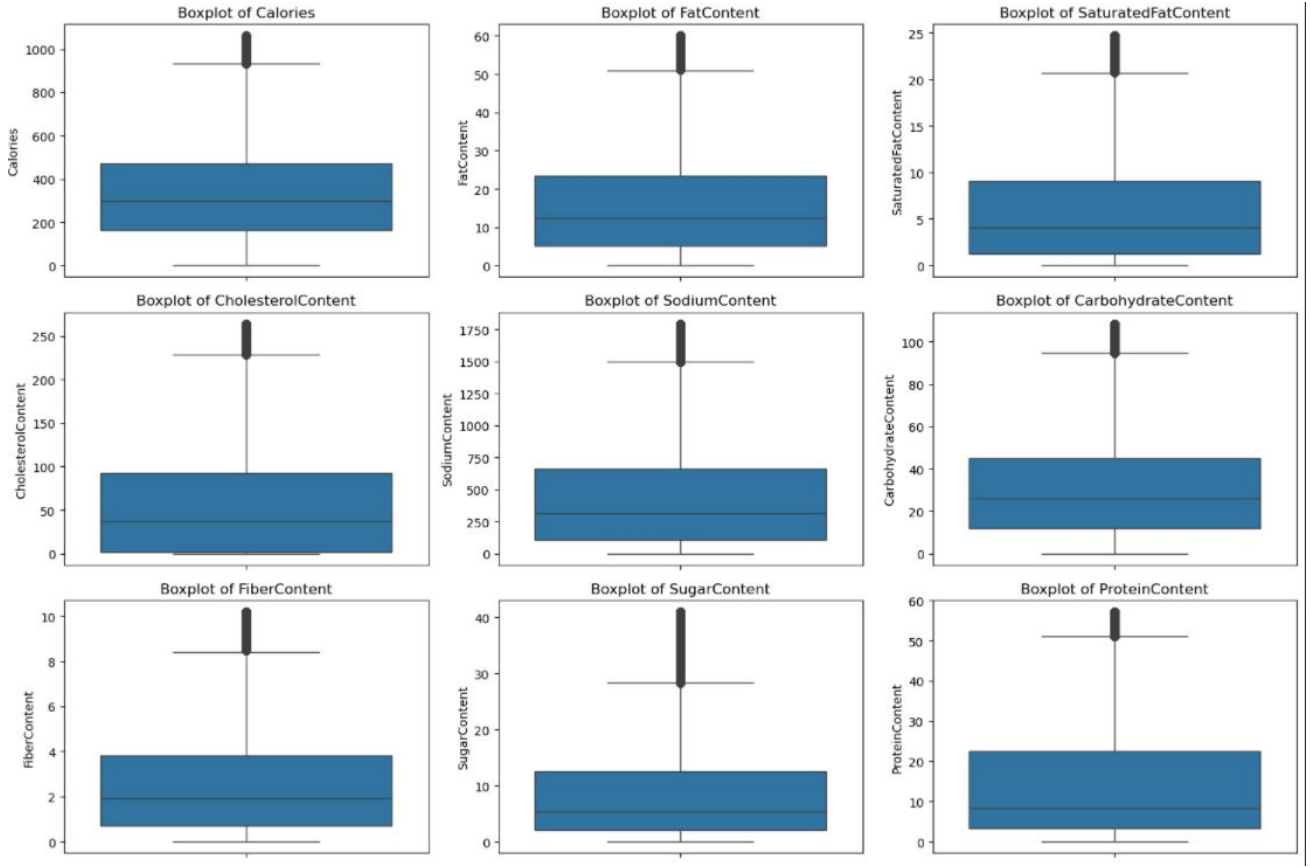


Health Personalization
Match recipes to preferences and nutrition needs, helping users meet health goals with dietary accommodations.

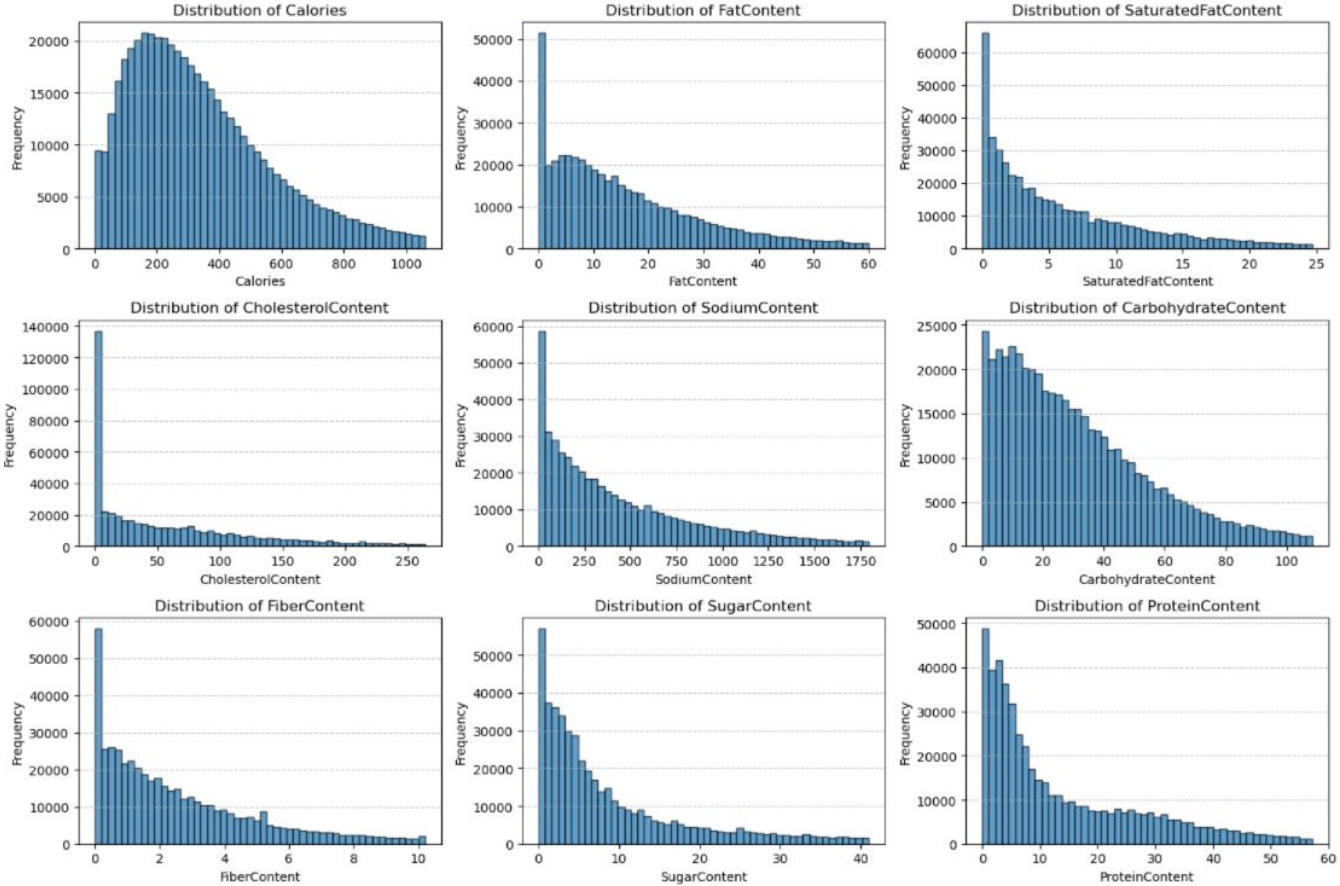
Medical Support
Enable healthcare pros to create custom meal plans for patients with diabetes and heart disease.

Business Value
Allow food businesses to offer personalized menu recommendations based on customer preferences.

Nutritional Component Distribution After Outlier Removal

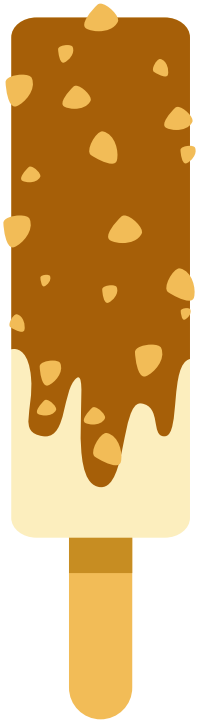


Distribution plot



Data Distribution Analysis

1. Most nutrients display right-skewed patterns, with most foods containing minimal cholesterol, sodium, saturated fat and fiber.
2. Calorie distribution forms a bell curve, mostly concentrated in the 200-400 calorie range.
3. Carbohydrate distribution is gradual, indicating significant variation across foods.
4. Protein distribution shows most foods have low protein content, with few high-protein recipes.



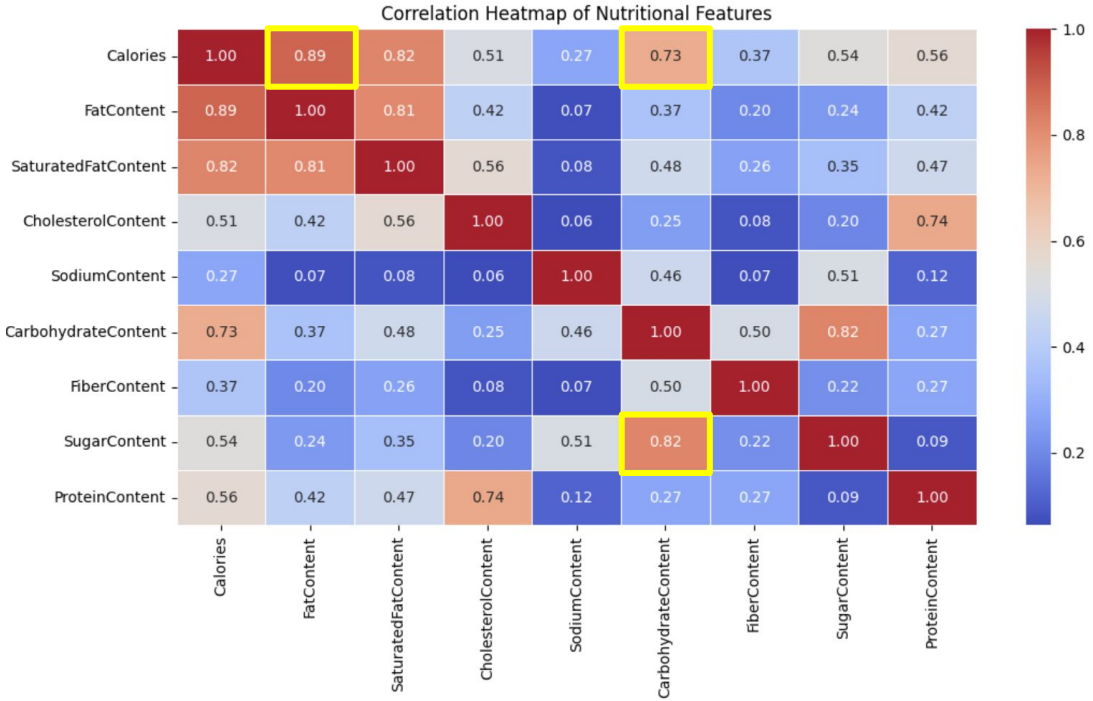
Interesting Insights

- How do different nutritional components interact with each other?
- What are some frequently seen ingredients in the dataset? What are some frequently seen keywords in a 5 star recipe review?
- How does nutrition contents and calories influence ratings?
- How has nutrition content changes over years?

- Above analysis are considered in building the diet recommendation system.

Interesting Insights

- How do different nutritional components interact with each other?

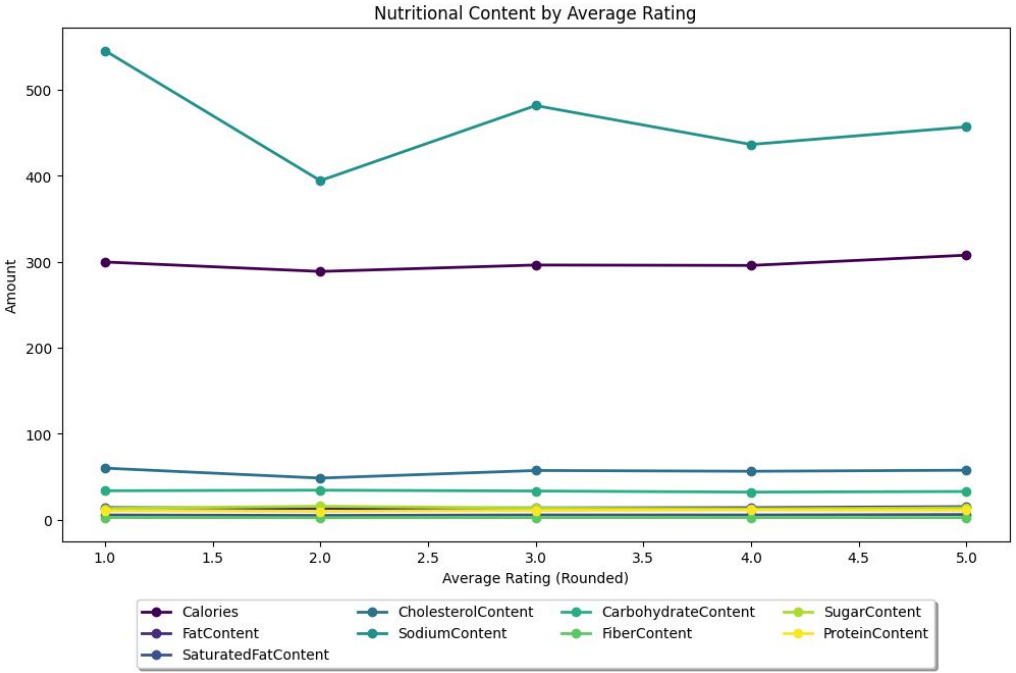


Interesting Insights

- What are some frequently seen ingredients in the dataset?
- What are some frequently seen keywords in a 5 star recipe review?

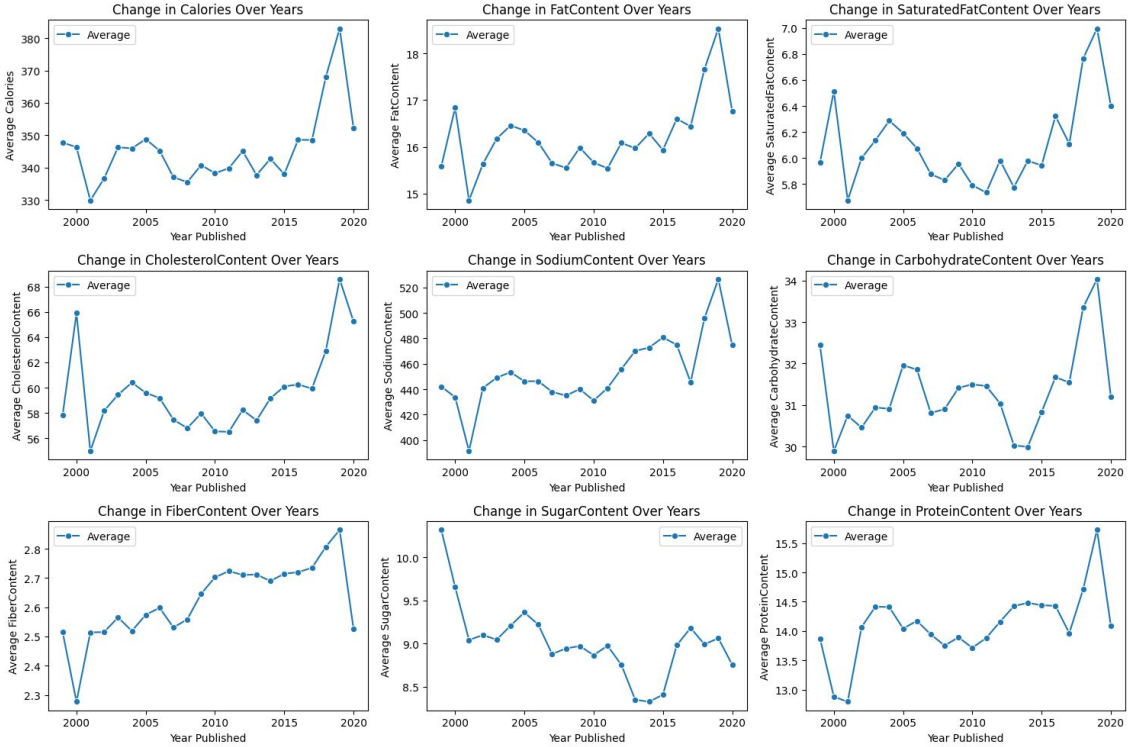
Interesting Insights

- How does nutrition contents and calories influence ratings?



Interesting Insights

- How has nutrition content changes over years?



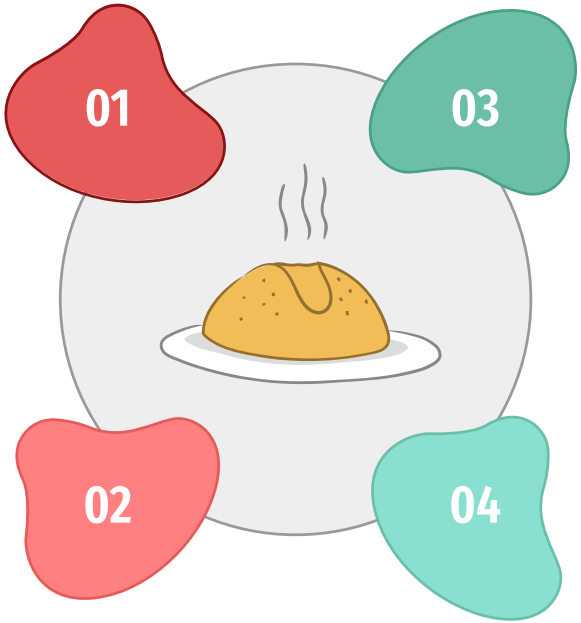
Recommendation Algorithm Framework

Data Collection

Extract and load recipe data into a pandas dataframe

Feature Engineering

Extract nutrition info from raw data for recommendations



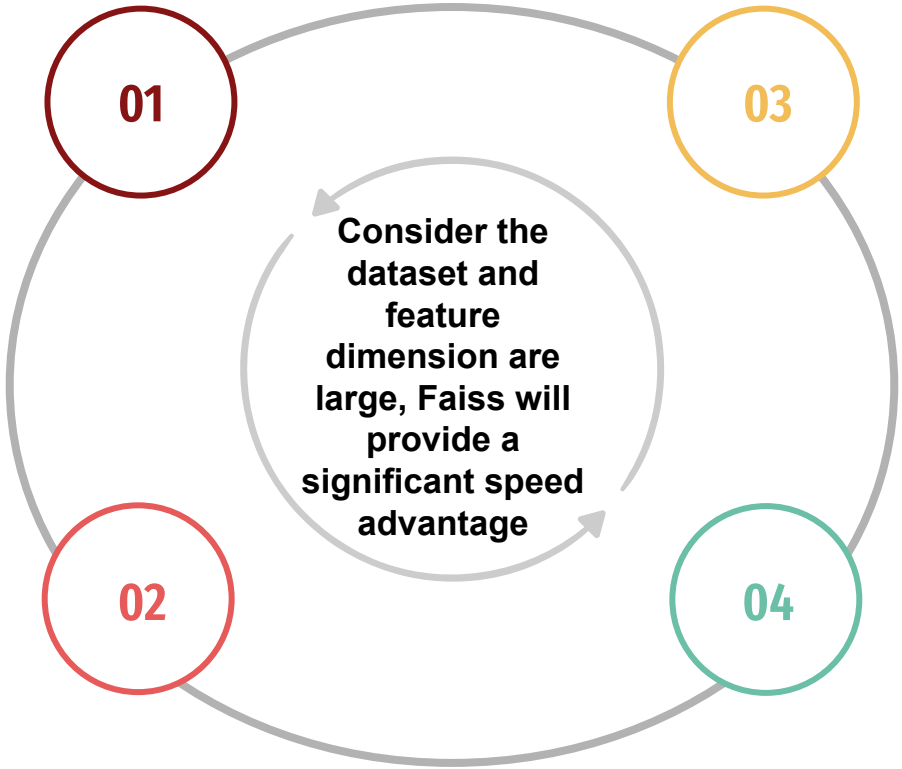
Recommendation Model

Implement content-based recommendation using KNN with cosine similarity

Evaluation & Feedback

Define user preferences, obtain and display recommended recipes

KNN VS FAISS


01
Import the FAISS library
02
Create a FAISS index based on L2 distance
03
Search for the 5 recipes closest to the user's preferences
04
Retrieve recommendation results
Consider the dataset and feature dimension are large, Faiss will provide a significant speed advantage

Refined recommendation algorithm

01

input calories requirement & nutrition contents

02

input user's height, weight, health target and the ingredients

Original Design

A more straightforward recommendation system

```
## test case
user_preferences = {
    "Calories": 500,
    "FatContent": 0.2,
    "SaturatedFatContent": 0.1,
    "CholesterolContent": 0.1,
    "SodiumContent": 0.2,
    "CarbohydrateContent": 0.3,
    "FiberContent": 0.5,
    "SugarContent": 0.1,
    "ProteinContent": 0.8
}
recommended_recipes = recommend_recipes_knn(data, user_preferences, top_n=5)
## recommend based on KNN with cosine similarity
recommended_recipes
```

```
### Test case
height = 1.75  # height(m)
weight = 70    # weight(kg)
goal = "fat_loss"  # health target: "fat_loss", "muscle_gain", "maintain"
ingredients = {"chicken", "broccoli", "rice"}  # ingredients that user has

# recommend recipes
recommended_recipes = recommend_recipes(height, weight, goal, ingredients, data, top_n=5)
```

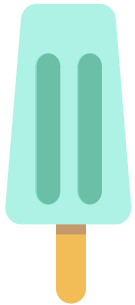
Refined recommendation algorithm

Data Collection

RecipeIngredientsPart
&
Calories

Recommendation model

Ingredient similarity
&
needed calorie calculation
based on user's input



Feature Engineering

Preprocess RecipeIngredientParts by doing text cleaning and standardization

Ranking & reranking

Calculate ingredient similarity and calorie difference to get recipes with least calories based on specific ingredients

Demo

Input testcase

height = 1.75 m, weight = 70 kg,
goal = fat_loss,
ingredients = {“chicken”, “broccoli”, “rice”}

	Name	Calories	FatContent	SaturatedFatContent	CholesterolContent	SodiumContent	CarbohydrateContent	FiberContent	SugarContent	ProteinContent	RecipeIngredientParts	RecipeInstructions	IngredientSimilarity	CalorieDiff
373932	Frozen Baby Meal 4 of 6 (Chicken, Rice & B...	21.9	0.0	0.0	0.0	0.1	4.8	0.1	0.0	0.4	c("chicken", "rice")	c("Cook chickend and shred well", "Add cream o...	0.816497	1326.1
476581	Chicken Broccoli Rice Casserole	634.8	15.1	7.5	33.1	781.2	25.8	2.3	1.4	17.4	c("rice", "chicken breasts", "broccoli", "ched...	c("Combine all ingredients in a large bowl and...	0.707107	713.2
328637	Easy Chicken and Rice	594.2	16.1	4.0	93.6	312.5	76.0	3.9	3.5	34.6	c("water", "white rice", "chicken meat", "broc...	c("set oven at 350* F.", "spray oven dish with...	0.707107	753.8
311702	Megan G's Broccoli Soup	252.0	2.1	0.5	0.0	683.1	48.2	3.5	3.7	10.2	c("chicken broth", "broccoli", "rice", "parmes...	c("Let the broth come to a boil; add the brocc...	0.707107	1096.0
177183	Finicky Feline Diet	96.6	2.5	0.7	26.2	58.6	8.1	0.6	0.5	9.8	c("cooked rice", "broccoli", "carrot", "chicke...	"Process all ingredients in a food processor o...	0.707107	1251.4

Further exploration on recipe reviews

Step 2

Select the best recipe based on Rating and review number

Step 1

Use the recipe index to get their reviews

Step 3

Do sentiment analysis on the reviews: find the most positive and suggestive one

Demo

01

Trace Reviews

Input testcase

height = 1.75 m, weight = 65 kg,
goal = muscle_gain
ingredients = {“chicken”, “carrot”, “potato”}

```
### Test case
height = 1.75  # height(m)
weight = 65    # weight(kg)
goal = "muscle_gain"  # health target: "fat_loss", "muscle_gain", "maintain"
ingredients = {"chicken","carrot","potato"}  # ingredients that user has

# recommend recipes
recommended_recipes = recommend_recipes(height, weight, goal, ingredients, data, top_n=10)
```

```
selected = review[review["RecipeId"].isin(recommended_recipes["RecipeId"])]
```

Demo

02

Choose top-rated recipe

```
selected = review[review["RecipeId"].isin(recommended_recipes["RecipeId"])]
ratings_per_recipe = selected.groupby("RecipeId")["Rating"].apply(list).head(5)
ratings_per_recipe = selected.groupby("RecipeId")["Rating"].agg(['mean', 'count'])

# Calculate a weighted score (you can adjust the factor to balance between the two metrics)
ratings_per_recipe['weighted_score'] = ratings_per_recipe['mean'] * np.log(ratings_per_recipe['count'])
top_3_recipes = ratings_per_recipe.sort_values(by='weighted_score', ascending=False).head(3)
top_3_recipes
```

RecipeId	mean	count	weighted_score
43222	4.574074	54	18.245909
152322	4.750000	12	11.803307
52615	4.300000	10	9.901116

RecipeId	Name	Calories	FatContent
43222	Easiest Chicken Pot Pie	311.4	16.7

Demo

03

Sentiment Analysis

Find the most positive and the most suggestive review
“What other people tried with this recipe?”
“What can I learn from the review?”

```
def clean_text(text):
    text = re.sub(r'[^a-zA-Z\s]', '', str(text).lower())
    return text
def get_sentiment(text):
    return TextBlob(str(text)).sentiment.polarity

def detect_suggestion_score(text):
    suggestion_keywords = [
        'would', 'could', 'might', 'suggest', 'recommend',
        'try', 'maybe', 'perhaps', 'alternative', 'modify',
        'change', 'adjust', 'swap', 'replace', 'improve'
    ]
```

Most Suggestive Review:
Suggestion Score: 6
Review:
Very easy and flavorful recipe. I did **swap out the can of cream of potato soup for a can of cream** of chicken soup which added the chickeny goodness a pot pie needs. To replace the potatoes that were omitted from not adding the cream of potato soup **I simply cut up a potato into little cubes & boiled them for 12 minutes** then added them to my thawed veggie mix. Lastly I added 1/2 of a 10 oz. tub of Philadelphia Cooking Creme (Reduced Fat Italian Cheese & Herb) & mixed it into the whole mess. **These little tweaks added a lot of flavor plus a rich & creamy texture that my whole family loved**. I would have given it 5 stars if I didn't have to "adjust" the recipe as much as I did but I will certainly make this again!

Thanks!