IE 551: ADV TOPICS: STATISTICS FOR MACHINE LEARNING
PROJECT 2: RECIPE RECOMMENDATION SYSTEM

GROUP MEMBERS:

NAME: SRI HARSHA SAMAYAMANTHULA NETID: ss4305
NAME: SAI RAM BALU NETID: sb2310

ABSTRACT:

•Many people worldwide are becoming more worried about their health and how they live in today's world. Just avoiding junk food and exercising isn't enough. It's also important to eat a balanced diet. A balanced diet depends on your height, weight, and age. Eating right can help you stay at a healthy weight, reduce the risk of diseases like cancer and heart disease, and improve overall health when combined with exercise.

•However, there aren't many advanced projects for recommending food or diets. So, We thought of making a recipe recommendation system using machine learning for this purpose.

RECIPE recommendation system

•Recipe recommendation system plays a crucial role in encouraging healthier eating habits. This engine analyses the nutritional details and ingredient lists of different foods to suggest personalized choices to users. What makes this approach beneficial is its capacity to consider each person's unique dietary needs and tastes, including allergies and food aversions. By offering users customized suggestions, a content-based recommendation system can guide them toward healthier eating decisions, ultimately contributing to improved health.

•Furthermore, by suggesting a wide range of nutritious foods, this type of engine can inspire users to explore new food options, reducing the monotony often associated with restrictive diets. This variety helps individuals maintain a balanced

diet, which is likely to have a favorable impact on their long-term health and well-being.

Data description:

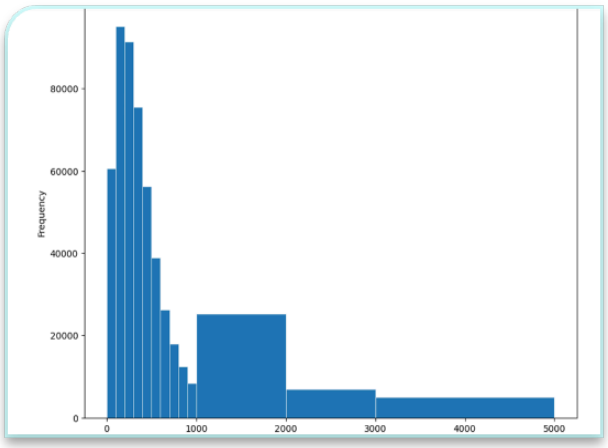
•The dataset used in this project is scraped from Food.com, featuring over 500,000 recipes and 1,400,000 reviews. It provides a comprehensive overview of various recipes, complete with details such as cooking and preparation times, ingredients, and nutritional content. The primary focus is on nutritional information, which includes attributes like calories, fat content, saturated fat, cholesterol, sodium, carbohydrates, fiber, sugar, and protein. Given the size of the dataset, it offers a rich source of information for analysis and model building.

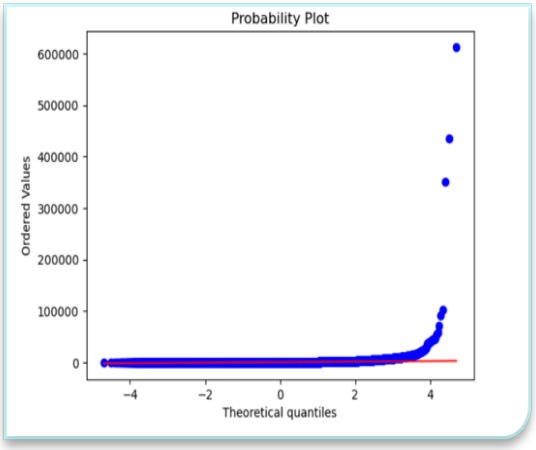
```
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 522517 entries, 0 to 522516
Data columns (total 28 columns):
 #   Column                      Non-Null Count   Dtype
---  ------                      --------------   -----
 0   RecipeId                    522517 non-null  int64
 1   Name                        522517 non-null  object
 2   AuthorId                    522517 non-null  int64
 3   AuthorName                  522517 non-null  object
 4   CookTime                    439972 non-null  object
 5   PrepTime                    522517 non-null  object
 6   TotalTime                   522517 non-null  object
 7   DatePublished               522517 non-null  object
 8   Description                 522512 non-null  object
 9   Images                      522516 non-null  object
 10  RecipeCategory              521766 non-null  object
 11  Keywords                    505280 non-null  object
 12  RecipeIngredientQuantities  522514 non-null  object
 13  RecipeIngredientParts       522517 non-null  object
 14  AggregatedRating            269294 non-null  float64
 15  ReviewCount                 275028 non-null  float64
 16  Calories                    522517 non-null  float64
 17  FatContent                  522517 non-null  float64
 18  SaturatedFatContent         522517 non-null  float64
 19  CholesterolContent          522517 non-null  float64
 20  SodiumContent               522517 non-null  float64
 21  CarbohydrateContent         522517 non-null  float64
 22  FiberContent                522517 non-null  float64
 23  SugarContent                522517 non-null  float64
 24  ProteinContent              522517 non-null  float64
 25  RecipeServings              339606 non-null  float64
 26  RecipeYield                 174446 non-null  object
 27  RecipeInstructions          522517 non-null  object
dtypes: float64(12), int64(2), object(14)
memory usage: 111.6+ MB
```

Data Analysis and Visualization:

•To understand the distribution of key nutritional attributes in the dataset, we have created a histogram to visualize the frequency of calories across the recipes. This visualization uses bins to categorize the calorie content, allowing us to see where most recipes fall in terms of calorie count. Additionally, we generated a probability plot (also known as a Q-Q plot) to examine how the distribution of calorie content compares with a normal distribution. This provides insights into the skewness or

outliers in the data, indicating that some recipes contain significantly higher calorie content than others.





Data Transformation:

•Given the large number of recipes and varying nutritional content, it was necessary to apply some transformations to the dataset. We filtered the data based on specific nutritional limits to focus on recipes that align with general dietary guidelines. The applied limits were:

•Calories: less than 2000, Fat content: less than 100, Saturated fat content: less than 13

•Cholesterol content: less than 300, Sodium content: less than 2300, Carbohydrates: less than 325

•Fiber content: less than 40, Sugar content: less than 40, Protein content: less than 200

•By applying these filters, we reduced the dataset to a more manageable subset for further analysis. The filtered data was then standardized using the StandardScaler to ensure consistent scaling across all attributes.

Methodology:

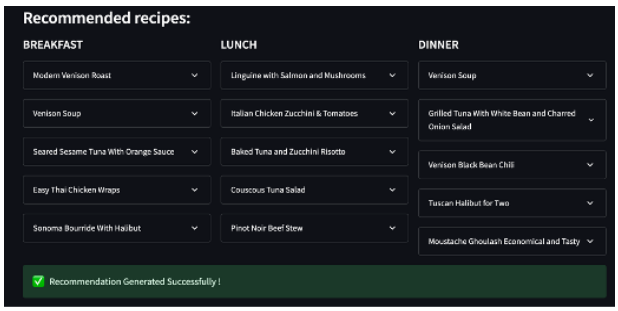
•The recommendation engine for this project is built using the Nearest Neighbors algorithm, a type of unsupervised learning used for neighbour searches. Given the dataset size and desired simplicity, I used the brute-force approach with cosine similarity as the metric for measuring distances between recipes. This method is computationally efficient for smaller datasets and provides a straightforward way to identify similar recipes based on their nutritional content.

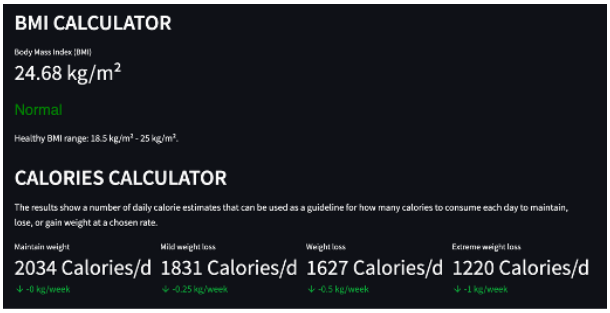
•To build the recommendation model, I used a pipeline that included the StandardScaler for data normalization and the NearestNeighbors algorithm with cosine similarity. The parameters used for this model included a neighbor count of 5 and no distance information returned. This setup allows for a straightforward computation of the nearest neighbors for any given input

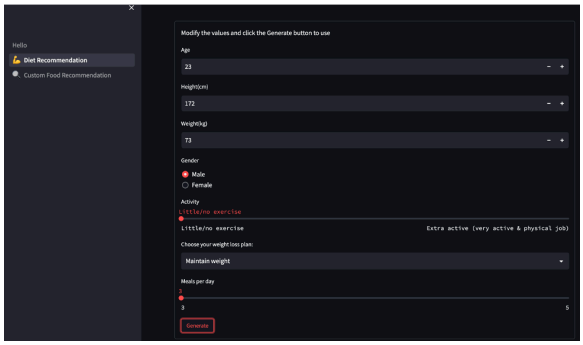
ONLINE accessibility:

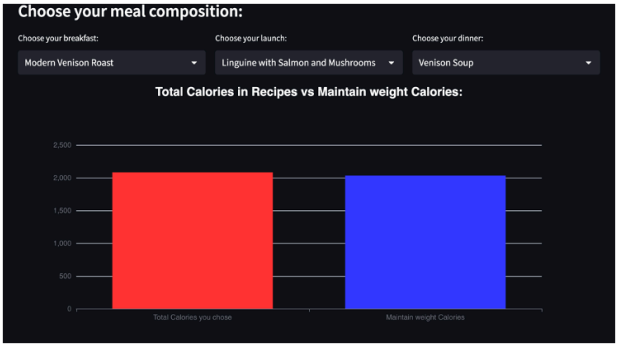
•Once the model was developed, we focused on making our work user-friendly by designing a user interface. To achieve this, we used Streamlit, an open-source platform that allows for rapid prototyping and interactive data-driven applications on the frontend.

•For the backend, we opted for FASTAPI, a modern framework known for its speed and simplicity. In the backend, we imported the model, and whenever a user requests a recipe recommendation, FASTAPI processes the request, communicates with the model, and generates suggestions based on the user's input.


Recommended recipes:
BREAKFAST
LUNCH
DINNER
Modern Venison Roast
Linguine with Salmon and Mushrooms
Venison Soup
Venison Soup
Italian Chicken Zucchini & Tomatoes
Grilled Tuna With White Bean and Charred Onion Salad
Seared Sesame Tuna With Orange Sauce
Baked Tuna and Zucchini Risotto
Venison Black Bean Chili
Easy Thai Chicken Wraps
Couscous Tuna Salad
Tuscan Halibut for Two
Sonoma Bourride With Halibut
Pinot Noir Beef Stew
Moustache Ghoulash Economical and Tasty
Recommendation Generated Successfully !


BMI CALCULATOR
Body Mass Index (BMI)
24.68 kg/m²
Normal
Healthy BMI range: 18.5 kg/m² - 25 kg/m².
CALORIES CALCULATOR
The results show a number of daily calorie estimates that can be used as a guideline for how many calories to consume each day to maintain, lose, or gain weight at a chosen rate.
Maintain weight
2034 Calories/d
-0 kg/week
Mild weight loss
1831 Calories/d
-0.25 kg/week
Weight loss
1627 Calories/d
-0.5 kg/week
Extreme weight loss
1220 Calories/d
-1 kg/week


Hello
Diet Recommendation
Custom Food Recommendation
Modify the values and click the Generate button to use
Age
23
Height(cm)
172
Weight(kg)
73
Gender
Male
Female
Activity
Little/no exercise
Little/no exercise
Extra active (very active & physical job)
Choose your weight loss plan:
Maintain weight
Meals per day
3
5
Generate


Choose your meal composition:
Choose your breakfast:
Choose your launch:
Choose your dinner:
Modern Venison Roast
Linguine with Salmon and Mushrooms
Venison Soup
Total Calories in Recipes vs Maintain weight Calories:
2,500
2,000
1,500
1,000
500
0
Total Calories you chose
Maintain weight Calories


Nutritional Values:
Calories
FatContent
SaturatedFatContent
CholesterolContent
SodiumContent
CarbohydrateContent
FiberContent
SugarContent
ProteinContent

Customised recipe recommendation:

•This page offers personalized recipes based on user-specified nutrient values. A frontend slider allows users to adjust nutrient amounts such as fats, sodium, proteins, and more. As users set their desired nutrient levels, the system recommends suitable recipes.

•Additionally, users can enter specific ingredients, ensuring the recommendations align with their preferences while meeting the nutrient criteria. A chart library is utilized to visualize the nutrient content of each recommended recipe through pie charts.