

MAIS 202 - PROJECT DELIVERABLE 1

Project name: Personalized Disease-Based Recipe Recommendation System

Team Members: Anousha Mahboubi Pour, Lamees Amin Khan (Team 8)

1) Overview

We are developing a web-based application which helps users find healthy recipes based on their medical conditions and available ingredients at home. This app will ensure users receive dietary recommendations that aligns with their health needs making it easier for them to decide what they can eat.

2) Choice of Dataset

We will be using the following datasets to train our model:

- **Diet Recommendations Dataset (Kaggle):** It provides dietary restrictions for diseases such as obesity, diabetes, and hypertension which will help our AI model generate recipes based on the recommended nutritional values for each disease. <https://www.kaggle.com/datasets/ziya07/diet-recommendations-dataset>
- **USDA Food Database:** This dataset contains nutritional values for individual ingredients. The nutritional values for the ingredients that the user will input can be extracted from here. <https://fdc.nal.usda.gov/>
- **Kaggle Recipe Dataset:** A structured dataset containing thousands of recipes with nutritional values. Using this dataset, the AI model will provide recipes based on what ingredients are available. <https://www.kaggle.com/datasets/shuyangli94/food-com-recipes-and-user-interactions>

3) Methodology

Data Preprocessing: We will standardize ingredient names and remove duplicates to ensure uniformity across different datasets. Recipes that violate dietary restrictions, such as high-sugar desserts for diabetics, will be filtered out. Additionally, we will convert recipe attributes into structured numerical data. By leveraging the Diet Recommendations Dataset, we will map health conditions to specific dietary restrictions. Finally, we will use information from the Recipe Dataset to match suitable recipes with health constraints.

Machine Learning Model: We will make use of decision trees to determine what recipe suits best based on the user's health condition. The model will analyze the nutritional content of each recipe which contains the ingredients user has. Based on predefined dietary restrictions, it

will classify recipes as either recommended or not recommended. The system will also rank the best-matching recipes to optimize user experience. Decision trees provide clear interpretation and are ideal for this size of datasets, but they are less efficient for larger datasets if we would have used datasets for more than 3 diseases.

Evaluation Metric: To measure if the AI model correctly classified the recipe as recommended or not recommended based on the dietary restrictions, we can make use of confusion matrix since it provides breakdown of true positives, false positives, true negatives, and false negatives, precision, recall and accuracy for the overall correctness of the model.

4) Application

Our long term plan is to convert this into a mobile application, but for now we will implement our model into a web app with a GUI which will ask the user to select what disease they have out of the three options they are provided, and to provide a list of ingredients they have available. The model will match the user's dietary restrictions and ingredients available to the recommended recipes. The output will display the recommended recipes each with their nutritional breakdown that would be healthy for the person to eat.