Report Meal Recommendation Project
Muhammad Tayyab

Introduction

1.1 Overview

Weight management is a crucial aspect of maintaining overall health and well-being. With the rise of obesity and related health issues globally (Caballero, 2007), there is an increasing need for effective and sustainable weight management strategies. Traditional approaches to weight management, such as generalized diet plans and exercise regimens, often fail to account for individual differences in nutritional needs, metabolic rates, and lifestyle factors. As a result, these approaches may not be effective for everyone.

The recommendation system is designed as a web application, allowing users to easily input their personal details and dietary preferences to receive tailored meal suggestions. Effective weight management relies heavily on understanding and managing calorie intake, which plays a crucial role in reducing weight by ensuring a caloric deficit is maintained (Finer, 2012).
This App will help users to enter key parameters such as age, height, weight, activity level (ranging from Sedentary to Super Active), sex, and their desired weight loss plan, which includes options like Maintain Weight, Mild Weight Loss, Moderate Weight Loss, and Extreme Weight Loss.

To provide accurate recommendations, the system calculates both the Body Mass Index (BMI) and Basal Metabolic Rate (BMR) from the user's inputs. BMI helps categorize the user's weight status (Katherine M. Flegal, et al., 2013), guiding appropriate dietary adjustments, while BMR indicates the number of calories needed at rest to maintain basic physiological functions. Understanding BMR allows the system to determine the user's daily caloric needs, which are adjusted according to their weight loss goals and activity level to achieve a caloric deficit. Additionally, users can specify dietary preferences, such as low-carb, high-protein, low-saturated fats, or low sugars, which are weighted in the recommendation process. By integrating BMI and BMR calculations with these preferences, the system delivers personalised meal suggestions to support effective and sustainable weight management.

The core of the recommendation system is built on machine learning models that analyze these inputs to suggest meals that align with the user's nutritional requirements. The target variable for the models is the food name, while the feature variables include nutritional components such as calories, fat, saturated fats, cholesterol, sodium, carbohydrates, fiber, sugar, and protein. The algorithms—Nearest Neighbor Regressor with KD-tree with Manhattan (L1 norm) distance metric is compared with the brute force algorithm with Euclidean distances metric to determine the most effective method for generating recommendations.

In addition to meal recommendations, the system includes a custom diet generator. This feature allows users to specify desired nutritional values using sliders, such as calorie content, carbohydrates, proteins, sugars, fats, and cholesterol. Based on these user-defined parameters, the system will recommend the five to ten best-matching foods, providing a highly tailored dietary suggestion that meets specific nutritional goals.

1.2 Problem Statement

The problem addressed by this project is the inadequacy of traditional, one-size-fits-all dietary recommendations in effectively managing weight. Standard diet plans often provide generalised guidelines that fail to account for individual differences in nutritional needs, lifestyle, and personal health goals, leading to suboptimal results and poor adherence to dietary regimens. As individuals vary significantly in factors such as age, sex, activity level, weight management goals, and dietary preferences, there is a critical need for personalised dietary recommendations that are tailored to these unique characteristics.

Existing dietary recommendations do not effectively integrate individual-specific data to provide customised meal plans that align with personal weight management goals, whether that be weight maintenance, mild to extreme weight loss, or adapting to particular dietary preferences like low-carb or high-protein diets. This lack of personalisation often results in dissatisfaction with food choices and failure to achieve desired weight outcomes, underscoring the limitations of a generalized approach to nutrition.

Furthermore, the current landscape of dietary planning lacks integration of advanced technologies, such as machine learning, which can analyze and interpret complex datasets to generate tailored recommendations. This gap presents an opportunity to leverage machine learning algorithms to create a more responsive and customised approach to meal planning, which adapts to the user's changing needs and preferences over time.

To address this gap, this project aims to develop a recommendation system, which is a app that delivers tailored dietary suggestions based on comprehensive user profiles. By leveraging machine learning algorithms to analyze user inputs—including demographic details, activity levels, and specific dietary preferences—the system will generate customised meal plans that better align with individual nutritional needs and weight management objectives.

1.4 Scope and Limitations

1.4.1 Scope

The scope of this project encompasses the development and evaluation of a recommendation system tailored for weight management, targeting adult users. The project involves several key components: data collection and preprocessing, algorithm implementation, system integration, and performance evaluation. System integration is achieved by creating a application that incorporates user inputs and weight management goals to deliver customised meal suggestions. The primary aim is to provide a user-friendly platform that offers tailored meal plans based on individual user profiles and preferences. Another scope of the system allows users to compare their daily required calorie intake with the calories from their selected foods. Additionally, there is a custom food recommender integrated into the platform, enabling users to specify desired nutritional values.

1.4.2 Limitations

Target Population: The system will not be specifically designed for adult users and may not be suitable for children or adolescents, as their Body Mass Index (BMI) calculations and nutritional requirements differ significantly from those of adults.
Health Conditions: The system will not account for specific health conditions such as heart disease, diabetes, or other chronic illnesses, which require specialised dietary

management. Users with such conditions may need personalised nutrition plans developed by healthcare professionals, which are beyond the scope of this project.

Data Availability: The effectiveness of the recommendation system is highly dependent on the quality, availability, and accuracy of the nutritional data used in the dataset as well as the information provided by users

Generalisation and User Preferences: The system may not fully capture the diverse range of user preferences and cultural dietary practices. While the system allows for some degree of customisation based on user inputs, it may not accommodate all individual preferences or dietary restrictions, potentially limiting its applicability for certain user groups.

Chapter 2: Literature Review

2.1 Introduction to Recommendation Systems in Healthcare

Recommendation systems have become integral tools in various sectors, including healthcare, where they are utilized to manage information overload and provide tailored advice to users. As healthcare increasingly adopts data-driven approaches, the application of recommendation systems in areas like diet, fitness, and disease management has gained significant attention. These systems leverage machine learning algorithms to analyze large datasets and deliver Personalised recommendations, thereby improving the effectiveness of health interventions. In the realm of dietary management, such systems are particularly valuable as they can adapt to individual nutritional needs, preferences, and health conditions, offering users more relevant and actionable Advice.

2.2 Dietary Recommendation Systems

Several studies have explored the use of recommendation systems specifically for dietary planning and health management. The DASH Diet Recommendation System proposed by (Sookrah, et al., 2019) focuses on hypertensive patients, using content-based filtering combined with machine learning to provide dietary recommendations tailored to manage blood pressure. While effective for a specific condition, this system lacks the flexibility to adapt to other health concerns or broader dietary goals.

Similarly, (Mogaveera, et al., 2021) developed an E-Health Monitoring System with Diet and Fitness Recommendation using Machine Learning that integrates diet and fitness recommendations for patients with chronic conditions such as diabetes and thyroid issues. This system uses a Decision Tree algorithm to personalize recommendations based on user health data. However, its focus is primarily on predefined health conditions rather than general weight management, and it does not account for user preferences such as macronutrient customization. Another paper by (Agapito, et al., 2016) introduced DIETOS: A recommender system for adaptive diet monitoring and personalized food suggestion. DIETOS builds user health profiles through real-time questionnaires and adjusts dietary recommendations dynamically based on these profiles. Although DIETOS offers a more adaptable approach than the aforementioned systems, it still emphasizes managing chronic diseases rather than specifically targeting weight loss.

2.3 Predictive Models in Dietary Interventions

Predictive analytics have also played a significant role in dietary management, particularly in forecasting weight outcomes based on dietary interventions. (Babajide, et al., 2020) in their conference paper “A Machine Learning Approach to Short-Term Body Weight Prediction in a Dietary Intervention Program” examined the effectiveness of various machine learning models, including Artificial Neural Networks (ANN) and Random Forests (RF), in predicting short-term body weight changes during a dietary intervention program. Their study demonstrates the potential of machine learning to enhance the accuracy of weight predictions. However, the focus remains on prediction rather than real-time, Personalised dietary adjustments based on user preferences or exercise levels.

In their paper, (Kaur, et al., 2022) further expanded on this by investigating the risk of obesity using advanced machine learning algorithms such as Gradient Boosting and XGBoost. Their study also applied the Nearest Neighbor method for meal planning aimed at reducing obesity. While their approach aligns more closely with the goals of Personalised weight management, it still primarily focuses on obesity risk prediction rather than offering flexible, user-driven customization in dietary recommendations.

2.4 Techniques and Algorithms in Personalised Diet Planning

Across these studies, a variety of machine learning algorithms have been employed to enhance dietary recommendation systems. From content-based filtering in the DASH diet system to Decision Trees in E-Health Monitoring and advanced models like Gradient Boosting in obesity risk prediction, the choice of algorithm significantly impacts the system’s ability to personalize and adapt to individual needs.

However, the systems discussed above generally fall short in offering comprehensive personalization features that allow users to filter diets by specific macronutrient content (e.g., low carbs, high protein, high fiber) or to adjust calorie intake based on exercise levels. These are critical 10factors in effective weight management, where dietary needs can vary widely depending on individual goals, physical activity, and metabolic rates.

2.5 Critical Analysis of the existing works

The DASH Diet Recommendation System is a notable example of a system that effectively addresses the needs of a specific patient group hypertensive individuals. By employing content-based filtering combined with machine learning, the system offers Personalised dietary recommendations aimed at managing blood pressure. Its practical implementation through a mobile application enhances accessibility and user engagement.

However, the system's focus on hypertension limits its generalization to a broader audience, such as individuals aiming for weight loss. Additionally, the DASH system lacks the flexibility for users to filter dietary recommendations based on specific macronutrient goals, such as low carbs or high protein, which are crucial for Personalised weight management. The static nature of its recommendations further limits its adaptability to changes in the user’s health status or lifestyle, such as varying exercise levels.

In contrast, the comprehensive review by (Yue, et al., 2021) on “An Overview of Recommendation Techniques and Their Applications in Healthcare” provides an extensive overview of recommendation techniques, including content-based, collaborative filtering, and hybrid methods, and their applications across multiple healthcare scenarios. This paper offers valuable insights into the challenges of implementing recommendation systems in healthcare, such as ensuring data privacy, maintaining accuracy, and engaging users.

However, its broad theoretical focus and lack of specific case studies or applications limit its direct relevance to weight management or Personalised diet planning. While the review is thorough in its scope, it falls short of providing practical examples of how these techniques are applied in real-world settings, which diminishes its utility for understanding the practical challenges in developing such systems.

(Mogaveera, et al., 2021) offer a more integrated approach with their E-Health Monitoring System, which combines health monitoring with Personalised diet and fitness recommendations. The system uses a C4.5 Decision Tree algorithm to tailor recommendations based on user health data, providing a holistic solution for managing chronic conditions such as diabetes and thyroid issues. This integration of health monitoring and Personalised recommendations is a significant strength, as it addresses multiple aspects of health management simultaneously.
However, the system’s focus on predefined chronic conditions limits its applicability to a broader population, including those seeking weight management without underlying health issues. Moreover, the lack of real-time adaptation to changes in the user’s lifestyle, such as exercise levels, reduces the system’s effectiveness in dynamic, everyday situations. While the system includes both diet and fitness recommendations, it does not dynamically adjust caloric intake based on exercise levels, a critical feature for accurate weight management.

The DIETOS system, represents a significant advancement in adaptive dietary recommendations. By using dynamic health profiles created through real-time questionnaires, DIETOS provides Personalised food suggestions that adapt to changes in the user’s health status. The involvement of medical professionals in developing these profiles adds credibility and ensures that the recommendations are medically sound. Additionally, DIETOS's inclusion of a catalog of regional foods makes the system culturally relevant and more relatable to its target users.
Despite these strengths, DIETOS, like many other systems, primarily targets individuals with chronic diseases, limiting its appeal to a broader audience focused on weight management. The system’s reliance on extensive real-time data collection may also be burdensome for users, potentially raising privacy concerns. Furthermore, while DIETOS is adaptive, it does not allow users to filter recommendations based on specific dietary goals, such as macronutrient content, nor does it adjust caloric intake based on exercise levels, which are essential features for effective weight management.

In the realm of predictive analytics, (Babajide, et al., 2020) explore the application of various machine learning models, including Artificial Neural Networks (ANN) and Random Forests (RF), in predicting short-term body weight changes during a dietary intervention program. Their study demonstrates the potential of machine learning to improve the accuracy of weight predictions, which is invaluable for assessing the effectiveness of dietary interventions. The detailed comparison of different models provides insights into their relative performance in dietary contexts, showcasing the predictive power of advanced algorithms.
However, the system’s focus on prediction rather than real-time, Personalised dietary adjustments limits its direct utility for users seeking immediate, actionable advice. Additionally, the study’s focus on short-term predictions within a specific dietary program may not address the needs of users looking for long-term weight management solutions. The absence of user-driven customization or feedback mechanisms further limits user engagement and adherence to the recommended dietary plans.

Finally, the on Predicting Risk of Obesity and Meal Planning by (Kaur, et al., 2022) expands the focus to obesity risk assessment using advanced machine learning algorithms such as Gradient Boosting and XGBoost. This study successfully integrates obesity risk prediction with meal planning, providing a comprehensive approach to managing and preventing obesity. The use of Nearest Neighbor methods for meal planning aligns with the methodology of our project, adding relevance and context to our chosen approach.

However, while effective in predicting obesity, the system's primary focus on risk assessment may overshadow the importance of providing flexible, user-driven dietary recommendations for ongoing weight management. Additionally, the system's reliance on a dataset from the UCI ML repository may limit its generalizability to diverse populations, and the real-world applicability of the model remains to be fully validated.

Overall, while the existing systems demonstrate significant advancements in Personalised health management, several limitations persist, particularly in terms of personalization, adaptability, and user engagement. Many systems are tailored to specific health conditions or predictive outcomes, often at the expense of broader applicability to weight management. Few systems offer the flexibility to tailor recommendations based on user-defined dietary goals or dynamically adjust caloric intake according to exercise levels.

My project, which employs Nearest Neighbor Regressor for Personalised meal recommendations specifically aimed at weight management, addresses these gaps. By allowing users to filter diets based on specific macronutrient content and adjust recommendations according to their exercise

levels, my system offers a higher level of personalization and adaptability than existing models. This approach not only enhances user engagement but also provides a practical and scalable solution for effective weight management, distinguishing it from previous work in the field.

2.6 Gaps in the existing work and Future Directions

The existing literature highlights significant advancements in dietary recommendation systems, particularly in managing chronic conditions and predicting weight outcomes. However, there remains a gap in developing a system that offers truly Personalised meal recommendations tailored specifically for weight loss. Most current systems do not allow users to filter their diets according to specific macronutrient goals or to dynamically adjust caloric intake based on exercise levels.
This gap is where the current project makes a novel contribution. Users can filter their meal options based on criteria such as low carbs, high protein, or low fats, ensuring that the recommendations align with their specific weight loss strategies. Additionally, the system adjusts caloric recommendations based on the user's exercise level, providing a more holistic and user-centric approach to weight management.
This project advances the field by combining the predictive power of machine learning with the flexibility of user-driven customization, addressing the limitations of existing systems. It offers a scalable and adaptive solution that not only predicts weight outcomes but actively supports users in achieving their weight loss goals through Personalised dietary guidance.

Data Collection

The success of any machine learning project, particularly in the realm of recommendations, hinges on the quality and comprehensiveness of the data used. For this project, food and nutrition data were sourced from a publicly available dataset on Kaggle (Anon., 2024), which is known for its extensive and diverse repositories of structured data. The selected dataset contains detailed information on 522,517 recipes, categorized into 312 different food categories. Each recipe entry in this dataset includes multiple attributes such as cooking times, servings, ingredients, nutrition facts, and preparation instructions, among other details. These attributes provide a rich foundation for building a recommendation system tailored to individual nutritional needs and preferences.

3.3.1 Data Exploration and Preliminary Processing

To begin the process of building the diet recommendation system, the initial focus was on understanding and exploring the recipe dataset, which forms the foundation of the recommendation engine.

3.3.2 Selection of Relevant Data

Given the large number of features in the original dataset, the first step in the data exploration phase was to identify and extract only the most relevant columns for the purpose of the project. The columns selected were as follows:
• **Name**: This was considered as the target column, as it represents the recipe names, which are essential for identifying the recommended recipes.
• **Nutritional Features**: These columns are crucial as they provide detailed nutritional information that directly impacts diet and weight loss recommendations:

1. Calories
2. FatContent
3. SaturatedFatContent
4. CholesterolContent
5. SodiumContent
186. CarbohydrateContent
7. FiberContent
8. SugarContent
9. ProteinContent

· **RecipeIngredientParts**: Lists the ingredients used in each recipe, which are critical for understanding the composition of the dish and for generating recommendations.
• **CookTime, PrepTime, TotalTime**: These columns provide insights into the time required to prepare and cook the recipes, which can be useful for personalizing recommendations based on the user's time constraints.
• **RecipeInstructions**: Contains step-by-step instructions for preparing the recipes, ensuring that the recommended dishes can be easily followed by the user.
• **Images**: While primarily for visual representation, this column could enhance the user experience by providing a visual cue for each recommended dish.
By focusing on these relevant columns, the dataset was streamlined to include only the data necessary for effective analysis and model building. This step not only reduced the computational load but also ensured that subsequent analyses were more targeted and meaningful.

3.3.3 Data Cleaning

After selecting the relevant columns for the diet recommendation system, the next critical step was to clean the data to ensure its quality and suitability for analysis. This phase involved two primary tasks: removing duplicate entries and filtering the data to include only recipes that fall within recommended daily nutrient intake ranges.

3.3.4 Removing Duplicate Recipes

The dataset contained over half a million recipes, making it essential to ensure that each entry was unique to avoid redundancy and bias in the recommendation system. Duplicate recipes, particularly those with identical names, were identified and removed to ensure that each recipe

was represented only once. This was accomplished using the following operation:
By dropping these duplicates, the dataset was refined to include only distinct recipes, thus preventing any single recipe from skewing the analysis or recommendations.

3.3.5 Filtering Recipes by Nutritional Content

Given the focus on weight loss and healthy eating, it was important to filter out recipes that exceeded recommended daily nutrient intake levels. For this purpose, the dataset was filtered to retain only those recipes whose nutritional content fell within specific daily thresholds. These thresholds were based on standard dietary guidelines (NHS, 2022) and included:
• **Calories**: Maximum 2000 kcal
• **Fat Content**: Maximum 100 grams
• **Saturated Fat**: Maximum 13 grams
• **Cholesterol**: Maximum 300 milligrams
• **Sodium**: Maximum 2300 milligrams
• **Carbohydrates**: Maximum 325 grams
• **Fiber**: Maximum 40 grams
• **Sugar**: Maximum 40 grams
• **Protein**: Maximum 200 grams
By applying these nutritional filters, the dataset was tailored to include only those recipes that align with healthy dietary standards, ensuring that the recommendations generated by the system are conducive to weight loss and overall health.

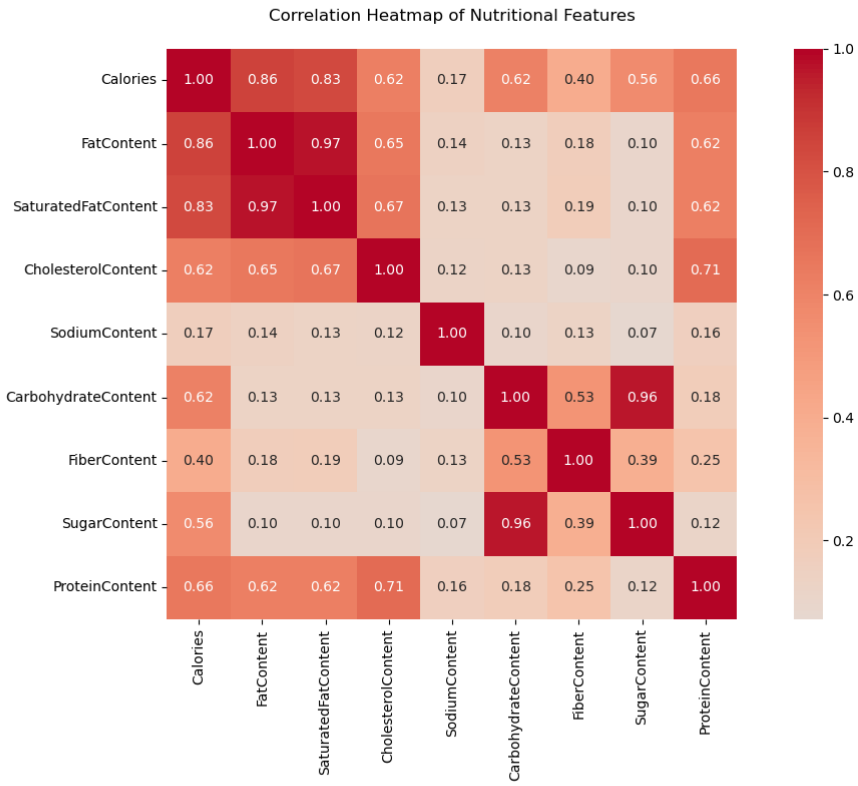
Following the data cleaning process, the dataset was significantly refined to ensure it was ready for further analysis and modeling. The updated dataset now contains 316,872 unique recipe entries, each with comprehensive information about the recipe's nutritional content, preparation details, and other relevant attributes.

The cleaned dataset consists of 16 columns, which provide a range of information necessary for the diet recommendation system. Below is a summary of the dataset's structure:

20• **Total Entries (Rows):** 316,872
• **Total Columns:** 16
3.3.6 Data Types and Completeness
• **Numerical Data**: The dataset includes nine numerical columns (`float64` type) that represent various nutritional metrics. These are fully populated with no missing values, ensuring a robust basis for analysis.
• **Text Data**: Seven columns contain textual data (`object` type), including the recipe name, ingredient list, cooking/preparation times, instructions, and image paths. The '`CookTime`' column has some missing values, which could be addressed in future preprocessing steps if necessary.

3.4 Exploratory Data Analysis (EDA)

To gain deeper insights into the relationships between the various nutritional features in the dataset, correlation analysis was conducted which is shown in the Figure 2. This step was crucial in understanding how different nutritional components are related to one another, which can inform the development of a more effective diet recommendation system.



3.4.1 Correlation Analysis

The first step in this analysis involved calculating the Pearson correlation coefficients between the nutritional variables. The correlation matrix was then visualized using a heatmap, which provides an intuitive representation of the relationships between the variables.
The heatmap above illustrates these correlations, where each cell represents the correlation coefficient between two variables. The color scale ranges from dark blue (indicating a negative correlation) to bright yellow (indicating a strong positive correlation). The values closer to 1 imply a strong positive correlation, values closer to -1 indicate a strong negative correlation, and values around 0 suggest no significant correlation.

Key Observations

• **Calories**: This variable shows a strong positive correlation with FatContent (0.77), CarbohydrateContent (0.71), and ProteinContent (0.69). This indicates that recipes higher in calories tend to also be high in these nutrients.
• **FatContent and SaturatedFatContent**: As expected, FatContent and SaturatedFatContent are strongly correlated (0.77), which is typical since saturated fat is a component of total fat.
• **CholesterolContent**: There is a moderate correlation between CholesterolContent and SaturatedFatContent (0.51) and ProteinContent (0.68), suggesting that recipes high in cholesterol also tend to be higher in these components.
• **CarbohydrateContent and FiberContent**: The correlation between CarbohydrateContent and FiberContent is moderate (0.59), which is consistent with the fact that dietary fiber is a component of carbohydrates.
• **SugarContent**: This variable has relatively low correlations with most other nutrients, but it has a small positive correlation with CarbohydrateContent (0.39).
This heatmap provided a clear visual summary of how these nutritional features interact, allowing for better understanding and guiding further steps in the analysis and model development. It

highlights the relationships that need to be considered when making dietary recommendations, ensuring that the system can provide balanced and healthy options based on user needs.