OPEN ACCESS
Indo-JC
Indonesia Journal of Computing
ISSN 2460-9056
socj.telkomuniversity.ac.id/indojc

Ind. Journal on Computing
Vol. 9, Issue. 2, August 2024. pp. 83-96
doi:10.34818/indojc.2024.9.2.959

# Diet and Physical Exercise Recommendation System Using a Combination of K-Means and Random Forest

Muhammad Ilham Hafizha[#1], Z. K. A. Baizal[*2]

*# School of Computing, Telkom University*
*Bandung, Indonesia*
[1] milhamhamfizh@student.telkomunivesity.ac.id
[2] baizal@telkomunivesity.ac.id

**Abstract**

Public health has become a significant focus in the modern era due to the increasing number of people suffering from various diseases. Unhealthy eating habits and lack of physical activity are often linked to health issues, including obesity. Several studies have developed food recommendation systems for individuals with obesity using K-Means and Random Forest algorithms to provide recommendations based on specific user aspects. However, these studies do not offer physical activity recommendations to address a lack of fitness or exercise. This research develops a diet and physical exercise recommendation system for individuals with obesity using a combination of K-Means and Random Forest. The system categorizes and classifies foods and physical activities, providing personalized recommendations based on user data input. These recommendations consider body mass index (BMI), age, weight, and food preferences like vegan or non-vegan. The system provides diet and exercise plans tailored to the individual needs of each user. The system's accuracy is evaluated using the Mean Absolute Percentage Error (MAPE) metric, with the highest accuracy for food recommendations being 99.03% for non-vegan lunch and the lowest being 70.74% for vegan breakfast. The MAPE for exercise recommendations is consistently at 26.35%, indicating a stable accuracy of 73.65%. The testing results show that the system accurately recommends diet and physical exercise plans for each user.

**Keywords**: Public Health, Obesity, Unhealthy Diet, Lack of Physical Activity, BMI, Diet Recommendation System, Physical Exercise, K-Means, Random Forest, Mean Absolute Percentage Error

**Abstrak**

Kesehatan masyarakat menjadi fokus utama di era modern karena meningkatnya jumlah penderita berbagai penyakit. Pola makan yang tidak sehat dan kurangnya aktivitas fisik sering dikaitkan dengan masalah kesehatan, termasuk obesitas. Beberapa penelitian telah mengembangkan sistem rekomendasi makanan untuk individu dengan obesitas menggunakan algoritma K-Means dan Random Forest untuk memberikan rekomendasi berdasarkan aspek spesifik pengguna. Namun, penelitian-penelitian tersebut tidak memberikan rekomendasi aktivitas fisik untuk mengatasi kurangnya kebugaran atau olahraga. Penelitian ini mengembangkan sistem rekomendasi diet dan latihan fisik untuk individu dengan obesitas menggunakan perpaduan K-Means dan Random Forest. Sistem ini mengkategorikan dan mengklasifikasikan makanan dan aktivitas fisik, serta memberikan rekomendasi yang dipersonalisasi berdasarkan input data pengguna. Rekomendasi ini mempertimbangkan berbagai faktor, seperti indeks massa tubuh (IMT), usia, berat badan, dan preferensi makanan seperti vegan atau non-vegan. Sistem ini memberikan rencana diet dan latihan fisik yang disesuaikan dengan kebutuhan individu setiap pengguna. Akurasi sistem dievaluasi menggunakan metrik Mean Absolute Percentage Error (MAPE), dengan akurasi tertinggi untuk rekomendasi makanan adalah 99,03% untuk makan siang non-vegan, dan terendah adalah 70,74% untuk sarapan vegan. MAPE untuk rekomendasi latihan fisik konsisten di angka 26,35%,

Received on April, 2024. Accepted on August, 2024

menunjukkan akurasi stabil sebesar 73,65%. Hasil pengujian menunjukkan bahwa sistem merekomendasikan rencana diet dan latihan fisik secara akurat bagi setiap pengguna.

**Kata Kunci**: Kesehatan Masyarakat, Obesitas, Pola Makan Tidak Sehat, Kurangnya Aktivitas Fisik, Sistem Rekomendasi Diet, Latihan Fisik, IMT, K-Means, Random Forest, Mean Absolute Percentage Error

## I. Introduction

In this modern era we face serious health challenges, particularly those related to unhealthy eating habits and lack of physical activity. Various studies have shown that poor eating habits are a major factor behind numerous health issues, including obesity. According to UNICEF's conceptual framework, it is known that the combination of unhealthy diets and lack of physical activity significantly contributes to the increase in obesity cases [1]. This is supported by recent data from the World Health Organization, which shows that global obesity cases have more than doubled from 1980 to 2014. The data indicates that approximately 13% of the world's adult population, comprising 11% of men and 15% of women, are obese. There are over 1.9 billion overweight adults, with more than 600 million of them being obese. The primary factors for this increase are the consumption of high-calorie, fat-rich foods and the decline in physical activity, particularly due to predominantly sedentary work [2].

With the rapid advancement of technology in today's era, information spreads quickly and can be accessed from anywhere. Although much information is available on healthy eating patterns and physical exercise, many people still need more personalized guidance and support to achieve their health and fitness goals. Additionally, with busy schedules and conflicting priorities, finding the time and motivation to prioritize healthy habits poses challenges. This includes information on managing obesity, which is crucial for treating and preventing diseases. However, much of the available information is general and lacks specificity. For example, information on diet and exercise recommendations often needs to be tailored to individual circumstances, thus providing insufficient guidance that is both precise and effective.

Various studies have been conducted on the development of food recommendation systems for obese individuals, including a study in India that developed a system using the K-Means and Random Forest algorithms to provide diet and exercise plan recommendations to users [3]. This system allows users to input their age, gender, and weight. The system then includes food recommendations based on the user's Body Mass Index (BMI). According to the research findings, this system has improved the accuracy and efficiency of recommendations.

The previous research recommended breakfast, lunch, and dinner based on the user's Body Mass Index (BMI). However, it did not provide supporting information, such as physical activity recommendations to address fitness or physical inactivity issues. Therefore, we developed a diet and exercise recommendation system simultaneously. This system will implement the K-Means algorithm in classifying food data into three main categories: breakfast, lunch, and dinner. The K-Means algorithm, which is an iterative clustering method, is used to separate the dataset into homogeneous clusters. By reducing the variation within clusters and increasing the differences between clusters, this research aims to improve classification accuracy, allowing users to input their age, gender, and weight. Furthermore, the Random Forest algorithm is used to assist in creating classes from the dataset. The Random Forest algorithm is a collection of decision trees combined to recommend diet and exercise. This system is designed to provide diet and exercise recommendations based on the user's Body Mass Index (BMI) to provide a comprehensive solution for users with tailored recommendations based on their BMI. The system has various benefits for people. It provides personalized recommendations based on an individual's BMI, making it more relevant and effective than general guidelines. By combining diet and physical activity recommendations, the system helps users to adopt an overall healthy lifestyle. The results of this study are expected to help address the problems that users face in maintaining and improving their health by choosing diet and exercise menus in their daily activities.

## II. LITERATURE REVIEW

Diet and exercise recommendation systems have become a popular research topic in recent years. This research focuses on developing systems that can provide diet and physical exercise recommendations tailored to individual needs based on various factors such as age, food type, and BMI. In this context, several studies have explored the use of K-Means and Random Forest algorithms.

Singh Kardam et al. (2021) [3] implemented the K-Means and Random Forest algorithms in testing a diet recommendation system because both algorithms have complementary advantages in the classification and grouping of data. The K-Means algorithm was used to cluster food items into different categories, such as breakfast, lunch, and dinner, allowing the system to organize food items based on calories and relevant food categories. The Random Forest algorithm was then used for further classification, leveraging multiple decision trees to improve accuracy. Thus, implementing these two algorithms enables the system to provide accurate and personalized recommendations tailored to the user's nutritional needs and preferences.

Golagana et al. (2023) [4] emphasized that the K-Means algorithm effectively groups users based on their physical details and health conditions, enabling the system to provide more personalized diet recommendations. This grouping is important as it facilitates the creation of meal plans that align with the user's calorie needs and food preferences. Furthermore, Random Forest is known for its ability to handle high-dimensional data and provide accurate predictions, which is highly useful in predicting user food preferences. This algorithm can handle data complexity and extract important patterns to make precise recommendations. The combination of these two algorithms allows the proposed diet recommendation system to identify similar user groups and predict meal plans that align with individual nutritional needs, thereby improving the quality of recommendations provided to users.

Shah et al. (2023) [5] applied the K-Means and Random Forest algorithms to harness the strengths of both methods in data analysis and accurate predictive model creation. The K-Means algorithm was used to cluster food data based on nutritional characteristics such as calories, allowing the system to filter and recommend foods that align with the user's calorie needs. This clustering helps simplify the recommendation process by categorizing foods into similar groups, making it easier to identify food choices that align with the user's dietary goals. Random Forest is a powerful classification algorithm used to predict the food categories that best suit user preferences. This algorithm combines the results of many decision trees to improve accuracy. With its ability to handle large and complex datasets as well as diverse features, Random Forest is well-suited for tasks such as food classification and diet recommendations that require consideration of many variables. The implementation of these two algorithms aims to create a recommendation system that can provide highly personalized diet and exercise advice, ultimately helping individuals achieve their health and fitness goals more effectively.

### *A. Recommendation System*

Navastara et al. (2019) [6] stated that a recommendation system is a system that can provide users with suggestions for beneficial items/products. Jaja et al. (2020) [7] described a recommendation system as a system that helps users cope with information overload by providing specific recommendations that are expected to meet users' desires and needs. Muhammad Rizqi Az Zayyad and Arrie Kurnawardhani (2021) [8] noted that recommendation systems have been utilized as an effective strategy to manage the vast amount of available information and to provide item recommendations that match users' preferences. Fadlil and Mahmudy (2007) [9] explained that this system operates by collecting data from users both directly and indirectly. Direct data collection can be done by asking users to rate an item, rank their favorite items, choose the best item from a selection, or list the items they like and dislike the most. Indirect data collection involves observing user behavior, such as items viewed by users on e-commerce websites or collecting transaction data from online stores. The collected data is then processed using specific algorithms to predict which items will be useful for users based on the provided information. The results are returned to the users in the form of appropriate item recommendations.

### B. Body Mass Index (BMI)

Body Mass Index (BMI) is a method of measuring weight adjusted for height, calculated by dividing the weight in kilograms by the square of the height in meters (kg/m²). Licenziati et al. (2022) [11] stated that the Body Mass Index (BMI) uses the following formula:

$$\text{BMI} = \frac{\text{Weight}}{\text{Height}} \tag{1}$$

According to the NHLBI [12] stated that the results of the Body Mass Index (BMI) calculation are classified based on the Asia-Pacific Criteria classification in Table I.

TABLE I
Classification Of Asian BMI Criteria

| Asia-Pacific Classification | Body Mass Index |
|---|---|
| Underweight | < 18,5 |
| Normal | 18,5 – 22,9 |
| Overweight | 23 – 24,9 |
| Obesitas I | 25 – 29,9 |
| Obesitas II | ≥ 30 |

The results of Body Mass Index (BMI) calculations can be used as a reference to classify a person's weight into several categories, such as underweight, normal weight, overweight, obesity I, and obesity II.

## III. Research Method

This section presents the research methodology used in designing and implementing the recommendation system for diet and physical exercise. The system utilizes a combination of K-Means and Random Forest algorithms to provide personalized recommendations. The methodology includes data collection, preprocessing, algorithm selection, model training, and evaluation. Additionally, Fig 1 illustrates the system design of the recommendation system. In Fig 1, the process begins by directing the user to a website containing data such as age, weight, mealtime category, vegan/non-vegan, and height. After the user inputs the data, the system processes the user input by preprocessing and then processing the data using the K-Means and Random Forest algorithms. After the model produces predictions, an evaluation is done using the Mean Absolute Percentage Error (MAPE) test metric. Then, food and exercise data suitable for the user's condition will be retrieved and displayed to the user as a result of the diet and exercise recommendations. The flowchart of the system process depicted in Fig 2 is shown below.

The flowchart in Fig 2 outlines the sequential steps from data input to recommendation output, emphasizing the systematic approach employed in this system. The preprocessed data is grouped into several sub-clusters based on the K-Means algorithm in the section. The best sub-cluster is selected to form the desired number of clusters. After clustering, the data is used in the Random Forest model. After splitting the data into training and testing data, the Random Forest algorithm determines the required number of trees, takes a random sample, and creates trees to generate predictions. The prediction results from the Random Forest model are then evaluated. This research methodology ensures a robust framework for developing and deploying an adequate diet and physical exercise recommendation system, contributing to improved health outcomes and user satisfaction.

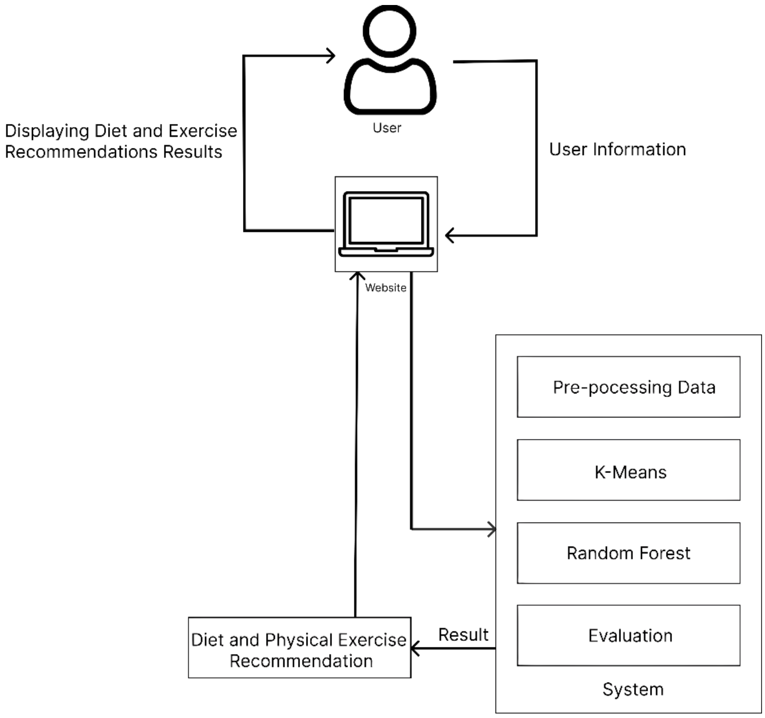


Fig 1. The System Design

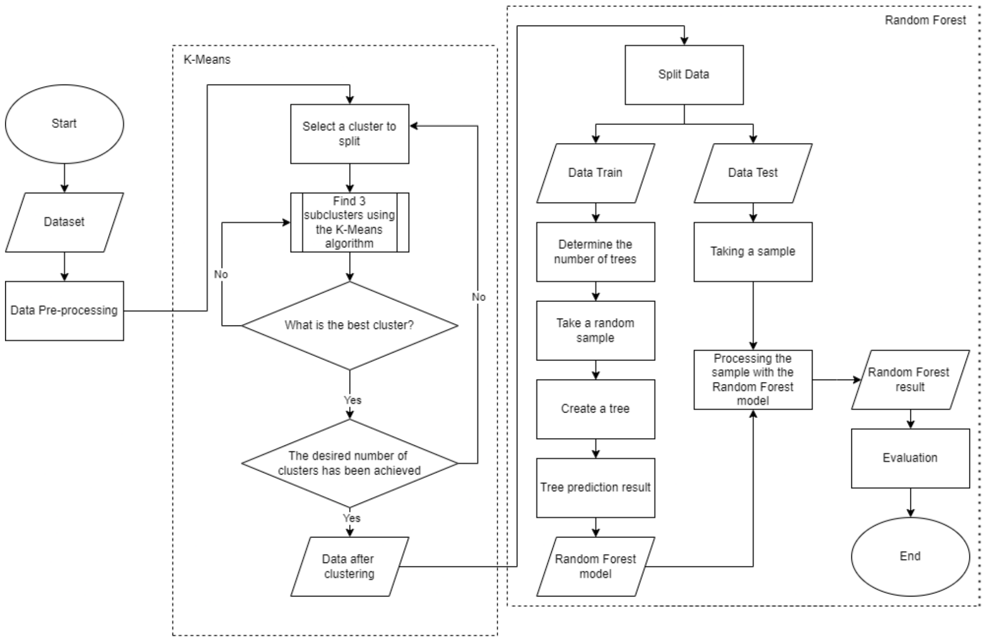


Fig 2. Flowchart System Process

## *A. Dataset*

In the data preparation stage, we utilized publicly available datasets collected from verified sources, particularly from Kaggle. For this research, we selected the Diet and Physical Exercise Datasets, which have been accessed and used by many previous studies. They are considered suitable for analyzing eating patterns based on predefined Body Mass Index (BMI) categories: Underweight, Normal, Overweight, and Obese. The Diet Dataset consists of 89 food entries covering various attributes such as Fooditems, Breakfast, Lunch, Dinner, VegNovVeg, Calories, Fats, Proteins, Iron, Calcium, Sodium, Potassium, Carbohydrates, Fiber, VitaminD, and Sugars. Meanwhile, the Physical Exercise Dataset consists of 3994 entries, including Date, Workout Name, Duration, Exercise Name, Set Order, Weight, Reps, Distance, and Seconds. The use of these datasets has been a common practice in previous research and is included to support the validity and reliability of this study's analysis. Here are the datasets that have been collected:

TABLE II
Diet Dataset

| Food_items | Breakfast | Lunch | Dinner | VegNonVeg | Calories | Fats | Proteins | Iron | Calcium | Sodium | Pottasium | Carbohydrates | Fibre | VitaminD | Sugar |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Asparagus Rebus | 0 | 1 | 1 | 0 | 22 | 0.2 | 2.4 | 0.91 | 23 | 14 | 224 | 4.1 | 2 | 0 | 1.3 |
| Alpukat | 1 | 0 | 0 | 0 | 160 | 15 | 2 | 0.55 | 12 | 7 | 485 | 8.5 | 6.7 | 0 | 0.7 |
| Pisang | 1 | 0 | 0 | 0 | 89 | 0.3 | 1.1 | 0.26 | 5 | 1 | 358 | 23 | 2.6 | 0 | 12 |
| Roti Bagels | 0 | 1 | 1 | 0 | 250 | 1.5 | 10 | 2.76 | 20 | 439 | 165 | 4.1 | 0 | 0 | 6.1 |
| Buah Beri | 1 | 0 | 0 | 0 | 349 | 0.4 | 14 | 6.8 | 190 | 298 | 77 | 77 | 13 | 0 | 46 |
| ....... | ....... | ....... | ....... | ....... | ....... | ....... | ....... | ....... | ....... | ....... | ....... | ....... | ....... | ....... | ....... |

TABLE III
Exercise Dataset

| Date | Workout Name | Duration | Excersice Name | Set Order | Weight | Reps | Distance | Seconds | Notes | Workout Notes | RPE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 13/09/2021 | PUSH 1 | 1h 12m | Bench Press (Barbell) | 1 | 35.0 | 10 | 0 | 0 | | Focusing on CHEST, SHOULDERS and TRICEPS | 9 |
| 13/09/2021 | PUSH 1 | 1h 12m | Bench Press (Barbell) | 2 | 35.0 | 9 | 0 | 0 | | | 10 |
| 13/09/2021 | PUSH 1 | 1h 12m | Bench Press (Barbell) | 3 | 35.0 | 8 | 0 | 0 | | | 10 |
| 13/09/2021 | PUSH 1 | 1h 12m | Bench Press (Barbell) | 4 | 35.0 | 7 | 0 | 0 | | | 10 |
| 13/09/2021 | PUSH 1 | 1h 12m | Incline Cable Chest Fly | 1 | 12.5 | 12 | 0 | 0 | | | 9.5 |
| 13/09/2021 | PUSH 1 | 1h 12m | Incline Cable Chest Fly | 2 | 12.5 | 10 | 0 | 0 | | | 10 |
| 13/09/2021 | PUSH 1 | 1h 12m | Incline Cable Chest Fly | 3 | 12.5 | 8 | 0 | 0 | | | 10 |
| 13/09/2021 | PUSH 1 | 1h 12m | Incline Cable Chest Fly | 4 | 12.5 | 7 | 0 | 0 | | | 10 |
| 13/09/2021 | PUSH 1 | 1h 12m | Overhead Press (Dumbbel) | 1 | 12.0 | 10 | 0 | 0 | | | 8 |
| ....... | ....... | ....... | ....... | ....... | ....... | ....... | ....... | ....... | ....... | ....... | ....... |

### B. *Data Pre-processing*

In the next step, the data is separated based on mealtime categories, such as breakfast, lunch, and dinner. Each food item and its corresponding data index are stored in a separate list for each mealtime category. Then, the data is further processed in several steps, including the transpose method, which converts rows into columns. After the data is processed, the next step is calculating each user's Body Mass Index (BMI). This process determines the user's age category based on the predefined age ranges. Then, the BMI calculation is performed using the formula BMI = Weight / Height $(m)^2$. Based on the calculated BMI value, users are categorized into five predefined BMI categories: Underweight, Normal, Overweight, and Obese, based on BMI value ranges.

### C. *K-Means*

The labeling process in the diet and exercise recommendation system using the K-Means algorithm is carried out in the following steps. First, food data is separated based on mealtimes, such as breakfast, lunch, and dinner. Then, each corresponding data is stored in matrix form for each mealtime category. The K-Means process is then applied to these matrices with a predefined number of clusters, in this case, three clusters. The clustering results, which include cluster labels for each food data, are printed for each mealtime category. This process assigns cluster labels to each food category according to the K-Means results, allowing for grouping foods based on their calorie values. These cluster labels can then be used to recommend a balanced diet according to the user's nutritional needs.

### D. *Random Forest*

The process of predicting labels using the Random Forest algorithm involves several steps. First, a Random Forest Classifier or Regressor model is created by initializing a RandomForestClassifier or RandomForestRegressor object with specific parameters such as the number of trees (n_estimators). Next, the model is trained (fit) using training data (X_train and y_train). After the model is trained, test data (X_test) is provided for predictions. Predictions are made using the prediction method on the trained model object, resulting in predicted labels for each test data sample. The prediction results are then printed for evaluation, allowing an understanding of the model's performance and its fit with the given test data.

### E. *Evaluation*

The results of the generated recommendations will be evaluated using a specific method that can be measured based on the errors or discrepancies formed from the prediction results. The error value can be obtained using the Mean Absolute Percentage Error (MAPE) calculation metric. Mean Absolute Percentage Error (MAPE) is the average of the absolute percentage differences between actual data and predicted data. The lower the Mean Absolute Percentage Error (MAPE) value, the better the prediction model performance. The formula to calculate Mean Absolute Percentage Error (MAPE) is as follows:

$$MAPE = \frac{\sum_{t=1}^{n} \frac{|x_t - y_t|}{x_t}}{n} \times 100\% \quad (2)$$

$x_t$ = Actual value

$x_t$ = Predicted value

$n$ = Total number of data

Based on the Mean Absolute Percentage Error (MAPE) formula, which depicts the error value of the predictive model, the accuracy of the model can be obtained with the following calculation formula:

$$Accuracy = 100\% - MAPE \quad (3)$$

## IV. RESULTS AND DISCUSSION

### A. *Data Pre-processing*

In this process, data is grouped based on mealtime categories: breakfast, lunch, and dinner. Each food item and its corresponding data index are stored in separate lists for each mealtime category. After grouping the data according to mealtime categories, the data is further processed with several steps, one of which is using the transpose method to convert rows into columns. Rows of data are extracted based on predefined indices to retrieve a subset of data containing only the specified columns. This process removes columns for breakfast, lunch, dinner, and veg/non-veg. Then, the obtained subset of data is transposed back to its original form to facilitate further analysis. The result of the processed data is printed, providing food data based on meal categories and their nutritional values that can be used for further analysis. The result of the process will be displayed in the image below :

| | Food_items | Breakfast | Lunch | ... | Fibre | VitaminD | Sugars |
|---|---|---|---|---|---|---|---|
| 1 | Avocados | 1 | 0 | ... | 6.7 | 0 | 0.70 |
| 2 | Bananas | 1 | 0 | ... | 2.6 | 0 | 12.00 |
| 4 | Berries | 1 | 0 | ... | 13.0 | 0 | 46.00 |
| 8 | American cheese | 1 | 0 | ... | 0.0 | 0 | 0.00 |
| 9 | Coffee | 1 | 0 | ... | 0.0 | 0 | 0.00 |
| 10 | Corn | 1 | 1 | ... | 2.7 | 0 | 7.70 |
| 12 | Grapes | 1 | 0 | ... | 11.0 | 0 | 6.30 |
| 13 | Milk | 1 | 0 | ... | 0.0 | 0 | 0.00 |
| 14 | Cashew Nuts | 1 | 0 | ... | 3.3 | 0 | 5.90 |

Fig 5. Breakfast Food Data

| | 1 | 2 | 4 | ... | 81 | 82 | 87 |
|---|---|---|---|---|---|---|---|
| Food_items | Avocados | Bananas | Berries | ... | Banana Chips | Honey | Chocolate milk |
| Calories | 160 | 89 | 349 | ... | 519 | 304 | 535 |
| Fats | 15.0 | 0.3 | 0.4 | ... | 34.0 | 0.0 | 30.0 |
| Proteins | 2.0 | 1.1 | 14.0 | ... | 2.3 | 0.3 | 7.7 |
| Iron | 0.55 | 0.26 | 6.8 | ... | 1.25 | 0.42 | 2.35 |
| Calcium | 12.0 | 5.0 | 190.0 | ... | 18.0 | 6.0 | 189.0 |
| Sodium | 7 | 1 | 298 | ... | 6 | 4 | 79 |
| Potassium | 485.0 | 358.0 | 77.0 | ... | 536.0 | 52.0 | 372.0 |
| Carbohydrates | 8.5 | 23.0 | 77.0 | ... | 58.0 | 82.0 | 59.0 |
| Fibre | 6.7 | 2.6 | 13.0 | ... | 7.7 | 0.2 | 3.4 |
| VitaminD | 0 | 0 | 0 | ... | 0 | 0 | 0 |

Fig 6. Breakfast Food Data After Processing

| | Food_items | Breakfast | Lunch | ... | Fibre | VitaminD | Sugars |
|---|---|---|---|---|---|---|---|
| 0 | Asparagus Cooked | 0 | 1 | ... | 2.0 | 0 | 1.30 |
| 3 | Bagels made in wheat | 0 | 1 | ... | 4.1 | 0 | 6.10 |
| 5 | Brocolli | 0 | 1 | ... | 2.8 | 0 | 0.60 |
| 6 | Brown Rice | 0 | 1 | ... | 3.4 | 0 | 0.00 |
| 7 | Cauliflower | 0 | 1 | ... | 3.3 | 0 | 0.00 |
| 10 | Corn | 1 | 1 | ... | 2.7 | 0 | 7.70 |
| 15 | Onions | 0 | 1 | ... | 1.7 | 0 | 4.20 |
| 17 | Pasta canned with tomato sauce | 0 | 1 | ... | 0.9 | 0 | 4.00 |
| 19 | Peas | 0 | 1 | ... | 5.7 | 0 | 5.70 |
| 21 | Pumpkin | 0 | 1 | ... | 1.1 | 0 | 2.10 |

Fig 7. Lunch Food Data

| | 0 | ... | 79 |
|---|---|---|---|
| Food_items | Asparagus Cooked | ... | Oyster cooked |
| Calories | 22 | ... | 159 |
| Fats | 0.2 | ... | 4.0 |
| Proteins | 2.4 | ... | 29.0 |
| Iron | 0.91 | ... | 4.9 |
| Calcium | 23.0 | ... | 6.0 |
| Sodium | 14 | ... | 81 |
| Potassium | 224.0 | ... | 409.0 |
| Carbohydrates | 4.1 | ... | 0.0 |
| Fibre | 2.0 | ... | 0.0 |
| VitaminD | 0 | ... | 0 |

Fig 8. Lunch Food Data After Processing

| | Food_items | Breakfast | Lunch | ... | Fibre | VitaminD | Sugars |
|---|---|---|---|---|---|---|---|
| 0 | Asparagus Cooked | 0 | 1 | ... | 2.0 | 0 | 1.3 |
| 3 | Bagels made in wheat | 0 | 1 | ... | 4.1 | 0 | 6.1 |
| 5 | Brocolli | 0 | 1 | ... | 2.8 | 0 | 0.6 |
| 6 | Brown Rice | 0 | 1 | ... | 3.4 | 0 | 0.0 |
| 7 | Cauliflower | 0 | 1 | ... | 3.3 | 0 | 0.0 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 83 | Chocolate Icecream | 0 | 0 | ... | 1.2 | 0 | 25.0 |
| 84 | Vanilla Ice cream | 0 | 0 | ... | 0.7 | 0 | 21.0 |
| 85 | Strawberry Icecream | 0 | 0 | ... | 0.9 | 0 | 0.0 |
| 86 | Marshmallows | 0 | 0 | ... | 0.1 | 0 | 58.0 |
| 88 | Rice Pudding | 0 | 0 | ... | 0.7 | 0 | 0.0 |

Fig 9. Dinner Food Data

| | 0 | ... | 88 |
|---|---|---|---|
| Food_items | Asparagus Cooked | ... | Rice Pudding |
| Calories | 22 | ... | 376 |
| Fats | 0.2 | ... | 0.1 |
| Proteins | 2.4 | ... | 2.7 |
| Iron | 0.91 | ... | 1.79 |
| Calcium | 23.0 | ... | 14.0 |
| Sodium | 14 | ... | 366 |
| Potassium | 224.0 | ... | 5.0 |
| Carbohydrates | 4.1 | ... | 91.0 |
| Fibre | 2.0 | ... | 0.7 |
| VitaminD | 0 | ... | 0 |

Fig 10. Dinner Food Data After Processing

### *B. K-Means*

In this process, food data is separated based on mealtimes, namely breakfast, lunch, and dinner, and then processed to obtain calorie data. K-Means model with 3 clusters representing low, medium, and high-calorie foods is applied to the calorie data. K-Means clustering is performed to group food calorie data into 3 clusters in this process. Cluster 0 is marked with the number 0 and contains relatively low-calorie data, Cluster 1 is marked with the number 1 and contains medium-calorie data, while Cluster 2 is marked with the number 2 and contains high-calorie data. Breakfast, lunch, and dinner foods produce cluster label results through the K-Means clustering. The results of this K-Means process are shown in Fig 11.

```
Breakfast
[0 1 2 0 1 0 0 0 0 0 0 1 1 0 1 1 0 0 1 1 1 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 1
 2 0 0 0]
Lunch
[2 0 0 0 0 0 2 0 0 2 2 1 1 1 2 2 0 0 2 0 0 0 2 0 0 0 0 0 0 2 0 0 0 0 0 0 0
 0 1 2 2 1 0 0]
Dinner
[2 0 0 0 0 0 0 0 2 0 0 2 0 2 2 2 1 2 2 2 2 2 2 0 0 2 0 0 0 2 0 0 0 0 0 2 0
 0 0 0 0 0 1 0 2 1 2 2 0 1 2 1 0 0 0 0 0 0 0 2]
```

Fig 11. K-Means Results

### *C. Random Forest*

In the Random Forest process for diet planning, the first step is to read the nutrition distribution dataset and transform it into a data frame, which is then transposed so that each row represents a type of food and each column represents a nutritional attribute such as calories, protein, and fat. This nutrition data is used to develop a model to calculate the recommended food nutrition values and ensure they align with the user's nutritional needs. The data is then grouped based on BMI and user age categories and transformed into an array where each row represents a type of food, and each column represents a nutritional attribute. This process is carried out for each BMI and age category to generate feature data (X_train) and label data (y_train) used for model training. Afterward, the data is split into training and testing datasets. A Random Forest Classifier model with 100 estimators is used to train the training data and produce a model capable of predicting food recommendations. Once the model is trained, the testing data is processed using the Random Forest model to generate suitable predictions for low-calorie foods.

```
Avocados
Bagels made in wheat
Berries
Brocolli
Cauliflower
```

Fig 12. Random Forest Diet Sample Results

In the Random Forest process for physical training, the first step is to read and convert the physical training data into a data frame, where features such as weight, set order, and number of repetitions are used as variables X. In contrast, the physical training labels are stored in variable y and converted into numeric form using LabelEncoder. The data is then split into training and testing datasets. The RandomForestRegressor model trains the training data and generates predictions based on the testing data. After training, the model processes the testing data and produces predictions for physical training, including information on set order and number of repetitions based on weight.

```
Best exercise: Chest Fly (Dumbbell), Set Order: 1, Reps: 8
Other Recommendation :Bench Press (Barbell), Set Order: 1, Reps: 10
```

Fig 13. Random Forest Physical Exercise Sample Result

*D. Evaluation*

In this step, the evaluation result of accuracy measurement using MAPE (Mean Absolute Percentage Error) is shown in the table below.

TABLE IV
Evaluation Result

| Mealtime Category | Veg/Non Veg | MAPE Diet | Diet Accuracy | MAPE Physical Exercise | Physical Exercise Accuracy |
|---|---|---|---|---|---|
| Breakfast | Vegan | 29.26% | 70.74% | 26.35% | 73.65% |
| Breakfast | NonVegan | 7.80% | 92.20% | 26.35% | 73.65% |
| Lunch | Vegan | 9.75% | 90.25% | 26.35% | 73.65% |
| Lunch | NonVegan | 0.97% | 99.03% | 26.35% | 73.65% |
| Dinner | Vegan | 6.34% | 93.66% | 26.35% | 73.65% |
| Dinner | NonVegan | 8.29% | 91.71% | 26.35% | 73.65% |

Table IV shows the evaluation results of accuracy measurement using MAPE (Mean Absolute Percentage Error) for diet and exercise recommendations based on user input. The results indicate that the diet recommendations for Vegans have a higher MAPE than Non-Vegans across all mealtimes. The highest accuracy for diet recommendations is for Non-vegetarians at lunchtime, at 99.03%, while the lowest accuracy is for Vegans in the morning, at 70.74%. The MAPE for exercise recommendations is consistently at 26.35%, and the accuracy for exercise recommendations is stable at 73.65%, indicating relatively stable accuracy for exercise recommendations. Overall, this system shows fairly good accuracy in recommending diets and physical exercises.

### E. System Testing Results

In this process, the system is developed using the Python programming language. After the model is formed, a user interface is created to connect the user with the ongoing process. In Fig 14, the homepage of the diet and physical training recommendation system is Displayed.

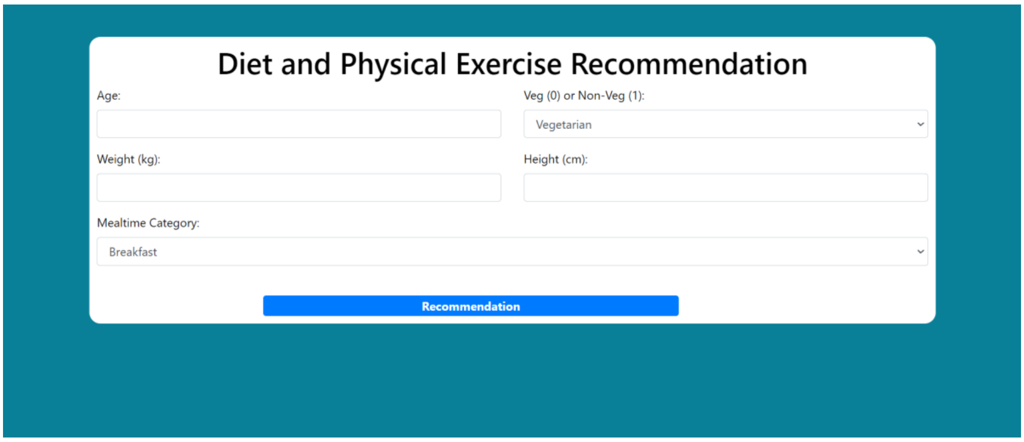


Fig 14. Input Display of Diet and Physical Exercise Recommendation System

In the system interface shown in Fig 14, users can input their personal information, such as age, weight, vegan status, height, and desired mealtime category, for the recommendation process. After that, the user will receive recommendations, as shown in Fig 15. Here is the display of the diet and physical training recommendation results page according to the user input, showing the user-provided information, diet recommendations, and physical training recommendations.

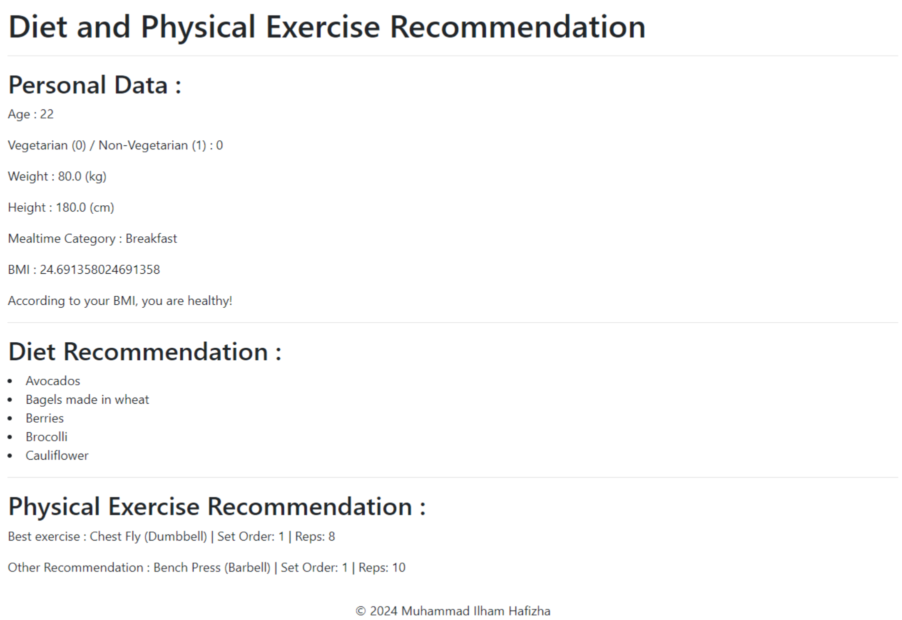


Fig 15. Display of Diet and Physical Exercise Recommendation Results

The system can benefit society by offering personalized diet and exercise recommendations tailored to individual needs, promoting healthier lifestyles, and preventing obesity-related health problems. Its advanced algorithms provide accurate and data-driven insights, making health advice more accessible and practical. The system supports holistic health and accommodates dietary preferences. Overall, the system encourages healthier and more active communities, improving public health outcomes.

## V. CONCLUSION

In this study, we developed an integrated diet and physical exercise recommendation system using a combination of K-Means and Random Forest algorithms. The system consists of several stages, including data preparation from public datasets, data processing by separating them based on mealtime categories, Body Mass Index calculation, K-Means for clustering food data based on calories, Random Forest for recommending diet and exercise plans, and evaluating accuracy using Mean Absolute Percentage Error (MAPE). The test results indicate that the system has fairly good accuracy in recommending diet and exercise plans, with the diet recommendation accuracy for Non-Vegans higher than Vegans at all meal times and the exercise recommendation accuracy relatively stable in all categories with a MAPE of 26.35% and an accuracy of 73.65%. For further research, this system must be continued to improve its accuracy and usability by expanding the dataset and enhancing the user interface and experience. With ongoing development, this system is expected to provide greater benefits to the community.

## ACKNOWLEDGMENT

The author would like to thank Telkom University for their support and the opportunity to conduct this research. We also extend our thanks to previous researchers who have inspired the development of the concept and methodology of this research. Additionally, we would like to convey our utmost gratitude to all parties involved in supporting this research, both physically and intellectually.

## REFERENCES

[1] P. Widya Hapsari, “The Application of UNICEF'S 2020 Conceptual Framework of Maternal and Child Nutrition in Indonesia.” [Online]. Available: https://www.researchgate.net/publication/349161215.

[2] "Obesity and overweight." World Health Organization. Accessed September 12, 2015. http://www.who.int/mediacentre/factsheets/fs311/en/.

[3] Kardam, S. S., Yadav, P., Thakkar, R., & Ingle, A. (2021). Website on diet recommendation using machine learning. Int Res J Eng Technol (IRJET), 2021, 3708-3711.

[4] Golagana, R., Sravani, V., Reddy, T. M., & Kavitha, C. H. (2023). Diet Recommendation System Using Machine Learning. Journal of Health and Nutrition, 15(2), 233-245.

[5] Shah, J., Yadav, V., Pandey, S., Patel, Y., & Patil, M. S. (2023). Sports Physical Training and Diet Plan. International Journal of Advances in Engineering and Management (IJAEM), 5(4), 857-860.

[6] Navastara, D. A., Mursidah, E., Gonti, Y. A., Wahyuni, D., Wiyadi, P. D. S., & Suadi, W. (2018). Clustering Topik Penelitian Berbasis Unsupervised Learning Untuk Rekomendasi Koleksi Pustaka Di Perpustakaan Its. Jurnal Ilmiah Teknologi Informasi, 17(2), 125-134.

[7] Jaja, V. L., Susanto, B., & Sasongko, L. R. (2020). Penerapan Metode Item-Based Collaborative Filtering Untuk Sistem Rekomendasi Data MovieLens. d'CARTESIAN: Jurnal Matematika dan Aplikasi, 9(2), 78-83.

[8] Zayyad, M. R. A., & Kurniawardhani, A. (2021). Penerapan Metode Deep Learning pada Sistem Rekomendasi Film. AUTOMATA, 2(1).

[9] Fadlil, J., & Mahmudy, W. F. (2007). Pembuatan Sistem Rekomendasi Menggunakan Decision Tree dan Clustering. Jurnal Ilmiah Kursor, 3(1).

[10] Shah, M., Degadwala, S., & Vyas, D. (2022, February). Diet recommendation system based on different machine learners: A review. In 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS) (pp. 290-295). IEEE.

[11] Licenziati, M. R., Ballarin, G., Iannuzzo, G., Lonardo, M. S., Di Vincenzo, O., Iannuzzi, A., & Valerio, G. (2022). A height-weight formula to measure body fat in childhood obesity. Italian Journal of Pediatrics, 48(1), 106.

[12] Lim, J. U., Lee, J. H., Kim, J. S., Hwang, Y. I., Kim, T. H., Lim, S. Y., ... & Rhee, C. K. (2017). Comparison of World Health Organization and Asia-Pacific body mass index classifications in COPD patients. International journal of chronic obstructive pulmonary disease, 2465-2475.

[13] Phanich, M., Pholkul, P., & Phimoltares, S. (2010, April). Food recommendation system using clustering analysis for diabetic patients. In 2010 international conference on information science and applications (pp. 1-8). IEEE.

[14] Yang, Y., & Koenigstorfer, J. (2020). Determinants of physical activity maintenance during the Covid-19 pandemic: a focus on fitness apps. Translational behavioral medicine, 10(4), 835-842.

[15] Toledo, R. Y., Alzahrani, A. A., & Martinez, L. (2019). A food recommender system considering nutritional information and user preferences. IEEE Access, 7, 96695-96711.

[16] Abdulaziz, M., Al-motairy, B., Al-ghamdi, M., & Al-qahtani, N. (2021). Building a personalized fitness recommendation application based on sequential information. International Journal of Advanced Computer Science and Applications, 12(1).

[17] Singh, S., Phulphagar, D., Shelke, P., Khan, S., & Sonawane, K. USER-TO-USER BASED RECOMMENDATION SYSTEM FOR FITNESS FREAKS.

[18] Ezin, E., Kim, E., & Carrascosa, I. P. (2018, October). 'Fitness that Fits':-A Prototype Model for Workout Video Recommendation. In 12th ACM Conference on Recommender Systems.

[19] Hemaraju, S., Kaloor, P. M., & Arasu, K. (2023, May). Yourcare: A diet and fitness recommendation system using machine learning algorithms. In AIP Conference Proceedings (Vol. 2655, No. 1). AIP Publishing.

[20] Gouthami, B., & Gangappa, M. (2020). Nutrition diet recommendation system using user's Interest. International Journal of Advanced Research in Engineering and Technology, 11(12), 2910-2919.

[21] Cho, J., & Kim, S. (2020). Personal and social predictors of use and non-use of fitness/diet app: Application of Random Forest algorithm. Telematics and Informatics, 55, 101301.

[22] Zhao, Z., Arya, A., Orji, R., & Chan, G. (2020, August). Physical activity recommendation for exergame player modeling using machine learning approach. In 2020 IEEE 8th International Conference on Serious Games and Applications for Health (SeGAH) (pp. 1-9). IEEE.

[23] Chen, Q., & Lee, S. (2021). A machine learning approach to predict customer usage of a home workout platform. Applied Sciences, 11(21), 9927.

[24] Tran, T. N. T., Felfernig, A., Trattner, C., & Holzinger, A. (2021). Recommender systems in the healthcare domain: state-of-the-art and research issues. Journal of Intelligent Information Systems, 57(1), 171-201.

[25] Venkatachalam, P., & Ray, S. (2022). How do context-aware artificial intelligence algorithms used in fitness recommender systems? A literature review and research agenda. International Journal of Information Management Data Insights, 2(2), 100139.

[26] Aquino, J. R. V., Barrientos, L. J. M., Coloma, M. C., Lacar, L. V. S., & Acang, J. P. A. (2017). Data Analytics Enforced Content-Based Fitness Recommender System for Individuals without Allergies and Ailments. Journal of Software Engineering: Theories and Practices, 2(1), 11-16.

[27] Mogaveera, D., Mathur, V., & Waghela, S. (2021, January). e-Health monitoring system with diet and fitness recommendation using machine learning. In 2021 6th International Conference on Inventive Computation Technologies (ICICT) (pp. 694-700). IEEE.

[28] Ramyaa, R., Hosseini, O., Krishnan, G. P., & Krishnan, S. (2019). Phenotyping women based on dietary macronutrients, physical activity, and body weight using machine learning tools. Nutrients, 11(7), 1681.

[29] Nagaraj, P., Muneeswaran, V., & Deshik, G. (2022, August). Ensemble Machine Learning (Grid Search & Random Forest) based Enhanced Medical Expert Recommendation System for Diabetes Mellitus Prediction. In 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 757-765). IEEE.

[30] Nagaraj, P., Muneeswaran, V., Dharanidharan, A., Balananthanan, K., Arunkumar, M., & Rajkumar, C. (2022, April). A prediction and recommendation system for diabetes mellitus using XAI-based lime explainer. In 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) (pp. 1472-1478). IEEE.