IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

AI-Based Personalized Diet Recommendation Using Machine Learning

Siddhi Kumbhar[1] and Ms. Shevti Chandan[2]
[1]Student, Master of Computer Applications
[2]Assistant Professor, Master of Computer Applications
Sadhu Vaswani Institute of Management Studies for Girls, Pune
Savitribai Phule Pune University, Pune, Maharashtra, India

Abstract: *Personalizing diet plans considering the age group, lifestyle and physical activity, meals, food and diet preferences, and diet goals is challenging. This paper proposes a machine-learning based personalized diet recommendation framework, which employs a Random Forest classifier to predict the diet goal of the user followed by retrieving of suitable foods corresponding to the predicted diet goal. The study utilizes a primary questionnaire dataset of 50 records and eight user-related features, namely Age Group, Gender, Occupation, Physical Activity, Meals per Day, Outside Food, Diet Type and Preferred Food Categories. The target variable consists of three classes, namely Weight Loss, Maintain Weight and Weight Gain. The categorical variables underwent one-hot encoding followed by the 80:20 stratified split to train and test an ensemble of 150 trees with max depth 8. On the 10 record test set, the model achieved 60.00% accuracy, 61.67% weighted precision, 60.00% weighted recall and 60.00% weighted F1-score. The prediction model was further explained using SHAP (SHapley Additive exPlanations) for interpretability. A Streamlit-powered research dashboard was developed to demonstrate the machine-learning model-development process and display the predicted diet goal along with the SHAP values. The research is an academic prototype and should not be recommended as a substitute for professional nutritionist assistance*

Keywords: Personalized Diet; Machine Learning; Random Forest; Diet Goal Prediction; SHAP; Explainable AI; Food Recommendation; Nutrition Dataset

I. INTRODUCTION

Personalized nutrition focuses on providing dietary suggestions according to an individual's characteristics instead of applying the same diet plan to everyone. Factors such as physical activity, meal frequency, outside food consumption, dietary preferences, and food choices can influence an individual's dietary requirements.

Machine learning can be used to discover patterns in such user information and estimate an appropriate diet goal. In this research, Random Forest is selected because it combines several decision trees and can handle relationships among different structured input variables.

Prediction alone does not explain why a particular result is produced. Therefore, SHAP is included to examine the contribution of input features to the model's decision. This improves the interpretability of the proposed approach.

The proposed framework follows two main stages. First, questionnaire information is processed by the machine learning model to predict one of three diet goals: weight loss, maintain weight, or weight gain. Second, the predicted goal is used to retrieve suitable food items from a secondary nutrition dataset. A Streamlit interface brings these components together by displaying the prediction, food recommendations, and explanation information.

II. OBJECTIVES OF THE STUDY

The study aims to:

• collect and prepare questionnaire-based primary data for personalized diet prediction;

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, October 2026 Impact Factor: 8.2

- identify useful user characteristics for predicting diet goals;
- develop a Random Forest model for three diet-goal categories;
- evaluate the model using accuracy, precision, recall, F1-score, and confusion matrix;
- apply SHAP to interpret model behavior at both overall and individual levels;
- connect the predicted goal with a secondary nutrition dataset;
- provide suitable food suggestions through an integrated research dashboard.

III. LITERATURE REVIEW

3.1 Personalized Diet Recommendation System Using Machine Learning, 2024

This research applies content-based filtering for generating food and recipe recommendations according to user preferences. Nearest-neighbor techniques and cosine similarity are used to identify similar food items. In contrast, the present study treats diet goal prediction as a supervised classification problem and uses Random Forest for this purpose.

3.2 Unveiling Neural Networks for Personalized Diet Recommendations, 2024

This work investigates neural network approaches for personalized dietary recommendations. It also considers a secondary error-prediction model based on food-logging information. The present research follows a simpler approach using questionnaire data and focuses on an interpretable Random Forest model rather than a neural network architecture.

3.3 Diet Recommendation System Using K-Means Clustering Algorithm of Machine Learning, 2024

The study applies K-Means clustering to dietary preference information and combines the resulting groups with a nutritional food database. Although both studies use machine learning for dietary recommendation, the proposed work uses supervised learning to predict predefined diet goals, whereas K-Means follows an unsupervised clustering approach.

3.4 SousChef System for Personalized Meal Recommendations: A Validation Study, 2022

This study presents a personalized meal recommendation approach based on recipes and optimization techniques. Its findings demonstrate the value of tailoring meal suggestions to individual requirements. The proposed research extends this concept by combining diet-goal prediction, food retrieval, and model interpretation within one prototype.

3.5 Systematic Review of Nutritional Recommendation Systems, 2021

This review examines existing nutritional recommendation systems and summarizes major research directions and challenges. It highlights the importance of personalization, suitable data, and effective recommendation techniques. The present work addresses these requirements through a structured questionnaire dataset, machine learning prediction, nutrition data, and explainable AI.

IV. RESEARCH GAP

Existing studies have investigated several techniques for dietary recommendation, including content-based recommendation, similarity-based methods, K-Means clustering, recipe optimization, and neural network approaches. However, fewer simple academic prototypes combine diet-goal classification, food recommendation, and prediction explanation in one workflow.

To address this gap, the proposed framework uses Random Forest to classify the user's diet goal and then connects the predicted category with a structured nutrition dataset. SHAP is additionally used to make the model's output easier to interpret. Thus, the proposed approach combines prediction, recommendation, and explainability in a single research-oriented framework.

V. RESEARCH METHODOLOGY

The research methodology is structured around the following steps: data collection; data quality assessment; feature selection; data preparation; train-test split; model training; SHAP explanation; predicted diet goal; secondary nutrition-

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, October 2026 Impact Factor: 8.2

data matching; dashboard display. The methodology follows the organization of the study through introduction, literature review, research gap, methodology, model-development findings, interpretation, limitations, and conclusion.

5.1 Research Design

The study implements a quantitative, machine-learning-based experimental design, where the primary questionnaire responses are used as the modelling data, and the secondary nutrition dataset serves as the recommendation data. The model's performance is validated on an unseen test set, whereas the SHAP explainer is applied to the trained model to provide the interpretation.

5.2 Data Collection and Feature Selection

The primary data were collected through a structured questionnaire related to dietary habits, lifestyle, food preferences, and diet goals. The final dataset contains 50 records and 18 questionnaire fields. From these fields, eight features were selected for the Random Forest model: Age Group, Gender, Occupation, Physical Activity, Meals per Day, Outside Food, Diet Type, and Preferred Food Categories. The target variable is Diet Goal, comprising three categories: Weight Loss, Maintain Weight, and Weight Gain.

No.	Model Feature	Role
1	Age Group	Input
2	Gender	Input
3	Occupation	Input
4	Physical Activity	Input
5	Meals per Day	Input
6	Outside Food	Input
7	Diet Type	Input
8	Preferred Food Categories	Input
9	Diet Goal	Target

Table 1. Final features used for Random Forest diet-goal prediction

Some questionnaire fields, such as Most Important Meal, Preferred Cuisine, Meal Plan Preference, Diet Planning Difficulty, Usefulness Rating, and Suggestions, were not included in the final prediction model since they were not part of the selected model feature set.

Target class distribution in the 50-record primary dataset: Weight Loss = 20, Maintain Weight = 15, Weight Gain = 15.

Diet Goal	Records	Percentage
Weight Loss	20	40%
Maintain Weight	15	30%
Weight Gain	15	30%
Total	50	100%

Table 1A. Target class distribution

5.3 Data Preparation

Before model training, the collected data were checked for consistency and suitable data types. Most of the selected variables were categorical, so one-hot encoding was used to convert them into numerical form. Meals per Day was treated as a numerical feature. The preprocessing operation was included in the machine learning pipeline so that the same transformation could also be applied when a new user enters information through the dashboard.

The final primary dataset contains 50 records. The current dataset does not contain Height, Weight, or Body Mass Index (BMI) as model inputs, so BMI was not included in the final experiment and no BMI values were artificially created. BMI can be considered in future research if reliable height and weight information is collected.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

5.4 Train-Test Split

To evaluate the performance of the developed model, the dataset was divided into training and testing subsets. An 80:20 stratified split was used with random state 42. As a result, 40 records were used for training and 10 records were kept for testing. Stratification was applied to maintain the representation of the three diet-goal classes in both subsets.

Item	Value
Total records	50
Training records	40
Testing records	10
Split	80:20 stratified
Random state	42
Target classes	Weight Loss / Maintain Weight / Weight Gain

Table 2. Dataset split and target configuration

5.5 Feature Engineering and Model Development

The selected questionnaire variables were transformed into a model-ready format before training. Categorical features were converted using one-hot encoding, while Meals per Day was retained as a numerical feature. The resulting feature matrix was provided to a Random Forest classifier.

The Random Forest model was configured with 150 decision trees and a maximum depth of 8. Balanced class weighting was used to provide appropriate consideration to the different target classes. The random state was fixed at 42 to make the experiment reproducible.

Parameter	Final Setting
Algorithm	Random Forest Classifier
Trees	150
Maximum Depth	8
Class Weighting	Balanced
Random State	42
Preprocessing	One-hot encoding + numeric feature
Prediction Target	Diet Goal

Table 3. Final Random Forest model configuration

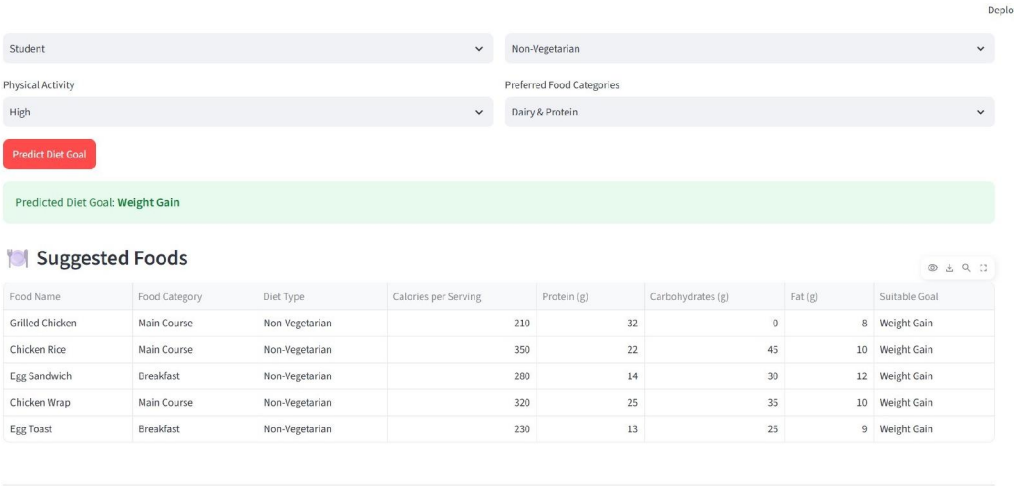
Feature	Data Type	Preprocessing	Role
Age Group	Categorical	One-hot encoding	Input
Gender	Categorical	One-hot encoding	Input
Occupation	Categorical	One-hot encoding	Input
Physical Activity	Categorical	One-hot encoding	Input
Meals per Day	Numerical	Numeric	Input
Outside Food	Categorical	One-hot encoding	Input
Diet Type	Categorical	One-hot encoding	Input
Preferred Food Categories	Categorical	One-hot encoding	Input
Diet Goal	Categorical	Target label	Output

Table 3A. Feature engineering matrix

The trained model was integrated into the Streamlit research dashboard. The dashboard accepts the same eight input features used during model training and applies the saved preprocessing pipeline before generating the prediction. After the Diet Goal is predicted, the system retrieves suitable food records from the secondary nutrition dataset. The

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

dashboard also includes the SHAP-based explanation so that prediction, recommendation, and interpretation can be demonstrated together.



Food Name	Food Category	Diet Type	Calories per Serving	Protein (g)	Carbohydrates (g)	Fat (g)	Suitable Goal
Grilled Chicken	Main Course	Non Vegetarian	210	32	0	8	Weight Gain
Chicken Rice	Main Course	Non-Vegetarian	350	22	45	10	Weight Gain
Egg Sandwich	Breakfast	Non-Vegetarian	280	14	30	12	Weight Gain
Chicken Wrap	Main Course	Non-Vegetarian	320	25	35	10	Weight Gain
Egg Toast	Breakfast	Non-Vegetarian	230	13	25	9	Weight Gain

Figure 1. Integrated research dashboard showing model input, predicted diet goal and suitable food recommendations.

VI. SHAP-BASED EXPLAINABLE MODEL INTERPRETATION

SHAP, or SHapley Additive exPlanations, was used to make the Random Forest predictions easier to understand. Instead of showing the final Diet Goal, SHAP helps to identify how the different input features influenced the model's decision.

The SHAP explainer was applied to the trained Random Forest model after the categorical variables had been encoded. Since one-hot encoding creates several transformed variables from categorical features, the corresponding SHAP contributions were considered together to understand the influence of the original questionnaire features.

Global SHAP analysis provides an overview of which features have greater influence on the model's predictions. In the present experiment, Physical Activity, Preferred Food Categories, and Meals per Day showed comparatively higher aggregate importance, followed by Diet Type and Outside Food.

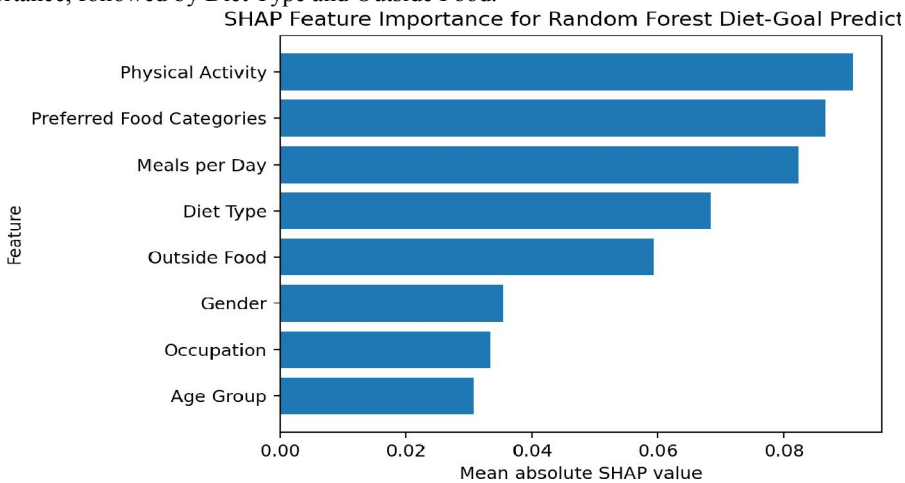


Figure 2. Global SHAP feature importance for the Random Forest diet-goal model.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

6.1 Individual SHAP Prediction

An individual prediction was also examined to demonstrate how SHAP can explain a specific model output. The selected example had the following characteristics: Age Group 18–25, Female, Student, High Physical Activity, 3 Meals per Day, Often Outside Food, Non-Vegetarian Diet Type, and Dairy & Protein as the preferred food category.

For this input combination, the model predicted Weight Gain. The approximate class probabilities were 19.89% for Weight Loss, 44.36% for Weight Gain, and 35.76% for Maintain Weight. Since Weight Gain had the highest probability, it was selected as the final prediction.

The local SHAP explanation provides a feature-level view of the factors contributing to this particular prediction.

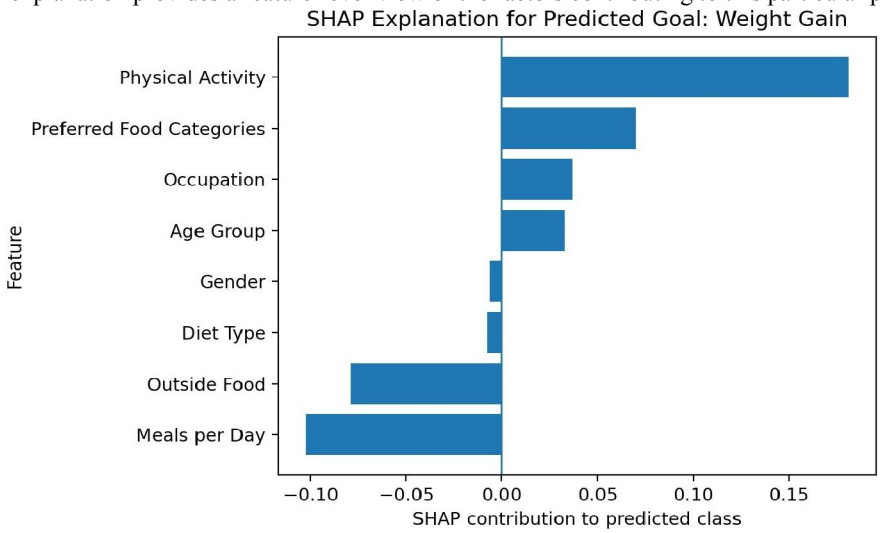


Figure 3. Local SHAP explanation for the demonstrated Weight Gain prediction.

VII. RECOMMENDATION PROCESS

After the Random Forest model predicts the user's Diet Goal, the recommendation stage uses the secondary nutrition dataset to retrieve suitable food items. The nutrition dataset contains information such as Food Name, Food Category, Diet Type, Calories per Serving, Protein, Carbohydrates, Fat, and Suitable Goal.

The predicted Diet Goal is matched with the Suitable Goal field in the secondary dataset. The matching food records are then displayed through the dashboard. This approach separates the prediction task from the food-retrieval task and allows the recommendation component to use structured nutritional information.

Field	Purpose
Food Name	Identifies the recommended food item
Food Category	Groups food items
Diet Type	Checks dietary compatibility
Calories per Serving	Provides energy information
Protein	Provides protein information
Carbohydrates	Provides carbohydrate information
Fat	Provides fat information
Suitable Goal	Matches the predicted Diet Goal

Table 4A. Secondary nutrition dataset matrix

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

VIII. MODEL EVALUATION AND RESULTS

The developed Random Forest model was evaluated using the 10-record test set. The model achieved 60.00% accuracy, 61.67% weighted precision, 60.00% weighted recall, and 60.00% weighted F1-score.

Metric	Score	Meaning
Accuracy	60.00%	Overall proportion of correct predictions
Weighted Precision	61.67%	Weighted correctness of predicted classes
Weighted Recall	60.00%	Weighted proportion of actual classes identified
Weighted F1-Score	60.00%	Weighted balance of precision and recall

Table 4. Random Forest performance on the 10-record test set

8.1 Class-wise Performance

The class-wise results show that the model performed differently on the three Diet Goal categories. Weight Gain achieved a precision, recall, and F1-score of 0.67. Maintain Weight had a precision of 0.50, recall of 0.67, and F1-score of 0.57. Weight Loss obtained a precision of 0.67, recall of 0.50, and F1-score of 0.57.

Class	Precision	Recall	F1-Score	Support
Maintain Weight	0.50	0.67	0.57	3
Weight Gain	0.67	0.67	0.67	3
Weight Loss	0.67	0.50	0.57	4

Table 5. Class-wise classification report

8.2 Confusion Matrix

Actual / Predicted	Maintain Weight	Weight Gain	Weight Loss
Maintain Weight	2	1	0
Weight Gain	0	2	1
Weight Loss	2	0	2

Table 6. Confusion matrix for the 10-record hold-out test set

The model correctly classified two of the three Maintain Weight cases, two of the three Weight Gain cases, and two of the four Weight Loss cases. The remaining cases were classified into another diet-goal category. These variations are expectable to some extent because the primary dataset contains only 50 records and the test set only 10 records.

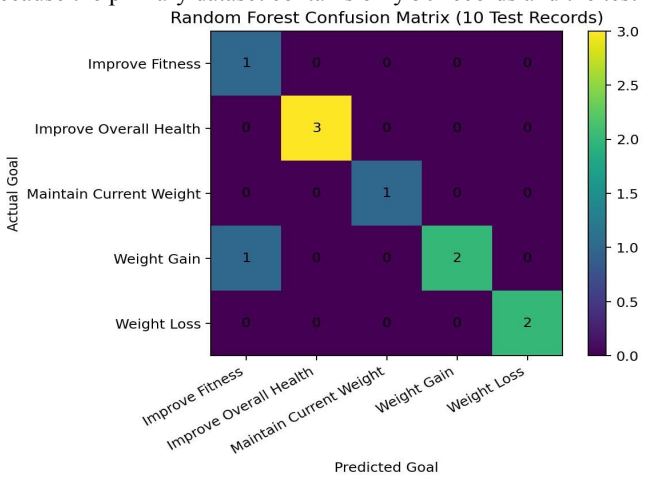


Figure 4. Confusion matrix visualization.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

Actual Class	Correct Predictions	Misclassified Cases	Interpretation
Maintain Weight	2 of 3	1	One case predicted as Weight Gain
Weight Gain	2 of 3	1	One case predicted as Weight Loss
Weight Loss	2 of 4	2	Two cases predicted as Maintain Weight

Table 7. Confusion matrix interpretation

8.3 Feature Importance and SHAP Findings

Two approaches were considered for understanding feature influence: Random Forest feature importance and SHAP-based importance. Random Forest feature importance provides an importance score based on the decision trees, while SHAP gives a more detailed explanation of how features contribute to individual predictions.

For this research, SHAP is particularly useful since it provides both an overall view of feature influence and an explanation for individual predictions. This makes the model more understandable than presenting only the final predicted Diet Goal.

Figure 5. SHAP-based feature importance for the Random Forest diet-goal model.

9. Proposed Personalized Diet Recommendation Workflow

The proposed framework begins with collecting questionnaire information from the user. The data are then checked and transformed into a suitable format for machine learning. After feature encoding, the Random Forest model predicts the user's Diet Goal. SHAP is then used to explain the prediction. The predicted goal is matched with the secondary nutrition dataset, and suitable food records are retrieved and displayed through the dashboard.

Primary Questionnaire Data → Data Preparation → Feature Encoding → Random Forest Prediction → SHAP Explanation → Predicted Diet Goal → Secondary Nutrition Dataset → Suitable Food Recommendations → Integrated Dashboard

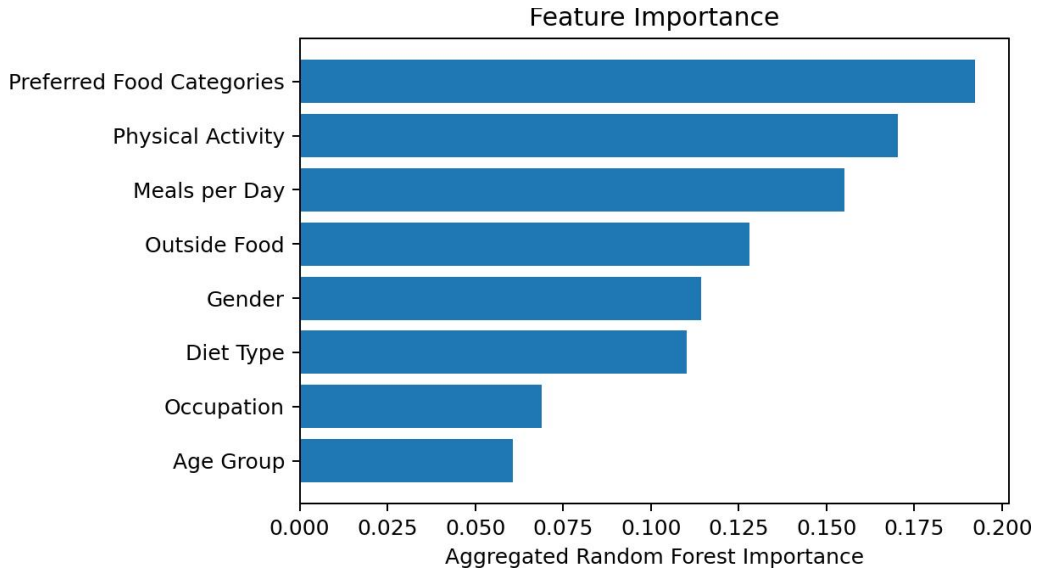


Figure 6. Proposed personalized diet recommendation workflow.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

X. FINDINGS AND DISCUSSION

The developed prototype demonstrates that questionnaire-based lifestyle and dietary information can be used to predict a broad diet goal using a Random Forest classifier. The model successfully handled the three target categories: Weight Loss, Maintain Weight, and Weight Gain.

The model achieved 60.00% accuracy on the 10-record test set. Although this result is useful for demonstrating the feasibility of the proposed approach, it should not be interpreted as a general performance measure for larger populations because the dataset is small.

The use of SHAP adds an explainability layer to the prediction process. Instead of presenting only the predicted Diet Goal, the system can also show the contribution of individual features. This improves the transparency of the research prototype.

The secondary nutrition dataset extends the system beyond prediction by providing suitable food records according to the predicted goal. The integrated dashboard brings these components together and demonstrates the complete workflow from user input to prediction, explanation, and food retrieval.

XI. LIMITATIONS

The present study has several limitations. First, the primary dataset contains only 50 records, with only 10 records used for final testing. Therefore, the results may change when the model is trained and evaluated on a larger and more diverse dataset.

The dataset may not fully represent different age groups, occupations, lifestyles, and dietary patterns. The current experiment also does not use Height, Weight, or BMI as model inputs.

The predicted Diet Goal represents a generalized machine-learning classification and should not be treated as a medical or clinical prescription. Similarly, SHAP explains the relationships learned by the model but does not prove that a particular feature causes a specific diet goal.

The food suggestions are retrieved from the secondary nutrition dataset and should therefore not be considered professionally prescribed meal plans. Finally, the model has not yet been validated using an independent external dataset.

XII. FUTURE SCOPE

Future research can improve the proposed system by collecting a larger number of questionnaire responses from a more diverse population. Additional validated nutritional variables such as Height, Weight, and BMI may also be included if reliable data are available.

The model can be further evaluated using k-fold cross-validation and external validation. Other machine-learning and recommendation algorithms can be compared with Random Forest to identify more suitable approaches.

The recommendation component can also be improved by considering nutritional factors, dietary restrictions, and a larger professionally validated food database.

XIII. CONCLUSION

This research presented an AI-Based Personalized Diet Recommendation framework using Machine Learning. The proposed system uses questionnaire-based lifestyle and dietary information to predict one of three broad diet goals: Weight Loss, Maintain Weight, and Weight Gain.

A Random Forest classifier was developed using eight selected input features. The final experiment used 50 primary records, with 40 records for training and 10 records for testing. The model achieved 60.00% accuracy, 61.67% weighted precision, 60.00% weighted recall, and 60.00% weighted F1-score.

SHAP was incorporated to improve the interpretability of the model and provide both global and individual prediction explanations. A secondary nutrition dataset was then connected with the predicted diet goal to retrieve suitable food records.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

The Streamlit research dashboard demonstrates the complete workflow in an integrated interface. Although the current results are useful as an academic prototype, the small dataset limits generalization. Therefore, larger datasets, additional validated nutritional variables, cross-validation, external validation, and professionally reviewed nutrition information should be considered in future research.

Overall, the study demonstrates how machine learning, explainable AI, and structured nutrition data can be combined to create a basic personalized diet recommendation framework. The system is intended for academic and research demonstration and should not replace advice from qualified nutrition professionals.

REFERENCES

[1] "Personalized Diet Recommendation System Using Machine Learning," 2024.
[2] "Unveiling Neural Networks for Personalized Diet Recommendations," 2024.
[3] "Diet Recommendation System Using K-Means Clustering Algorithm of Machine Learning," 2024.
[4] "SousChef System for Personalized Meal Recommendations: A Validation Study," 2022.
[5] S. A. de la Torre et al., "Systematic Review of Nutritional Recommendation Systems," Applied Sciences, vol. 11, no. 24, 12069, 2021