A Food Recommender System Considering User Preferences and Medical Information(Allergens)

Molly Priya(2022UCA1869) Computer Science and Engineering in AI

ABSTRACT The World Health Organization identifies the overall increasing of noncommunicable diseases as a major issue, such as premature heart diseases, diabetes, and cancer. Unhealthy diets have been identifiedas the important causing factor of such diseases. In this context, personalized nutrition emerges as anew research field for providing tailored food intake advices to individuals according to their physical, physiological data, and further personal information. Specifically, in the last few years, several types of research have proposed computational models for personalized food recommendation using nutritional knowledge and user data. In this paper, we present a food recommender system designed to generate personalized meal suggestions by incorporating both user preferences and medical information (allergens like gluten, dairy etc.). Traditional recommender systems often focus on taste and preferences, but our approach integrates critical health factors such as medical conditions, dietary restrictions, and nutritional requirements. The system utilizes machine learning algorithms to analyze user-specific data, including allergies, medical history (e.g., diabetes, hypertension), and individual taste preferences. By employing clustering techniques, such as K-nearest neighbors (KNN), and advanced filtering mechanisms, our model recommends foods that are not only appealing but also aligned with the user's health goals. The system adapts over time as new data about the user's preferences and medical conditions is introduced. Experimental results show the effectiveness of the proposed system in providing healthier, more personalized food recommendations, making it a valuable tool for users aiming to maintain or improve their health while enjoying their meals.

INDEX TERMS Daily meal plan recommendation, user preferences, nutritional information, multi-criteria decision making, recommender systems.

I. INTRODUCTION

The World Health Organization estimates that noncommunicable diseases such as cardiovascular diseases,cancer, chronic respiratory diseases and diabetes, are responsible for 63% of all deaths worldwide [39]. Furthermore, it also points out that such diseases are preventable through effective interventions that tackle shared risk factors such as the unhealthy diets. In this context, whereas a one-size-fits-all approach may fail, personalized nutrition can benefits consumers to adhere to a healthy, pleasurable, and nutritional diet when it is closely associated to individual parameterssuch as the physical and psychological characteristics including health status, phenotype and genotype, the consumer's needs and preferences, behavior, lifestyle, as well as budget.Personalized nutrition can be used for different target groups from healthy people to patients such as malnourished people, vulnerable groups, people with allergies or noncommunicable diseases, including cancer.Personalized nutrition has been formally deined as the healthy eating advice, tailored to suit an individual based on genetic data, and alternatively on personal health status, lifestyle, nutrients intake and phenotypic data. Regarding the cost of genetic data management, in the last few years there have been an increasing in the research efforts focused on the management of these alternative data with this aim in mind . Specically, several computational solutions have been proposed with the goal of healthy eating advice.

- The **menu planning problem** has been a focus of research for over 50 years, and it continues to be an active area of exploration, especially in developing personalized menu generation frameworks. In the past few years, two distinct research clusters have emerged, emphasizing different aspects of personalized healthy menu generation:
- **Building complex information models** for personalized services. These models often incorporate tools such as flow charts, inference engines, medical questionnaires, and prescriptions processing, with a significant emphasis on semantic information modeling using ontologies. These efforts aim to create robust information sources to drive nutritional recommendations. However, these models often face challenges in terms of generalization and complexity, particularly in scaling across diverse user needs.

- **Nutritional information processing** focuses on utilizing available nutritional data, tackling But these approaches typically lack integration of individual preferences and real-time adaptability to the user's unique requirements.

A notable shortcoming of these approaches is their limited focus on the **user's preferences**, a crucial factor in creating personalized nutrition plans. Moreover, many existing solutions treat nutrition as a component of larger wellness platforms, rather than as the central focus. The integration of nutritional principles into computational models has also been insufficiently explored, and while semantic modeling holds promise, it often struggles with generalization across different contexts.

My project aims to address these limitations by developing an **Allergen-based Food Recommendation System**, which incorporates both **personal preferences** and **nutritional principles** into its recommendation model. Specifically, my project will focus on the following research questions:

a) Does incorporating the user's **allergy profile** and preferences improve personalized menu recommendations?
b) How does integrating **nutritional principles** into the recommendation process enhance the overall effectiveness of the menu planning system?

To explore these questions, I will employ **recommender systems (RSs)**, a proven tool for personalizing user experiences in contexts of information overload. RSs have been widely used in domains such as e-commerce, e-learning, and tourism to provide tailored recommendations. However, their application in food recommendation systems, especially those considering both nutritional and allergen-related needs, remains relatively underexplored

This project will advance the state of food RSs by building a model that integrates **nutritional data, allergy management**, and **user preferences** using a combination of **multi-criteria decision-making (MCDM)** approaches and **machine learning** algorithms, including **K-Nearest Neighbors (KNN)**. This hybrid system will offer personalized, safe meal recommendations that meet the user's daily nutritional requirements while avoiding allergens.

The novelty of this project lies in the following:

a) **Multi-Dimensional Input**: Unlike traditional recommendation systems that primarily focus on ingredients or user preferences, this project combines **nutritional information** (calories, fats, proteins, etc.) with allergy data. This allows for more comprehensive and tailored recommendations.
b) **Dynamic Allergy Handling**: The system dynamically adjusts its recommendations by taking into account the specific **allergies** of the user. Using advanced **machine learning algorithms** like **Random Forest**, the system not only predicts allergies associated with foods but also filters out foods that may be unsafe for the user, offering a **highly personalized** experience.
c) **Nutritional and Allergy-based Filtering**: The project integrates the power of **SMOTE** (Synthetic Minority Over-sampling Technique) to address class imbalance in the allergy dataset, ensuring that rare allergy types do not get overlooked. This results in a more **inclusive** and **diverse set of recommendations** for all users, regardless of the prevalence of their specific allergies.
d) **PCA for Dimensionality Reduction**: The use of **Principal Component Analysis (PCA)** in feature extraction helps reduce the complexity of the data while preserving essential variance, improving the efficiency of the recommendation process. This contributes to faster and more accurate predictions.
e) **Recipe Recommendation Based on Multiple Factors**: While many recommendation systems focus solely on **ingredient-based matching**, this project integrates both **ingredient features** (via **TF-IDF Vectorization**) and **nutritional information**. This approach allows the system to recommend recipes that not only match nutritional needs but also comply with allergy restrictions.
f) **Cross-Validation for Robustness**: The use of **cross-validation** techniques helps ensure the model generalizes well and avoids overfitting, making it more robust in real-world scenarios. The model's accuracy is consistently tested against different folds of the data to ensure reliable and trustworthy recommendations.
g) **Real-Time Personalization**: The recommendation system can be personalized in real-time as users input their nutritional and allergy information. This makes the system adaptable to changing dietary needs, evolving user preferences, or updated allergy information.

The remainder of this proposal outlines the architecture of the allergen-based food recommendation system, including data preparation, **multi-criteria decision analysis**, and optimization-based menu recommendations. We will demonstrate the system's effectiveness through a case study and present a detailed analysis of the results.

II. RESEARCH METHODOLOGY:

II.1. PROCESS

Allergen-based food recommender systems are gaining importance due to the growing need for personalized meal planning for individuals with allergies. Existing studies in the food recommendation domain largely focus on dietary preferences, but many overlook the critical aspect of allergen avoidance. This project tackles the challenge of recommending safe, balanced meals by considering users' **allergy profiles**, alongside their height, weight, age, and activity levels.
The system is designed to provide tailored meal suggestions that align with a user's Total Daily Energy Expenditure (TDEE), while excluding allergenic ingredients. In addition, the system tracks users' nutritional intake, helping them maintain healthy eating habits. This study provides an overview of recommendation techniques used to suggest safe and nutritious meals for individuals with specific dietary needs, addressing the gap in existing recommender systems.

II.2. COLLECTION OF DATA

Data collection focuses on gathering nutritional information and allergen data for a wide variety of food items. The data is compiled from user-reported allergies, lifestyle details, and daily activity records. The framework is designed to inspect the collected data for accuracy and compare it with historical data to ensure consistency. This data is essential for building a comprehensive dataset that supports personalized food recommendations. The collection process is followed by data pre-processing, where techniques are applied to clean and normalize the dataset for use in the machine learning model.

II.3. DATA PREPARATION

The data preparation stage involves pre-processing the raw data to ensure it is ready for analysis by machine learning algorithms. Irrelevant attributes, such as foods that do not meet the user's dietary needs or contain allergens, are filtered out. This step is critical to ensuring that the model works with clean, relevant data, allowing for accurate food recommendations that meet both nutritional and allergy-related constraints.

II.4. CLUSTERING PROCESS

In this stage, the user data is clustered to identify patterns and similarities in dietary preferences and nutritional needs. The k-means algorithm is then applied, with k representing the number of clusters that produce the best results. This clustering process helps in segmenting users based on their unique profiles, making it easier to deliver personalized food recommendations.

II.5. PRE-PROCESSING

Pre-processing is a vital task in preparing the dataset for machine learning algorithms. The raw data is transformed into a format suitable for training models, ensuring that irrelevant information is removed, and essential features are highlighted. Without proper pre-processing, training on raw data can lead to poor results, especially in allergen-based recommendation systems where precision is key. This stage optimizes the dataset for machine learning by resolving inconsistencies, handling missing values, and enhancing the overall quality of the data.

III. THE GENERAL ARCHITECTURE FOR FOOD RECOMMENDATION:

The architecture of your Food Recommendation System focuses on providing personalized food recommendations by taking into account the nutritional information and user allergies. The core flow of the system involves:

1. Data Preparation:
    - The system begins by loading the dataset, typically an Excel file containing detailed information about various food items, including their nutritional values and associated allergies.
    - The file, foodrecmergedallergen.xlsx, contains columns like Food_items, Calories, Fats, Proteins, Iron, Calcium, Sodium, Potassium, Carbohydrates, Fibre, VitaminD, Sugars, and Food_allergy.
2. Normalization of Numerical Features:
    - The nutritional features such as Calories, Fats, Proteins, etc., are normalized using StandardScaler to scale the data to a standard range (zero mean and unit variance). This ensures that the features have equal importance when feeding them into machine learning models.
    - X_numerical represents the normalized nutritional data.
3. Model Training (K-Nearest Neighbors - KNN):
    - The KNN model (K-Nearest Neighbors) is used for finding similar food items based on their nutritional values.

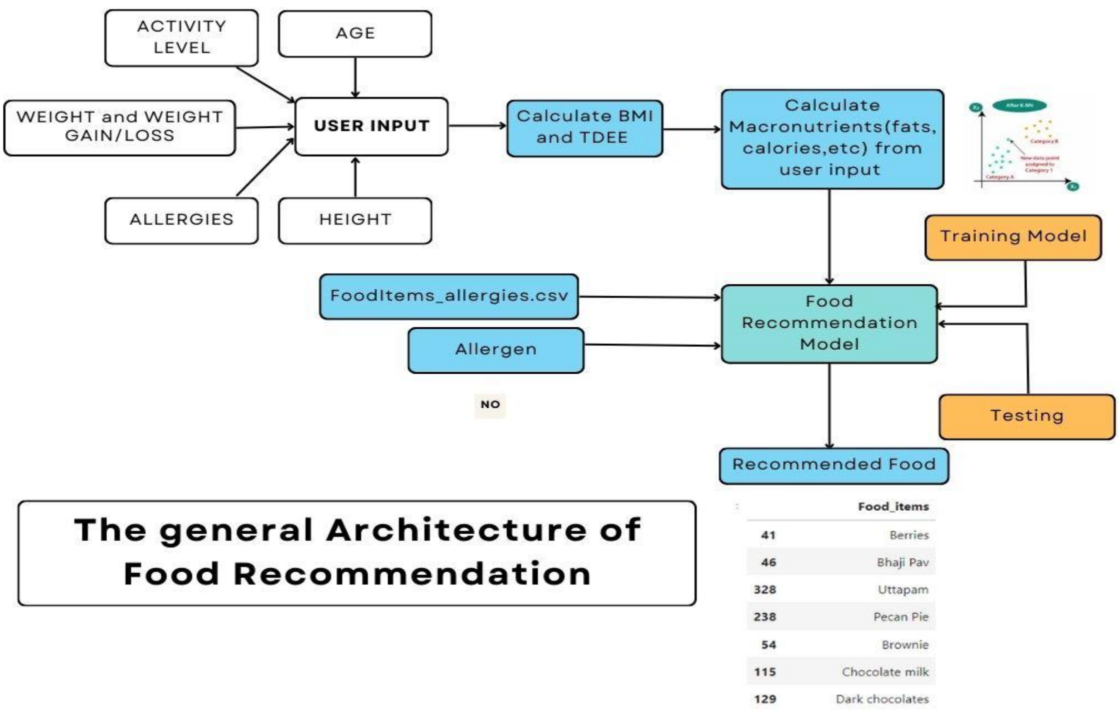
- It is trained using the NearestNeighbors class from scikit-learn, where the Euclidean distance metric is used to measure similarity. The model is trained on the normalized nutritional features of the food items.
- This step ensures that the model can find the closest (most similar) food items based on nutritional data for a given input.

4. Recipe Recommendation:

   - The function recommend_recipes(input_features, user_allergy) is where the core recommendation process happens.
   - The input features (user's nutritional preferences) are scaled using the same StandardScaler that was used during training. This ensures consistency between the model's training data and the new user input.
   - Using the KNN model, the system finds the top K (11) closest food items that best match the user's nutritional input.
   - The recommendations returned by KNN are filtered to exclude any food items that match the user's allergies. The allergy information is taken into account by checking the Food_allergy column in the data.
   - The result is a set of filtered recommendations that don't contain ingredients the user is allergic to, ensuring a safe and tailored recommendation.

5. Return Results:

   - The function returns a dictionary containing the Food_items, image_url, and Food_allergy for each recommended recipe. This makes the recommendations easy to display in a user interface, providing details like the name of the food item, an image of the food, and any associated allergy information.



The general Architecture of Food Recommendation

IV. IMPLEMENTATION METHODOLOGIES:

In this section, we describe the methodologies used to implement the K-Nearest Neighbors (KNN) and Random Forest algorithms for the food recommendation system. Both techniques are leveraged to provide personalized recommendations based on the user's nutritional preferences while taking allergies into account.

1. **K-Nearest Neighbors (KNN) :** K-Nearest Neighbors is a simple, instance-based machine learning algorithm that is commonly used for classification and regression tasks. It operates by finding the closest points (neighbors) in the

training data to a given query point and uses the labels of those neighbors to predict the class (or value) of the query point.

Methodology for KNN-based Food Recommendation:

- Data Preprocessing:
  - Numerical Features: The dataset includes various nutritional features, such as Calories, Fats, Proteins, Iron, Calcium, and others. These features are normalized using StandardScaler to ensure that each feature has zero mean and unit variance. This step prevents features with larger values (e.g., calories) from disproportionately influencing the model.
  - Feature Scaling: Nutritional input data from users is scaled using the same StandardScaler used for training data, ensuring that the scale of input data is consistent with the model's training data.
- Model Training:
  - The KNN model is trained using the scikit-learn NearestNeighbors class, with Euclidean distance as the metric for similarity measurement. Euclidean distance is suitable for continuous features like nutritional values.
  - The model is trained on the normalized nutritional features (e.g., Calories, Fats, etc.), where each food item is treated as a point in a high-dimensional space. The n_neighbors parameter is set to 11 to find the 10 nearest neighbors, including the query point itself.
- Recommendation:
  - Given a user input with nutritional preferences (e.g., calorie intake, fat content), the system scales these input features using StandardScaler and uses the KNN model to find the top 10 most similar food items.
  - Allergy Filtering: Once the nearest neighbors are identified, the list of recommended food items is filtered to exclude items that contain the user's allergies. This step ensures that all recommended items are safe for the user to consume.
  - Result: The system returns the names, images, and allergy information of the top 10 recommended food items, ensuring the results align with the user's preferences and restrictions.

Advantages of KNN for Food Recommendations:

- Simplicity: KNN is simple to understand and easy to implement. It does not require a training phase, which is beneficial in certain scenarios.
- Non-parametric: KNN makes no assumptions about the underlying data distribution, making it flexible for diverse datasets.
- Effective for Similarity-Based Recommendations: KNN is particularly useful in scenarios where the recommendation system is based on finding similarities between items, such as nutritional content in food items.

2. **Random Forest:**Random Forest is an ensemble learning method that aggregates the predictions of multiple decision trees to improve model accuracy and robustness. It is widely used for both classification and regression tasks due to its ability to handle complex data patterns and provide high predictive performance.
   Methodology for Random Forest-based Food Recommendation:
   - Data Preprocessing:
     - Similar to KNN, numerical features (nutritional values such as Calories, Fats, Proteins, etc.) are scaled using StandardScaler to ensure uniformity in the scale of the input data.
     - Textual Features (e.g., food item descriptions) are processed using the TF-IDF Vectorizer, which converts text data into numerical vectors representing word importance relative to the entire corpus. This transforms food descriptions into features that can be used by the Random Forest model.
   - Feature Engineering:
     - Numerical Data: Nutritional features are standardized to make sure the Random Forest algorithm does not bias certain features over others due to their scale.

    - Textual Data: Food descriptions are transformed using TF-IDF (Term Frequency-Inverse Document Frequency), capturing the relevance of individual terms (e.g., ingredients) for each food item.

- Model Training:
    - A Random Forest Classifier is trained using the processed data. It is an ensemble method that constructs multiple decision trees and aggregates their predictions to make a final decision. The model is fitted using n_estimators = 100 (number of decision trees) and random_state = 42 to ensure reproducibility.
    - The model leverages both the nutritional and textual features for training, capturing complex relationships between the food items and their properties.

- Recommendation:
    - Given a user's nutritional input, the features are processed, and the model predicts the class (allergy type) of the food items.
    - Class Probabilities: The Random Forest model provides not only a prediction (i.e., the allergy type) but also the probability of each class. This helps in ranking food items based on how likely they are to match the user's nutritional preferences.
    - Allergy Filtering: After generating the predictions, food items that contain allergens are filtered out, ensuring that the recommended foods are safe for the user.
    - Result: The system returns the top recommended food items, filtered by the user's allergies and sorted by the probability of being a good match for the user's input.

Advantages of Random Forest for Food Recommendations:

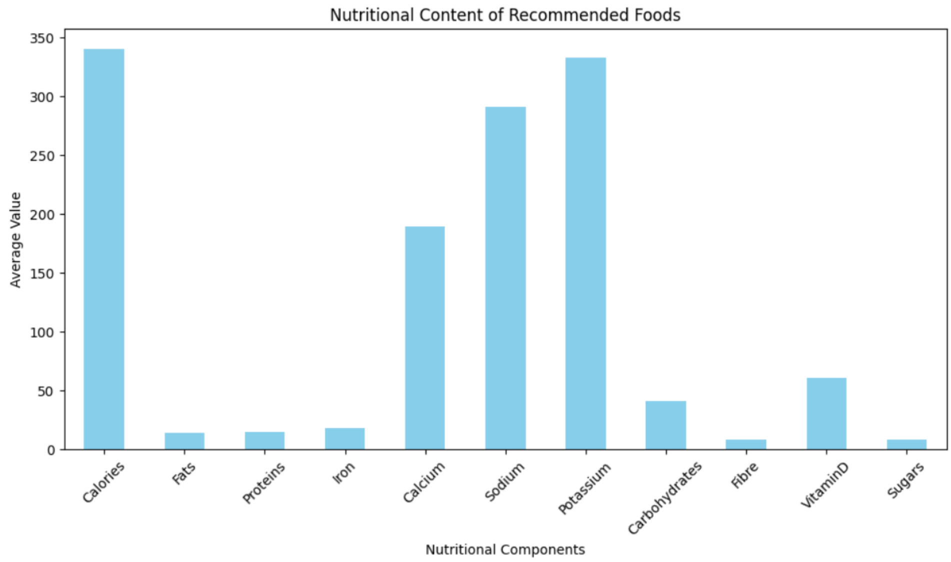
- Handles Complex Data: Random Forest can capture complex interactions between features (e.g., nutritional values and food descriptions) better than simpler models like KNN.
- Robustness: It is less prone to overfitting compared to individual decision trees, making it a reliable model even with noisy data.
- Feature Importance: Random Forest can provide insights into which features are most important for making predictions, offering interpretability.

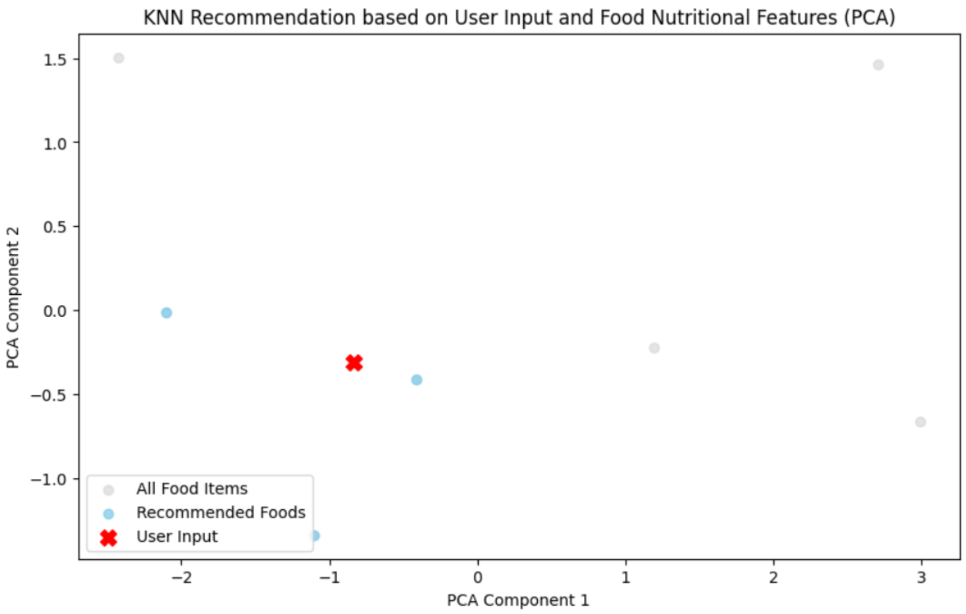
Comparison of KNN and Random Forest:KNN is based on similarity, meaning that it finds food items that are numerically closest to the user's input. It is more intuitive but less powerful when dealing with complex feature interactions. Random Forest, on the other hand, creates a more generalized model that can handle intricate relationships between different features (e.g., nutritional data and text descriptions). It is more flexible and can handle larger datasets with higher accuracy.

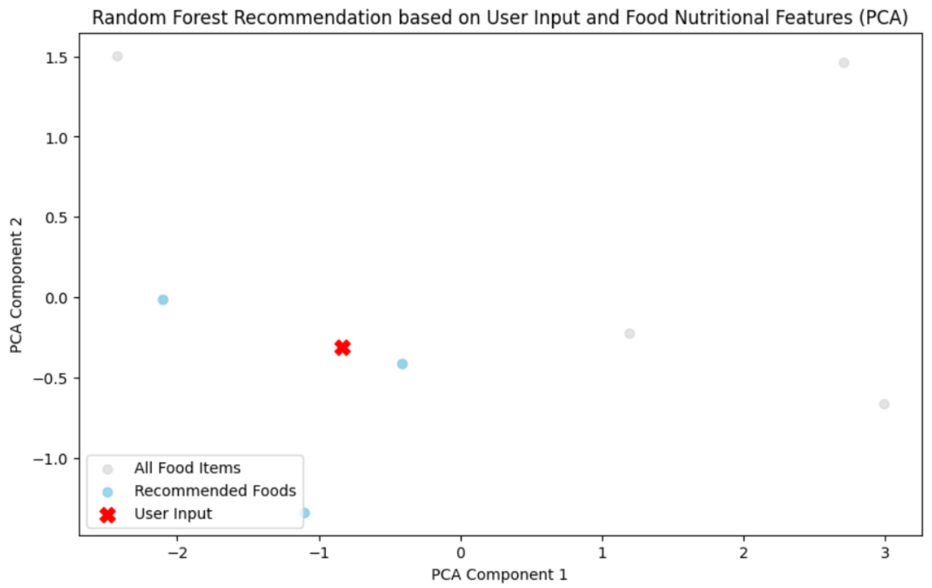
Both models are effective for food recommendation, but Random Forest might offer better predictive performance when the dataset contains complex relationships that cannot be easily captured by the KNN algorithm.

The food recommendation system utilizes KNN for finding similar food items based on nutritional preferences and Random Forest for providing more robust predictions that consider both numerical and textual data. The combination of these algorithms ensures that the system delivers personalized, safe, and accurate food recommendations for users with dietary restrictions and allergies.

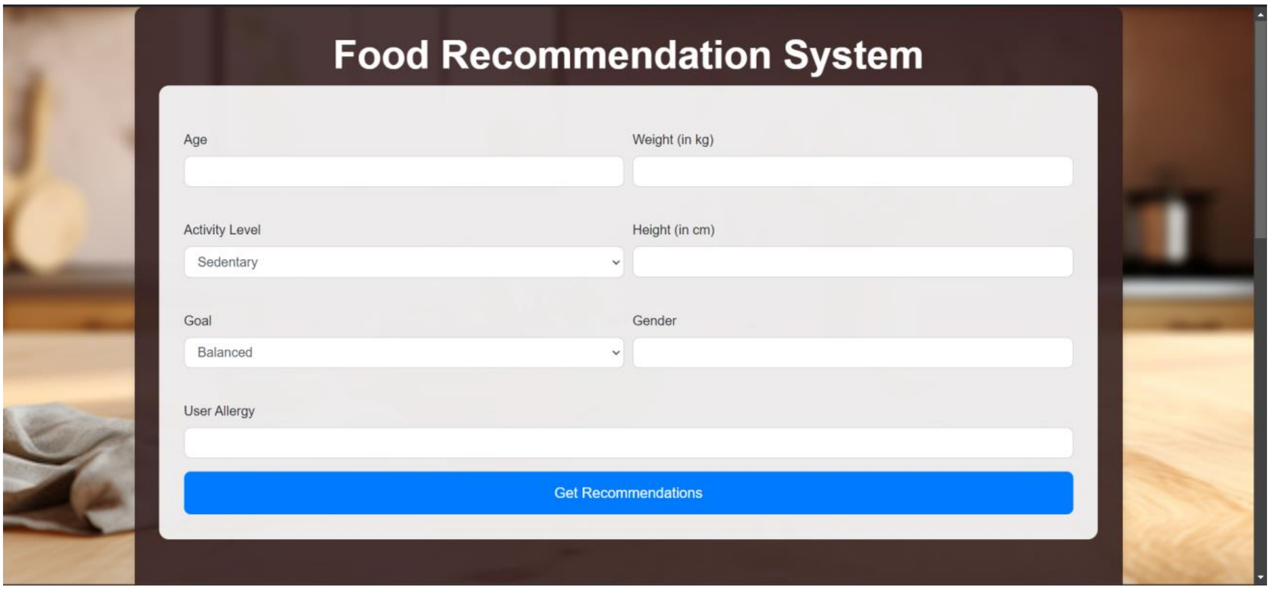
V. VISUAL REPRESENTATION:



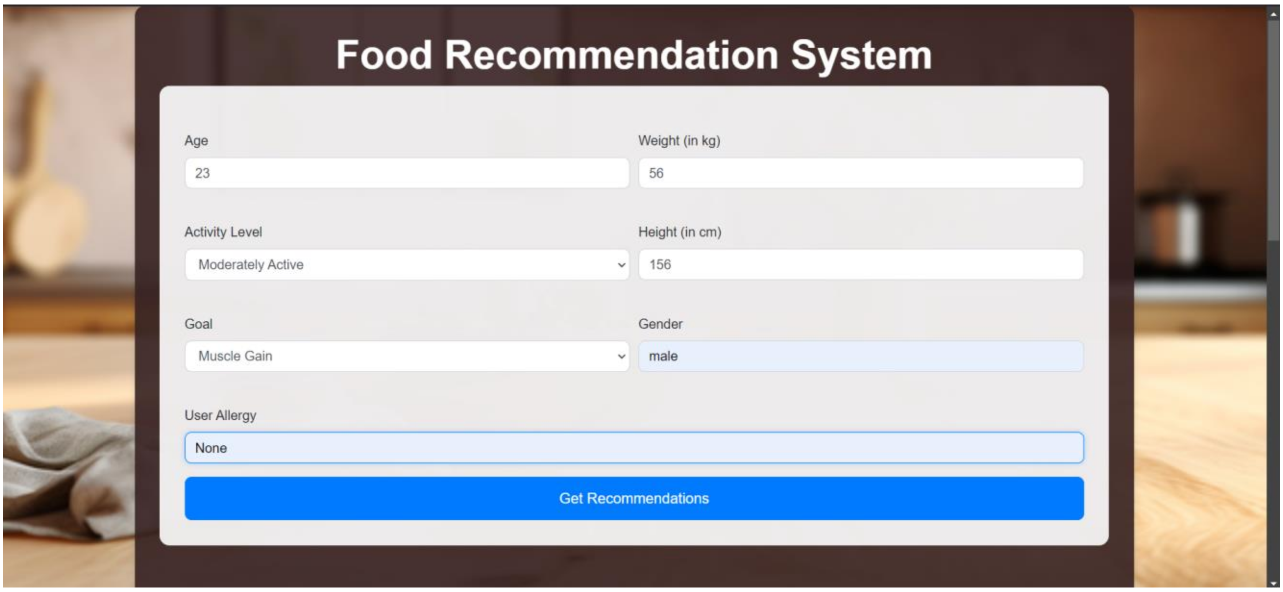




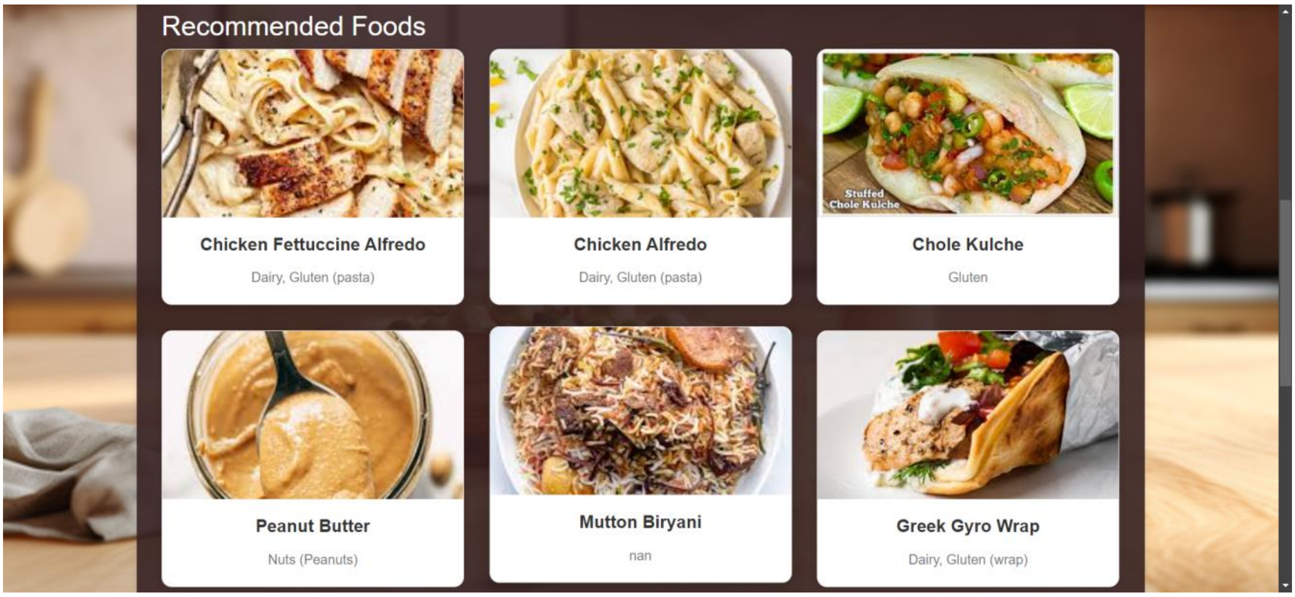
VI. OUTPUTS:



Food Recommendation System
Age
Weight (in kg)
Activity Level
Sedentary
Height (in cm)
Goal
Balanced
Gender
User Allergy
Get Recommendations



Food Recommendation System
Age
23
Weight (in kg)
56
Activity Level
Moderately Active
Height (in cm)
156
Goal
Muscle Gain
Gender
male
User Allergy
None
Get Recommendations



Recommended Foods
Chicken Fettuccine Alfredo
Dairy, Gluten (pasta)
Chicken Alfredo
Dairy, Gluten (pasta)
Stuffed Chole Kulche
Chole Kulche
Gluten
Peanut Butter
Nuts (Peanuts)
Mutton Biryani
nan
Greek Gyro Wrap
Dairy, Gluten (wrap)

VII. CONCLUSION:

The **Allergen-based Food Recommendation System** offers a personalized solution to the complex challenge of meal planning for individuals with food allergies. By integrating advanced machine learning algorithms like **K-Nearest Neighbours (KNN)** with comprehensive nutritional and allergen data, the system ensures that users receive meal suggestions that are not only safe but also nutritionally balanced. The inclusion of **user-specific data**—such as height, weight, age, and dietary goals—enables the system to tailor recommendations precisely to each user's needs, enhancing both safety and health outcomes. The project's core strengths lie in its ability to dynamically adapt to users' evolving health requirements, while its user-friendly interface makes it accessible to individuals with varying technical knowledge. As the system continues to learn from user feedback, it becomes more accurate in predicting safe and nutritious meals. This not only improves the quality of life for those with allergies but also has broader applications in healthcare, nutrition, and food service industries, where personalized nutrition is becoming increasingly important. Ultimately, this system paves the way for safer and more efficient meal planning, addressing both dietary needs and allergy risks, while contributing to overall well-being.

VIII. REFERENCES:

[1] (WenjieWang, Ling-Yu Duan, Hao Jiang, Peiguang Jing, Xuemeng Song, and Liqiang Nie. 2021. Market2Dish:Health-aware Food Recommendation. ACM Trans. Multimedia Comput. Commun. Appl. 17, 1, Article 33 (April 2021),acm.org

[2] (Reetu Singh, Research Scholar Computer Science and Engineering Department, Motilal Nehru National Institute of Technology, Allahabad, Prayagraj, (Uttar Pradesh); Pragya Dwivedi Assistant Professor Computer Science and Engineering Department Motilal Nehru National Institute of Technology Allahabad, Prayagraj, (Uttar Pradesh), India), https://www.researchgate.net/publication/374418599 **(October 2023**)

[3] Mehrdad Rostami, Kamal Berahmand, Saman Forouzandeh, Sajad Ahmadian, Vahid Farrahi, Mourad Oussalah Center for Machine Vision and Signal Analysis (CMVS), Faculty of Information.Received 27 March 2023; Received in revised form 15 September 2023**,** www.elsevier.com /locate/ neucom (science direct)**; Accepted 21 January 2024**

[4] CELESTINE IWENDI 1, (Senior Member, IEEE), SULEMAN KHAN2, JOSEPH HENRY ANAJEMBA 3, (Member, IEEE), ALI KASHIF BASHIR 4, (Senior Member, IEEE), AND FAZAL NOOR5 Received December 21, 2019, accepted January 17, 2020, date of publication January 21, 2020, date of current version **February 17, 2020**.

[5] Saman Forouzandeh a, ∗, Mehrdad Rostami b, Kamal Berahmand c, Razieh Sheikh pour. Accepted **18 December 2023(sciencedirect)**

[6] RACIEL YERA TOLEDO1, AHMAD A. ALZAHRANI2, AND LUIS MARTíNEZ 3, (Member, IEEE) date of publication **July 17,2021**

[7] Nazar Oleksiv 1, Oleh Veres 1, Andrii Vasyliuk 1, Ihor Rishnyak 1 and Lyubomyr Chyrun 2,1 Lviv Polytechnic National University, S. Bandera Street, 12, Lviv, 79013, Ukraine,2 Ivan Franko National University of Lviv, University Street, 1, Lviv, 79000, Ukraine acm.org,date of published **May 12–13, 2022**

[8] Reema Golagana, V. Sravani, T. Mohan Reddy, IV year B.Tech students, Dept. of Computer Science Engineering, Raghu Engineering College, Visakhapatnam, Andhra Pradesh, India **Vol-13, Issue-4, April 2023**

[9] Shubham Singh Kardam1, Pinky Yadav2, Raj Thakkar3, Prof Anand Ingle4, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056,Volume: 08 Issue: 04 **| Apr 2021** www.irjet.net

[10] Abolfazl ajami,Dr. babak Teimourpour,published may 21 2021

[11] Bauer J, Nanopoulos A (2014). Recommender systems based on quantitative implicit customer feedback. Decision Support Systems, 68:77-88.

[12] Trang Tran TN, Atas M, Felfernig A, Stettinger M (2018). An overview of recommender systems in the healthy food domain. Journal of Intelligent Information Systems, 50:501-26.

[13] Min W, Jiang S, Jain R (2019). Food recommendation: Framework, existing solutions, and challenges. IEEE Transactions on Multimedia, 22(10):2659-71.

[14] Mika S (2011). Challenges for nutrition recommender systems. InProceedings of the 2nd Workshop on Context Aware Intel. Assistance, Berlin, pp. 25-33.

[15] Burke L, Deakin V (2010). Clinical sports nutrition. McGraw Hill.

[16] Kaminsky L (2006). ACSM's resource manual for Guidelines for exercise testing and prescription. CQUniversity