TAILORED DIET RECOMMENDATION SYSTEM USING KNN

B. Bhaskara Rao
bhaskararaobrd@gmail.com

S. Karimulla Basha
kareem768@gmail.com

M. Madhavan
21095a3204@rgmcet.edu.in

J. Mahesh
20091a3225@rgmcet.edu.in

B.S. Rohith Kumar
20091a3238@rgmcet.edu.in

Department of Computer Science and Engineering (Data Science)
Rajeev Gandhi Memorial College of Engineering and Technology, Nandyal, 518501, AP, India

Abstract— In recent years, there has been a notable increase in chronic diseases, affecting a significant portion of both elderly and younger populations. The World Health Organization advocates for maintaining a proper diet to promote better health. This paper introduces a personalized food recommendation system tailored to individual dietary needs and preferences. Utilizing an extensive dataset sourced from Kaggle, specifically "Food.com - Recipes and Reviews," which is available under the CC0: Public Domain license, comprising 522,517 recipes across 312 different categories, encompassing diverse nutritional information. We investigate the effectiveness of various machine learning algorithms, including K-Nearest Neighbors (KNN), Random Forest (RF), Decision Tree (DT), and Linear Regression (LR), in predicting optimal food choices. After rigorous experimentation involving training and testing these algorithms, K-Nearest Neighbors (KNN) emerges as the most suitable model for our dataset. Our primary focus is on meeting the unique dietary preferences of users by providing recommendations for specific food items, accompanied by visualized nutrient values, ingredient lists, and preparation instructions. Additionally, we introduce a novel feature that suggests food items based on individual Body Mass Index (BMI) and Basal Metabolic Rate (BMR), offering a more personalized food recommendation approach. Our research contributes to the development of personalized diet recommendation systems, furnishing users with tailored and informative suggestions to support their dietary objectives and preferences.

Index Terms— tailored diet recommendation, machine learning algorithms, personalized nutrition guidance, dietary preferences, health objectives, content-based filtering, data preprocessing, feature engineering, K-Nearest Neighbors (KNN).

I. INTRODUCTION

EVERYONE wants live their complete life healthier till they die, there might be some illness and diseases comes along with age, desires to lead a healthy life until their twilight years, but in recent times, a concerning trend has emerged where chronic diseases afflict not just the elderly but also middle-aged and even younger individuals. These diseases pose increasing risks as people age. The World Health Organization (WHO) has issued a comprehensive report titled "DIET, NUTRITION AND THE PREVENTION OF CHRONIC DISEASES" [1], which addresses various aspects of diet, nutrition, and the mitigation of chronic illnesses. The report highlights the challenges to public health, including shifting dietary patterns, reduced physical activity due to sedentary lifestyles, an aging population, and the detrimental effects of tobacco and alcohol consumption. The joint WHO/FAO expert consultation underscores the importance of addressing diet, nutrition, and physical activity collectively to combat chronic diseases effectively. While the primary focus was on setting dietary and nutritional targets, the significance of physical exercise was also emphasized. The report emphasizes how adopting healthy dietary and exercise habits across the lifespan can mitigate the global burden of chronic diseases. Several recommendation systems for dietary management exist, each serving distinct purposes. For instance, [2] introduces a patient diet recommendation system leveraging machine learning within the context of the Internet of Medical Things (IoMT). Similarly, [3] presents a diet recommendation system tailored to individual requirements. [4] utilizes machine learning and big data for meal recommendations based on human behavior, while [6] provides a platform for personalized expert nutrition recommendations considering both nutritional information and user preferences. Each of these systems employs different methodologies and addresses diverse use cases. In our project, we focus on a content-based recommendation system using data sourced from the "Food.com - Recipes and Reviews" dataset available under the CC0: Public Domain license. This dataset comprises a vast collection of 522,517 recipes across 312 categories. We experiment with various machine learning models such as KNN, RF, DT, LR and evaluate their performance using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) Score. Apart from the core Personalized Diet Recommendation System, we also included one more additional feature to generate diet recommendation based on BMR [11] which will be a generalized diet recommendation system. The sub sequent sections of this paper are structured as follows: Section II reviews existing work in the domain of diet recommendation. Section III delineates our methodology, including system architecture, data collection, pre-processing, algorithms, training, and testing. Section IV presents our experimental results. Section V outlines potential future research directions, and finally, Section VI concludes the paper.

II. Literature Survey

C. Iwendi et al. [2] introduce an efficient patient diet recommendation system leveraging machine learning models in the context of the Internet of Medical Things (IoMT). Their study highlights the integration of IoMT technologies to enhance personalized dietary guidance for patients. By employing machine learning algorithms, the proposed system aims to deliver tailored diet recommendations, contributing to optimized patient health management. This research underscores the potential of IoMT and machine learning in healthcare practices, particularly in dietary interventions and patient care.

Shah, Degadwala, and Vyas [3] diet recommendation system for dietary advice to individual needs using various machine learning techniques. Their research highlights the urgent need for personalized nutrition in today's health-conscious society, showcasing how machine learning can transform dietary recommendations.

Lambay and Mohideen [4] proposed an innovative hybrid diet recommendation approach, blending machine learning with big data analytics to provide personalized dietary guidance. Their research emphasizes the importance of utilizing the advantages of different algorithms such as Naïve Approach, K-Nearest Neighbors (KNN) and Latent Factor Model(LFM) by preparing a hybrid model to improve recommendation accuracy.

Islam et al. [5] introduced a meal recommendation system based on human behavior analysis, transforming the way dietary suggestions are made. Authors used an hierarchical ensemble method applied to predict affectivity upon multiple feature extraction methods and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is used to generate a food list based on the predicted affectivity.

Chen et al. [6] presented the PERSON - Personalized Expert Recommendation System for Optimized Nutrition, platform for personalized expert nutrition recommendations. A personalized expert recommendation system for optimized nutrition is introduced in this paper, which performs direct to consumer personalized grocery product filtering and recommendation. Their study emphasizes the importance of expert knowledge and optimization techniques in tailoring dietary guidance to individual needs.

Yera Toledo, Alzahrani, and Martínez [7] developed a food recommender system that considers both nutritional information and user preferences, how individuals make dietary choices. Their research underscores the significance of integrating nutritional content with user-centric approaches for a comprehensive food recommendation experience.

Sarker [8] explored the development of recommendation systems, offering valuable insights into machine learning algorithms' real-world applications. The paper provides a comprehensive overview of different types of real time data and compatible machine learning algorithms. [11,12] authors explored that are utilizing in Vehicular Ad hoc networks.

III. METHODOLOGY

A) System Architecture:

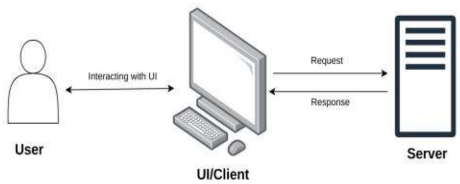

Figure 1. Client-Server Architecture.

The Figure 1 shows architecture diagram of a visual representation of a client-server system, where a user interacts with a user interface (UI) to make requests to a server. The user inputs their preferences and other information. The UI then sends a request to the server, which processes the request and generates a response.

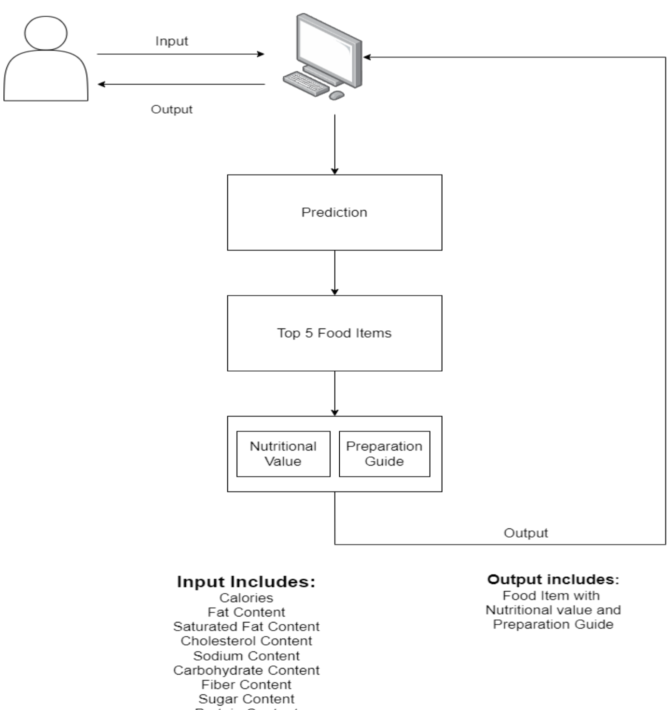

Figure 2. Tailored Diet Recommendation System Architecture.

The Figure 2 shows Architecture Diagram of a Tailored Diet Recommendation System Architecture, which includes:

Input: This section list of types of nutritional information that the system can process. These include calories, fat content, saturated fat content, cholesterol content, sodium content, carbohydrate content, fiber content, sugar content, and protein content.

Prediction: Based on the input nutrient values the internal diet recommendation system will make a prediction. The predicted food items would be top 5 food items based on the nutrients nearest values.

Nutritional Value and Preparation Guide: In the predicted food items, along with nutrient values, preparation of the food item is also given.

Output: Output contains recommended food items with nutritional value and preparation process of the food item.

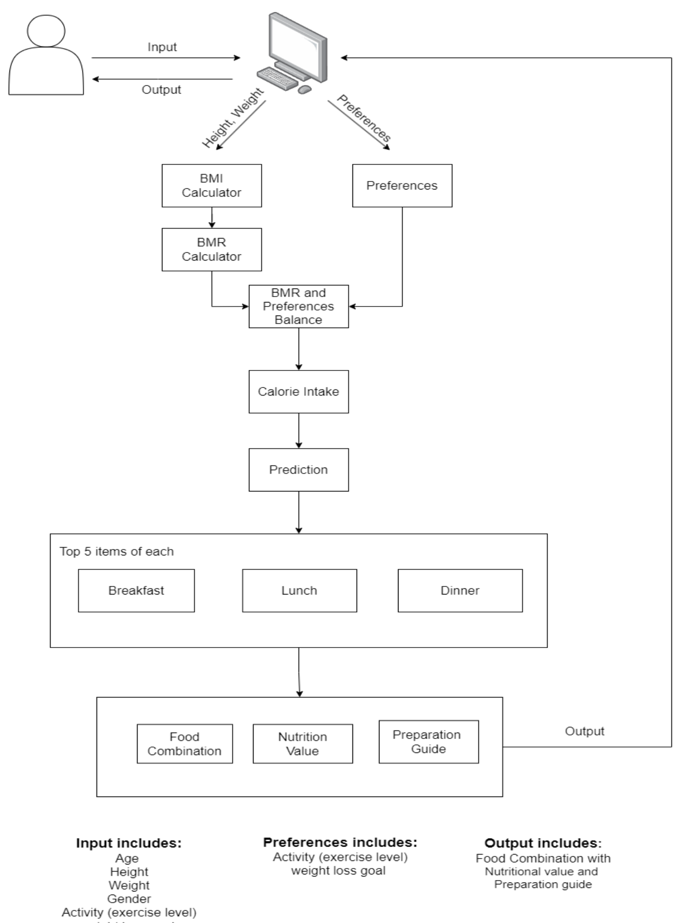

Figure 3. Generalized Diet Recommendation System with BMI and BMR.

The Figure 3 shows the architecture diagram of generalized diet recommendation system that based on BMI (Body Mass Index), BMR (Basal Metabolic Rate) and user preferences in terms of activity level. Provides recommendations for food combinations and nutrition based on user preferences and input. The diagram is divided into several sections, each representing a different component of the system.

Input: This section includes the user's information such as age, height, weight, gender, activity level, and weight loss goal. This information is used to calculate the BMI and BMR to provide personalized recommendations.

BMI Calculator: This component calculates the user's BMI based on the input provided. The BMI is used to determine the user's weight category (underweight, normal weight, overweight, or obese) and provide appropriate recommendations.

BMR Calculator: This component calculates the user's BMR based on their age, height, weight, gender, and activity level input. The BMR is used to determine the user's minimum number of calories a person needs to consume to maintain their body weight, and it is affected by several factors, including age, gender, body size and provide recommendations for food combinations that meet those needs.

Preferences: This section includes the user's preferences such as activity level and weight loss goal. This information is used to provide personalized recommendations for food combinations and nutrition.

BMR and Preferences Balance: This section combines the user's BMR calculation with their preferences, such as weight loss goals, to generate personalized nutrition recommendations.

Calorie Intake: This section calculates the user's daily calorie intake based on their BMR and weight loss goal. This information is used to provide recommendations for food combinations that meet the user's calorie needs.

Prediction: This section predicts the user's food items combination based on BMR and preferences balance based on their input and preferences.

Meal Categories: This section includes the different meal categories (Breakfast, Lunch, and Dinner) where the user can input the food items they consume for each meal.

Food Combination with Nutrition and Preparation guide: This section provides recommendations for food combinations with nutrition based on the user's input and preferences along with process of preparing the food item.

Output: This section displays the output of the system, including the user's BMI, calorie intake, and recommended food combinations with nutrition and preparation of the recommended food items.

B) Data Collection:

Table 1. Dataset

Sl.No	FEATURE	TYPE
1	RecipeId	Numeric
2	Name	Categorical
3	Calories	Numeric
4	FatContent	Numeric
5	SaturatedFatContent	Numeric
6	CholesterolContent	Numeric
7	SodiumContent	Numeric
8	CarbohydrateContent	Numeric
9	FiberContent	Numeric
10	SugarContent	Numeric
11	ProteinContent	Numeric

Dataset Description:

The Table 1 shows dataset utilized for this project is sourced from Kaggle and is known as "Food.com - Recipes and Reviews." This dataset is available under the CC0: Public Domain license. It comprises a vast collection of 522,517 recipes spanning 312 different categories. Each recipe entry provides comprehensive information such as cooking times, servings, ingredients, nutritional values, instructions, and more. However, for the purpose of this project, only select features relevant to the diet recommendation system have been chosen. Including the name of the recipe, recipe category (categorical type), and various nutritional components such as calories, fat content, saturated fat content, cholesterol content, sodium content, carbohydrate content, fiber content, sugar content, and protein content. The name of the recipe serves for food item identification and prediction, while the recipe category and nutritional values are crucial for the recommendation system's functioning. The dataset's numeric features provide essential data for analysis and prediction, aiding in the development of a robust and effective diet recommendation system tailored to individual preferences and nutritional needs.

At application level using some of the existing helpful features in the dataset which includes CookTime, PrepTime, TotalTime, RecipeIngredientParts and RecipeInstructions.

B)Data Pre-Processing

At data pre-processing stage, firstly all the irrelevant features for this project were removed, and made up the data with essential features such as calories, fat content, saturated fat content, etc. Checked for any duplicate values present and removed as found. Verified for any null values present and used appropriate techniques in handling nan values such as replacing with appropriate value, mean imputation, median imputation, frequency imputation and removing null data. Some of the features having outliers, boxplot is used to detect and handle outliers. Different features having different scale of values for normalization used Standard Scalar technique to make it easier for the training and testing.

C) Algorithms

From the reference [2], which is a similar diet recommendation system based for patients details, the authors proposed few algorithm including machine learning and deep learning models which are logistic regression, naïve bayes, Recurrent Neural Network (RNN), Multilayer Perceptron (MLP), Gated Recurrent Units (GRU), and Long Short-Term Memory (LSTM). The authors with the dataset they have they were considering a classification model and by comparing all performance metrics of classification model, they found that LSTM was performed well among all the other mentioned models.

The dataset and the approach that we are following is little different than what the authors of [2] has followed, we are constructing a regression model with the nutrient features that we have in our dataset, so for that reason we are considering the algorithms K-Nearest Neighbors (KNN), Random Forest (RF), Decision Tree (DT), Linear Regression (LR). By verifying with different regression model performance metrics, KNN is a better appropriate algorithm than rest of the other mentioned algorithms. More details about the proposed algorithms as following.

Proposed Algorithms:

K-Nearest Neighbors (KNN):

K-Nearest Neighbors (KNN) is a non-parametric algorithm used for classification and regression tasks. In the context of diet recommendation system, KNN works by identifying the k-nearest neighbors to a given data point based on a similarity measure (e.g., Euclidean distance). In the case of dietary recommendations, it is a simple yet effective algorithm, particularly useful when the underlying data has a clear structure and distinct clusters

Random Forest (RF):

Random Forest (RF) is an ensemble learning method that operates by constructing a multitude of decision trees during training and outputting the mode (classification) or mean prediction (regression) of the individual trees. RF is well-suited for handling high-dimensional data and mitigating overfitting. In the context of our diet recommendation system, RF can capture complex relationships between dietary factors and health outcomes, providing robust recommendations based on these relationships.

Decision Tree (DT):

Decision Trees (DT) are tree-like structures where each internal node represents a "decision" based on a feature, each branch represents the outcome of that decision, and each leaf node represents the final decision or prediction. DT is a simple and interpretable algorithm that can handle both categorical and numerical data. In the context of diet recommendation, DT can be used to create a hierarchy of decision rules based on nutritional content, dietary preferences, and health objectives to recommend suitable diets.

Linear Regression (LR):

Linear Regression involves modeling the relationship between a dependent variable and one or multiple independent variables through a linear approach. It assumes a linear relationship between the variables and aims to find the best-fitting line to describe this relationship. In the context of our diet recommendation system, Linear Regression can help identify linear dependencies between dietary factors (e.g., calorie intake, fat content) and health outcomes (e.g., weight loss, cholesterol levels). It provides insights into how changes in dietary habits may impact health, allowing for personalized recommendations aligned with specific health goals.

D) Training & Testing

For this Project we are taking four machine learning models which can best serve the purpose of diet recommendation, which includes K-Nearest Neighbors (KNN), Random Forest (RF), Decision Tree (DT), and Linear Regression (LR). Our objective is to determine the most effective model for predicting nutrient values in food items based on their nutritional content. We divided our dataset into training and testing sets, utilizing 70% of the data for training and 30% for testing. The features used for training include Calories, FatContent, SaturatedFatContent, CholesterolContent, SodiumContent, CarbohydrateContent, FiberContent, SugarContent, and ProteinContent. Each model was trained using these features to learn the combinations and patterns in the data, aiming to predict the nearest nutritional values for a given food item. Each model was trained using 70% of the dataset, with the nine features mentioned earlier as inputs. During training, the models learned the relationships between the nutritional features to predict the nearest nutrient values for a given food item. The remaining 30% of the dataset was used for testing the trained models. The models predicted the nutrient values for the test data, which were then compared with the actual values to evaluate the performance of each model.

Subsequently, a variety of evaluation metrics including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) are employed to assess the performance of each algorithm by comparing the predicted data by the model against actual data present the data. This meticulous evaluation ensures that our system can accurately predict nutritional content while minimizing errors and optimizing recommendations for individual users.

With this approach our diet recommendation system, ensuring reliable and accurate dietary suggestions tailored to each user's unique needs and preferences.

IV. Experiment Results

A. Comparison Graphs → Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R2) Score.

Mean Absolute Error (MAE): Mean Absolute Error (MAE) quantifies the average absolute variance between the predicted and observed values. It serves as a metric for evaluating the average magnitude of prediction errors, irrespective of their direction. Smaller MAE values signify enhanced predictive accuracy. Performance evaluation of different proposed machine learning models of MAE shown in Figure 4.

Mean Squared Error (MSE): Mean Squared Error (MSE) measures the average squared disparity between the predicted values and the actual values. Squaring the errors penalizes larger errors more heavily than smaller ones. MSE is commonly used in regression tasks and provides a measure of the variance of the errors. Smaller MSE values signify enhanced predictive accuracy. Performance evaluation of different proposed machine learning models of MSE shown in Figure 5.

Root Mean Squared Error (RMSE): RMSE is the square root of the MSE and represents the average magnitude of the errors in the same units as the target variable. It provides a more interpretable measure of the prediction errors compared to MSE. Like MAE and MSE, lower RMSE values indicate better predictive accuracy. Performance evaluation of different proposed machine learning models of RMSE shown in Figure 6

R-squared (R2) Score: The R-squared (R2) Score is a statistical metric that signifies the proportion of the variability in the dependent variable (target) elucidated by the independent variables (features) within the model. It spans from 0 to 1, with 1 signifying a perfect fit. R-squared closer to 1 indicates that the model explains a larger proportion of the variance in the target variable, implying better predictive performance. Performance evaluation of different proposed machine learning models of R-Squared (R2) shown in Figure 7

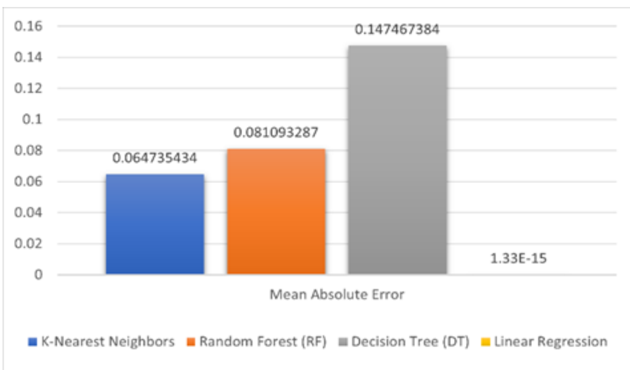

Figure 4. MAE Graph

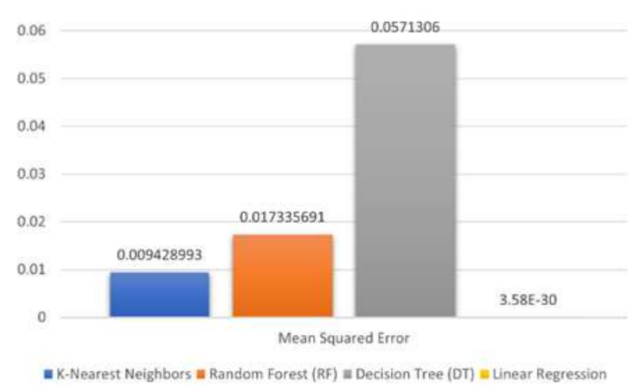

Figure 5. MSE Graph

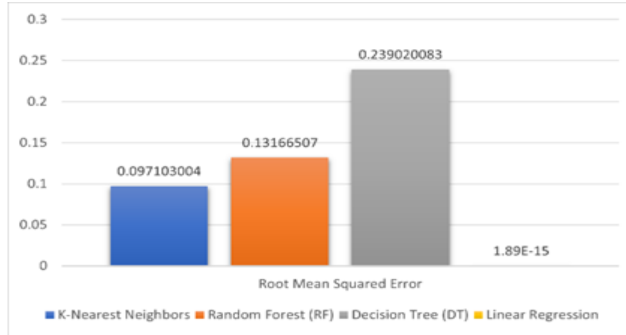

Figure 6. RMSE Graph

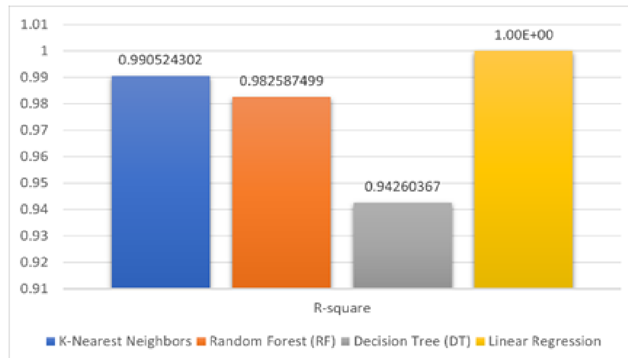

Figure 7. R-squared Graph

Based on the experimental results with different performance metrics including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and R-squared (R2) Score for different machine learning algorithms including K-Nearest Neighbors (KNN), Random Forest (RF), Decision Tree (DT), and Linear Regression (LR). Choosing K-Nearest Neighbors (KNN) is more appropriate for the performance it has, though we have one more algorithm which is LR but it might lead to overfitting. The remining two algorithms RF and DT are performing lesser than KNN.

V. Future Work

While this paper provides comprehensive details for a tailored recommendation system, this work can be continued further by extending current technique of recommendation from content based recommendation to collaborative filtering based recommendation system, which eventually collect users actions in the data their preferences of food category (ex:

people continuously taking vegetarian food most likely vegans, by having this information eliminate recommending food items having meat content), this will take significant time and data usage by users to track and efficiently maintain their preferences.

VI. Conclusion

In conclusion, this project has demonstrated the effectiveness of utilized machine learning algorithm K-Nearest Neighbors (KNN) in developing a tailored diet recommendation system. By leveraging Kaggle dataset "Food.com - Recipes and Reviews." addressed the need for personalized nutrition guidance and generalized nutritious food in today's people food habits where they don't have any awareness of how to get their food with their preferences.

References

[1] Diet, nutrition and the Prevention of Chronic Diseases: Report of a joint WHO/FAO expert consultation, Geneva, 28 January - 1 February 2002 (no date) World Health Organization. At https://www.who.int/publications-detail-redirect/924120916X (Accessed: 09 April 2024).

[2] C. Iwendi, S. Khan, J. H. Anajemba, A. K. Bashir and F. Noor, "Realizing an Efficient IoMT-Assisted Patient Diet Recommendation System Through Machine Learning Model," in IEEE Access, vol. 8, pp. 28462-28474, 2020, doi: 10.1109/ACCESS.2020.2968537.

[3] Shah, M., Degadwala, S. and Vyas, D. (2022) 'Diet recommendation system based on different machine learners', International Journal of Scientific Research in Computer Science, Engineering and Information Technology, pp. 01–10. doi:10.32628/cseit228249.

[4] Lambay, M.A. and Mohideen, S.P. (2022) 'A hybrid approach-based diet recommendation system using ML and Big Data Analytics', Journal of Mobile Multimedia [Preprint]. doi:10.13052/jmm1550-4646.1864.

[5] T. Islam, A. R. Joyita, M. G. R. Alam, M. Mehedi Hassan, M. R. Hassan and R. Gravina, "Human-Behavior-Based Personalized Meal Recommendation and Menu Planning Social System," in IEEE Transactions on Computational Social Systems, vol. 10, no. 4, pp. 2099-2110, Aug. 2023, doi: 10.1109/TCSS.2022.3213506.B. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished.

[6] Chen, C.-H. *et al.* (2018) 'Person—personalized expert recommendation system for optimized nutrition', *IEEE Transactions on Biomedical Circuits and Systems*, 12(1), pp. 151–160. doi:10.1109/tbcas.2017.2760504.

[7] R. Yera Toledo, A. A. Alzahrani and L. Martínez, "A Food Recommender System Considering Nutritional Information and User Preferences,"in IEEE Access, vol.7, pp.96695–96711,2019, doi: 10.1109/ACCESS.2019.2929413.

[8] Sarker, I.H. (2021b) Machine learning: Algorithms, real-world applications and research directions - SN computer science, doi: 10.1007/s42979-021-00592-x

[9] Cheng, D. et al. (2021) 'Why dataset properties bound the scalability of Parallel Machine Learning Training Algorithms', IEEE Transactions on Parallel and Distributed Systems, 32(7), pp. 1702–1712. doi:10.1109/tpds.2020.3048836.

[10] Harris, J.A. and Benedict, F.G. (1918) 'A biometric study of human basal metabolism', Proceedings of the National Academy of Sciences, 4(12), pp. 370–373. doi:10.1073/pnas.4.12.370.

[11] S. Karimulla Basha, T. N Shankar, "Fuzzy logic-based forwarder selection for efficient data dissemination in VANETs," Wireless Networks, The Journal of Mobile Communication, Computation and Information, vol.27,3, pp.2193-2216, Feb 2021,

[12] S. Karimulla Basha, T. N. Shankar, "Fuzzy Logic Based Multi-Hop Broadcasting in High-Mobility VANETs," International Journal of Computer Science and Network Security, vol. 21, no. 3, pp. 165-171, March 2021.