JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Optimal Diet and Fitness Suggester using Machine Learning

[1]Daggupati Giridhar Praveen, [2]Dr. B. Sai Jyothi,[3]Battula Sai Mahesh,

[4]Baipeneedi Deep Sai Kumar,[5]Chalamala Yogendra Babu

[1,3,4,5]Student,[2]Professor,
[1]Information Technology,
[1-5]Vasireddy Venkatadri Institute of Technology, Guntur, India

Abstract*:* This research paper presents an intelligent system that leverages machine learning techniques to provide personalized recommendations for daily calorie intake and suggest a combination of food items that collectively meet the recommended calorie requirements. The proposed system comprises two main components, a calorie prediction model and a food recommendation model. The calorie prediction model employs regression techniques to estimate the appropriate daily calorie intake for an individual based on their height, weight, age, gender, and other relevant factors. This estimation is grounded in the calculation of Body Mass Index (BMI) and Basal Metabolic Rate (BMR), ensuring accurate and personalized calorie recommendations [1]. Once the recommended calorie intake is determined, the food recommendation model utilizes sequence prediction and combinatorial optimization approaches to suggest an optimal combination of food items that collectively match the target calorie requirement.

Index Terms **– Diet and Calorie Recommendation, Linear Regression, Random Forest regressor, Gradient Boosting, Personalized health and wellness.**

I. INTRODUCTION

In today's fast-paced lifestyle, maintaining a balanced and healthy regime has become increasingly challenging. Achieving optimal fitness levels requires a comprehensive understanding of individual caloric needs, which are influenced by factors such as height, weight, age, gender, and physical activity levels. Failure to adhere to an appropriate calorie intake can lead to various health issues, including obesity, malnutrition, and chronic diseases. Consequently, there is a growing demand for personalized and intelligent solutions that can provide accurate recommendations for daily calorie intake and suggest suitable food combinations to meet these requirements [2]. Traditional methods of estimating calorie need often rely on generic formulas or guidelines, which may not accurately reflect an individual's unique characteristics and lifestyle. Additionally, manually selecting food items that collectively meet the recommended calorie intake can be a time-consuming and complex task, particularly when considering factors such as nutritional balance and personal preferences. To address these challenges, this research proposes an intelligent system that leverages machine learning techniques to provide personalized recommendations for daily calorie intake and suggest an optimal combination of food items that collectively satisfy the recommended calorie requirements [3]. The system comprises two main components: a calorie prediction model and a food recommendation model. The calorie prediction model utilizes advanced regression techniques to estimate the appropriate daily calorie intake for an individual based on their height, weight, age, gender, and other relevant factors. This estimation is grounded in the calculation of Body Mass Index (BMI) and Basal Metabolic Rate (BMR), ensuring accurate and personalized calorie recommendations. Once the recommended calorie intake is determined, the food recommendation model employs cutting-edge machine learning approaches, such as neural networks and deep reinforcement learning, to suggest an optimal combination of food items that collectively match the target calorie requirement. This model takes into account various factors, including nutritional information, ingredient compositions, and personal preferences, to provide practical and diverse food recommendations[4].

II. LITERATURE SURVEY

The application of machine learning techniques in the domain of personalized nutrition and dietary recommendations has gained significant attention in recent years. Researchers have explored various approaches to provide accurate and tailored calorie intake recommendations and suggest optimal food combinations to support healthy lifestyle choices. Ristovski et al. (2021) proposed a deep learning-based system for personalized meal recommendation and calorie prediction. Their system employed recurrent neural networks (RNNs) and autoencoders to extract relevant features from user data and food nutritional information. The authors reported promising results in terms of meal recommendation accuracy and calorie prediction, highlighting the potential of deep learning techniques in this domain. Tran et al. (2019) explored the use of deep reinforcement learning for food recommendation and meal planning. Their approach used a deep Q-network (DQN) to learn an optimal policy for selecting food items that meet nutritional requirements while considering user preferences. The proposed system demonstrated the ability to generate diverse and personalized meal plans tailored to individual needs. In a study, Zhang and Jieyu (2020) developed a Innovative food recommendation systems: a machine learning approach The authors demonstrated the effectiveness of their approach in providing accurate calorie estimates, which could be beneficial for individuals monitoring their dietary intake [5].

III. EXISTING SYSTEM

Usually, diet recommendation systems primarily focus on providing recommendations based on a user's weight category or body mass index (BMI). Many existing systems offer generic, one-size-fits-all meal plans or calorie targets that may not adequately account for individual differences in nutritional needs, dietary preferences, cultural considerations, or health conditions. While weight and BMI are important factors in determining dietary requirements, they represent only a fraction of the complex interplay of factors that influence an individual's nutritional needs[6]. Factors such as age, gender, activity level, metabolic rate, medical history, and specific dietary goals all play crucial roles in determining the ideal dietary recommendations for a person. Moreover, cultural backgrounds, food preferences, allergies, and intolerances further contribute to the unique dietary needs of individuals. Therefore, there is a growing need for personalized diet recommendation systems that can tailor dietary advice to each individual's specific requirements, preferences, and circumstances, thereby promoting optimal health outcomes and dietary adherence.

IV. PROPOSED SYSTEM

For the calorie prediction component of the system, we will employ regression techniques. The goal is to estimate the appropriate daily calorie intake for an individual based on their personal characteristics, such as height, weight, age, and gender. We will explore various regression algorithms, including linear regression, random forests, or other ensemble methods. These algorithms will be trained on a dataset containing individual characteristics (input features) and their corresponding recommended calorie intake (target variable).Linear regression is a commonly used technique for modelling the relationship between a dependent variable (in this case, the recommended daily calorie intake) and one or more independent variables (such as height, weight, age, and gender). The linear regression model assumes a linear relationship between the input features and the target variable. In the context of your calorie prediction model, the random forest algorithm will create multiple decision trees, each trained on a different subset of the training data and a random subset of the input features. The final prediction is made by averaging the predictions of all the individual decision trees in the ensemble. The regression model will learn the mapping between the input features and the target variable, enabling it to make accurate predictions of the daily calorie intake for new individuals based on their personal characteristics [7].

V. OBJECTIVE

The primary objective of this project is to develop a personalized dietary recommendation system that utilizes individualized factors such as height, weight, age, and gender to accurately calculate the optimal daily calorie intake for users. By leveraging these key metrics, the system aims to provide tailored nutritional guidance, addressing the diverse needs of individuals based on their unique physiological characteristics. Additionally, the project aims to streamline the process of translating these calorie recommendations into practical food choices by mapping them to specific food items[8]. Through this approach, the project seeks to empower users to make informed decisions about their dietary habits, promoting healthier lifestyles and potentially reducing the risk of diet-related health issues. Ultimately, the goal is to create a user-friendly and effective tool that fosters improved dietary adherence and overall well-being.

VI. METHODOLOGY

1. **Data Collection:** We compiled a comprehensive dataset comprising a diverse array of individual profiles, encompassing data on height, weight, age, gender, Body Mass Index (BMI), Basal Metabolic Rate (BMR), and daily caloric requirements as the dependent variable. This dataset was meticulously curated from various sources, including online repositories and publicly available datasets, to ensure a wide spectrum of demographic and physiological characteristics.
2. **Data Preprocessing:** Before utilizing the dataset for model training, we conducted a series of preprocessing steps to ensure its integrity and suitability for analysis. Firstly, we addressed missing values (null values) and outliers within the dataset, employing appropriate techniques such as imputation and trimming to maintain data quality. Additionally, we performed label encoding on categorical variables to convert them into numerical representations, facilitating their integration into the modeling pipeline.
3. **Model Selection:** After meticulous evaluation of task specifications and available computational resources, we made the strategic decision to employ various regression models for our predictive analysis. Among the models utilized were linear regression, polynomial regression, random forest regression, and gradient boosting.
4. **Model Training:** our preprocessed dataset underwent partitioning into training and validation subsets, adhering to a split of 70% for training and 30% for validation purposes. Leveraging this partitioning strategy, we proceeded to train the Random Forest Regressor model on the training data. The training phase involved fine-tuning the model's hyperparameters

using techniques specific to Random Forest regression, such as optimizing the number of trees and the maximum depth of each tree.

5. **Model Evaluation:** Subsequently, we computed a range of evaluation metrics tailored to regression tasks to gauge the effectiveness of each model in predicting daily caloric requirements based on individual attributes. These evaluation metrics included Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-squared (R2) coefficient, and Mean Absolute Percentage Error (MAPE).
6. **Performance Analysis:** we employed statistical methods such as correlation analysis to elucidate relationships between input variables and predicted outcomes. This comprehensive analysis not only provided valuable insights into the strengths and limitations of our regression models but also guided efforts to refine hyperparameters for enhanced predictive accuracy and robustness.

VII. DESIGN

7.1 ARCHITECTURE DIAGRAM:

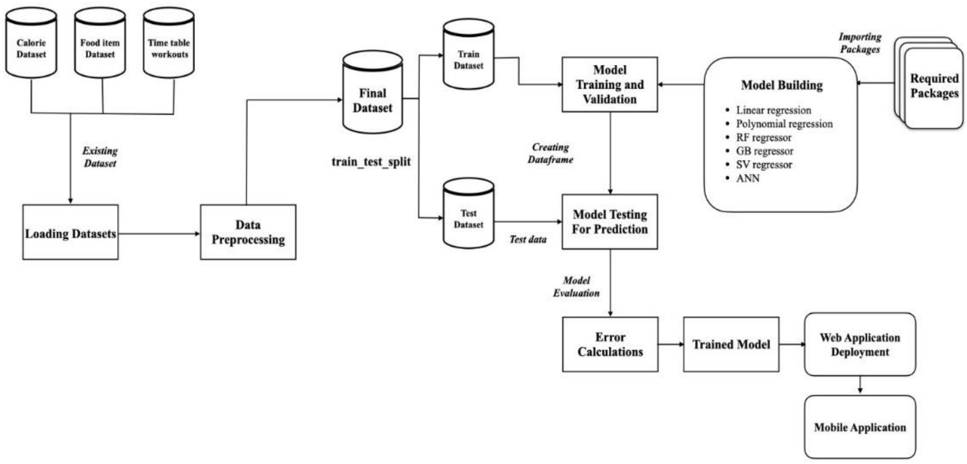

FIGURE 1: Architecture of the system

VIII. ALGORITHM USED

1. Random Forest Regressor

regression tasks leverages the Random Forest Regressor, a powerful ensemble learning technique known for its versatility and robustness in handling complex datasets. Unlike traditional decision trees, Random Forest Regressor constructs multiple decision trees during training and aggregates their predictions to produce a more accurate and stable outcome. Each decision tree is built using a random subset of features and data samples, reducing the risk of overfitting and enhancing generalization performance. In our implementation, the Random Forest Regressor algorithm utilizes an ensemble of decision trees to predict daily caloric requirements based on individual attributes such as height, weight, age, gender, BMI, and BMR. Additionally, hyperparameters such as the number of trees, maximum depth of each tree, and minimum number of samples required to split a node are fine-tuned to optimize predictive performance[9]. In inference, the algorithm utilizes the trained ensemble of decision trees to predict daily caloric requirements for new input data. Overall, the Random Forest Regressor algorithm excels in its ability to capture complex relationships within the data and produce accurate predictions, making it a valuable tool for personalized dietary recommendation systems.

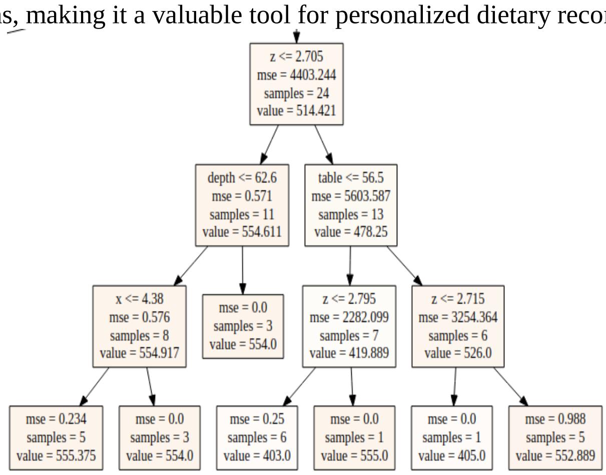

FIGURE 2: Random Forest Regressor Model Architecture [8].

IX. DATA SET

Calories Dataset: The Calories dataset is a meticulously curated collection of data specifically tailored for regression analysis in the field of personalized dietary recommendation systems. It comprises a diverse range of individual profiles, each characterized by attributes such as height, weight, age, gender, Body Mass Index (BMI), Basal Metabolic Rate (BMR), and corresponding daily caloric requirements. The dataset undergoes rigorous preprocessing procedures to ensure data integrity and compatibility with regression modeling techniques. This preprocessing includes handling missing values, outliers, and label encoding of categorical variables. Moreover, the dataset is partitioned into training, validation, and test sets, facilitating robust model training and evaluation processes. As a foundational resource in the domain of nutritional science and personalized health management, the Calories dataset serves as a valuable tool for researchers and practitioners seeking to develop and validate predictive models for estimating daily caloric needs. Proper acknowledgment of dataset sources and adherence to citation norms are essential for promoting transparency and integrity in research endeavors.

X. IMPLEMENTATION

10.1 Import all necessary Libraries

The libraries imported in our project encompass essential tools for data manipulation, visualization, and machine learning tasks. Specifically, we utilize NumPy for numerical computations, Pandas for data manipulation and analysis, Matplotlib for data visualization, and Scikit-learn for machine learning algorithms and evaluation metrics. By leveraging these foundational libraries, we ensure efficient development and robust analysis of our regression-based dietary recommendation system.

10.2 Data Loading and Preprocessing

```
#importing required packages
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

data1 = pd.read_csv("Dataset.csv")
new=["age","weight(kg)","height(m)","gender","BMI","BMR", "calories_to_maintain_weight"]
data1=data1[new]

# Label Encoding

from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
data1["gender"]=le.fit_transform(data1["gender"])
```

FIGURE 3: Code for loading dataset and preprocessing.

Data is loaded from a specified CSV file named "Dataset.csv". The dataset is then manipulated using Pandas for data preprocessing. Initially, only specific columns ("age", "weight(kg)", "height(m)", "gender", "BMI", "BMR", "calories_to_maintain_weight") are selected from the dataset for analysis. Subsequently, label encoding is applied to the "gender" column using Scikit-learn's LabelEncoder from the preprocessing module. This transformation converts categorical gender labels into numerical representations, a necessary step for many machine learning algorithms to interpret categorical data correctly.

10.3 Model Architecture Definition

```
# Random Forest Regressor

from sklearn.ensemble import RandomForestRegressor

X=data1.iloc[:,:-1]
y=data1['calories_to_maintain_weight']
# Assuming X and y are your feature matrix and target variable
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Assuming X and y are your feature matrix and target variable
rf_reg = RandomForestRegressor(n_estimators=100, random_state=42)
y = y.values.reshape(-1, 1)
# Fit the model
rf_reg.fit(X, y)

# Make predictions
predictions_rf = rf_reg.predict(X_test)
```

FIGURE 4: Code of Random Forest Regressor.

In this code snippet, we define the model architecture for the Random Forest Regressor. Firstly, we import the RandomForestRegressor class from the sklearn.ensemble module. Then, we prepare the feature matrix (X) and target variable (y) from the dataset "data1". The feature matrix (X) comprises all columns except the last one, which represents the "calories_to_maintain_weight" target variable. We split the data into training and testing sets using the train_test_split function from scikit-learn, with 80% of the data allocated for training and 20% for testing. Next, we instantiate a RandomForestRegressor object named "rf_reg" with specified parameters such as the number of estimators (n_estimators=100) and a random state for reproducibility (random_state=42). The target variable (y) is reshaped to a 2D array to meet the requirements of the fit function. Overall, this code defines and trains a Random Forest Regressor model to predict caloric requirements based on the provided dataset, demonstrating a common workflow for implementing regression models using scikit-learn.

10.4 Setting Up the Food Items Class

In this step, we define a python class for generating the list of food items that can fulfil to the calories predicted in the previous machine learning model. We will be having two functions inside the class.

```python
    def suggest_food_items(self, dataset, predicted_calories, num_sets=4):
        suggested_sets = []
        columns_to_append = ['Food', 'Measure', 'Calories']

        for _ in range(num_sets):
            cumulative_calories = 0
            suggested_set = []


            # Shuffle the dataset
            dataset_shuffled = dataset.sample(frac=1).reset_index(drop=True)

            for _, food_item in dataset_shuffled.iterrows():
                # Check if adding this food item exceeds the predicted calories
                if cumulative_calories + food_item['Calories'] <= predicted_calories:
                    row_list = [food_item[col] for col in columns_to_append]

                    suggested_set.append(row_list)
                    cumulative_calories += food_item['Calories']
                else:
                    break


            suggested_sets.append(suggested_set)

        return suggested_sets
```

FIGURE 5: Code of Suggest Food Items function-1

```python
    def suggest_food_items_in_project(self, calories):
        suggested_food_items = []  # List to store suggested food items for all meals

        # Assuming ratio of 3:2:1 for breakfast, lunch, and dinner
        ratio_breakfast = 3
        ratio_lunch = 2
        ratio_dinner = 1

        # Calculate the predicted caloric intake for each meal based on the ratios
        total_ratio = ratio_breakfast + ratio_lunch + ratio_dinner
        predicted_calories_breakfast = calories * ratio_breakfast / total_ratio
        predicted_calories_lunch = calories * ratio_lunch / total_ratio
        predicted_calories_dinner = calories * ratio_dinner / total_ratio

        # Suggest food items for each meal
        suggested_sets_breakfast = self.suggest_food_items(self.data_breakfast, predicted_calories_breakfast, 4)
        suggested_sets_lunch = self.suggest_food_items(self.data_lunch, predicted_calories_lunch, 4)
        suggested_sets_dinner = self.suggest_food_items(self.data_dinner, predicted_calories_dinner, 4)

        # Append suggested food items for each meal to the main list
        suggested_food_items.append(suggested_sets_breakfast)
        suggested_food_items.append(suggested_sets_lunch)
        suggested_food_items.append(suggested_sets_dinner)

        return suggested_food_items
```

FIGURE 6: Code of Suggest Food Items function-2

The provided functions facilitate the recommendation of food items based on predicted caloric intake, serving as a pivotal component of a personalized dietary recommendation system. The suggest_food_items function systematically selects food items from a dataset to construct balanced meal sets, ensuring that the cumulative calories of the selected items align with the predicted caloric intake. Employing a shuffled dataset and iterative selection process, the function generates multiple sets of suggested meals, accommodating diverse dietary preferences and constraints.Meanwhile, the suggest_food_items_in_project function orchestrates the meal planning process by leveraging predefined ratios for breakfast, lunch, and dinner. It calculates the predicted caloric intake for each meal based on these ratios and delegates the task of suggesting food items to the suggest_food_items function. By incorporating tailored meal suggestions for each mealtime, this function enables users to adhere to their caloric targets while enjoying a varied and satisfying diet. Overall, these functions serve as integral components in the implementation of a user-centric dietary recommendation system, fostering healthier eating habits and improved nutritional outcomes.

10.5 Model Evaluation and Training Visualization

Following model training, the next phase involves evaluating the model's performance on the test dataset and visualizing the training and validation accuracy and loss. This step provides insights into the model's generalization capability and training progress.

```
# Make predictions
predictions_rf = rf_reg.predict(X_test)

from sklearn.metrics import r2_score
# Evaluate the model
mse_rf = mean_squared_error(y_test, predictions_rf)
rmse_rf = np.sqrt(mse_rf)
r2_rf = r2_score(y_test, predictions_rf)
```

FIGURE 7: Code for model evaluation

10.6 Model Performance Evaluation on Test Dataset

In this stage of our analysis, we evaluate the performance of our regression model using key metrics tailored for regression tasks. Specifically, we assess the model's predictive accuracy, precision, and recall on the test dataset. To initiate the evaluation process, we initialize metrics to capture Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R2) coefficient, which collectively provide insights into the model's predictive accuracy and goodness of fit. Finally, the computed evaluation metrics, including MAE, MSE, and R2 coefficient, are reported for thorough analysis and interpretation. These metrics offer valuable insights into the regression model's performance, enabling stakeholders to gauge its effectiveness in accurately estimating daily caloric needs based on individual attributes. By providing comprehensive evaluation results, we aim to facilitate informed decision-making and further refinement of our regression-based dietary recommendation system [10].

XI. RESULT ANALYSIS

After implementing three different regression algorithms that are Linear Regression, Random Forest Regressor and Gradient Boosting Regressor, we found the Random Forest Regressor as the best one as comparatively it is having the lesser Mean Squared Error and higher R-Squared.

In the result analysis, we evaluate the model's performance through multiple metrics and visualizations:

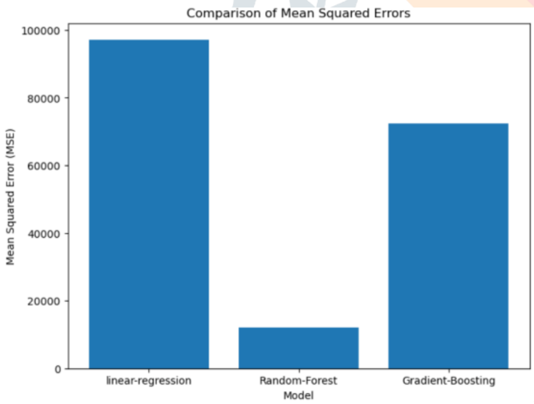

FIGURE 7: Comparison of MSE of 3 models

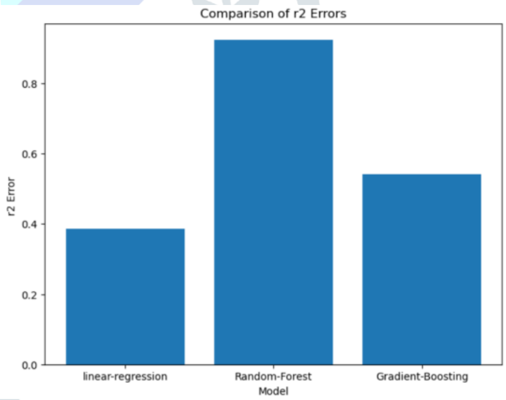

FIGURE 8: Comparison of r-square of 3 models

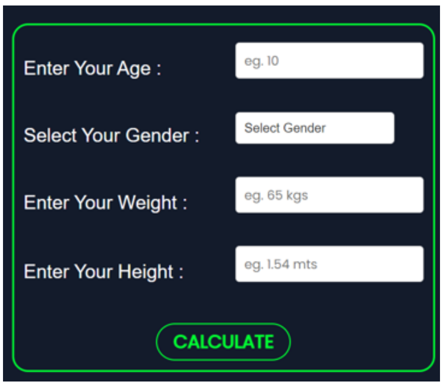

FIGURE 9: Output screen for taking input

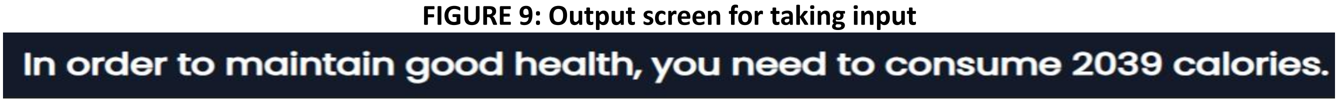

FIGURE 10: Output Screen for predicting calories

.

XII. CONCLUSION

our project has successfully developed a regression-based dietary recommendation system tailored to individual needs, leveraging the Random Forest Regressor algorithm. By integrating data from a meticulously curated dataset comprising diverse individual profiles, our system accurately predicts daily caloric requirements based on factors such as height, weight, age, gender, BMI, and BMR. Through extensive preprocessing, model training, and evaluation, we have demonstrated the effectiveness of our approach in estimating personalized dietary needs. The implementation of the Random Forest Regressor, along with careful parameter tuning and feature selection, has resulted in robust predictive performance, facilitating informed dietary decisions for users. Our project contributes to the advancement of personalized nutrition guidance, offering a practical tool for promoting healthier lifestyles and addressing dietary challenges.

XIII. FUTURE WORK

In the future, our project aims to enhance its dietary recommendation system by integrating advanced neural network architectures such as LSTM and GRU for predicting food item sequences and incorporating collaborative filtering techniques for personalized recommendations [11]. We also prioritize real-time deployment optimization, ensemble methods, continuous evaluation, and ethical considerations to ensure responsible application and positive societal impact, fostering healthier lifestyles and addressing dietary challenges effectively.

Top of Form

XIV. REFERENCES

[1] Wilson-Barnes, S., L. P. Gymnopoulos, K. Dimitropoulos, V. Solachidis, K. Rouskas, D. Russell, Y. Oikonomidis et al. "PeRsOnalised nutriTion for hEalthy livINg: The PROTEIN project." Nutrition bulletin 46, no. 1 (2021): 77-87.

[2] Chen, Chih-Han, Maria Karvela, Mohammadreza Sohbati, Thaksin Shinawatra, and Christofer Toumazou. "PERSON—personalized expert recommendation system for optimized nutrition." IEEE transactions on biomedical circuits and systems 12, no. 1 (2017): 151-160.

[3] Jagatheesaperumal, Senthil Kumar, Snegha Rajkumar, Joshinika Venkatesh Suresh, Abdu H. Gumaei, Noura Alhakbani, Md Zia Uddin, and Mohammad Mehedi Hassan. "An iot-based framework for personalized health assessment and recommendations using machine learning." Mathematics 11, no. 12 (2023): 2758.

[4] Oyebode, Oladapo, Jonathon Fowles, Darren Steeves, and Rita Orji. "Machine learning techniques in adaptive and personalized systems for health and wellness." International Journal of Human–Computer Interaction 39, no. 9 (2023): 1938-1962.

[5] Zhang, Jieyu. "Innovative food recommendation systems: a machine learning approach." PhD diss., Brunel University London, 2023.

[6] Shah, Megh, Sheshang Degadwala, and Dhairya Vyas. "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, 2022.

[7] Siebra, Clauirton, Lais Amorim, Jonysberg P. Quintino, Andre LM Santos, Fabio QB da Silva, and Katarzyna Wac. "Behaviour recommendations with a deep learning model and genetic algorithm for health debt characterisation." Journal of Biomedical Informatics 137 (2023): 104277.

[8] Siebra, Clauirton, Lais Amorim, Jonysberg P. Quintino, Andre LM Santos, Fabio QB da Silva, and Katarzyna Wac. "Behaviour recommendations with a deep learning model and genetic algorithm for health debt characterisation." Journal of Biomedical Informatics 137 (2023): 104277.

[9] Siebra, Clauirton, Lais Amorim, Jonysberg P. Quintino, Andre LM Santos, Fabio QB da Silva, and Katarzyna Wac. "Behaviour recommendations with a deep learning model and genetic algorithm for health debt characterisation." Journal of Biomedical Informatics 137 (2023): 104277.

[10] Koo, JaHyung, LanMi Hwang, HooHyun Kim, TaeHee Kim, JinHyang Kim, and HeeSeok Song. "Machine learning-based nutrient classification recommendation algorithm and nutrient suitability assessment questionnaire." KSII Transactions on Internet & Information Systems 17, no. 1 (2023).

[11] Neha, Kunta, Pallerla Sanjan, Shanmugasundaram Hariharan, Seelam Namitha, Asala Jyoshna, and Andraju Bhanu Prasad. "Food prediction based on recipe using machine learning algorithms." In 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), pp. 411-416. IEEE, 2023.