EMPLOYING MACHINE LEARNING FOR FOOD RECOMMENDATION

Sejal gupta
Dept. of computer sci & engg
Galgotias university
Noida, uttar Pradesh

Harsh jaiswal
Dept. of computer sci & engg
Galgotias university
Noida, uttar pradesh

Sahil kumar
Dept. of computer sci & engg
Galgotias university
Noida, uttar pradesh

Abdul Aleem
Associate Professor
Dept. of computer sci & engg
Galgotias university
Noida, uttar pradesh

ABSTRACT

This paper introduces a machine learning based diet recommendation system aiming to suggests what foods people should eat based on their needs. It tries to understand and help each person with what they should eat to stay healthy. Beyond considering user-specific factors such as their disease and other details like age, gender, weight, and the nutrients in their food, such as calories, protein, fat, sodium, fiber, and cholesterol, the system uses BMI (Body Mass Index) as an important measure to check the user is underweight, overweight or obesity. The machine learning techniques used here for recommendation are

k means and random forest. K-means and Random Forest are used to handle details about what people eat, their health signals,

and what they prefer. K-means puts similar eating habits in order, and Random Forest helps in suggesting things accurately based on this organized information. Both methods work together to give personalized and accurate advice about what's best for each person's diet.

1.INTRODUCTION

introduction According to WHO, since 1975, the number of people dealing with obesity worldwide has shot up a lot. In 2016, more than 1.9 billion adults (18 years and older) were facing weight issues. Out of this massive number, over 650 million were dealing with obesity. Think about it, 39% of adults were too heavy, and 13% were in the obese category. What's surprising is that in many places, more people die because of being overweight or obese than because of being underweight. But here's a positive point that we can prevent obesity, underweight or overweight problem just by taking a balanced and diverse diet that meets the body's energy, protein, vitamin, and mineral requirements. Even though many people know eating healthy is crucial, but life gets too busy that they don't want to think too hard about making meals. And here this system play an important role by helping people to switch up their eating habits and pick healthier foods. Its primary aim is to offer personalized dietary guidance, suggesting foods that match each person's unique needs, ultimately promoting a healthier way of life. Utilizing

advance technology and an in-depth comprehension of individual factors like health conditions, age, gender, and nutritional data encompassing calories, protein, fat, sodium, fiber, and cholesterol content, the system tries to provide personalized dietary needs. Within this system, Body Mass Index (BMI) serves as a crucial measure in evaluating an individual's health status, determining whether they fall within a healthy weight range or face issues like being underweight, overweight, or obese. Alongside BMI, the system employs complex machine learning techniques such as the K-means clustering and Random Forest algorithms. These methods form the core of the recommendation system, efficiently organizing and understanding dietary habits and health markers. K-means clustering assumes a central role in categorizing and arranging dietary patterns, identifying clear similarities among food preferences, thereby contributing to the construction of a structured dietary database. Simultaneously, the Random Forest algorithm enhances the accuracy and personalization of suggestions. By merging information about a person's food likes, health status, and nutritional necessities, the system crafts tailored recommendations, bridging the gap between personalized dietary guidance and optimal health results. The cooperative interplay between K-means clustering and Random Forest techniques enables the system to furnish precise and personalized dietary suggestions. Through amalgamating comprehensive dietary insights and health indicators, the system endeavors to offer thorough and accurate recommendations tailored to meet the distinct dietary needs of each individual.

1.1 PROBLEM STATEMENT

The number of people eating more and more fast food is really alarming, they are getting various health issues such as diabetes, obesity, and increase in blood pressure etc. Hence it has become very essential for people to have a good balanced nutritional healthy diet. But in this fast pace generation not everyone has the time and money to spend on personal dietitian and nutrition who will look upon and take care of their health by advising them a healthy diet plan according to the individual personal information. The food recommendation system will suggest them balanced diet food, which will give a satisfactory result for healthy life.

2.OBJECTIVES

1.The objective to study this system is to consider the important aspects of the user’s lifestyle & make sure that these factors are incorporated while the system is working on the solution to propose a recommended healthy and nutritious diet to the user.

2.A healthy lifestyle can live only by intaking a good nutritious healthy diet & doing a moderate amount of physical activity like Yoga & Playing Outdoor Games. A Healthy Diet & Physical Activity can surely help in maintaining a healthy weight.

3.A Good nutritious healthy diet can benefits more than just managing the healthy weight. A good nutritious diet keeps your vitals good which give you strength and you feel joyous & fall sick very rarely.

3.EXISTING SYSTEM

There are several works proposed on recommendations system for diet and food. These systems are used for recommending the foods in different terms like Food Recommendations, Menu Recommendations, Diet Plan Recommendation, Recipe Recommendation & Health Recommendations for Specific Diseases. Majority of these system works on the preferences of the user extracted from different sources like user rating and online food bills ordered by the user.

Earlier the systems were only designed to provide the list of food for a healthy and safe

diet and the insulin of diabetic patients change on the hourly basis and they cannot eat the recommended food and in some cases the user may be allergic to the food recommended to the user.

The systems were proposed to provide breakfast, lunch, dinner and they started to provide the ingredients for the food they want to eat and the number of calories they need to take to maintain a healthy and balanced lifestyle . they were also given a preferrable option that they want to lose weight or they want to gain weight so the recommendation system will suggest those food which will be low in carbs to reduce body weight and suggest those foods which are rich in carbohydrates so that a user can gain body weight.

This research introduces those techniques that will neglect the food that a user is allergic to, it would not be suggested to the user. Only those foods will be recommended which will be highly acceptable by the user's body.

In some case the food suggested to a diabetic patient may lead to a serious concern as mentioned earlier that the insulin level changes on an hourly basis. So, now the foods will have the option of when the user has eaten their last meal and what kind of meal, it will help the food recommendor to predict a more accurate food for the user.

Maintaining a healthy lifestyle with ease will be available to any user without hiring a personal dietician and it will be more easy to cook those meals because the ingredients of the food will be provided to you and the time taken to prepare that will also be mentioned.

4.PROPOSED SYSTEM

4.1 K-MEANS CLUSTERING

The k means algorithm is a unsupervised machine learning algorithm and an iterative algorithm that tries to partition the dataset into pre-defined distinct nonoverlapping subgroups (clusters) where each data point belongs to only one group. It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training.

It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters such that the sum of the squared distance between the data points and the cluster's centroid is at the minimum.

4.1.1 EQUATION

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - C_j \right\|^2$$

This is the formula of squared error function :

Where-

J=objective function

K=number of clusters

n=number of cases

x_i=case i

c_j=centroid for cluster j

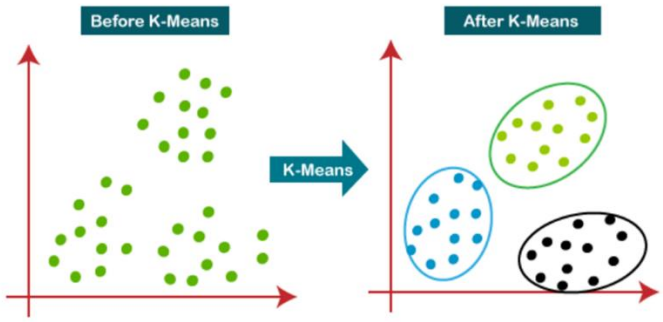

Fig-1 : K-means Algorithm

4.2RANDOM FOREST ALGORITHM

Random Forest algorithm is CLASSIFIED UNDER Supervised classification algorithm in MACHINE LEARNING. In this algorithm we

create a forest in which the trees should be random , but we don't perform the algorithm in reference to the decision tree, we don't calculate the information gain or gini index in this case , we make random trees and the length of the tree will decide the accuracy of the prediction If you input a training dataset with targets and features into the decision tree, it will formulate some set of rules. These rules can be used to perform predictions. . There is a direct relationship between the number of trees in the forest and the results it can get.

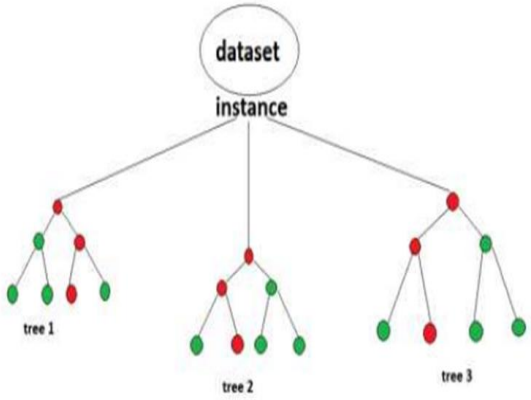

Fig-2: Random Forest algorithm

4.3 **K-NEAREST NEIGHBORS ALGORITHM**

KNN is a machine learning algorithm to find clusters of similar users based on and make predictions using

cosine similarity. We can understand it's working with the help of following steps:

I. For actualizing any calculation, we require an information set. So amid the primary step of KNN, we must stack the preparing as well as test data.
II. Next, we have to select the esteem of K i.e. the closest information focuses. K can be any integer.
III. For each point within the test data do the following –
i. Calculate the remove between test information and each push of preparing data with the assistance of cosine similarity.
ii. Now, based on the removed esteem, sort them in rising order.
iii. Next, it'll select the best K lines from the sorted array.
iv. Now, it'll relegate a lesson to the test point based on the foremost visit lesson of these rows.

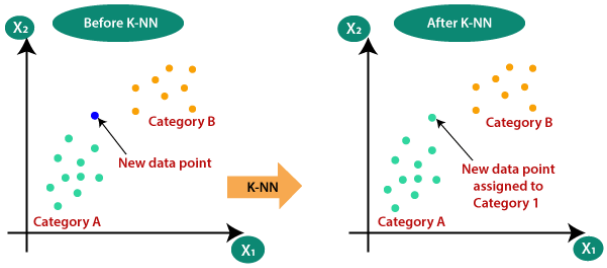

Fig-3: KNN Algorithm

4.4 Gaussian Naïve Bayes

probability and statistics based on the application of Bayorama Teorama [7]. The Bayorama Teorama equation is as follows [8].

P(H|E) : The conditional probability of a hypothesis H occurs if the evidence is provided. P(E|H) : The probability that a proof E occurs will affect the hypothesis H.

P(H) : The initial (prior) hypothesis of H hypothesis occurs regardless of any evidence. P(E) : The initial probability (prior) of evidence E occurs regardless of the hypothesis / other evidence.

4.5 Linear Regression

Whenever we've gained information with various factors, one vital inquiry is the manner by which the factors are connected. For instance, we could want the connection between individuals' loads and levels, or review time and grades. Relapse is a bunch of procedures for assessing connections, and we'll zero in on them in our framework. In this paper, we'll zero in on finding one of the most straightforward kinds of relationship: direct. This process is unsurprisingly called linear regression, and it has many applications. In this project the dependent feature is BMI and independent features are weight (kg) and height (m*2).

4.6 Support vector Machine

By and large the utilization of AI calculation is utilized for the suggestion. In this paper, the help vector machine will recommend the eating regimen. SVM is a managed AI

calculation which works in view of the idea of choice planes that characterizes choice limits. A decision boundary separates the objects of one class from the object of another class. Support vectors are the data points which are nearest to the hyper-plane as shown in figure5. Kernel function is used to separate non-linear data by transforming input to a higher dimensional space. Gaussian radial basis function kernel is used in our proposed method.

5.IMPLEMENTATION AND DESIGN

5.1 User flow

The user flow will happen in various steps like :

- The user will request food by entering the details like weight, age and other asked details.
- Then the machine learning model will use the k-means clustering algorithm to cluster the foods to be recommended to the user .
- After that the user will be recommended food according to the daytime or nighttime by predicting the food by random forest algorithm .

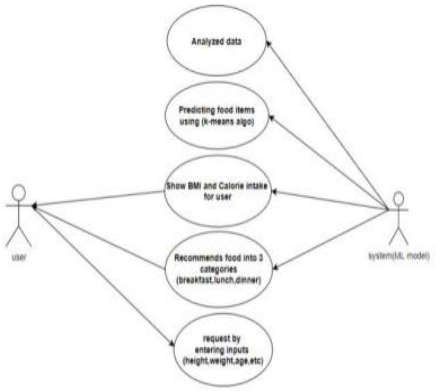

Fig- 4 : UR Diagram

5.2 SYSTEM ARCHITECTURE

In this, we will be dividing our food under several categories like breakfast , lunch , snacks , dinner. And we will then recommend food according to our user needs which would be healthy too.

1. The system will ask the necessary information about the user like their Name, Age, Gender, Height & Weight etc.

2.Now, the details will be going through the ML model in following ways:

2.1.ML Model will use K-Means algorithm for clustering to classify the food on the basis of calories.

2.2.Now, ML Model will use Random Forest Classifier to classify the food and do predict the food based on the input.

3.Now, after the analysis of all the data the system model will give response by showing the user's BMI & their corresponding health state like Overweight, Healthy Or Fit.

4.Then the system will be recommending diet to the user into four categories (Breakfast, Lunch, Evening Snacks & Dinner) based on their input.

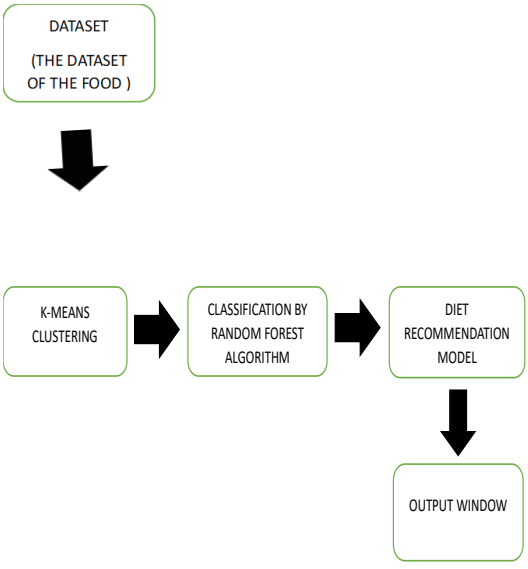

Fig – 5: System workflow

6.RESULT & ANALYSIS

The result will provide nutritious food and help in maintaining the healthy lifestyle of a person . here , we will have a input window where the user has to input some values into it and then suggest food according to their needs .

The output window will also show the ingredients and the time taken to prepare that food and also the calories required.Finally, the random forest algorithm was selected as the most suitable algorithm for the prediction system according to the analysis.

We did a initial grid search which was performed to choose the right set of hyperparameters.

	model	best_score	best_params
0	random_forest	0.770909	{'max_depth': 24}
1	kNeighbor	0.767273	{'n_neighbors': 4}
2	decision_tree	0.760000	{'max_depth': 1, 'min_samples_leaf': 1}
3	gradient_boosting	0.760000	{'n_estimators': 1}
4	SVC	0.760000	{'C': 1, 'gamma': 'scale'}
5	GaussianNB	0.345455	{}
6	LogisticRegression	0.738182	{'C': 1, 'solver': 'newton-cg'}

Fig-6 : Results in initial grid search

A line plot is made to compare the mean accuracy scores to the leave-one-out cross-validation result, with error bars used to show the minimum and maximum values for each result distribution. The results suggest that k=10 alone might be a bit overly optimistic and that k=11 would be a more realistic estimate in Random forest.

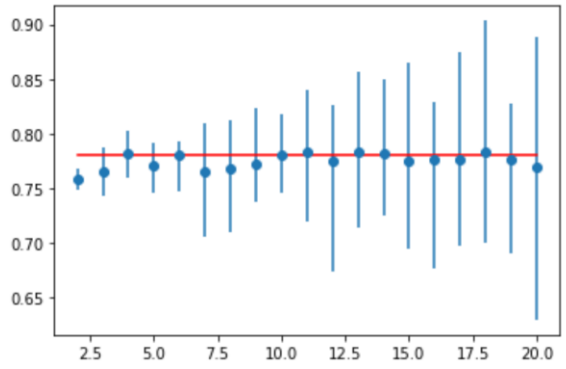

Fig-7 : Variation of mean accuracy for cross-validation K-values with error bars (Blue) vs. the ideal case (red) in Random forest.

The below plot shows how the classification accuracy varied according to the number of trees in the Random forest. These types of analyses were done for all the algorithms with different hyperparameters to select the most suitable algorithm with optimal hyperparameters.

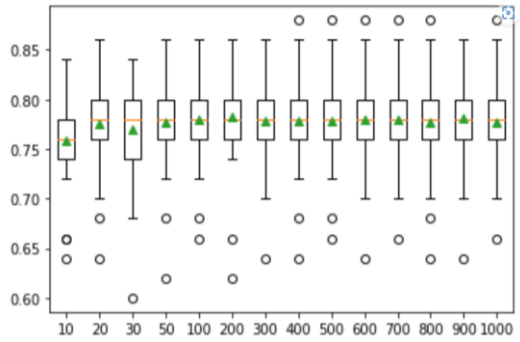

Fig-8: Number of trees vs. classification accuracy in Random forest

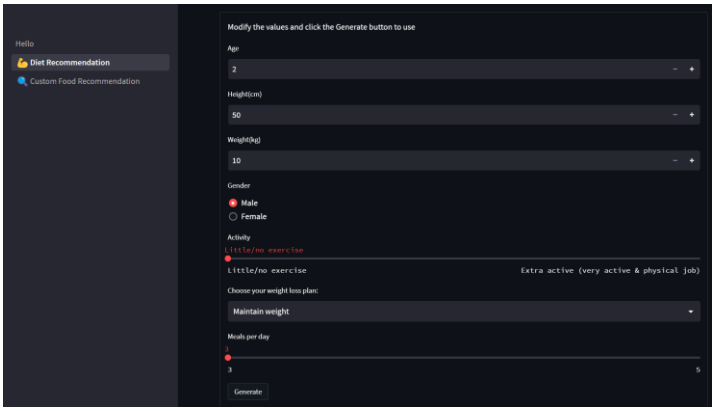

Fig -9 : Input window

Fig-10: Output window

7.CONCLUSION & FUTURE SCOPE

The emerging technologies like machine learning and artificial intelligence playing a important part in the development of the IT (Information Technology) industries. We have

made use of these technologies and create a website for people who are consult about their diet and want to lead a healthy life. The importance of nutritional guidance and it will also help the user to maintain a healthy lifestyle . The recommendation will also helps in discarding the foods that you are allergic to and is to those patients whose body blood pressure and insulin level may change drastically and it be really helpful to the user to make any food according to your diet in just minutes and take the proper amount of calories as per needed by the body.

8.REFERENCES

[1] P. Khaire, R. Suvarna, Ashraf, "Virtual Dietitian: An Android based Application to Provide Diet", (IRJET), Volume: 07 Issue: 01 Jan 2020

[2] T.N.T. Tran, M. Atas, Alexander Felfernig, MartinStettinger, "An overview of recommender systems in the healthy food domain", J Intell Inf Syst (2018)

[3] S. Singh, S. Bait, J. Rathod, Nileema Pathak, "Diabetes Prediction Using Random Forest Classifier And Intelligent Dietician", IRJET, Volume: 07 Issue: 01 Jan 2020

[4] Shubham Singh Kardam, Pinky Yadav, Raj Thakkar, Anand Ingle, "Website on Diet Recommendation Using Machine Learning", IRJET, Volume: 08 Issue: 04 Apr 2021

[5] CELESTINE IWENDI , SULEMAN KHAN , JOSEPH HENRY ANAJEMBA , ALI KASHIF BASHIR , FAZAL NOOR, "Realizing an Efficient IoMT-Assisted Patient Diet Recommendation System Through Machine Learning Model", IEEE date of publication January 21, 2020.

[6] Aberg, Johan. (2006). " Dealing with Malnutrition: A Meal Planning System for Elderly".

[7] H. Zhang, Z. X. Cao, M. Li, Y. Z. Li, and C. Peng, —Novel naïve Bayes classification models for predicting the carcinogenicity of chemicals,‖ Food Chem. Toxicol., vol. 97, pp. 141–149, 2016.

[8] E. Prasetyo, Data Mining Konsep dan aplikasi menggunakan matlab. Yogyakarta: Andi, 2012.