Advances in Computational Sciences and Technology
ISSN 0973-6107 Volume 16, Number 1 (2023) pp. 35-43

https://dx.doi.org/10.37622/ACST/16.1.2023.35-43

Diet Recommendation System Based On Vitamin Intake

K. Sneha Reddy
Assistant Professor, G. Narayanamma Institute of Technology & Science (For Women)
email id: ksreddy26@gnits.ac.in

K. Chetna Reddy
Students, G. Narayanamma Institute of Technology & Science (For Women)
email id: chetureddy740@gmail.com

Preethika Chennareddy
Students, G. Narayanamma Institute of Technology & Science (For Women)
email id: pcreddy2001@gmail.com

Nuzhath Jahan
Students, G. Narayanamma Institute of Technology & Science (For Women)
email id: nuzhath3873@gmail.com

Abstract

A WHO study states that insufficient and unbalanced food intake causes about 9% of heart attack deaths, about 11% of coronary heart disease deaths and 14% of gastrointestinal cancer deaths worldwide. In addition, about 0.25 billion children suffer from various types of deficiency ranging from vitamin A to vitamin K deficiency, 0.2 billion people suffer from iron deficiency (anemia) and 0.7 billion people suffer from iodine deficiency. The main goal of this work is to recommend a menu to different individuals. The Diet recommendation system deals with the large volume of information present from the dataset.

In this project, a custom data set is prepared based on different high and low values of vitamins from (vitamin A, B, C, D, E, K) and the features are divided from normal and abnormal states of vitamins and the labels are divided into 0 and 1 as normal and abnormal. Another data set is prepared based on a combination of different vitamins and their deficiency and recommended foods according to which vitamin is deficient. In this project, multiple classifier algorithms (KNN, decision tree, random forest, logistic regression, voting classifier) are used and ensemble algorithm is used to combine multiple algorithms and train a new algorithm. The prediction is displayed using a flask web application that detects vitamin deficiency and recommends the type of food to be consumed in different combinations.

Keywords: Diet Recommendation, Machine Learning, Vitamins, Deficiency, Food.

1 Introduction

Nowadays, a person suffers from various health problems such as fitness problems, improper diet, mental problems, etc. Various studies show that improper and insufficient dietary intake is the main cause of various health problems and diseases. A WHO study states that insufficient and unbalanced diet causes about 9% of heart attack related deaths, around 11% of coronary heart disease related deaths and 14% of gastrointestinal cancer related deaths worldwide. In addition, about 0.25 billion children are Vitamin A deficient, 0.2 billion people are iron deficient (anemia), and 0.7 billion people are iodine deficient. The main goal of this project is to recommend a menu to different kinds of individuals. A recommendation system deals with the large amounts of information present by filtering the most relevant data based on user-provided data and other factors that cater to the user's preferences and interest. It detects the match between user and item and attributes similarities between users and items for recommendations based on their physical aspects (age, gender, height, weight, body fat percentage), preferences (weight loss or gain). The recommendation process basically has three phases, which are the information gathering phase, the learning phase, and the recommendation phase. First, information is gathered about a particular problem and various solutions related to that problem are categorized. After the collection of information comes the learning phase in which various conclusions are drawn from the collected information and in the last phase i.e. the recommendation phase, an output is given in which various recommendations are made. In our project, the recommendation output is based on the user's physical aspects, preferences, and body mass index (BMI). A balanced diet is an important aspect of a healthy lifestyle for people. Along with a balanced diet, regular exercise is essential for a healthy life. Nutrition and health are often overlooked today. Most people suffer from diabetes, heart disease, cancer, stroke, etc. Diseases are almost directly related to unhealthy eating habits. So our body needs nutrients to stay healthy and food provides the essential nutrients that keep us from getting sick. A healthy and balanced diet usually contains vitamins, minerals, proteins, healthy fats, proteins, carbohydrates and fiber. A healthy food pyramid is a combination of a plant-based diet and a reasonable amount of animal products. These include vegetables, grains, fruits, oils and sweets, dairy products, meat and beans. In general, one is not aware of the main causes of deficiency or excess of various vital substances, such as calcium, protein and vitamins, and how to normalize these substances with a balanced diet. With the advantage of technology, people can lead a healthier lifestyle. In this project, build a system that aims to recommend appropriate nutritional intake to its users based on body mass index (BMI) and food data preferences.

2 Related Work

Raciel yera toledo designed a food recommendation system based on nutritive information and user preferences. Meal plan for users recommended using user preferences. This tool handles both user likings and nutritional information. Vijay Jaiswal, who suggests healthy eating habits, eating habits and the amount of calories burned, nutrient intake and other aspects can be using data mining tools. In this tool, the hidden patterns and habits of customers when taking food are found from various data sources. In this tool decision tree learning algorithm and Random Tree algorithms are used on various datasets. H. Jiang designed a system for calculating daily calorie needs. The knapsack algorithm is used to suggest user dietary combinations. Unlike other diabetic diet recommendation systems, this system can rank recommended diet combinations using the TOPSIS algorithm according to the user's diet. Jung-Hyun Lee designed a customized dietary recommendation service for treating heart disease. This service provides customers with customized general information, family medical history, seasonal food intake. Rung-Ching Chen made a recipe ontology that describes the treatment of some common diseases with accurate food recommendations, and an inference engine for customer health condition, and the recipe ontology can be used for precise food priority recipe recommendations. FidelsonTanzil uses the ABC algorithm to get data from the database based on the user's requirements. Kmean and SOM algorithms are used on the datasets. Mohd Afisi designed the ABC algorithm in Data Mining and successfully tested it against six traditional classification algorithms, and ABC proved to be a suitable recommendation algorithm. Xiaoyan Gao proposed a food recommendation problem based on recipe recommendation factors for user choice. Using a neural network-based solution to Ordered Diet recommendations.

Authors Ingmar Weber and Palakorn Achanananuparp [1] attempted to gain insights into predicting the success of machine leaning, a diet that would help people trying to stay healthy and fit by monitoring their dietary intake. The authors used the public food logs of more than 4,000 longstanding active MyFitnessPal users to study the characteristics of failed diets. Specifically, the authors trained a machine learning model to predict repeatedly exceeding or falling short of one's own set daily caloric goals, and then looked at which characteristics contributed to the model's predictions, focusing on "quantified self" data. The authors realized that the classification performance was enough and the token-based model performed better than the category-based model and used such data viabily for deeper data mining.

Nadish Shah and Ishani Shah [2] presented a design of a healthy food intake based on web data mining to find hidden patterns and business strategies from their customer and web data to track eating habits and recommend food kinds that will improve overall health and avoid the types of food that increase the risk of disease. The authors used data mining algorithms such as classification, clustering, association rules, etc. in the data mining process to get useful information about various people's eating habits. The nutritional composition of each type of food was analyzed and the fat, energy, and vitamin content of the recipe was determined. Then, using a classification mining algorithm, they processed the composition and obtained a result of whether the food was healthy or not and custom recommendations were designed for every person.

Through this paper, authors Aine P. Hearty and Michael J. Gibney [3] have shown how a food level coding system can be studied using data mining techniques. They assessed the appropriateness of supervised data mining methods to predict an attribute of diet quality based on dietary intake using a food-based coding system and a novel meal-based coding system. The authors used food consumption databases from the Northern Ireland Food Consumption Survey 1997–1999. A Healthy Eating Index (HEI) score was created. Artificial neural networks (ANNs) and decision trees predicted HEI quintiles based on diet combinations. As a result, the ANN had a higher accuracy than the decision tree based on its capability to predict HEI. However, based on the food coding system, the decision tree had more accuracy than ANN. Data mining was used by Christy Samuel Raju, Sanchit V Chavan, Karan Pithadia, Shraddha Sankhe, Prof. Sachin Gavhane [4] to develop the Fitness Advisor System. The "Fitness Advisor" made by the authors was a desktop app that gave the user advice according to his/her weight related problems by diagnosing the same and spreading awareness about the health risks. The authors considered various factors in the system, such as height, weight, body type, size, gender, smoking habits, drinking habits, health conditions, physical activity, sleep schedule, etc. The combination of clustering, association and classification algorithms to successfully provide the best possible advice to the user's problem was used by the authors. The Apriori algorithm was used by the authors to generate association rules. The final output of the system was expert advice on diet and exercise.

3 Existing Systems

A data-based food recommendation system is designed that recommends easy recipes based on preferences already entered by the user. The user's favoured recipes are divided into ingredients, which are allotted ratings based on saved user preferences. Recipes with appropriate ingredients are recommended. The authors do not take into account nutritional factors and the balance of the diet. Moreover, there is a chance of identical recommendations since the user's preferences may not change daily. Brands and latent factors are used for Android centered food recommendation system. The system will suggest a individualized recipe to the user centered on the brands and rankings listed in the user likings. The advocated system uses latent symptom vectors and matrix factorization in its algorithm. Prediction accurateness is attained by using brands that thoroughly match recommendations to user preferences. However, the authors do not reflect on nutrition in order to balance the user's intake according to his requirements. Disease prediction was made based on symptoms.

4 Architecture

Fig 4.1 contains the architecture of the proposed system. A system architecture or systems architecture is the conceptual model that defines a system's structure, behaviour, and more views. An architecture description is a formal description and representation of a system. Organized in a way that supports reasoning about the structures and behaviours of the system. The three-layered software architecture (a three tired architecture) emerged in the 1990s to overcome the restrictions of the two-layered architecture. The third layer (middle layer server) is between the client (user interface) and the server (data management) components. This middle layer offers process administration where business logic and rules are implemented and can house hundreds of users (as compared to only 100 users of the two layered architecture) by offering functions such as queuing, application execution, and database staging. The three-layered architecture is used when an operative distributed client/server design is needed that delivers (when compared to the two-layered) amplified performance, reusability, maintainability, flexibility, and scalability while hiding the intricacy of distributed processing from the user. These attributes have made three tier architectures an accepted choice for Internet applications and net-centric information systems.

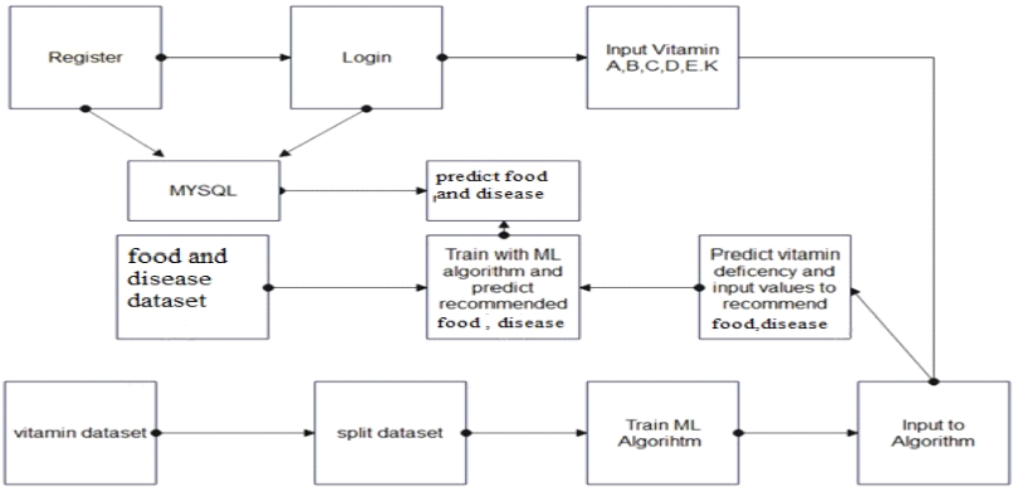

Fig 4.1: Proposed work flow

4.1 Dataset

The vast majority of data sets available today relate to vitamin deficiencies and food recommendations; extensive data collection and analysis of this aspect is only now beginning to gain traction. We were able to use data from the US Department of Agriculture (USDA) system as a primary source for our research and the machine learning models we developed. They deliver management on food, agriculture, natural sources, nutrition and related issues centered around public policy, the best available science and effective management. They have an image to provide economic opportunity via innovation to aid rural America; promote agricultural production that better nourishes Americans while helping to nourish others around the world; and protect America’s natural resources by conservation, restoration of forests, watershed improvements, and healthy private working land. Food, Nutrition, and Consumer Services work to influence the nation's agricultural wealth to reduce food insecurity and improve nutrition security in the USA. Its operating agency, the Food and Nutrition

Service, manages federal domestic nutrition assistance programs and includes the Center for Nutrition Policy and Promotion, which associates scientific research to consumer nutritional needs via science-based dietary guidelines, nutrition policy coordination, and nutrition education.. The datasets used for our analysis were obtained from USDA's FoodCentral. FoodData Central[6] is an integrated research-oriented data system that provides expanded data on nutrients and other food components, as well as links to sources of related agricultural, food, food supplement, and other data. FoodData Central can be used and has benefits for a variety of users, including researchers, policy makers, academics and educators, nutrition and health professionals, product developers and others.

4.2 Module Description

The data is divided into different sets and then trained for different models. The data set was first divided into a training set (80%) and a pre-training set (20%). The pre-training set was divided into pre-training (80%) and pre-test (20%). Now the training set is further divided into train (80%) and verification (20%). This train set is again divided into a train set (80%) and a test set (20%). So now I have separate sets for train validation and testing that don't overlap. The pretrain set was used to find the best models for the given dataset. In this phase the machine learning algorithms are initialized and using this information algorithm the train values are given to the algorithm which will know what are the features and what are the labels. Then the data is modeled and stored as a pickle file in the system that can be used for prediction. The dataset is trained using multiple algorithms and the accuracy of each model is calculated and the best model is used for prediction. In this phase, new data is received as input and the trained models are loaded using pickle and then the values are pre-processed and passed to the predictive function to find out the result that will be displayed in the web application.

5 Implementation

Fig 5.1 gives a brief explanation of how algorithms are used in the proposed system. The first step in the machine learning process is to get the data. This will depend on the type of data you are gathering and the source of data. This can be either static data from an existing database or real-time data from an IoT system or data from other repositories.

The second stage deals with data cleaning. The real-world data is often shambolic, redundant, or is missing elements. In order to feed data into the machine learning model, we need to first clean, prepare and manipulate the data. This is the most critical step in the machine learning work flow and takes up the most time as well. Having to clean data means that you can get a more accurate model in the future.

Data can be in any format - CSV, XML, JSON, etc. After cleaning the data, you need to then convert these data into valid formats that can be fed onto the machine learning platform. Finally, these datasets are further divided into training and testing datasets. The training dataset is used to train the model. The testing dataset is used to validate the model.

The next step in the machine learning workflow is to train the model. A machine learning algorithm is used on the training dataset to train the model. This algorithm leverages mathematical modeling to learn and predict behaviors. These algorithms can fall into three broad categories - binary, classification, and regression.
Next comes the evaluation, the evaluation process is needed to check whether is well-trained or competent. Through this method, you will easily get to test your model against data that were never released. This happens just to ensure how the model responds to the data it hasn't come across yet. Evaluation is ideally done to analyze how the model might perform in real-time.
During the training session, there are multiple parameters to be considered. For each parameter, they should be able to specify or define what makes a model suitable for your use, else you might find yourself wasting your time or tweaking parameters for a longer duration of time. This happens to check whether or not there is still room for improvement in the training model. This is easily done by tuning certain parameters – learning rate or how many times have the trained model runs during the training session. Once the model is trained, deploy and pipeline it to production for application consumption.
The machine learning process that we have outlined here is a fairly standard process. As you go through this process on your own with your own problems, you will start to discover a few more machine learning steps that might work for you.

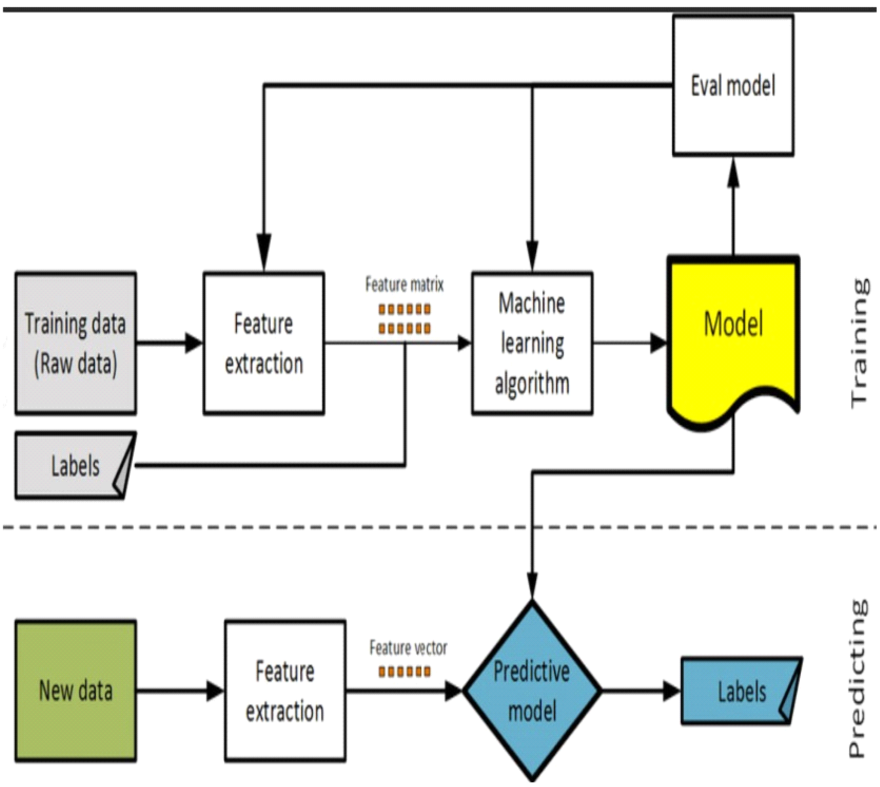

Fig 5.1: Algorithm used in the proposed system

6 Results

We used datasets primarily from USDA's FoodCentral to evaluate our models. In machine learning, classifier performance measures[5] are used to judge how well algorithms perform for a given data set. These measures provide us with a framework for evaluating the strength and limitations of model predictions. A true positive is the total number of times the model correctly predicts positive results. A false positive is the number of times the model incorrectly predicts as a positive. True Negatives is the total number of times the model correctly predicted negative results. False negatives are the total number of times the model incorrectly predicted as negative. Accuracy is a ratio that measures the effectiveness of a model in predicting positive labels from all positive predictions. As we can see in Figure 6.1 below, it represents the classification report obtained from the sklearn.metrics package. It provides details on precision, recall, f1-score, support and accuracy. We can clearly observe that the accuracy is 98% for the built model and considered data set. Accuracy is 98%. Recall averages 98%. The f1 score also averages 98%.

	precision	recall	f1-score	support
0	0.98	0.99	0.99	111
1	0.98	0.97	0.97	60
accuracy			0.98	171
macro avg	0.98	0.98	0.98	171
weighted avg	0.98	0.98	0.98	171

Fig 6.1: Classification Report

7 Conclusion

We have created a website which recommends the food items and predicts vitamin deficiency in which we have implemented prediction by taking input as vitamins and their deficiency. For training of the system, the initial process involves the dataset preparation of food items depending upon the vitamin deficiency. The prediction of various food recommendation along with disease prediction is done, depending upon which are essential for the type of vitamin deficient. After the training is performed, using logistic regression, the nearest food items are predicted which best suited for the appropriate diet. Our diet recommendation system allows users to basically get the desired healthy diet on the bases of vitamin deficiency and disease prediction system can give updates to user changes of disease with predicted vitamin deficient values.

References

[1] Phanich, M., Pholkul, P., & Phimoltares, S., “Food recommendation system using clustering analysis for diabetic patients,” in Proc. of International Conference on Information Science and Applications, pp. 1-8, IEEE, April 2010. Article.

[2] Ge, M., Elahi, M., Fernaández-Tobías, I., Ricci, F., & Massimo, D., “Using tags and latent factors in a food recommender system,” in Proc. of the 5th International Conference on Digital Health, pp. 105-112, ACM., May 2015.

[3] Freyne, J., & Berkovsky, S., “Evaluating recommender systems for supportive technologies,” User Modeling and Adaptation for Daily Routines, pp. 195-217, Springer London, 2013.

[4] Prof. Prajkta Khaire, Rishikesh Suvarna, Ashraf Chaudhary, “Virtual Dietitian: An Android based Application to Provide Diet”, International Research Journal of Engineering and Technology (IRJET),Volume: 07 Issue: 01 | Jan 2020

[5] Powers, David M. W. (2011). "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation". Journal of Machine Learning Technologies. 2 (1): 37–63.