Diet Recommendation System Using Machine Learning

Preprint · December 2023

DOI: 10.13140/RG.2.2.20744.19207

1 author:

Mohammed Shaheel
Bournemouth University
1 PUBLICATION **0** CITATIONS

SEE PROFILE

Diet Recommendation System Using Machine Learning

Mohammed Shaheel
MSc:Digital Health and Artificial Intelligence
Bournemouth University
Bournemouth,UK
kpmshaheel@hotmail.com

***Abstract*—Development of a recommendation system with help of Machine learning techniques to promote healthy eating habits and easy access for user for a tailored diet recommendation which can help them with better food choice and improves health. An automized unsupervised model which benefits users who don't have access to a personal nutritionist or may not afford such a one there by helping to improve general health.**

I. INTRODUCTION

One of the leading problems around the world is unhealthy heating habits and lack of awareness of what to eat based on their body conditions which further leads to development of chronic diseases. For instance, if you visit a food store, for example, Subway, they just provide you with your allergen list, which confuses the customer at times. Most of time people are aware of their blood levels and sugar, but unaware of content of food that can affect these levels. So here we try an automized recommendation system with the help of machine learning techniques to recommend various food based on their contents with user input such as height, weight etc and even to get a personalized diet plan based on this value. Many users from underdeveloped regions don't have access to personal nutritionists or may not be able to afford such facility. Key advantage of such a system is that it can take consider of individual's dietary restriction and preferences and provide a tailor-made service for that individual. For early training of the model, dataset named food.com is available at Kaggle where it's possible to train and test the system model on thousands of food data values. Further, for interaction of User a web based front end will be developed for user input and recommendation output.

II. PROBLEM STATEMENT

Fast food consumption is alarmingly high, which has led to the consumption of unhealthy food. This leads to high blood pressure, diabetes, obesity, and other health problems. Therefore, it is now absolutely necessary for people to have a decently adjusted diet that is balanced, nutritious, and healthful. But not everyone in today's world has the time or means to pay for a personal nutritionist or dietitian to assess their health and provide a balanced diet plan tailored to their individual needs. We talked about the person's unhealthy eating habits in this report and tried to create a recommendation system which automatically provides them a way to improve their diets and live better lives.

III. LITERATURE SURVEY

Thi Ngoc Trang Tran et al. (2021) [1] demonstrate how their approach can be applied to provide an extensive overview of research on healthcare recommender systems: Furthermore, that it provides information for proposed scenarios and methodologies. This includes advice on nutrition, diets,medication plans, health services, and various assistance from medical experts. To ensure that students fully understand recommendation systems, they also provide real-world examples.

Wenbin Yue et al [2] state that the purpose of his work is to offer a thorough analysis of conventional recommendation techniques and their uses in the healthcare industry. Three main recommendation methodologies are stated here: content-based recommendation , collaborative filtering and hybrid recommendation. After that, a discussion regarding five application scenarios pertaining to health has been provided. These scenarios include, but are not limited to, dietary advice, lifestyle advice, training advice, patient and physician decision-making, and illness prediction. Ultimately, a number of the most important issues facing this evolving field are acknowledged, along with strong arguments.

Gao et al. (2017) [3] introduced a framework for a personalized diet system. The system incorporated usage of Bayesian personalized ranking and matrix factorization tounderstand what a specific user prefer over a large volume of data. The results pointed out that performance of such a system is much better than conventional collaborative filtering techniques

The Nourishment Recommendation Framework (NRF), which was proposed by Anonnya Banerjee et al. [4], gathers user input from children, evaluates the information, and generates an output that, in the end, produces a better diet plan. Its goal is to give kids between the ages of 8 and 13 nutrient-dense meals that are matched to their age, development, gender, and medical records.

A machine learning technique is used by M. Geetha et al. [5] to analyze a person's body and predict their future health status over the course of a calendar year based on their pre-medical history. It provides diet recommendation systems that consider both historical and current food consumption data, and it makes more reliable diet report recommendations than it did previously.

Ramni Harbir Singh, Sargam Maurya, et al.[6] used cosine similarity and KNN to create a movie recommendation system. This study describes a technique that offers users generic recommendations based on a movie's genre and/or

popularity. Many deep learning strategies are used in the Content-Based Recommender System's implementation. This work also sheds light on the challenges faced by content-based recommendation systems and our attempts to overcome them.

For diabetic patients, a Food Recommendation System (FRS) [7] is suggested, which employs SelfOrganizing Map and K-mean clustering for food clustering analysis. The suggested method suggests the replacement foods based on food parameters and nutrition.

IV. METHODOLOGY

The novelty in this approach is that the machine learning model will learn based data on USDA food data base where up-to-date, vendor-specific nutrition, allergen and ingredient information of varous food items are availabl*e*. The idea was to run different machine learning algorithms on this data to learn what category of food item would be appropriate for a particular user based on values he given through the front end and predict a food item suitable for the user.

A. Datasets

The recommendation system is based on USDA nutrition information database. Every food item's nutrition information is maintained in the USDA database. The USDA Foods Database is available for state agencies and school districts to access up-to-date, vendor-specific nutrition, allergen and ingredient information for direct delivered USDA Foods for child nutrition programs offered on the USDA FOOD AVAILABLE LIST[8] .A sample of dataset being used are shown in Fig 1 and Fig 2.

Food_item	Breakfast	Lunch	Dinner	VegNovVe	Calories
Asparagus	0	1	1		22
Avocados	1	0	0	0	160
Bananas	1	0	0	0	89
Bagels ma	0	1	1	0	250
Berries	1	0	0	0	349
Brocolli	0	1	1	0	25
Brown Ric	0	1	1	0	362
Cauliflowe	0	1	1	0	32
American	1	0	0	0	331
Coffee	1	0	0	0	2
Corn	1	1	1	0	97
Dark choc	0	0	1	0	556

Figure 1: Sample food dataset

Calories	Fats (gm)	Proteins(g	Iron(mg)	Calcium(m	Sodium(m
160	15	2	0.55	12	7
89	0.3	1.1	0.26	5	1
349	0.4	14	6.8	190	298
331	24	20	0.84	497	966
2	0	0.3	0.02	2	1
97	1.4	3.3	0.55	2	253
93	2.1	5.6	2.63	2	9
97	6.9	3.8	0.12	2	52
553	44	18	6.68	2	12

Figure 2: Sample nutrient dataset

B. Data Preprocessing

- Normalization: Machine learning models perform better when all of the values are brought into a single scale through the use of min-max normalisation. Because the dataset contains a variety of scale values—some are single digit values, some features have two digest values, and some features have three digest values—normalization is the process of condensing the dataset to a single range of values. The min-max scaling technique is used in this study to normalise data that fall between 0 and 1 on the scale.

- Data encoding is carried out following the removal of inconsistent and duplicate values from the dataset. Subsequently, the nominal attributes are converted into numerical values. Performance will be improved since machine learning models' backend operations are carried out on numerical values prior to being implemented using a machine learning model. Before being given to data encoding process, the non-numeric data included in the database was converted to numeric data. Using machine learning techniques, backend computations were performed on numerical values instead of nominal values prior to supplying input to the suggested model..

C. Feature Selection.

For feature selection we make use of recursive feature elimination approach. In this approach a ranking process is introduced .Feature are ranked based on model's coefficient or feature's importance approach further fits the model and starts to eliminate each feature based on lowest ranking through a recursive fashion. The process involve developing a model with remaining features after repeatedly removing least significant parts until desired number of features is obtained.[9] RFE attempts to eliminate any dependencies or collinearity that may exist in the model.

D.Machine Learning

1) **K-means** clustering is a popular unsupervised machine learning algorithm in various applications, including diet recommendation systems. K means clustering can be

used to put individuals who are similar together in a dataset of users who have each disclosed information about their dietary restrictions, allergies, and food preferences. For each cluster, this can assist in creating customised meal plans, recipes, and diet regimens. Each user is assigned to the closest cluster using K-means clustering, which is based on how similar they are.

2) **Random Forest** is a popular machine-learning algorithm used in various content basedrecommendation systems. It works especially well for tackling non-linear and high-dimensional situations[10]. A supervised learning algorithm that is a member of the decision tree family is called Random Forest. To produce a more reliable and accurate prediction model, Random Forest integrates several decision trees, in contrast to conventional decision trees, which only use one tree. Using Random Forest, a recommendation system can build a model that forecasts a user's preferences based on their prior behaviour as well as the behaviour of users who are similar to them. Large datasets can be analysed by the algorithm, which can then be utilised to find trends and insights to provide tailored suggestions.

3) **Decision Tree:** A decision tree's primary objective is to divide a population of data into more manageable segments. They are supervised learning methods without parameters that are applied to regression and classification. The goal is to build a model that, by learning basic decision rules deduced from the data's attributes, can forecast the value of an output class.

4) The **Naive Bayes approach**, which is based on the Bayes theorem, is an efficient supervised learning strategy for classification tasks. It's a probabilistic classifier, which means it generates predictions based on the chances of something existing**.** The core idea of this classifier is based on bayes theorem. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

V . WORKFLOW

Following fig.3, show the proposed workflow of the recommendation system. Here dataset in form of csv obtained from USDA Database is employed to data preprocessing unit. Data preprocessing unit carries out three preprocessing stages : Normalization, encoding and visualization. Once preprocessing is completed predictions are made by recommendation system with help of all four machine learning algorithms. Further each model is evaluated basis on various model assessment metrics

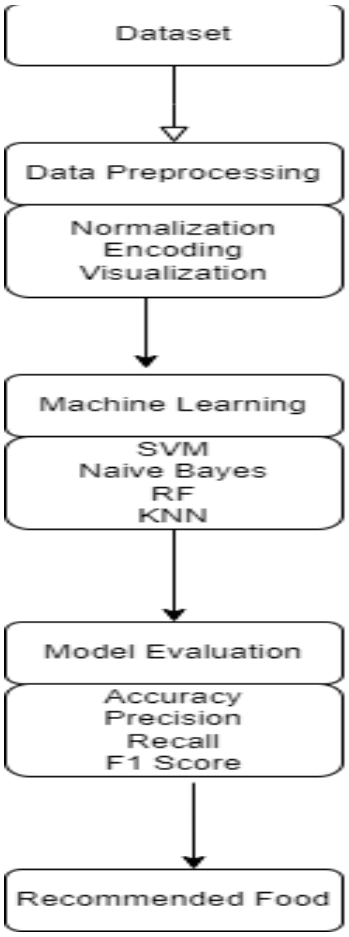

Fig. 3.Workflow of the proposed model

VI.MODEL EVALUATION

For model evaluation mainly four various assessment metrics are used

Precision : It can be defined as the actual correct prediction divided by the total number of predictions the model made. To put it simply, our model predicts that 7 out of 10 patients will have cancer, but only 3 of the predicted 7 patients actually have heart disease, meaning that the precision in this instance is 3/7 = 0.428.

Recall : Recall is determined by dividing the total number of true positives and false negatives by the number of true positives in a two-class classification problem. Let's say there are seven real cancer patients, but our model only predicts that five of them will have cancer. In this scenario, the recall is 5/7=0.714.

Accuracy : It can be calculated by dividing the total number of correctly classified examples by the total number of classified examples.

F1 Score : The precision and recall weighted average is called the F1 score. It also takes into account false positives and false negatives, as we are aware in precision and recall[11]. Generally speaking, F1 score is more valuable than accuracy, particularly in cases where the class distribution is uneven.

When the cost of false positives and false negatives is comparable, accuracy performs best. It is preferable to consider both Precision and Recall if the costs associated with false positives and false negatives are significantly different.

User inputs his personal data such as age weight height, blood, sugar etc. This data is further run against all the four machine learning models which gives prediction. Based on model evaluation process which include above assessment metrics carried out after predictions prediction of best model-based evaluation score is displayed for user.

VII. RESULT

Here diet recommendation system focuses on prediction developed by the four machine learning models incorporated in the recommendation system.. The recommendation system ask inputs from user end regarding user personal details. This includes user's Age,height,weight,blood,sugar level etc. This data is used to get a personalized food recommendation for the particular user. Here the data given by user is run against all the four machine learning models which gives prediction. Based on model evaluation process carried out after predictions, prediction of best model based evaluation score is displayed for user

VIII.PROPOSED FRONT END

An application can be further developed using the Fast API framework, which allows for the creation of fast and efficient web APIs. When a user makes a request to the API (user data, nutrition data...) the model is used to generate a list of recommended food similar/suitable to his request (data) which are then returned to the user via the API.

The application's front-end is made with Streamlit. Streamlit is an open source app framework in Python language[12]. It helps to create web apps for data science and machine learning in a short time. It is compatible with major Python libraries such as scikit-learn, Keras, PyTorch, SymPy(latex), NumPy, pandas, Matplotlib etc. For our case the front-end is composed of three web pages. The home page, which serves as an introduction and welcome page. The user can access the page with diet recommendations by using the left sidebar navigation. The user fills in his age, weight, height, blood type, sugar level, and other details on the diet recommendation page, and the page generates a diet recommendation based on the data provided. Additionally, the user can further specify their food preferences using nutritional values through the custom food recommendation..

IX .CONCLUSION

A content-based recommendation system which can produce food recommendations to a user based on his input. Further built upon to get even a personalized diet plan for user. Paper utilizes unsupervised learning techniques for training, testing and recording the response. At the end various metrics will be used to record and test model's accuracy such as F1score, recall etc... Furthermore, develop a front-end, utilizing web development technologies for user interaction with the system. The end goal is to design a system where users can get food and diet recommendations based on input, such as height and weight through the interface.

X . ACKNOWLEDGMENT

I would like to express my deepest gratitude to my Unit leader for research methods and professional issues, Dr Philip Davies, for his continuous guidance and expertise. His valuable insights and feedback have been instrumental in shaping the direction of this research. I would also like to thank my colleagues for their dedicated efforts and collaborative spirit.

XI.REFERENCES

[1]]Thi Ngoc Trang Tran, Alexander Felfernig, Christoph Trattner and Andreas Holzinger, "Recommender systems in the healthcare domain: state-of-the-art and research issues", Journal of Intelligent Information Systems, vol. 57, 2020, pp. 171–201, doi: 10.1007/s10844-020- 00633-6.

[2] W. Yue, Z. Wang, J. Zhang and X. Liu, "An Overview of Recommendation Techniques and Their Applications in Healthcare", *IEEE/CAA J. Autom. Sin.*, vol. 8, no. 4, pp. 701-717, 2021. H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[3] Jun Gao, Ninghao Liu, Mark Lawley, and Xia Hu, "An Interpretable Classification Framework for Information Extraction from Online

[4] A. Banerjee and N. Nigar, "Nourishment Recommendation Framework for Children Using Machine Learning and Matching Algorithm", *2019 Int. Conf. Comput. Commun. Informatics ICCCI 2019*, pp. 1-6, 2019

[5] M. Geetha, C. Saravanakumar, K. Ravikumar and V. Muthulakshmi, "Human Body Analysis and Diet Recommendation System using Machine Learning Techniques", 2021.

[6] Ramni Harbir Singh, Sargam Maurya, Tanisha Tripathi, Tushar Narula, Gaurav Srivastav," Movie Recommendation System using Cosine Similarity and KNN",2020,((IJEAT), DOI: 10.35940/ijeat.E9666.069520

[7] 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

[8] US Department Agriculture, https://www.fns.usda.gov/usda-fis/usda-foods-database

[9] Avcontentteam, https://www.analyticsvidhya.com/blog/2023/05/recursive-feature-elimination/

[10] Tony yiu, https://towardsdatascience.com/understanding-random-forest-58381e0602d2

[11] Dipesh Silwal, https://www.linkedin.com/pulse/confusion-matrix-accuracy-precision-recall-f1-score-measures-silwal

[12] Zakaraia narjis, https://github.com/zakaria-narjis/Diet-Recommendation-System