Identification Of Vitamin Deficiency and Recommendation of Rich Vitamin Food Using Machine Learning Techniques

***K.V. Satyanarayana, [1]Gangireddy Pujitha, [2]Battula Vishal, [3]Indukuru Pranay Varma**

4th YEAR , Department of Computer Science, Raghu Engineering College

*vsatyanarayana.kalahasthi@gmail.com,[1]19981a0550@raghuenggcollege.in,[2]19981a0523@raghuenggcollege.in,[3]19981a0558@raghuenggcollege.in

Abstract

According to a WHO report, Insufficient and unbalanced food consumption, is to blame for 14% of deaths from gastrointestinal cancer. 11% of ischemic heart disease fatalities and 9% of heart attack deaths occur worldwide. Moreover, iodine deficiency affects 0.7 billion people worldwide, 0.2 billion due to an iron shortage (anemia), and 0.25 billion children, who among other things lack vitamins A through K. The primary goal of this initiative is to offer a personalised diet plan that takes into account each person's needs. The system in use uses a wide variety of info from numerous datasets..

To accomplish this goal, two datasets were prepared. The first data set was built on a variety of high and low vitamin levels, including those for vitamins A, B, C, D, E, and K. and the conditions of vitamins were further classified into normal and pathological conditions Labels were separated into the numerals 0 and 1, which correspond to normal and abnormal circumstances, respectively. By analyzing the interplay of different vitamins and their inadequacies, the second set of data was compiled. Then depending on the insufficiency of a particular vitamin, the appropriate nourishment was identified.

In this project,They included KNN, decision trees, random forests, logistic regression, and vote classifier, among other classifier algorithms. To enhance performance, an ensemble algorithm was implemented, combining several algorithms to create a new one. Each algorithm's accuracy was computed, and the most accurate one used for prediction. A Flask web application was used to present the predictions and suggest suitable food types for various combinations based on the diagnosed vitamin deficiency.

Keywords: Machine Learning , Vitamins (A,B,C,D,E,K) , Food Recommendation , Vitamin Insufficiency.

INTRODUCTION

People are currently dealing with a variety of health difficulties, including bad eating habits, fitness issues, and mental health problems. According to research, eating poorly and insufficiently is a major contributor to a number of illnesses and health issues. According to a World Health Organization study, inadequate and unbalanced food intake is to blame for 14% of gastrointestinal cancer deaths, 11% of deaths from ischemic heart disease, and 9% of deaths from heart disease worldwide. A further 250 million children suffer from a vitamin A deficit, 200 million from anaemia due to iron deficiency, and 700 million from iodine inadequacy. The main goal of this study is to offer nutritional suggestions for various people. The suggested system handles enormous volumes of data by selecting the most important information based on user-provided data and other variables that take into consideration user preferences and interests. It determines if users and products are similar to one another and assigns them based on physical characteristics like age, gender, height, weight, and body fat percentage, as well as desires like weight loss or growth, and then provides the right

suggestions. The information gathering phase, the learning phase, and the recommendation phase are the three stages of the recommendation process. It begins by gathering data on a particular issue and classifying the numerous fixes connected to it. Second, it goes through the learning phase, where different conclusions are made using the data gathered. Ultimately, numerous recommendations are given throughout the process of evaluation. Based on the user's physical condition, preferences, and body mass index, our project recommends an output (BMI).A balanced diet is critical to leading a healthy lifestyle. In addition to a balanced diet, regular physical activity is vital to staying healthy. Nutrition and health are often neglected these days, with most people suffering from diseases such as diabetes, heart disease, cancer, stroke, etc., which are directly linked to poor eating habits. In order to maintain good health, our bodies rely on essential nutrients provided by food. A range of foods, including vitamins, minerals, protein, healthy fats, carbohydrates, and fibre, make up a balanced diet. Vegetables, grains, fruits, oils and sweets, dairy products, meat, and legumes are all included in a well-balanced dietary pyramid, along with vegetables, grains, fruits, and oils and sweets. Yet, many people might not be aware of the reasons for excesses or shortfalls of vital nutrients like calcium, proteins, and vitamins, or how to balance them through a healthy diet. In our project, we aim to develop a system that suggests appropriate nutritional intake to users based on their BMI and grocery store data preferences. BMI calculates weight status categories, including underweight, healthy weight, overweight, and obese. Grocery data comprises seasonal foods, user-processed foods, plant-based foods, and animal products.

LITERATURE REVIEW

A food suggestion system that considers dietary data and user preferences is provided by Raciel Yera Toledo. recommendations for the user's menus based on their preferences. This tool controls user preferences and dietary data. Vijay Jaiswal uses data mining tools to suggest healthy eating habits, eating habits and calorie consumption scores for nutrient intake and more. In this tool, hidden patterns and eating habits of customers are uncovered from different data sources. In this Decision Tree Learning Algorithm tool, a random tree algorithm is used for different data sets. H. Jiang proposed a system for calculating daily caloric needs. The Backpack algorithm is used to recommend the user's diet combination. Unlike other diabetic diet recommendation systems, this system can use the TOPSIS algorithm to sort recommended diet combinations based on the user's food intake. Jung-Hyun Lee introduces personalized diet recommendation service for heart disease management. This service provides customers with general personal information, family medical history, seasonal food intake. In their research, Rung-Ching Chen creates a recipe ontology that precisely outlines the therapies for common diseases, providing dietary recommendations and a reasoning engine to cater to the individual health status of customers. This ontology can be utilized to suggest appropriate recipe recommendations based on a person's food priorities. FidelsonTanzil employs the ABC algorithm to extract pertinent information from the database to cater to the user's requirements. The data set is analyzed using

Kmean and SOM algorithms. Mohd Afisi conducts a successful test of the ABC algorithm along with six traditional classification algorithms in Data Mining, affirming that it is a fitting algorithm for recommendation. Xiaoyan Gao introduces a food recommendation system that factors in user-selected recipe recommendation preferences, utilizing a neural network-based solution for Ordered Diet Recommendation. Lastly, in their research, Data mining approaches are investigated by Ingmar Weber and Palakorn Achananuparp [1] in order to understand meal suggestions and patterns from customer and network data, using algorithms such as classification, grouping, association rules to attempt and gain insight by predicting a lean diet - successful dieting will help people stay healthy by monitoring food intake. In order to research the traits of failing diets, the scientists examined the diet diaries of more than 4,000 long-term active My FitnessPal members. They developed a model using machine learning to predict whether users had met or failed to meet their daily calorie objectives, and they looked at the variables that affected the algorithm's predictions, with a focus on "self-quantified" data. The authors found that token-based models outperformed category-based models in terms of classification performance, and identified opportunities for further data exploration.By means of web data mining, they discovered concealed patterns and enterprise tactics from customer and network data. This involved keeping tabs on dietary preferences and proposing types of food that could enhance well being while steering clear of those that accentuated the likelihood of sickness. To gather pertinent data on people's eating patterns, the authors used data mining methods including classification, grouping, and association rules. They estimated the proportions of fat, calories, and vitamins in the recipe after examining the nutritional makeup of each type of food. Next, they used a classification extraction algorithm to determine whether or not the diet was healthful, making tailored recommendations for each person in line with. The authors of this study, Aine P. Hearty and Michael J. Gibney [3], show how to investigate meal coding systems using data mining approaches. The authors evaluated the effectiveness of supervised data mining approaches in forecasting a component of diet quality based on food intake using two coding systems, including a food-based coding system and a novel meal-based coding system. They developed a Healthy Eating Index (HEI) score using the food consumption database from the Northern and Southern Ireland Food Consumption Survey, which was carried out between 1997 and 1999. The HEI quintile was predicted using artificial neural networks (ANN) and decision trees based on food pairings. According to the results, ANNs are just slightly better at predicting HEI than decision trees. Nonetheless, decision trees beat artificial neural networks in terms of the meal coding system. The study was carried out by CHRISTY SAMUEL RAJU, SANCHIT V CHAVAN, KARAN PITHADIA, SHRADDHA SANKHE, and PROF using data mining methods. Also participating was SACHIN GAVHANE [4].creation of a fitness guidance system. The author created "Fitness Adviser," a computer programme that gives users guidance based on their concerns about their weight by accurately diagnosing and raising awareness of health issues. Height, weight, body shape, gender, smoking, alcohol use, health status,

physical activity, sleep duration, and other variables were all evaluated by the authors. The authors combined clustering, association, and classification algorithms to efficiently provide the best expert advice for user queries. To produce association rules, they used the Apriori algorithm. At the end, the system generates nutrition and exercise advice from specialists.

METHOD

KNN Algorithm

A flexible method, the K-Nearest Neighbors (KNN) method[6] can be used to solve both regression and classification problems, It provides advantages including interpretability, processing speed, and predictive ability. This machine learning technique uses data from prior datasets for example-based learning and applies a distance function, such as the Manhattan or Euclidean distance, based on the weighted average of the target values of the k nearest neighbours, determine the target value for new samples. To balance the prediction's bias and variance, the value of k can be changed. KNN has the advantage of not requiring any training or optimization, but its use of data samples for prediction increases its complexity and time requirements. This overview focuses on KNN techniques for early prediction of dietary recommendations, using vitamins as the input parameter for recommending high-quality foods. Unknown values of vitamins can be predicted by calculating the Euclidean distance between them and their nearest known values. Several distance functions, with the Euclidean metric being the most popular one, can be employed to calculate the separation between points in feature space.

Classification Algorithms

The section you've been waiting for is now here: teaching machine learning algorithms. To gauge how well our algorithm is working, I started by splitting our balanced dataset into an 80/20 train-test split. To prevent overfitting, I employed the widely used k-fold cross-validation resampling technique. This entails dividing the training set into k folds, fitting the model on k-1 folds, and then making predictions on the last fold. For each fold, this procedure is repeated, and the resulting predictions are averaged. Let's examine some of the most well-known classification algorithms in more detail to see which one did the best with our data : Classification trees, Support Vector Classifiers, Logistic Regression, Linear Discriminant Analysis, K Nearest Neighbors (KNN), Random Forest Classifiers, and K Nearest Neighbors (KNN).

Logistic Regressions

A statistical model known as the logistic model uses the event's log-odds as a linear combination of independent variables to estimate the likelihood of an event. Contrarily, a logistic model's parameters are estimated using the regression analysis approach known as logistic regression. [7]
Equation for logistic regression: y = e(b0 + b1*x) / (1 + e(b0 + b1*x)). where x is the projected outcome and y is the input value. The bias or intercept term is b0, and the single input value's coefficient is b1 (x)

Linear Discriminant Analysis(LDA)

A linear model known as linear discriminant analysis (LDA) is used for both classification and dimensionality

reduction. It has a lengthy history and is frequently used for feature extraction in pattern classification applications. Originally formulated by Fisher in 1936 for two classes, LDA was later generalised for many classes by C.R Rao[8] in 1948. LDA's main goal is to project data in a way that maximises variability across classes and decreases variability within classes from a D-dimensional feature space to a D' (D > D')-dimensional space.

Arg max J(W) is equal to (M1 - M2)2/S12 + S22.

IN THE NUMERATOR:

$$(M_1 - M_2)^2$$

$$= (W^T m_1 - W^T m_2)(W^T m_1 - W^T m_2)^T \quad = W^T (m_1 - m_2)(m_1 - m_2)^T W$$

$$= W^T S_b W \qquad \ldots\ldots\ldots \quad (2)$$

S_b = between class scatter

FOR DENOMINATOR:

Class Scatter before projection

$$S_i = \sum_{x(n) \in C_i} (x(n) - m_i)(x(n) - m_i)^T$$

Scatter for projected samples:

$$S_i^2 = \sum_{x(n) \in C_i} (W^T x(n) - M_i)(W^T x(n) - M_i)^T$$

$$M_i = W^T m_i$$

With little re-arrangement we will get:

$$S_i^2 = W^T \left(\sum_{x(n) \in C_i} (x(n) - m_i)(x(n) - m_i)^T \right) W$$

$$S_i^2 = W^T S_i W$$

So,

$$S_1^2 + S_2^2 = W^T (S_1 + S_2) W = W^T S_W W \quad \ldots\ldots\ldots (3)$$

S_W = within class scatter

Classification trees

When classifications or predictions of outcomes are present in the data mining work, classification tree methods[9] (also known as decision tree methods) are advised, and the objective is to produce rules that are simple to understand and can be converted into SQL or a natural query language.

Support vector classifier(svc)

A popular supervised learning technique for both classification and regression problems

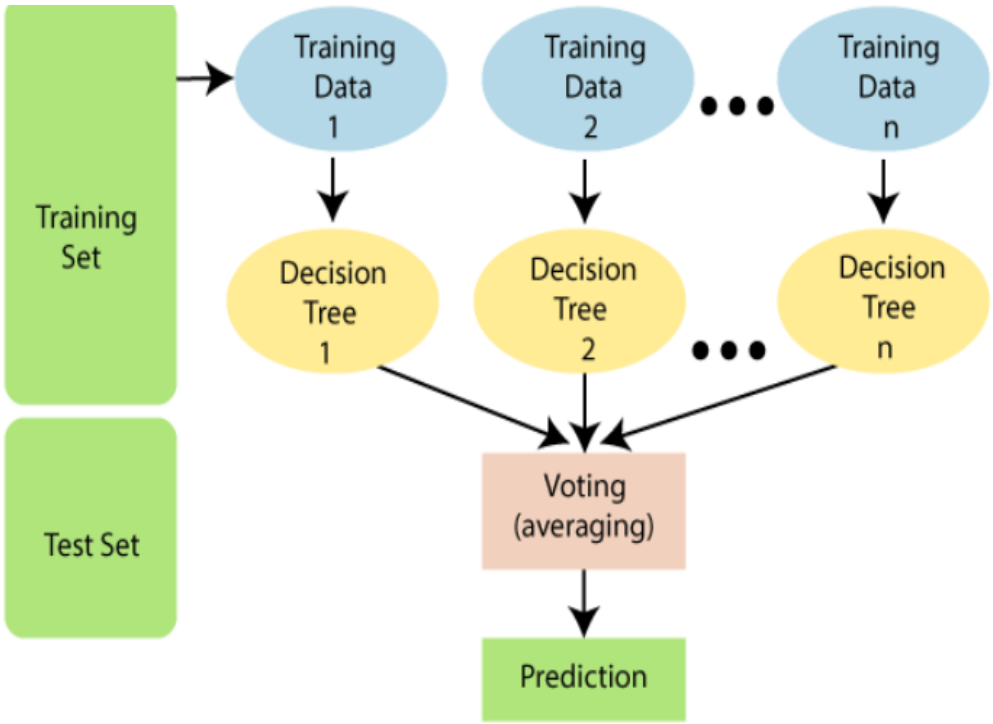
is called a Support Vector Machine (SVM). But, in machine learning, classification is where it is most frequently utilised. The primary goal of the SVM method is to identify the ideal boundary or decision line that may effectively divide the n-dimensional space into several categories, allowing us to appropriately categorise fresh data pieces in the future. A hyperplane is the name given to this ideal decision boundary. [10]

B(0) + sum (ai * (x,xi)) = f (x)

Random forest

Popular supervised learning algorithms known as random forests are associated with machine learning methods. They can be applied to ML issues involving classification and regression[11]. Ensemble learning is the idea behind random forests, It mixes various classifiers to address complicated issues and enhance model performance.

Operation of RFA:

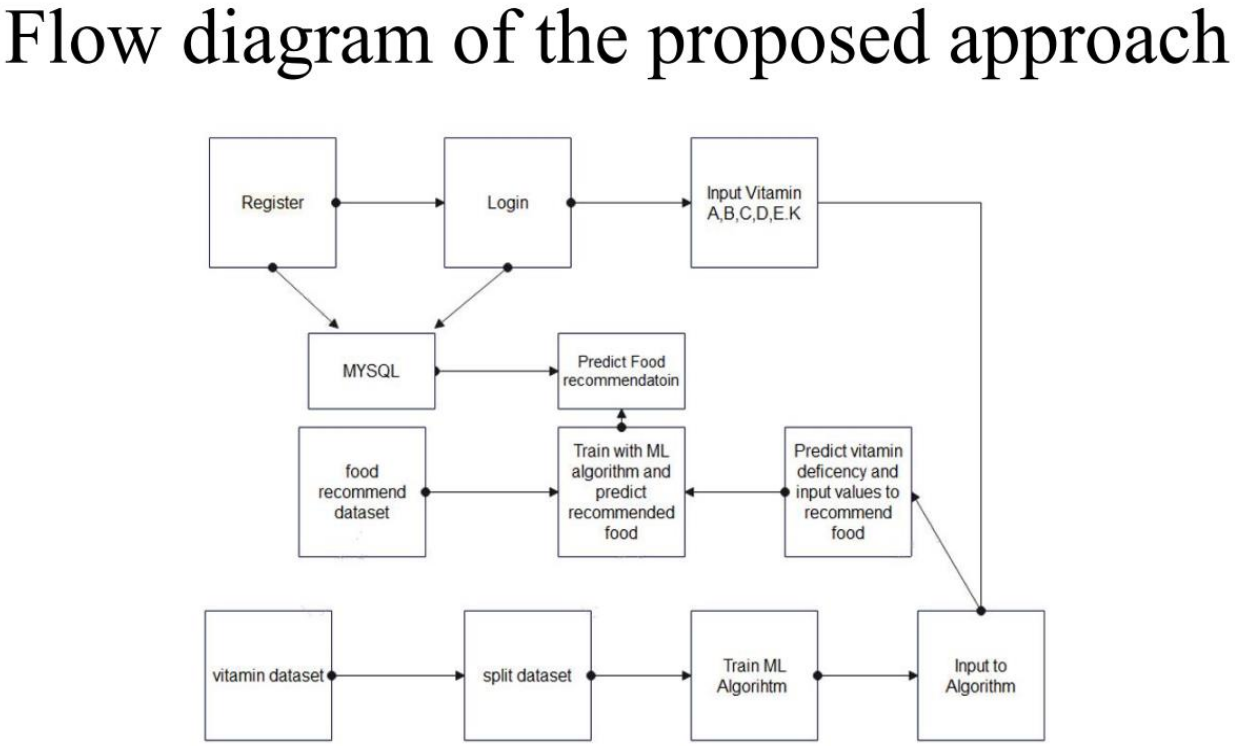


RF formula:

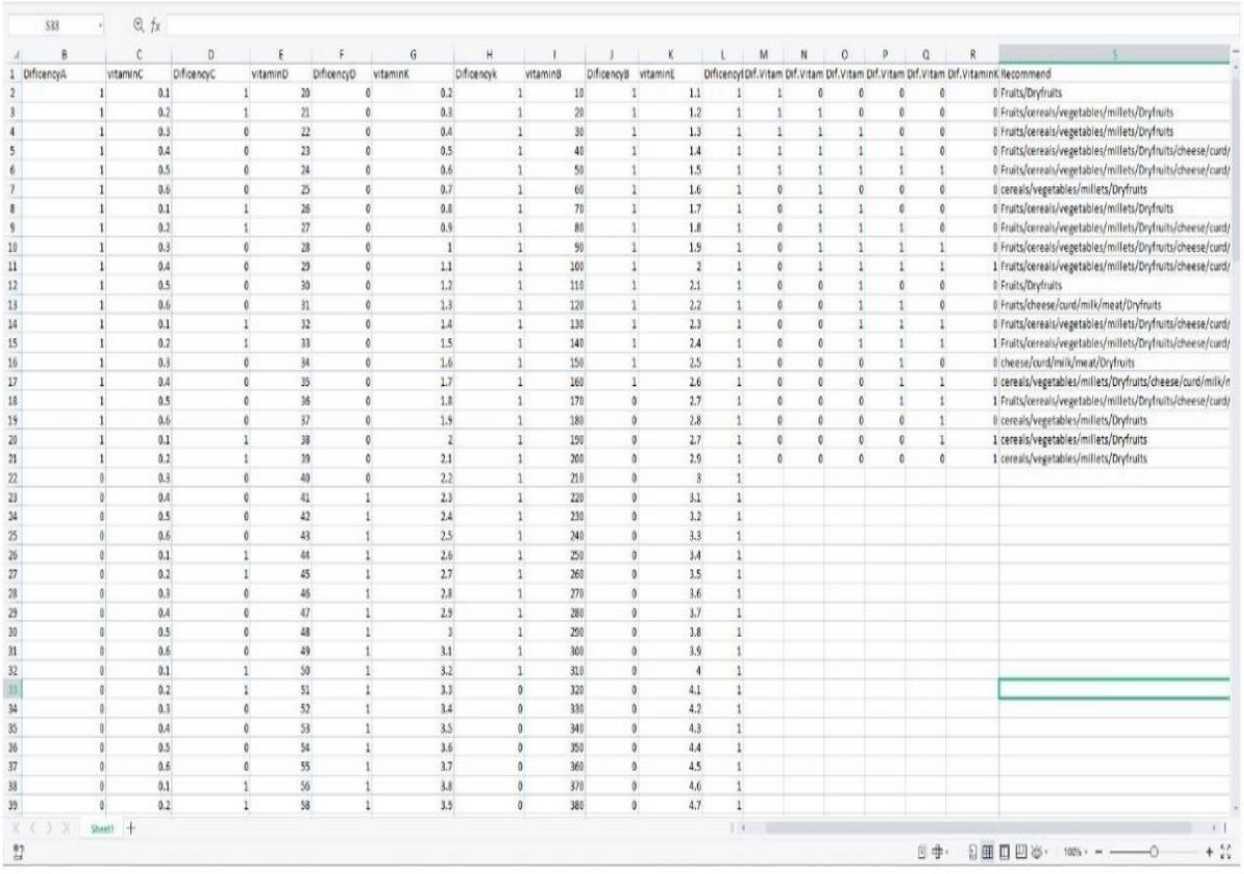
$$MSE = \frac{1}{N} \sum_{i=1}^{N} (fi - yi)^2$$

Where N is the number of data points, fi is the value returned by the model and yi is the actual value for data point i.

FLOW CHART:



DATA SETS :



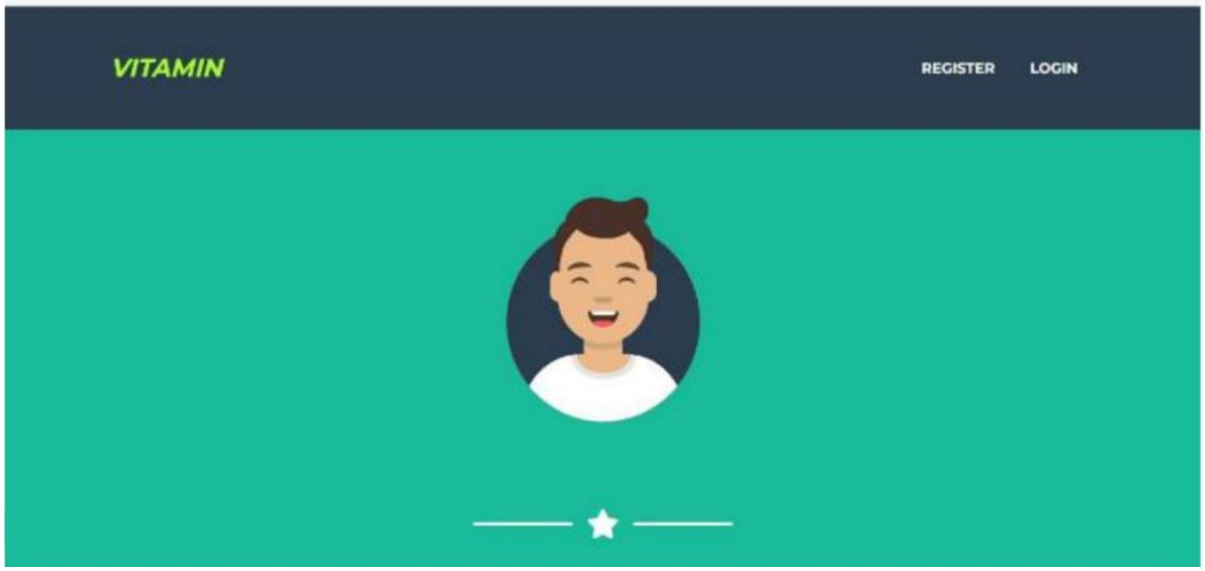
DificencyA	vitaminC	DificencyC	vitaminD	DificencyD	vitaminK	Dificencyk	vitaminB	DificencyB	vitaminE	DificencyE	Dif.Vitam A	Dif.Vitam B	Dif.Vitam C	Dif.Vitam D	Dif.Vitam E	Dif.VitaminK	Recommend
1	0.1	1	20	0	0.2	1	10	1	1.1	1	1	0	0	0	0	0	Fruits/Dryfruits
1	0.2	1	21	0	0.3	1	20	1	1.2	1	1	1	0	0	0	0	Fruits/cereals/vegetables/millets/Dryfruits
1	0.3	0	22	0	0.4	1	30	1	1.3	1	1	1	1	0	0	0	Fruits/cereals/vegetables/millets/Dryfruits
1	0.4	0	23	0	0.5	1	40	1	1.4	1	1	1	1	1	0	0	Fruits/cereals/vegetables/millets/Dryfruits/cheese/curd/
1	0.5	0	24	0	0.6	1	50	1	1.5	1	1	1	1	1	1	0	Fruits/cereals/vegetables/millets/Dryfruits/cheese/curd/
1	0.6	0	25	0	0.7	1	60	1	1.6	1	0	1	0	0	0	0	cereals/vegetables/millets/Dryfruits
1	0.1	1	26	0	0.8	1	70	1	1.7	1	0	1	1	0	0	0	Fruits/cereals/vegetables/millets/Dryfruits
1	0.2	1	27	0	0.9	1	80	1	1.8	1	0	1	1	1	0	0	Fruits/cereals/vegetables/millets/Dryfruits/cheese/curd/
1	0.3	0	28	0	1	1	90	1	1.9	1	0	1	1	1	1	0	Fruits/cereals/vegetables/millets/Dryfruits/cheese/curd/
1	0.4	0	29	0	1.1	1	100	1	2	1	0	1	1	1	1	1	Fruits/cereals/vegetables/millets/Dryfruits/cheese/curd/
1	0.5	0	30	0	1.2	1	110	1	2.1	1	0	0	1	0	0	0	Fruits/Dryfruits
1	0.6	0	31	0	1.3	1	120	1	2.2	1	0	0	1	1	0	0	Fruits/cheese/curd/milk/meat/Dryfruits
1	0.1	1	32	0	1.4	1	130	1	2.3	1	0	0	1	1	1	0	Fruits/cereals/vegetables/millets/Dryfruits/cheese/curd/
1	0.2	1	33	0	1.5	1	140	1	2.4	1	0	0	1	1	1	1	Fruits/cereals/vegetables/millets/Dryfruits/cheese/curd/
1	0.3	0	34	0	1.6	1	150	1	2.5	1	0	0	0	1	0	0	cheese/curd/milk/meat/Dryfruits
1	0.4	0	35	0	1.7	1	160	1	2.6	1	0	0	0	1	1	0	cereals/vegetables/millets/Dryfruits/cheese/curd/milk/n
1	0.5	0	36	0	1.8	1	170	0	2.7	1	0	0	0	1	1	1	Fruits/cereals/vegetables/millets/Dryfruits/cheese/curd/
1	0.6	0	37	0	1.9	1	180	0	2.8	1	0	0	0	0	1	0	cereals/vegetables/millets/Dryfruits
1	0.1	1	38	0	2	1	190	0	2.7	1	0	0	0	0	1	1	cereals/vegetables/millets/Dryfruits
1	0.2	1	39	0	2.1	1	200	0	2.9	1	0	0	0	0	0	1	cereals/vegetables/millets/Dryfruits
0	0.3	0	40	0	2.2	1	210	0	3	1							
0	0.4	0	41	1	2.3	1	220	0	3.1	1							
0	0.5	0	42	1	2.4	1	230	0	3.2	1							
0	0.6	0	43	1	2.5	1	240	0	3.3	1							
0	0.1	1	44	1	2.6	1	250	0	3.4	1							
0	0.2	1	45	1	2.7	1	260	0	3.5	1							
0	0.3	0	46	1	2.8	1	270	0	3.6	1							
0	0.4	0	47	1	2.9	1	280	0	3.7	1							
0	0.5	0	48	1	3	1	290	0	3.8	1							
0	0.6	0	49	1	3.1	1	300	0	3.9	1							
0	0.1	1	50	1	3.2	1	310	0	4	1							
0	0.2	1	51	1	3.3	0	320	0	4.1	1							
0	0.3	0	52	1	3.4	0	330	0	4.2	1							
0	0.4	0	53	1	3.5	0	340	0	4.3	1							
0	0.5	0	54	1	3.6	0	350	0	4.4	1							
0	0.6	0	55	1	3.7	0	360	0	4.5	1							
0	0.1	1	56	1	3.8	0	370	0	4.6	1							
0	0.2	1	58	1	3.9	0	380	0	4.7	1							

TEST CASES :

TC ID	Condition Being Tested	Expected Result	Result
1	Check for null values an duplicate records	Display the duplicate records and null values	Passed
2	Extract features using vectorisation	Display the features of the data	Passed
3	Check if the data is not empty and display records	Show error if it is empty else display the data	Passed
4	Display the number of records for training and test	Display the count of training and test data	Passed

GUI APPLICATION :

Home Page:



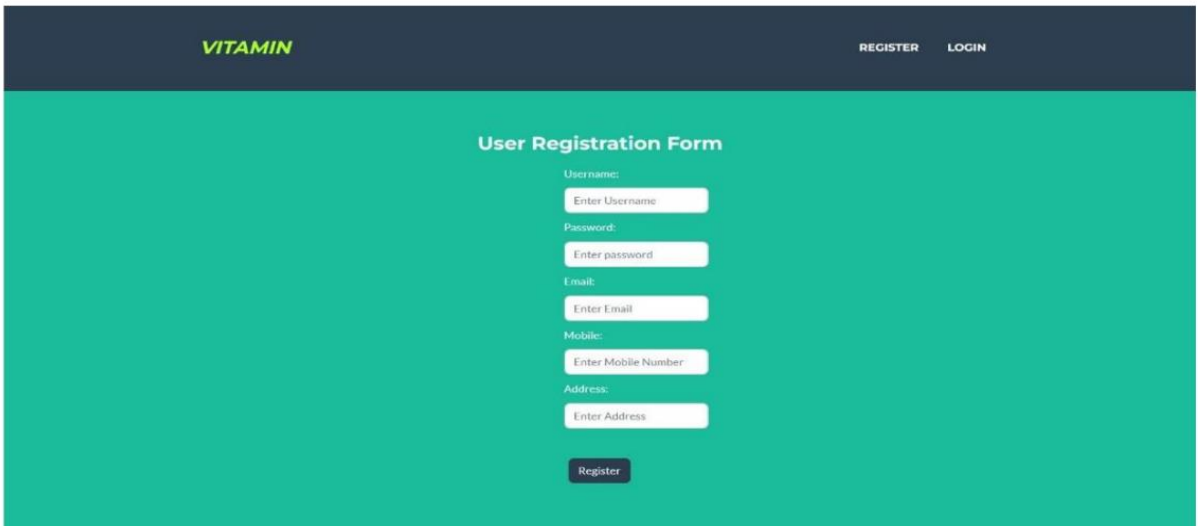
Step1- Register or Login: Home page contains two options on the top right corner. One is Register and another is Login. New user must choose Register option and existing user choose Login option.

Step2-User Registration Form: It contains various fields including Username, Password, Email, Mobile, Address. After entering all these details then you click Register button.

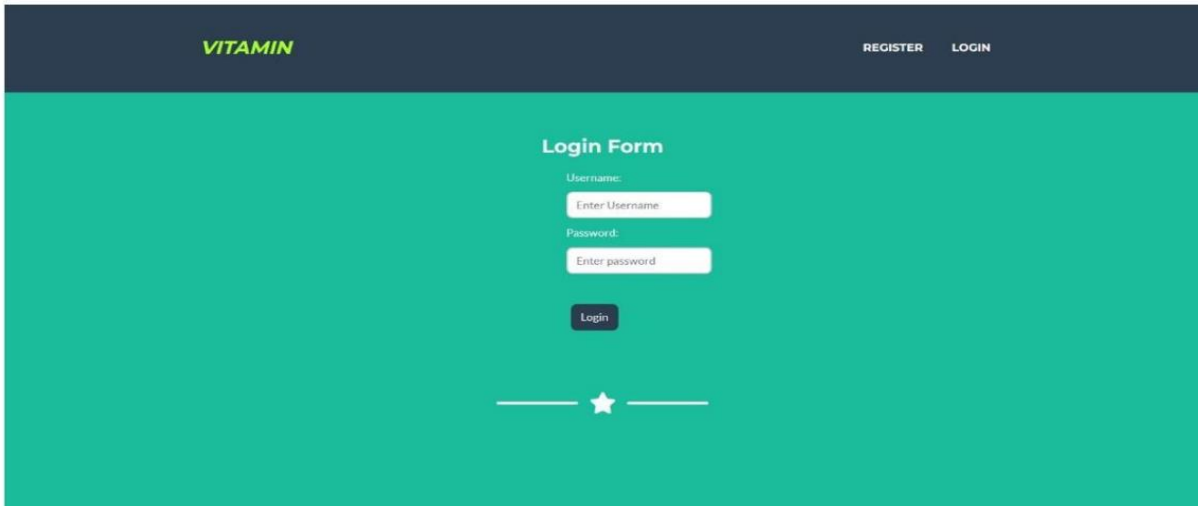
Step3- Login Form: Login form consists of 2 fields. One is username and another is password. After entering username and password then you click Login button.

Step4- Vitamin Form: After login you will see the vitamin form.

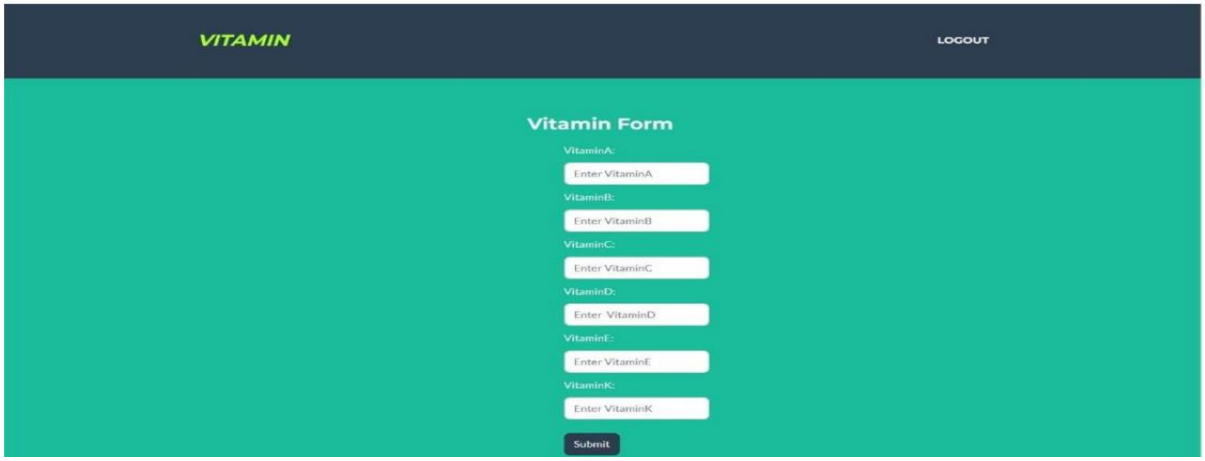
Sample Output 1



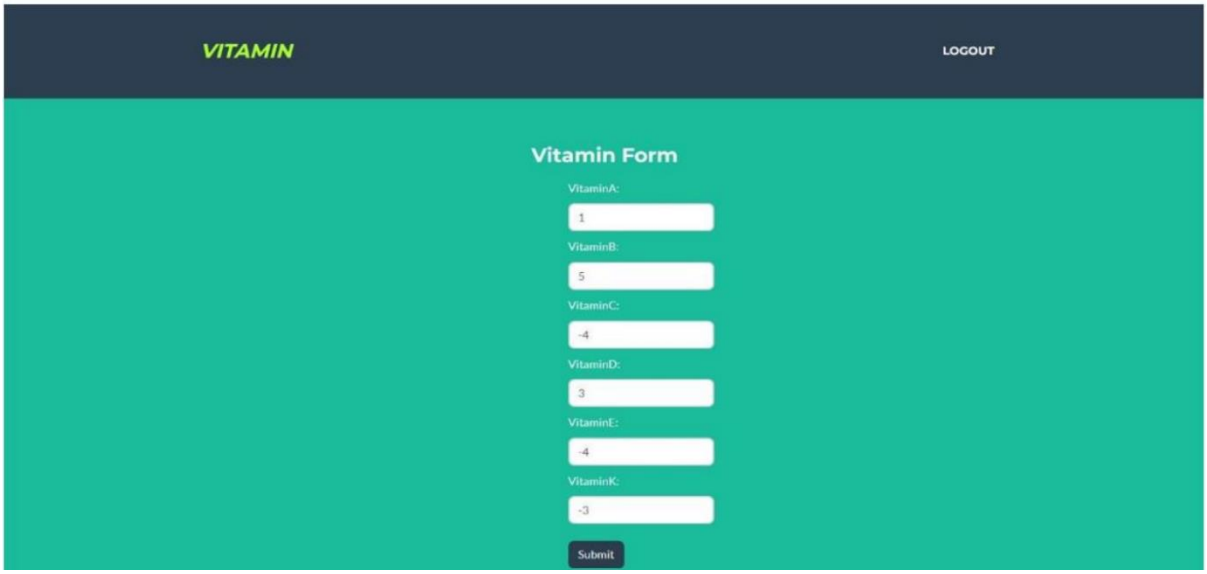
Sample Output 2



Sample Output 3



Sample Output 4



Step5- Enter Input Values in Vitamin Form: The vitamin form contains 6 vitamin names (A,B,C,D,E,K) and their corresponding boxes. In that boxes we have to enter the input values and then click submit. We pass these input values as arguments to the functions. The function algo_predict x(vit x1) , x may be 1,2,3,4,5 and x1 may be like A,B,C,D,E,K is useful for identifying whether the vitamin is deficient or in the range.

The another function algo_reco (difa,difb,difc,difd,dife,difk) is useful for suggests some food.

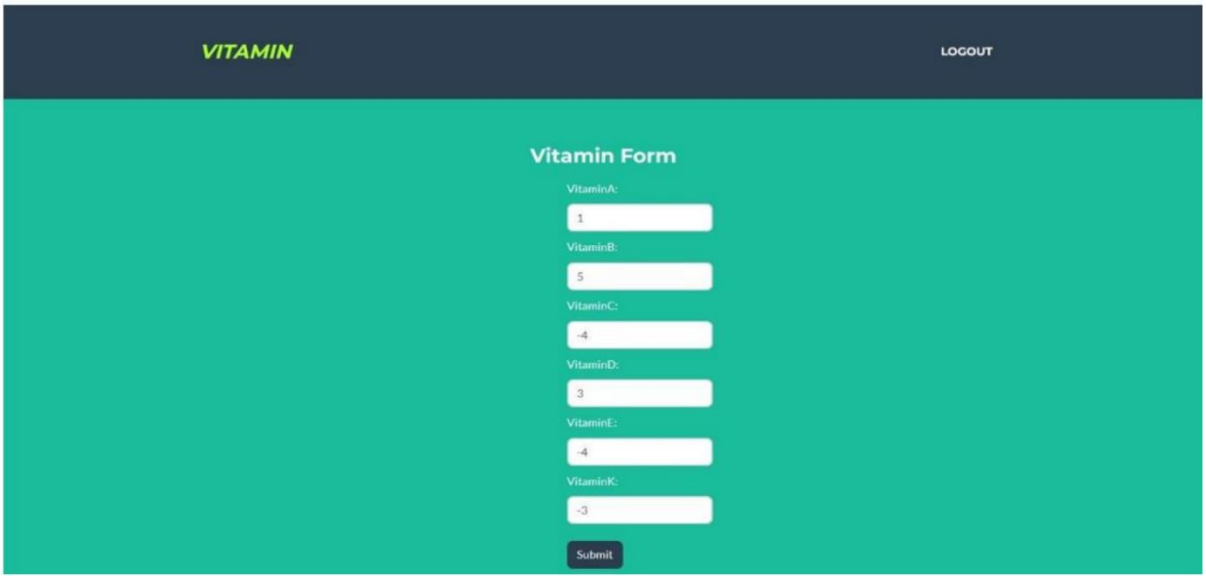
Sample Output 5

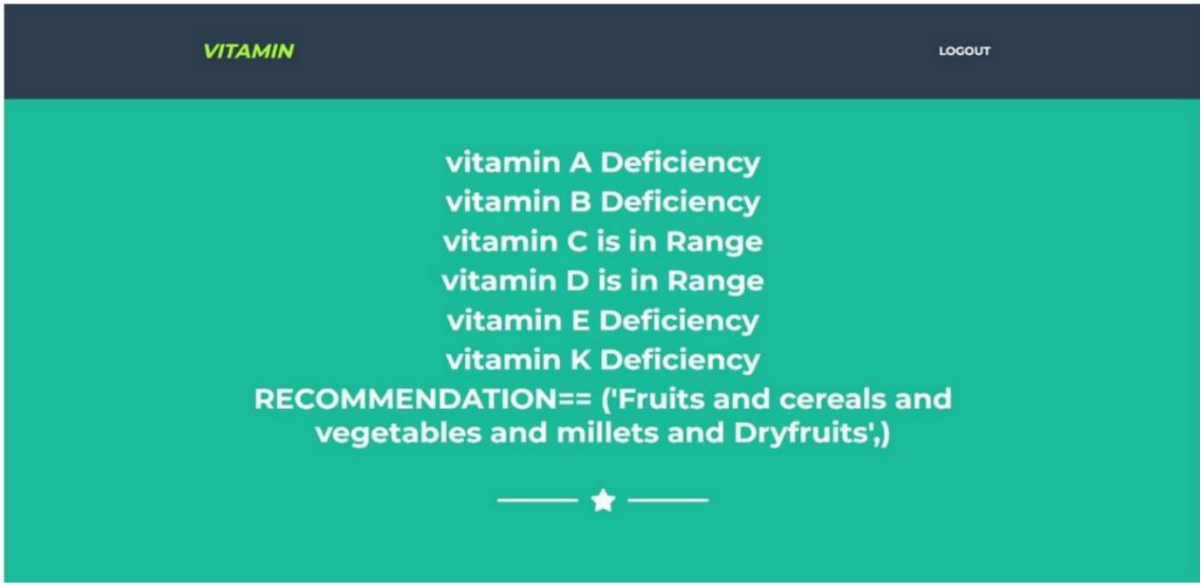
Step6- Check the Vitamin Deficiency and Diet Suggestion using Machine Learning Algorithms: The computer detects vitamin deficiency and proposes some food when the user enters vitamin values. If the input value is negative then it is a deficiency. Based on the input values it tells whether the vitamin is deficient or in the range decrypt-ed successfully. Similarly you can

enter any message and perform encryption and decryption.

RESULTS

Case1 : The vitamin form contains 6 vitamin names (A,B,C,D,E,K) and their corresponding boxes. In that boxes we have to enter the input values and then click submit. The computer detects vitamin deficiency and proposes some food when the user enters vitamin values. If the input value is negative then it is a deficiency.





Conclusion

We have created a website that recommends foods and predicts vitamin deficiencies, and we are implementing predictions by taking vitamin intake and their deficiencies. To train the system, the initial process was to prepare a data set of foods based on vitamin deficiencies. Different dietary recommendations are provided depending on the type of vitamin deficiency.After clustering is complete, a random forest classifier is employed to determine the food option that is most suitable for a balanced diet. Our system for

dietary recommendations enables users to obtain a healthy diet tailored to their specific vitamin deficiencies.

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] Shivani Singh, Sonal Bait, Jayashree Rathod, Prof. Nileema Pathak," Diabetes Prediction Using Random Forest Classifier And Intelligent Dietician " , International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 01 | Jan 2020.

[6] https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning

[7] https://www.javatpoint.com/logistic-regression-in-machine-learning

[8]https://www.analyticsvidhya.com/blog/2021/08/a-brief-introduction-to-linear-discriminant-analysis/

[9] https://www.frontlinesystems.com/classification-tree

[10] https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

[11] https://www.javatpoint.com/machine-learning-random-forest-algorithm