

IMPROVED PREDICTIVE LEARNING APPROACHES IN MEDICINAL SERVICES FOR CUSTOMIZED DIET SUGGESTION FRAMEWORK

A Project Report

Submitted to the Faculty of Engineering of
**JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA,
KAKINADA**

In partial fulfillment of the requirements for the award of the Degree of
BACHELOR OF TECHNOLOGY

In
COMPUTER SCIENCE AND ENGINEERING

By

V. S. N. S. GAYATHRI

(18481A05L6)

R. LAKSHMI NAVEEN KUMAR

(18481A05I4)

SHAIK SHABANA

(18481A05J8)

V. RAHUL NADH

(18481A05N1)

Under the guidance of

Dr. S. NARAYANA, M.Tech, PhD

Professor & Mentor(AS&A), Department of CSE



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SESHADRI RAO GUDLAVALLERU ENGINEERING COLLEGE
(An Autonomous Institute with Permanent Affiliation to JNTUK, Kakinada)
SESHADRI RAO KNOWLEDGE VILLAGE
GUDLAVALLERU – 521356
ANDHRA PRADESH
2021-2022**

SESHADRI RAO GUDLAVALLERU ENGINEERING COLLEGE

(An Autonomous Institute with Permanent Affiliation to JNTUK, Kakinada)
SESHADRI RAO KNOWLEDGE VILLAGE, GUDLAVALLERU

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING



CERTIFICATE

This is to certify that the project report entitled **“IMPROVED PREDICTIVE LEARNING APPROACHES IN MEDICINAL SERVICES FOR CUSTOMIZED DIET SUGGESTION FRAMEWORK”** is a bonafide record of work carried out by **V. S. N. S. GAYATHRI (18481A05L6), SHAIK SHABANA (18481A05J8), R. LAKSHMI NAVEEN KUMAR (18481A05I4), V. RAHUL NADH (18481A05N1)** under the guidance and supervision of **Dr. S. NARAYANA** in the partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Computer Science and Engineering of Jawaharlal Nehru Technological University Kakinada, Kakinada** during the academic year 2021-2022.

Project Guide
(Dr. S. NARAYANA)

Head of the Department
(Dr. M. BABU RAO)

External Examiner

ACKNOWLEDGMENT

The satisfaction that accompanies the successful completion of any task would be incomplete without the mention of people who made it possible and whose constant guidance and encouragement crown all the efforts with success.

We would like to express our deep sense of gratitude and sincere thanks to **Dr. S. Narayana Professor & Mentor(AS&A)**, Department of Computer Science and Engineering for his constant guidance, supervision and motivation in completing the project work.

We feel elated to express our floral gratitude and sincere thanks to **Dr. M. Babu Rao**, Head of the Department, Computer Science and Engineering for his encouragement all the way during analysis of the project. His annotations, insinuations and criticisms are the key behind the successful completion of the project work.

We would like to take this opportunity to thank our beloved principal **Dr. G. V. S. N. R. V. Prasad** for providing a great support for us in completing our project and giving us the opportunity for doing projects.

Our Special thanks to the faculty of our department and programmers of our computer lab. Finally, we thank our family members, non-teaching staff, attendants and our friends, who had directly or indirectly helped and supported us in completing our project in time.

Team members

V. S. N. S. Gayathri(18481A05L6)
Shaik Shabana (18481A05J8)
R. Lakshmi Naveen Kumar(18481A05I4)
V. Rahul Nadh(18481A05N1)

CONTENTS

TITLE	PAGENO
ABSTRACT	I
LIST OF FIGURES	II
LIST OF ABBREVIATIONS	III
CHAPTER 1: INTRODUCTION	01-03
1.1 INTRODUCTION	01-02
1.2 OBJECTIVES OF THE PROJECT	02
1.3 PROBLEM STATEMENT	02-03
CHAPTER 2: LITERATURE REVIEW	04-06
CHAPTER 3: PROPOSED METHOD	07-33
3.1 METHODOLOGY	07-14
3.2 IMPLEMENTATION	14-31
3.3 DATA PREPARATION	32-33
CHAPTER 4: RESULTS AND DISCUSSION	34-38
CHAPTER 5: CONCLUSION AND FUTURE SCOPE	39
BIBLIOGRAPHY / REFERENCES	40-41
LIST OF POs AND PSOs	42-43
PROJECT OUTCOMES	44
Mapping of Project outcomes with graduated POs and PSOs	44
Published paper in the project	45-51

ABSTRACT

In this cutting-edge world many human beings are struggling from extraordinary kinds of ailments and illnesses. A learn about through WHO reviews that insufficient and imbalanced consumption of meals motives round 9% of coronary heart assault deaths, about 11% of ischemic coronary heart ailment deaths, and 14% of gastrointestinal most cancers deaths worldwide. Moreover, round 0.25 billion teens are struggling from extraordinary sorts of nutrient deficiency specifically from Vitamin-A to Vitamin-K deficiency, 0.2 billion human beings are struggling from Iron deficiency (Anemia), and 0.7 billion human beings are struggling from Iodine deficiency. The primary goal of this undertaking is to propose a food regimen to one-of-a-kind folks the use of the datasets that are organized based totally on the aggregate of a number nutritional vitamins and their deficiency and meals to be endorsed based totally on which diet is deficient. In this mission a couple of classifier algorithms are used (KNN, Decision tree, Random forest, Logistic regression, Voting classifier). Ensembled algorithm is used to mix more than one algorithms and instruct a new algorithm. Accuracy of every algorithm is calculated and the high-quality algorithm is used for prediction purposes. Prediction is proven the use of flask net utility which will observe deficiency of diet and endorse kind of meals to be taken on a range of combos.

Keywords : Vitamin, Deficiency, Diet, Decision tree, Random forest

LIST OF FIGURES

FIG NO	FIGURE	PAGE NO
3.1	System Architecture	08
3.2	General Structure of Decision Tree	11
3.3	Working of the Random Forest Algorithm	13
3.4	Data Flow Diagram	15
3.5	Data Preprocessing	16
3.6	Testing & Training	17
3.7	Initializing Multiple Algorithms and training with Logistic Regression	18
3.8	Predict Data	19
3.9	Downloading process of Anaconda	27
3.10	License Agreement to download Anaconda	27
3.11	Advanced Installation options	28
3.12	Successful installation of Anaconda	28
3.13	Anaconda Prompt	29
3.14	Installing Python libraries in Anaconda prompt	29
3.15	Get a link to run app	30
3.16	Paste the link to run app	31
3.17	Vitamin Dataset	32
3.18	Recommend Dataset	33
4.1	Home page of the website	34
4.2	Registration page of a website	35
4.3	Login page of the website	36
4.4	Vitamin Form	37
4.5	Submit option for vitamin form	37
4.6	Vitamin Deficiency Prediction	38
4.7	Food Recommendation based on the vitamin deficiency	38

LIST OF ABBREVIATIONS

Abbreviation	Explanation
KNN	K Nearest Neighbors
np	NumPy
pd	Pandas
HTML	HyperText Markup Language
CSS	Cascading Style Sheets
OLS	Ordinary Least Squares
sklearn	Scikit Learn
ML	Machine Learning
URL	Uniform Resource Locator
WWW	World Wide Web

CHAPTER-1

INTRODUCTION

1.1 INTRODUCTION

The recommendation process has basically three stages that are Information Collection Phase, Learning Phase and Recommendation Phase. The information is firstly collected about a particular problem and the various solutions related to that problem are categorized. After the collection of information the Learning Phase comes in which various conclusions are made out of that information which is gathered and in the last phase i.e., Recommendation Phase an output is given in which various recommendations are made.

Balanced nutrition is an important aspect of a healthy lifestyle for people. Along with a balanced diet, regular physical exercise is crucial for a healthy life. Nowadays nutrition and health are often overlooked. The majority of people suffer from diabetes, heart disease, cancer, stroke etc. The diseases are almost directly related to unhealthy eating habits. So, our body needs nutrients to stay healthy, and food supplies essential nutrients that stop us from getting sick. A healthy, balanced diet will usually include vitamins, minerals, protein, healthy fats, proteins, carbohydrates, and fiber. A healthy food pyramid is a combination of plant foods, moderate amounts of animal products. Which includes vegetables, grains, fruits, oils and sweets, dairy, meat and beans. Generally, a person remains unaware of major causes behind deficiency or excess of various vital substances, such as calcium, proteins, and vitamins, and how to normalize such substances through a balanced diet. With the advantage of technology, the people can live a healthier lifestyle. The fast-food consumption rate is alarmingly high and this consequently has led to the intake of unhealthy food. This leads to various vitamin deficiency issues such as E, A,B,C,D,K, an 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 paced generation not everyone has the time and money.

By getting details of user present vitamin values this application can automate the process of detection of vitamin deficiency and food recommendation system using machine learning. Aim is to create our own vitamin deficiency data set for A,B,C,D,E,K and create food recommendation dataset and train both datasets with the five machine learning algorithms and find accuracy of each algorithm and use best accuracy model in a web application where vitamin deficiency and food recommendation is predicted based on given input values.

In this project its own data set is prepared based on various high and low values of vitamins from (vitamin a, b, c, d, e, k) and features are divided from normal and abnormal conditions of vitamins and labels are divided into 0 and 1 as normal and abnormal. Another dataset is prepared based on a combination of various vitamins and their deficiency and food to be recommended based on which vitamin is deficient. In this project multiple classifier algorithms are used (knn, decision tree, random forest, logistic regression, voting classifier) ensemble algorithm is used to combine multiple algorithms and train a new algorithm. Accuracy of each algorithm is calculated and the best algorithm is used for prediction purposes. Prediction is shown using flask web application which will detect deficiency of vitamins and recommend types of food to be taken on various combinations.

1.2 OBJECTIVES OF THE PROJECT

- To develop a model where vitamin deficiency and food recommendation is predicted based on given input values. Aim is to create our own vitamin deficiency data set for A,B,C,D,E,K and create food recommendation dataset and train both dataset with five machine learning algorithms and find accuracy of each algorithm and use best accuracy model in a web application where vitamin deficiency and food recommendation is predicted based on given input values.
- To attain accuracy higher than the existing models for predicting vitamin deficiency and recommending food for it.
- To overcome drawbacks of the existing system in recommending the food or diet based on the vitamin deficiency values.

1.3 PROBLEM STATEMENT

The fast-food consumption rate is alarmingly high and this consequently has led to the intake of unhealthy food. Content based food recommender system is proposed which recommends food recipes according to the preferences already given by the user. The preferred recipes of the user are fragmented into ingredients which are assigned ratings for the ingredients according to the stored users' preferences. The recipes with the matching ingredients are

recommended. The authors do not consider the nutrition factors and the balance in the diet. Moreover, chances of identical recommendation are also present because the preference of the users may not change on a daily basis so it causes imbalance in the diet which results in the nutrient deficiency.

This leads to various vitamin deficiency issues such as E, A,B,C,D,K , an increase in blood pressure, obesity, diabetes. Hence it has become very essential for people to have a good balanced nutritional healthy diet. But in this fast paced generation not everyone has the time and money to spend on a personal dietitian and nutritionist who will look upon and take care of their health by advising them a healthy diet plan according to the individual personal information. By getting details of user present vitamin values this application can automate the process of detection of vitamin deficiency and food recommendation system using machine learning. In this project its own data set is prepared based on various high and low values of vitamins from (vitamin a, b, c, d, e, k) and features are divided from normal and abnormal conditions of vitamins and labels are divided into 0 and 1 as normal and abnormal. Another dataset is prepared based on a combination of various vitamins and their deficiency and food to be recommended based on which vitamin is deficient.

CHAPTER-2

LITERATURE REVIEW

2.1 LITERATURE REVIEW

Raciel yera Toledo proposed a food recommender system considering nutritional information and user preferences. The meal plan for the user recommended using users' preferences. This tool manages both user preferences and nutritional information. Vijay Jaiswal proposes healthy food habits, eating patterns and calories burned count can be intake of nutrients and so on using the data mining tools. In this tool the hidden patterns and customer food taking habits are found from different data sources. In this tool decision tree learning algorithm, Random Tree algorithms are used on different datasets. H. Jiang proposed a system to calculate the daily calorie demand. The Knapsack algorithm is used for recommended diet combinations of users. Different from other diabetic diet recommendation systems, this system can rank the recommended diet combinations using the TOPSIS algorithm according to the user's food nutrition. Jung-Hyun Lee proposed a customized diet recommendation service managing heart diseases. This service provides customers customized general information, family history of diseases, seasonal food intakes. Rung-Ching Chen constructed a recipe ontology that defines some common diseases healing with a variety of food recommendations and an inference engine for customer health condition and a recipe ontology can be used for proper recipe recommendations on food priorities. FidelsonTanzil uses ABC algorithm to extract information from the database according to the user's requirements. Kmean and SOM algorithms are used on datasets. Mohd Afisi projected the ABC algorithm in Data Mining and tested compared to six traditional classification algorithms successfully and ABC proved as a suitable algorithm for recommendation. Xiaoyan Gao proposed the food recommendation problem on user choice recipe recommendation factors. By using a neural network-based solution on Ordered diet Recommendation.

The authors INGMAR WEBER and PALAKORN ACHANANUPARP [1] made an attempt to gain insights from machine learning - diet success prediction which would help people trying to stay fit and healthy by keeping a track on their dietary intake. The authors used public food diaries of more than 4,000 long-term active MyFitnessPal users to study the

characteristics of an unsuccessful diet. Concretely, authors trained a machine learning model to predict repeatedly being over or under self-set daily calorie goals and then look at which features contribute to the model's prediction, where research was centered around “quantified self ” data. The authors observed that classification performance was sufficient and the token-based model performed better than the category-based model and used such data feasibly for more in-depth data mining.

NANDISH SHAH and ISHANI SHAH [2] presented a proposal of healthy food habits and eating system based on web data mining, to discover hidden patterns and business strategies from their customer and web data, which would track eating habits and recommend the types of food that will improve the health and avoid the types of food that raise the risk of illness. The authors used data mining algorithms like classification, clustering, association rules, etc. in the data mining process to extract useful information about people’s eating habits. The nutritive structure of each kind of food was analyzed and the fat, energy, vitamin percentage in the recipe was calculated. Then they used the classification mining algorithm to process the composition data and give out the result whether the diet is healthy or not. As a result, personalized recommendations were suggested for each person.

How a coding system at the meal level might be analyzed by using data mining techniques was demonstrated by the authors AINE P. HEARTY AND MICHAEL J. GIBNEY [3] through this article. They evaluated the usability of supervised data mining methods to predict an aspect of dietary quality based on dietary intake with a food-based coding system and a novel meal-based coding system. The authors used Food consumption databases from the NorthSouth Ireland Food Consumption Survey 1997–1999. A healthy eating index (HEI) score was developed. Quintiles of the HEI based on combinations of foods were predicted by Artificial neural networks (ANNs) and decision trees. As a result, the ANN had a slightly higher accuracy than did the decision tree in relation to its ability to predict HEI. However, on the basis of the meal coding system, the decision tree had higher accuracies than did the ANN. Data mining was used by CHRISTY SAMUEL RAJU, SANCHIT V CHAVAN, KARAN PITHADIA, SHRADDHA SANKHE, PROF. SACHIN GAVHANE [4] to develop a Fitness Advisor System. “Fitness Advisor” developed by authors was a desktop application that advised the user according to his/her problem associated with body weight by an efficient diagnosis of the same and spreading proper awareness about the health hazards. The authors

considered different factors in the system such as height, weight, body type, sex, smoking, drinking, health condition, physical activity, sleeping hours etc. A combination of clustering, association and classification algorithms to effectively deliver the best possible expert advice to the user's problem was used by authors. The Apriori algorithm was used by authors for generating association rules. The final output of the system was expert's advice in terms of diet and exercise.

2.2 EXISTING SYSTEM

- Content based food recommender system is proposed which recommends food recipes according to the preferences already given by the user. The preferred recipes of the user are fragmented into ingredients which are assigned ratings according to the stored users' preferences. The recipes with the matching ingredient are recommended. The authors do not consider the nutrition factors and the balance in the diet. Moreover, chances of identical recommendation are also present because the preference of the user may not change on a daily basis.
- Tags and latent factors are used for android based food recommender systems. The system recommends personalized recipes to the user based on tags and ratings provided in user preferences. The proposed system used latent feature vectors and matrix factorization in their algorithm. Prediction accuracy is achieved by use of tags which closely match the recommendations with users' preferences. However, the authors do not consider nutrition in order to balance the diet of the user according to his needs.
- The above-mentioned diet recommendation systems are specifically dealing with some diseases or related to balance the diet plans. In case of food recommendation for specific diseases, the systems recommend different foods for patients without knowing the level of disease which may vary in different cases and cause severe effects on patients. Similarly, in the case of food recommendations to balance the diet, nutrition factors are ignored which are very much important to recommend food and a balanced diet.

CHAPTER 3

PROPOSED METHOD

3.1 METHODOLOGY

3.1.1 Proposed System :

The System works in a Machine Learning Environment, we use multiple machine learning algorithms to check accuracy of vitamin deficiency and food recommendation and the best model is used for prediction in flask web application. When a user enters vitamin values, the algorithm will predict deficiency in vitamins and recommend food.

- 1)It automates the process of vitamin deficiency detection and food recommendation.
- 2)Previous datasets are used for training and testing.
- 3) Accuracy of the model is improved compared to existing methods.

The data is divided into different sets and then trained for different models. The dataset was first divided into a training set (80%) and pre-training set(20%). The pre-training set was divided into pre-train(80%) and pre-test(20%). Now, the training set is further divided into train(80%) and validation set(20%). This train set is again divided into train(80%) and test set(20%). So, now I have train validation and test sets separate which are nonoverlapping. The pretrain set was used to find the best models for the given dataset. We took the best 4 models using the pretest set. Their performance was compared based on their mean absolute errors. Once the best 4 models were obtained, hyperparameters for these models were tuned and the best parameter was selected. In this project its own data set is prepared based on various high and low values of vitamins from (vitamin a, b, c, d, e, k) and features are divided from normal and abnormal conditions of vitamins and labels are divided into 0 and 1 as normal and abnormal. Another dataset is prepared based on a combination of various vitamins and their deficiency and food to be recommended based on which vitamin is deficient. Multiple classifier algorithms are used (knn, decision tree, random forest, logistic regression, voting classifier) ensemble algorithm is used to combine multiple algorithms and train a new algorithm. Accuracy of each algorithm is calculated and the best algorithm is used for prediction purposes. Prediction is shown using a flask web application which will detect deficiency of vitamins and recommend types of food to be taken on various combinations.

3.1.2 System Architecture :

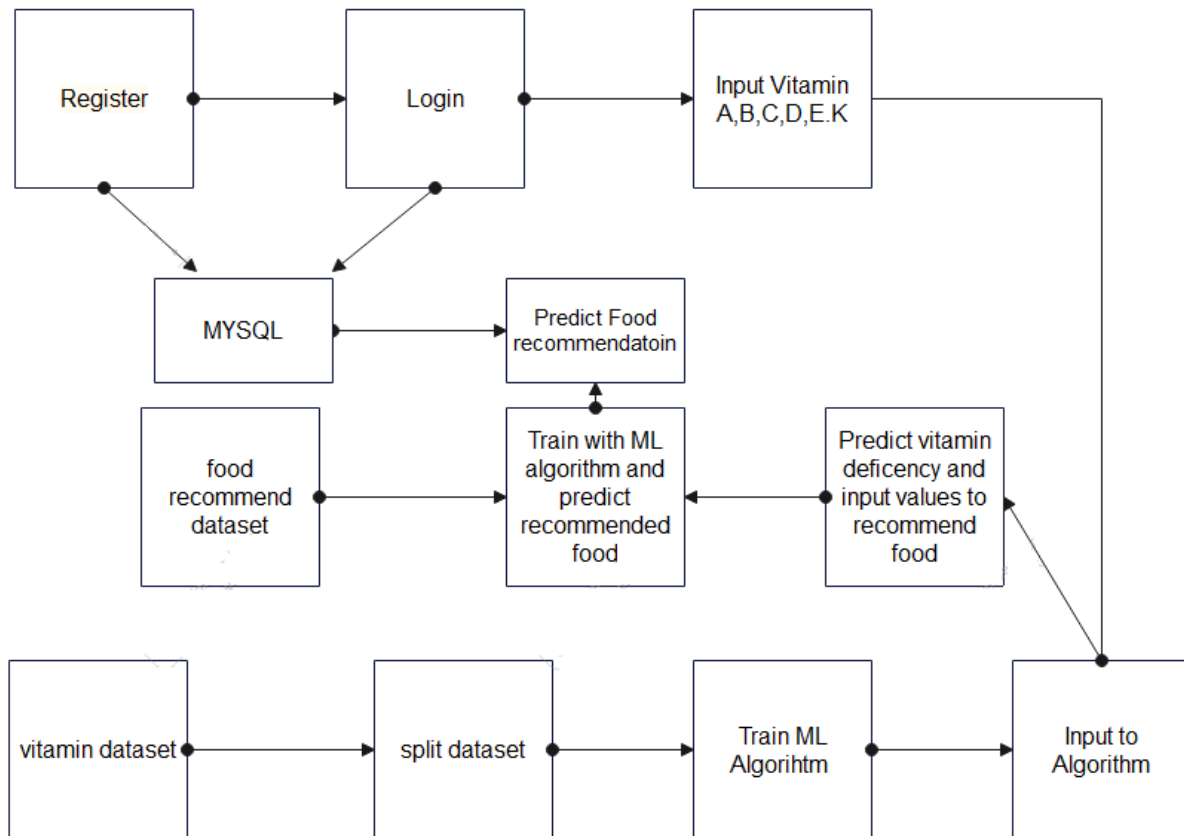


Figure 3.1 : System Architecture

A system architecture or systems architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system. Organized in a way that supports reasoning about the structures and behaviors of the system.

3.1.2.1 Three-Tier Architecture :

The three-tier software architecture (a three-layer architecture) emerged in the 1990s to overcome the limitations of the two-tier architecture. The third tier (middle tier server) is between the user interface (client) and the data management (server) components. This middle tier provides process management where business logic and rules are executed and can accommodate hundreds of users (as compared to only 100 users with the two tier architecture) by providing functions such as queuing, application execution, and database

staging. The three tier architecture is used when an effective distributed client/server design is needed that provides (when compared to the two tier) increased performance, flexibility, maintainability, reusability, and scalability, while hiding the complexity of distributed processing from the user. These characteristics have made three layer architectures a popular choice for Internet applications and net-centric information systems.

3.1.2.2 Advantages of Three-Tier :

- Separates functionality from presentation.
- Clear separation – better understanding.
- Hanges limited to well defined components.
- Can be running on WWW.
- Effective network performance.

3.1.3 KNN Classifier Algorithm :

The K-nearest neighbor method can be used for both regression and classification predictive problems. This method helps in interpreting output, calculating time and predictive power. Machine learning techniques are used in various fields. KNN is also one of the machine learning methods. This is also called a method of sample-based learning. This will contain the data of past datasets and can be used while predicting the new datasets. This will apply to functions called distance functions like Manhattan or Euclidean distance. This can be used to compute distance from samples to all other training samples. It calculates the target value for new samples. The target value will be the weighted sum of target values of the k nearest neighbors. The valve of K can be directly proportional to the prediction. Whenever the valve of K is small this indicates there is high variance and there is low bias. If the valve of the K is larger than this indicates that there is low variance and high bias. The main advantage of this KNN is it does not require any training or optimization. This KNN uses data samples when predicting the new datasets. Hence it is having higher complexity and also more time consumption.

This work represents a review of K-NN technique for the early prediction of food recommendation. K-NN analysis is used for predicting the unknown parameter from the known parameters. In this work we are considering vitamins as input parameters which are the main parameters to be considered for a good food recommendation, although there are

many other factors that can be considered [17]. The unknown value of vitamins can be predicted from the nearest known values of the nearest neighbors by calculation of Euclidean distance between them. Then we would be able to predict the type of food for given vitamin parameters. To measure the distance between points in a feature space, various distance functions can be used, in which the Euclidean distance function is the most widely used one[18]. Let p and q are represented as feature vectors. To calculate the distance between p and q , the Euclidean metric is generally used if $a=(a_1, a_2)$ and $b=(b_1, b_2)$ then the distance is given.

3.1.4 Logistic Regression :

This is a classification function that uses class for building and uses a single multinomial logistic regression model with a single estimator. Logistic regression usually states where the boundary between the classes exists, also states the class probabilities depend on distance from the boundary, in a specific approach. This moves towards the extremes (0 and 1) more rapidly when the data set is larger. These statements about probabilities make logistic regression more than just a classifier. It makes stronger, more detailed predictions, and can be fit in a different way; but those strong predictions could be wrong. Logistic regression is an approach to prediction, like Ordinary Least Squares (OLS) regression. However, with logistic regression, prediction results in a dichotomous outcome [13]. Logistic regression is one of the most commonly used tools for applied statistics and discrete data analysis. Logistic regression is linear interpolation.

3.1.5 Decision Tree :

- Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.
- In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node.
Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

- The decisions or the test are performed on the basis of features of the given dataset.
- It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.
- It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.
- In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.
- A decision tree simply asks a question, and based on the answer (Yes/No), it further splits the tree into subtrees.

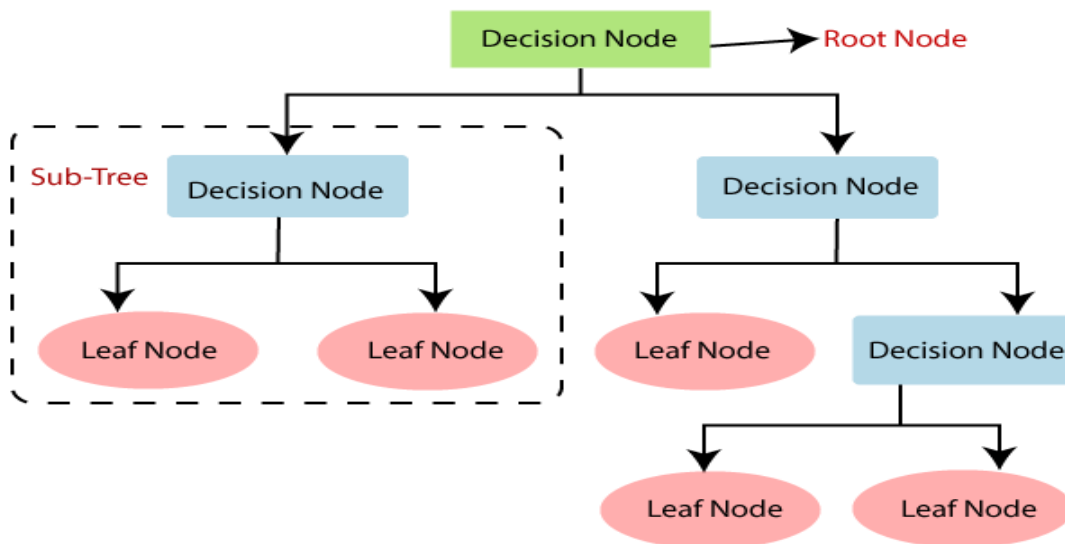


Figure 3.2 : General Structure of Decision Tree

There are various algorithms in Machine learning, so choosing the best algorithm for the given dataset and problem is the main point to remember while creating a machine learning model. Below are the two reasons for using the Decision tree:

- Decision Trees usually mimic human thinking ability while making a decision, so it is easy to understand.
- The logic behind the decision tree can be easily understood because it shows a tree-like structure.
- **Root Node** : Root node is from where the decision tree starts. It represents the entire dataset,

which further gets divided into two or more homogeneous sets.

- **Leaf Node** : Leaf nodes are the final output node, and the tree cannot be segregated further after getting a leaf node.
- **Splitting** : Splitting is the process of dividing the decision node/root node into sub-nodes according to the given conditions.
- **Branch/Subtree** : A tree formed by splitting the tree.
- **Pruning** : Pruning is the process of removing the unwanted branches from the tree.
- **Parent/Child node** : The root node of the tree is called the parent node, and other nodes are called the child nodes.

In a decision tree, for predicting the class of the given dataset, the algorithm starts from the root node of the tree. This algorithm compares the values of the root attribute with the record (real dataset) attribute and, based on the comparison, follows the branch and jumps to the next node. For the next node, the algorithm again compares the attribute value with the other sub-nodes and moves further. It continues the process until it reaches the leaf node of the tree.

3.1.6 Random Forest :

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

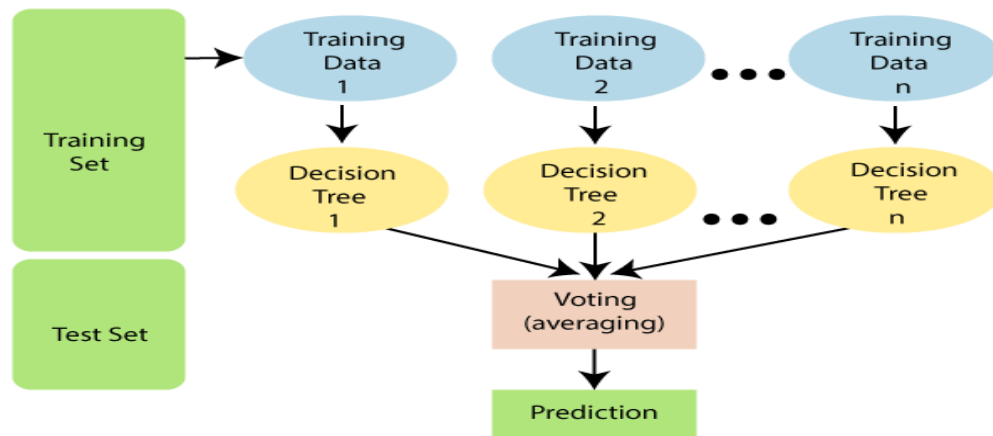


Figure 3.3 : Working of the Random Forest Algorithm

Random Forest works in two-phase first is to create the random forest by combining N decision trees, and second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps and diagram:

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (Subsets).

Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

There are mainly four sectors where Random forest mostly used :

1. Banking: Banking sector mostly uses this algorithm for the identification of loan risk.
2. Medicine: With the help of this algorithm, disease trends and risks of the disease can be identified.
3. Land Use: We can identify the areas of similar land use by this algorithm.
4. Marketing: Marketing trends can be identified using this algorithm.

Advantages of Random Forest :

- Random Forest is capable of performing both Classification and Regression tasks.
- It is capable of handling large datasets with high dimensionality.
- It enhances the accuracy of the model and prevents the overfitting issue.

3.1.7 Voting Classifier :

It is a machine learning estimator that trains various base models or estimators and predicts on the basis of aggregating the findings of each base estimator. The aggregating criteria can be combined with a decision of voting for each estimator output. A Voting Classifier is a machine learning model that trains on an ensemble of numerous models and predicts an output (class) based on their highest probability of chosen class as the output. It simply aggregates the findings of each classifier passed into Voting Classifier and predicts the output class based on the highest majority of voting. The idea is instead of creating separate dedicated models and finding the accuracy for each of them, we create a single model which trains by these models and predicts output based on their combined majority of voting for each output class.

3.2 IMPLEMENTATION

3.2.1 Software and Hardware Requirements :

The software and hardware components of a computer system that are required to install and use software efficiently. The software manufacturer will list the system requirements on the software package. If your computer system does not meet the system requirements then the software may not work correctly after installation. This web application may also be accessed using its URL. System requirements for operating systems will be hardware components, while other application software will list both hardware and operating system requirements. System requirements are most commonly seen listed as minimum and recommended requirements. The minimum system requirements need to be met for the software to run at all on your system, and the recommended system requirements, if met, will offer better software usability.

3.2.1.1 Software Requirements :

- Operating system : Windows XP/7/10.
- Coding Language : Python
- Tool : Anaconda
- Interface : Jupyter notebook
- Database Server : SQLyog
- Front-End Design : HTML,CSS,JS

3.2.1.2 Hardware Requirements :

- System : Intel(R) Core(TM) i3-7020U CPU @ 2.30GHz
- Hard Disk : 1 TB.
- Input Devices : Keyboard, Mouse
- RAM : 4 GB.

3.2.2 Module Description :

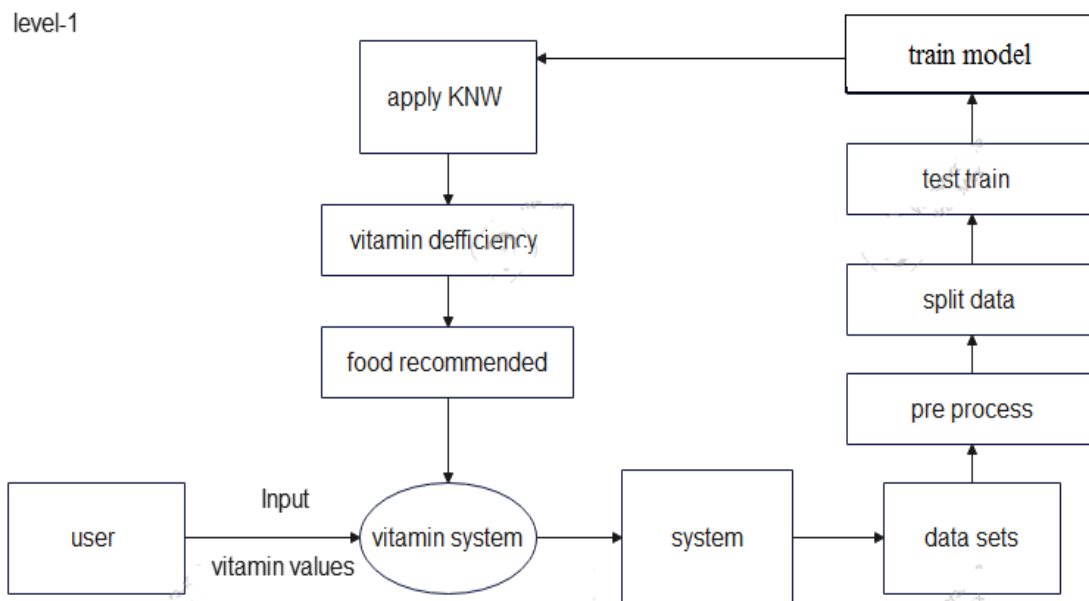


Figure 3.4 : Data Flow Diagram

Data collection :

- In this project we are using vitamin dataset and food recommendation dataset which is prepared based on min and max vitamin values from the test results and features are min and max values of vitamin a, b, c,d,e,k values and labels are deficiency and non-deficiency.
- Based on the vitamin deficiency food data set is prepared with various combinations. In this feature are vitamin deficient values and labels are types of food.

Data preprocessing :

- Features are extracted from the data set and stored in a variable as xtrain variable and labels are stored in y train variable. Data is preprocessed by standard scalar function and new features and labels are generated.

```
12 from sklearn.metrics import classification_report, confusion_matrix
13 from sklearn.metrics import accuracy_score
14
15 def process_tweet(tweet):
16     tweet = tweet.lower()
17     tweet = re.sub('@[^\s]+', '', tweet)
18     tweet = re.sub('((www\.[^\s]+)|(https?://[^\s]+))', ' ', tweet)
19     tweet = re.sub(r'\d+', ' ', str(tweet))
20     tweet = re.sub('&quot;', ' ', tweet)
21     tweet = emoji(tweet)
22     tweet = re.sub(r'[a-zA-Z]\b', ' ', str(tweet))
23     #for word in tweet.split():
24     #    if word.lower() in contractions:
25     #        tweet = tweet.replace(word, contractions[word.lower()])
26     tweet = re.sub(r'[^\w\s]', ' ', str(tweet))
27     tweet = re.sub(r'(.)\1+', r'\1', tweet)
28     tweet = re.sub(r'\s+', ' ', str(tweet))
29     return tweet
30
31 def emoji(tweet):
32     # Smile -- :) , : ) , :-), (:, ( : , (-: , :) , :D
33     tweet = re.sub(r'(:|:s?|:)|(:|:s?|:)|(:|:s?|:)|(:|:s?|:)|(:|:s?|:)', ' positiveemoji ', tweet)
34     # Laugh -- :D , :D , :-D , XD , X-D , XD , X-D
35     tweet = re.sub(r'(:|:s?|:)|(:|:s?|:)|(:|:s?|:)|(:|:s?|:)|(:|:s?|:)', ' positiveemoji ', tweet)
36     # Love -- <3 , :*
37     tweet = re.sub(r'(<3|:*)', ' positiveemoji ', tweet)
38     # Wink -- ;-), ;), ;-D, ;D, (;, (-: , @-
39     tweet = re.sub(r'(;|:|:s?|:)|(:|:s?|:)|(:|:s?|:)|(:|:s?|:)|(:|:s?|:)', ' positiveemoji ', tweet)
40     # Sad -- :-( , : ( , : ( , )- , : / , : /
41     tweet = re.sub(r'(:|:s?|:)|(:|:s?|:)|(:|:s?|:)|(:|:s?|:)|(:|:s?|:)', ' negativeemoji ', tweet)
42     # Cry -- :,( , : ( , : (
43     tweet = re.sub(r'(:|:s?|:)|(:|:s?|:)|(:|:s?|:)|(:|:s?|:)|(:|:s?|:)', ' negativeemoji ', tweet)
44     return tweet
45
```

Figure 3.5 : Data Preprocessing

Testing & Training :

- In this stage data is sent to the testing and training function and divided into four parts x test train, and y test train. Train variables are used for passing to algorithms whereas test variables are used for calculating accuracy of the algorithm.

```

38 # Wink -- ;-), ;), ;-D, ;D, (;, (-; , @-)
39 tweet = re.sub(r'(;;-?\)|;-?D|\\(-?;|@-\\))', ' positiveemoji ', tweet)
40 # Sad -- :-(, :(, :(, );, )-:, :-/, :-/
41 tweet = re.sub(r'(:\s?\\(|:-\\(|\\)\s?:|\\)-:|:-/|:-\\|)', ' negativeemoji ', tweet)
42 # Cry -- :, (, :(, :"(
43 tweet = re.sub(r'(:|\\(|:\\(|:"\\(|:"\\(|)', ' negativeemoji ', tweet)
44 return tweet
45
46
47 def prediction_result(datav):
48     total_data = pd.read_csv("food.csv", encoding="ISO-8859-1")
49
50     pd.set_option('display.max_colwidth', -1)
51     tweet = total_data.columns.values[2]
52     sentiment = total_data.columns.values[1]
53     total_data['processed_tweet'] = np.vectorize(process_tweet)(total_data[tweet])
54
55     count_vectorizer = CountVectorizer(ngram_range=(1,2)) # Unigram and Bigram
56     final_vectorized_data = count_vectorizer.fit_transform(total_data['processed_tweet'])
57     from sklearn.model_selection import train_test_split
58     X_train, X_test, y_train, y_test = train_test_split(final_vectorized_data, total_data[sentiment],
59                                                         test_size=0.2, random_state=69)
60
61     comment2 = [datav]
62     check = count_vectorizer.transform(comment2).toarray()
63     model_naive = MultinomialNB().fit(X_train, y_train)
64     predicted_naive = model_naive.predict(check)
65     print(predicted_naive)
66     navacc=model_naive.score(X_test,y_test)*100
67     print(classification_report(Y_test,predicted_naive))
68     print(confusion_matrix(Y_test,predicted_naive))
69     print("Model Accuracy ")
70     print(navacc)
71
72     if predicted_naive == 1:
73         result = 'Success'

```

Figure 3.6 : Testing & Training

Initializing Multiple Algorithms and training with Logistic Regression :

- In this stage machine learning algorithms are initialized and train values are given to the algorithm by this information algorithm will know what are features and what are labels. Then data is modeled and stored as a pickle file in the system which can be used for prediction.
- Data set is trained with multiple algorithms and accuracy of each model is calculated and best model is used for prediction


```
41 ypredict = clf.predict(x)
42 print("dddddddddddddddddddddddddddd")
43 #print(x[0])
44 print(ypredict)
45 return ypredict[0]
46
47 def algo_predict1(vitb):
48     x = np.array([vitb]).reshape(1, -1)
49     optimum = pd.read_excel("VITAMINDATASET.xlsx")
50
51     values = np.array([ optimum.vitaminB])
52
53     X = values.reshape(-1, 1)
54     Y = optimum.DificencyB
55
56     #print(optimum.vitaminA)
57     from sklearn.model_selection import train_test_split
58     X_train, X_test, y_train, y_test = train_test_split(X, Y,
59                                                         test_size=0.2)
60
61     #clf = RandomForestClassifier()
62     clf = KNeighborsClassifier(n_neighbors = 1, weights='uniform', algorithm='auto')
63
64     #clf=linear_model.LogisticRegression(fit_intercept=False)
65     #clf = DecisionTreeClassifier(criterion = "entropy", random_state = 1, splitter='best')
66     # Training the classifier
67     clf.fit(X_train, y_train)
68     acc=clf.score(X_test,y_test)*100
69     print(acc)
70     ypredict = clf.predict(x)
71     print("dddddddddddddddddddddddddddd")
72     #print(x[0])
73     print(ypredict)
74     return ypredict[0]
75
76
```

Figure 3.7 : Initializing Multiple Algorithms and training with Logistic Regression

Predict data :

- In this stage new data is taken as input and trained models are loaded using pickle and then values are preprocessed and passed to predict function to find out the result which is shown on web application.

```
126 @app.route("/track")
127 def track():
128     username = request.args.get('username')
129     vita = request.args.get('vita')
130     vitb = request.args.get('vitb')
131     vitc = request.args.get('vitc')
132     vitd = request.args.get('vit')
133     vite = request.args.get('vite')
134
135     data = vit_info(username,vita,vitb,vitc,vitd,vite)
136     print("dddddddddddddddddddddddddd")
137     print(data)
138     print("dddddddddddddddddddddddddd")
139     print(data)
140     return render_template("viewimage.html",m1="sucess",data=data)
141
142 #-----Track-----
143 # @app.route("/track")
144 # def track():
145 #     name = request.args.get('name')
146 #     iname = request.args.get('iname')
147
148 #     data = image_info(iname)
149 #     print("dddddddddddddddddddddddddd")
150 #     print(data)
151 #     data = v_image(data)
152 #     print("dddddddddddddddddddddddddd")
153 #     print(data)
154 #     return render_template("viewdata.html",m1="sucess",users=data)
155
156 #-----Update Item-----
157
158 if __name__ == "__main__":
159     app.run(debug=True, host='127.0.0.1', port=5000)
160
```

Figure 3.8 : Predict Data

3.2.3 Technology Description :

3.2.3.1 Python :

Python Language is used for this website.

Python is a popular programming language. It was created by Guido van Rossum, and released in 1991.

It is used for:

- web development (server-side),
- software development,
- mathematics,
- system scripting.

3.2.3.1.1 Why Python :

- ✓ Python works on different platforms (Windows, Mac, Linux, Raspberry Pi).

- ✓ Python has a simple syntax similar to the English language.
- ✓ Python has syntax that allows developers to write programs with fewer lines than some other programming languages.
- ✓ Python runs on an interpreter system, meaning that code can be executed as soon as it is written. This means that prototyping can be very quick.
- ✓ Python can be treated in a procedural way, an object-oriented way or a functional way.

To develop python and machine learning libraries we use **JUPYTER NOTEBOOK** in **ANACONDA**.

3.2.3.2 Anaconda :

Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine Learning applications, Large- scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. It is developed and maintained by Anaconda, Inc. The distribution includes data-science packages suitable for Windows, Linux, and macOS. Packaged versions are required and are managed by the package management system anaconda. This package manager was spun out as a separate open-source package as it ended up being useful on its own and for other things than Python.

3.2.3.2.1 What's included with Anaconda?

- An open-source package and environment management system called Conda, which
- makes it easy to install/update packages and create/load environments.
- Machine learning libraries like TensorFlow, scikit-learn and Theano.
- Data science libraries like pandas, NumPy and Dask .
- Visualization libraries like Bokeh, Datashader , matplotlib and Holoviews .
- Jupyter Notebook, a shareable notebook that combines live code, visualizations and text.

3.2.3.2.2 Jupyter Notebook :

Jupyter is a free, open-source, interactive web tool known as a computational notebook, which researchers can use to combine software code, computational output, explanatory text and multimedia resources in a single document. Computational notebooks have been around

for decades, but Jupyter in particular has exploded in popularity over the past couple of years. This rapid uptake has been aided by an enthusiastic community of user-developers and a redesigned architecture that allows the notebook to speak dozens of programming languages — a fact reflected in its name, which was inspired, according to co-founder Fernando Pérez, by the programming languages Julia (Ju), Python (Py) and R.

3.2.3.3 SQLyog :

SQLyog is a powerful MySQL development and administration solution, trusted by 2.5 million users around the world. With SQLyog, organizations have a tool that enables database developers, administrators, and architects to visually compare, optimize, and document schemas.

3.2.3.4 HTML :

The Hyper Text Markup Language or HTML is the standard markup language for documents designed to be displayed in a web browser. It can be assisted by technologies such as Cascading Style Sheets (CSS) and scripting languages such as JavaScript. Web browsers receive HTML documents from a web server or from local storage and render the documents into multimedia web pages. HTML describes the structure of a web page semantically and originally included cues for the appearance of the document.

3.2.3.5 CSS :

Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a markup language such as HTML. CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript. CSS is designed to enable the separation of presentation and content, including layout, colors, and fonts. This separation can improve content accessibility; provide more flexibility and control in the specification of presentation characteristics; enable multiple web pages to share formatting by specifying the relevant CSS in a separate. CSS file, which reduces complexity and repetition in the structural content; and enable the CSS file to be cached to improve the page load speed between the pages that share the file and its formatting.

3.2.3.6 JavaScript :

JavaScript often abbreviated JS, is a programming language that is one of the core technologies of the World Wide Web, alongside HTML and CSS. Over 97% of websites use JavaScript on the client side for web page behavior, often incorporating third-party libraries. All major web browsers have a dedicated JavaScript engine to execute the code on users' devices.

3.3.3.7 Python and Machine Learning Libraries :

import numpy as np :

- NumPy is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more.

import pandas as pd :

- Pandas is a high-level data manipulation tool developed by Wes McKinney. It is built on the Numpy package and its key data structure is called the DataFrame. DataFrames allow you to store and manipulate tabular data in rows of observations and columns of variables.

import matplotlib :

- Matplotlib is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack. It was introduced by John Hunter in 2002.

import sklearn :

- It is a library in Python that provides many unsupervised and supervised learning algorithms. We are using it for splitting the dataset into train and test sets.

import imblearn :

- In machine learning, while building a classification model we sometimes come to situations where we do not have an equal proportion of classes. This is called a class imbalance. All machine learning models are designed in such a way that they should attain maximum accuracy but in these types of situations, the model gets biased towards the majority class and will, at last, reflect on precision and recall. So how to build a model on these types of data set in a manner that the model should correctly classify the respective class and not get biased.
- To get rid of these imbalance class issues few techniques are used called the Imblearn Technique that is mainly used in these types of situations. Imblearn techniques help to either up sample the minority class or down sample the majority class to match the equal proportion. Through this article, we will discuss imblearn techniques and how we can use them to do upsampling and downsampling. For this experiment, we are using Pima Indian Diabetes data since it is an imbalance class data set. The data is available on Kaggle for downloading.

import sklearn.metrics as m :

There are 3 different APIs for evaluating the quality of a model's predictions:

Estimator score method : Estimators have a score method providing a default evaluation criterion for the problem they are designed to solve. This is not discussed on this page, but in each estimator's documentation.

Scoring parameter : Model-evaluation tools using cross-validation (such as `model_selection.cross_val_score` and `model_selection.GridSearchCV`) rely on an internal scoring strategy. This is discussed in the section The scoring parameter: defining model evaluation rules.

Metric functions : The metrics module implements functions assessing prediction error for specific purposes. These metrics are detailed in sections on Classification metrics, Multilabel ranking metrics, Regression metrics and Clustering metrics.

Finally, Dummy estimators are useful to get a baseline value of those metrics for random predictions.

from sklearn.cluster import KMeans :

- Clustering of unlabelled data can be performed with the module sklearn.cluster.
- Each clustering algorithm comes in two variants: a class that implements the fit method to learn the clusters on train data, and a function that, given train data, returns an array of integer labels corresponding to the different clusters. For the class, the labels over the training data can be found in the labels_ attribute.

from imblearn.under_sampling import RandomUnderSampler :

This bias in the training dataset can influence many machine learning algorithms, leading some to ignore the minority class entirely. This is a problem as it is typically the minority class on which predictions are most important.

One approach to addressing the problem of class imbalance is to randomly resample the training dataset. The two main approaches to randomly resampling an imbalanced dataset are to delete examples from the majority class, called undersampling, and to duplicate examples from the minority class, called oversampling.

- Random resampling provides a naive technique for rebalancing the class distribution for an imbalanced dataset.
- Random oversampling duplicates examples from the minority class in the training dataset and can result in overfitting for some models.
- Random undersampling deletes examples from the majority class and can result in losing information invaluable to a model.

from sklearn.preprocessing import StandardScaler :

- Standardize features by removing the mean and scaling to unit variance
- The standard score of a sample x is calculated as:
$$z = (x - u) / s$$
- where u is the mean of the training samples or zero if with_mean=False, and s is the standard deviation of the training samples or one if with_std=False.
- Centering and scaling happen independently on each feature by computing the relevant statistics on the samples in the training set. Mean and standard deviation are then stored to be used on later data using transform.

- Standardization of a dataset is a common requirement for many machine learning estimators: they might behave badly if the individual features do not more or less look like standard normally distributed data (e.g. Gaussian with 0 mean and unit variance).

from sklearn.model_selection import train_test_split :

- Before discussing train_test_split, you should know about Sklearn (or Scikit-learn). It is a Python library that offers various features for data processing that can be used for classification, clustering, and model selection.
- Model_selection is a method for setting a blueprint to analyze data and then using it to measure new data. Selecting a proper model allows you to generate accurate results when making a prediction.
- To do that, you need to train your model by using a specific dataset. Then, you test the model against another dataset.
- If you have one dataset, you'll need to split it by using the Sklearn train_test_split function first.
- train_test_split is a function in Sklearn model selection for splitting data arrays into two subsets: for training data and for testing data. With this function, you don't need to divide the dataset manually.
- By default, Sklearn train_test_split will make random partitions for the two subsets. However, you can also specify a random state for the operation.

import seaborn as sns :

- Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.
- For a brief introduction to the ideas behind the library, you can read the introductory notes. Visit the installation page to see how you can download the package and get started with it. You can browse the example gallery to see what you can do with seaborn, and then check out the tutorial and API reference to find out how.
- To see the code or report a bug, please visit the GitHub repository. General support questions are most at home on stackoverflow or discourse, which have dedicated channels for seaborn.

from sklearn.model_selection import cross_val_score :

- Learning the parameters of a prediction function and testing it on the same data is a methodological mistake: a model that would just repeat the labels of the samples that it has just seen would have a perfect score but would fail to predict anything useful on yet-unseen data. This situation is called overfitting. To avoid it, it is common practice when performing a (supervised) machine learning experiment to hold out part of the available data as a test set $X_{\text{test}}, y_{\text{test}}$. Note that the word “experiment” is not intended to denote academic use only, because even in commercial settings machine learning usually starts out experimentally. Here is a flowchart of typical cross validation workflow in model training. The best parameters can be determined by grid search techniques.

import tensorflow as tf :

To do machine learning TensorFlow, you are likely to need to define, save, and restore a model.

A model is, abstractly:

- A function that computes something on tensors (a forward pass)
- Some variables that can be updated in response to training

from tensorflow.keras import layers, models, optimizers, callbacks, regularizers, metrics :

When to use a Sequential model. A Sequential model is appropriate for a plain stack of layers where each layer has exactly one input tensor and one output tensor.

A Sequential model is not appropriate when:

- Your model has multiple inputs or multiple outputs
- Any of your layers has multiple inputs or multiple outputs
- You need to do layer sharing

from sklearn import metrics :

- Accuracy classification score. In multilabel classification, this function computes subset accuracy: the set of labels predicted for a sample must exactly match the corresponding set of labels in y_{true} .

3.2.4 Installation Steps :

3.2.4.1 Anaconda Installation Steps :

Steps to install anaconda on windows :

Step1 : Visit www.anaconda.com and download anaconda. After downloading click on the downloaded file. Then this page will appear.

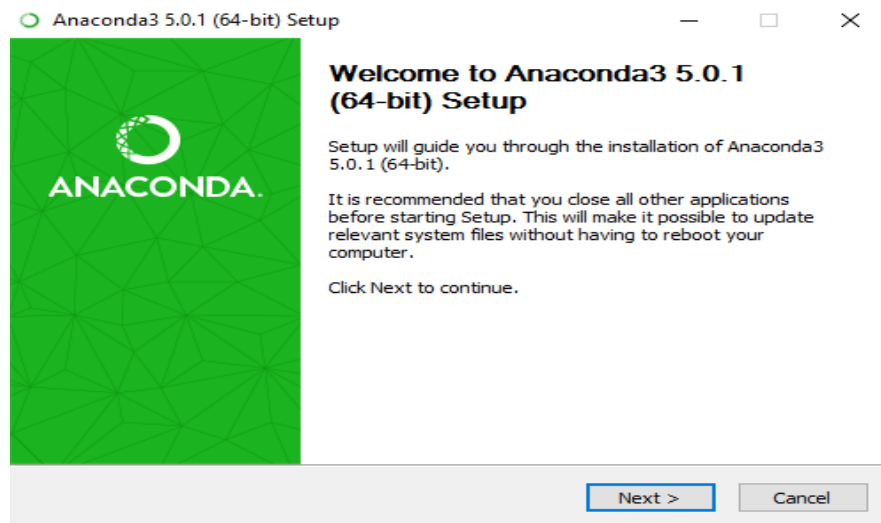


Figure 3.9 : Downloading process of Anaconda

Step2 : At the beginning of the install, you need to click Next to confirm the installation.

Step3 : Then agree to the license.

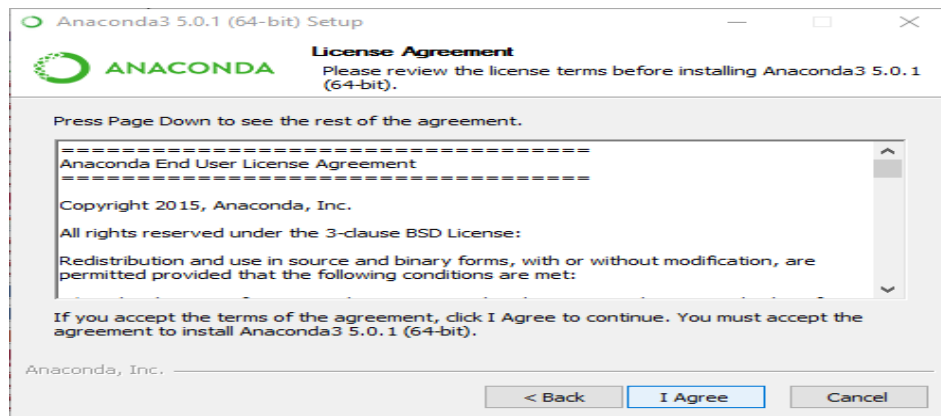


Figure 3.10 : License Agreement to download Anaconda

Step4 : At the advanced Installation Options screen, We recommend that you do not check “Add Anaconda to my PATH environment variable”.

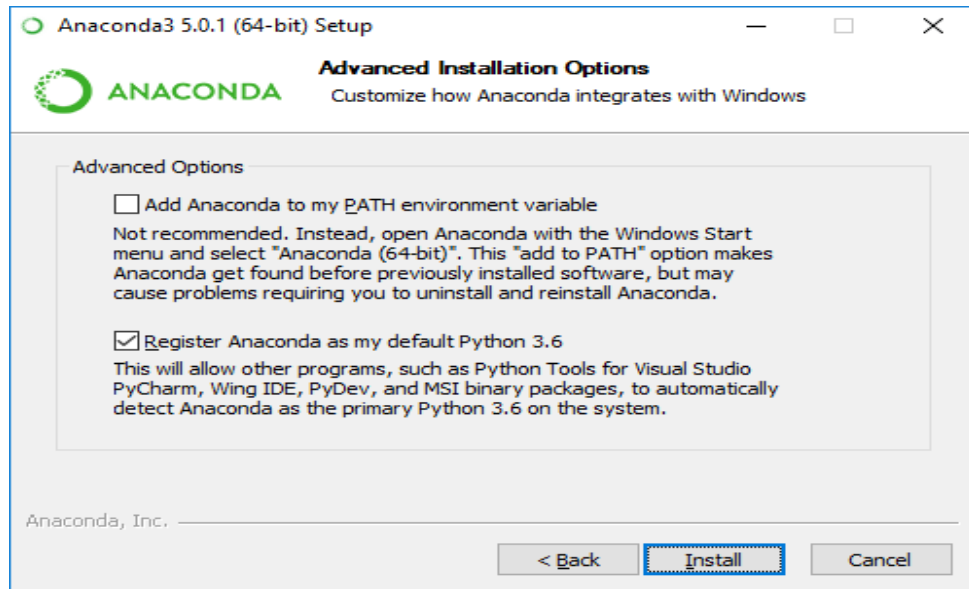


Figure 3.11 : Advanced Installation options

Step5 : After a successful installation you will see the “Thanks for installing Anaconda” dialog box.

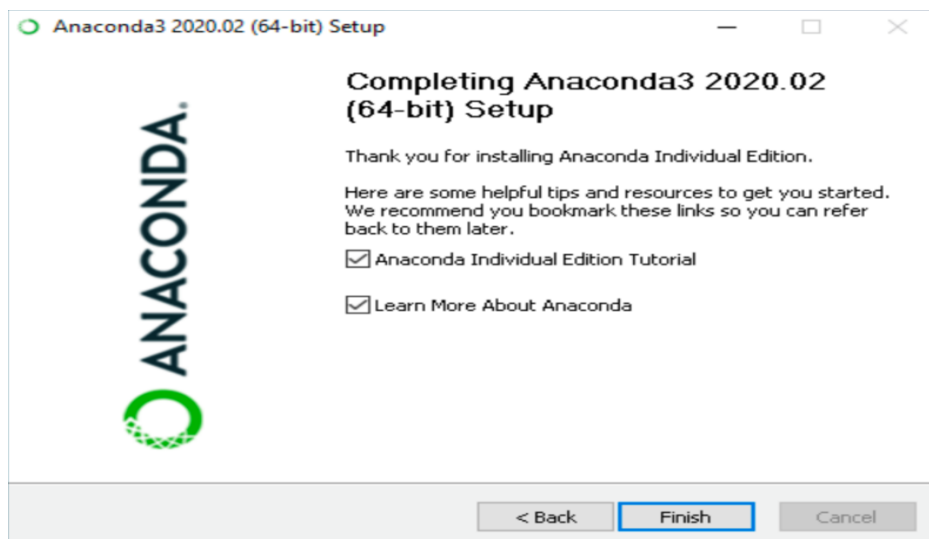


Figure 3.12 : Successful installation of Anaconda

Step6 : Click on finish. Installation of anaconda is completed.

3.2.4.2 Installing libraries in Anaconda :

Step1 : Go to the windows start menu and open the anaconda prompt.

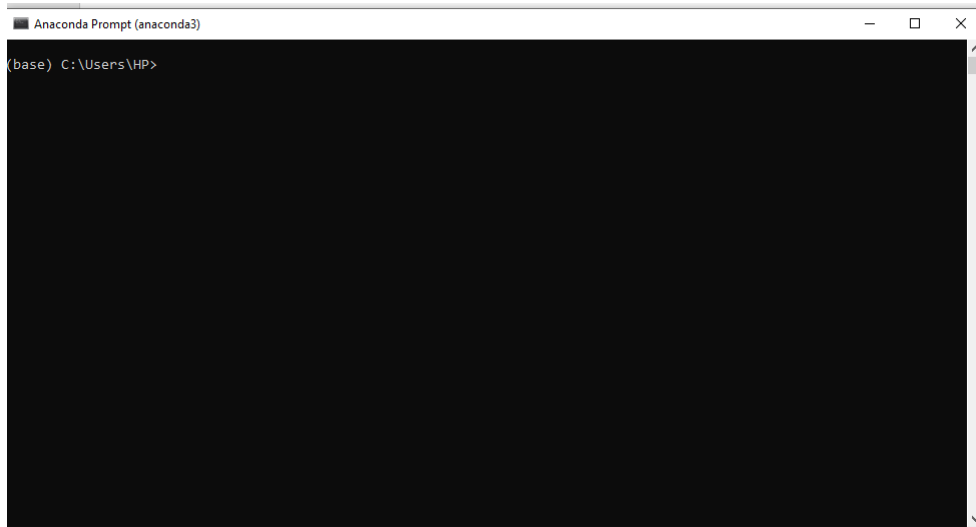


Figure 3.13 : Anaconda Prompt

Step2 : Install numpy, matplotlib, keras, mysqlldb etc.

Syntax: pip install library name,

Example: pip install numpy.

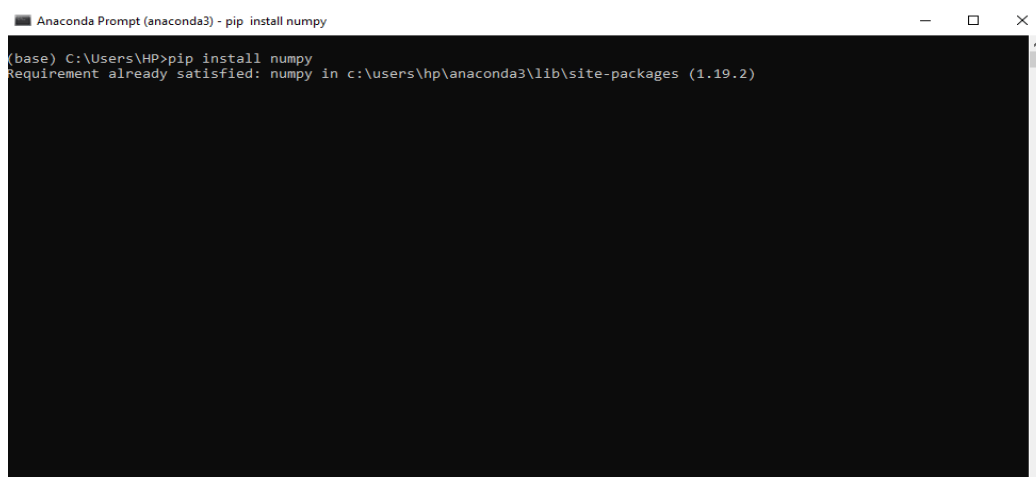


Figure 3.14 : Installing Python libraries in Anaconda prompt

3.2.5 Procedure for Execution :

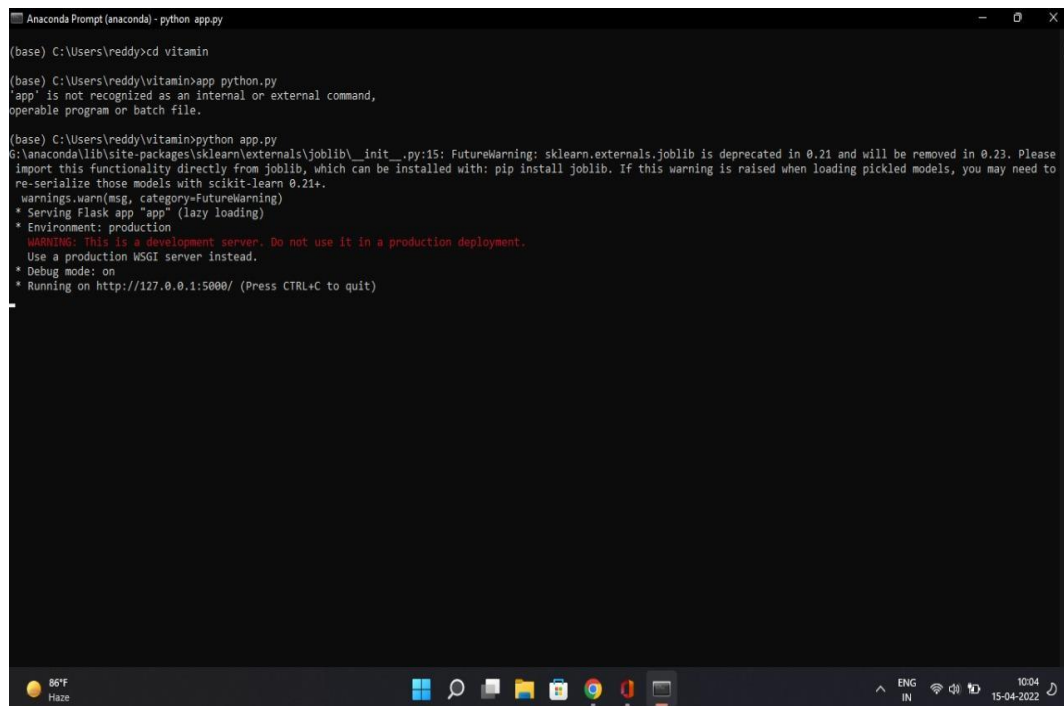
We can execute using the anaconda app. Below are the steps for execution,

Step1 : Open the anaconda prompt.

Step2 : Now execute the command “cd vitamin” to change the current directory to the location where our app.py file is stored.

Step3 : Now execute the app.py file.

Step4 : You will get a link as shown in the below figure.



```
Anaconda Prompt (anaconda) - python app.py
(base) C:\Users\reddy>cd vitamin
(base) C:\Users\reddy\vitamin>python app.py
'app' is not recognized as an internal or external command,
operable program or batch file.
(base) C:\Users\reddy\vitamin>python app.py
G:\anaconda\lib\site-packages\sklearn\externals\joblib\__init__.py:15: FutureWarning: sklearn.externals.joblib is deprecated in 0.21 and will be removed in 0.23. Please
import this functionality directly from joblib, which can be installed with: pip install joblib. If this warning is raised when loading pickled models, you may need to
re-serialize those models with scikit-learn 0.21+.
warnings.warn(msg, category=FutureWarning)
* Serving Flask app "app" (lazy loading)
* Environment: production
  WARNING: This is a development server. Do not use it in a production deployment.
  Use a production WSGI server instead.
* Debug mode: on
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```

Figure 3.15 : Get a link to run app

Step5 : Paste the link in any web browser.

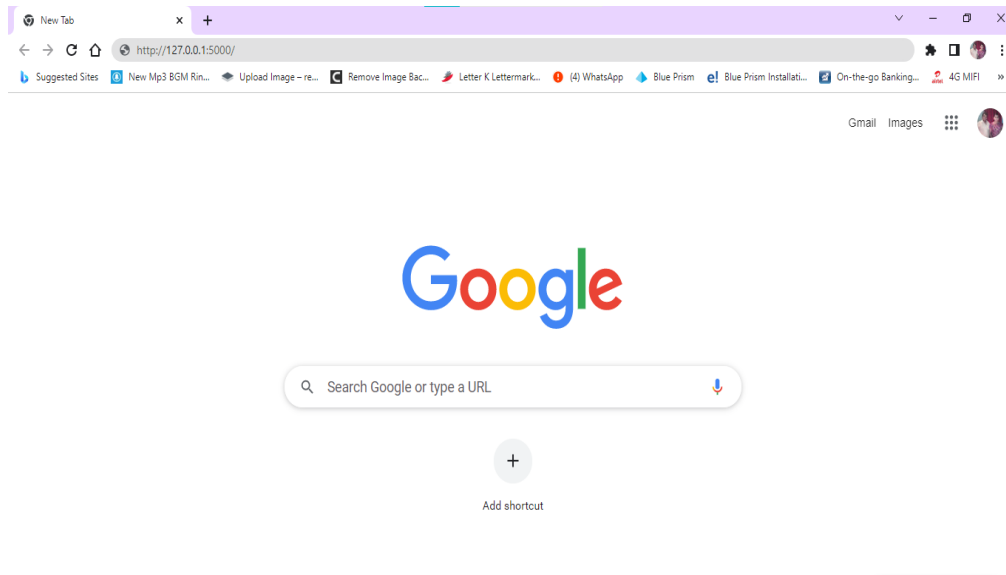


Figure 3.16 : Paste the link to run app

Step6 : Then the app will open.

3.3 DATA PREPARATION

There are two datasets. They are :

1. Vitamin Dataset
2. Recommend Dataset

Vitamin Dataset contains vitamin A, B, C, D, E, K values and their calculated deficiency values and food to be recommended.

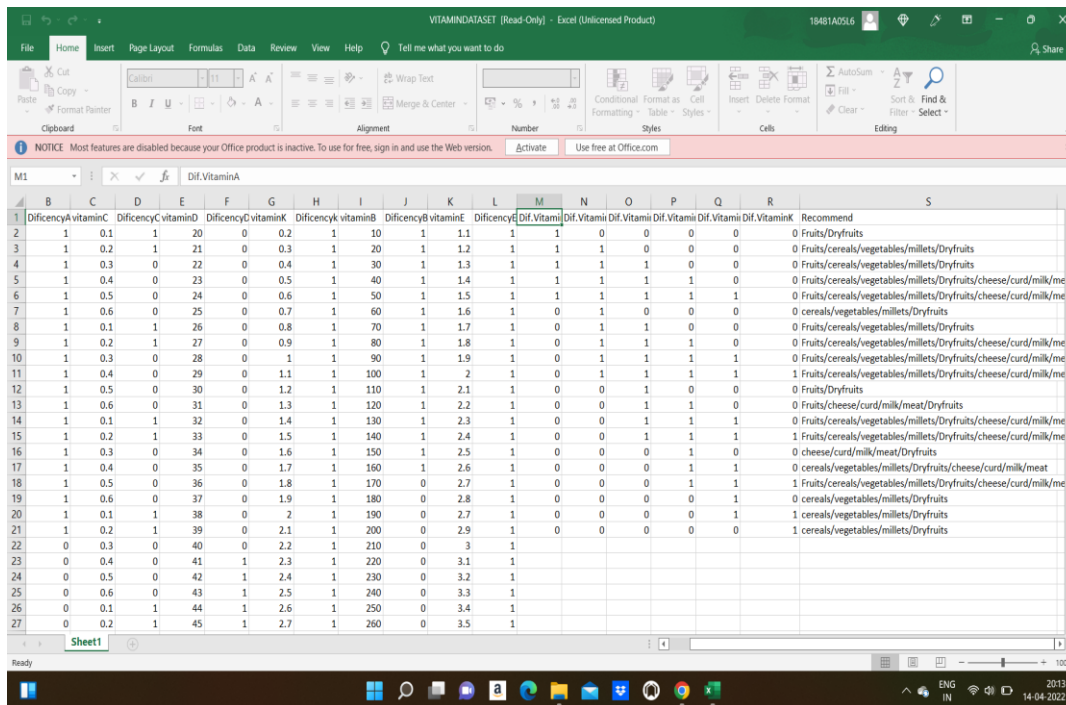


Figure 3.17 : Vitamin Dataset

Recommend Dataset contains parameters or values for recommending the food for the predicted deficiencies in vitamin A, B, C, D, E, K values.

Improved Predictive Learning Approaches in Medicinal Services for Customized Diet Suggestion Framework

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
166	1	1	1	1	1	1	0	4															
167	0	1	0	0	0	0	0	6															
168	0	1	1	0	0	0	0	2															
169	0	1	1	1	0	0	0	4															
170	0	1	1	1	1	0	0	4															
171	0	1	1	1	1	1	1	4															
172	0	0	1	0	0	0	0	1															
173	0	0	1	1	0	0	0	2															
174	0	0	1	1	1	0	0	4															
175	0	0	1	1	1	1	1	4															
176	0	0	0	1	0	0	0	3															
177	0	0	0	1	1	0	0	5															
178	0	0	0	1	1	1	1	4															
179	0	0	0	0	1	0	0	6															
180	0	0	0	0	1	0	0	1	6														
181	0	0	0	0	0	0	1	6															
182	1	0	0	0	0	0	0	1															
183	1	1	0	0	0	0	0	2															
184	1	1	1	0	0	0	0	2															
185	1	1	1	1	0	0	0	4															
186	1	1	1	1	1	1	0	4															
187	0	1	0	0	0	0	0	6															
188	0	1	1	0	0	0	0	2															
189	0	1	1	1	0	0	0	4															
190	0	1	1	1	1	1	0	4															
191	0	1	1	1	1	1	1	4															
192	0	0	1	0	0	0	0	1															

Figure 3.18 : Recommend Dataset

CHAPTER 4

RESULTS AND DISCUSSION

4.1 HOME PAGE :

Home page contains “Register and Login” options. Login option used to login into the Vitamin Deficiency and Food Recommendation System and Register option is used to register in the Vitamin Deficiency and Food Recommendation System.

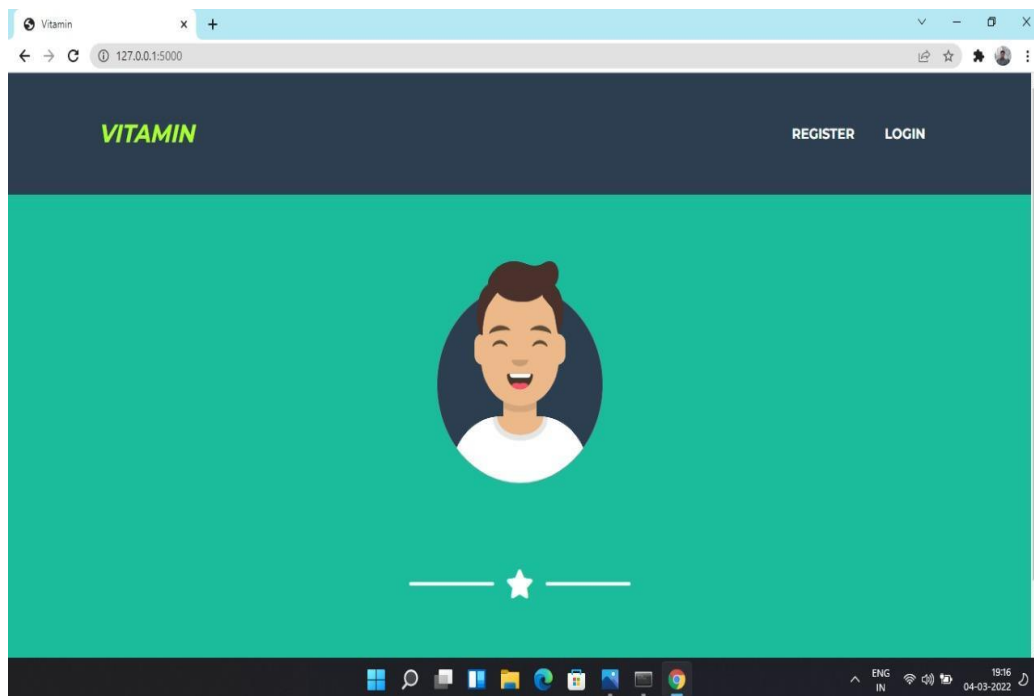


Figure 4.1 : Home page of the website

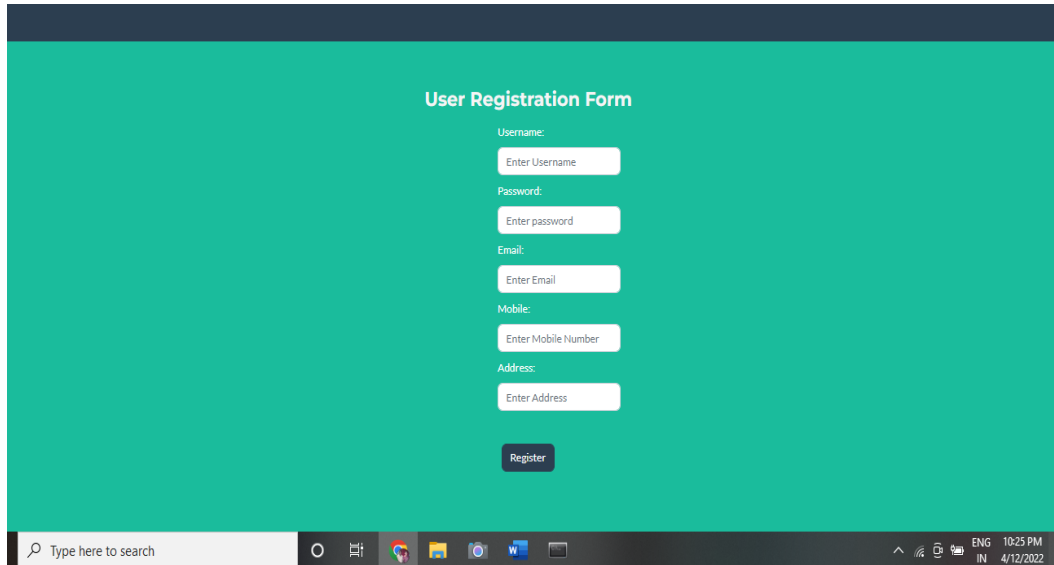
4.2 REGISTRATION PAGE :

The new user will register by clicking on the Register option. By clicking on the Register button, user details are stored on the database.

Registration form contains

- Username
- Password
- Email

- Mobile Number
- Address.



The image shows a web browser window displaying a "User Registration Form". The form is centered on a teal background. It includes input fields for Username, Password, Email, Mobile, and Address, each with a placeholder text "Enter [field name]". Below these fields is a dark blue "Register" button. The browser's taskbar at the bottom shows a search bar, several application icons, and system status icons on the right indicating the time as 10:25 PM on 4/12/2022.

Figure 4.2 : Registration page of a website

4.3 LOGIN PAGE :

Every user has his/her own credentials to login into the Vitamin Deficiency and Food Recommendation System. Login Form contains username and password.

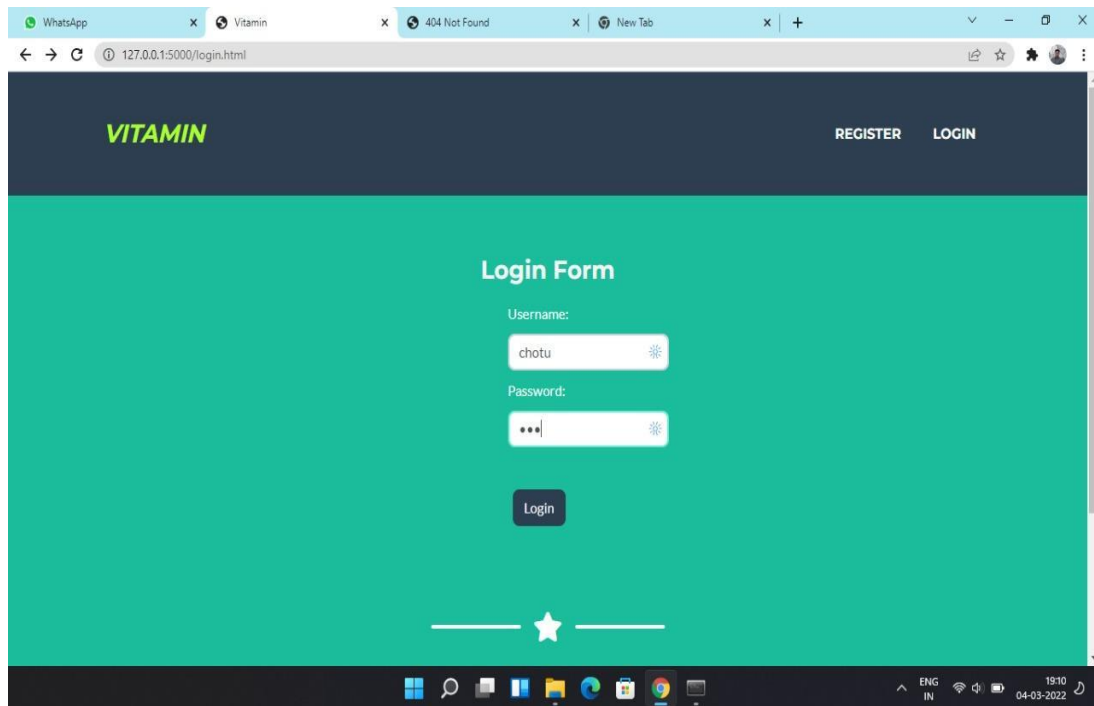


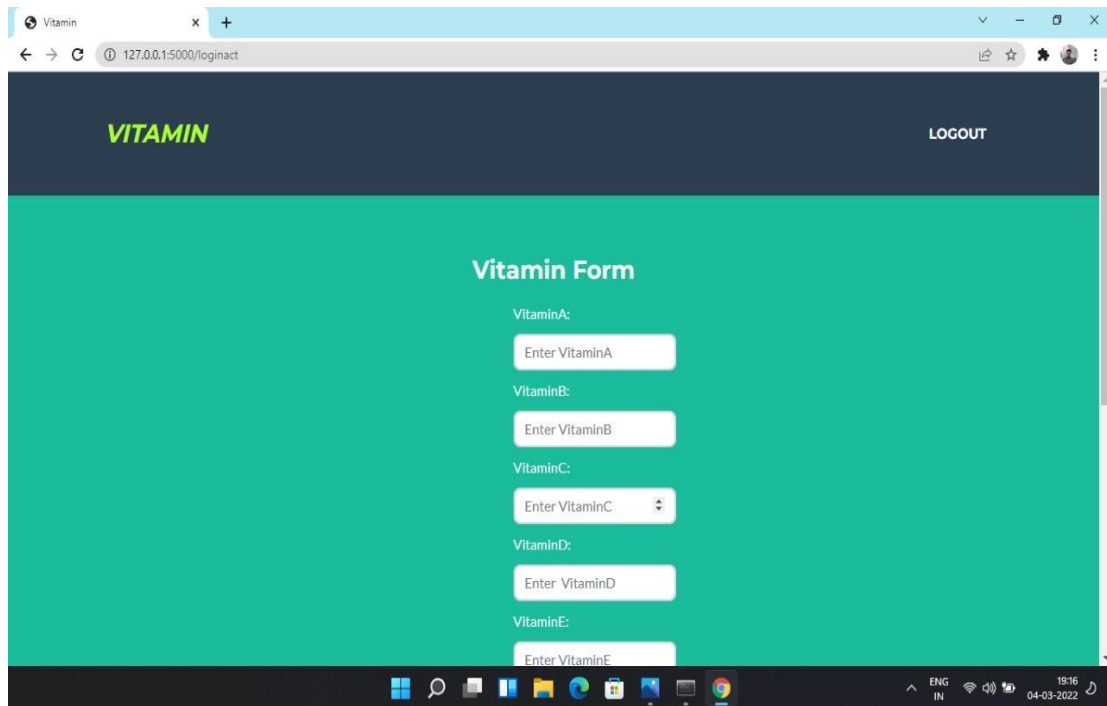
Figure 4.3 : Login page of the website

After login to the Vitamin Deficiency and Food Recommendation System , It will redirect to the page below.

This page contains

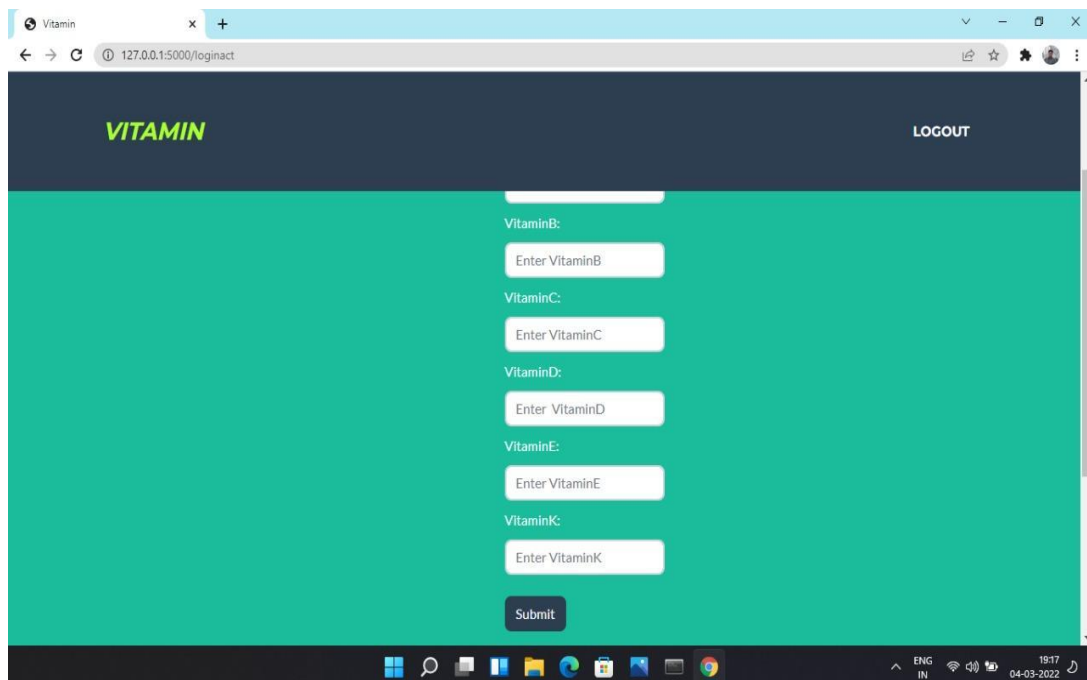
- Vitamin Form
- Logout

In the vitamin form, the user has to enter his/her vitamin values for vitaminA, vitaminB, vitaminC, vitaminD, vitaminE and vitaminK and click the submit button.



The screenshot shows a web browser window with the address bar displaying "127.0.0.1:5000/loginact". The page has a dark blue header with the word "VITAMIN" in green on the left and "LOGOUT" in white on the right. The main content area has a teal background and is titled "Vitamin Form" in white. Below the title, there are six input fields, each with a label and a placeholder text: "VitaminA: Enter VitaminA", "VitaminB: Enter VitaminB", "VitaminC: Enter VitaminC", "VitaminD: Enter VitaminD", "VitaminE: Enter VitaminE", and "VitaminF: Enter VitaminF". The browser's taskbar at the bottom shows various application icons and system status icons.

Figure 4.4 : Vitamin Form



This screenshot shows the same "Vitamin Form" as Figure 4.4, but with an additional input field for "VitaminK: Enter VitaminK" and a dark blue "Submit" button at the bottom. The layout and header remain consistent with the previous figure. The browser's taskbar at the bottom shows the same set of application and system icons.

Figure 4.5 : Submit option for vitamin form

After clicking the submit button, vitamin deficiency for the vitamins A, B, C, D, E and K is

predicted and the food to be taken to normalize the vitamin values is recommended.

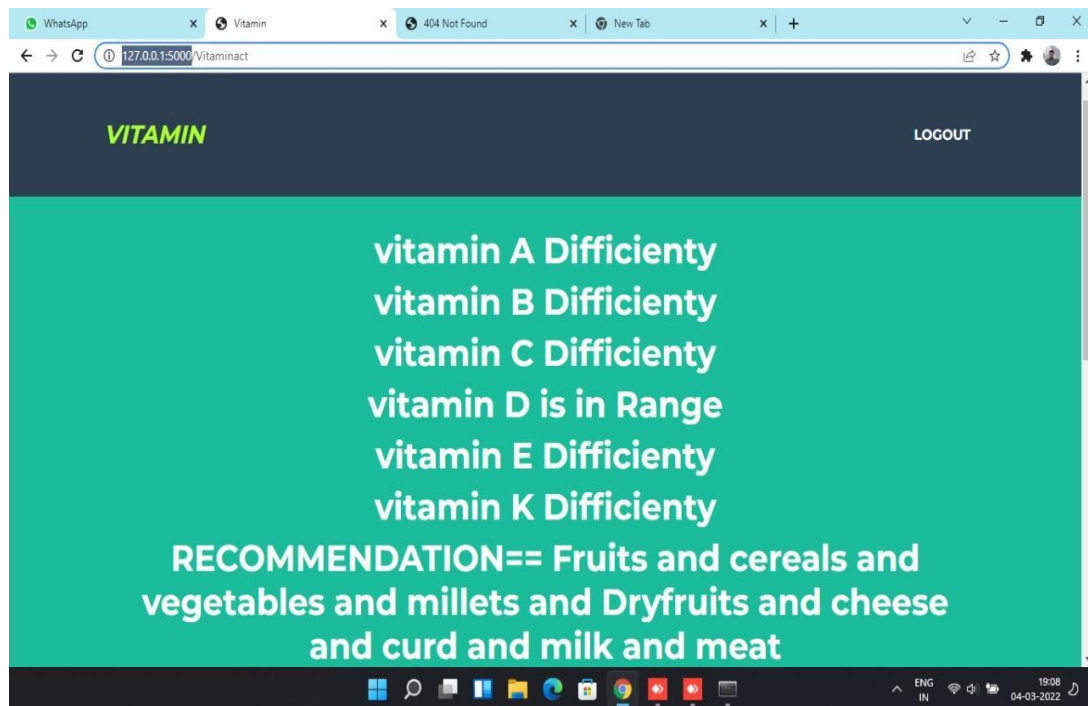


Figure 4.6 : Vitamin Deficiency Prediction

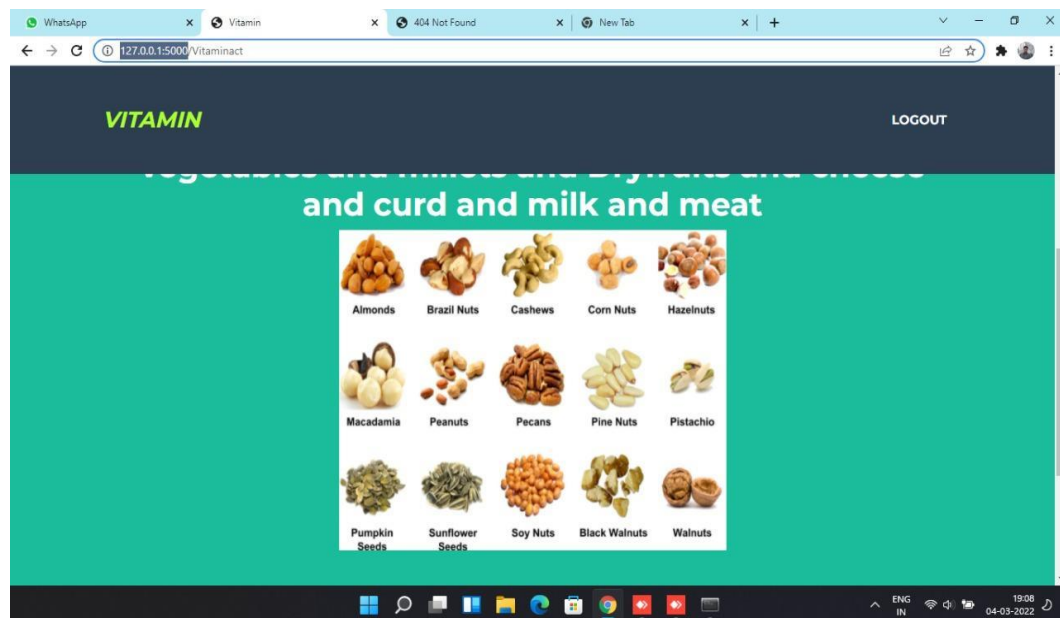


Figure 4.7 : Food Recommendation based on the vitamin deficiency

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

5.1 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 recommendations depending upon which are essential for the type of vitamin deficient. After the clustering is performed, using Random Forest classifier, 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 basis of vitamin deficiency.

5.2 FUTURE SCOPE :

Predicting vitamin deficiency and recommending food accurately at individual level is quite a challenging task to care for organizations for a sustainable healthcare system. This project developed a machine learning model that can incorporate the heterogeneous information within datasets for two times predictions in vitamin deficiency and recommending food. Utilizing the embedding learning processes, our model eliminates the reliance on domain knowledge to handcraft medical codes into semantic similar categories. These findings highlight the potential of our model to provide better guidance in medicinal services. This system has a lot of potential in the future with a few improvements. This mechanism may also be enhanced to predict deficiency and recommend food for other nutrients : carbohydrates, proteins, minerals etc. This System focuses more on the user-friendliness. It may be made available to anyone in the region and will remain available in a variety of languages.

BIBLIOGRAPHY / REFERENCES

- [1] V. Nallarasan, L. Anand, J. Prabakaran, “Improved Predictive Learning Approaches for Customized Diet Suggestion Framework in Medicinal Services”, Journal of Xi'an University of Architecture & Technology, Volume 12, Issue 7, 2020
- [2] Phanich, M., Pholkul, P., & Phimoltare, 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.
- [3] Ge, M., Elahi, M., Fernández-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.
- [4] Freyne, J., & Berkovsky, S., “Evaluating recommender systems for supportive technologies,” User Modeling and Adaptation for Daily Routines, pp. 195-217, Springer London, 2013.
- [5] 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
- [6] 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
- [7] Pratiksha Ashok.Naik Information Technology SDMCET Dharwad, India,” Intelligent Food Recommendation System Using Machine Learning”, International Journal of Innovative Science and Research Technology, Volume 5, Issue 8, August – 2020
- [8] Shubham Singh Kardam, Pinky Yadav, Raj Thakkar, Prof Anand Ingle,” Website on Diet Recommendation Using Machine Learning”, International Research Journal of Engineering and Technology (IRJET), Volume: 08 Issue: 04 | Apr 2021
- [9] Raciél Yera Toledo, Ahmad A. Alzahrani, Luis Martinez, “A Food Recommender

System Considering Nutritional Information and User Preferences”, IEEE Access, Volume 7, 2019

[10] Suyash Maheshwari, Manas Chourey, “Recipe Recommendation System using Machine Learning Models”, International Research Journal of Engineering and Technology(IRJET), Volume 6, Issue 9, Sep 2019

[11] L. Yang, C.-K. Hsieh, H. Yang, J. P. Pollak, N. Dell, S. Belongie, C. Cole, and D. Estrin, “Yum-Me: A personalized nutrient-based meal recommender system”, ACM Trans. Inf. Syst., vol. 36, no. 1, pp. 1-31, Jul.2017.

SESHADRI RAO GUDLAVALLERU ENGINEERING COLLEGE

(An Autonomous Institute with Permanent Affiliation to JNTUK, Kakinada)
Seshadri Rao Knowledge Village, Gudlavalleru

Department of Computer Science and Engineering

Program Outcomes (POs)

Engineering Graduates will be able to:

- 1. Engineering knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
- 2. Problem analysis:** Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
- 3. Design/development of solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
- 4. Conduct investigations of complex problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions., component, or software to meet the desired needs.
- 5. Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
- 6. The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
- 7. Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

- 8. Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
 - 9. Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
 - 10. Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
 - 11. Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
- Life-long learning:** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Program Specific Outcomes (PSOs)

PSO1 : Design, develop, test and maintain reliable software systems and intelligent systems.

PSO2 : Design and develop web sites, web apps and mobile apps.

PROJECT PROFORMA

Classification of Project	Application	Product	Research	Review
	√			

Note: Tick Appropriate category

Project Outcomes	
Course Outcome (CO1)	Identify and analyze the problem statement using prior technical knowledge in the domain of interest.
Course Outcome (CO2)	Design and develop engineering solutions to complex problems by employing systematic approach.
Course Outcome (CO3)	Examine ethical, environmental, legal and security issues during project implementation.
Course Outcome (CO4)	Prepare and present technical reports by utilizing different visualization tools and evaluation metrics.

Mapping Table

CS1537 : MAIN PROJECT															
Course Outcomes	Program Outcomes and Program Specific Outcomes														
	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12		PSO 1	PSO 2
CO1	3	3	1					2	2	2				1	1
CO2	3	3	3	3	3			2	2	2		1		3	3
CO3	2	2	3	2	2	3	3	3	2	2	2			3	
CO4	2		1		3				3	3	2	2		2	2

Note: Map each project outcomes with POs and PSOs with either 1 or 2 or 3 based on level of mapping as follows:

1-Slightly (Low) mapped 2-Moderately (Medium) mapped 3-Substantially (High) mapped

COPY RIGHT



ELSEVIER
SSRN

2022IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 13th Apr 2022. Link

[:http://www.ijiemr.org/downloads.php?vol=Volume-11&issue=ISSUE-04](http://www.ijiemr.org/downloads.php?vol=Volume-11&issue=ISSUE-04)

DOI: 10.48047/IJIEMR/V11/I04/21

Title **Improved Predictive Learning Approaches in Medicinal Services for Customized Diet Suggestion Framework**

Volume 11, Issue 04, Pages: 131-136

Paper Authors

Dr. S. Narayana, V. S. N. S. Gayathri , Shaik Shabana, R. Lakshmi Naveen Kumar, V. Rahul Nadh



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per **UGC Guidelines** We Are Providing A Electronic Bar Code

Improved Predictive Learning Approaches in Medicinal Services for Customized Diet Suggestion Framework

^[1]Dr. S. Narayana, ^[2]V. S. N. S. Gayathri, ^[3]Shaik Shabana, ^[4]R. Lakshmi Naveen Kumar, ^[5]V. Rahul Nadh

^[1] Professor & Mentor (AS&A), CSE Department of SR Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India-521356.

^[2] ^[3] ^[4] ^[5] Student, CSE Department of SR Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India-521356.

satyala1976@gmail.com, vsngayathri@gmail.com,
shabanashabbughf@gmail.com, reddynaveenkumar67@gmail.com,
vikkurthyrahulnadh@gmail.com

ABSTRACT

In this cutting-edge world many human beings are struggling from extraordinary kinds of ailments and illnesses. A learn about through WHO reviews that insufficient and imbalanced consumption of meals motives round 9% of coronary heart assault deaths, about 11% of ischemic coronary heart ailment deaths, and 14% of gastrointestinal most cancers deaths worldwide. Moreover, round 0.25 billion teens are struggling from extraordinary sorts of nutrient deficiency specifically from Vitamin-A to Vitamin-K deficiency, 0.2 billion human beings are struggling from Iron deficiency (Anemia), and 0.7 billion human beings are struggling from Iodine deficiency. The primary goal of this undertaking is to propose a food regimen to one-of-a-kind folks the use of the datasets that are organized based totally on the aggregate of a number nutritional vitamins and their deficiency and meals to be endorsed based totally on which diet is deficient. In this mission a couple of classifier algorithms are used (KNN, Decision tree, Random forest, Logistic regression, Voting classifier). Ensembled algorithm is used to mix more than one algorithms and instruct a new algorithm. Accuracy of every algorithm is calculated and the high-quality algorithm is used for prediction purposes. Prediction is proven the use of flask net utility which will observe deficiency of diet and endorse kind of meals to be taken on a range of combos

Keywords : *Vitamin, Deficiency, Diet, Decision tree, Random forest*

1. INTRODUCTION

In this venture its personal records set is organized based totally on a range of excessive and low values of nutritional vitamins from (vitamin a,b,c,d,e,k) and points are divided from ordinary and ordinary stipulations of nutritional vitamins and labels are divided into 0 and 1 as regular and abnormal. Another

dataset is organized based totally on an aggregate of quite a number nutritional vitamins and their deficiency and meals to be endorsed based totally on which nutrition is deficient. In this venture a couple of classifier algorithms are used (knn, selection tree, random forest, logistic regression, vote casting classifier) ensembled algorithm is used to

mix a couple of algorithms and teach a new algorithm. Accuracy of every algorithm is calculated and the nice algorithm is used for prediction purposes. Prediction is proven the use of flask internet software which will observe deficiency of nutritional vitamins and endorse sorts of meals to be taken on a number combinations.

The suggestion method has essentially three ranges that are Information Collection Phase, Learning Phase and Recommendation Phase. The data is first off amassed about a precise trouble and the more than a few options associated to that hassle are categorized. After the series of records the Learning Phase comes in which more than a few conclusions are made out of that records which is gathered and in the ultimate segment i.e., Recommendation Phase an output is given in which a number of hints are made.

Balanced diet is an essential factor of a wholesome way of life for people. Along with a balanced diet, everyday bodily workout is fundamental for a wholesome life. Nowadays vitamin and fitness are regularly overlooked. The majority of humans go through from diabetes, coronary heart disease, cancer, stroke etc. The ailments are nearly at once associated to unhealthy ingesting habits. So, our physique wishes vitamins to continue to be healthy, and meals elements necessary vitamins that quit us from getting sick. A healthy, balanced weight-reduction plan will normally encompass vitamins, minerals, protein, wholesome fats,

proteins, carbohydrates, and fiber. A healthful meals pyramid is a mixture of plant foods, reasonable quantities of animal products. Which consists of vegetables, grains, fruits, oils and sweets, dairy, meat and beans. Generally, a individual stays unaware of fundamental reasons at the back of deficiency or extra of quite a number integral substances, such as calcium, proteins, and vitamins, and how to normalize such materials via a balanced diet. With the gain of technology, the humans can stay a healthier lifestyle. The fast-food consumption charge is alarmingly excessive and this hence has led to the consumption of unhealthy food. This leads to more than a few nutrition deficiency problems such as E, A, B, C, D, K make bigger in blood strain etc. Hence it has emerged as very quintessential for human beings to have an exact balanced dietary wholesome diet. But in this quickly paced era no longer every person has the time and money. By getting important points of consumer existing vitamin values this utility can automate the technique of detection of diet deficiency and meals suggestion device the usage of laptop learning. Aim is to create our personal diet deficiency records set for A,B,C,D,E,K and create meals advice dataset and teach each dataset with 5 computer gaining knowledge of algorithms and locate accuracy of every algorithm and use fine accuracy mannequin in a internet utility the place nutrition deficiency and meals suggestion is estimated based totally on given enter values.

2. LITERATURE SURVEY

Raciel yera toledo proposed a meals recommender machine thinking about dietary data and person preferences. The meal sketch for the consumer encouraged the use of users' preferences. This device manages each consumer preferences and dietary information. Vijay Jaiswal proposing a wholesome meals habits, consuming patterns and energy burned rely can be consumption of vitamins and so on the use of the records mining tools. In this device the hidden patterns and patron meals taking habits are determined from specific information sources. In this device choice tree getting to know algorithm, Random Tree algorithms are used on distinct datasets. H. Jiang proposed a machine to calculate the everyday calorie demand. The Knapsack algorithm is used for endorsed weight-reduction plan combos of users. Different from different diabetic weight loss plan suggestion systems, this machine can rank the encouraged weight-reduction plan mixtures the use of TOPSIS algorithm in accordance to user's meals nutrition. Jung-Hyun Lee proposed a personalized eating regimen advice carrier managing coronary heart diseases. This carrier presents clients personalized customary information, household records of diseases, seasonal meals intakes. Rung-Ching Chen developed a recipe ontology that defines some frequent ailments restoration with a range of meals hints and an inference engine for patron fitness situation and a recipe ontology can be used for appropriate recipe suggestions on

food priorities. FidelsonTanzil makes use of ABC algorithm to extract facts from database in accordance to user's requirements. K MEAN and SOM algorithms are used on datasets. Mohd Afisi projected ABC algorithm in Data Mining and examined in contrast to six normal classification algorithms efficaciously and ABC proved as an appropriate algorithm for recommendation. Xiaoyan Gao proposed the meals advice trouble on person preference recipe advice factors. By the use of a neural network-based answer on Ordered weight loss plan Recommendation.

The authors INGMAR WEBER and PALAKORN ACHANANUPARP [1] made an strive to achieve insights from laptop leaned - food regimen success prediction which would assist human beings attempting to continue to be match and wholesome via maintaining a song on their dietary intake. The authors used public meals diaries of greater than 4,000 long-term lively MyFitnessPal customers to find out about the traits of an unsuccessful diet. Concretely, authors skilled a computer studying mannequin to predict persistently being over or beneath self-set day by day calorie desires and then seem to be at which aspects make a contribution to the model's prediction, the place lookup used to be founded round "quantified self " data. The authors discovered that classification overall performance was once ample and the token-based mannequin carried out higher than the category-based mannequin and used such

records feasibly for greater in-depth information mining.

NANDISH SHAH and ISHANI SHAH [2] introduced a suggestion of wholesome meals habits and ingesting machine based totally on net information mining, to find out hidden patterns and enterprise techniques from their patron and net data, which would song ingesting habits and endorse the kinds of meals that will enhance the fitness and keep away from the sorts of meals that elevate the hazard of illness. The authors used facts mining algorithms like classification, clustering, affiliation rules, etc. in the records mining procedure to extract beneficial facts about people's consuming habit. The nutritive shape of every variety of meals used to be analysed and the fat, energy, nutrition proportion in the recipe used to be calculated. Then they used the classification mining algorithm to system the composition facts and supply out the end result whether or not the food plan is healthful or not. As a result, customized tips had been counselled for every person.

3.PROPOSED SYSTEM

The System works in a Machine Learning Environment, we use a couple of computers getting to know algorithms to take a look at accuracy of diet deficiency and meals suggestion and the satisfactory mannequin is used for prediction in flask internet application. When a consumer enters diet values, the algorithm will predict deficiency in nutritional vitamins and advise meals

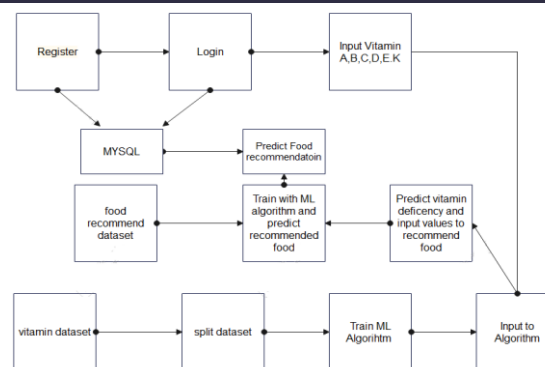


Figure-1: System Architecture

3.1 MODELS

3.1.1 DATASET:

In this project we are using vitamin dataset and food recommendation dataset which is prepared based on min and max vitamin values from the test results and features are min and max values of vitamin a, b, c, d, e, k values and labels are deficiency and non-deficiency.

3.1.2 PRE PROCESSING:

Features are extracted from the data set and stored in a variable as an xtrain variable and labels are stored in y train variable. Data is pre processed by standard scalar function and new features and labels are generated.

3.1.3 METHODOLOGY:

As seen from the above figure, we can see how the data is divided into different sets and then trained for different models. • The dataset was first divided into a training set (80%) and pre-training set(20%). • The pre-training set was divided into pre-train(80%) and pre-test(20%) • Now, the training set is further divided into train(80%) and validation set(20%). This train set is again divided into

train(80%) and test set(20%). So, now I have train validation and test sets separate which are nonoverlapping. • The pretrain set was used to find the best models for the given dataset. We took the best 4 models using the pretest set. Their performance was compared based on their mean absolute errors. • Once the best 4 models were obtained, hyperparameters for these models were tuned and the best parameter was selected.

3.2 MODULES IMPLEMENTATION

Data collection:

- In this project we are using vitamin dataset and food recommendation dataset which is prepared based on min and max vitamin values from the test results and features are min and max values of vitamin a, b, c, d, e, k values and labels are deficiency and non-deficiency.
- Based on the vitamin deficiency food data set is prepared with various combinations. In this feature are vitamin deficient values and labels are types of food.

Data preprocessing:

- Features are extracted from the data set and stored in a variable as xtrain variable and labels are stored in y train variable. Data is preprocessed by standard scalar function and new features and labels are generated.

Testing training:

- In this stage data is sent to the testing and training function and divided into four parts x test train, and y test train. Train

variables are used for passing to algorithms whereas test variables are used for calculating accuracy of the algorithm.

Initializing Multiple Algorithms and training with Logistic regression:

- In this stage machine learning algorithms are initialized and train values are given to the algorithm by this information algorithm will know what are features and what are labels. Then data is modelled and stored as a pickle file in the system which can be used for prediction.
- Data set is trained with multiple algorithms and accuracy of each model is calculated and best model is used for prediction

Predict data:

- In this stage new data is taken as input and trained models are loaded using pickle and then values are pre processed and passed to predict function to find out the result which is shown on web application.

4.RESULTS AND DISCUSSION

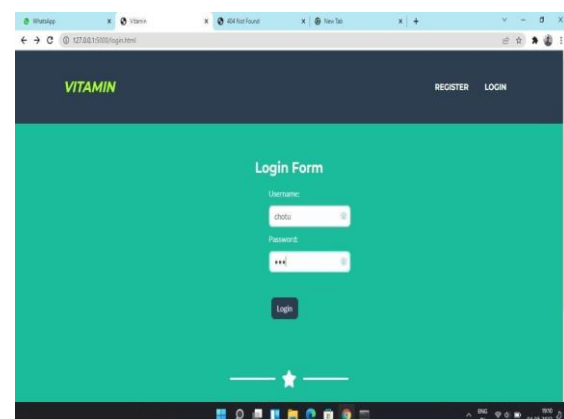


Figure-2: In the above screen shot user can login by using user name and password

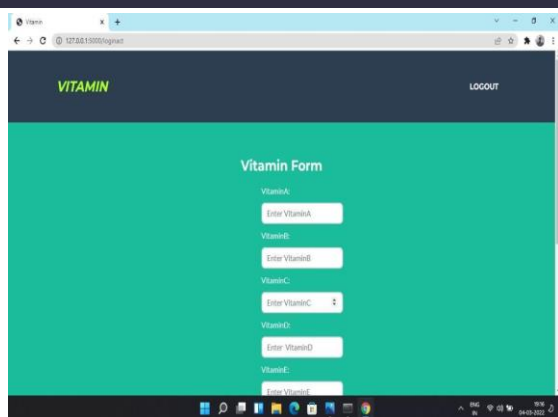


Figure-3:in the above screen user entering vitamin values

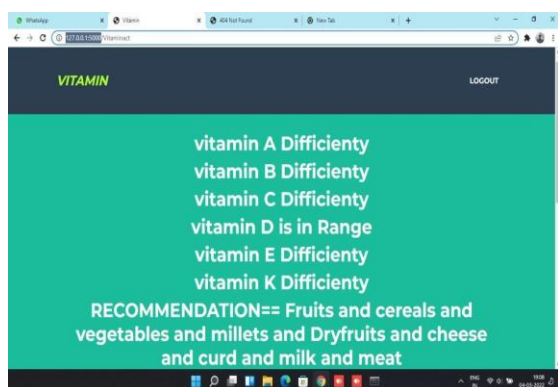


Figure-4: Result Page

5. CONCLUSION

We have created an internet site which recommends the meals gadgets and predicts diet deficiency in which we have carried out prediction by using taking enter as nutritional vitamins and their deficiency. For education of the system, the preliminary method includes the dataset education of meals gadgets relying upon the nutrition deficiency. The prediction of more than a few meals pointers relying

upon which are fundamental for the kind of diet deficient. After the clustering is performed, the usage of Random Forest classifier, the nearest meals gadgets are estimated which fantastic perfect for the excellent diet. Our food plan advice gadget approves customers to essentially get the favored wholesome weight-reduction plan on the foundation of diet deficiency.

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., Fernández-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. Prajka 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