VIT UNIVERSITY
(Estd. u/s 3 of UGC Act 1956)
VELLORE ■ CHENNAI
www.vit.ac.in

School of Computer Science and Engineering

DECLARATION

We hereby declare that the project entitled **"Personalized Health Plan using Nutrient Comparison"** submitted by us to the School of Computer Science and Engineering, VIT University, Vellore-14 in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Computer Science and Engineering** is a record of bonafide work carried out by us under the supervision of **Meenakshi S P, Assistant Professor (Senior).** We further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma of this institute or of any other institute or university.

Signature
Aakash Suman (13BCE0439)

Signature
Anupam Sinha (13BCE0425)

VIT UNIVERSITY
(Estd. u/s 3 of UGC Act 1956)
VELLORE ■ CHENNAI
www.vit.ac.in

School of Computer Science and Engineering

CERTIFICATE

The project report entitled "**Personalized Healthcare using Nutrient Comparison**" is prepared and submitted by **Aakash Suman (13BCE0439) and Anupam Sinha (13BCE0425).** It has been found satisfactory in terms of scope, quality and presentation as partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Computer Science and Engineering** in VIT University, India.

Guide
(Name & Signature)

Internal Examiner
(Name & Signature)

External Examiner
(Name & Signature)

ACKNOWLEDGEMENT

In course of our final research project, we would like to thank our guide **Prof. Meenakshi SP, Assistant Professor (Senior), VIT University, Vellore** who gave us her constant support and inspiration in successful completion of our project. Her necessary guidance in sorting out our doubt related to the various phases of the project were of extreme importance. We express our heart full gratitude to **Prof. Senthil Kumar R, HOD, School of Computing Science and Engineering, VIT University, Vellore** for his valuable advice in completing this project successfully. We are greatly indebted to our head of department of SCOPE, **Prof. Arun Kumar T, Dean, School of Computer Science and Engineering, Vellore Institute of Technology** for his support. The successful implementation of project depends on the perseverance, hard work and strong determination with combined effort of proper knowledge. We would also like to express our respect and regards to our Chancellor **Dr. G. Vishwanthan** who provided us necessary facilities which is required to accomplish our project. We would like to extend our support to our friends and family for their necessary help that benefited us indirectly or directly with the completion of the project. Last but not the least, we would like to thank **The Almighty** who kept his showers of blessing upon us to keep us fit and healthy throughout.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Expansion
PCA	Principal Component Analysis
SOM	Self Organized map
S-coefficient	Silhouette Coefficient
DBSCAN	Density-based spatial clustering of applications with noise
BIRCH	Balanced iterative reducing and clustering using hierarchies
USDA	United States Department of Agriculture
SSE	Sum of squared error
DB Index	Davies-Bouldin Index

ABSTRACT

Eating nutritious food does not only benefit a person healthily but also mentally. It helps to keep our body immune to certain kind of diseases by ensuring a healthy diet. There are several kinds of nutrition deficiency problems that can occur if one does not pay attention to type of food one is in taking. One of the most common aspect of nutrition deficiency that occurs in most of the people is diabetes. It is a type of disease that occur due to excessive intake of sugar. Another type of disease that is caused due to deficiency of iron and vitamin B12 is anemia. Several neural disease can occur if one is not taking adequate amount of protein in their diet. So, in our work we have focused on exploiting data science in the health field particularly in food nutrition sector. While carrying our research we found out that not many work has been carried out in this sector and it needs to be done for the betterment of society. Here we have tried to come out with a healthy nutrition diet by comparing the nutritional values of foods and provide some efficient food items for a user. This comparison is being done by various algorithms in data analytics sector and the best is selected. For the base of the food products, we have used a standard nutrition chart provided by the renowned United States Department of Agriculture also known as USDA.

1. INTRODUCTION

1.1 THEORETICAL BACKGROUND

In past few years, data science has revolutionized the informational sector and the ways statistics are used to arrive at optimum conclusion. After all, data management in any field of discipline is worth a key to control the flow and direction of its concerns. In the food sector, for example, there are set of essential nutrients which one has categorized according to convenience of their study. In biology, food is decomposed into various types of biomolecules; enzymes, carbohydrates, lipids to name few. Their classification and systematic study generates a lot of data to be kept and analyzed. For an average person, it is not easy to simply go through and decide what suits him better. Especially, if one is suffering from some illness. Now, worthless to mention, there are thousands of disease in which data science can be used to model one's proper requirement. In our work, we've mostly focused upon Diabetes, as having 371 million people being affected and counting, it is worth a concern. Our project aims use of data science and its methods such as K-means clustering, Self- organizing Maps to classify food intake as per nutrition values recommended by USDA.

1.2 MOTIVATION

We are so thankful to the mother earth for providing us with all the required things to sustain life on this planet. And in this modern era, it is just next to impossible to imagine our life without technology. We are so surrounded by technology say whether it is the smartphones we carry everywhere, the train ticket booking center, the vending machine, social networking and many more. With this advent of technology, huge amount of data is being generated every day. Earlier, some traditional data management techniques were used to handle the data generated likes of database management system or the file management system. But now, using those traditional methods for data analysis is a tedious task. It has become very slow and inaccurate as the data generation has reached beyond threshold. So, in order to cope with this problem, many new tools and algorithms have come into existence to analyze. As data science is necessary in almost all the fields, however it has much been exploited in the case of healthcare. We were

very keen to use some of the data science algorithms in healthcare and come up with some suitable results beneficial to the mankind.

1.3 AIM OF THE PROPOSED WORK

Our work focuses to recommend a diet plan to the user comparing the nutrient constituents in the food items and have an alternate item with similar taste and healthier option. Particularly, we have concentrated more on people suffering from diabetes because it's the case where the patients need to take optimum quantity of sugar and burn optimum amount of calories. It cannot be too low and it cannot be too high. We have used different data mining techniques and find the clusters of various food items and compare those techniques to find the best suitable match for the food products.

1.4 OBJECTIVES

The objectives of our work are given below:

- Recommend the best food items amongst the given list of food data.
- Categorizing the food products according to the nutritional food value.
- Exploiting data science in the health sector.
- Finding out the best data mining and clustering techniques in comparing the nutritional value of food products.

2. LITERATURE SURVEY

In our work we have tried to devise a food recommendation system which would use a food nutrition dataset to find out the best food clusters. We have applied various data mining techniques to group the food dataset in various cluster and then comparing the algorithms and finding out best from them.

2.1. SURVEY OF THE EXISTING WORK

Diabetes is a type of disease which is due to high level of sugar in blood. In our body ,pancreas secretes Insulin which helps in controlling the blood sugar level but due to some metabolic reasons if pancreases are not been able to secrete the insulin then the blood sugar level of our body gets out of control and in turn, it leads to Diabetes.

Diabetes can be genetic or due to improper dietary habits. We can control later by taking proper diet amongst the different food items available by comparing their nutritional values. Here, we have chosen diabetes because it is a disease which is highly dependent on food intake.

In 2010, a food recommendation system on the basis of nutrient comparison is made which used SOM (self-organizing map) to map n-dimension data to reduced dataset. The datasets are very high in dimension it becomes cumbersome to apply data analytical algorithm and the results which comes are also infeasible. Therefore, it is necessary to apply some algorithms which reduces the dimension of the data without changing the information of the data. It maps n-dimensional data to m dimensional data where n<m where m is 2 in most of the condition. It uses vector quantization in which low dimensional data are computed for already applied prototypes. Also in SOM there is a parameter setting for the neighborhood width i.e. lambda. In this system later K-Mean algorithm is implemented to cluster the food items according to eight nutrition parameters which is main factors effecting diabetes. Here similar food items are grouped in two categories which are categorization according to food characteristic and nutrition affecting diabetes .The food dataset affecting diabetes was grouped in 6 categories which are Normal Food (NF), Limited Food (LF), Avoidable Food (AF), Enriching Food(EF), Preferable Food(PF) and Average Food(AF).

As the name suggest, every food category have the corresponding features. For example, Normal Food contains food items which are low in starch, sugar and carbohydrates and there intake don't effect the blood sugar level too much. Meat, fish, vegetables, sea foods and sugar food substances comes under normal food category. Limited food contains the food items which contains carbohydrates in large amount. Foods like rice, pasta, noodles and starchy food comes under this category. Avoidable food contains the food items which contains carbohydrates in large amount and it also primarily affect the blood sugar level to a greater extent. Sweets, Sweet Fruits, drinks etc. which contains sugar comes under this category. This aims to encourage the widest variety of food items which is allowed to be eaten. The food dataset are categorized in two different ways: The nutrients are rated according to particular scores after taking consult from nutritionist. The users are recommended best food items from each group using distance comparison between various items in the cluster.

In 2015, an analysis of PCA (Principal Component Analysis) was conducted on Nutritional values of fruit and vegetable and cluster analysis was done. The richness of nutrition of particular food items are calculated and it is later replaced by the substituted food item required. By using dimension reduction method, the multiple parameters are recombined to form reduced into new dimensional parameters without letting any information getting lost in the process which enhances the process of analysis. Here, a sample of vegetables and fruits were analyzed and the principal component is calculated which later on help in finding a principal vector for each item. In dataset if there are multidimensional attributes then it becomes very difficult to interpret and map them. Also, it takes lot of time to apply algorithm on the dataset. Thus, to reduce the time and noise we use PCA. It uses linear technique for the dimension reduction to reduce the data dimension using a parameter principal component of the data. In PCA, we subtract the mean value from each of the data dimension and then the covariance matrix is calculated. If the reduced data have n dimension then the Covariance matrix have dimension n*n. The Eigen Vector along with values of covariance matrix is calculated which enable us to extract lines that characterize the data. Principal Component value of the data is found out. Later, the new dataset is derived later using vector components. In our work we have used PCA to reduce the dataset to more readable of data form.

In 2011, a study was made on the BIRCH clustering method to optimize its parameters. It is an unsupervised algorithm which is used for clustering datasets .It performs hierarchical clustering over the large dataset and it can efficiently cluster dataset in single scan. The datasets are clustered in incremented manner also it dynamically processes multi-dimensional data points. In this algorithm decision is made without scanning all the data points and only important points are scanned. It follows incremental model which doesn't require the complete dataset. Birch cluster has advantage over other algorithm when the dataset is very large and it don't take the case where data-set is not able to fit in main memory. It also don't consider the heuristic weighing based on the distance between the data points. The Birch algorithm has been studied and implemented on our dataset to see the results in the subsequent section.

The various parameters involved in birch Clustering algorithm are branching factor, selection of separator line and quality threshold. Most of the normal clustering approaches don't consider these aspects and the cluster generated by them are not of good quality. But, Birch algorithm can be applied on the large dataset also and results of which is also efficient. We found out in our survey that main feature changes when parameters like maximal distance between various points, number of cluster (k),branching factor, normal distribution of points ,etc. changes. The branching factor and threshold value has direct impact on the results. Therefore, proper implementation of BIRCH algorithm can give us efficient clustering results on very large dataset.

In 2006, a study was made on the implementation of DBSCAN algorithm on large datasets. From the observations it came that there was a drawback of DBSCAN i.e. it takes a large amount of time when it is implemented on large dataset. So, this problem was solved can be solved by applying the algorithm on only, a few selected part of datasets i.e. prototypes but this solution also had a demerit that the solution can deviate from the original result. Therefore, to solve these problem we devise a Nobel method in which we tried to coarse the level to reduce the time requirement and the result also don't get deviated from the original result. DBSCAN stands for Density-based spatial clustering of applications with noise. The points in dataset are group together that are close together and reachable from each other and the points which are outlier point are low-density points. In this algorithm points are classified as three categories that are core points, density-reachable points and outliers. A point is called core point if there are at least min points at a distance ε (maximum radial distance from point) and the point is called directly reachable. A point is called reachable point if there is

intermediate points from starting point to that point i.e. intermediate points should be core points. Outlier points are the points which cannot be reached from any other points. Unlike algorithms like k-means, specification of number of clusters (k) is not required. It uses the concept of noise and it is very robust. The parameters min points and ε is set by domain experts. We tried to implement and find out the observation on our dataset. This method is implemented and its results has been discussed in subsequent section. The prototypes which are derived are of two types using leaders clustering method. This method takes less time.

K-Mean is also one of the most versatile algorithm which help us in clustering the datasets according to similarity in characteristics. K-means divides the clusters where distance between the data items within the same group should be least and distance between the data items within the different group should be maximum i.e. intra cluster distance should be minimum and inter cluster should be maximum. It is also flexible type of algorithm in which we can choose number of clusters (k) according to dataset for which the result is more efficient. The value of k can be calculated by various parameters available like Davies-Bouldin index or plotting a graph of clusters and the inter-cluster distance. The best value is chosen as k. Euclidean distance is the metric used to find the clusters and points. K mean is an iterative type of algorithm in centroid of the clusters are recalculated till we reach a point where there is insignificant change in values.

Silhouette Coefficient is a performance estimating parameter of clusters. It is a method which help interpretation of clusters and later validates the result for their consistency. It shows how efficiently the cluster lies under its cluster. It measures the extent to which each object lies within its cluster as compared to other cluster. The range of S-coefficient lies between -1 to 1.The value closer to 1 shows that the object matches to its cluster and value near to -1 means the object don't matches to its cluster. So, we get the idea whether the datasets is properly clustered or not. The S-coefficient is calculated using any of the mentioned distance metric like Manhattan or Euclidean distance. Therefore, it is very useful comparing various algorithm and find out the best algorithms amongst them.

2.2. GAPS IDENTIFIED IN THE SURVEY

In previous works they have used SOM (Self-Organizing Maps) for mapping high-dimensional space to map it to low-dimensional space. But, when the dataset is very

large it takes too much time to map it. So, it is too much time consuming and not efficient. There are many more algorithms present for dimension reduction .Some comparing the results of them will help us in finding the more optimal solution. Here, we have applied PCA (Principal Component Analysis) to reduce the dimension of the dataset in which we convert n-dimensional parameter to reduced dimension by finding principal components orthogonal to each other. It speed up our clustering process. In previous works only a particular algorithm was applied food dataset but, here in our work we have applied various clustering algorithm and made a comparison amongst them using a particular parameter which help us in knowing which algorithm is good for the dataset. Also, in previous work in diabetes the food items were classified as normal food, limited food and avoidable food which are not very specific we can further classify food in various food cluster as enriched food, unaccountable food, etc. The earlier work was mostly focused on diabetes but there are many more disease which are dependent on dietary habits like anemia, etc. So, we could apply our study on those disease also. In previous work the dataset used was not so large like around 400 elements which would not cover whole of the food category .Thus, we can use a dataset which contain around dataset>5000 elements which would cover variety of dataset which would enable user to choose food item according to their food choices.

3. OVERVIEW OF THE PROPOSED SYSTEM

3.1 INTRODUCTION AND RELATED CONCEPTS

Personalized health plan using nutrition comparison uses various data mining techniques likes clustering algorithms and some of the dimension reduction techniques. The implementation is done using python language and some of the pre-defined libraries for data mining techniques written in python were used. Some of the algorithms that were used are given below:

Clustering Algorithms:

- K-Mean Clustering: k-means is an algorithm which help us in clustering the datasets according to similarity in characteristics. K-means divides the clusters where distance between the data items within the same group should be least and distance between the data items within the different group should be maximum i.e. intra cluster distance should be minimum and inter cluster should be maximum. It is also flexible type of algorithm in which we can choose number of clusters (k) according to dataset for which the result is more efficient. The value of k can be calculated by various parameters available like Davies-Bouldin index or plotting a graph of clusters and the inter-cluster distance

- BIRCH Clustering: It performs hierarchical clustering over the large dataset and it can efficiently cluster dataset in single scan. The datasets are clustered in incremented manner also it dynamically processes multi-dimensional data points. In this algorithm decision is made without scanning all the data points and only important points are scanned. It follows incremental model which doesn't require the complete dataset. Birch cluster has advantage over other algorithm when the dataset is very large and it don't take the case where data-set is not able to fit in main memory. It generally scans data in single go.

Dimension Reduction techniques:

- Self -Organizing Maps:

  It maps n-dimensional data to m dimensional data where n<m where m is 2 in most of the condition. It uses vector quantization in which low dimensional data

are computed for already applied prototypes. Also in SOM there is a parameter setting for the neighborhood width i.e. lambda

- PCA algorithm:

  It uses linear technique for the dimension reduction to reduce the data dimension using a parameter principal component of the data. In PCA, we subtract the mean value from each of the data dimension and then the covariance matrix is calculated. If the reduced data have n dimension then the Covariance matrix have dimension n*n. The Eigen Vector along with values of covariance matrix is calculated which enable us to extract lines that characterize the data. Principal Component value of the data is found out.

3.2 ARCHITECTURE

The architecture diagram of the system is given below:

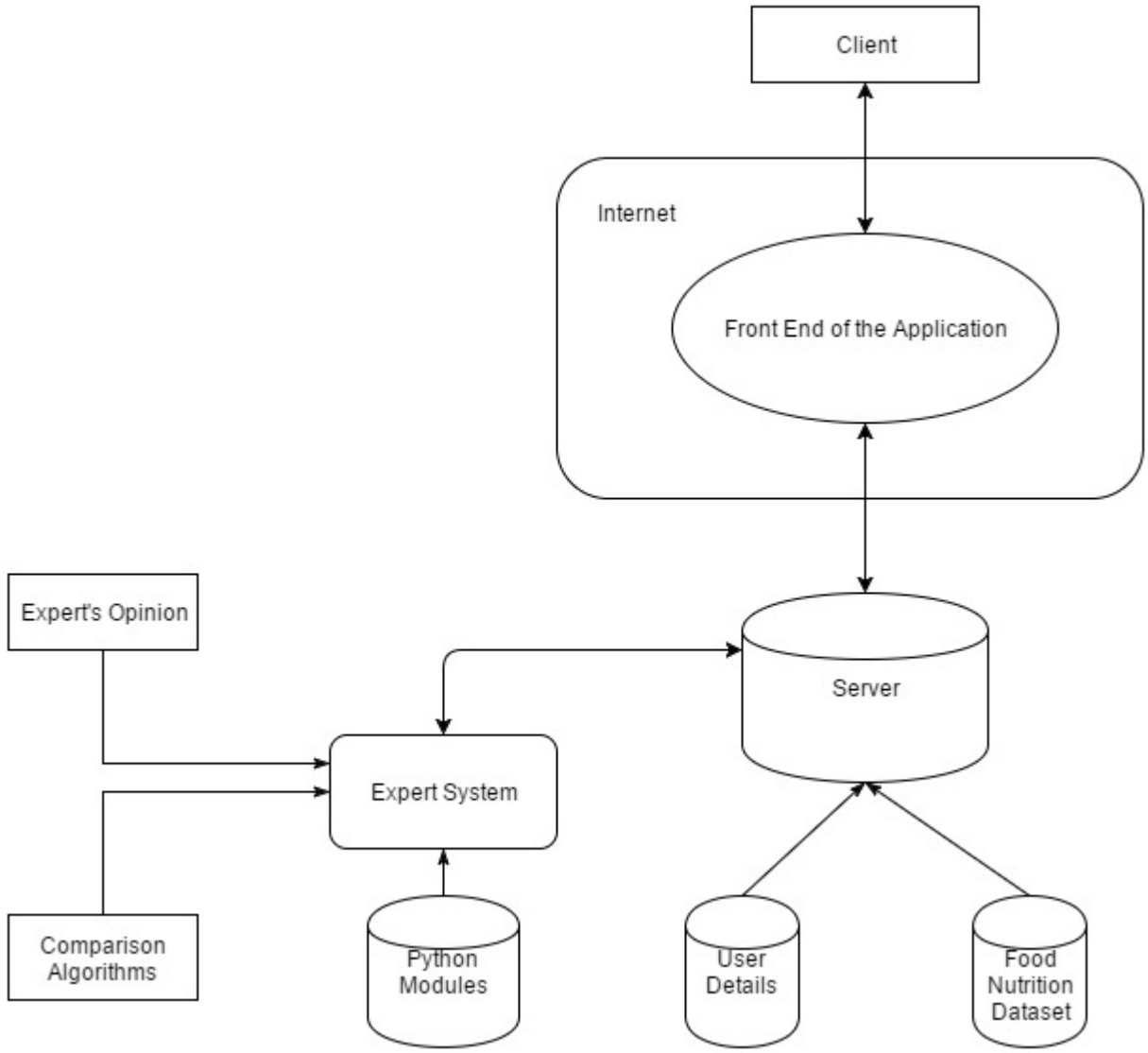


Fig 3.1

3.3 PROPOSED SYSTEM MODEL

The proposed system model can be explained using some UML diagrams. Some of them are:

1) Use Case Diagram:

The use case diagram of the system is given below:

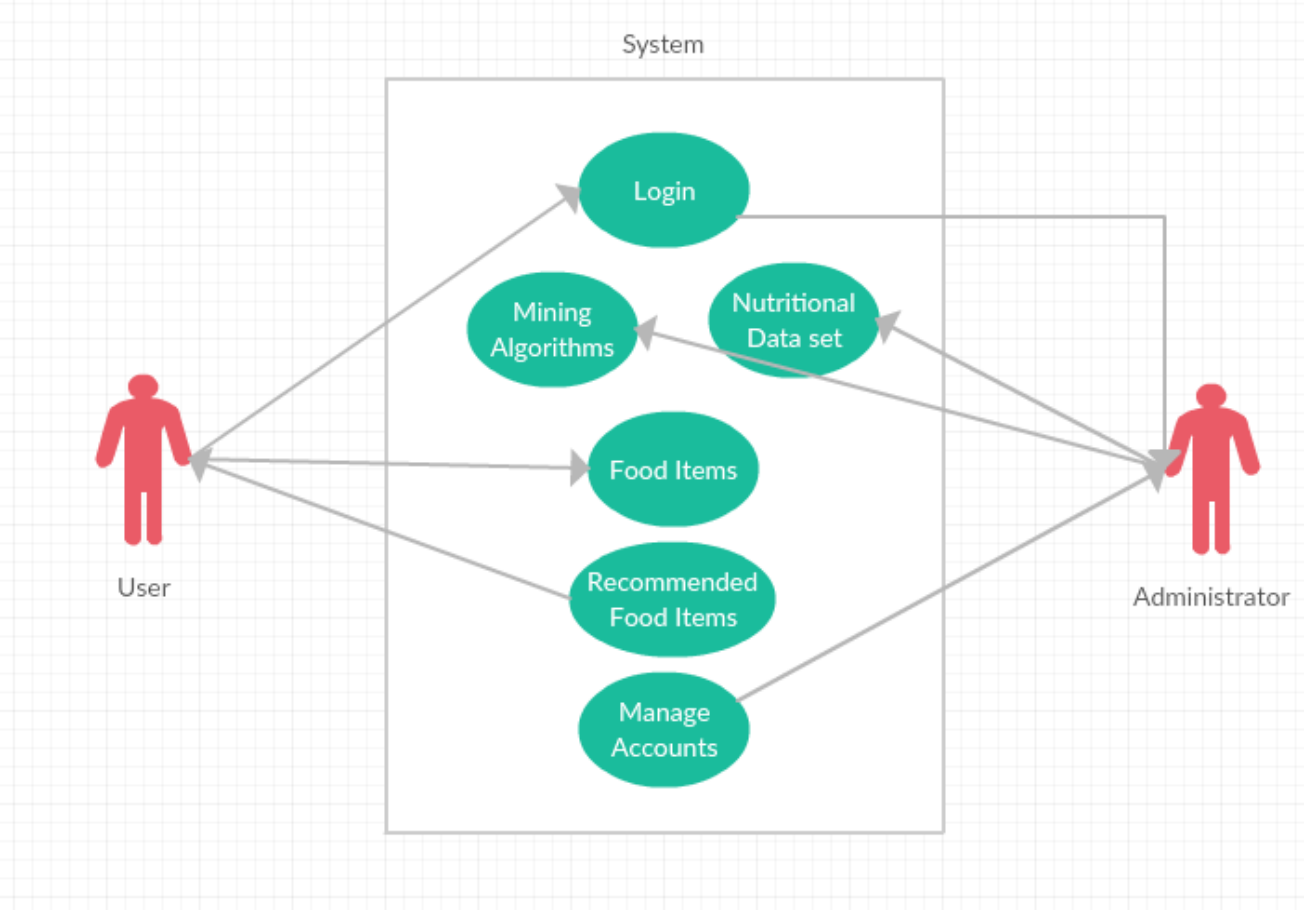


Fig 3.2

2) Activity Diagram:

The activity diagram for the proposed system is given below:

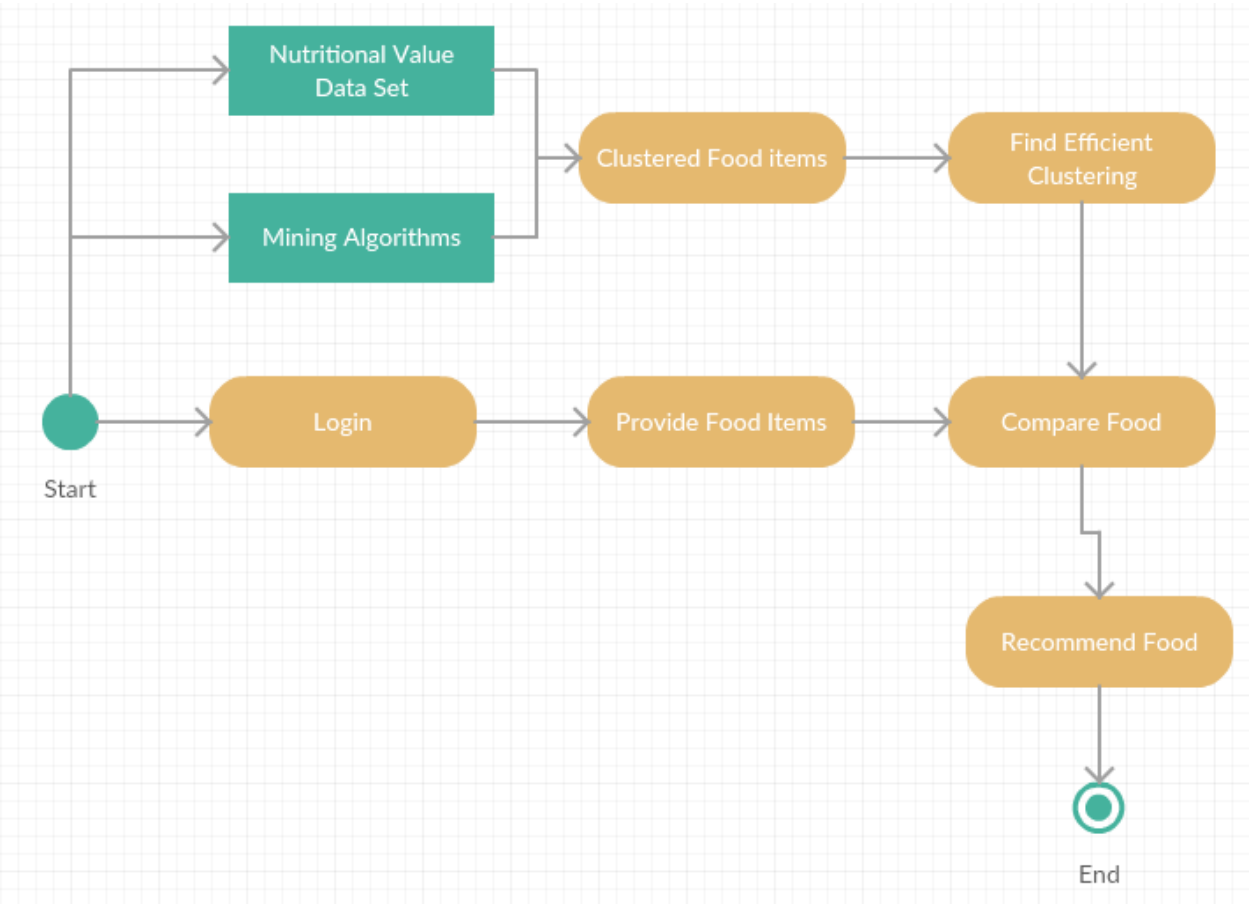


Fig 3.3

4. PROPOSED SYSTEM ANALYSIS AND DESIGN

4.1 INTRODUCTION

As the pictures shown in the previous diagrams suggest, the client will interact with the system using the front end of the application. The system will be hosted online to allow to its users from anywhere. Some of the python modules will be used for carrying the data mining techniques likes of scipy libraries and numpy libraries. These techniques will help in clustering the data and reducing the dimension of the dataset which can later be used as comparison for the food item provided by the client. Now, as the client enters the food items for comparison, the system will search for the common clusters and compare the nutritional values and provide the result for what food should be taken and how much it should be taken.

4.2 REQUIREMENT ANALYSIS

4.2.1 Functional Requirements

4.2.1.1 Product Perspective:

The Personalized health care system focuses on comparing the nutritional values of the food items and suggest the best of the food items. This system mostly concentrates on people suffering from diabetes so as to provide the optimal amount of nutrition content in the food the user needs to take. For this system, we will be needing a standard dataset of nutritional values of food provided by some of the trusted institutes. These data should be present in .csv format so that the system can efficiently recognize. Further, many of the third party application likes of rapid miner and MS Excel should be used for getting the desired results.

4.2.1.2 Product Features

The features of the system are given below:

- The user will be able to access the system with the help of a username and password.

- The dataset used for the project is provided by the USDA. It has around 10,000 food items along with their nutritional value.

- The data will be divided into different clusters using different algorithms and the best will be identified.

- The comparison will be done according to the food items provided by the user.

- The comparison will be based on the basis of having the nearest possible clusters.

4.2.1.3 User Characteristics

The interactive user interface will be sufficient to enable its user to use the system with ease. Further, a help icon will be provided to read the directions for using the system. Users without any previous understanding of computer systems will be able to adapt themselves for the UI.

4.2.1.4 Assumptions & Dependencies

- The modules included in python for different mining algorithms yields the accurate results.

- The food entered by the user is present in the dataset. Else, fuzzy logic techniques will be used to find the most appropriate food.

- The algorithms used have already be previously used in different fields.

- The scatter plots and the graphs are dependent on the types of algorithms used

4.2.2 Non Functional Requirements

4.2.2.1 Product Requirements

- Efficiency: In order to calculate the efficiency of the system, many algorithms will be used to find the relationships between the data items and for the comparison of the food products which thereby will help in analyzing the speed

and memory. The analysis of the results will be done using various pictorial representation to have a clear insight into the system.

- Portability: The system will be globally hosted. User from any part of the world can enjoy the ease of accessing the system by just having a proper internet connection and any type of browsers.

- Usability: The end user will be able to access the system without any previous training. Instructions for using the system will be provided in the help menu. Efforts will be made to keep the front end attractive and simple as it can be.

- Reliability: The mean time between failures for the system will be 3 hours.

4.2.3 Operational Requirements

- Economic: The system will be free for its user to access it. The part where it is being deployed should have healthy economy where people can afford a computer.

- Environmental: The deployment focus mostly will done in those areas where people are generally health conscious and where people have mineral deficiency diseases.

- Political: Government's help will be required in order to promote the system where it can reach the masses.

- Ethical: Online security for the domain will be required to prevent the system from being hacked and falling into wrong hands

- Health and Safety: People's health is the main objective of the project. No negligence will be given in this sector.

- Legality: The dataset that is being used is verified by USDA. Efforts should be made to give proper credits to all other third parties in order to gain their belief.

4.2.4 System Requirements

4.2.4.1 Hardware Requirements

- CPU:(2 x Intel Core 2(2.66 Ghz,128K cache)
- RAM:2GB
- Minimum database space:10GB

4.2.4.2 Software Requirements:

- PhpMyAdmin:

  It is an open source tool to handle the database with the use of web browser. It can be used create a table, delete the data of it, modify the values and managing the granted permissions.
- HTML, CSS, JavaScript:

  These tools would be used to create the frontend i.e. the webpage through which the user would enter the value .CSS would add various styles and lucrative look to the webpage. JavaScript would handle the dynamic part of the webpage like moving graphics, audio, and video. It would also help in client side scripting by checking any type
  Of discrepancies in data.

- Python Modules:
  - SciPy Libraries: It would contain the modules which would help in interpolation, optimization, functions and machine learning tools for training datasets.

    It would contain the clustering libraries like K-means and SVM which would be used in Unsupervised Learning.
  - Numpy Libraries: It is a python library which allows user to add large, multi-dimensional arrays and matrices. It has many high level mathematical functions.
  - Pandas Library: It is a python software library which would be used for data manipulation and analysis of tables. It would be useful for reshaping and pivoting of datasets.

- MS Excel **:**

  MS Excel is spreadsheet tools which would help filtering the collected data from the form in specified CSV (Comma Separated Values) which would be later used as our datasets.

- Web Browsers
  - Internet Explorer 10,11 & Edge
  - Firefox(recent versions)
  - Chrome, Safari(recent versions)

5. RESULT AND DISCUSSION

5.1 RESULTS

5.1.1 Cleaning of dataset and removing the missing values:

We have used USDA food dataset which contains around 8000 food items. The dataset at the beginning had some missing values. So that missing values were replaced by zero to avoid any type of discrepancy. After this, we have changed the elements in dataset to "float-64" data format. Then, we selected important features from the dataset i.e. Carbohydrates, Energy, Protein, Fiber, Vitamin E, Thiamin and Vitamin C.

	CARBOHYDRATES	ENERGY	PROTEIN	FIBER	VITAMIN E	THIAMIN	VITAMIN C
0	0.06	717	0.85	0.0	2.32	0.005	0.0
1	0.06	717	0.85	0.0	2.32	0.005	0.0
2	0.00	876	0.28	0.0	2.80	0.001	0.0
3	2.34	353	21.40	0.0	0.25	0.029	0.0
4	2.79	371	23.24	0.0	0.26	0.014	0.0
5	0.45	334	20.75	0.0	0.24	0.070	0.0
6	0.46	300	19.80	0.0	0.21	0.028	0.0
7	3.06	376	25.18	0.0	0.00	0.031	0.0
8	1.28	403	24.90	0.0	0.29	0.027	0.0
9	4.78	387	23.37	0.0	0.00	0.046	0.0
10	2.57	394	23.76	0.0	0.28	0.015	0.0
11	3.38	98	11.12	0.0	0.08	0.027	0.0
12	4.61	97	10.69	0.2	0.04	0.033	1.4
13	6.66	72	10.34	0.0	0.01	0.023	0.0
14	3.66	86	11.83	0.0	0.04	0.041	0.0
15	2.72	72	12.39	0.0	0.01	0.021	0.0
16	4.07	342	5.93	0.0	0.29	0.020	0.0
17	1.43	357	24.99	0.0	0.24	0.037	0.0
18	4.09	264	14.21	0.0	0.18	0.154	0.0
19	1.55	389	25.60	0.0	0.27	0.021	0.0
20	42.65	466	9.65	0.0	0.00	0.315	0.0
21	2.22	356	24.94	0.0	0.24	0.030	0.0
22	0.36	413	29.81	0.0	0.28	0.060	0.0
23	0.49	327	20.05	0.0	0.23	0.080	0.0
24	0.68	373	24.48	0.0	0.26	0.015	0.0
25	2.19	300	22.17	0.0	0.19	0.030	0.0
26	2.47	318	21.60	0.0	0.21	0.016	0.0
27	2.77	254	24.26	0.0	0.14	0.018	0.0
28	3.83	302	25.96	0.0	0.37	0.101	0.0
29	1.12	368	23.41	0.0	0.26	0.013	0.0
...	...	...	...	...	...	...	...

Table 5.1 USDA dataset used in Food Recommendation System.

5.1.2 Normalization of food dataset:

The food dataset have different attributes having values ranging to some values. So, the data needs to be converted into dataset ranging from (0, 1).

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Fig 5.1 Formula used in normalization

ROW	ITEM	CARBOHY	ENERGY	PROTEIN	FIBER	VITAMIN	THIAMIN	VITAMIN C
1001	BUTTER,W	0.0006	0.7949	0.009624	0	0.010065	0.000214	0
1002	BUTTER,W	0.0006	0.7949	0.009624	0	0.010065	0.000214	0
1003	BUTTER OI	0	0.971175	0.00317	0	0.012148	0.000043	0
1004	CHEESE,BL	0.0234	0.391353	0.242301	0	0.001085	0.001241	0
1005	CHEESE,BF	0.0279	0.411308	0.263134	0	0.001128	0.000599	0
1006	CHEESE,BF	0.0045	0.370288	0.234941	0	0.001041	0.002995	0
1007	CHEESE,CA	0.0046	0.332594	0.224185	0	0.000911	0.001198	0
1008	CHEESE,CA	0.0306	0.416851	0.2851	0	0	0.001326	0
1009	CHEESE,CH	0.0128	0.446785	0.281929	0	0.001258	0.001155	0
1010	CHEESE,CH	0.0478	0.429047	0.264606	0	0	0.001968	0
1011	CHEESE,CC	0.0257	0.436807	0.269022	0	0.001215	0.000642	0
1012	CHEESE,CC	0.0338	0.108647	0.125906	0	0.000347	0.001155	0
1013	CHEESE,CC	0.0461	0.107539	0.121037	0.002532	0.000174	0.001412	0.000583
1014	CHEESE,CC	0.0666	0.079823	0.117074	0	0.000043	0.000984	0
1015	CHEESE,CC	0.0366	0.095344	0.133945	0	0.000174	0.001754	0
1016	CHEESE,CC	0.0272	0.079823	0.140285	0	0.000043	0.000898	0
1017	CHEESE,CF	0.0407	0.379157	0.067142	0	0.001258	0.000856	0
1018	CHEESE,ED	0.0143	0.395787	0.282948	0	0.001041	0.001583	0
1019	CHEESE,FE	0.0409	0.292683	0.160892	0	0.000781	0.006588	0

Table 5.2 dataset after normalization of data.

5.1.3 Dimension Reduction:

In dataset we have many parameters. So, this makes very difficult for us to interpret them, and therefore we need to reduce the dimension of the dataset. We have used PCA (Principal component Analysis) which uses principal of linear combination of vector to map it on to low-dimensional space. It becomes very easier for low –dimensional data to map them on 2-dimensional space.

5.1.4. Clustering of food items:

Clustering is a process in which we group food items according to similarity in their features and positioning them in same group.

We have applied 3 clustering algorithm to clusters food dataset:

K-Mean Clustering:

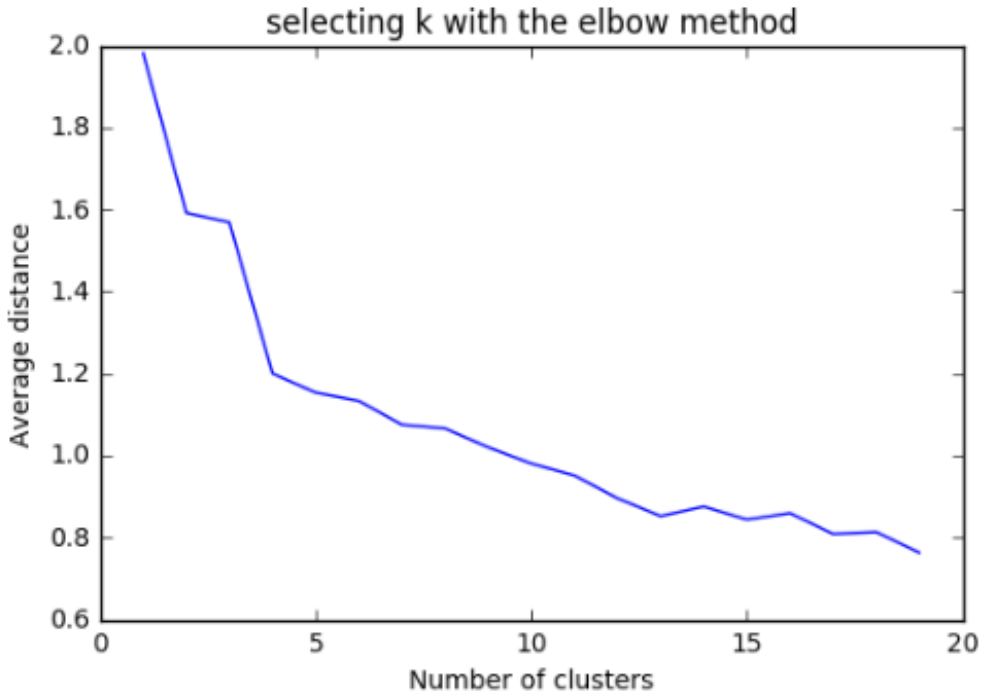
K-means divides the clusters where distance between the data items within the same group should be least and distance between the data items within the different group should be maximum i.e. intra cluster distance should be minimum and inter cluster should be maximum.

Here, k (number of clusters) is calculated using the graph of average distance of cluster elements and value of k. The point in graph where we find a sharp bend and has less average distance is considered to be our k and to validate the number of clusters in sample is to simply run the algorithm till a particular range like 1 to 20 and, for each value of k the sum of squared error (SSE) is calculated.

The SSE for each value of K is then plotted in graph and the point where the line chart looks like an arm or have sharp bent is elbow point is the optimal value of k

By observing the graph we can find out the value of k:

Fig 5.2 Graph between average distance and k values.



Davies-Bouldin Index:

It is a metrics for evaluating the clustering algorithms. The less magnitude of DB index indicates good cluster as we get the cluster compact in nature and having their center far way. It does not exhibit any trend with respect to number of clusters.

For above experiment Davies-Bouldin index is coming as this for different values of k:

K(value)	DB index
3	-1.6
4	-1.18
5	-1.19
6	-0.810
7	-0.751
8	-0.766
9	-0.823
10	-0.993
11	-1.01

Table 5.3 Davies-Bouldin Index table

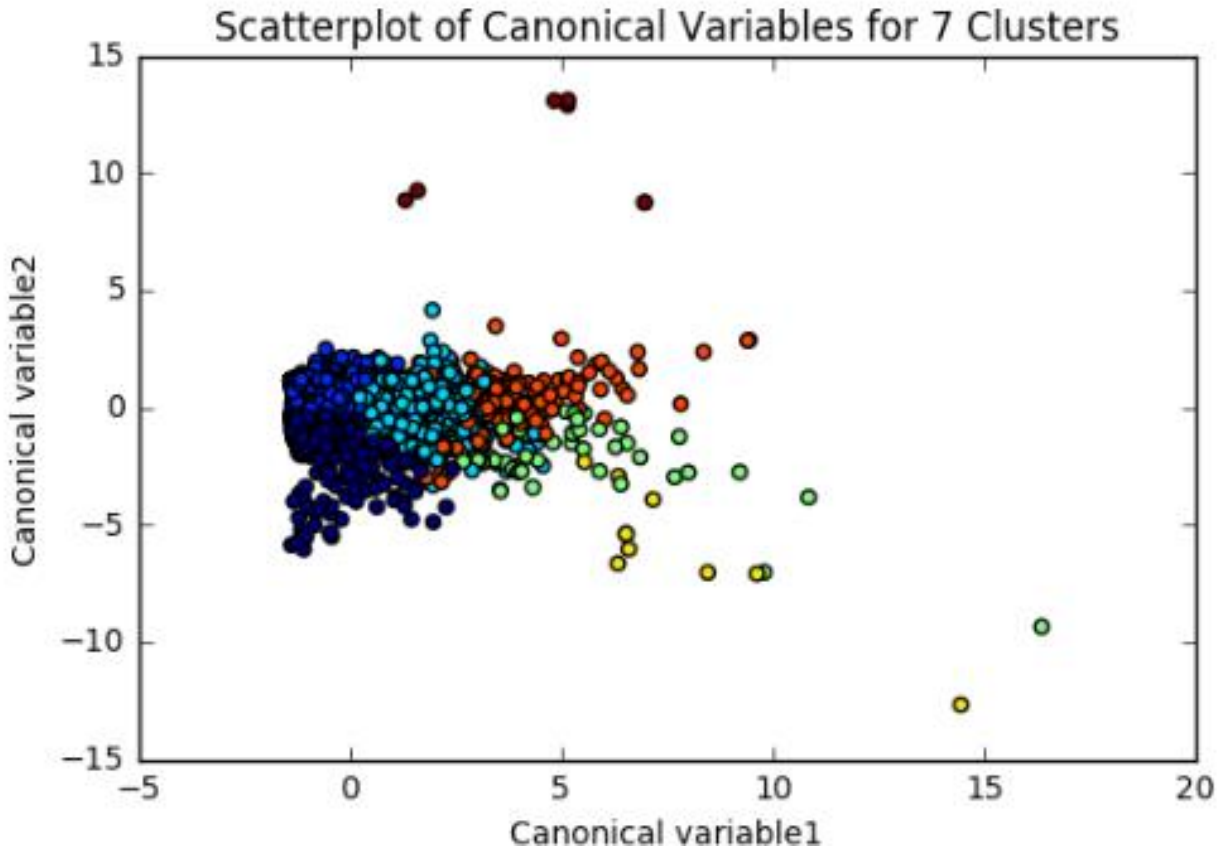


Fig 5.3 K-mean Clustering graph

DBSCAN Algorithm:

The points in dataset are group together that are close together and reachable from each other and the points which are outlier point are low-density points.In this algorithm points are classified as three categories that are core points,density-reachable points and outliers. As, our dataset is very large so, we have applied the DBSCAN here.

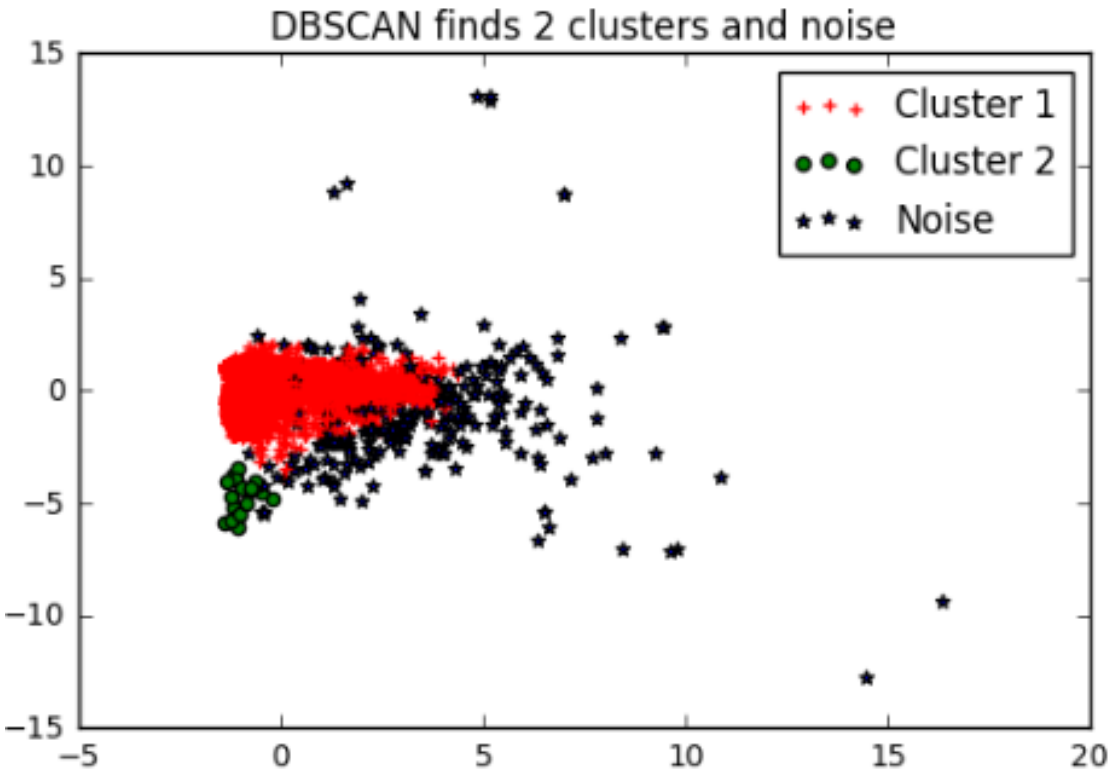


Fig 5.4 DBSCAN Clustering graph

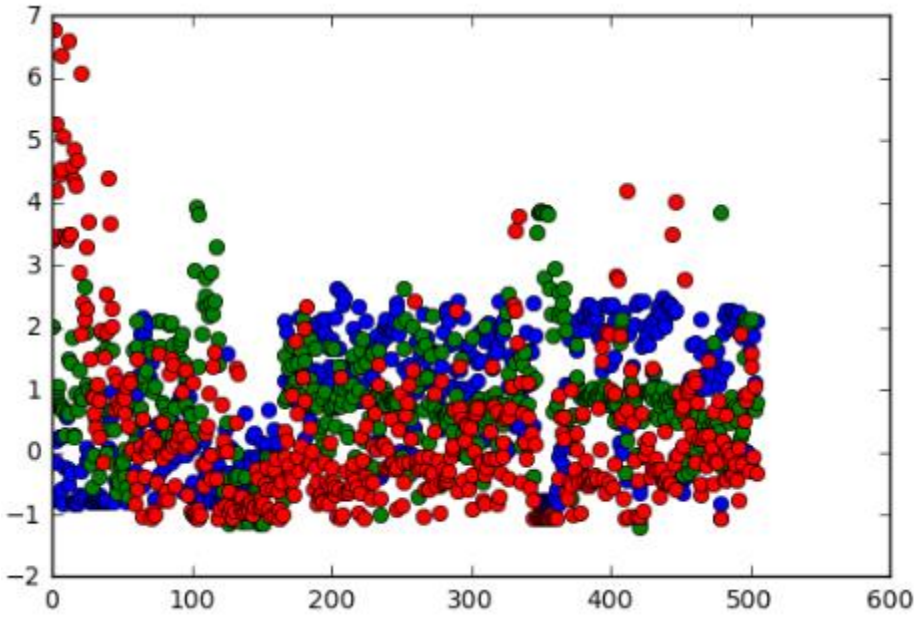
It requires 2 parameters:

i. ε (maximum radial distance from point): It is calculated by the observing average cluster distance graph and number of clusters parameter.

ii. Minimum samples: It is the value of k where there is a sharp bend in the graph SSE and k.

BIRCH Algorithm:

As our dataset is too large so, we are applying it to observe the results. Here We have chosen branching factor as 50.



5.5 BIRCH clustering graph

5.1.5. Clustering Algorithm Comparison:

Here ,we are comparing algorithm on the basis of best categorization of food i.e. in which food are grouped in particular clusters according to their nutritional value here, time is not considered a factor in comparison.

We have chosen **silhouette coefficient** as a parameter

It shows how efficiently the data points lies under its cluster. It measures the extent to which each object lies within its cluster as compared to other cluster. The range of S-coefficient lies between -1 to 1.The value closer to 1 shows that the object matches to its cluster and value near to -1 means the object don’t matches to its cluster.

The values of silhouette coefficient for 3 algortihms are

Algorithm	Silhouette Coefficient
K-mean	0.44
Birch	0.21
DBSCAN	0.32

Table 5.4 Silhouette coefficient table

K-mean has highest positive silhouette coefficient value so it is best.

5.1.6 .Cluster Categorization:

We have used clusters obtained from the K-mean as our database and classified them according to their nutritional value.

We have obtained 7 clusters and they are classified into 6 food categories:

a. Limited Food:

The food items which have moderate amount of sugar and carbohydrates and lacks in protein and vitamins are called limited food.

b. Avoidable Food:

The food items which have excess of fat, complex carbohydrates and sugar in it are called avoidable food.

c. Normal Food:

The food items which contains all nutrients in moderate amount are called Normal Food.

d. Enriching Food:

The food items which contains protein in high amount are called Enriching Food.

e. Preferable Food:

The food items which contains all important nourishing nutrient in high amount are called preferable food.

f. Average food:

The food items which contains mainly fibers and neither carbohydrates and sugar in high amount and also lacks in protein and vitamins .So, there intake don’t affect our health too much .These food items are called Average food.

5.1.7 Working of Food Recommendation System:

5.1.7.1 User Authentication Module:

The first step involve in project is to login to the account using username and password. If user credentials don't matches then he is not allowed to sign in. It can be seen in figure given below.

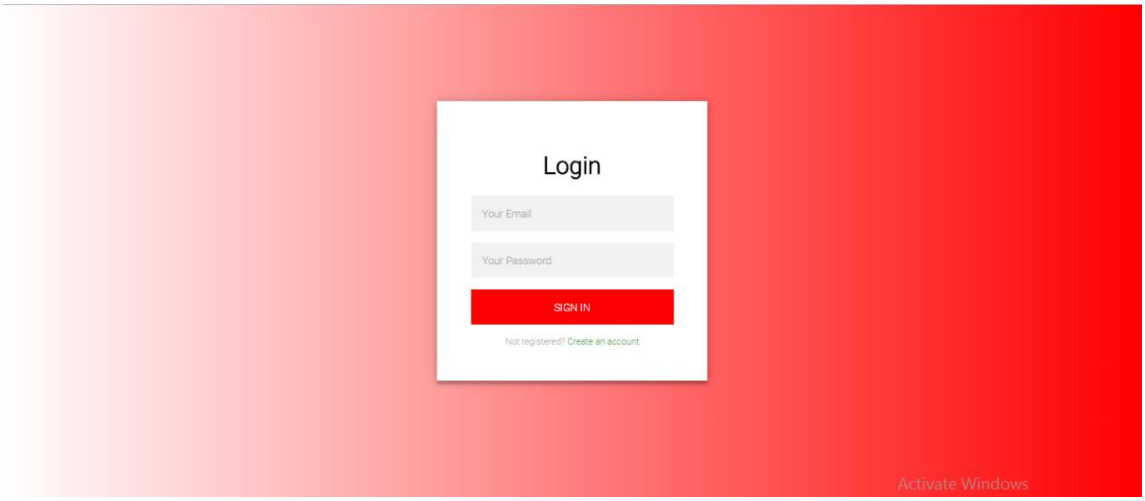


Fig 5.6 Login page

5.1.7.2 New User Registration:

Fig shows the registration form which any need to fill whenever he opens food recommendation system homepage. It has 5 rows which needs to be filled by user i.e. Name, email, age, phone number and password.

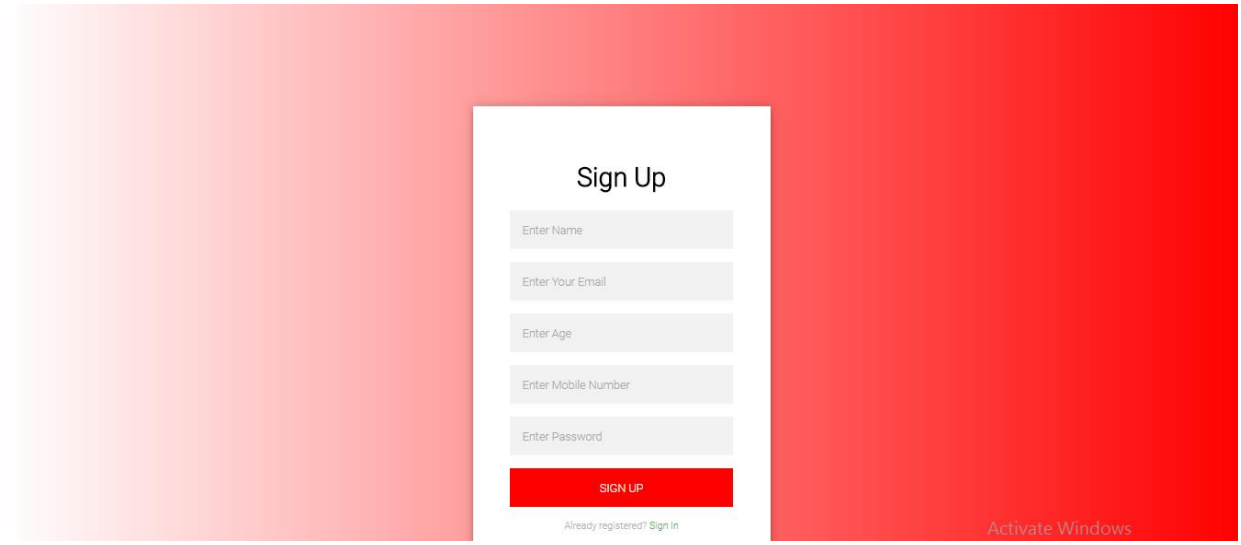


Fig 5.7 Sign up page

5.1.7.3 Food category selection:

When a user login to the system then they are provided with six food category according to their choices. It can be seen in the figure and he just has to click on it to go on to next page.

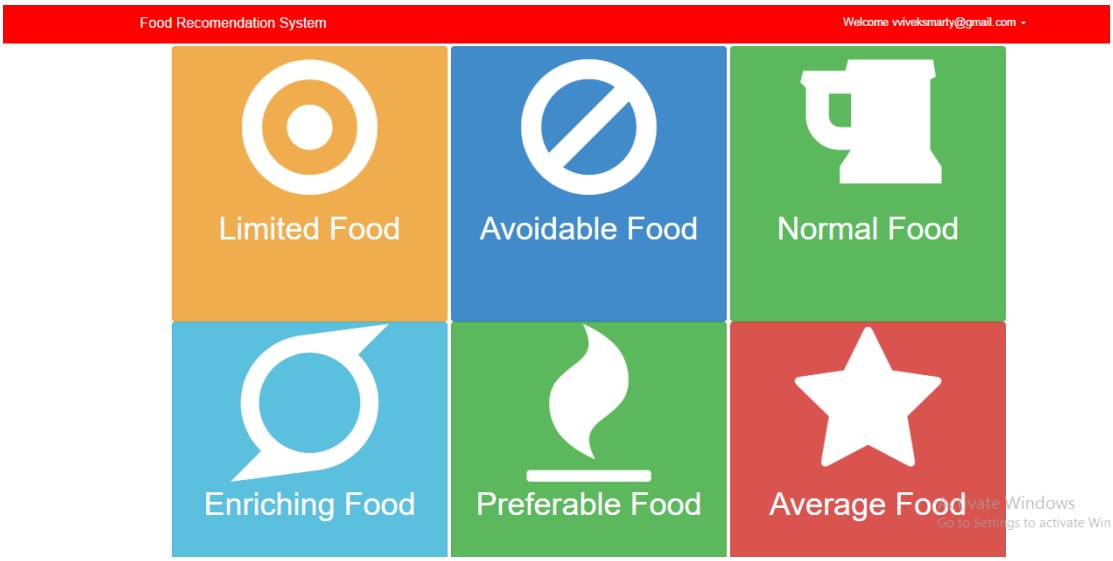


Fig 5.8 Food Recommendation Homepage

5.1.7.4 Food Group selection:

After selecting food category user need to choose the food choice that he needs to choose from particular food group like chicken, beans, etc. It can be seen in figure given below. User need to tick the checkbox and click on submit.

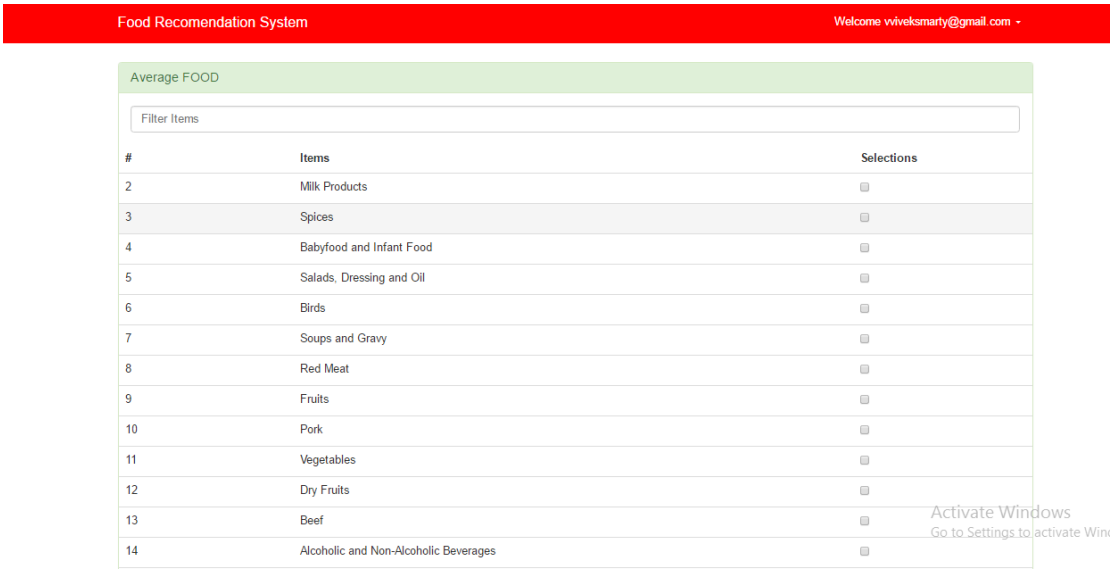


Fig 5.9 Food group searching page

5.1.7.5 Filtering the food result:

The food items have many food items so, user can filter his desired choice food items by typing the name .The food items would come according to the substring value. It can be seen in given figure

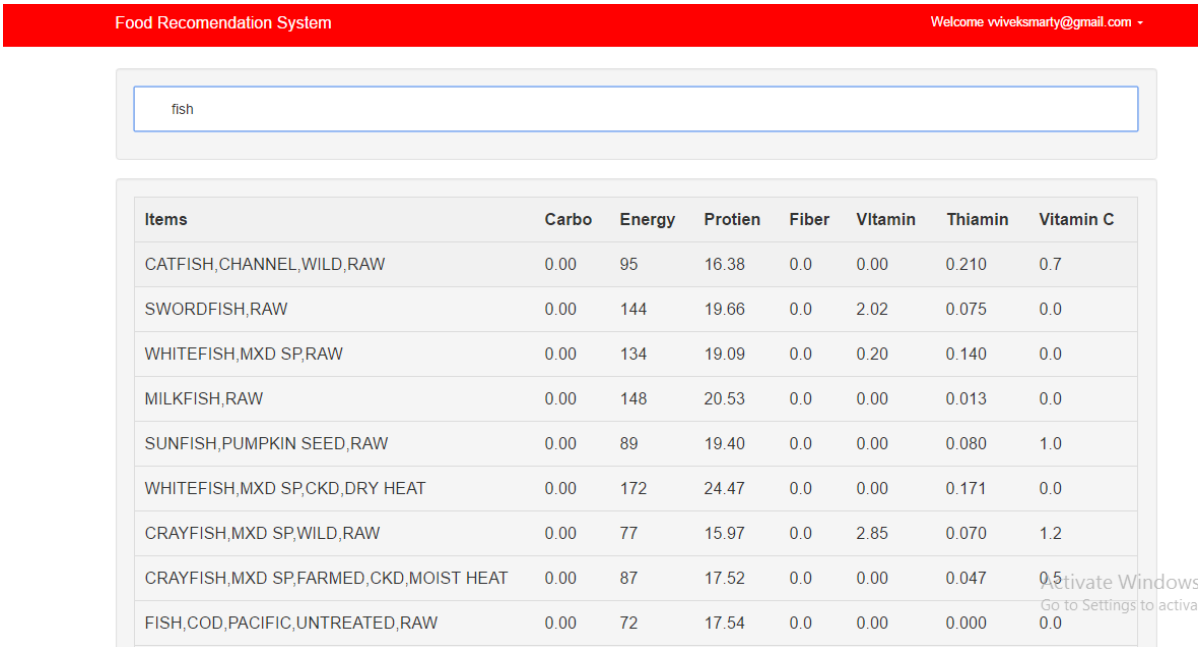


Fig 5.10 Food item filtering page

5.1.7.6 Database module:

We have used phpMyAdmin of wamp to store our data and create our local server.

We have 11 tables in our database. They are given below:

- Cluster0
- Cluster1
- Cluster2
- Cluster3
- Cluster4
- Cluster5
- Cluster6

  The cluster table contains the row number of the food item present in particular cluster.

- Data_Set:

  The dataset table is the parent table which has all its values like group number, food items, and nutrient value and used to reference other table.

- Food choice:

  It classify food into different groups like beans, milk products, etc.

- Pref:

  It contains the information that whether any food item is veg or non-veg.

- Users:

It is the table which contains the information about the users.

The database tables can be seen in figure given below

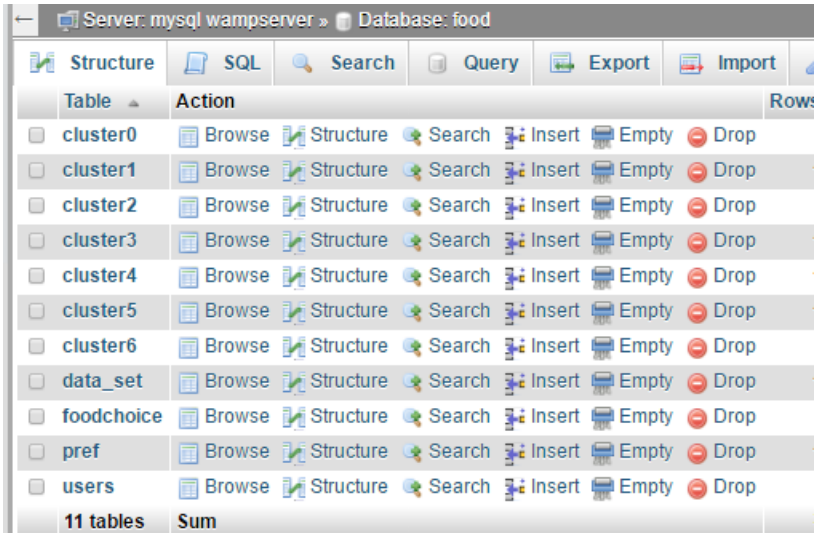


Fig 5.11 Food dataset

5.2 Sample Test Cases

T_ID	Test Case Description	Input	Expected Output	Actual Output	Result
1	Logging in without any details	Null	Please enter the details	Please enter the details	Success
2	Logging in without sign up	Username: abc@gmail.com Password: 123567	Username or password is incorrect	Incorrect Credentials, try again	Success
3	Logging in without providing password	Username: abc@gmail.com Password: (NULL)	Enter password	Please enter your password	Success
4	Logging in without email id and without password	Username:(NULL) Password:123456	Enter email address	Please enter your email address	Success
5	Redirecting to signup page	Click “create an account”	Redirected to signup page	Redirected to signup page	Success
6	Entering Username without other credentials	Name: abc Other details: NULL	Enter other details	Please enter email id, password, etc.	Success

7	Entering email id without other credentials	Email id: abc@gmail.com Other details: NULL	Enter other details	Please enter password, etc.	Success
8	Entering password without other credentials	Password: 123456 Other details: NULL	Enter other details	Please enter password, etc.	Success
9	Entering email id without correct format	Email: abc.gmail.com	Incorrect Email format	Incorrect Email format	Success
10	Entering password without correct format	Password: 123	Password is too short	Password must be greater than 6 characters	Success
11	Entering age other than numbers	Age: abc	Enter valid age	Age should be in numbers	Success
12	Entering mobile number other than just numbers	Mobile Number: ac123	Enter valid number for mobile number	Mobile number should be in numbers	Success
13	Entering all the required credentials without any error	Name: abc Email: abc@gmail.com Password: 12345678 Age: 20 Mobile Number: 9659895622	User successfully registered	User successfully registered	Success
14	Redirecting to login page after signup	Click “Click here to login”	Redirected to login page	Redirected to login page	Success
15	Logging in with correct email id and incorrect password	Email: abc@gmail.com Password: 123456	Username or password incorrect	Incorrect Credentials	Success
16	Logging in with incorrect email id and correct password	Email: abc@msil.com Password: 12345678	Username or password incorrect	Incorrect Credentials	Success

17	Logging with correct credentials	Email: abc@gmail.com Password: 12345678	Logged in successfully and redirected to homepage	Redirected to homepage	Success
18	Redirecting to limited food page	Clicking “limited food” icon	Redirected to limited food page	Redirected to limited food page	Success
19	Redirecting to homepage	On Clicking “Food Recommendation System”	Redirected to homepage	Redirected to homepage	Success
20	Redirecting to avoidable food page	Clicking “avoidable food” icon	Redirected to avoidable food page	Redirected to avoidable food page	Success
21	Redirecting to normal food page	Clicking “normal food” icon	Redirected to normal food page	Redirected to normal food page	Success
22	Redirecting to enriching food page	Clicking “enriching” icon	Redirected to enriching food page	Redirected to enriching food page	Success
23	Redirecting to preferable food page	Clicking “preferable food” icon	Redirected to preferable food page	Redirected to preferable food page	Success
24	Redirecting to average food page	Clicking “average food” icon	Redirected to average food page	Redirected to average food page	Success
25	Selecting no food category in limited food page	Click submit without check in any option	No result should be displayed	No results	Success
26	Selecting only “Milk Products” in limited food	Check “Milk Products”	Recommended milk products	Milk products get displayed	Success
27	Selecting all the food items in limited food	Check all items	Display all the recommended food results	All the recommended food are displayed	Success
28	Selecting no food category in avoidable food	Click submit without check in any option	No result should be displayed	No results	Success
29	Selecting only “Spices” in avoidable food	Check “Spices”	Recommended Spices	Spices get displayed	Success

30	Selecting all the food items in avoidable food	Check all items	Display all the recommended food results	All the recommended food are displayed	Success
31	Selecting no food category in normal food	Click submit without check in any option	No result should be displayed	No results	Success
32	Selecting only “Fruits” in normal food	Check “Fruits”	Recommended Fruits	Fruits get displayed	Success
33	Selecting all the food items in normal food	Check all items	Display all the recommended food results	All the recommended food are displayed	Success
34	Selecting no food category in enriching food	Click submit without check in any option	No result should be displayed	No results	Success
35	Selecting only “Birds” in enriching food	Check “Birds”	Recommended Birds	Birds get displayed	Success
36	Selecting all the food items in enriching food	Check all items	Display all the recommended food results	All the recommended food are displayed	Success
37	Selecting no food category in Preferable food	Click submit without check in any option	No result should be displayed	No results	Success
38	Selecting only “Fruits” in Preferable food	Check “Vegetables”	Recommended Vegetables	Vegetables get displayed	Success
39	Selecting all the food items in preferable food	Check all items	Display all the recommended food results	All the recommended food are displayed	Success
40	Checking search filter in result page	Type “abc” in search box	Every food item with substring “abc” shall be displayed	No result with substring: “abc” found	Success
41	Checking search filter in result page	Type “chili” in search box	Every food item with substring	Food items with substring	Success

			"chili" shall be displayed	"chili" get displayed	
42	Checking random results	Check "Milk products" & "spices" in limited foods and clicking submit	Results should be in random order	All the recommended food items are randomized	Success
43	Redirecting to login page after clicking logout	Click logout option	Logged out successfully	Logged out successfully and redirected to login page	Success

5.3 Summary of the Result

We have taken USDA food data as our dataset on which we applied normalization and eliminated missing parameters to cleanse the data. After that, we have applied PCA algorithm on it to reduce its dimension so that it reduces our calculation time and it would be easier to map it on the 2-d space. After dimension reduction, we implemented three clustering techniques K-mean, DBSCAN and BIRCH to cluster food items in same group according to nutrients. We used silhouette coefficient as a parameter to find out which is the best suited clustering method for our dataset. We find out that for K-mean, silhouette coefficient was coming better than other algorithm. We made clusters of dataset using k-mean clustering. The value of k i.e. the number of clusters were coming best for k-mean. We made a database in cluster tables, main database table, user, food choice and preferences tables were added. We have divided food items in groups according to their categories like milk products are categorized in one table, fish products in other. The seven clusters are categorized in 6 food groups according to the nutrition in it. They are normal food, recommended food, limited food, avoidable food, enriched food and average food. Now, using front end and applying proper queries we can recommend particular food to user according to their choice what they choose in the Food Recommended System. We have a text box also available through which we can get the desired results according to the food choice typed in it. Therefore, food recommendation system helps users to select food items according to his taste and also keeping him aware that which food items is nutritious and which is not.

6. CONCLUSION, LIMITATIONS AND SCOPE FOR FUTURE WORK

6.1 CONCLUSION

In present world health is one of the most important issue every person is facing. Food intake is one of the most important factor effecting this because most of the diseases are caused to bad food habits.

Food recommendation system can help in spreading awareness amongst people about the nutritious value of the food which they eat. It classifies food into different categories which gives user clue that that these food items are desirable for them and other food items which are not desirable for them. It also requires a thorough study of various clustering and dimension reduction algorithm which is best suited for a proper categorization of food items. Therefore, Food recommendation system would help users to eat food items of their choices.

6.2 LIMITATIONS

It would only give rough idea about food item and there is no individual comparison between foods items present in a cluster.

It would require different clusters for different types of diseases as every disease is caused by different types of nutrition parameters.

The food items differ from region to region. So, we require different dataset for different regions.

6.3 SCOPE FOR FUTURE WORK

The present work is only limited to particular region it can be extended according to various demographic parameters in future. Also, this work mostly focuses on diabetes So we increase our work to various different disease which is caused due to deficiency of nutrient. Also in future we can apply some machine learning tools which would help in recommending food items which most preferable for person according to the input value feed by user.

7. APPENDIX

Appendix A

```
#Python libraries used for algorithms

from pandas import Series, DataFrame
import pandas as pd
import numpy as np
import matplotlib.pylab as plt
from sklearn.cross_validation import train_test_split
from sklearn import preprocessing
from sklearn.cluster import KMeans



#input data from file
data=pd.read_csv("c:\\users\\anupam\\desktop\\B.csv")
#making all dataframe name uppercase
data.coloumns=map(str.upper,data.columns)



#data cleaning
data_clean=data.dropna()



#subset the clustering variables
cluster=data_clean[['CARBOHYDRATES','ENERGY','PROTEIN','FIBER','VITAMI
N E','THIAMIN','VITAMIN C']]
cluster.describe()
print(cluster)



#Standardize
clustervar=cluster.copy()
```

```
clustervar['CARBOHYDRATES']=preprocessing.scale(clustervar['CARBOHYDRATES'].astype('float64'))
clustervar['ENERGY']=preprocessing.scale(clustervar['ENERGY'].astype('float64'))
clustervar['PROTEIN']=preprocessing.scale(clustervar['PROTEIN'].astype('float64'))
clustervar['FIBER']=preprocessing.scale(clustervar['FIBER'].astype('float64'))
clustervar['VITAMIN E']=preprocessing.scale(clustervar['VITAMIN E'].astype('float64'))
clustervar['THIAMIN']=preprocessing.scale(clustervar['THIAMIN'].astype('float64'))
clustervar['VITAMIN C']=preprocessing.scale(clustervar['VITAMIN C'].astype('float64'))


#SPLITTING OF DATA in training and test set
#clus_train,clus_test=train_test_split(clustervar,test_Size=0.3)


#k mean cluster analysis for 1-20 clusters
from scipy.spatial.distance import cdist
clusters=range(1,20)
meandist=[]

for k in clusters:
    model=KMeans(n_clusters=k)
    model.fit(clustervar)
    clusassign=model.predict(clustervar)
meandist.append(sum(np.min(cdist(clustervar,model.cluster_centers_,'euclidean'),axis=1))/clustervar.shape[0])

#elbow method to find the k
f1=plt.figure()
f2=plt.figure()
plt.plot(clusters,meandist)
plt.xlabel('Number of clusters')
plt.ylabel('Average distance')
```

```
plt.title('selecting k with the elbow method')
plt.show()


#Interpret 7 cluster solution
model3=KMeans(n_clusters=7)
model3.fit(clustervar)
clusassign=model3.predict(clustervar)

from sklearn.decomposition import PCA
pca_2=PCA(2)
plot_coloumns=pca_2.fit_transform(clustervar)
plt.scatter(x=plot_coloumns[:,0],y=plot_coloumns[:,1],c=model3.labels_,)
plt.xlabel('Canonical variable1')
plt.ylabel('Canonical variable2')
plt.title('Scatterplot of Canonical Variables for 7 Clusters')
plt.show()
print(plot_coloumns[:,0])
print(plot_coloumns[:,1])
```

APPENDIX B

#Login and Signup page

```
<html>
<head>
   <title>Food Recomendation System</title>
</head>
<link rel="stylesheet" href="login/style.css" type="text/css" />
<body>
```

```
<div class="login-page">
  <div class="form">
    <form class="register-form" method="POST" >

     <?php
      if ( isset($errMSG) ) {

       ?>
       <div class="form-group">
            <div class="alert alert-<?php echo ($errTyp=="success") ? "success" :
$errTyp; ?>">
       <span class="glyphicon glyphicon-info-sign"></span> <?php echo $errMSG; ?>
             </div>
            </div>
             <?php
     }
     ?>
<h1>Sign Up </h1>

            <span><?php echo $nameError; ?></span>
            <input type="text" name="name" class="form-control" placeholder="Enter
Name" maxlength="50" value="<?php echo $name ?>" required />
            <span ><?php echo $emailError; ?></span>
            <input type="email" name="email" class="form-control" placeholder="Enter
Your Email" maxlength="40" value="<?php echo $email ?>" required />
            <input type="text" name="age" class="form-control" placeholder="Enter
Age" maxlength="50" value="<?php echo $age ?>" required/>
            <input type="text" name="mobile" class="form-control" placeholder="Enter
Mobile Number" maxlength="50" value="<?php echo $mobile ?>" required/>
            <span><?php echo $passError; ?></span>
            <input type="password" name="pass" class="form-control"
placeholder="Enter Password" maxlength="15" required />
            <button type="submit" class="btn btn-block btn-primary" name="btn-
signup">Sign Up</button>
```

```
    <p class="message">Already registered? <a href="#">Sign In</a></p>
   </form>



<form method="post" class="login-form" action="<?php echo
htmlspecialchars($_SERVER['PHP_SELF']); ?>" autocomplete="off">
      <?php
        if ( isset($errMSG) ) {
      ?>
     </span> <?php echo $errMSG; ?></span>

      <?php
          }
      ?>

      <h1>Login </h1>

    <span><?php echo $emailError; ?></span>
    <input type="email" name="email" class="form-control" placeholder="Your
Email" value="<?php echo $email; ?>" maxlength="40" />
    <span><?php echo $passError; ?></span>
    <input type="password" name="pass" class="form-control" placeholder="Your
Password" maxlength="15" />
    <button type="submit" class="btn btn-block btn-primary" name="btn-login">Sign
In</button>

<p class="message">Not registered? <a href="#">Create an account</a></p>
   </form>
  </div></div>
```

#navigation bar throughout the system

```
<div class="navbar-wrapper">
   <div class="container-fluid">
      <nav class="navbar navbar-fixed-top">
         <div class="container">
            <div class="navbar-header">
               <button type="button" class="navbar-toggle collapsed" data-
toggle="collapse" data-target="#navbar" aria-expanded="false" aria-
controls="navbar">
               <span class="sr-only">Toggle navigation</span>
               <span class="icon-bar"></span>
               <span class="icon-bar"></span>
               <span class="icon-bar"></span>
               </button>
               <a class="navbar-brand" href="#">Food Recomendation System</a>
            </div>

            <div id="navbar" class="navbar-collapse collapse">

               <ul class="nav navbar-nav pull-right">
                  <li class=" dropdown"><a href="#" class="dropdown-toggle active"
data-toggle="dropdown" role="button" aria-haspopup="true" aria-
expanded="false">Welcome <?php echo $userRow['userEmail']; ?>  <span
class="caret"></span></a>
                     <ul class="dropdown-menu">
                        <li><a href="login/logout.php?logout">Log Out</a></li>
                     </ul>
                  </li>
               </ul>
            </div>
         </div>
      </nav></div></div>
```

#homepage

```
<div class="container" align="center">
    <div class="row">
      <div class="col-lg-12">

        <p>
          <a class="btn btn-sq-lg btn-warning" href="categories/limited.php" >
             <i class="fa fa-dot-circle-o fa-5x"></i><br/>
            Limited Food
          </a>
          <a  class="btn btn-sq-lg btn-primary" href="categories/avoidable.php" >
           <i class="fa fa-ban fa-5x"></i><br/>
         Avoidable Food
          </a>
          <a class="btn btn-sq-lg btn-success" href="categories/normalVeg.php">
           <i class="fa fa-beer fa-5x"></i><br/>
          Normal Food
          </a>
          <a class="btn btn-sq-lg btn-info" href="categories/enrich.php" >
           <i class="fa fa-superpowers fa-5x"></i><br/>
          Enriching Food
          </a>
                              </a>
          <a class="btn btn-sq-lg btn-success" href="categories/prefer.php" >
           <i class="fa fa-fire fa-5x"></i><br/>
          Preferable Food
          </a>
                              </a>
          <a class="btn btn-sq-lg btn-danger" href="categories/average.php" >
           <i class="fa fa-star fa-5x"></i><br/>
          Average Food
          </a>
```

```
        </p>
       </div>
  </div>
```

#limited food page

```
<form method="POST" action="../outputs/limitedfoodresult.php">
        <table class="table table-hover" id="dev-table">
          <thead>
            <tr>
              <th>#</th>
              <th>Items</th>
              <th></th>
              <th>Selections</th>

            </tr>
          </thead>
          <tbody>
            <tr>
              <td>1</td>
              <td>Milk Products</td>
              <td></td>
              <td>
                    <input type="hidden" name="milk" value="0" />
                    <input type="checkbox" id="" value="1" name="milk"></td>
            </tr>
            <tr>
              <td>2</td>
              <td>Spices</td>
              <td></td>
              <td>
                    <input type="hidden" name="spices" value="0" />
                    <input type="checkbox" id="" value="2" name="spices"></td>
            </tr>
```

```
<tr>
 <td>3</td>
 <td>Baby and Infant's Food</td>
 <td></td>
 <td>
      <input type="hidden" name="baby" value="0" />
      <input type="checkbox" id="" value="3" name="baby"></td>
</tr>
 <tr>
 <td>4</td>
 <td>Salads, Dressing and Oil</td>
 <td></td>
 <td>
      <input type="hidden" name="salads" value="0" />
      <input type="checkbox" id="" value="4" name="salads"></td>
</tr>
                  <tr>
 <td>5</td>
 <td>Soup and Gravy</td>
 <td></td>
 <td>
      <input type="hidden" name="soup" value="0" />
      <input type="checkbox" id="" value="6" name="soup"></td>
</tr>
<tr>
 <td>6</td>
 <td>Cereals and Derivatives</td>
 <td></td>
 <td>
      <input type="hidden" name="cereals" value="0" />
      <input type="checkbox" id="" value="8" name="cereals"></td>
</tr>
<tr>
 <td>7</td>
```

```
  <td>Fruits</td>
  <td></td>
  <td>
        <input type="hidden" name="fruits" value="0" />
        <input type="checkbox" id="" value="9" name="fruits"></td>
 </tr>
 <tr>
  <td>8</td>
  <td>Pork</td>
  <td></td>
  <td>
        <input type="hidden" name="pork" value="0" />
        <input type="checkbox" id="" value="10" name="pork"></td>
 </tr>
<tr>
  <td>9</td>
  <td>Vegetables</td>
  <td></td>
  <td>
        <input type="hidden" name="veg" value="0" />
        <input type="checkbox" id="" value="11" name="veg"></td>
 </tr
<tr>
  <td>10</td>
  <td>Dry Fruits</td>
  <td></td>
  <td>
        <input type="hidden" name="dry" value="0" />
        <input type="checkbox" id="" value="12" name="dry"></td>
 </tr>
 <tr>
  <td>11</td>
  <td>Beef</td>
  <td></td>
```

```
 <td>
      <input type="hidden" name="beef" value="0" />
      <input type="checkbox" id="" value="13" name="beef"></td>
</tr>
<tr>
 <td>12</td>
 <td>Alcoholic and Non Alcoholic Beverages</td>
 <td></td>
 <td>
      <input type="hidden" name="alcohol" value="0" />
      <input type="checkbox" id="" value="14" name="alcohol"></td>
</tr>
 <tr>
 <td>13</td>
 <td>Legumes and Beans</td>
 <td></td>
 <td>
      <input type="hidden" name="legumes" value="0" />
      <input type="checkbox" id="" value="16" name="legumes"></td>
</tr>
<tr>
 <td>14</td>
 <td>Breads, Cookies and Rolls</td>
 <td></td>
 <td>
      <input type="hidden" name="breads" value="0" />
      <input type="checkbox" id="" value="18" name="breads"></td>
</tr>
<tr>
 <td>15</td>
 <td>Ice cream, Candies and Snacks</td>
 <td></td>
 <td>
      <input type="hidden" name="ice" value="0" />
```

```
        <input type="checkbox" id="" value="19" name="ice"></td>
  </tr>
  <tr>
   <td>16</td>
   <td>Rice, Pasta & Noodles</td>
   <td></td>
   <td>
        <input type="hidden" name="rice" value="0" />
        <input type="checkbox" id="" value="20" name="rice"></td>
  </tr>
  <tr>
   <td>17</td>
   <td>Fast Foods</td>
   <td></td>
   <td>
        <input type="hidden" name="fast" value="0" />
        <input type="checkbox" id="" value="21" name="fast"></td>
  </tr>
   <tr>
   <td>18</td>
   <td>Rolls</td>
   <td></td>
   <td>
        <input type="hidden" name="rolls" value="0" />
        <input type="checkbox" id="" value="22" name="rolls"></td>
  </tr>

   <tr>
    <td><input type="submit" class="btn btn-primary" name=""></td>
    <td></td>
    <td></td>
    <td></td>
   </tr>
 </tbody>
```

```
        </table>
          </form>
```

#limited food results page

```
<?php
$con=mysqli_connect("localhost","root","suman123","food");
// Check connection
if (mysqli_connect_errno())
{
echo "Failed to connect to MySQL: " . mysqli_connect_error();
}


$milk=$_POST['milk'];
 $spices=$_POST['spices'];
 $baby=$_POST['baby'];
 $salads=$_POST['salads'];
 $soup=$_POST['soup'];
 $cereals=$_POST['cereals'];
 $fruits=$_POST['fruits'];
 $pork=$_POST['pork'];
 $veg=$_POST['veg'];
 $dry=$_POST['dry'];
 $beef=$_POST['beef'];
 $alcohol=$_POST['alcohol'];
 $legumes=$_POST['legumes'];
 $breads=$_POST['breads'];
 $ice=$_POST['ice'];
 $rice=$_POST['rice'];
 $fast=$_POST['fast'];
 $rolls=$_POST['rolls'];
```

```
$result = mysqli_query($con,"select * from data_set where ROW in (select ROW
from cluster5) and (g_no=$milk or g_no=$spices or g_no=$baby or g_no=$salads or
g_no=$cereals or
g_no=$fruits or g_no=$pork or g_no=$veg or g_no=$dry or g_no=$beef or
g_no=$alcohol or g_no=$legumes or g_no=$breads or g_no=$ice or g_no=$rice or
g_no=$fast or g_no=$rolls)");

echo " <div class='container'>
     <div class='well'>

<table id='myTable'>
  <tr class='header'>
    <th style='width:40%;">Items</th>
    <th >Carbo</th>
    <th >Energy</th>
    <th >Protien</th>
    <th >Fiber</th>
    <th >VItamin</th>
    <th >Thiamin</th>
    <th >Vitamin C</th>

  </tr>
";
while($row = mysqli_fetch_array($result))
{
echo "  <tbody>
            <tr>";
echo "<td>" . $row['ITEMS'] . "</td>";
echo "<td>" . $row['CARBOHYDRATES'] . "</td>";
echo "<td>" . $row['ENERGY'] . "</td>";
echo "<td>" . $row['PROTEIN'] . "</td>";
```

```
echo "<td>" . $row['FIBER'] . "</td>";
echo "<td>" . $row['VITAMIN_E'] . "</td>";
echo "<td>" . $row['THIAMIN'] . "</td>";
echo "<td>" . $row['VITAMIN_C'] . "</td>";
echo "</tr>";
}
echo " </tbody>
        </table>
        </div>
        </div>";
mysqli_close($con);
?>

#filtering in result page

<script>
function myFunction() {
  var input, filter, table, tr, td, i;
  input = document.getElementById("myInput");
  filter = input.value.toUpperCase();
  table = document.getElementById("myTable");
  tr = table.getElementsByTagName("tr");
  for (i = 0; i < tr.length; i++) {
    td = tr[i].getElementsByTagName("td")[0];
    if (td) {
      if (td.innerHTML.toUpperCase().indexOf(filter) > -1) {
        tr[i].style.display = "";
      } else {
        tr[i].style.display = "none";
      }
    }
  }
}
</script>
```

8. REFERENCES

1. Joshua M. Krbez,Adnan Shaout, Fuzzy nutrition system, IJIRCE(2013)
2. Akshi kumar, Pulkit tanwar, Saurabh nigam, Survey and evaluation of food recommendation systems and techniques, International Conference on Computing for Sustainable Global Development (3592-3596),2016.
3. NidalIsmael, Mahmoud Alzaalan and WesamAshour, International Journal of Artificial Intelligence and Applications for Smart Devices, 2014
4. Balqees Al-Thuhli,Halima Al-Alawi , International Business Information Management Association Conference, Volume: 2013
5. K. Mallikharjuna Rao, N. Balakrishna, N. Arlappa , A. Laxmaiah and G.N.V. Brahmam, Diet and Nutritional Status of Women in India,Kamala-Raj 2010.
6. Maiyaporn Phanich, Phathrajarin Pholkul, and Suphakant Phimoltares, Food Recommendation System Using Clustering Analysis for Diabetic Patients, IEEE ,2010
7. P. Viswanath, Rajwala Pinkesh ,*l*-DBSCAN : A Fast Hybrid Density Based Clustering Method,IEEE,2016.
8. Xiuhua Liang, Guangming Deng, Bin Yan, Fruit and VegetableNutritionValue Assessment and Replacement Based on the Principal ComponentAnalysis and Cluster Analysis, Applied Mathematics, 1620-1629,2015.
9. Yasir Mehmood, Mudassar Abbas, Xi Chen, and Timo Honkela, Self-Organizing Maps of Nutrition, Lifestyle and Health Situation in the World, springer, 2011.
10. László Kovács, László Bednarik, Parameter Optimization for BIRCH PreClustering Algorithm, International Symposium on Computational Intelligence and Informatics(2011).
11. Anil K.Jain, Richard C. Dubes , Algortihms for clustering data, Prentice-Hall, Inc. Upper Saddle River, NJ, USA ©1988
12. A.K. Jain, M. N. Murty , P. J. Flynn, ACM Computing Surveys (CSUR), Volume 31 Issue 3, Sept. 1999
13. Hui Zou, Trevor Hastie & Robert Tibshirani, Sparse principal component analysis,Journal of computational and graphical statistics,Volume 15,2006

14. Park MH., Hong JH., Cho SB. (2007) Location-Based Recommendation System Using Bayesian User's Preference Model in Mobile Devices. In: Indulska J., Ma J., Yang L.T., Ungerer T., Cao J. (eds) Ubiquitous Intelligence and Computing. UIC 2007. Lecture Notes in Computer Science, vol 4611. Springer, Berlin, Heidelberg
15. Burke R. (2007) Hybrid Web Recommender Systems. In: Brusilovsky P., Kobsa A., Nejdl W. (eds) The Adaptive Web. Lecture Notes in Computer Science, vol 4321. Springer
16. Shiva Nadi, Mohammad Hossein Saraee and Ayoub Bagheri," A Hybrid Recommender System for Dynamic Web Users", *International Journal Multimedia and Image Processing(IJMIP),* vol. 1, no. 1, 2011
17. D. Riecken," Personalized Views of Personalization. *"Communications of the ACM,* vol. 43, no. 8, pp.27 -28 2000
18. Chi-Chun Lo, Ding-Yuan Cheng, Chi-Hua Chen. "A Semantic Web Methodology for Situation-Aware Curative Food Service Recommendation System", .*Computer Science and Software Engineering, 2008 International Conference* on, pp.444-447, 12-14 Dec. 2008.
19. Yoosoo Oh, Ahyoung Choi, Woontack Woo. "U-BabSang: a contextaware food recommendation system", *J. Supercomput*, 2010, vol. 54, no. 1, pp. 61-81.
20. Jong-Hun Kim, Jung-Hyun Lee, Jee-Song Park, Young-Ho Lee, KeeWook Rim. "Design of Diet Recommendation System for Healthcare Service Based on User Information", *Fourth International Conference on Computer Sciences and Convergence Information Technology, ICCIT'09*, pp. 516-518, 24-26 Nov. 2009.
21. Jill Freyne, Shlomo Berkovsky, "Intelligent food planning: personalized recipe recommendation*", Proceedings of the 15th international conference on Intelligent user interfaces*, February 07-10, 2010, Hong Kong, China
22. Wahidah Husain, Lee Jing Wei, Sooi Li Cheng and Nasriah Zakaria. "Application of data mining techniques in a personalized Diet recommendation system for cancer patients", *Published in Humanities, Science and Engineering (CHUSER), 2011 IEEE Colloquium*, 5-6 Dec. 2011, pp. 239 – 244.

23. Chang Shing Lee , M. H. Wang and H. Hagras, "A type-2 fuzzy ontology and its application to personal diabetic-diet recommendation", *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 2, pp.374 -395, February 2010
24. Achmad Arwan, Bayu Priyambadha, Riyanarto Sarno, Mohamad Sidiq, Heri Kristiantol, "Ontology and semantic matching for diabetic food recommendations", *Published in Information Technology and Electrical Engineering (ICITEE), 2013 International Conference*, Yogyakarta, 7-8 Oct. 2013, pp. 170 - 175.
25. Shilpa Dharkar, Anand Rajavat, "Performance analysis of healthy Diet recommendation System using Web data mining", *International Journal of Scientific & Engineering Research (IJSER) Volume 3*, Issue 5, May2012.
26. hreya B. Ahire, Harmeet kaur khanuja, "A personalized framework for health care recommendation", *published in Computing Communication Control and Automation (ICCUBEA), 2015 International Conference, Pune*, 26-27 Feb. 2015, pp. 442 – 445.
27. The Basic of a Healthy Diabetes Diet. http://diabetes.webmd.com/diabetes-diet-healthy-diet-basics
28. . Han, M. Kamber, and J. Pei, Data Mining Concepts and Techniques, Morgan Kaufmann, pp. 383-403, 2006.
29. C. S. Lee, M. H. Wang, W. C. Sun, and Y. C. Chang, "Intelligent Healthcare Agent for Food Recommendation at Tainan City," Proc. IEEE Int'l Conf Systems, Man, and Cybernetics, pp.1465-1470, October 2008.
30. J. B. MacQueen : "Some methods for classification and analysis of multivariate observations," Proc. 5th Berkeley Symp. Mathematical Statistics and Probability, vol. 1, pp.281-297, 1967.
31. J. P. Bantle, J. Wylie-Rosette, A. L. Albright, C. M. Apovian, N. G. Clark, M. J. Franz, B. J. Hoogwerf, A. H. Lichtenstein, E. Mayer-Davis, A. D. Mooradian, and M. L. Wheeler, "Nutrition recommendations and interventions for diabetes:a position statement of the american diabetes association," Diabetes care, vol. 31, pp. 61-78, January 2008.
32. Chang Shing Lee , M. H. Wang and H. Hagras, "A type-2 fuzzy ontology and its application to personal diabetic-diet recommendation", *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 2, pp.374 -395, February 2010.

33. https://sites.google.com/site/dataclusteringalgorithms/k-means-clustering-algorithm
34. https://en.wikipedia.org/wiki/K-means_clustering
35. https://en.wikipedia.org/wiki/BIRCH
36. https://en.wikipedia.org/wiki/Self-organizing_map
37. https://en.wikipedia.org/wiki/Principal_component_analysis
38. http://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html
39. http://scikit-learn.org/stable/modules/generated/sklearn.cluster.Birch.html
40. http://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html
41. https://www.datacamp.com/community/tutorials/machine-learning-python#gs.lb8P8kk
42. http://madhukaudantha.blogspot.in/2015/04/density-based-clustering-algorithm.html
43. https://www.w3schools.com/howto/howto_css_signup_form.asp
44. https://www.w3schools.com/howto/howto_css_login_form.asp
45. https://www.w3schools.com/php/php_forms.asp
46. https://www.w3schools.com/php/func_mysqli_connect.asp