International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 4
April 2018

Personalized Healthcare Management using R Programming

Nikita B. Kothari
Bharati Vidyapeeth Collegeof Engineering, Navi Mumbai.
Email id: kotharinikita0@gmail.com

Prof. Shankar M. Patil
Bharati Vidyapeeth Collegeof Engineering, Navi Mumbai
Email id: smpatil2k@gmail.com

ABSTRACT

Now a day's Electronic gadgets and fast foods have become an indispensable part of our daily lives. We eat unhealthy foods, live a sedentary life style, spent most of our time on electronic gadgets, lack of sleep, increase of stress and do almost no exercise which gives us a diseased body. These few life style related factors if not managed properly can lead us to obesity and diabetes at very younger age. This paper will focus on prevention of these diseases by using prediction. This prediction is done based on the regions and region included here is major part of Mumbai. The prediction will be done using classification technique. So, if the users of this system know their current health quotient then they can easily modify their lifestyle and eating habits thus, avoiding these diseases in future leading to a diseased free, healthy and long life.

KEYWORDS-Classification, healthy quotient, Naive Bayes, Diabetes, Obesity.

1.INTRODUCTION

Data mining [6] is the process of analysing the existing data and extracting useful information out of it. Sometimes data mining is used to predict the future based on the existing data. It finds the relation between the existing data and based on the relation it predicts the outcome of the remaining data. Data mining has been used to uncover patterns from the large amount of stored information and then used to build predictive models.

Diabetes[8] is a life-long disease that affects the way your body handles glucose, a kind of sugar, in your blood.

What Causes Diabetes?

Your pancreas makes a hormone called insulin. It's what lets your cells turn glucose from the food you eat into energy. People with type 2 diabetes make insulin, but their cells don't use it as well as they should. Doctors call this insulin resistance.

At first, the pancreas makes more insulin to try to get glucose into the cells. But eventually it can't keep up, and the sugar builds up in your blood instead.

Usually a combination of things causes type 2 diabetes, including:

Genes. Scientists have found different bits of DNA that affect how your body makes insulin.

Extra weight. Being overweight or obese can cause insulin resistance, especially if you carry your extra pounds around the middle. Now type 2 diabetes affects kids and teens as well as adults, mainly because of childhood obesity.

The existing systems [1,2,3,4,5] and devices monitor the health condition of a patient, using sensors or predict the chances of diseases based on reports and other medical parameters but there is lack of system that predicts the chances of Obesity and Diabetes in near future based on lifestyle and food habits of a person.This system will help users determine their chances of getting obesity or diabetes in future if they continue to follow the current lifestyle and food habits.This system will benefit common users and they can be aware beforehand and proper preventive measures can be taken.

"Data mining is purely data-driven; this feature is important in healthcare."

2.PROPOSED SYSTEM

The proposed system will overcome the limitations of the existing systems. This system will predict the health quotient of the user based on the lifestyle he/she lives i.e. how much healthy a person is at this moment and what are the diseases he/she can get in near future if the user continues to live the same type of life style. Some of the reasons for any user falling ill is the lifestyle the person lives, the food and the timing of having food, not doing any exercise, not taking enough sleep and not taking breaks from work for relaxing.

This system will also give suggestion to avoid those diseases. Suggestions of food to have and lifestyle to adapt in order to lead a healthy life. This system will also give suggestions of doctors if required. This system will also show its users recent advancements in the field of medical sciences.

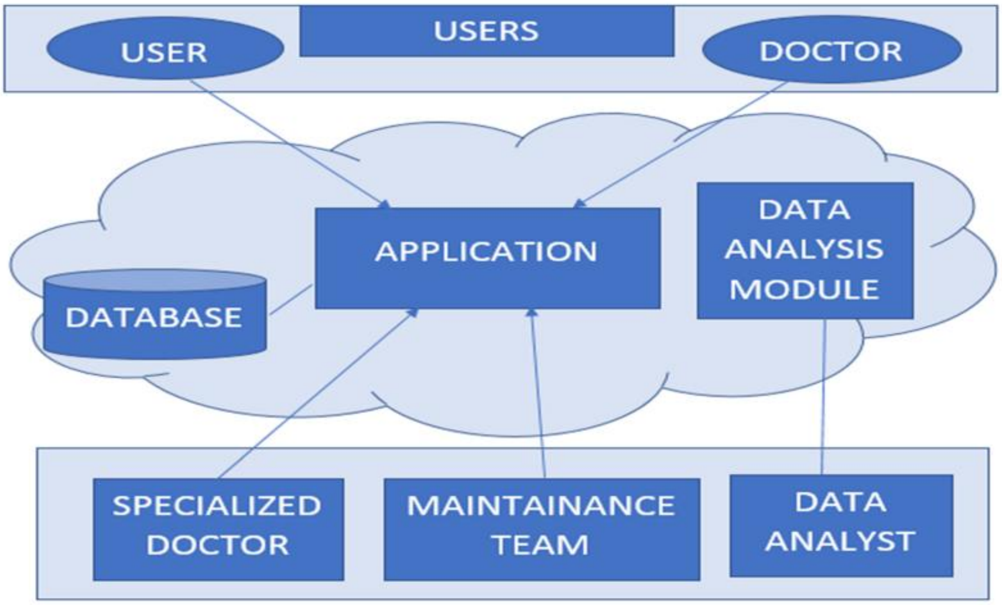


Fig1.Block Diagram of Proposed system.

Data Analysis Module

This module will consist of the prediction module designed in R. The model is designed using Naive Bayes classification Algorithm. The classification algorithm is applied on the data present in the database.

3.DATA COLLECTION.

This will include a set of questions asked to the user which includes age, sex, height, weight, questions related to meals like type of food user take in breakfast and dinner, no. of hours a person works, gap between works, no of hours sleep user takes, number of hours spent on electronic gadgets and exercise/ yoga related questions.

This questionnaire was made in google forms and shared on various social sites and the data was collected.

Questions included in the questionnaire:

Personal Information

Name, Age, Height, Weight, City, Profession, Gender, Email Id.

Lifestyle and Normal Routine Related Question: -

1. Do You Skip Any of You Meal?
   - Yes
   - No
2. If You Skip Meal, Then Which Meal You Skip Breakfast, Lunch, Evening Snacks, Dinner?
   - Breakfast
   - Lunch
   - Evening Snacks
   - Dinner
   - None
3. What Is the Time Gap Between You Any Two Meals?
   - 2 Hours
   - 3Hours
   - 4 Hours
   - 5 or more than 5 Hours
4. What Is the Time at Which You Have Breakfast?
5. What Is the Time at Which You Have Dinner?
6. What Food Do You Have for Breakfast?
   - Dropdown List of Food Items generally consumed in Mumbai region for breakfast.
7. What Food Do You Have in Dinner?
   - Dropdown List of Food Items generally consumed in Mumbai region for Dinner.
8. How Many Hours Do You Work in A Day?
9. How Many Hours You Spent on Electronic Gadget?
10. How Many Hours of Sleep You Take?
11. Do You Do Exercise or Yoga or Any Other Forms of Physical Activity (Dancing, Swimming)?
    - Yes
    - No
12. If Yes Then How Many Hours You Do Exercise or Yoga?

What Do You Do for Relaxing Yourself/Relieve Stress?

4.DATA MINING

4.1. DATA PREPROCESSING

- Data cleaning
  - It includes filling missing values, inconsistent data handling and noisy data handling.
  - Since proper care was taken while designing the google forms that is all fields important for data mining was made required and proper data type was selected

So, there is no noisy data or inconsistent data.

- Data integration :-Data from various sources that is Various social sites collected was integrated in a single excel sheet.
- Data transformation: -The data type of time was transformed as per excel sheet format.

Data reduction

- Data reduction

This was used for selecting the attributes for prediction.

E.g.: -For prediction of obesity only name, age height, weight and fat content were used.

- Dimensionality reduction — Remove unimportant attributes.

Only required Fields for prediction were selected for next step.

- concept hierarchy generation

Recursively reduce the data by collecting and replacing low level concepts (such as numeric values for age) by higher level concepts (such as young, middle-aged, or senior).

The implementation of this has been explained in section 3.3

4.2. UNSUPERVISED TO SUPERVISED DATA CONVERSION IN EXCEL

The data collected through survey was converted to an excel file and this data will be used for further analysis purpose. Based on the data collected some part of project supervised learning can be applied while other part has unsupervised learning so here we convert unsupervised attributes to supervised by assigning class labels to it. This is done using formulae (IF ELSE IN EXCEL) in excel.

FIG2.SkipMealClass

Skip Meal Class				
THRESHOLD VALUE	**NO_MEALS**	**BREAKFAST**	**LUNCH**	**DINNER**
class	NO	BAD_HIGH	BAD_MEDIUM	BAD_LOW
Points	4	1	2	3

FIG3.TimeGapClass

Time Gap		
THRESHOLD VALUE	**<=3 HOUR**	**>3 HOURS**
Class	Ideal	Not Ideal
Points	2	1

FIG4.DinnerTimeClass

DinnerTime		
THRESHOLD VALUE	**BEFORE 8 PM**	**AFTER 8 PM**
Class	IDEAL_DTIME	NOT_IDEAL_DTIME
Points	4	2

FIG5.BreakfastTimeClass

BreakfastTime		
THRESHOLD VALUE	**7AM-9AM**	**ANY OTHER TIME**
Class	IDEAL_BTIME	NOT_IDEAL_BTIME
Points	4	1

FIG6.WorkDayClass

WorkDay		
THRESHOLD VALUE	**<=8 HOURS**	**>8 HOURS**
Class	W_IDEAL	NOT_W_IDEAL
Points	2	1

FIG7.ElectronicClass

Electronic		
THRESHOLD VALUE	**<=3 HOURS**	**>3 HOURS**
Class	E_IDEAL	NOT_E_IDEAL
Points	4	1

FIG8.SleepTimeClass

SLEEPTIME		
THRESHOLD VALUE	6 HOURS	>6 HOURS
Class	IDEAL_SLEEP	NOT_IDEAL_SLEEP
Points	2	1

FIG9.WorkoutClass

WORKOUT		
Class	GOOD	BAD
Points	4	1

FIG10SugarLevelClass

SUGARLEVEL			
Class	HIGH		NORMAL
SUGARINTAKE GENDER WISE	MALE	FEMALE	
	SUGAR>37.9GM	SUGAR>25GM	
Points	1		5

FIG11.ObesityCLass

OBESITYCLASS				
class	UNDERFAT	IDEAL	OVERFAT	UNDERFAT
Points	4	3	2	1

FIG12.HealthQuotientClass

HEALTHQUOTIESTCLASS			
class	UNHEALTHY	MODERATELY_HEALTHY	HEALTHY
Points	1	3	5

FIG13.HealthQuotientCalculation

HEALTHQUOTIENTCALCULATION			
CLASS	HEALTHY	UNHEALTHY	MODERATELY_HEALTHY
POINTS	TOTAL>20	TOTAL<12	12<TOTAL<20

FIG14.DiabetesLevelCalculation

DIABETES LEVELCALCULATION			
DIABETES LEVELS	HIGH	MEDIUM	LOW
POINTS	TOTAL<=6	6<TOTAL<=10	TOTAL<10

5.IMPLEMENTATION

5.1 CLASSIFICATION

■ Classification algorithmspredicts categorical class labels (discrete or nominal) while prediction techniques are used for predicting continuous valued functions.

Classification—A Two-Step Process

Model construction: describing a set of predetermined classes

- Each tuple/sample is assumed to belong to a predefined class, as determined by the class label attribute
- The set of tuples used for model construction is training set
- The model is represented as classification rules, decision trees, or mathematical formulae

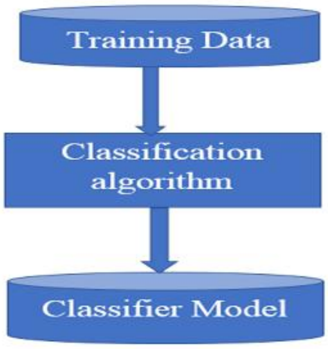


Fig.15.Model Construction

Model usage: for classifying future or unknown objects

- Estimate accuracy of the model
- The known label of test sample is compared with the classified result from the model
- Accuracy rate is the percentage of test set samples that are correctly classified by the model
- Test set is independent of training set, otherwise over-fitting will occur
- If the accuracy is acceptable, use the model to classify data tuples whose class labels are not known

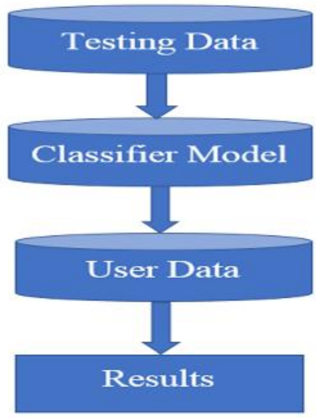


Fig.16.Model Usage

5.2. OBESITY PREDICTION.

Obesity Calculation Formulae

Metric BMI Formula

BMI = weight (kg) ÷ height^2 (m^2)

Class labels are assigned based on below table.

Fig.17. classification of obesity.

CLASSIFICATION	MEASUREMENT(kg/m^2)
Underweight	<18.50
Normal	18.50-24.90
Overweight	>=25.00
Obese	>=30.00

5.3. CALCULATION OF BODY FAT PERCENTAGE

Body fat percentage =1.2*BMI+.23*age-5.4-10.8*gender

*Gender: female=0 male=1

This is because body fat is 10% greater in women than in male.

Class labels are assigned based on below table

FIG.18. Body Fat Percentage Table

	Underfat	Ideal	Overfat	Obese
Women				
Age 20-39	<21%	21% to 33%	34% to 39%	>39%
Age 40-59	<23%	23% to34%	35% to 40%	>40%
Age60-79	<24%	24% to 35%	36% to 42%	>42%
Men				
Age 20-39	<8%	8% to 19%	20% to 25%	>25%
Age 40-59	<11%	11% to 21%	22% to 28%	>28%
Age60-79	<13%	13% to 24%	25% to 30%	>30%

5.4. CONTENT OF SUGAR TO BE CONSUMED EVERYDAY

According to the American Heart Association (AHA), the maximum amount of added sugars you should eat in a day are [7]:

Men: 150 calories per day (37.5 grams or 9 teaspoons).

Women: 100 calories per day (25 grams or 6 teaspoons).

5.5. NAIVE BAYES IMPLEMENTATION IN R STUDIO

5.5.1. OBESITY PREDICTION

Attribute Subset selection.

The pre-processed data from section 4.1 and using section 5.2 and 5.3 is used for the prediction.

The decision of attributes required was selected on the basis of the suggestions given by the dietitian Dr Pallavi (Sion Hospital, Mumbai, Maharashtra, India).

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 4
April 2018

A separate excel sheet is made of the attributes needed for the prediction.

	A	B	C	D	E
1	AGE	GENDER	HEIGHT	WEIGHT	FAT
2	20	FEMALE	5.3	50	IDEAL
3	22	FEMALE	5.3	45	UNDERFAT
4	20	FEMALE	5.5	70	IDEAL
5	20	FEMALE	5.4	42	UNDERFAT
6	20	FEMALE	5.4	56	IDEAL
7	20	FEMALE	5.8	52	UNDERFAT
8	20	FEMALE	4.1	55	OBESE
9	21	FEMALE	5.4	55	IDEAL

FIG19.Attributes Selected for Obesity Prediction

Steps for prediction of Obesity:

I. Import the excel sheet (FIG19) in R studio.
II. Create training and test datasets using the above data.
III. Model Construction(FIG.15): -Perform the Naïve Bayes Classification algorithm present in Caret Library of R studio.
IV. Plot the model using plot function of R studio.
V. Model Usage(FIG.16)-On the test dataset created in step II model created in step III will be applied. This will lead to Obesity prediction of test datasets.
VI. Then next step is to measure the accuracy of the above Classification algorithm applied. For this Confusion Matrix is made in r studio.
VII. Now, Once the model is ready and tested our new user data will be fed into another excel and it will be imported in the R studio. Then this new data will be Classified using the model created in step III.
VIII. The result of the model (Obesity Levels-Overfat, Underfat, Obese, Ideal) will be mailed to the respective user.
IX. The step I to Step VIII will be repeated for every new user so for convenience its made into a R script.

The accuracy of the algorithm varies as random sample of data are selected as training and test dataset each time.

5.5.2. DIABETES PREDICTION

The pre-processed data from section 4.1 and 4.2 is used for the prediction.

Attribute Subset selection.

The decision of attributes required was selected on the basis of the suggestions given by the dietitian Dr Pallavi (Sion Hospital, Mumbai, Maharashtra, India).

A separate excel sheet is made of the attributes needed for the prediction.

Initially the data received from user is converted into class labels(supervised) and then these labels are converted into points using section 4.2

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 4
April 2018

	A	B	C	D
1	SugarLevelPoints	HealthQuotientPoints	FatPoints	DIABETES
2	1	3	2	HIGH
3	1	3	4	MEDIUM
4	1	3	3	MEDIUM
5	5	3	3	MEDIUM
6	1	1	4	HIGH
7	5	3	3	MEDIUM
8	1	3	4	MEDIUM
9	5	3	1	LOW
10	5	3	3	MEDIUM

FIG20.Attributes selected For Diabetes Prediction

Steps for prediction of Diabetes:

I. Import the excel sheet (FIG19) in R studio.
II. Create training and test datasets using the above data.
III. Model Construction(FIG.15): -Perform the Naïve Bayes Classification algorithm present in Caret Library of R studio.
IV. Plot the model using plot function of R studio.
V. Model Usage(FIG.16)-On the test dataset created in step II model created in step III will be applied. This will lead to Diabetes prediction of test datasets.
VI. Then next step is to measure the accuracy of the above Classification algorithm applied. For this Confusion Matrix is made in R studio.
VII. Now, Once the model is ready and tested our new user data will be fed into another excel and it will be imported in the R studio. Then this new data will be Classified using the model created in step III.
VIII. The result of the model (Diabetes Chances-Low, Medium, High) will be mailed to the respective user.
IX. The step I to Step VIII will be repeated for every new user so for convenience it's made into a R script.

The accuracy of the algorithm varies as random sample of data are selected as training and test dataset each time.

6.RESULTS.

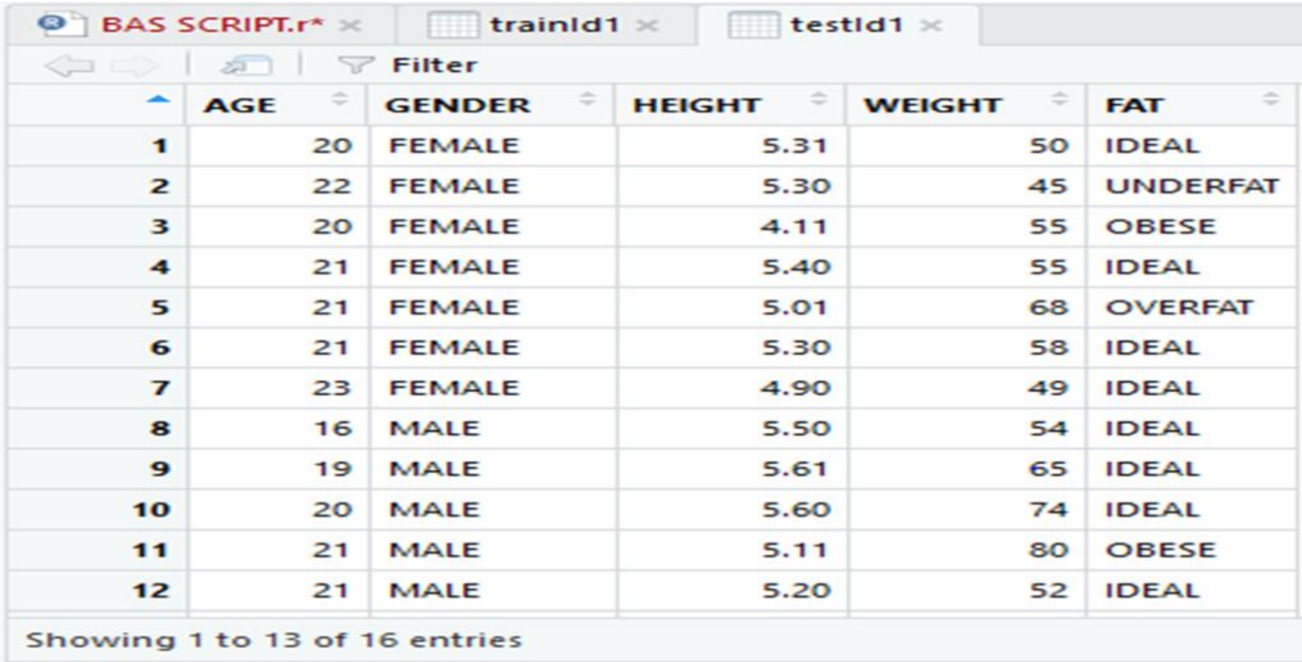
BAS SCRIPT.r* trainId1

Filter

	AGE	GENDER	HEIGHT	WEIGHT	FAT
1	20	FEMALE	5.500000	70.0	IDEAL
2	20	FEMALE	5.400000	42.0	UNDERFAT
3	20	FEMALE	5.400000	56.0	IDEAL
4	20	FEMALE	5.800000	52.0	UNDERFAT
5	21	FEMALE	5.200000	56.0	IDEAL
6	21	FEMALE	5.180000	50.0	IDEAL
7	21	FEMALE	5.800000	64.0	IDEAL
8	21	FEMALE	5.200000	56.0	IDEAL
9	21	FEMALE	5.100000	63.0	IDEAL
10	21	FEMALE	5.000000	50.0	IDEAL
11	21	FEMALE	5.200000	59.0	IDEAL
12	21	FEMALE	5.400000	54.0	IDEAL

Showing 1 to 13 of 48 entries

FIG21.Train Dataset for Obesity



BAS SCRIPT.r* | trainId1 | testId1

Filter

	AGE	GENDER	HEIGHT	WEIGHT	FAT
1	20	FEMALE	5.31	50	IDEAL
2	22	FEMALE	5.30	45	UNDERFAT
3	20	FEMALE	4.11	55	OBESE
4	21	FEMALE	5.40	55	IDEAL
5	21	FEMALE	5.01	68	OVERFAT
6	21	FEMALE	5.30	58	IDEAL
7	23	FEMALE	4.90	49	IDEAL
8	16	MALE	5.50	54	IDEAL
9	19	MALE	5.61	65	IDEAL
10	20	MALE	5.60	74	IDEAL
11	21	MALE	5.11	80	OBESE
12	21	MALE	5.20	52	IDEAL

Showing 1 to 13 of 16 entries

FIG22.Test Dataset for Obesity

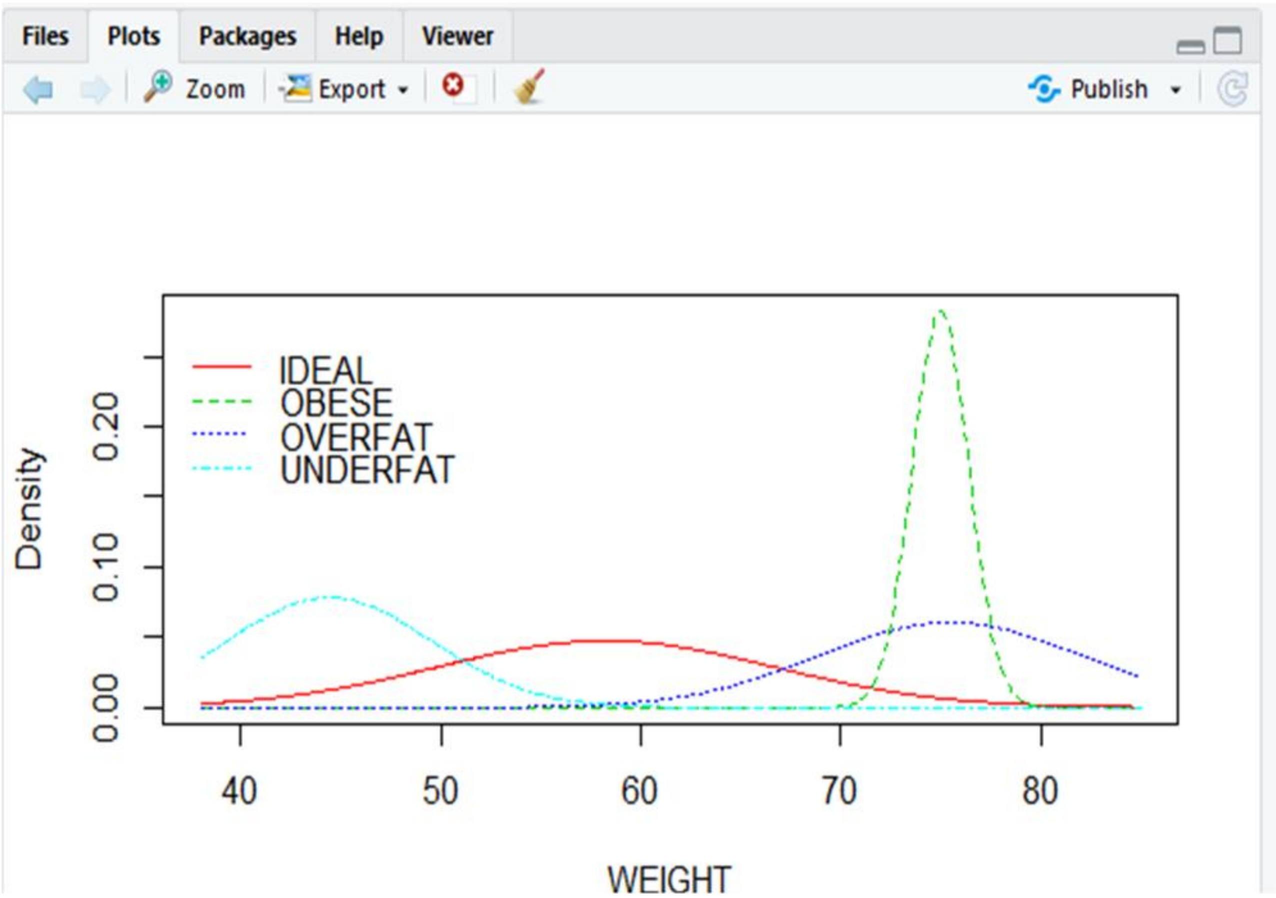


FIG23.Graph for Train Dataset of Obesity

IJETSR ISSN:2394-3386

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 4
April 2018

BAS SCRIPT.r* | train

Filter

1	UNDERFAT
2	UNDERFAT
3	UNDERFAT
4	IDEAL
5	IDEAL
6	IDEAL
7	IDEAL
8	IDEAL
9	IDEAL
10	IDEAL
11	OVERFAT
12	IDEAL
13	IDEAL

Showing 1 to 13 of 16 entries

FIG24.Prediction Result of Train Dataset for Obesity

BAS SCRIPT.r* | PREDICTION | PREDICT_OBESITY | trainId

Filter

	SugarLevelPoints	HealthQuotientPoints	FatPoints	DIABETES
1	1	3	2	HIGH
2	1	3	4	MEDIUM
3	1	3	3	MEDIUM
4	1	1	4	HIGH
5	5	3	3	MEDIUM
6	5	3	1	LOW
7	5	3	3	MEDIUM
8	1	3	3	MEDIUM
9	5	3	3	MEDIUM
10	5	5	3	MEDIUM
11	5	3	3	MEDIUM
12	5	3	3	MEDIUM

Showing 1 to 13 of 55 entries

FIG25.Train Dataset For Diabetes

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 4
April 2018

	SugarLevelPoints	HealthQuotientPoints	FatPoints	DIABETES
1	5	3	3	MEDIUM
2	1	3	4	MEDIUM
3	1	3	4	MEDIUM
4	5	3	3	MEDIUM
5	5	5	3	MEDIUM
6	5	3	3	MEDIUM
7	5	3	4	MEDIUM
8	5	3	3	MEDIUM
9	1	3	4	MEDIUM

Showing 1 to 9 of 9 entries

FIG26.Test Dataset for Diabetes

1	LOW
2	HIGH
3	HIGH
4	LOW
5	MEDIUM
6	LOW
7	LOW
8	LOW
9	HIGH

Showing 1 to 9 of 9 entries

FIG27.Prediction Result of Test Dataset for Diabetes

7.CONCLUSION.

This system will provide users the chances of Obesity and Diabetes (High, Medium, Low) in future. Thus, users can be aware beforehand only, which in turn will help users improve the quality of life.

For future work this model can be extended further by adding more features to website like adding doctors connect feature (For immediate support from doctors), personalized do's and don'ts list, online appointment booking, medical report analysis and online treatment. Also, along with the website an android and iOS application can be developed. This infrastructure can be hosted on cloud for providing better services.

8.REFERENCES.

1. D. West, ``How mobile devices are transforming healthcare," IssuesTechnol. Innov., vol. 18, no. 1, pp. 111, 2012.
2. A. Assaad and D. Fayek, ``General hospitals network models for the support of e-health applications," in Proc. IEEE/IFIP Netw. Oper. Manage.Symp. (NOMS), Apr. 2006, pp. 14.
3. Classification Algorithms of Data Mining, Indian Journal of Science and Technology, Vol 9(39), DOI: 10.17485/ijst/2016/v9i39/102065, October 2016
4. M. Aminian and H. R. Naji, ``A hospital healthcare monitoring system using wireless sensor networks," J. Health Med. Inf., vol. 4, p. 121, Feb. 2013.
5. Big data, big knowledge: big data for personalised healthcare: IEEE journal of biomedical and health informatics (volume 19, issue 4, July 2015.
6. Implementing WEKA for medical data classification and early disease prediction, 3rd IEEE International Conference on "Computational Intelligence and Communication Technology" (IEEE-CICT 2017)
7. https://www.healthline.com/nutrition/how-much-sugar-per-day#section3.
8. https://www.webmd.com/diabetes/type-2-diabetes#1
9. Data mining in healthcare: decision making and precision, Database Systems Journal vol. VI, no. 4/2015
10. Big Data Predictive Analysis: Using R Analytical Tool, International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud).
11. Predictive Analysis of Diabetic Patient Data Using Machine Learning and Hadoop, International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud).
12. Prediction on Diabetes Using Data Mining Approach, Pardha Repalli, Oklahoma State University.