DATA MINING

Cape Breton University

Submitted to: Prof. Dr Ebrahim Sharifi

Course Section: MGSC 5125 - 20

Team number – 03

PROJECT TITLE:

Estimation of Obesity Levels Based on Eating Habits and Physical Condition

GROUP MEMBERS

NAME	STUDENT ID	EMAIL ID
Chaitra Dudhiman Muni Mohan	0281012	cbu22cbzs@cbu.ca
Nikita Dhaval Tailor	0282701	cbu22cfwp@cbu.ca
Ruchitaben Patel	0282738	cbu22cfvr@cbu.ca
Prachi Dhimantkumar Gajjar	0285437	cbu22cljh@cbu.ca
HarinadhReddy Devarapalli	0277145	cbu22btld@cbu.ca

TABLE OF CONTENT

FIGURES:

TABLES:

1. ABSTRACT:

Obesity is a global health concern associated with various health issues. It has become very important to understand the various factors contributing to the obesity levels. This project approaches exploratory, predictive, and prescriptive analysis of estimating obesity level type based on various factors such as gender, weight, height, smoking habits, alcohol consumption, frequency of physical activities, number of main meals in a day, water consumption frequency, vegetables included in the food, etc. We have used supervised machine learning algorithm to develop a predictive model K-Nearest Neighbours (KNN) and evaluated the metrics such as accuracy, precision, recall (sensitivity), and F1 score. Also, through an unsupervised machine learning algorithm, we have developed K-means and analysed the clusters to classify individuals into distinct obesity levels. This descriptive, predictive, and prescriptive analysis gives us a better insight into understanding the relationship between eating habits, physical conditions, and obesity type, and the results of this project can offer personalized suggestions for individuals at risk of obesity and also for future research in the healthcare management system

Keywords: Obesity levels, exploratory, predictive, prescriptive, K-Nearest Neighbours (KNN), K-means

2. INTRODUCTION:

Excessive accumulation of body fat in individuals is known as Obesity. Various health issues such as cardiovascular diseases, diabetes, chronic diseases, and cancers are linked with obesity. Nowadays, obesity is commonly found in adults, adolescents, and children. Obesity has turned into an epidemic concern as it steadily increasing across the world. Genetics, lifestyle, and physical conditions are the main contributing factors to obesity. Obesity is posing a threat to the world and the healthcare system. Understanding the patterns leading to obesity is important for devising effective prevention and management strategies for the healthcare system.

In this project, we aim to understand the factors contributing to obesity levels and also to estimate the obesity levels based on eating habits and physical condition in individuals of Mexico, Peru, and Colombia through the three analytics domains i.e., exploratory, predictive, and prescriptive analysis. The results of this analysis will be important findings for public health interventions, healthcare policy formulation, and personalized wellness initiatives to prevent obesity-related complications and promote long-term health and well-being.

In addition, our project uses the flexibility of the R programming language to perform advanced statistical analyses, machine learning algorithms, and data visualization techniques. R is a powerful

tool for every step of the data science process, from preparing and enhancing data to training and evaluating models. This allows us to easily combine different approaches and draw valuable insights from the data.

The outcome of our project empowers policymakers, healthcare practitioners, and community leaders in Mexico, Peru, and Colombia by combining predictive analytics with expertise and stakeholder engagement. We provide evidence-based insights for designing targeted interventions and reducing obesity's impact. By working together and making data-driven decisions, we hope to bring about positive change and enhance health outcomes for those affected by obesity.

3. OBJECTIVE:

The main objective of our projects is to implement supervised and unsupervised learning techniques to explore and predict obesity levels based on their eating habits and physical conditions and also to give an optimal solution based on the analysis.

4. DATASET DESCRIPTION:

We have gathered the estimation of obesity level type levels based on their eating habits and physical conditions data set from the UCI Machine Learning repository.

The dataset consists of the lifestyle and eating habits of individuals from Peru, Mexico, and Colombia. The dataset has 17 attributes, 2111 records, and 7 class labels. The UCI machine learning repository collected 77 percent of the data from WEKA tools and SMOTE filters.

- Gender – male or female (Categorical)
- Age – age of the individual (Continuous)
- Weight - the weight of the individual (Continuous)
- Height – the height of the individual (Continuous)
- family_history_with_overweight – genetics of the family (Binary)
- FAVC – frequency of high-calorie food (Binary)
- FCVC – vegetables present in the food (Integer)
- NCP – number of main meals per day (Continuous)
- CAEC – any food or snack taken in between the main meals (Categorical)
- SMOKE – smoking patterns (Binary)
- CH20 – water intake per day (Continuous)
- SCC – calories monitored per day (Binary)

- FAF – frequency of physical activities (Continuous)
- TUE – time spent on the screen space (Integer)
- CALC – alcohol consumption patterns (Categorical)
- MTRANS – means of transport used (Categorical)
- NObeyesdad – obesity level type; Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II, Obesity Type III (Categorical)

5. DATA PRE-PROCESSING:

Data preprocessing is an important step in any of the data analysis.

It involves cleaning, transforming, and preparing raw data into a format such that the integrity and consistency of our data are maintained. In this step, we are pre-processing and cleaning our data to make it suitable for implementing our necessary models.

1. Loading the necessary packages.

   Figure 1: Showing loading of the necessary packages.

```
# Load libraries
library(factoextra)
library(cluster)
library(outliers)
library(class)
library(caret)
```

2. Loading our dataset

   Figure 2: Showing loading of the dataset

```
> # Read the dataset
> obesity_data <- read.csv("C:/Users/Vishal/Desktop/ObesityDataSet_raw_and_data_sinthetic .csv", stringsAsFactors = TRUE)
```

3. Checking for any missing values. Our dataset estimation of obesity based on eating habits and physical conditions has no missing values.

Figure 3: Checking for missing values.

```
> colSums(is.na(obesity_data))
                          Gender                              Age                           Height
                               0                                0                                0
                          Weight family_history_with_overweight                             FAVC
                               0                                0                                0
                            FCVC                              NCP                             CAEC
                               0                                0                                0
                           SMOKE                             CH2O                              SCC
                               0                                0                                0
                             FAF                              TUE                             CALC
                               0                                0                                0
                          MTRANS                       NObeyesdad
                               0                                0
```

4. The data set consists of numerical, continuous, categorical, ordinal, and binary data. Hence, to convert the categorical features into numeric codes we have used the factor method in R programming. Also, here we have assigned our class labels into factors.

Figure 4: Converting categorical to numeric using factor.

```
obesity_data$Gender = factor(obesity_data$Gender, levels = c('Female','Male'), labels = c(1,2))
obesity_data$family_history_with_overweight = factor(obesity_data$family_history_with_overweight,
                                                     levels = c('yes','no'), labels = c(1,0))

obesity_data$FAVC = factor(obesity_data$FAVC, levels = c('yes','no'), labels = c(1,0))
obesity_data$SMOKE = factor(obesity_data$SMOKE, levels = c('yes','no'), labels = c(1,0))
obesity_data$SCC = factor(obesity_data$SCC, levels = c('yes','no'), labels = c(1,0))

obesity_data$CAEC = factor(obesity_data$CAEC, levels = c('no', 'Sometimes','Frequently','Always'),
                           labels = c(0,1,2,3))

obesity_data$CALC = factor(obesity_data$CALC, levels = c('no', 'Sometimes','Frequently','Always'),
                           labels = c(0,1,2,3))

obesity_data$MTRANS = factor(obesity_data$MTRANS, levels = c('Public_Trans', 'Walking','Automobile','Motorbike','Bike'),
                             labels = c(1,2,3,4,5))
```

Also, here we have assigned our class labels (Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II, Obesity Type III) into factors.

Figure 5: Assigning our class labels to numeric codes in R

```
obesity_data$NObeyesdad = factor(obesity_data$NObeyesdad, levels = c('Insufficient_Weight', 'Normal_Weight',
                                                                     'Overweight_Level_I','Overweight_Level_II', 'Obesity_Type_I',
                                                                     'Obesity_Type_II','Obesity_Type_III'), labels = c(1,2,3,4,5,6,7))
```

5. We have used the round () function to round numerical values to a specified number of decimal places or the nearest integer.

Figure 6: Using the round () function

```
obesity_data$Age = as.numeric(format(round(obesity_data$Age, 0)))
obesity_data$FCVC = as.numeric(format(round(obesity_data$FCVC, 0)))
obesity_data$NCP = as.numeric(format(round(obesity_data$NCP, 0)))
obesity_data$CH2O = as.numeric(format(round(obesity_data$CH2O, 0)))
obesity_data$FAF = as.numeric(format(round(obesity_data$FAF, 0)))
obesity_data$TUE = as.numeric(format(round(obesity_data$TUE, 0)))
```

6. We have used 'outliers' to deal with the detection and manipulation of outliers

Figure 7: Using outliers

```
#Outlier Analysis
age_outlier <- grubbs.test(obesity_data$Age)
age_outlier

#Outlier Analysis
Height_outlier <- grubbs.test(obesity_data$Height)
Height_outlier

#Outlier Analysis
Weight_outlier <- grubbs.test(obesity_data$Weight)
Weight_outlier
```

6. TRAINING AND TESTING DATASET:

Split the obesity level type dataset into two subsets namely, the training set and the testing set to evaluate the models' performance. The training subset allows us to train our model and the testing subset will measure the performance. Our data was divided into ratios of 70% to 30% using set.seed(5000). Where 70% represent the training dataset while 30% represent the testing dataset.

Figure 8: Splitting dataset

```
#split data
#Setting seed

set.seed(5000)
ind <- sample(2, nrow(obesity_data), replace = T, prob = c(0.7, 0.3))
train <- obesity_data[ind == 1,]
test <- obesity_data[ind == 2,]
```

7. METHODOLOGY:

The methodology of our project is,

- **Descriptive analysis:** Our study aims to uncover practical information from the dataset, revealing trends and connections that could guide specific actions. Exploratory analysis for predicting the obesity level type helps in visualizing the data to investigate and understand the general relationships, patterns, and distribution of the data.

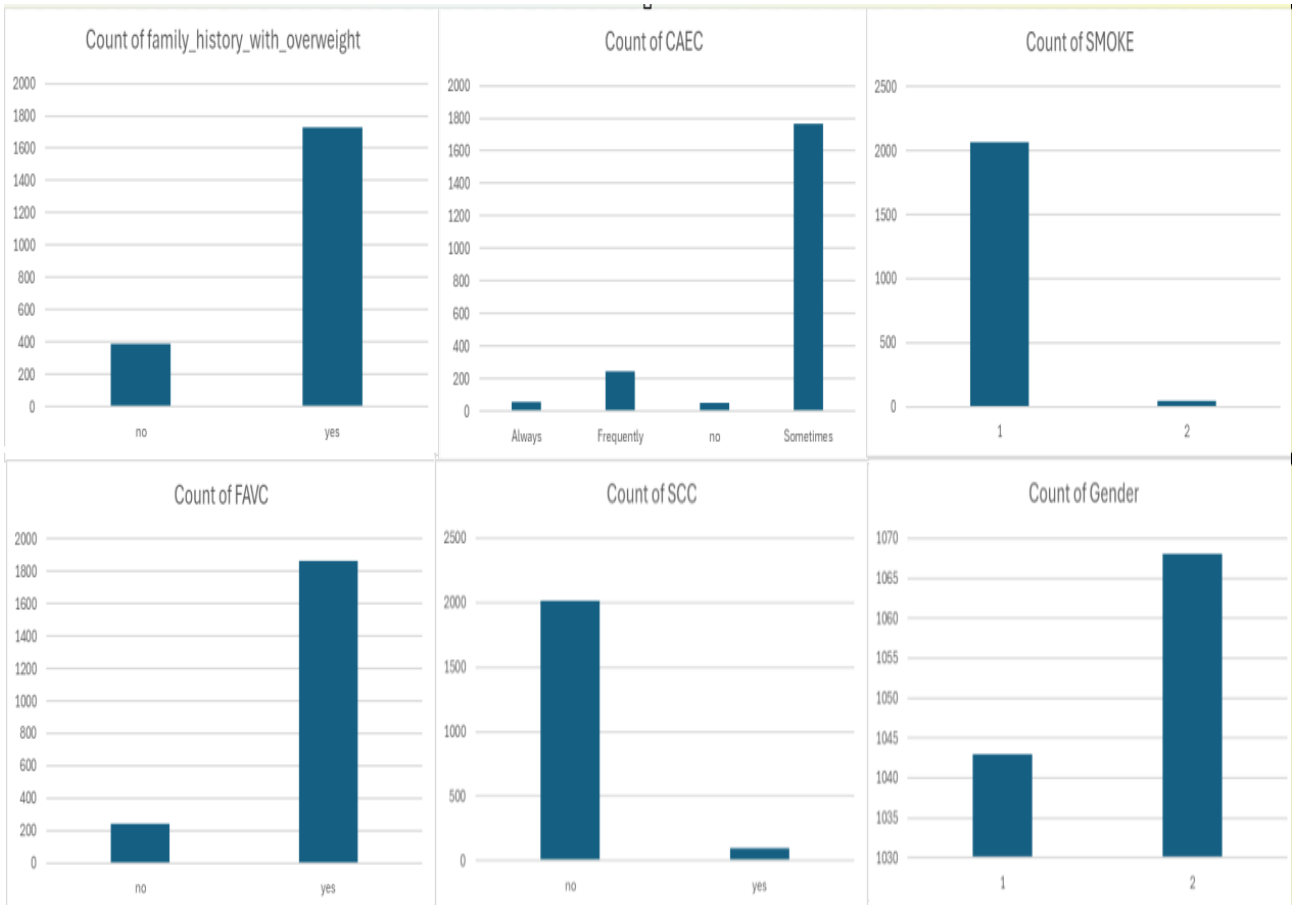
- **Predictive analysis:** Our focus is on predicting obesity rates using various factors like eating habits, exercise levels, personal data, and economic status through the supervised learning model K-Nearest Neighbours (KNN). By teaching these models with past data and testing their

accuracy through thorough validation, we hope to create reliable predictive models that can effectively pinpoint individuals at risk of obesity.

- **Prescriptive analysis:** Our project utilizes the unsupervised learning technique K-means clustering and dimensionality reduction to uncover hidden patterns in the data and identify different subgroups of individuals with similar obesity profiles. By grouping the data into clusters with shared characteristics, we hope to gain a better understanding of the diverse risk factors of obesity and develop targeted interventions. Through data analysis and visualizations using tools like Tableau, we aim to shed light on complex relationships and trends, providing insight into the underlying causes of obesity in various cultural settings.

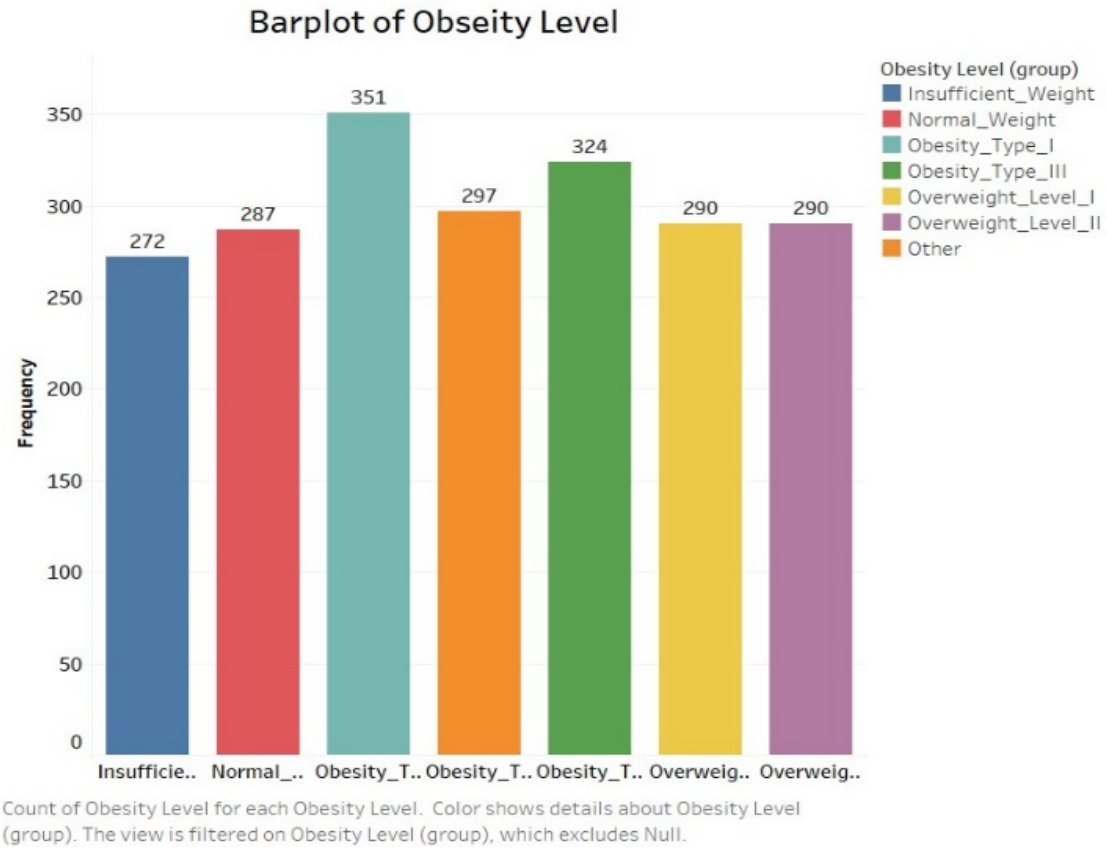
8. DESCRIPTIVE ANALYSIS:

Figure 9: Distribution of attributes



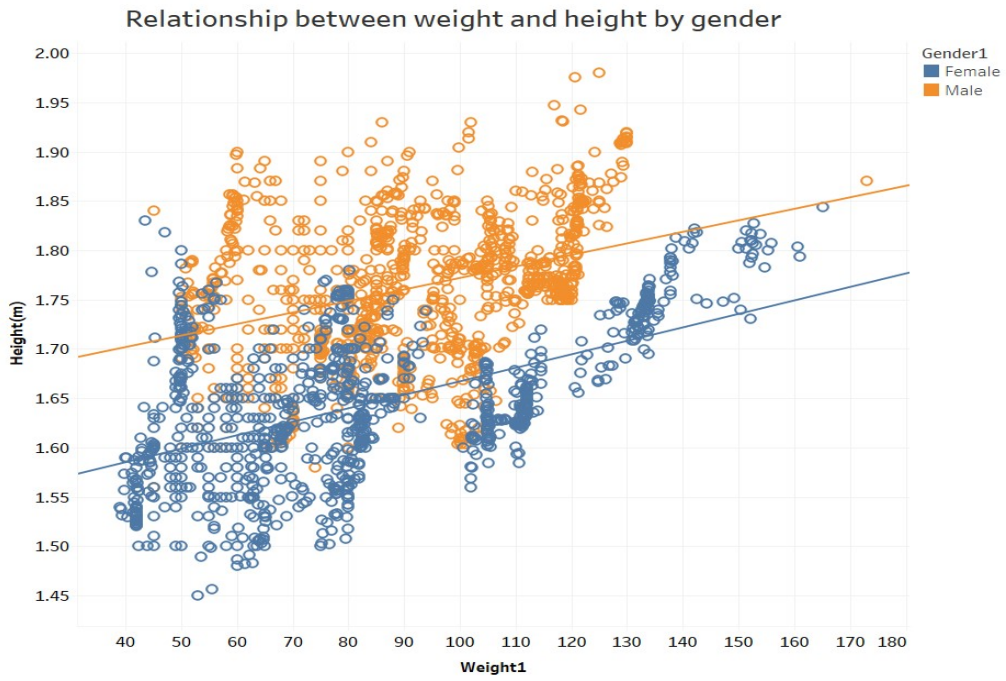
The above visualization is a bar graph representing the distribution of attributes. This bar graph represents different attributes of obesity to the frequency or distribution of each attribute. This visualization helps us to understand each attribute and its contribution towards obesity levels in a better way.

Figure 10: Obesity level bar graph



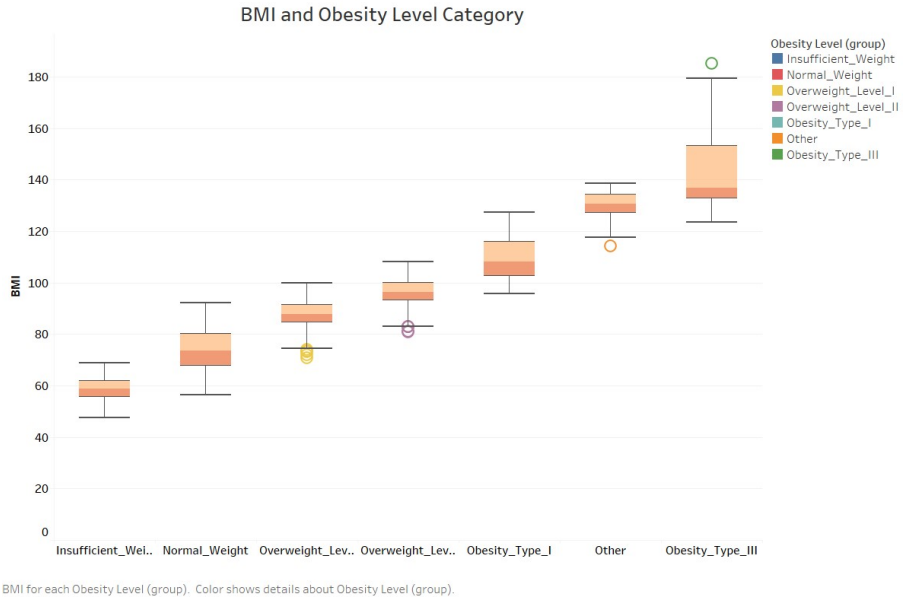
This bar chart gives us a visual representation of the number of individuals in each obesity level category. Type 1 obesity has the highest incidence and is more common in the population. The underweight type is the least common among the population.

Figure 11: Relation between weight and height by gender



The above figure represents the relationship between weight and height by gender. Both genders show an upward trend for weight and height. Also, the females may have a slightly faster rate of weight gain or height increase compared to males over time as the regression line is slightly steeper for females when compared to males. Also, the data points for weight in males are more clustered than the data points for weight in females.

Figure 12: Relation with BMI and Obesity Level Category



The above boxplot represents the relationship between the body mass index (BMI) and obesity level category. Body mass index (BMI) is calculated by, Body mass index (BMI) = (weight) / (height)*2 Here, the medians of each category are separated by similar intervals of around 10 kg/m².

Figure 13: Correlation Heatmap

The above figure is a graphical representation of the heatmap. Here, we have considered the attributes such as age, height, weight, FCVC, CH2O, FAF, and TUE to understand the correlation between the variables. Each cell in the heat map shows the correlation coefficient between two variables. The brighter colour red indicates a positive correlation whereas blue colour indicates a negative correlation. Thus, the colour intensity in the above heatmap shows the strength of the correlation.

9. PREDICTIVE ANALYSIS:

K-Nearest Neighbors (KNN)

KNN is a supervised learning algorithm used for classification and prediction. It operates based on the principle of similarity, assuming that similar data points tend to belong to the same class or have similar output values. It predicts the class of a data point by majority voting of its k nearest neighbors.

Steps involved in KNN:

1. Input the training dataset
2. Value k= 5 is set
3. Calculate the distance using the Euclidean formula.
4. Sort the distances in ascending order
5. Determine the majority class among the k nearest neighbors.
6. Assign the testing data point to the determined class.

CONFUSION MATRIX:

A confusion matrix is used to evaluate the performance of the model. Each row and column of the confusion matrix are represented by actual and predicted values.

Figure14: Confusion matrix and statistics of KNN

```
Confusion Matrix and Statistics

          actual
predicted   1   2   3   4   5   6   7
        1  71  15   0   0   0   0   0
        2   1  56   2   0   0   0   0
        3   0  11  72   3   0   0   0
        4   0   3  10  70   3   0   0
        5   0   0   2   5  95   0   0
        6   0   0   0   1   4  76   2
        7   0   0   0   0   1   4 101

Overall Statistics

               Accuracy : 0.8898
                 95% CI : (0.8622, 0.9136)
    No Information Rate : 0.1694
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.8712
```

In the above confusion matrix, each column in the confusion matrix represents the instances in the actual class and each row in the confusion matrix represents the instances in the predicted class.

From the above confusion matrix, we can say that the diagonal of the matrix represents the true positive for each obesity type whereas the off-diagonal represents the misclassification i.e71 Insufficient Weight, 56 Normal Weight, 72 Overweight Level I, 70 Overweight Level II, 95 Obesity Type I,76 Obesity Type II, 101 Obesity Type III were accurate classifications where the rest off-diagonal were misclassification.

Also, the overall accuracy was found to be 88.98 %.

EVALUATION METRICS:

Accuracy, precision, recall, and f1- score were evaluated from the confusion matrix to evaluate the performance of the K-Nearest Neighbours (KNN)

(i) Accuracy: (TP + TN) / (TP + TN + FP + FN)

(ii) Precision: TP / (TP + FP)

(iii) Recall (Sensitivity): TP / (TP + FN)

(iv) F1 Score: 2 * (Precision * Recall) / (Precision + Recall)

Table 1: Precision, recall, and F1 score of obesity type

CLASS LABEL	PRECISION	RECALL	F1 Score
Insufficient Weight	98.61	82.56	89.87
Normal Weight	65.88	94.92	77.78
Overweight Level I	83.72	83.72	83.72
Overweight Level II	88.60	81.40	84.85
Obesity Type I	92.23	93.14	92.69
Obesity Type II	95.00	91.57	93.25
Obesity Type III	98.06	95.28	96.65

Figure15: Precision, recall, and F1 score of obesity type (KNN)

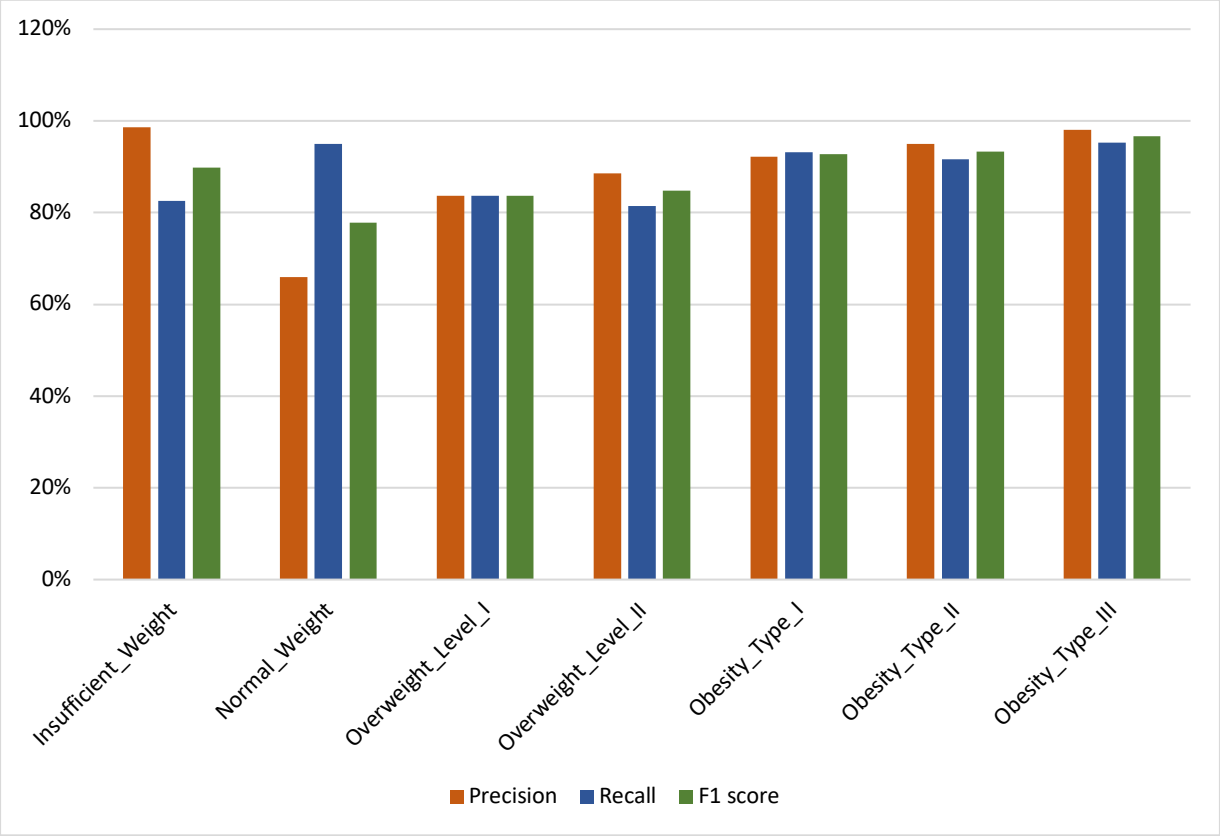
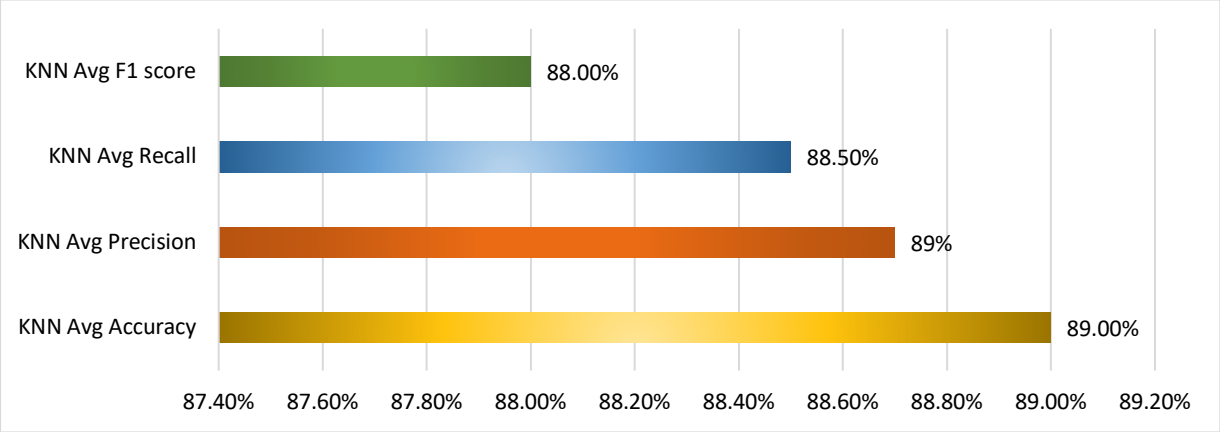


Figure16: Average accuracy, precision, recall, and F1 score of obesity type (KNN)



The overall average accuracy was 88%, precision was 89%, recall was 88.50% and F1 score was found to be 88.0%.

10. OBESITY LEVEL PREDICTION FOR NEW DATA USING KNN:

We have applied our dataset for predicting obesity levels for new data points. We used K=5 to predict the obesity of new users.

This involves creating a new data frame with numerical values representing features relevant to obesity prediction. The KNN model, trained on existing data, is then applied to this new data to predict the obesity level of individuals based on their feature values.

Figure17: Obesity level prediction for new data using KNN

```
# Create a new data frame with new numeric values for Prediction
new_data <- data.frame(
  Gender= c(2),
  Age=c(28),
  Height=c(1.8),
  Weight=c(140),
  family_history_with_overweight=c(1),
  FAVC=c(1),
  FCVC=c(1),
  NCP = c(3),
  CAEC = c(1),
  SMOKE = c(1),
  CH2O = c(1),
  SCC = c(0),
  FAF = c(1),
  TUE = c(0),
  CALC = c(2),  # How often do you drink alcohol?
  MTRANS=c(2),
```

Obesity Level Prediction using KNN

```
+ }
[1] "New Person is Obesity_Type_III"
>
```

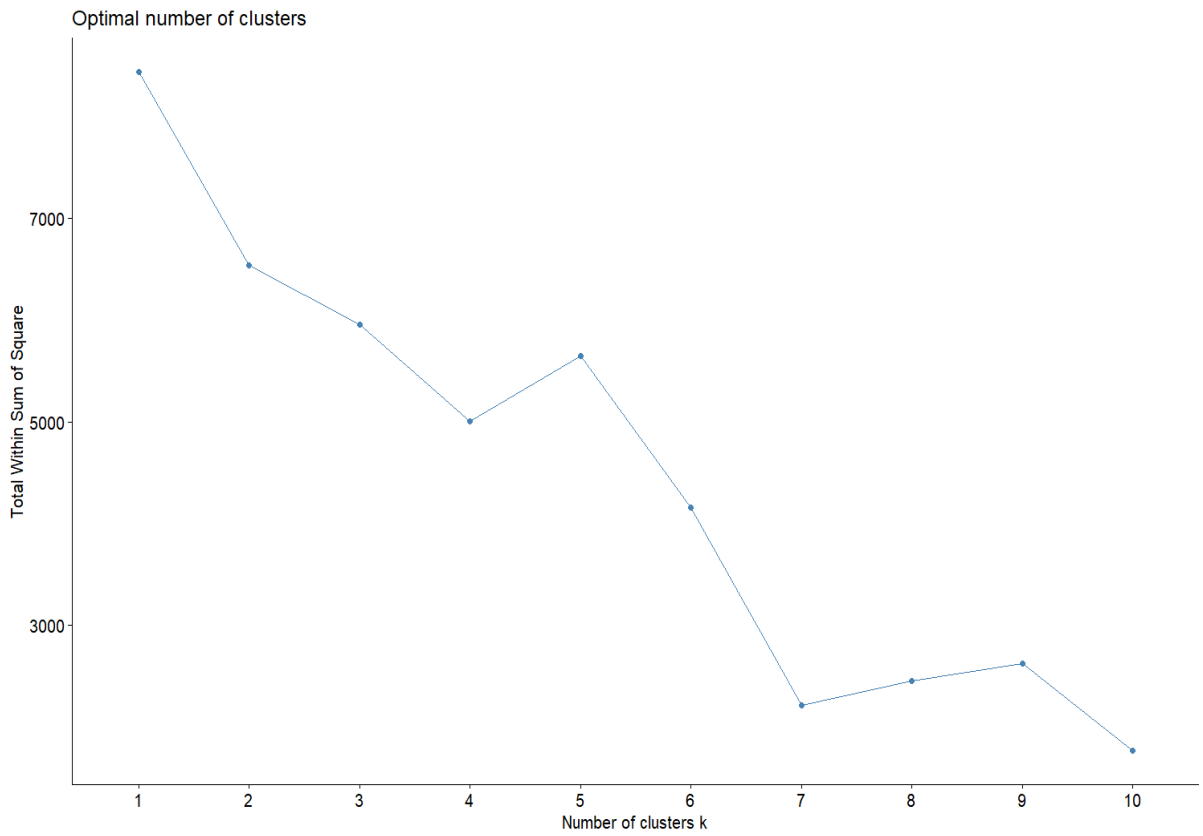
a

11. K-MEANS CLUSTERING:

K-means is an unsupervised learning algorithm used for clustering similar data points into groups. It helps identify patterns and segments within the data.

We have used K-Means Clustering in our analysis to cluster individuals based on their eating habits and physical conditions. We have used our analysis to cluster based on attributes such as eating habits (NCP, FAVC, FCVC, CALC).

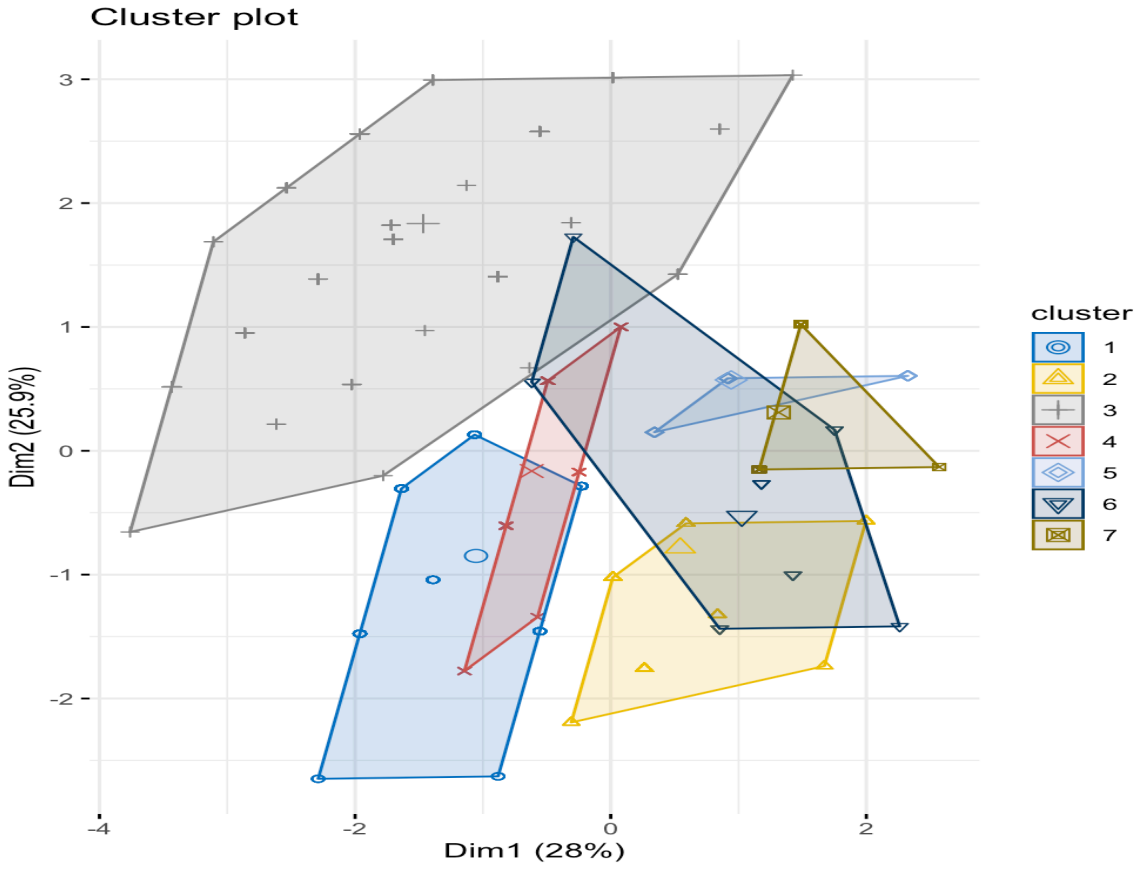
Figure18: Optimal number of clusters



From the graph, we found optimal clusters=7 for our dataset. Any new data points can be passed to predict which cluster they belong to.

12. K-MEANS RESULT

Figure19: K- means the result



13. CLUSTER ANALYSIS:

We have calculated the aggregate means of the clusters for analysis considering the attributes NCP, FAVC, FCVC, and CALC

Table 2: Cluster analysis for attributes NCP, FAVC, FCVC, and CALC

Cluster	NCP	FAVC	FCVC	CALC
1	0.426721	-0.36226	0.987325	0.592505
2	0.482374	-0.36226	-0.94795	0.617473
3	-1.71966	-0.36226	-0.84205	0.027251
4	-1.73169	2.759116	-0.08305	-0.48354
5	0.535851	2.759116	0.135552	-0.17691
6	0.352737	-0.36226	-0.0965	-1.41884
7	-1.92462	-0.36226	0.987325	-0.13489

Cluster 1: People in cluster 1 have a moderate consumption (42%) of main meals, (36%) lower consumption of high-calorie food, (98%) higher consumption of vegetables frequently, and a moderate alcohol intake (59%).

Cluster 2: People in cluster 2 have a moderate consumption (48%) of main meals, (36%) lower consumption of high-calorie food, (94%) higher consumption of vegetables frequently, and a moderate alcohol intake (62%).

Cluster 3: People in cluster 3 have a lower intake of main meals, lower consumption of high-calorie food, lower consumption of vegetables and consume alcohol.

Cluster 4: People in cluster 4 have an irregular diet with lower consumption of main meals, high intake of high-calorie food and less vegetables, and consume alcohol sometimes.

Cluster 5: People in cluster 5 have a good meal with higher consumption of high-calorie food, consume vegetables, and less alcohol intake.

Cluster 6: People in this cluster are characterized by a moderate consumption (35%) of main meals, indicating a balanced intake. However, they have a lower consumption of high-calorie foods (36%) and vegetables (10%). Their alcohol intake is relatively low (-14.2%)."

Cluster 7: People in Cluster 7 exhibit a low consumption (-19.2%) of main meals, indicating a reduced intake. They also show a lower consumption of high-calorie foods (36%) and fruits and vegetables (10%). Their alcohol intake is relatively low (-13%).

14. CLUSTER ASSIGNMENT TO NEW DATA:

Using the K-means algorithm we are assigning newly acquired data points to existing clusters. This assignment is typically based on the similarity or distance between the new data points and the centroids of the clusters. By determining which cluster a new data point belongs to, this method enables the grouping and analysis of new data based on patterns and characteristics observed in the original dataset's clusters. Based on Euclidean distance, we are determining which cluster center is closest to the new data point, aiding in its assignment to the appropriate cluster.

Figure 20: Cluster assignment to new data

```
# Create a new data frame with new numeric values
new_data <- data.frame(
  NCP = c(3),  # How many main meals do you have daily?
  FAVC = c(1),  # Do you eat high caloric food frequently?
  FCVC = c(0),  # Do you usually eat vegetables in your meals?
  CALC = c(1)  # How often do you drink alcohol?
)
```

Cluster Assignment to New User

```
> print("Cluster Assigned to New User: ")
[1] "Cluster Assigned to New User: "
> print(predicted_clusters)
[1] 5
> #Recommendations
```

Diet Plan Recommendations

```
> print("Diet Recommendation for New User: ")
[1] "Diet Recommendation for New User: "
> print(recommendation)
[1] "Limit number of main meals to moderate, reduced consumption of high calorie food, frequently consume vegetables."
>
```

15. PRESCRIPTIVE ANALYTICS & RECOMMENDATIONS:

Based, on our analysis and predictions, we have created diet recommendations for individuals according to the cluster they fall into. Diet plans are recommended based on cluster characteristics, providing personalized diet recommendations for each cluster.

- Cluster 1: Limit the number of main meals to moderate, reduce consumption of high-calorie food, and frequently consume vegetables.
- Cluster 2: Consume 2 main meals a day, reduce high-calorie intake, reduce alcohol intake
- Cluster 3: Lower the number of main meals in a day, high-calorie intake of food and alcohol.
- Cluster 4: Lower the number of meals in a day, eat vegetables, and consume less alcohol.
- Cluster 5: Eat a balanced 2 main meals a day considering calorie-rich food and including vegetables and can consume alcohol moderately
- Cluster 6: Moderate the number of meals in a day, reducing calories and increasing vegetable intake as well as lowering alcohol consumption.
- Cluster 7: Lower the frequency of meals in a day, lower the high-calorie intake of food and not consume alcohol

16. CONCLUSION:

Based on the exploratory analysis we have understood the distribution, relationship, and patterns of the attributes of obesity levels across the Peru, Mexico, and Columbia regions. The visualization of the attributes through exploratory analysis gives us a better insight into the lifestyle choices and physical conditions of the individuals in Peru, Mexico, and Columbia regions.

Further, in the predictive analysis, we evaluated the performance of the model K-Nearest Neighbours (KNN) and found the overall average accuracy to be 88%, precision 89%, recall 88.50%, and F1 score to be 88.0%. Through prescriptive analysis synthesizes evidence-based strategies into actionable recommendations, charting a course toward obesity prevention and intervention.

From the above descriptive, predictive, and prescriptive analysis, researchers and healthcare professionals can gain a better understanding of the demographics, obesity-causing habits, and lifestyles of individuals. These insights empower policymakers and healthcare providers to allocate resources effectively and implement preventive measures tailored to high-risk populations. This understanding can help them develop prevention and intervention strategies aimed at curbing obesity rates and promoting healthier communities.

17. REFERENCES:

1. Estimation of Obesity Levels Based On Eating Habits and Physical Condition. (2019). UCI Machine Learning Repository.
   Retrieved from: https://doi.org/10.24432/C5H31Z.

2. Asma Alqahtani, A., Albuainin, F., Alrayes, R., Al Muhanna, N., Alyahyan, E., & Aldahasi, E. (2021). Obesity Level Prediction Based on Data Mining Techniques. IJCSNS International Journal of Computer Science and Network Security, 21(3), 103.
   https://doi.org/10.22937/IJCSNS.2021.21.3.14

3. Faria Ferdowsy, K. S. (2021). A machine learning approach for obesity risk prediction,. *Current Research in Behavioral Sciences,*. Retrieved from
   https://www.sciencedirect.com/science/article/pii/S2666518221000401#bib0003

4. Rodolfo Cañas Cervantes, U. M. (2020,100472, ISSN 2352-9148). Estimation of obesity levels based on computational intelligence, *Informatics in Medicine Unlocked,, Volume 21,*.
   Retrieved from: (https://www.sciencedirect.com/science/article/pii/S2352914820306225)

5. Rifat Hossain, S.M. Hasan Mahmud, Md Altab Hossin, Sheak Rashed Haider Noori, Hosney Jahan,
   PRMT: *Predicting Risk Factor of Obesity among Middle-Aged People Using Data Mining Techniques*,Procedia Computer Science,Volume 132,2018,Pages 1068-1076,ISSN 1877-0509
   Retrieved from :https://doi.org/10.1016/j.procs.2018.05.022.
   https://www.sciencedirect.com/science/article/pii/S1877050918307543)

6. Garba, Salisu & Abdullahi, Marzuk & Wurnor, Nura. (2022). Sule Lamido University Journal of Science and Technology (SLUJST) Vol. 3 No. 1&2 [June, 2022], pp. 113-*121113Obesity Level ClassificationBased on Decision Tree and Naïve Bayes Classifiers.* SLU Journal of Science and Technology. 3. 113-121.
   Retrieved from:
   https://www.researchgate.net/publication/362044292_Sule_Lamido_University_Journal_of_Science_and_Technology_SLUJST_Vol_3_No_12_June_2022_pp_113-121113Obesity_Level_ClassificationBased_on_Decision_Tree_and_Naive_Bayes_Classifiers

7. Dugan, T. M., Mukhopadhyay, S., Carroll, A., & Downs, S. (2015). *Machine learning techniques for prediction of early childhood obesity.* Applied Clinical Informatics, 6(3), 506–520. https://doi.org/10.4338/ACI-2015-03-RA-0036

8. Daud, N., Mohd Noor, N. L., Aljunid, S. A., Noordin, N., & Fahmi Teng, N. I. M. (2019). Predictive Analytics: *The Application of J48 Algorithm on Grocery Data to Predict Obesity*. In 2018 IEEE Conference on Big Data and Analytics, ICBDA 2018 (pp. 1–6). Retrieved from: https://doi.org/10.1109/ICBDAA.2018.8629623

9. Yagin, F.H.; Gülü, M.; Gormez, Y.; Castañeda-Babarro, A.; Colak, C.; Greco, G.; Fischetti, F.; Cataldi, S. Estimation of Obesity Levels with a Trained Neural Network Approach optimized by the Bayesian Technique. *Appl. Sci.* **2023**, *13*, 3875. https://doi.org/10.3390/app13063875

10. De la Hoz Manotas, Alexis & De la Hoz Correa, Eduardo & Mendoza, Fabio & Morales, Roberto & Sanchez, Beatriz. (2019). Obesity Level Estimation Software based on Decision Trees. Journal of Computer Science. 15. 10. 10.3844/jcssp.2019.67.77.

11. Mondal PK, Foysal KH, Norman BA, Gittner LS. Predicting Childhood Obesity Based on Single and Multiple Well-Child Visit Data Using Machine Learning Classifiers. Sensors (Basel). 2023 Jan 9;23(2):759. doi: 10.3390/s23020759. PMID: 36679555; PMCID: PMC9865403.

12. C Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (The Morgan Kaufmann Series in Data Management Systems). 31, 371. http://www.amazon.com/Data-Mining-Techniques-ImplementationsManagement/dp/1558605525

13. Safaei M, Sundararajan EA, Driss M, Boulila W, Shapi'i A. A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity. Comput Biol Med. 2021 Sep;136:104754. doi: 10.1016/j.compbiomed.2021.104754. Epub 2021 Aug 16. PMID: 34426171.

14. Chatterjee A, Gerdes MW, Martinez SG. Identification of Risk Factors Associated with Obesity and Overweight-A Machine Learning Overview. Sensors (Basel). 2020 May 11;20(9):2734. doi: 10.3390/s20092734. PMID: 32403349; PMCID: PMC7248873.

15. Maswadi, K., Ghani, N. A., Hamid, S., & Rasheed, M. B. (2021). Human activity classification using Decision Tree and Naïve Bayes classifiers. Multimedia Tools and Applications, 80(14), 21709– 21726. https://doi.org/10.1007/s11042-020-10447-x

16. Thamrin, S. A., Arsyad, D. S., Kuswanto, H., Lawi, A., & Nasir, S. (2021). Predicting Obesity in Adults Using Machine Learning Techniques: An Analysis of Indonesian Basic Health Research 2018. Frontiers in Nutrition, 8(June), 1–15. https://doi.org/10.3389/fnut.2021.669155

17. Chang, R.L. and T. Pavlidis, 1977. Fuzzy decision tree algorithms. IEEE Trans. Syst. Man Cybernet., 7: 28-35. DOI: 10.1109/TSMC.1977.4309586

18. Nadkarni, S. and P.P. Shenoy, 2001. A Bayesian network approach to making inferences in causal maps. Eur. J. Operat. Res., 128: 479-498. DOI: 10.1016/S0377-2217(99)00368-9