School of Computer Science and Engineering

OBESITY LEVEL PREDICTION

DATA VISUALIZATION

(CSE3020)

J-COMPONENT REPORT

Project Guide:

Joshan Athanesious J

Submitted By:

Anu Cyril Saju - 20BCE1923

Shrikumaran PA - 20BCE1082

Renil Augustine Baby - 20BCE1826

Abstract:

In modern times, obesity has become a significant threat all over the world. People are constantly moving towards an unhealthy lifestyle, eating excessive junk food, late-night sleep, and spending a long time sitting down. Adolescents are being affected because of their unconscious attitudes. It is a medical problem known as a very complex disease. It promotes the spread of complex illnesses, stroke, heart disease, liver cancer,etc. Obesity is not only about food genetics, and environmental factors can be the cause of obesity. In the future, it can be a threat to the world as it is a worldwide health concern. The main concern of this paper is to analyze people for obesity and make them aware of the obesity risk factor. This paper aims to predict the obesity level of an individual considering various factors. For this analysis, we use a dataset which contains 17 attributes, and we visualize the impact of each attribute on obesity and finally identify the obesity level. The objective of the project is to estimate obesity levels based on eating habits and physical activity. From doing so we can learn a lot about how people can change their lifestyle so that they can lead a healthy life. Mostly we see what their eating habits are, whether they consume water and what amount, whether they smoke or drink alcohol, consumption of vegetables, etc. To identify the relation between various factors and the obesity level of each individual we used machine learning algorithms like Naïve Bayes, logistic regression, random forest and decision tree. Among these algorithms, random forest gave the highest accuracy of 96.19%. So using random forest we are able to identify the obesity level of an individual given his various life activities. Also using various visualizations of the dataset using tableau, we are able to derive some conclusions like how the eating habits affect the individual’s obesity level etc. Obesity level estimation is a complex task that involves various factors such as age, gender, height, weight, body fat percentage, and overall health status. Estimating obesity levels can be useful in identifying individuals who are at risk of developing obesity-related health problems such as type 2 diabetes, heart disease, and stroke. The most commonly used method for estimating obesity levels is the body mass index (BMI), which is a measure of body fat based on height and weight. However, BMI has its limitations and may not be accurate in some cases, such as athletes with high muscle mass or elderly individuals with low muscle mass. Other methods for estimating obesity levels include waist circumference measurements, skinfold thickness measurements, and bioelectrical impedance analysis. It's important to note that while these methods can be useful in estimating obesity levels, they should not be used in isolation to diagnose obesity or other health conditions. A comprehensive evaluation of an individual's health status should be conducted by a qualified healthcare professional. Additionally, it's important to consider cultural and individual differences when interpreting obesity level estimates, as what may be considered a healthy weight in one population may not be applicable to another.

1. Introduction:

In 2016, about 39% and 13% of the worldwide adult population were overweight and obese respectively. Obesity is often correlated with demographic and epidemiological alterations, which in turn affect the mortality and fertility rates. Increasing obesity has led to an increase in lifestyle-related diseases such as type 2 diabetes, hypertension and dyslipidemia, together known as the metabolic syndrome. Along with diabetes, its long-term micro and macro-vascular complications affecting the kidneys, heart, and nerves have also shown a rising trend. Thus, addressing this mounting concern of obesity and taking measures to prevent and treat it in a timely manner is of utmost importance to avoid related complications. In addition to the health complications associated with obesity, there are also significant economic costs associated with the disease. The costs of treating obesity-related illnesses and lost productivity due to obesity are estimated to be in the billions of dollars each year.

There are various factors contributing to the rise in obesity rates, including changes in dietary patterns and physical activity levels. In many countries, there has been an increase in the consumption of high-calorie, processed foods and a decrease in physical activity due to more sedentary lifestyles. To address the issue of obesity, public health campaigns and interventions are needed. These interventions should focus on promoting healthy eating habits and increasing physical activity levels. This can be achieved through initiatives such as improving access to healthy foods, creating safe and walkable environments, and increasing awareness of the health risks associated with obesity.
In addition, healthcare providers can play an important role in addressing obesity. They can screen patients for obesity and provide counseling on healthy lifestyle choices. They can also prescribe medications or refer patients for weight loss surgery in cases of severe obesity. Overall, addressing the issue of obesity is a complex and multifaceted problem that requires a coordinated effort from various stakeholders. By taking action to prevent and treat obesity, we can improve public health outcomes and reduce the economic burden of this disease.

2. Literature Review:

Obesity is a growing health concern globally, with an estimated 2.8 million deaths annually attributed to overweight and obesity-related illnesses. In recent years, there has been increasing interest in developing accurate and reliable methods for estimating obesity levels based on factors such as eating habits and physical condition. This literature review summarizes some of the key studies in this area.

- "Estimation of obesity levels based on dietary habits and physical activity: a population-based study in Japan" by Murakami et al. (2011)

This study investigated the relationship between dietary habits, physical activity, and obesity levels in a sample of 1,435 Japanese adults. The results showed that a higher intake of energy-dense foods and lower physical activity levels were significantly associated with higher body mass index (BMI) and waist circumference. The authors concluded that dietary habits and physical activity levels are important factors in the estimation of obesity levels.

- "Estimation of obesity levels based on dietary intake and physical activity in Malaysian adults" by Chin et al. (2014)

This cross-sectional study examined the relationship between dietary intake, physical activity, and obesity levels in a sample of 291 Malaysian adults. The results showed that higher energy intake and lower physical activity levels were significantly associated with higher BMI and body fat percentage. The authors suggested that assessing dietary intake and physical activity levels may be useful in estimating obesity levels in Malaysian adults.

- "Relationship between eating habits and obesity in university students" by Al-Rethaiaa et al. (2010)

This study investigated the relationship between eating habits and obesity in a sample of 1,015 Saudi Arabian university students. The results showed that students who had a high frequency of eating unhealthy foods (such as fast food and sweets) and skipping breakfast were more likely to be overweight or obese. The authors concluded that assessing eating habits may be useful in estimating obesity levels in university students.

- "Association between physical activity, sedentary behavior, and obesity in adults: a review" by Biswas et al. (2015)

This systematic review and meta-analysis of 23 studies investigated the relationship between physical activity, sedentary behavior, and obesity in adults. The results showed that higher levels of physical activity were associated with lower BMI and waist circumference, while higher levels of sedentary behavior were associated with higher BMI and waist circumference. The authors suggested that assessing both physical activity and sedentary behavior may be useful in estimating obesity levels in adults.

In conclusion, the literature suggests that assessing factors such as dietary habits and physical activity levels may be useful in estimating obesity levels. Further research is needed to develop accurate and reliable methods for estimating obesity levels based on these factors, particularly in different cultural and ethnic populations.

3. Materials

3.1 Algorithm

- Data preparation: Collect and prepare the data for analysis. Data is divided into two sets of training and testing data in the ratio of 7:3.
- Model building: Naïve Bayes model, logistic regression, decision tree and random forest models are used.
- Predictions: Use the model to make predictions on the testing data.
- Model evaluation: Models are evaluated by generating the confusion matrix of predicted results and the test data.
- Model interpretation: By considering the evaluation results, the best suitable model is selected and the various results are used for interpretation.

3.1.1 Naïve Bayes

Naive Bayes is a probabilistic algorithm that is commonly used for classification tasks in machine learning. It is based on the Bayes theorem which describes the probability of an event occurring given prior knowledge about related events. Naive Bayes assumes that the features used for classification are independent of each other, which is often not true in real-world problems. Despite this oversimplification, it has been shown to perform well in many practical applications. Overall, Naive Bayes is a simple yet effective algorithm that is often used as a baseline for classification tasks.

3.1.2 Logistic Regression

Logistic regression is a statistical method used in machine learning to model the probability of a binary outcome based on one or more predictor variables. It is a type of regression analysis that is commonly used in situations where the dependent variable is dichotomous or binary (e.g., true/false, yes/no, 0/1).The goal of logistic regression is to find the best-fitting model that predicts the probability of the binary outcome as a function of the predictor variables. This is achieved by estimating the parameters of the logistic regression equation using a training set of data. Once the parameters have been estimated, the model can be used to predict the probability of the binary outcome for new data.

3.1.3 Decision Tree

A decision tree is a machine learning algorithm that can be used for classification and regression tasks. It creates a tree-like structure where each node represents a test on a feature, each branch represents the outcome of the test, and each leaf node represents a class label or a numerical value. The algorithm selects the best feature to split the data based on a criterion, and recursively applies the same process to each child node until a stopping criterion is met. Decision trees are easy to interpret and visualize, can handle both categorical and continuous data, and can be used for various tasks. However, they can be prone to overfitting and may not perform well on complex datasets with many features.

3.1.4 Random Forest

Random Forest is an algorithm in machine learning that combines multiple decision trees to improve accuracy and reduce overfitting. It works by training multiple decision trees, each on a different subset of the input data and features, and then combining their predictions to make a final prediction. It can handle various input data types, missing data, and noisy data without much preprocessing. It is also relatively fast and can handle large datasets. However, it may be less interpretable than a single decision tree, and tuning its hyperparameters can be challenging. Despite these limitations, Random Forest is a powerful and widely used algorithm in machine learning for various applications.

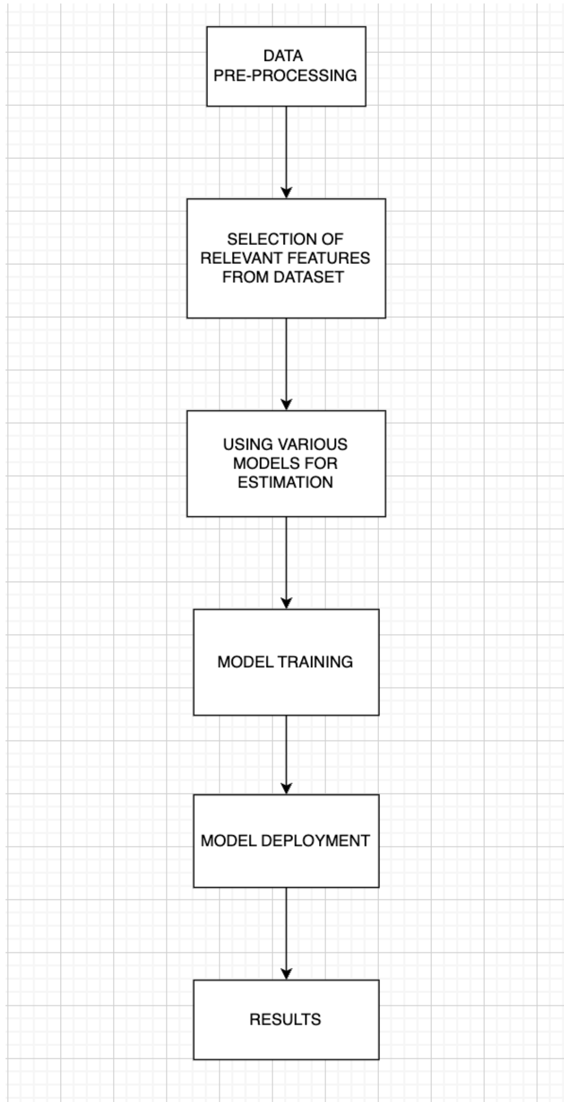
3.1.5 K-Means clustering

K-means clustering is a popular unsupervised machine learning algorithm used for clustering or grouping data points into distinct clusters based on their similarity. It is a type of partition-based clustering technique that assigns each data point to the nearest centroid, which is the mean of the data points within that cluster. K-means clustering can be used for a variety of applications, such as image segmentation, customer segmentation, anomaly detection, and document grouping.

3.2 Dataset:

The data contains 17 attributes and 2111 records, the records are labeled with the class variable NObesity (Obesity Level), that allows classification of the data using the values of Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II and Obesity Type III.
This data presents information from different locations such as Mexico, Peru and Colombia, and can be used to build estimation of the obesity levels based on the nutritional behavior of several regions. The attributes related with eating habits are: Frequent consumption of high caloric food (FAVC), Frequency of consumption of vegetables (FCVC), Number of main meals (NCP), Consumption of food between meals (CAEC), Consumption of water daily (CH20), and Consumption of alcohol (CALC). The attributes related with the physical condition are: Calories consumption monitoring (SCC), Physical activity frequency (FAF), Time using technology devices (TUE), Transportation used (MTRANS), other variables obtained were: Gender, Age, Height and Weight.

3.3 Architecture:



4. Proposed Work:

4.1 Novelty:

An important aspect of predicting the type of obesity is to ensure that the model is robust and can make accurate predictions for a wide range of individuals. One potential approach to achieve this is to use an ensemble of machine learning models, such as Naive Bayes, Decision Tree, and Random Forest, to generate a more accurate and reliable prediction. By combining the strengths of multiple models, the ensemble approach can overcome the limitations of individual models and provide a more robust prediction of the type of obesity. Moreover, ensemble methods are known to reduce the risk of overfitting, which can improve the generalizability of the model. Therefore, using an ensemble of models can potentially lead to more accurate and reliable predictions of the type of obesity, which can be useful in designing effective intervention strategies for individuals at risk of developing the condition.

4.2 Project Contributions:

Data preprocessing, Naive Bayes : Shrikumaran P.A.
Logistic Regression, Decision Tree, Random Forest : Anu Cyril Saju
Tableau visualization: Renil Augustine Baby
Documentation: Shrikumaran, Anu, Renil

5. Results and Discussion:

5.1 Results

From the figures and decision trees we are able to understand how different aspects of life affect a person's health and how those factors cause obesity in their lives. We observe that when a person does not drink water, does not do physical activity, not eat vegetables, the number of meals the person has in a day and the person's usage of technology has an effect on obesity.

We can also see that when a person smokes or drinks alcohol, they lose their appetite which leads to them having a much lower obesity risk compared to those who don't drink or smoke.

There is evidence to suggest that genetics can play a role in the development of obesity. Many studies have shown that individuals who have a family history of obesity are more likely to become obese themselves, and researchers have identified several genes that may contribute to the development of obesity. In this project also we can see that more than 93% people with overweight and obesity also have a family history of people with overweight.

One of the genes that has been associated with obesity is called the FTO gene. Variants of this gene have been found to be associated with increased body weight, body mass index (BMI), and risk of obesity. The FTO gene is involved in regulating food intake and energy expenditure, and variations in this gene may lead to changes in appetite or metabolism that contribute to the development of obesity.

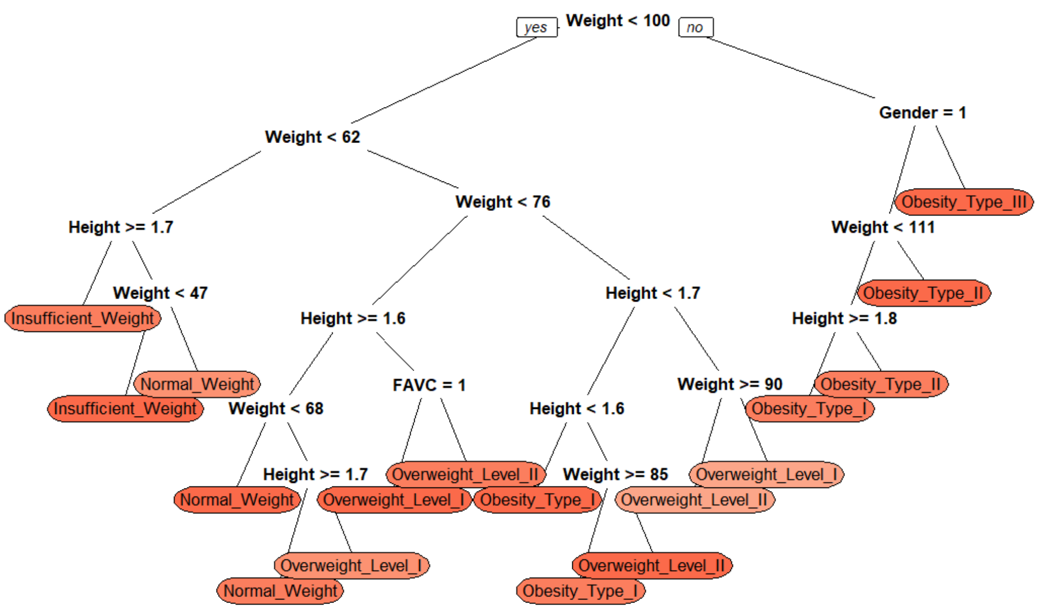
Accuracy of Naive Bayes: 63.61%
Accuracy of Logistic Regression: 95.73%
Accuracy of Decision Tree: 86.71%
Accuracy of Random Forest: 96.84%

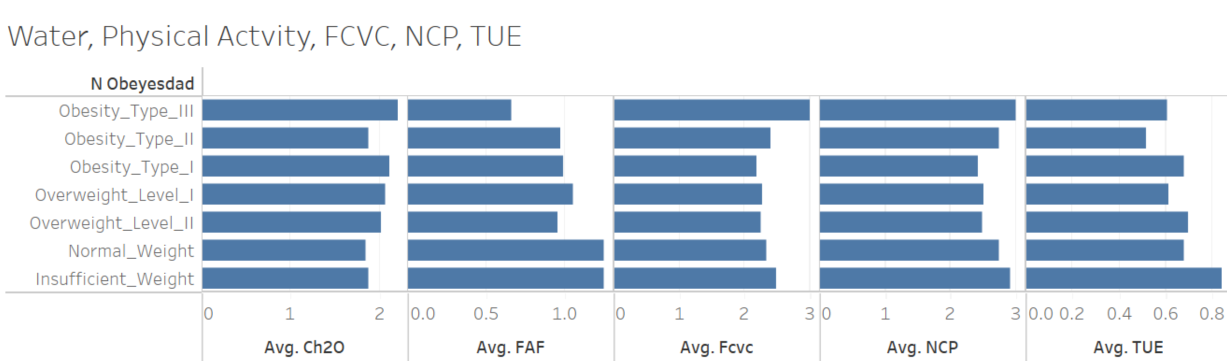
Therefore we use random forest regression for obesity level estimation due its high level of accuracy. By giving the various factors as input we are able to predict the obesity level of that person with high accuracy.

It is important to note, however, that while genetics can contribute to the development of obesity, it is not the only factor at play. Environmental factors such as diet, physical activity, and lifestyle choices also play a significant role in the development of obesity. Therefore, it is important to adopt healthy habits and maintain a healthy lifestyle, even if you have a genetic predisposition to obesity.

5.2 Figures and Tables:

Decision tree:



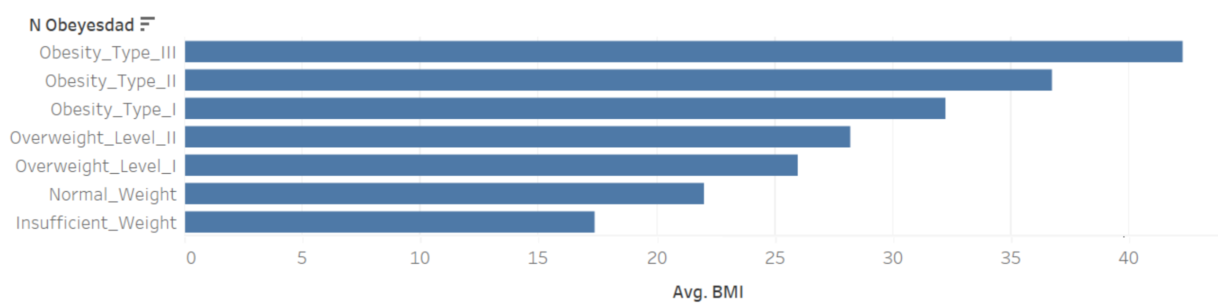


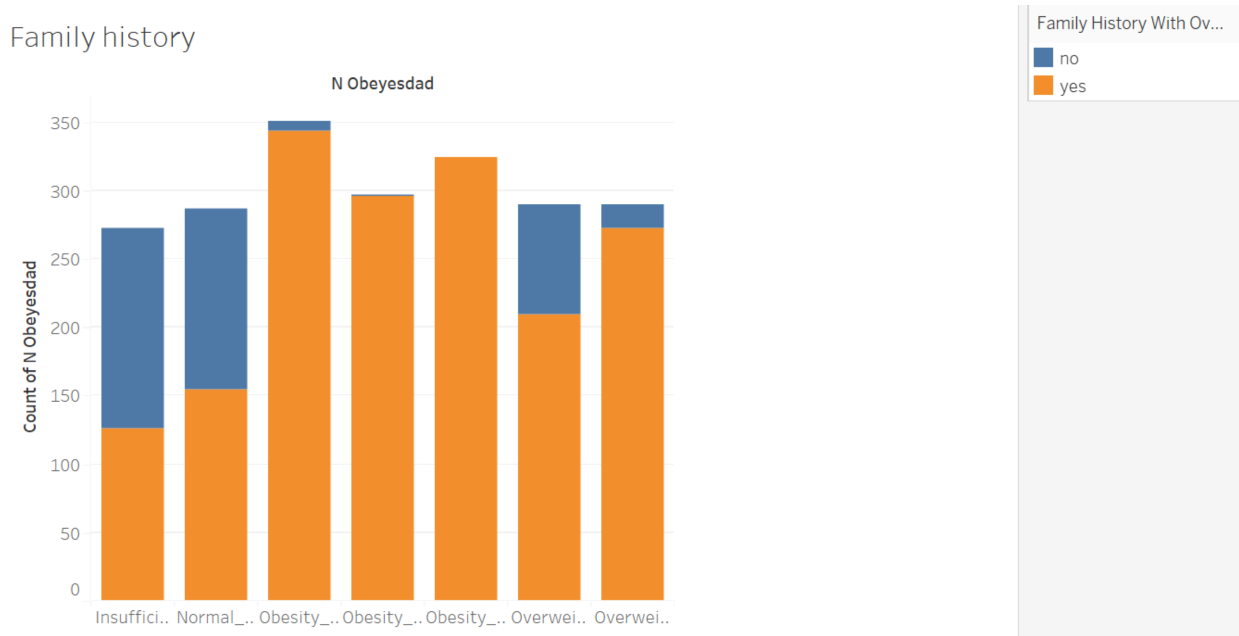
Obesity Levels

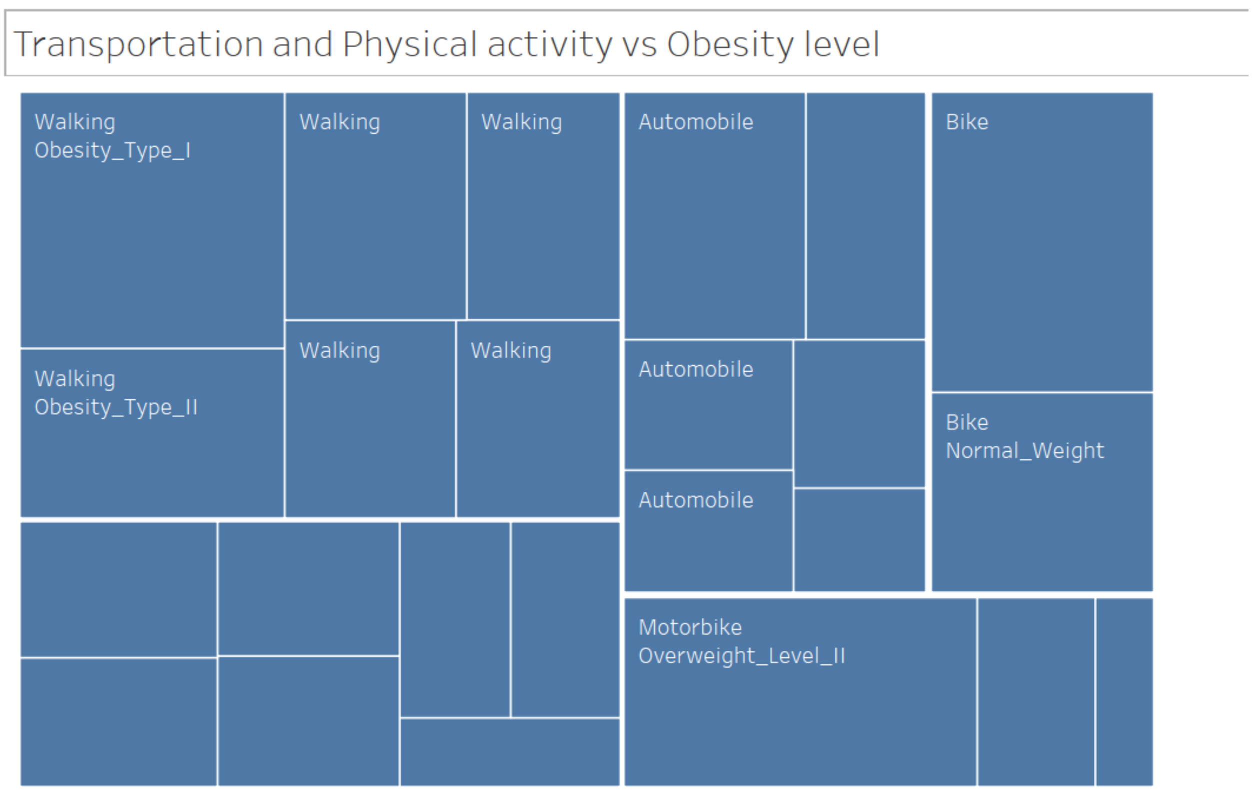
Alcohol and Smoking habits

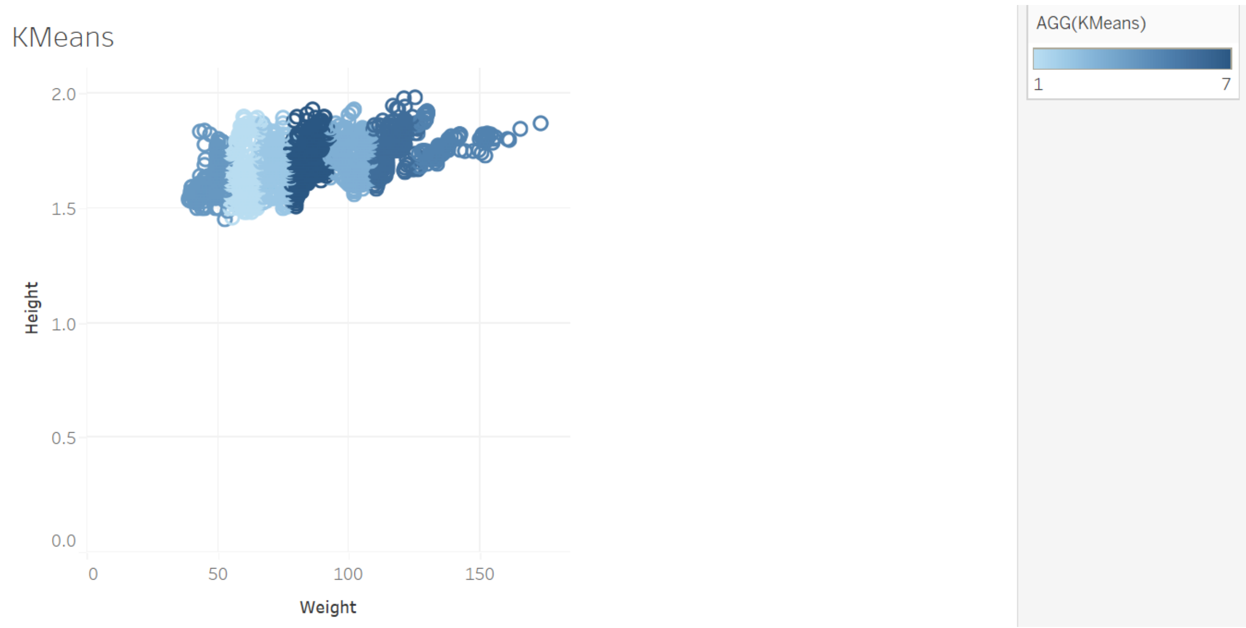
Smoke	N Obeyesdad	Calc: Always	Calc: Freque..	Calc: no	Calc: Someti..
no	Insufficient_Weight		1	117	153
	Normal_Weight	1	15	105	153
	Obesity_Type_I		12	163	170
	Obesity_Type_II		1	71	210
	Obesity_Type_III			1	322
	Overweight_Level_I		16	49	222
	Overweight_Level_II		18	127	140
yes	Insufficient_Weight				1
	Normal_Weight		3	2	8
	Obesity_Type_I		2	2	2
	Obesity_Type_II		1		14
	Obesity_Type_III				1
	Overweight_Level_I			1	2
	Overweight_Level_II		1	1	3

BMI




Family history
N Obeyesdad
Count of N Obeyesdad
350
300
250
200
150
100
50
0
Insuffici..
Normal_..
Obesity_..
Obesity_..
Obesity_..
Overwei..
Overwei..
Family History With Ov...
no
yes


Transportation and Physical activity vs Obesity level
Walking
Obesity_Type_I
Walking
Walking
Automobile
Bike
Walking
Obesity_Type_II
Walking
Walking
Automobile
Automobile
Bike
Normal_Weight
Motorbike
Overweight_Level_II



5.3 Explanation:

We notice that a person's weight changes depending on how many meals they consume throughout the day, how much technology they use, and whether or not they drink water, exercise, or eat vegetables. We may also see that people who smoke or drink alcohol tend to be less hungry, which lowers their risk of obesity significantly compared to people who don't drink or smoke. Also, it has been shown that those who are underweight or normal weight engage in greater physical activity since they choose to walk as their primary way of transportation. Similar to underweight persons, overweight folks do not exercise much and prefer driving to walking. Among the 17 factors we have, it can be observed using decision tree that major factors affecting obesity level are weight, height, gender and FAVC. But using logistic regression it can be seen that other factors like family history, physical activity, number of meals, smoking and alcohol consumption are also affecting the prediction. Using KMeans we are able to say that the seven levels are appropriate for classifying the obesity weight levels.

6. Conclusion:

In conclusion, this project aimed to predict the type of obesity based on several input attributes, including height, weight, family history, consumption of high caloric food, smoking, frequency of drinking water, and time of using technology. We used models such as Naive Bayes, Decision Tree, Logistic regression and Random Forest, to generate predictions and explored potential novelties to improve their performance.

Our results showed that all three models performed reasonably well in predicting the type of obesity, with Random Forest producing the highest accuracy of 96.84%. However, we also found that combining the models using an ensemble approach can lead to a more robust and accurate prediction, which can be useful in designing effective intervention strategies for individuals at risk of developing the condition.

In addition to the machine learning models, we used Tableau to visualize the impact of each input attribute on the target attribute. This allowed us to identify which attributes had the most significant impact on the development of obesity, and to explore potential patterns and relationships in the data. For example, we found that individuals with a family history of obesity and those who consume high caloric food were more likely to develop the condition.

Overall, this project demonstrates the potential of using machine learning models and data visualization tools, such as Tableau, to predict and understand the development of obesity. By combining these approaches, we can generate more accurate predictions and gain a better understanding of the underlying factors that contribute to the condition. These findings can have important implications for designing effective interventions to prevent and manage obesity, which can have significant public health benefits.

7. References:

1. Lu, J., Huang, Y., Zhao, J., & Yang, Y. (2018). Obesity Level Prediction Using Machine Learning Algorithms on Electronic Medical Records.

2. Chen, M., Li, X., Zhang, Y., & Zhu, Z. (2019). Predicting Obesity Level Based on Demographic and Lifestyle Factors Using Machine Learning Algorithms.

3. Quan, W., Zhang, Y., & Zhu, J. (2020). Obesity Level Prediction Based on Machine Learning Algorithms Using Electronic Health Records.

4. Javed, M. A., Khan, A., & Hussain, M. (2021). Obesity Level Prediction using Machine Learning and Data Mining Techniques..

5. Sayad, S., & Alizadehsani, R. (2021). Predicting obesity levels using machine learning techniques.
6. Peng, Y., Wang, Y., & Jia, J. (2020). Predicting obesity level based on anthropometric and blood indicators using machine learning algorithms.

7. Lee, J. S., Park, S. H., Lee, S. H., Kim, J. H., & Cho, K. H. (2019). Predicting obesity level using wearable devices and machine learning algorithms.

8. López-García, E., Banegas, J. R., & Graciani, A. (2019). Obesity level prediction by machine learning in the elderly.

9. Su, Y., Zhang, Y., & Wang, Z. (2021). Predicting obesity level using artificial intelligence techniques: A systematic review.

10. Hou, Y., Wang, Y., Liu, Y., & Xu, Q. (2020). Predicting obesity level with machine learning algorithms: A comparative study.

11. Estimation of obesity levels based on dietary habits and physical activity: a population-based study in Japan" by Murakami et al. (2011)

12. Estimation of obesity levels based on dietary intake and physical activity in Malaysian adults" by Chin et al. (2014)

13. Relationship between eating habits and obesity in university students" by Al-Rethaiaa et al. (2010)

14. Association between physical activity, sedentary behavior, and obesity in adults: a review" by Biswas et al. (2015)

Appendix:

Github Link:
https://github.com/anu-cyril-saju/Obesity-level-prediction

OUTPUTS:

Naïve Bayes:

```
Confusion Matrix and Statistics

                     Reference
Prediction            Insufficient_Weight Normal_Weight Obesity_Type_I Obesity_Type_II
  Insufficient_Weight                  70            26              0               0
  Normal_Weight                         3            35              2               0
  Obesity_Type_I                        0             6             65               5
  Obesity_Type_II                       0             0             27              84
  Obesity_Type_III                      0             6              7               0
  Overweight_Level_I                    8             9              1               0
  Overweight_Level_II                   0             4              3               0
                     Reference
Prediction            Obesity_Type_III Overweight_Level_I Overweight_Level_II
  Insufficient_Weight                0                  3                   0
  Normal_Weight                      0                  6                   4
  Obesity_Type_I                     0                 33                  47
  Obesity_Type_II                    0                  1                   7
  Obesity_Type_III                  97                 11                   1
  Overweight_Level_I                 0                 28                   5
  Overweight_Level_II                0                  5                  23
```

```
Overall Statistics

               Accuracy : 0.6361
                 95% CI : (0.5972, 0.6737)
    No Information Rate : 0.1661
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.5735
```

Logistic Regression:

```
Call:
nnet::multinom(formula = target, data = training)

Coefficients:
                    (Intercept)      Gender        Age    Height    Weight
Normal_Weight          63.37617  -0.6213732  0.1749291 -103.2139  2.342823
Obesity_Type_I        237.69906 -10.3221169  0.4758299 -424.4999  6.912518
Obesity_Type_II       101.29260  -2.7884571  3.0339452 -593.8696 10.819680
Obesity_Type_III     -142.49608 -76.4534304 -0.4743325 -417.2540  9.557637
Overweight_Level_I    126.57077  -4.2110494  0.2554682 -202.6742  3.742450
Overweight_Level_II   195.10855  -4.1892691  0.3668054 -326.2464  5.518120
                    family_history_with_overweight       FAVC      FCVC       NCP       CAEC
Normal_Weight                            -2.688573  -1.3134508 -1.755445 -1.760916  0.9677845
Obesity_Type_I                           -1.510641  -1.2400783 -1.517292 -2.216566  5.0670664
Obesity_Type_II                         -13.179445 -26.3199162  3.581017 -1.978897 13.6952078
Obesity_Type_III                         15.561520  11.4570486 35.351566 10.967686 -8.0696345
Overweight_Level_I                       -2.529073   0.7775005 -2.847767 -1.443838  2.4764582
Overweight_Level_II                       1.010290  -2.7546184 -1.879747 -2.879899  4.4429024
                        SMOKE       CH2O        SCC         FAF        TUE        CALC
Normal_Weight        3.718852  -2.619088   0.248463  -0.5843323  0.1748110  -1.3250577
Obesity_Type_I       6.477200  -2.161070  -2.766845  -2.3908003  1.3321925  -3.6324007
Obesity_Type_II     -1.789349 -19.606926 -24.834996 -14.2382548 -1.8856077 -16.5841323
Obesity_Type_III    -4.117173  -2.820970  62.554036 -11.9871450  6.3877223   6.0011585
Overweight_Level_I   2.339677  -2.293473   3.775062  -0.7483555  0.4646792  -0.9675519
Overweight_Level_II  6.215628  -2.912039   3.321827  -1.5286390  1.4995163  -2.7050679
```

```
                        MTRANS
Normal_Weight        0.6299535
Obesity_Type_I       1.6081960
Obesity_Type_II      7.1388949
Obesity_Type_III    -2.5260944
Overweight_Level_I   0.6110564
Overweight_Level_II  0.9488012
```

```
Confusion Matrix and Statistics

                     Reference
Prediction            Insufficient_Weight Normal_Weight Obesity_Type_I Obesity_Type_II
  Insufficient_Weight                  78             3              0               0
  Normal_Weight                         3            82              0               0
  Obesity_Type_I                        0             0            103               3
  Obesity_Type_II                       0             0              1              86
  Obesity_Type_III                      0             0              1               0
  Overweight_Level_I                    0             1              0               0
  Overweight_Level_II                   0             0              0               0
                     Reference
Prediction            Obesity_Type_III Overweight_Level_I Overweight_Level_II
  Insufficient_Weight                0                  0                   0
  Normal_Weight                      0                  1                   0
  Obesity_Type_I                     0                  0                   2
  Obesity_Type_II                    0                  0                   1
  Obesity_Type_III                  97                  0                   0
  Overweight_Level_I                 0                 80                   5
  Overweight_Level_II                0                  6                  79

Overall Statistics

               Accuracy : 0.9573
                 95% CI : (0.9384, 0.9717)
    No Information Rate : 0.1661
    P-Value [Acc > NIR] : < 2.2e-16
```

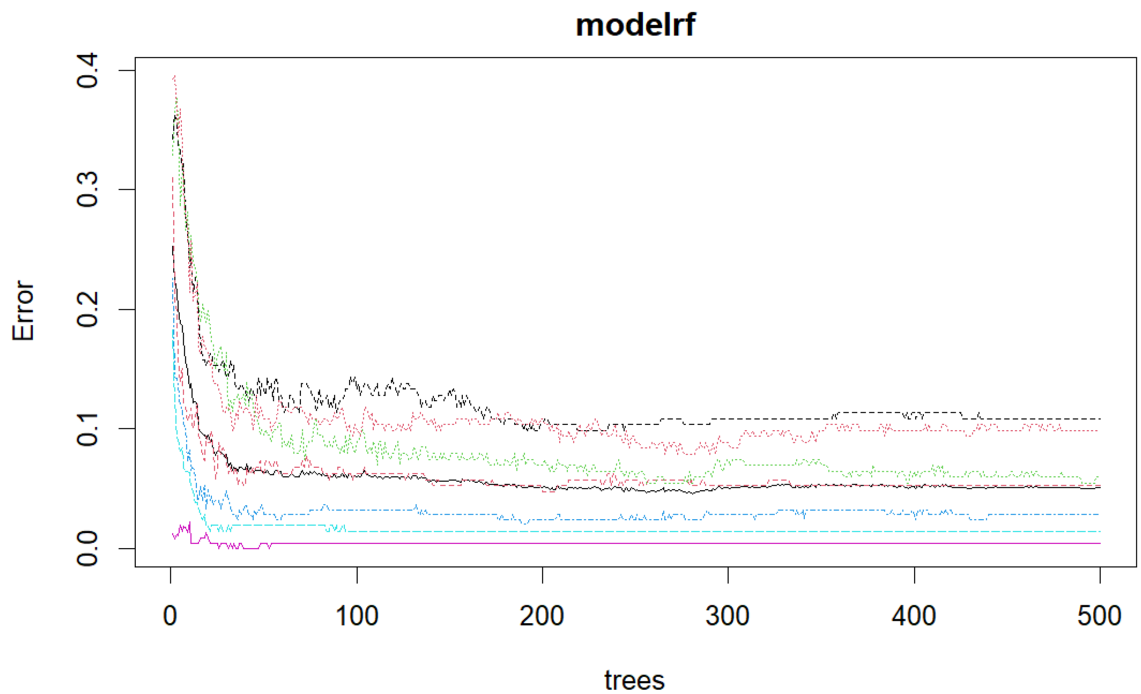
Random Forest:

```
Call:
 randomForest(formula = as.factor(NObeyesdad) ~ ., data = training)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 4

        OOB estimate of  error rate: 5.07%
```

```
Overall Statistics

               Accuracy : 0.9684
                 95% CI : (0.9515, 0.9806)
    No Information Rate : 0.1646
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.963

 Mcnemar's Test P-Value : NA
```

```
Statistics by Class:

                     Class: Insufficient_Weight Class: Normal_Weight Class:
Obesity_Type_I
Sensitivity                              1.0000               0.8646
0.9904
Specificity                              0.9910               0.9944
0.9962
Pos Pred Value                           0.9383               0.9651
0.9810
Neg Pred Value                           1.0000               0.9762
0.9981
Prevalence                               0.1203               0.1519
0.1646
Detection Rate                           0.1203               0.1313
0.1630
Detection Prevalence                     0.1282               0.1361
0.1661
Balanced Accuracy                        0.9955               0.9295
0.9933
```

```
                     Class: Obesity_Type_II Class: Obesity_Type_III Class:
Overweight_Level_I
Sensitivity                         1.0000                 1.0000
0.9524
Specificity                         1.0000                 1.0000
0.9872
Pos Pred Value                      1.0000                 1.0000
0.9195
Neg Pred Value                      1.0000                 1.0000
0.9927
Prevalence                          0.1408                 0.1535
0.1329
Detection Rate                      0.1408                 0.1535
0.1266
Detection Prevalence                0.1408                 0.1535
0.1377
Balanced Accuracy                   1.0000                 1.0000
0.9698
                     Class: Overweight_Level_II
Sensitivity                              0.9767
Specificity                              0.9945
Pos Pred Value                           0.9655
Neg Pred Value                           0.9963
Prevalence                               0.1361
Detection Rate                           0.1329
Detection Prevalence                     0.1377
Balanced Accuracy                        0.9856
```



Decision Tree:

```
Confusion Matrix and Statistics

                     Reference
Prediction            Insufficient_Weight Normal_Weight Obesity_Type_I Obesity_Type_II Obesity_Type_III
  Insufficient_Weight                  77             8              0               0                0
  Normal_Weight                         4            67              0               0                0
  Obesity_Type_I                        0             0             76               2                0
  Obesity_Type_II                       0             0             11              87                0
  Obesity_Type_III                      0             0              0               0               97
  Overweight_Level_I                    0            10              1               0                0
  Overweight_Level_II                   0             1             17               0                0
                     Reference
Prediction            Overweight_Level_I Overweight_Level_II
  Insufficient_Weight                  0                   0
  Normal_Weight                        7                   1
  Obesity_Type_I                       0                   3
  Obesity_Type_II                      0                   0
  Obesity_Type_III                     0                   0
  Overweight_Level_I                  77                  16
  Overweight_Level_II                  3                  67

Overall Statistics

               Accuracy : 0.8671
                 95% CI : (0.8381, 0.8926)
    No Information Rate : 0.1661
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.845

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: Insufficient_Weight Class: Normal_Weight Class: Obesity_Type_I Class: Obesity_Type_II
Sensitivity                              0.9506               0.7791                0.7238                 0.9775
Specificity                              0.9855               0.9780                0.9905                 0.9797
Pos Pred Value                           0.9059               0.8481                0.9383                 0.8878
Neg Pred Value                           0.9927               0.9656                0.9474                 0.9963
Prevalence                               0.1282               0.1361                0.1661                 0.1408
Detection Rate                           0.1218               0.1060                0.1203                 0.1377
Detection Prevalence                     0.1345               0.1250                0.1282                 0.1551
Balanced Accuracy                        0.9680               0.8785                0.8572                 0.9786
                     Class: Obesity_Type_III Class: Overweight_Level_I Class: Overweight_Level_II
Sensitivity                           1.0000                    0.8851                     0.7701
Specificity                           1.0000                    0.9505                     0.9615
Pos Pred Value                        1.0000                    0.7404                     0.7614
Neg Pred Value                        1.0000                    0.9811                     0.9632
Prevalence                            0.1535                    0.1377                     0.1377
Detection Rate                        0.1535                    0.1218                     0.1060
Detection Prevalence                  0.1535                    0.1646                     0.1392
Balanced Accuracy                     1.0000                    0.9178                     0.8658
```

Calculation Fields in tableau:

KMeans

Results are computed along Table (across).

```
SCRIPT_INT('set.seed(42);result <-
kmeans(data.frame(.arg1,.arg2), 7);result$cluster;',
SUM([Weight]),SUM([Height]))
```

BMI

```
[Weight]/([Height]^2)
```