

Estimation of Obesity Levels Analysis Report

1. Dataset Description

The "Estimation of Obesity Levels Based On Eating Habits and Physical Condition" dataset is designed for a classification task, aiming to predict an individual's obesity level. The target variable, 'NObesity', is categorical with seven distinct levels: Insufficient_Weight, Normal_Weight, Overweight_Level_I, Overweight_Level_II, Obesity_Type_I, Obesity_Type_II, and Obesity_Type_III.¹ The prediction is based on 17 input features that encompass various aspects of a person's lifestyle. These features include eating patterns (frequency of high-calorie food (FAVC), vegetable intake (FCVC), number of main meals (NCP), consumption of food between meals (CAEC)), physical activity frequency (FAF), time spent using technology devices (TUE), smoking habits (SMOKE), water consumption (CH2O), monitoring of calorie intake (SCC), frequency of alcohol consumption (CALC), family history of overweight, gender, age, height, weight, and transportation habits (MTRANS). Each row in the dataset represents an individual's profile with their corresponding obesity level. This dataset allows for a comprehensive analysis of the factors influencing obesity levels.

2. Key Findings

- Individuals with a family history of overweight are more likely to be in a higher obesity level category.
- Higher reported frequency of consuming high-caloric food (FAVC) is prevalent across all obesity levels.
- There is a trend suggesting that individuals with more frequent consumption of vegetables (higher FCVC) tend to be in lower obesity level categories.
- The average weight generally increases with higher obesity levels, as expected.
- Different modes of transportation ('MTRANS') show variations in the average weight of individuals using them. Walking and biking are associated with lower average weights compared to using automobiles or public transportation.
- Physical activity frequency (FAF) appears to be generally lower in individuals with higher obesity levels.
- Gender does not show a consistently strong differentiating effect across all obesity levels for factors like age or physical activity, although some variations exist within specific transportation modes and family history groups.

3. Insights from Visualizations

- **Histograms of Age, Height, and Weight:** The distributions of age and height appear relatively normal, while the weight distribution shows a right skew, indicating a larger number of individuals in the higher weight ranges.
- **Box Plots of Weight and Age by NObesity:** Box plots clearly illustrate the increasing median weight with higher obesity levels. The age distribution across different obesity levels shows some variation but no dramatic differences in central tendency.

- **Violin Plot of Weight by NObesity:** The violin plot provides a more detailed view of the weight distribution within each obesity level, revealing the spread and density of data points. It confirms the trend of increasing weight with higher obesity levels and also shows the variability within each category.
- **Scatter Plot of Height vs Weight by NObesity:** The scatter plot shows a positive correlation between height and weight, with different obesity levels tending to cluster in different regions of the plot. This visualization helps to see how individuals of varying heights and weights are categorized into different obesity levels.
- **Scatter Plot of Age vs Weight by NObesity:** This plot shows the distribution of individuals of different ages and weights across the obesity levels. While there's no strict linear relationship evident, certain age and weight ranges appear to be more associated with specific obesity categories.
- **Line Plot of Weight Trend with Age:** The line plot suggests a general trend of increasing weight with age, although this trend might vary depending on other lifestyle factors.
- **Line Plot of Weight Trend with Frequency of Vegetable Consumption:** This plot indicates a potential inverse relationship, where individuals reporting higher frequencies of vegetable consumption might have a tendency towards lower weights.
- **Pair Plot of Numerical Features by NObesity:** The pair plot provides a matrix of scatter plots and histograms for all numerical features, colored by 'NObesity'. This allows for the visual inspection of relationships between different pairs of features and how they relate to the obesity level categories. It helps in identifying potential correlations and separating characteristics of different obesity levels.
- **Heatmap of Correlation:** The correlation heatmap shows the linear relationships between the numerical features. 'Weight' shows a positive correlation with 'Height' and potentially with age. It also shows the correlation between other lifestyle factors. The target variable 'NObesity' (after potential encoding) would also show correlations with the input features.