CLASSIFICATION OF BMI CATEGORY

Based on Physical Conditions

&

Eating Habits

Presented by B239464
to
Health Policy Makers

OUTLINES

Introduction

Health Records

Exploratory Analysis

Fitted Model

Strengths & Weakness of Approach

Broad Implications

Conclusion / Recommendation

INTRODUCTION

This dataset include data for the estimation of obesity levels (BMI category) in individuals from the countries of Mexico, Peru and Colombia, based on their eating habits and physical condition.

77% of the data was generated synthetically using the Weka tool and the SMOTE filter, 23% of the data was collected directly from users through a web platform.

The dataset Contain 17 records and 2,111 observations.

SELECTED HEALTH RECORDS

EATING HABITS

- Frequent consumption of high caloric food (favc)

- Frequency of consumption of vegetables (fcvc)

PHYSICAL CONDITIONS

- Weight

- Height

- Family history with overweight (fho)

- Physical activity frequency (faf)

BMI CATEGORY

- Insufficient Weight

- Normal Weight

- Overweight Level I

- Overweight Level II

- Obesity Level I

- Obesity Level II

- Obesity Level III

EXPLORATORY ANALYSIS

Table 1: Summary Statistics

summary	bmi_cat	height	weight	faf	fcvc	favc	fho
count	2111	2111.00	2111.00	2111.00	2111.00	2111	2111
mean	NA	1.70	86.59	1.01	2.42	NA	NA
stddev	NA	0.09	26.19	0.85	0.53	NA	NA
min	Insufficient_Weight	1.45	39.00	0.00	1.00	no	no
max	Overweight_Level_II	1.98	173.00	3.00	3.00	yes	yes
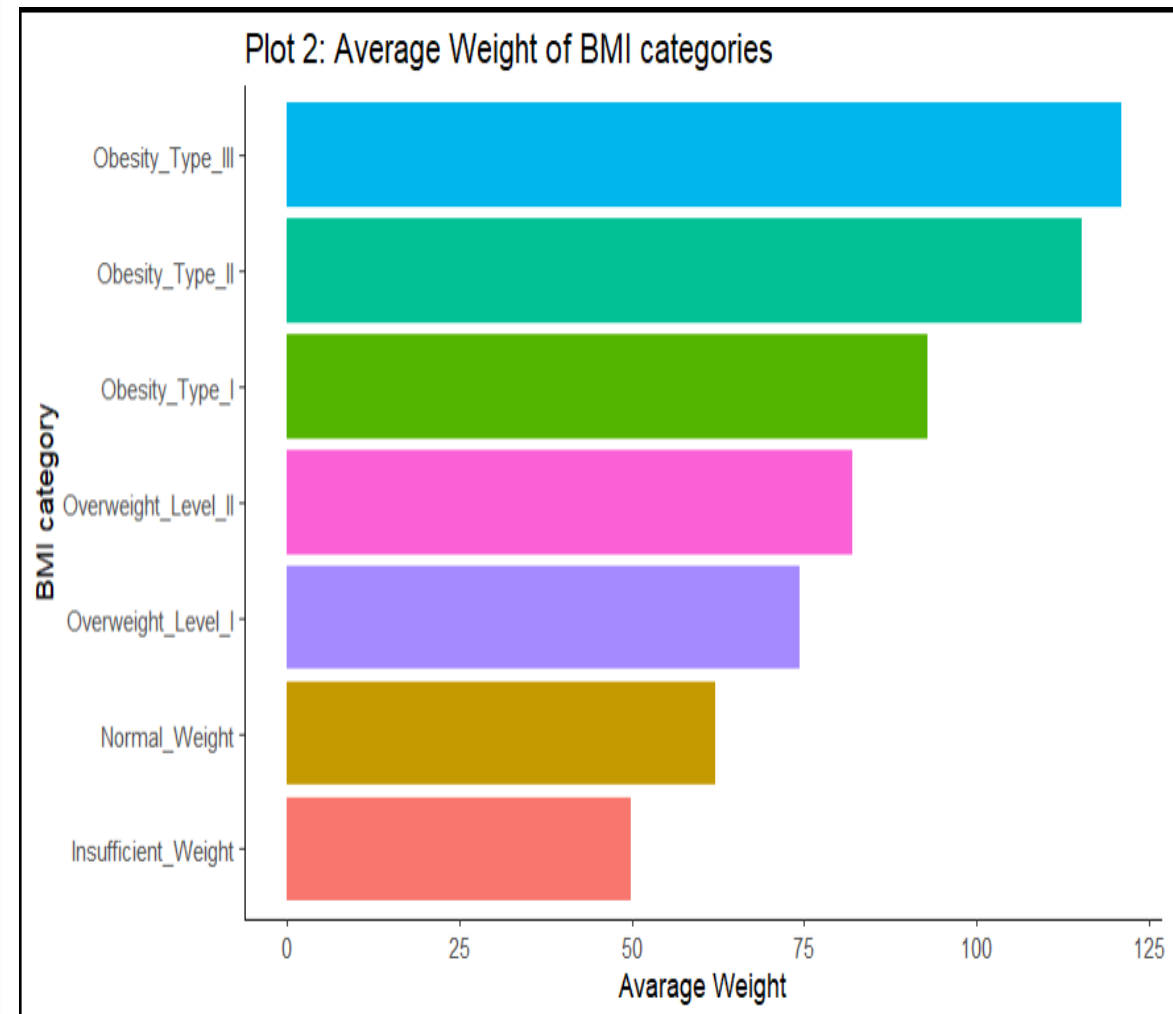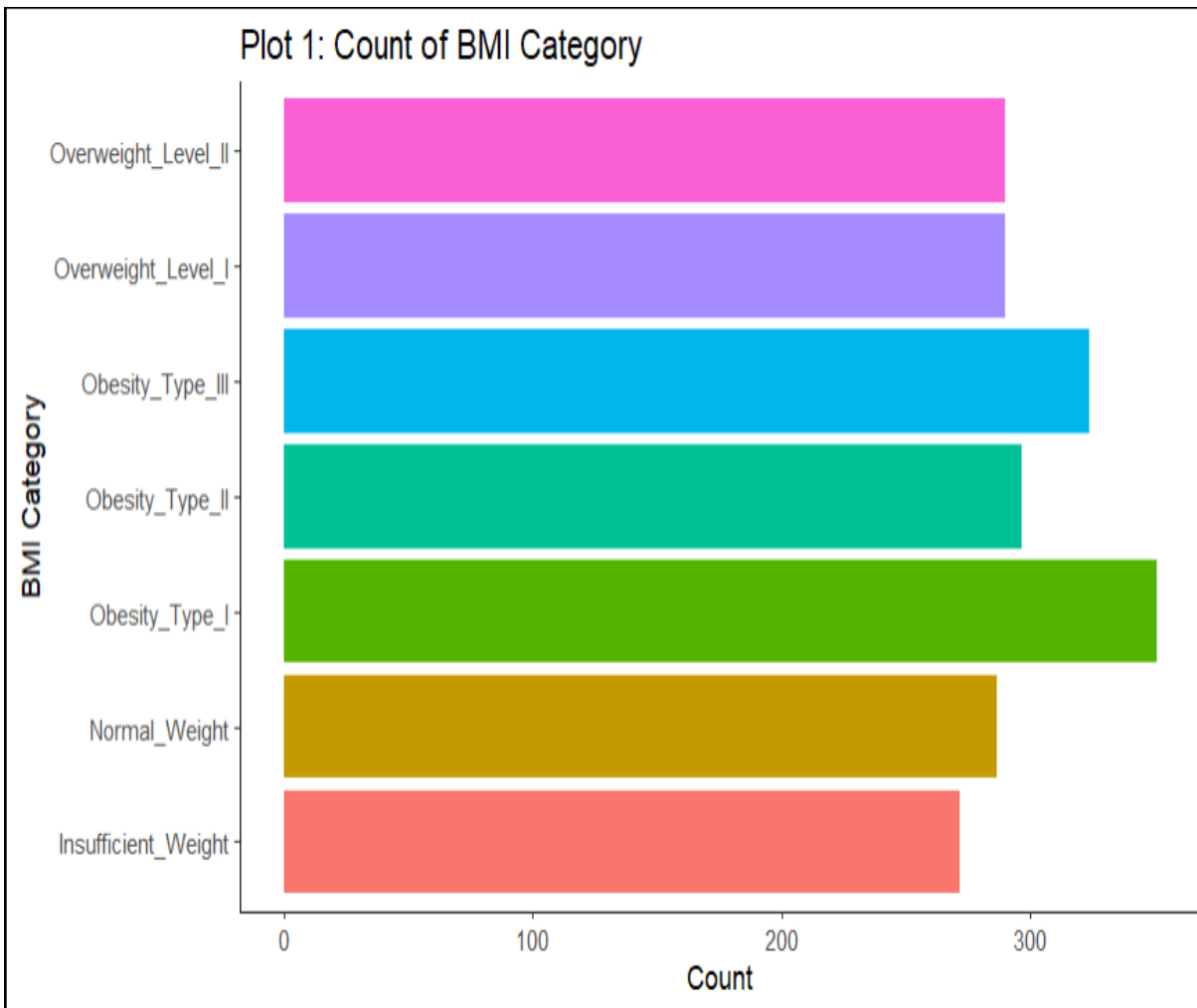
FITTED MODEL

A good model fit for predicting the class of several levels of BMI category would be a Random Forest Tree Model (RFT).

Why Use Random Forest Tree?

• Solid Principle

• Effective classification Voting

• Suitable for this data.

ml_random_forest(obesity_data, bmi_cat ~ height + weight + faf + fcvc + favc + fho, type = "classification")

MODEL RESULT

Table 5: Feature Importance

feature	importance
weight	0.51
height	0.22
fcvc	0.15
faf	0.07
fho__yes	0.03
favc__yes	0.02

PREDICTION ACCURACY

✓ 0.95 (95%)

PREDICTION F1 SCORE

✓ 0.95007 (95%)

Table 6: Top Row Predictions & lables on Test Data

bmi__cat	label	predicted_label	prediction
Insufficient_Weight	6	Insufficient_Weight	6
Insufficient_Weight	6	Insufficient_Weight	6
Insufficient_Weight	6	Insufficient_Weight	6
Normal_Weight	2	Normal_Weight	2
Normal_Weight	2	Insufficient_Weight	6
Normal_Weight	2	Normal_Weight	2

STRENGTH OF APPROACH

- Clear Variable Selection

- Robust Modelling Technique

- Detailed Model Evaluation

WEAKNESS OF APPROACH

- Synthetic Data Proportion

- Model Complexity

- High Computational Demand

BROAD IMPLICATIONS

Policy Development

Public Health Insights

Clinical Application

CONCLUSION / RECOMMENDATION

In summary, these findings highlight the effectiveness of Random Forest Model in predicting BMI category and the importance of some physical factors and eating habits in classifying BMI category.

The analysis provides valuable insights into the factors contributing to obesity, which can inform public health, policy strategies and interventions aimed at reducing obesity rates in the studied regions (Mexico, Peru and Colombia).