Estimation of Obesity Levels Based on Eating Habits and Physical Condition

Introduction

❖ Based on their dietary patterns and physical characteristics, this dataset estimates the prevalence of obesity in people from Mexico, Peru, and Colombia.

❖ The target record is labeled with the class variable NObesity (Obesity Level), which enables the data to Be classified using the values of Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II, and Obesity Type III. The data comprises 2111 records and 17 attributes.

❖ A study of 2,111 individuals between the ages of 14 and 61 served as the primary source of data for this research. A survey was used to get this data.

❖ Features of this project include frequent ingestion of foods high in calories, The frequency of vegetable eating, quantity of primary meals, Consumption

Objective

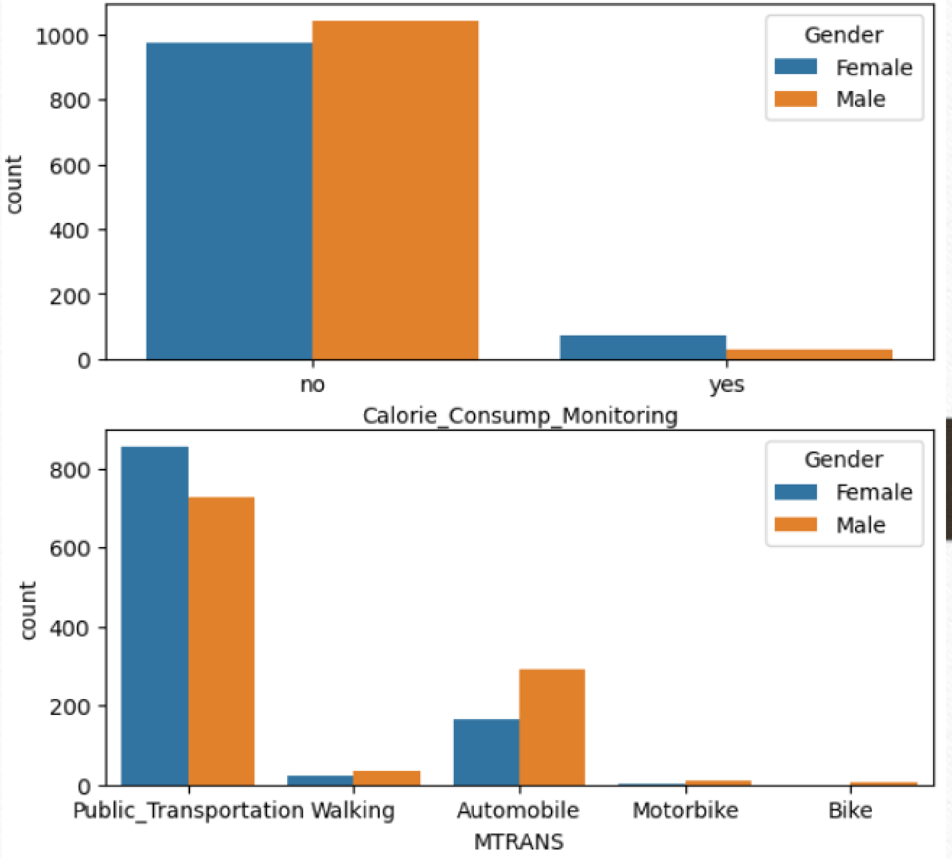
- People's quality of life can be negatively impacted by obesity, which can lead to several physical health issues. As a result, individuals are beginning to examine the causes of obesity and forecast when it may manifest.

- Because obesity leads to cardiovascular illnesses, it is becoming an issue in many developing countries.

- Therefore, this project's primary goal is to create a machine learning model that can predict an individual’s level of obesity based on their eating habits, physical condition, and other variables.

Modules & Workflow of project:

- Importing data
- Exploratory data analysis
- Handling missing values
- Handling outliers
- Handling categorical features
- Feature scaling
- Assumption check
- Feature selection
- Sampling the data
- Training different models
- Hyper tuning
- Result

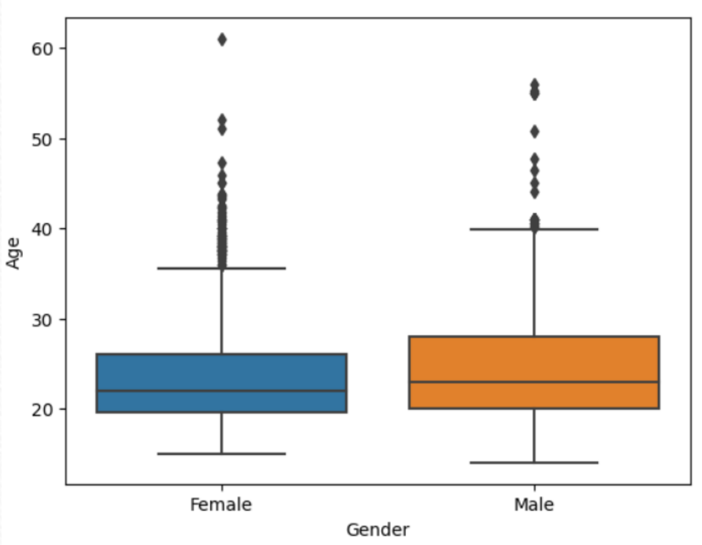


Tools / Platform & ML Algorithms

- ❖ I utilized the Anaconda distribution’s Jupyter Notebook for this Python coding assignment.

- ❖ Power BI was used to create dashboard.

- ❖ Implementing the Model: ML Algorithms: Random Forest, K Nearest Neighbor, Support Vector Classifier, Decision Tree, and Logistic Regression



Methodology:

- With the help of a computational intelligence model, this research aims to inform people about their risk of obesity based on a few important aspects of their lifestyle.
- **Data preprocessing and exploratory data analysis received between 50 and 60 percent of the time.**
- As this project, I used 80% of the data set to train the model and 20% for testing.
- For every method utilized, the following metrics were assessed: accuracy, AUC, F1-score, precision, and recall. The algorithm that performed the best in classifying the data was identified based on these metrics.

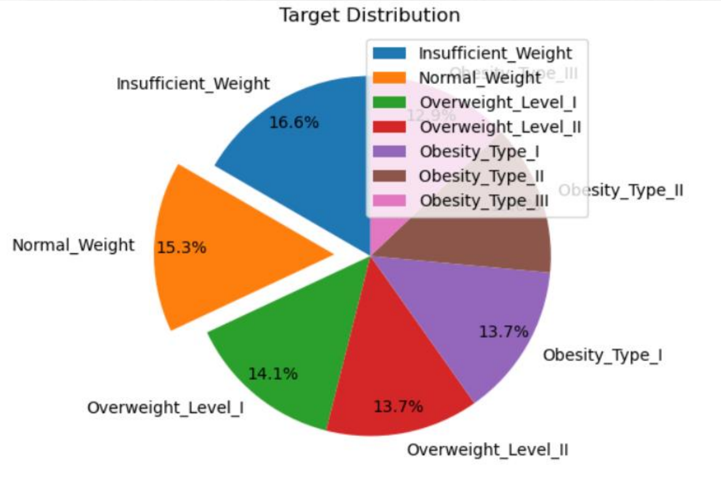


Figure: Dependent variable pie chart

Challenges:

- Unlike linear regression, the logistic regression algorithm just requires that the data be distributed normally. With the exception of the Number of Meals variable, all continuous attributes in the project were transformed to a normal distribution.
- Outliers were found in the dataset's columns for age, height, weight, and number of meals.
- The test set was being overfitted by practically every method.

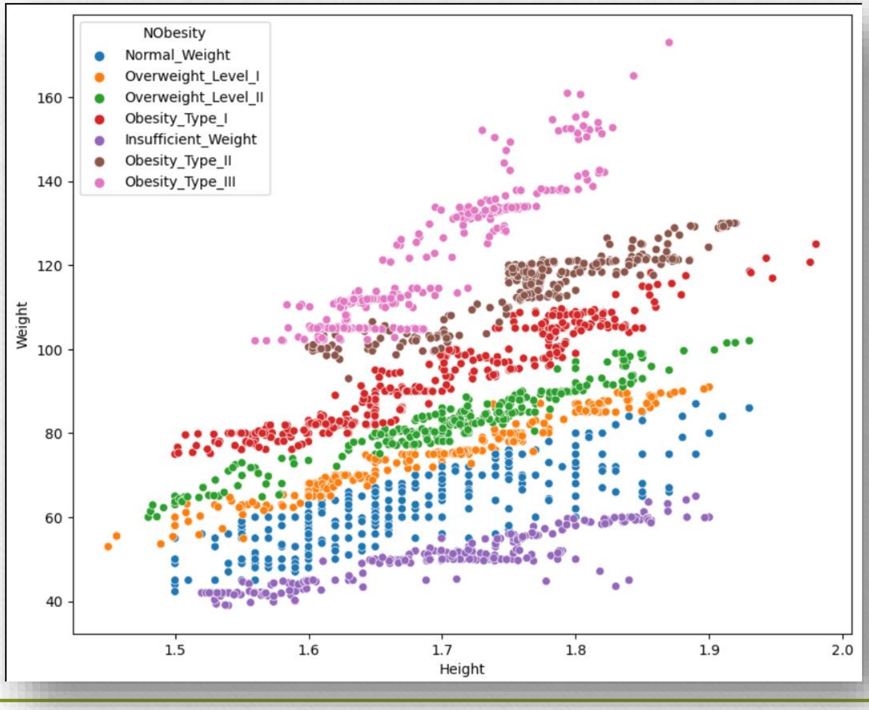


Figure: Weight vs. Height scatter plot

Advantages:

- This project uses both supervised and unsupervised classification approaches to provide a system based on machine learning algorithms.
- We have all seen the COVID-19 pandemic, and people are now more concerned about their health than they were previously. Additionally, there is a growing trend of online fitness and health tracking applications. Therefore, internet platforms may find this strategy quite helpful in contacting new clients.
- In order to reduce cardiovascular risk, this model can be very helpful when conducting online surveys and determining the prevalence of obesity in the population.

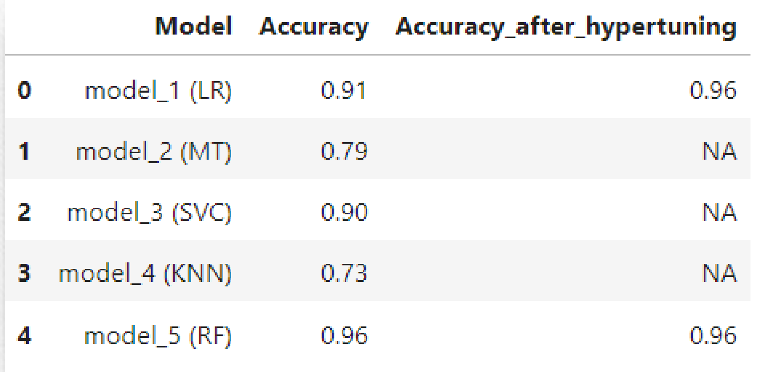
Disadvantages:

- This model was developed using data from surveys conducted in Colombia, Peru, and Mexico. It focuses on a specific demographic. Therefore, this model may not work as well in some areas.

Learnings:

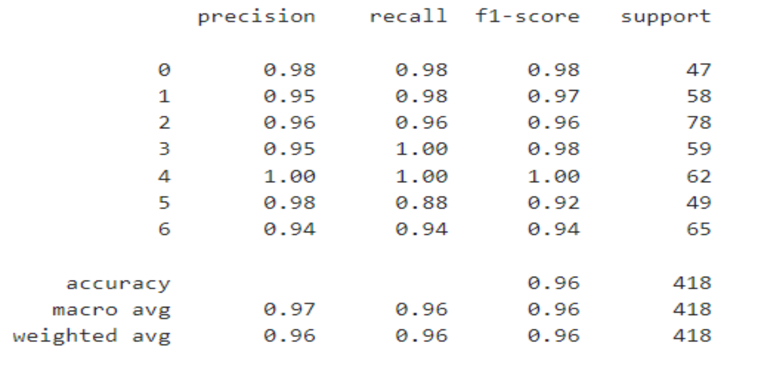
- Effective Time Management
- The Value of Preparing Data
- ML algorithm selection
- Hyper-tuning is important.

Result:



	Model	Accuracy	Accuracy_after_hypertuning
0	model_1 (LR)	0.91	0.96
1	model_2 (MT)	0.79	NA
2	model_3 (SVC)	0.90	NA
3	model_4 (KNN)	0.73	NA
4	model_5 (RF)	0.96	0.96

Figure- Result Table of all 5 models



```
              precision    recall  f1-score   support

           0       0.98      0.98      0.98        47
           1       0.95      0.98      0.97        58
           2       0.96      0.96      0.96        78
           3       0.95      1.00      0.98        59
           4       1.00      1.00      1.00        62
           5       0.98      0.88      0.92        49
           6       0.94      0.94      0.94        65

    accuracy                           0.96       418
   macro avg       0.97      0.96      0.96       418
weighted avg       0.96      0.96      0.96       418
```

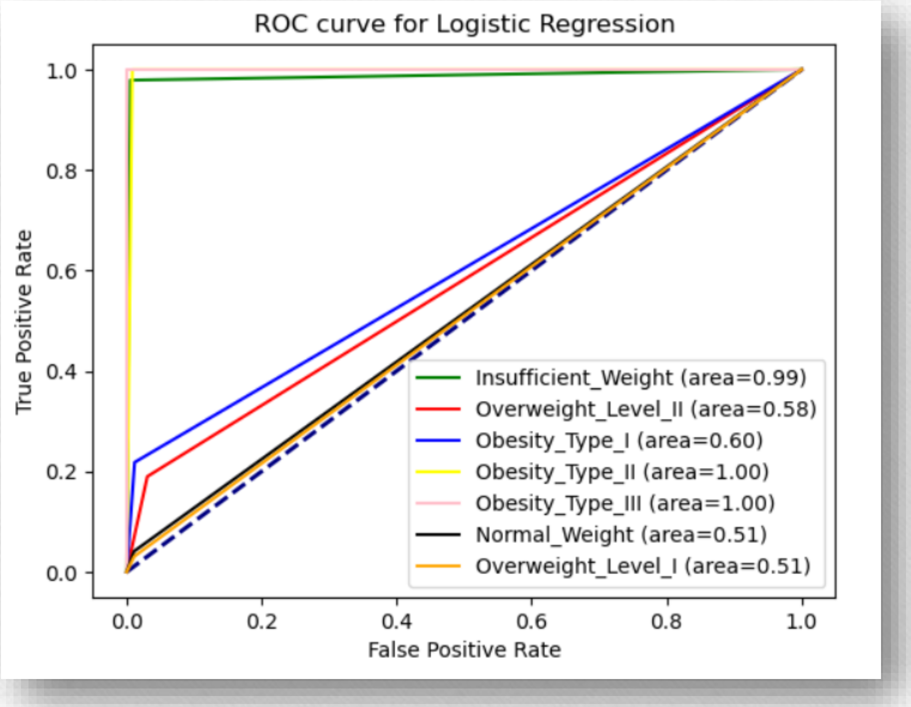
Figure- Classification Report of hyper tuned logistic regression model

- Models 1 and 5—that is, the Random Forest Classifier and Logistic Regression—performed better in predicting obesity status, as indicated in the data frame above.

- Prior to hypertuning, only logistic regression produced a generalized model with an accuracy of 91%; however, hypertuning increases this accuracy to 96%.

- In the case of Random Forest, the model is generalized with 96% accuracy following hypertuning.

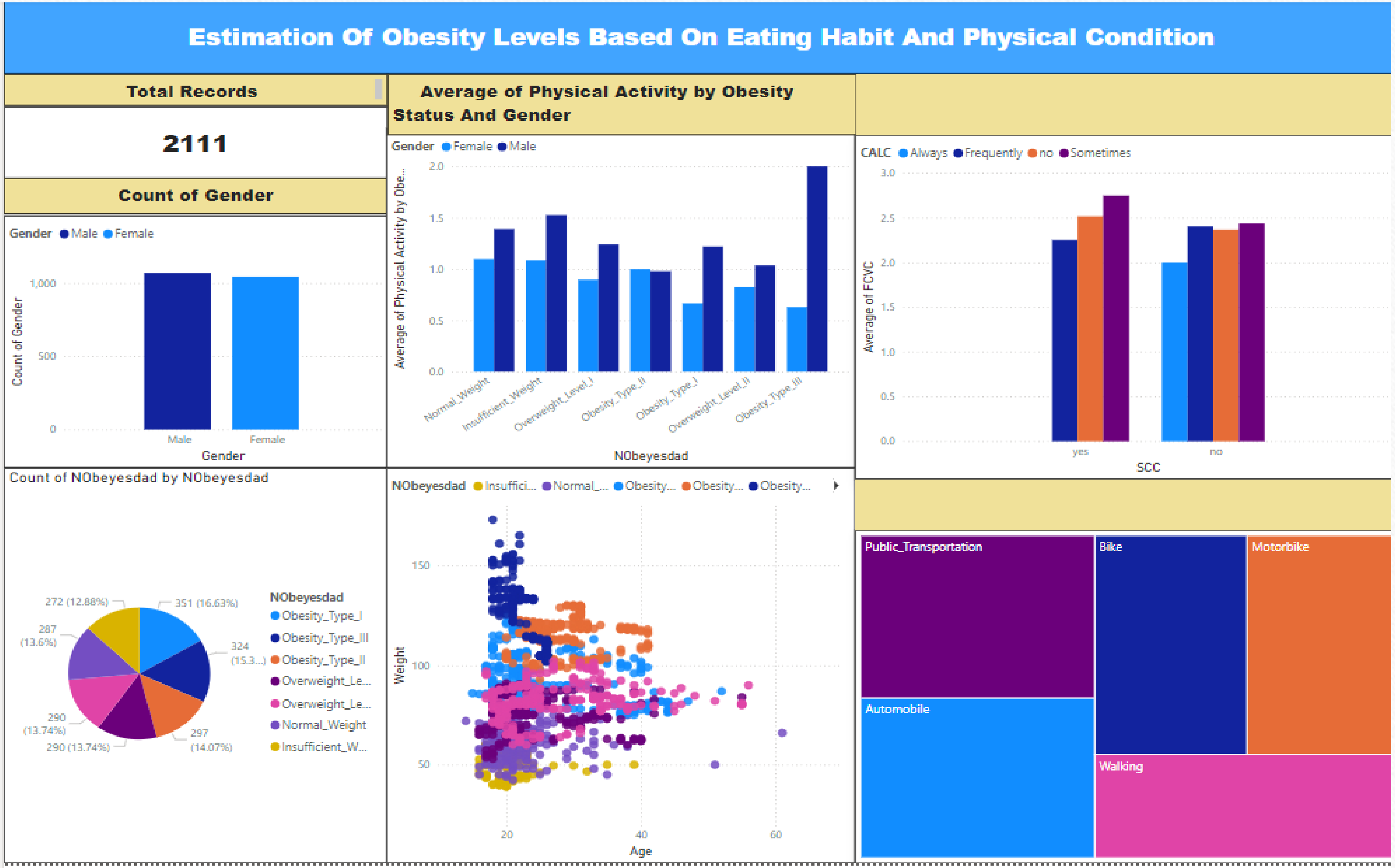
Result:



- ❖ With an accuracy of 96%, the models of random forest and logistic regression did quite well with this data.

- ❖ Based on the Logistic Regression ROC curve above, we may conclude that the model is more effective at predicting values for Insufficient Weight, Obesity Type II, and Obesity Type III.

Dashboard:



Estimation Of Obesity Levels Based On Eating Habit And Physical Condition

Total Records

2111

Count of Gender

Average of Physical Activity by Obesity Status And Gender

Count of NObeyesdad by NObeyesdad

- Link to my zipped file: https://drive.google.com/file/d/13FYh_oUg_ub_d9f-J3kGqheaOEg4IqB3/view?usp=drive_link

Many thanks for this greatest opportunity offered