FitVista

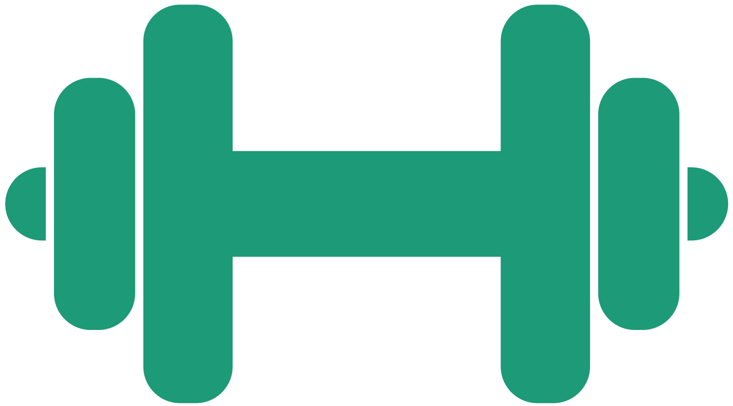
Sculpting Wellness, Shaping Lives

Group-14:
Sai Praneeth Veerapalli
Shubhang yadav Sandaveni
Swapna Vippaturi

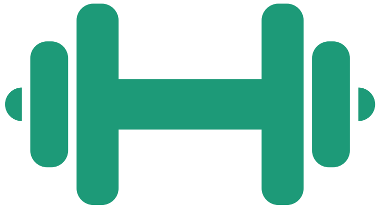
Contents

Executive Summary

Creating a predictive obesity model can help fitness professionals offer personalized solutions for weight management, attracting more clients looking for tailored fitness plans.



Background



The fitness industry has a big chance to thrive due to the growing focus on health and the rise of obesity-related health concerns

This aligns with the increasing demand for data-driven and personalized fitness solutions.

Core Team & Stakeholders

Our Role: Member of the Data Science team

Responsibility: Design and enhance the predictive obesity classification model

Internal Stakeholders: Collaborate with data engineers, and project managers

External Beneficiaries:

Fitness Professionals: Utilize the model for tailored client services

Healthcare Providers: Leverage insights for proactive health management

Individuals Seeking Fitness Guidance: Receive personalized recommendations

Scope and Objective

Project Deliverables:-

- **Obesity Classification Model**

Deliver a robust machine learning model for classifying clients into obesity risk categories based on lifestyle and behavioral factors.

- **Actionable Reports**

Provide regular, easy-to-interpret reports to stakeholders, offering insights into clients' obesity risk classifications and guiding professionals in tailoring fitness plans.

Future Scope (Not Delivering):-

- **Personalized Plans:** The project currently needs to include the delivery of personalized fitness plans. Recommending tailored plans is identified as a future scope for the pipeline.

Project Development Timeline

5–8 Oct.

Project Proposal (October 5-8, 2023): This initial phase was focused on defining the project scope, objectives, and deliverables.

9–16 Oct.

Data Collection (October 9-16, 2023): This stage involved gathering the necessary data for your project.

18–24 Oct.

Data Preprocessing (October 18-24, 2023): During this phase, the collected data was cleaned, transformed, and prepared for analysis.

24 Oct. – 2 Nov. 2023

Model Development (October 24 - November 2, 2023): Here, the focus was on developing the model based on the preprocessed data.

2–8 Nov.

Model Testing (November 2-8, 2023): This phase involved rigorously testing the model to ensure its accuracy and effectiveness.

9–14 Nov.

Performance Analysis (November 9-14, 2023)

Future Scope

Recommendation Systems

Development strategy

Complex Patterns: Classification leverages supervised learning to recognize complex patterns in lifestyle and behavioral data, enabling accurate identification of obesity risk categories.

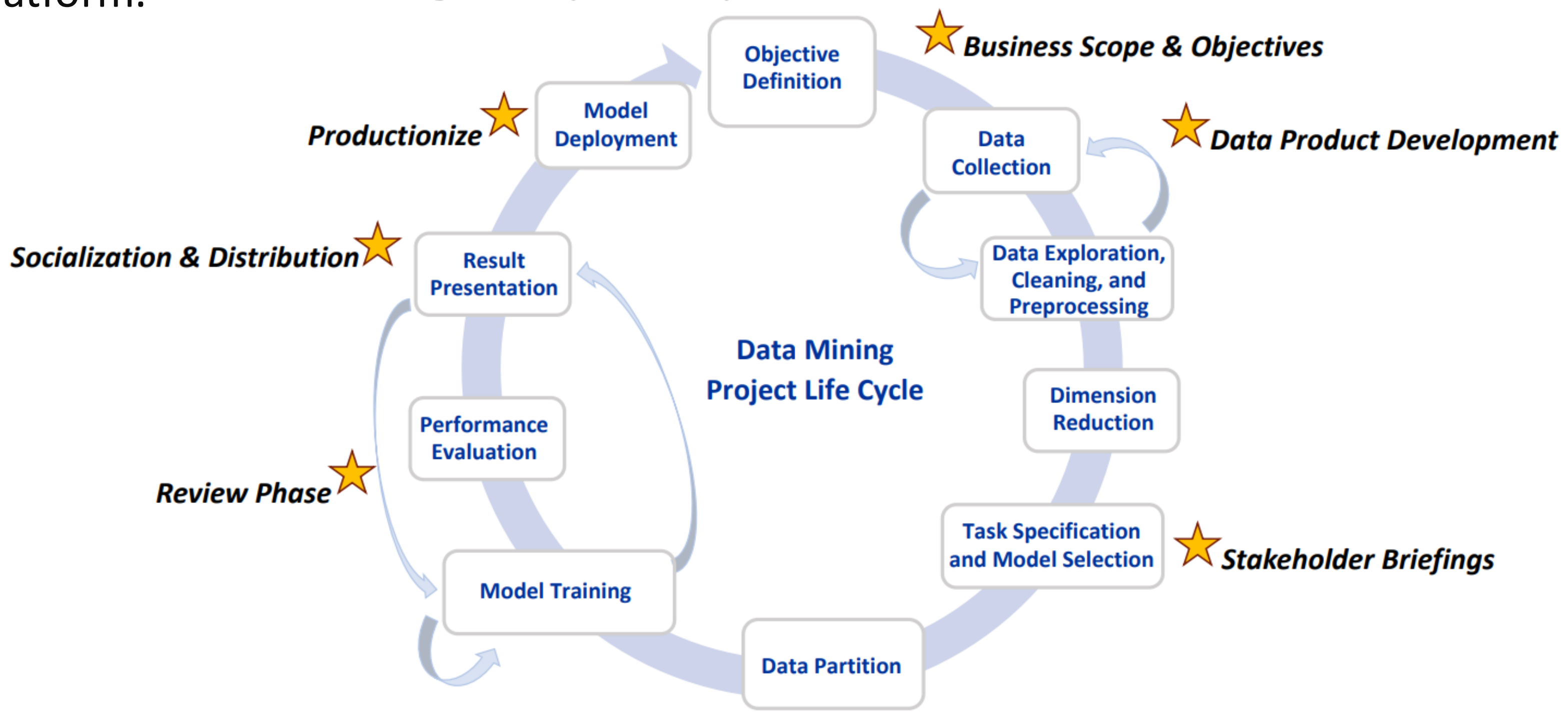
Personalized Interventions: By employing classification, we can tailor fitness plans based on individual risk levels, optimizing the effectiveness of interventions for weight management.

Actionable Insights: Classification provides actionable insights for fitness professionals, facilitating precise guidance and enhancing the overall impact on clients' health and wellness.

Development Details

This dataset comprises 2,111 entries and 17 features related to individuals' eating habits and physical condition from Mexico, Peru, and Colombia to estimate obesity levels. It includes seven classifications of weight status from Insufficient Weight to Obesity Type III. The majority (77%) of the data was synthesized using Weka and SMOTE, while the remainder (23%) was obtained directly from user inputs on a web platform.

Data Source: Estimation of obesity levels based on eating habits and physical condition. (2019). UCI Machine Learning Repository. https://doi.org/10.24432/C5H31Z.



Data Dictionary: Dataset Description: Obesity Dataset

The dataset contains information related to obesity, including demographic details, lifestyle factors, and health metrics. Each row in the dataset represents an individual.

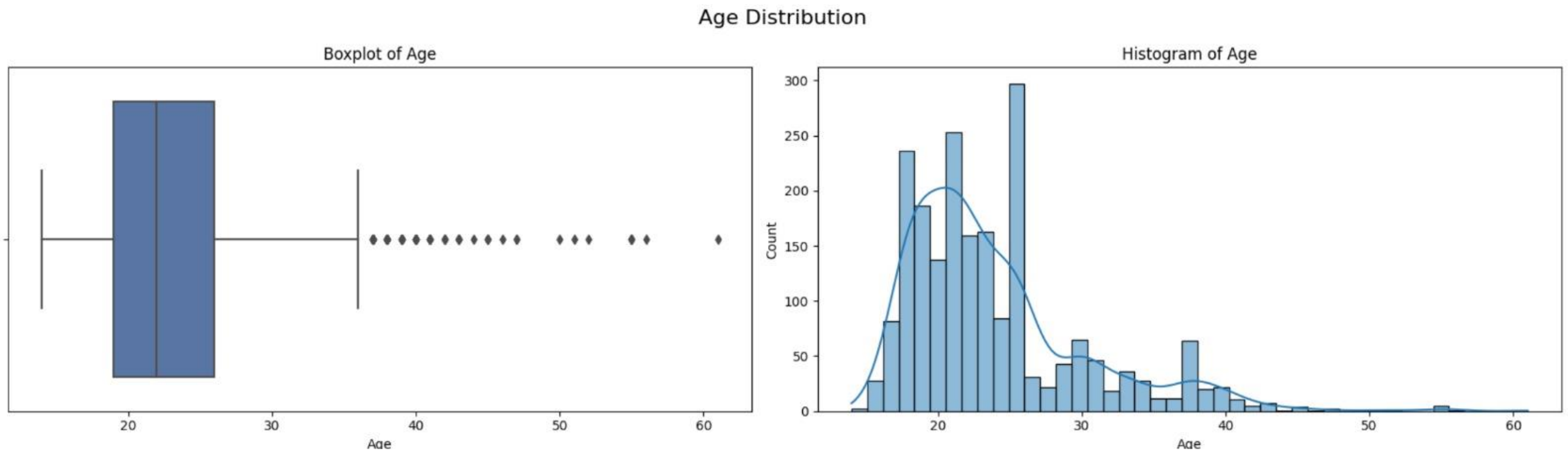
- **Gender:** Gender of the individual (Female/Male)
- **Age:** Age of the individual (Numeric value)
- **Height:** Height of the individual in meters (Numeric value)
- **Weight:** Weight of the individual in kilograms (Numeric value)
- **FamilyHistory_Overweight:** Whether a family member has suffered or currently suffers from overweight (Yes/No)
- **FAVC:** Frequency of consuming high-caloric food (Yes/No)
- **FCVC:** Frequency of eating vegetables in meals (Never/Sometimes/Always)
- **NCP:** Number of main meals consumed daily (Between 1 and 2/Three/More than three)
- **CAEC:** Whether the individual eats any food between meals (No/Sometimes/Frequently/Always)
- **SMOKE:** Smoking habit (Yes/No)
- **CH2O:** Daily water intake (Less than a liter/Between 1 and 2 L/More than 2 L)
- **SCC:** Whether the individual monitors daily calorie intake (Yes/No)
- **FAF**: Frequency of physical activity (I do not have/1 or 2 days/2 or 4 days/4 or 5 days)
- **TUE:** Time spent on technological devices (0 to 2 hours/3 to 5 hours/More than 5 hours)
- **CALC:** Frequency of alcohol consumption (I do not drink/Sometimes/Frequently/Always)
- **MTrans:** Preferred mode of transportation (Automobile/Motorbike/Bike/Public Transportation/Walking)

$$Mass\ body\ index = \frac{Weight}{height*height}$$

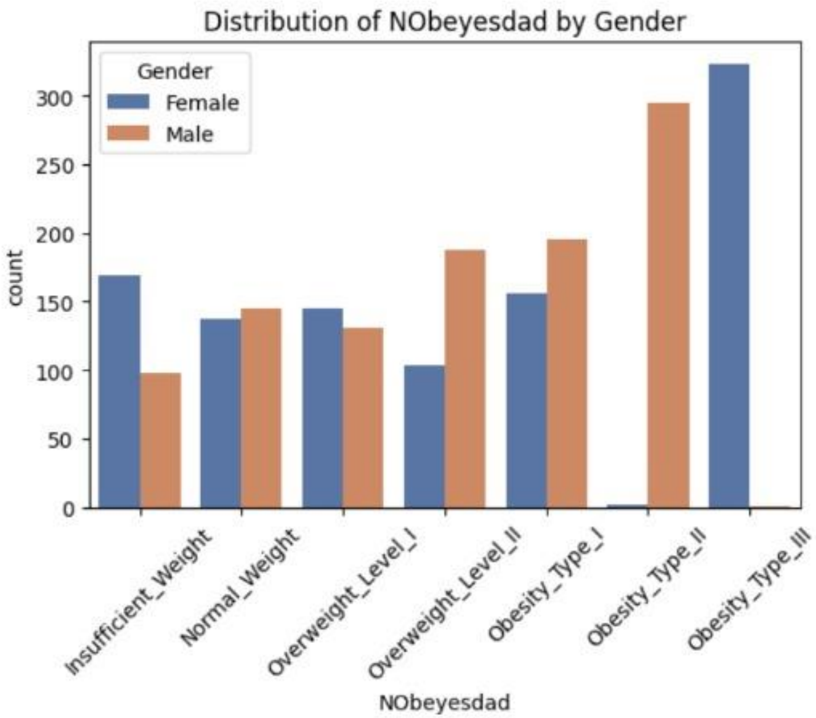
- Underweight Less than 18.5
- Normal 18.5 to 24.9
- Overweight 25.0 to 29.9
- Obesity I 30.0 to 34.9
- Obesity II 35.0 to 39.9
- Obesity III Higher than 40

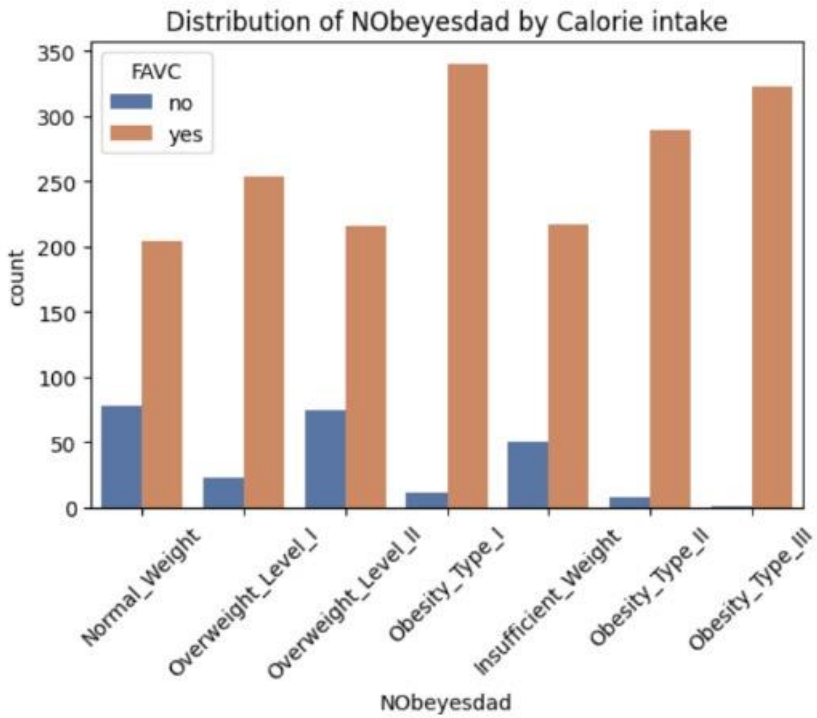
	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObeyesdad
0	Female	21.000000	1.620000	64.000000	yes	no	2.0	3.0	Sometimes	no	2.000000	no	0.000000	1.000000	no	Public_Transportation	Normal_Weight
1	Female	21.000000	1.520000	56.000000	yes	no	3.0	3.0	Sometimes	yes	3.000000	yes	3.000000	0.000000	Sometimes	Public_Transportation	Normal_Weight
2	Male	23.000000	1.800000	77.000000	yes	no	2.0	3.0	Sometimes	no	2.000000	no	2.000000	1.000000	Frequently	Public_Transportation	Normal_Weight
3	Male	27.000000	1.800000	87.000000	no	no	3.0	3.0	Sometimes	no	2.000000	no	2.000000	0.000000	Frequently	Walking	Overweight_Level_I
4	Male	22.000000	1.780000	89.800000	no	no	2.0	1.0	Sometimes	no	2.000000	no	0.000000	0.000000	Sometimes	Public_Transportation	Overweight_Level_II
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2106	Female	20.976842	1.710730	131.408528	yes	yes	3.0	3.0	Sometimes	no	1.728139	no	1.676269	0.906247	Sometimes	Public_Transportation	Obesity_Type_III
2107	Female	21.982942	1.748584	133.742943	yes	yes	3.0	3.0	Sometimes	no	2.005130	no	1.341390	0.599270	Sometimes	Public_Transportation	Obesity_Type_III
2108	Female	22.524036	1.752206	133.689352	yes	yes	3.0	3.0	Sometimes	no	2.054193	no	1.414209	0.646288	Sometimes	Public_Transportation	Obesity_Type_III
2109	Female	24.361936	1.739450	133.346641	yes	yes	3.0	3.0	Sometimes	no	2.852339	no	1.139107	0.586035	Sometimes	Public_Transportation	Obesity_Type_III
2110	Female	23.664709	1.738836	133.472641	yes	yes	3.0	3.0	Sometimes	no	2.863513	no	1.026452	0.714137	Sometimes	Public_Transportation	Obesity_Type_III

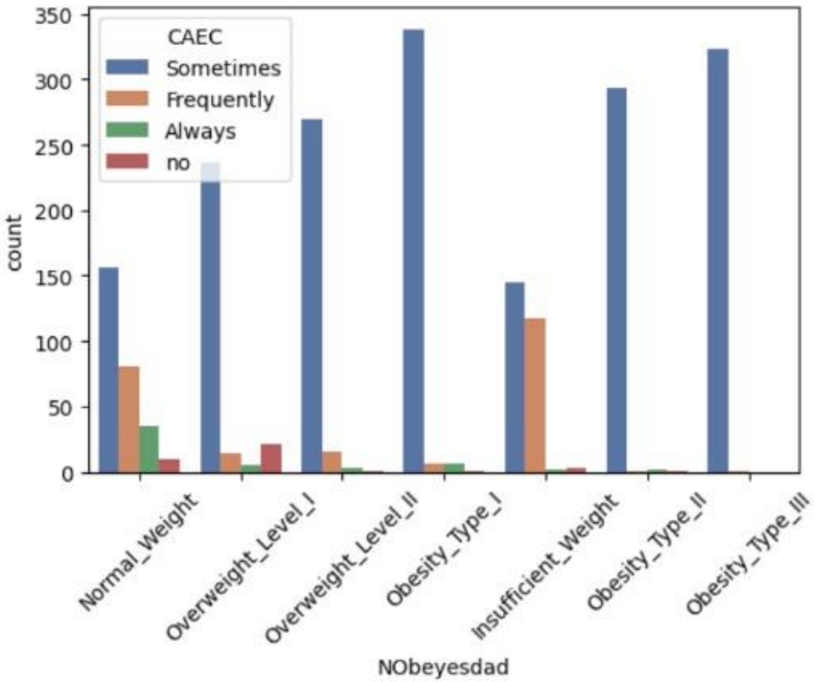
2087 rows × 17 columns



The outliers are observed in individuals aged 35 years or older; however, the majority of the dataset comprises individuals aged between 18 to 40. The histogram distribution is right-skewed so the mean age > median age.





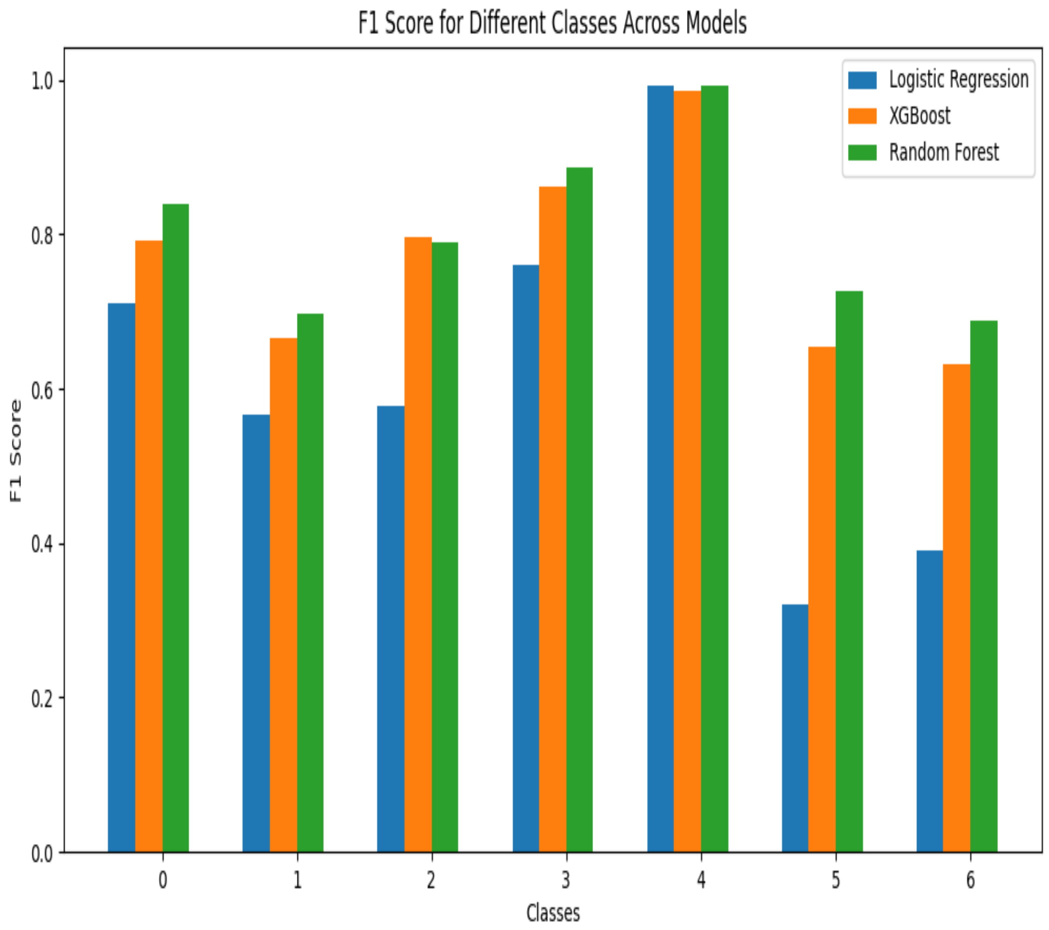


- Gender: the distribution across different obesity levels is similar between genders, but the number of women in 'Obesity_Type_III' and the number of men in ‘Obesity_type_II’ are higher.

- High-calorie intake: The bar plot shows that individuals who frequently consume high-calorie foods are more likely to fall into higher obesity categories. This shows a potentially strong correlation between high-calorie food consumption and obesity levels.

- This plot shows that those who eat between meals 'Sometimes' or 'Frequently' are more likely to have 'Normal_Weight' or 'Insufficient_Weight’. However, people who are 'Always' eating between meals are correlated with higher levels of obesity.

Results



<u>Class-Specific Performance: Models vary in performance across classes, indicating specific strengths for each.</u>

Random Forest: Random Forest model is well-balanced and effective at classifying all classes, suggesting its robustness and reliability in handling multi-class problems.
Logistic Regression: Logistic Regression performance drops noticeably, indicating that it may not handle complex, non-linear relationships as effectively.
XGBoost: XGBoost model performs quite well for some classes, it falls short for others, suggesting that the model may have overfitted to a particular class
Model Selection: Given the overall F1 scores, the Random Forest model is the most reliable choice for this classification task

Value

Data-Driven Insights into Obesity Levels

Preventive Healthcare

Real-World Integration

Future Scope

Targeted Health Programs

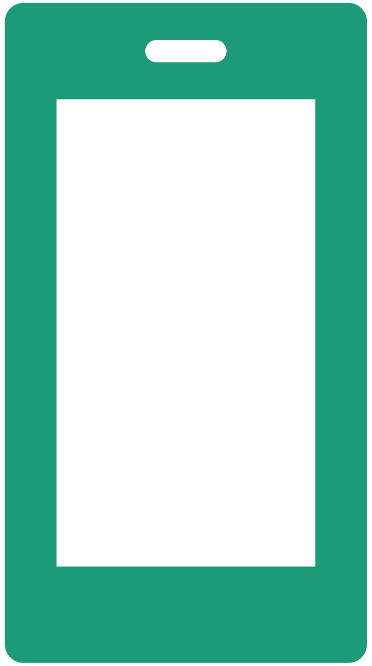
Expand Dataset

Personalized Fitness Plans & Dietary Guidance

Socialization & Distribution

Mobile Applications:

- Develop mobile applications for on-the-go access, enabling fitness professionals and individuals to view real-time updates and recommendations.
- Integrate push notifications for timely alerts and engagement



Thank You

- Let's commit to staying active, eating right, and living healthy.