# Obesity Level Prediction Project Report

## 1. Introduction

This project analyzes the obesity dataset to predict obesity levels based on demographic, lifestyle, and health-related features. It includes exploratory data analysis (EDA), feature engineering, data visualization, and machine learning models (Logistic Regression and Random Forest).

## 2. Dataset Overview

The dataset (obesity\_level.csv) contains information such as gender, age, height, weight, family history, dietary habits, physical activity, and transport. The target variable ('0be1dad') represents obesity levels, including categories like Insufficient Weight, Normal Weight, Overweight Levels, and Obesity Types.

## 3. Data Preprocessing

- Loaded the dataset using pandas  
- Checked for missing values and duplicates  
- Cleaned categorical columns (stripping spaces, lowercasing)  
- Created new features such as BMI and frequent\_fastfood indicator

## 4. Exploratory Data Analysis

- Distribution of obesity levels using countplots  
- Boxplots for BMI vs obesity level, physical activity vs obesity level  
- Correlation heatmap to check relationships among numerical features

## 5. Machine Learning Models

Two models were implemented and evaluated:  
  
1. Logistic Regression  
 - Used as a baseline classifier  
 - Evaluated using classification report and confusion matrix  
  
2. Random Forest Classifier  
 - Applied with 100 estimators and class balancing  
 - Provided improved accuracy compared to Logistic Regression

## 6. Results

The Random Forest model achieved better classification performance across most obesity levels. Confusion matrices and classification reports highlight the strengths and weaknesses of both models.

## 7. Conclusion

This project demonstrates how health and lifestyle factors can be used to predict obesity levels. Random Forest provided the best results, showing the importance of ensemble learning for imbalanced multiclass problems. Future work may include hyperparameter tuning, feature selection, and testing other models like Gradient Boosting or Neural Networks.