# Multinomial Logistic Regression Report

## 1. Introduction

Dataset: Obesity Levels Dataset  
  
Source: Kaggle  
URL: <https://www.kaggle.com/datasets/fatemehmehrparvar/obesity-levels>  
  
This dataset contains information related to obesity levels based on various lifestyle and medical attributes. The goal is to classify individuals into different obesity levels using multinomial logistic regression.

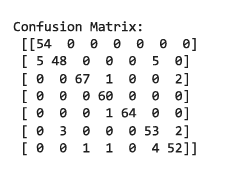
## 2. Data Preprocessing

Step 1: Load the Dataset & Inspect Structure  
The dataset was loaded and inspected for missing values.  
  
Dataset Shape: 2111 rows × 17 columns  
No missing values found.  
  
Step 2: Encoding Categorical Variables  
Categorical variables were encoded using Label Encoding.  
  
Step 3: Feature Scaling  
Numerical features were standardized using StandardScaler, except for the target(because it was a label from 1 to 7) variable.  
  
Step 4: Data Splitting  
The dataset was split into 80% training and 20% testing sets.

## 3. Model Implementation

Algorithm Used: Multinomial Logistic Regression  
Library: scikit-learn  
  
Hyperparameter Tuning: Used GridSearchCV to optimize parameters such as:  
- C (Regularization strength)  
- Solver (Optimizer)  
- Max\_iter (Maximum iterations)  
  
Best Parameters Found: {'C': 10, 'max\_iter': 300, 'solver': 'lbfgs'}

## 4. Results

Model Evaluation  
Accuracy: 0.9858 (98.58%)  
  
Confusion Matrix:  
  
  
Classification Report:  
A screenshot of a graph

AI-generated content may be incorrect.

**Visualizations:**  
1. Confusion Matrix Heatmap

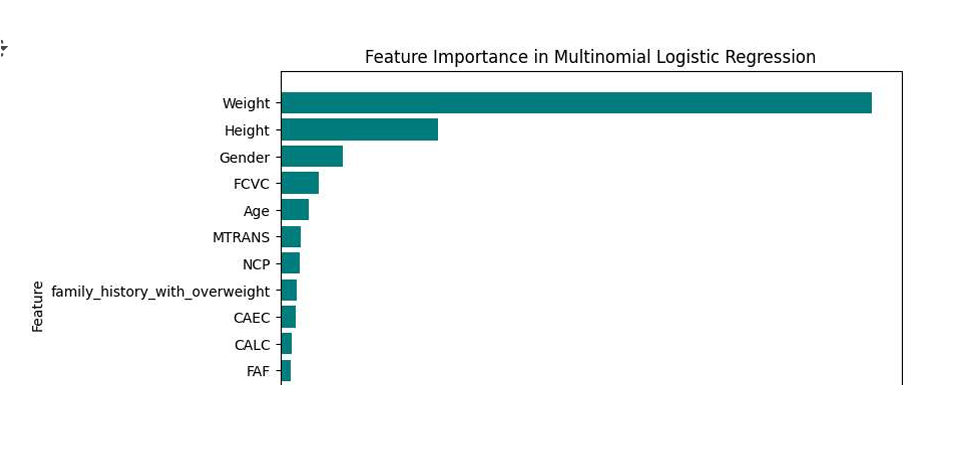
A diagram of heatmap

AI-generated content may be incorrect.  
2. ROC Curves for Each Class

A graph of a roc curve

AI-generated content may be incorrect.

3. Feature Importance Plot



## 5. Discussion

**Interpretation of Results**  
- The model achieved a high accuracy of 98.58%, indicating strong predictive performance.  
- The confusion matrix shows that most predictions were correct, with minimal misclassification.  
- The classification report highlights strong precision, recall, and F1-score across all classes.

## 6. Conclusion

The multinomial logistic regression model successfully predicted obesity levels with high accuracy.  
The confusion matrix and classification report confirmed strong classification performance.  
Future improvements can focus on handling class imbalance and exploring additional machine learning models.