Christian Kichian
Jason Hkayem
Boulos Aazar

USEK
HOLY SPIRIT UNIVERSITY OF KASLIK

Obesity Prediction Project Report

Purpose

The purpose of this project is to predict the obesity classification of individuals based on selected lifestyle and dietary features. This is intended to support public health analysis and provide useful insight into factors influencing body weight categories.

Data Used

The dataset used is `Obesity Data Set.csv`, which contains anonymized survey data on various personal, behavioral, and lifestyle factors. It includes features such as Weight, frequency of consuming food, and habits like family history of overweight and physical activity.

Outcome

The goal is to classify each individual into a specific obesity category labeled in the `NObeyesdad` column. For clarity, we limited the outcome to:

1. Insufficient Weight
2. Normal Weight
3. Overweight
4. Obese

This grouping consolidates multiple subcategories (there are various levels of obesity and overweight) into broader, more interpretable classes for analysis and public health insights.

Preprocessing

Selected the following features for prediction:
 - Weight
 - family_history_with_overweight
 - CAEC (eating between meals)
 - FAVC (frequent high-calorie food consumption)
 - FCVC (vegetable consumption frequency)
- Used one-hot encoding to convert categorical variables into numerical format
- Scaled all numeric features to a 0–1 range using MinMaxScaler

Machine Learning

- A `RandomForestClassifier` from scikit-learn was used
- The model was trained using the preprocessed feature matrix `X` and target labels `y`
- Data was not split into train/test because the focus was on preparing an integrated, in-notebook predictive model

Christian Kichian
Jason Hkayem
Boulos Aazar

Results

The model was successfully trained and is now ready for use directly within the notebook. Predictions can be made by preprocessing new inputs using the same encoding and scaling steps, then passing them into `model.predict()`.