

# Project: Obesity Prediction Analysis – ML

Author: Miro Zilaji, 9.5.2025

## Core Data Science Statement (PIKE) – Obesity Prediction Analysis

### (P) Problem – Relevant Question

Which personal lifestyle and demographic factors are most predictive of obesity levels in individuals, and how can we build a reliable machine learning model to classify people into specific obesity categories?

This question is essential for public health organizations and healthcare providers to identify risk patterns early and guide preventive interventions.

---

### (I) Intervention – Primary Approach

I used **pandas**, **NumPy**, **matplotlib**, and **seaborn** for data preprocessing and visualization.

The classification problem is addressed using a machine learning pipeline with the **AdaBoost ensemble algorithm**, applied with several weak learners including:

- **Decision Trees** (max\_depth 1, 3, 5),
- **Logistic Regression**,
- **Support Vector Classifier (SVC)**.

The models were trained on standardized features, and categorical variables were one-hot encoded. Model performance was evaluated using **accuracy**, **classification reports**, and **feature importance** analyses.

---

### (K) Kontrollintervention – Alternative Methods

As a control strategy, I trained a **Random Forest Classifier** to analyze feature importance independently and evaluate its standalone performance.

This provided a benchmark to compare against the ensemble-based AdaBoost classifiers, helping to validate whether boosting and different weak learners improved classification results.

---

### (E) Evidence – Target Outcome

The goal is to develop a **multi-class classifier** to predict obesity levels categorized in the `no_beyesdad` column.

The outcome is measured using:

- **Accuracy score** on test data,
- **Precision, Recall, F1-score** for each class,
- Visualization of **confusion matrices** and **feature importances**,
- Export of evaluation metrics to CSV for further interpretation and reporting.

This ensures interpretability, reproducibility, and actionable insights into key risk factors such as age, calorie intake, and physical activity.