Calories Burnt Prediction - Project Report

# 1. Data Description

The dataset used in this analysis contains 15000 entries and 9 features related to physical activities. The main features include age, gender, height, weight, duration of the activity, heart rate, and body temperature. The target variable is the number of calories burnt during the physical activity. Any missing values were handled during preprocessing to ensure a clean dataset for model training and evaluation.

# 2. Objectives

The main objective of this analysis is to predict the number of calories burnt during a physical activity based on multiple features, such as age, weight, duration of the exercise, and heart rate. The goal is to create a predictive model that accurately estimates calories burnt, which can be useful in fitness tracking applications.

# 3. Variations of Linear Regression Models

Several machine learning models were considered for this analysis, with a primary focus on regression models. While linear regression and Lasso regression were tested, the XGBoost Regressor (XGBRegressor) provided the best performance in terms of accuracy and error metrics. XGBRegressor is a powerful gradient boosting algorithm that handles complex relationships between the features and the target variable better than simple linear regression models.

# 4. Key Findings

The XGBRegressor model achieved excellent results, with an R-squared value of 0.9988 and a Mean Absolute Error (MAE) of 1.483. This indicates that the model fits the data very well and produces accurate predictions of calories burnt. Given the target variable's range (100-200), the low MAE suggests that the model is highly effective at predicting calories burnt.

# 5. Model Flaws and Future Action Plan

Despite the strong performance of the model, there are a few potential areas for improvement. First, the model may be overfitting due to the high R-squared value. Further cross-validation could help determine if the model generalizes well to new data. Additionally, feature engineering may enhance the model's performance, as some features may not have been fully optimized. In future iterations, alternative models such as Random Forest or Neural Networks could be explored to compare performance. Moreover, gathering more diverse data (e.g., varying types of physical activities) could improve the robustness of the model.

Tap [**here**](https://github.com/rufusclieve/Calories-Burnt-Prediction/tree/main) to check the project