Analyzing Fitness Patterns and Building Predictive Models using Gym Members Exercise Dataset

Team 27- tc3117, em3907, ap4613, tz2634, jy3344

Background & Problem Statement

In an era where sedentary lifestyles are increasingly prevalent, understanding and promoting effective fitness behaviors has become crucial for public health. Prior research has shown that physical activities significantly contribute to wellness outcomes across diverse age groups. As such, we aim to seek valuable insights by exploring the relationship between gym habits and their impact on overall health. By selecting a relevant dataset and applying appropriate machine learning techniques, we hope to (1) uncover key factors that influence health behaviors and outcomes and (2) propose effective fitness plans and strategies.

Given the comprehensive dataset of gym members' exercise routines, physical attributes, and fitness metrics, we seek to explore the following questions by leveraging machine learning techniques:

1. What are the most significant factors influencing fitness outcomes across different demographic groups?
2. How can we develop predictive models for personalized exercise recommendations that are tailored to individual characteristics and fitness goals?
3. What patterns in workout adherence and progression can be uncovered to inform strategies for improving long-term engagement with fitness routines?

Dataset: Gym Members Exercise Dataset (Kaggle)

The data consists of 973 samples of member's demographic and physiological attributes, providing a comprehensive view of their workout habits. It includes 15 columns such as age, gender, weight, height, maximum beats per minute (BPM), average BPM, resting BPM, workout session duration, calories burned, workout type, body fat percentage, daily water intake, workout frequency, experience level, and body mass index (BMI). These features offer valuable insights into fitness routines and health trends, enabling the exploration of relationships between various physiological factors and workout performance.

With this dataset, we can build machine learning models to predict fitness progression and key performance indicators. For instance, a regression analysis could predict calorie expenditure, while a classification task could determine the workout type based on demographic and physiological data.

Exploratory data analysis (EDA) will be performed to understand feature distributions, correlations, and trends, as well as identify outliers or missing data. Afterward, necessary preprocessing techniques will be applied. For example, we can analyze the correlation between physiological metrics such as weight, height, bpm, body fat percentage, and BMI to derive insights into how these factors impact fitness outcomes.

Proposed ML Techniques

To analyze the Gym Members Exercise Dataset, both supervised and unsupervised machine learning techniques can be applied for actionable insights:

Linear regression models the relationship between continuous target variables and predictor features. For example, it can predict calories burned or average heart rate based on factors such as age, weight, and workout frequency. Potential insights can be gained on how workout, body composition, and demographics affect energy expenditure. The model's interpretability makes it valuable for understanding the individual influence of features on target variables.

Gradient boosting is an ensemble technique that captures non-linear relationships between features. In this dataset, it can be used to predict outcomes such as calories burned or average heart rate, where multiple features interact in complex ways. Gradient boosting provides a more accurate and nuanced prediction than linear regression, making it suitable for understanding intricate fitness patterns.

Clustering, such as k-means and agglomerative clustering, can uncover hidden patterns by segmenting the data into groups based on features like workout type, session duration, experience level, and BMI. This approach will help categorize gym members into distinct fitness profiles, revealing common exercise patterns and supporting personalized fitness recommendations.

After building these models, we can rank features importance to identify the most significant factors influencing fitness outcomes across different demographic groups. Exercise recommendations can be done by multi-class classification on the workout type. Additionally, we could further do data augmentation using heuristic methods or association rule mining to uncover patterns specific to each individual.