Abstract

This project aims to predict the type of workout performed by gym members based on physiological and behavioral data such as heart rate, BMI, and session duration. Using a dataset from Kaggle, we perform data preprocessing, feature scaling, and model training. Several machine learning models (Decision Tree, Random Forest, and Support Vector Machine) are evaluated to determine the best-performing approach. The goal is to help gym users and trainers understand workout patterns and optimize exercise routines.

Related Work

Previous studies have applied machine learning techniques to analyze fitness and health-related data. For example, researchers have used wearable sensor data to predict calorie expenditure, heart rate patterns, and exercise types. Projects such as “Human Activity Recognition” using accelerometer data have demonstrated the effectiveness of Random Forest and SVM models for classification tasks. Our work follows a similar approach but focuses on gym-specific metrics like session duration, water intake, and experience level to predict workout type.

Planned Methodology

The methodology involves several stages:

1. **Data Preprocessing:** Handle missing values, encode categorical features (e.g., gender, workout type), and normalize numerical features using StandardScaler.
2. **Exploratory Data Analysis (EDA):** Identify feature correlations, visualize distributions, and detect outliers.

3. **Model Training:** Train and evaluate three supervised learning models — Decision Tree, Random Forest, and Support Vector Machine — using stratified train-test splits.
4. **Evaluation:** Compare models using metrics such as accuracy, F1-score, and confusion matrix to select the optimal classifier.

Proposed Implementation

The project is implemented in Python using libraries such as Pandas, Matplotlib, Seaborn, and scikit-learn. The dataset and scripts are maintained on GitHub for collaboration. The implementation includes a training pipeline and visualization components for insights into user performance and model behavior. Future improvements may include using deep learning or real-time data from wearable devices.