Northeastern University, Khoury College of Computer Science



CS 6220 Data Mining | Project Proposal

Due: Oct. 28, 2024

Chuanzhao Huang

[huang.chua@northeastern.edu](mailto:huang.chua@northeastern.edu)

Wenbo Qian

qian.wen@northeastern.edu

### 

### **Project Proposal: Analyzing Gym Member Exercise Habits to Improve Fitness Services in Schools**

#### **Objective:**

The objective of this project is to analyze gym members’ exercise habits and develop personalized fitness profiles for each group of users. This analysis will maximize users’ usage time and help decide the optimal strategy of equipment setting, thus providing schools with insights to enhance fitness services and increase student engagement in physical exercise.

#### 

#### **Motivation:**

Physical fitness plays a crucial role in students’ overall well-being, yet many schools struggle to provide tailored gym services that meet the diverse needs of their student body. This project aims to bridge this gap by offering data-driven insights into exercise habits, allowing schools to optimize their gym equipment and programs. By improving the relevance of fitness services, we expect to see a higher rate of student participation and better health outcomes. The solution will enable schools to provide personalized gym experiences based on data-driven user profiles, thus improving student satisfaction and fitness levels.

#### 

#### 

#### **Background:**

This problem arises from the increasing demand for personalized experiences in public services, including health and fitness. Schools, in particular, need to cater to a wide variety of student fitness levels and preferences, but typically lack the data and insights necessary to optimize their gym services. This project aims to analyze real-world gym data to understand user behaviors and generate fitness profiles. I will review literature on clustering techniques, user profiling, and exercise habit analysis to inform the methodology and ensure that the chosen algorithms (e.g., clustering, classification) are appropriate for the dataset and goals of the project.

#### **Proposed Approach:**

1. **Data Processing**: Clean and pre-process the dataset from Kaggle to categorize gym members into different user types based on exercise habits and demographic features.
   * Techniques: Data cleaning, feature engineering, normalization
2. **User Profiling**: For each user type, analyze their exercise habits by studying data like workout duration, heart rate, and exercise types. This will help build distinct profiles for different fitness behaviors.
   * Techniques: Clustering (e.g., K-means), Dimensionality reduction (e.g., PCA)
3. **Equipment Matching**: Based on the gathered insights, recommend the most suitable set of gym equipment for each user type, helping schools design personalized workout programs.
   * Techniques: Classification algorithms, Matching algorithms
4. **Evaluation**: Validate the clustering and profiling techniques using silhouette scores, cross-validation, and accuracy of the equipment-matching algorithms.

Key considerations include handling data imbalances, potential noise in heart rate readings, and the wide variety of user behaviors captured in the dataset. The goal is to deliver accurate and actionable insights that can be translated into practical improvements for school gym services.

#### **Value:**

This project will provide schools with valuable insights into the exercise habits of different types of users, allowing them to offer more personalized and effective fitness programs. By understanding which equipment sets work best for each user type, schools can optimize their resources and improve student participation in fitness activities, leading to healthier lifestyles and better student well-being.