Project Writeup: Personalized Activity/Workout Recommendation System

Team: ML Mavericks

Team Members:

- Soumya Bharathi Vetukuri
- Shubham Kothiya
- Yugm Patel

Homework Assignment Week 2: Fractal Clustering

Data Narrative

Primary Research Questions

1. What are the distinct patterns of physical activity among users, and how do these patterns relate to calorie burn and overall activity efficiency?

2. Can we identify optimal activity patterns (golden cluster) that represent the most effective combination of activity intensity, duration, and health outcomes?

1.1 Business Task

- Primary Objective: Identify and analyze effective fitness patterns to provide personalized activity recommendations

- Stakeholders: Fitness app users, health coaches, wellness program managers

- Business Value:

    - Optimize user engagement through targeted activity suggestions

    - Improve health outcomes through data-driven activity planning

    - Enable personalized fitness coaching at scale

1.2 ML Task

- Task Type: Unsupervised Learning (Clustering)
- Focus: Pattern recognition in daily activity data
- Specific Goals:
    1. Cluster users based on activity patterns
    2. Identify golden cluster representing optimal activity patterns
    3. Extract actionable insights from cluster characteristics

2. Dataset Analysis

2.1 Primary Dataset (Downloaded)

- Source: Kaggle FitBit Fitness Tracker Data
- File: dailyActivity_merged.csv(**Link: https://www.kaggle.com/datasets/arashnic/fitbit**)
- Size: 457 records
- Features: 15 columns including:
    - Activity metrics (steps, distance, calories)
    - Activity intensity levels
    - Time distribution across activity types
    - Temporal data (ActivityDate)

2.2 Secondary Dataset (Planned for Scraping)

- Source: Weather API (OpenWeatherMap)
- Target Data:
    - Daily weather conditions
    - Temperature

    - Precipitation
    - Humidity
    - Wind speed
- Purpose: Analyze environmental factors' impact on activity patterns
- Integration Plan: API calls using activity dates to match weather conditions

3. Fractal Clustering Implementation

3.1 Objective Functions

1. Activity Level and Efficiency Score
2. Activity Intensity and Consistency Score

4. Golden Cluster Analysis

4.1 Cluster Characteristics

The golden cluster (Cluster 2) exhibits:

- Size: 44 samples (9.6% of total)
- Average daily metrics:
    - Steps: 2,278
    - Calories: 1,017
    - Active Minutes: 9
    - Sedentary Minutes: 492

4.2 Quality Metrics

- Silhouette Score: 0.391
- SSE: 75.039
- Activity Level Score: 0.391
- Activity Intensity Score: 0.218

5. Conclusions and Insights

1. Activity Patterns:
    - Identified distinct user segments based on activity intensity
    - Golden cluster represents balanced activity pattern
    - Clear correlation between activity type and calorie burn
2. Optimization Opportunities:
    - Target interventions for sedentary users
    - Personalize recommendations based on cluster characteristics
    - Focus on consistency and gradual progression
3. Future Work:
    - Integrate weather data for environmental context
    - Develop personalized recommendation system
    - Implement real-time pattern recognition