Project Writeup: Personalized Activity/Workout Recommendation System

Team: ML Mavericks

Team Members:

- Soumya Bharathi Vetukuri
- Shubham Kothiya
- Yugm Patel

Homework Assignment Week 2 Final: Fractal Clustering V2

Data Narrative

Primary Research Questions

1. What are the optimal activity patterns that lead to better fitness outcomes?
2. How can we identify and characterize ideal user engagement patterns?
3. What features best predict successful fitness activity habits?

1.1 Business Task

- Primary Objective: Maximize user engagement and health outcomes through personalized activity recommendations
- Stakeholders: Fitness app users, health coaches, wellness program managers
- Value Proposition: Optimize user engagement, improve health outcomes, enable personalized coaching

1.2 ML Task

- Primary Focus: Pattern Recognition via Clustering
- Approach: Unsupervised learning to identify optimal activity patterns
- Goal: Discover natural groupings in user behavior and identify golden patterns

EDAV Insights

- Data Structure: Time-series activity data
- Features: 15 columns including activity metrics
- Distribution: Right-skewed activity patterns
- Correlations: Strong relationship between steps and active minutes

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

3.2 Fractal Distance Implementation

Replaced Euclidean distance with Fractal Distance in k-means:

3.3 Distribution Analysis

Analysis showed:

- Right-skewed distributions for activity metrics
- Non-normal distributions ($p < 0.05$ for all features)
- Strong correlation between TotalSteps and VeryActiveMinutes (0.70)
- Hierarchical structure in cluster dendrogram

3.4 Advanced Fractal Clustering

Implemented HDBSCAN for improved clustering:

- Handles varying density clusters
- Adapts to non-spherical shapes
- Robust to outliers
- Automatically determines optimal number of clusters
- Works well with right-skewed distributions

3.5 Enhanced Dataset Features

Added new columns based on analysis:

1. **activity_balance_score: Objective function 1 results**
2. **intensity_score: Objective function 2 results**
3. **is_golden_cluster: Binary classification target**
4. **fitness_score: Composite performance metric**

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

This cluster represents users who maintain a balanced and sustainable activity pattern, characterized by moderate but consistent activity levels. The relatively small cluster size suggests these are exemplar users whose patterns could serve as behavioral targets for others.

4.2 Quality Metrics

Silhouette Score: 0.391

- Indicates good cluster cohesion and separation
- Above average compared to other clusters

SSE: 75.039

- Low within-cluster variance
- Suggests consistent behavior patterns

Activity Level Score: 0.391

- Balanced distribution across activity intensities
- Meets WHO recommended activity guidelines

Activity Intensity Score: 0.218

- Shows effective energy expenditure
- Indicates sustainable activity patterns

5. Conclusions and Insights

5.1 Activity Patterns

- Identified distinct user segments based on activity intensity
    - Clear separation between sedentary and active users
    - Golden cluster represents balanced activity pattern
- Strong correlation between activity type and calorie burn
    - Higher intensity activities show better calorie efficiency
    - Consistent patterns lead to better outcomes

5.2 Optimization Opportunities

- Target interventions for sedentary users
    - Gradually increase activity levels
    - Focus on achievable milestones
- Personalize recommendations based on cluster characteristics
    - Adapt goals to user patterns
    - Provide intensity-appropriate suggestions
- Focus on consistency and gradual progression
    - Emphasize sustainable behavior changes
    - Monitor progress with multiple metrics