

# Project Proposal: Building a Rich Fitness Dataset and Recommendation System Using Machine Learning

By Team TrulyFit, AI Saturdays Lagos Cohort 9 - Flipped

---

## Problem Statement

Fitness datasets today are overly simplistic. A typical entry might say:

*“If you’re male, overweight, and want to burn fat, do cardio and eat chicken with rice.”*

This approach is limited because:

- It doesn’t specify exercise duration or intensity.
- It ignores available equipment.
- It lacks nutrition details like calories, protein, or allergens.
- It doesn’t account for personal preferences (vegetarian, halal, or local foods).
- It is too shallow for training advanced machine learning models.

In short, existing datasets cannot provide practical fitness advice nor serve as a solid foundation for serious ML research.

---

## Existing Solutions

Current platforms and datasets [1,2] provide general fitness recommendations, but most lack depth, personalization, and ML-readiness. They rely on rigid templates and do not integrate nutritional precision or exercise personalization. Our project differentiates itself by curating an enriched dataset, combining structured exercise and meal data, and applying ML methods to generate holistic fitness recommendations.

---

## Objectives

- Enrich the basic dataset with detailed fields: duration, intensity, equipment, calories, allergens, and cultural food swaps.
- Build a rule-based system for reliable baseline recommendations.
- Develop a nearest-neighbor recommender to suggest plans based on similar user profiles.
- Train a multi-output ML model to predict both exercise and meal templates from just three inputs: **Gender, Goal, and BMI**.
- Deploy results in a simple, interactive application.

---

## Proposed Dataset

We will start with an existing fitness dataset [URL] and augment it with new, richer features:

Datasets include :

1. Work Preference and Fitness goal dataset  
<https://www.kaggle.com/datasets/sumedh1507/fitness-and-workout-dataset>
2. Meal Plan and Exercise Schedules dataset  
<https://www.kaggle.com/datasets/kavindavimukthi/meal-plan-and-exercise-schedule-generator-goal-bmi>

We will also focus on

- **Exercise details:** type, duration, intensity, and required equipment.
- **Nutrition details:** calories, macronutrients, allergens, and substitution options.
- **Personalization factors:** diet preferences (vegetarian, halal, local foods).

Why This Dataset: By extending an existing dataset, we combine familiarity with innovation. The enriched dataset allows for practical use by individuals while also enabling meaningful ML experimentation.

---

# Proposed Methodology

## Data Sourcing & Cleaning

- Collect and structure exercise and meal plan data from sources like kaggle.
- Standardize units (minutes, calories, grams, etc.).
- Handle missing values and normalize formats.

## Exploratory Data Analysis (EDA)

- Visualize trends in BMI, gender, and goals.
- Explore exercise-nutrition correlations.

## Feature Engineering

- Encode categorical preferences (e.g., vegetarian).
- Generate derived features (calorie-to-protein ratio, intensity index).

## Modeling & Evaluation

- **Rule-based baseline:** Hard-coded mappings from Gender + Goal + BMI.
- **Nearest-neighbor recommender:** Suggests plans from similar users.
- **Multi-output ML model:** Predicts both exercise and meal templates simultaneously.
- Evaluate performance with accuracy, precision, and user-feedback metrics.

## Deployment

- Build an interactive app (API + Streamlit) where users input Gender, Goal, and BMI.
- Provide safe, transparent, and personalized recommendations instantly.

---

## Modeling Plan

We plan to test three main approaches:

1. **Rule-Based System:** Ensures reliable and interpretable baseline outputs.
2. **Nearest-Neighbor Recommender:** Captures similarity across users.
3. **Multi-Output ML Model:** Predicts both exercise and meal templates in a single pipeline.

Comparison will be based on predictive accuracy, personalization quality, and usability.

---

## Deployment Plan

The final solution will be deployed as a web application (Streamlit), backed by an API. Users can enter three simple inputs (gender, goal, and BMI) and instantly receive a combined exercise and nutrition plan, complete with practical details. The app may later be integrated with Hugging Face Spaces for public accessibility.

---

## Expected Outcomes

- A richer, ML-ready fitness dataset.
  - A recommendation system providing **personalized exercise and meal advice**.
  - An interactive, user-friendly web tool for individuals and researchers.
  - Insights from EDA on key factors affecting exercise and nutrition planning.
- 

## Community Impact

This project will empower individuals to access more **practical, safe, and personalized fitness guidance**. For the ML community, it creates a benchmark dataset suitable for recommender

systems and multi-output prediction research. The result is a dual benefit: better health outcomes and a stronger foundation for fitness AI innovation.

---

## Team Members

Lawani Abiola  
Manasi Thapa

---

## Acknowledgement

Kaggle for datasets

---

## References

- [1] Kaggle Fitness & Nutrition datasets:  
[<https://www.kaggle.com/datasets/sumedh1507/fitness-and-workout-dataset>]  
[2] Kaggle Meal Plan and exercise Schedule:  
[<https://www.kaggle.com/datasets/kavindavimukthi/meal-plan-and-exercise-schedule-gender-goal-bmi>]