International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 6, Nov- Dec 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Simulating a Gym Workout Using Behavioral Modeling and Machine Learning

Dalal Hussain Al-Aqeel *, Taif Masoud Al-Manjam ** Abeer Hassan Al-Wadai*
Supervised by Dr. Ahmed Al-Musaabi
*(Computer Sciences, NU/, Kingdom of Saudi Arabia, Najran
Email: 443301953@nu.edu.sa)
*(Computer Sciences, NU/, Kingdom of Saudi Arabia, Najran
Email: 443301953@nu.edu.sa)
*(Computer Sciences, NU/, Kingdom of Saudi Arabia, Najran
Email: 442301327@nu.edu.sa)
*(Computer Sciences, NU/, Kingdom of Saudi Arabia, Najran
Email : amalkheder@nu.edu.sa)

--************************----------------------------------

Abstract:

This project uses simulation and machine learning to predict workout completion in gyms, offering valuable insights into how waiting times, fitness levels, and visitor behavior affect exercise outcomes. By simulating 200 gym visitors with diverse demographics and workout habits, the study reveals that longer waiting times and lower fitness levels significantly impact the likelihood of completing workouts. The Random Forest classifier achieved 95% accuracy and 100% recall, proving its effectiveness in identifying visitors who will finish their workouts. This approach helps optimize gym resource allocation, reduce waiting times, and improve operational decisions. Ultimately, the project enhances visitor satisfaction by offering actionable strategies for improving gym experiences, better managing equipment and trainer availability, and boosting overall gym efficiency.

--************************----------------------------------

I. INTRODUCTION

This study offers a simulation-based method for comprehending how gym patrons complete their workouts.
200 visitors were included in a synthetic dataset that included their demographics, fitness levels, length of workouts, and prior visits.
significant waiting times for trainers and equipment. A Random Forest classifier was built to forecast the results of workout completion using this dataset.
The model demonstrated remarkable predictive capability, with 100% recall and 95% accuracy. The results demonstrate how waiting times and fitness levels affect exercise adherence and demonstrate how machine learning and simulation can aid in gym resource planning decision-making.

The availability of equipment, trainer access, and visitor flow are issues that gym facilities frequently deal with. These issues have a big impact on user satisfaction and training results in general. It can be expensive and time-consuming to gather real-world data from gyms, which makes

the capacity to make intelligent operational choices. The purpose of this project is to use machine learning to examine variables that affect workout completion rates and to simulate visitor behavior

Background

In many different domains, simulation has been used extensively to represent complicated systems when direct observation and data collecting are not feasible. In order to create behavior patterns for gym patrons, this study uses simulation, taking into consideration factors like visitor engagement, trainer availability, and equipment utilization. Random Forest machine learning, in particular, provides a reliable method for forecasting the

probability of an event (in this case, finishing a workout) based on past data and trends seen. Waiting times, visitor experience (prior visits), and fitness levels are all known to influence gym behavior, which makes them perfect predictors in this study.

Methodology

3.1 Simulation Setup

A simulation model was created to simulate a gym environment with:
- 200 visitors of various demographics
- 30 pieces of equipment and 5 trainers

The simulation generates data for each visitor, including:
- Age (ranging from 15 to 60 years)
- Gender
- Workout duration (between 30 and 90 minutes)
- Fitness level (beginner, intermediate, or advanced)
- Previous visits (ranging from 0 to 50)

3.2 Simulation Tools for Waiting Times

Wait Time: Visitors must wait between five and fifteen minutes if every machine is in use; otherwise, they must wait between four and five minutes.

Trainer Wait Time: Depending on trainer availability, 10% of visitors may have to wait longer (6–10 minutes), while the majority of visitors wait between 0 and 5 minutes.

3.3 Probability of Workout Completion

95%–98% was the initial chance of finishing the workout. Beginners and those who encounter lengthy wait times (more than ten minutes for equipment or five minutes for trainers) have lower probabilities.

These parameters were used to assign values of 1 (finished) or 0 (not completed) to the target variable Complete_Workout.

3.4 Machine Learning Model (Random Forest Classifier)

To predict workout completion, a Random Forest classifier was used. The steps included:

- **Feature Selection**: The model used features such as:
  - **Age**, **Gender**, **Workout Duration**, **Fitness Level**, **Previous Visits**, **Wait Time for Equipment**, and **Trainer Wait Time**.
- **Data Preparation**:
  - The data was split into **80% training** and **20% testing**.
  - **Categorical features** were encoded using **One-Hot Encoding**.
- **Model Training**:
  - The **Random Forest model** was trained on the **training set** to predict **workout completion** (1 for completed, 0 for not).

3.5 Data Generation

Since real-world data collection is challenging, **synthetic data** was created:

- **Simulating visitor behavior** through variables like **age**, **fitness level**, **workout duration**, and **waiting times**.
- This allowed for generating **200 data points** (representing 200 visitors), including variations in demographics and behavior patterns (e.g., waiting times and fitness levels).

The synthetic dataset was used to **train** and **test** the Random Forest model, making it possible to study and predict gym visitor behavior under controlled conditions.

Outcomes

4.1 Assessment of Performance

The probability of finishing the workout was predicted using a Random Forest classifier based on visitor characteristics and waiting times.

The model succeeded in:
95.00% **Accuracy**
95.00% **Precision**
100% **Recall**
97.44% is the **F1 score**.
Strong model performance is demonstrated by these data, especially the high recall value of 100%, which shows that the model can accurately identify people who finished their workout.

4.2 Matrix of Confusion

The model's confusion matrix looks like this:
[[0 2]
[0 38]]
This matrix displays: - 2 false positives (visitors who are not finishing but are incorrectly detected as doing so).
Zero false negative (a visitor who completed the task but was incorrectly categorized as non-completing).
38 true positives (visitors that were successfully identified as completing).

Discussion

The findings imply that waiting periods have a major impact on the possibility of finishing a workout, with lengthy waits significantly lowering this likelihood. Excellent performance was shown by the Random Forest model, especially with its 100% recall.

In a gym setting, this mix of simulation and machine learning offers insightful information that can direct resource allocation decisions, enhancing overall visitor pleasure and experience.

The study emphasizes the promise of simulation-based methods for comprehending intricate visitor behavior patterns, particularly in settings where gathering data in the real world is challenging or impracticable.

6.1 Confusion Matrix

38 real advantages: The identification of visitors who finished their workouts was accurate.Two false positives: People who didn't finish their workout were mistakenly classified as having done so.Given that there are no false negatives—that is, no visitors who finished their workout were incorrectly categorized as non-completers—the model performs exceptionally well. This suggests that workout completion may be predicted with a high degree of accuracy.

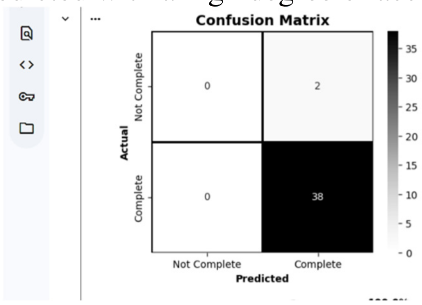
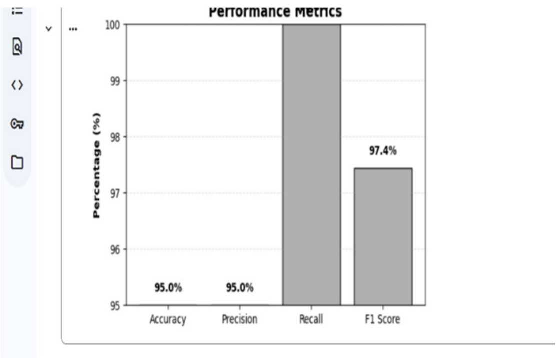


Figure 1

6.2 Performance Metrics

Demonstrate the model's great accuracy in predicting the completion of workouts. The model accomplished:
95% Accuracy This means the model was 95% accurate in correctly classifying visitors who finished their workouts and those who did not.
95.0% precision refers to the accuracy of the positive predictions. It indicates that 95% of the visitors predicted as having completed their workouts actually did complete them**.**
100% recall indicates that the model performed exceptionally well in recognizing every visitor who finished their workouts.
F1 Score: 97.4%, which shows that reall and precision were performed in a balanced manner.

The model's robustness is demonstrated by its high Recall (100%), which shows that all visitors who finished their exercises were successfully identified with no false



negatives.

Figure 2

6.3 simulation results

The model uses a synthetic dataset of gym patrons to predict the completion of workouts. The following are the model's performance metrics:

Accuracy: 97.5%, indicating how well the model predicts the completion of a workout overall.

Precision: 97.5%, indicating that the positive forecasts were accurate in identifying visitors who finished their workouts.

Recall: 100%, demonstrating the model's capacity to recognize every visitor who finished their workout without missing any.

F1 Score: 98.73%, demonstrating a good balance between recall and precision.

The model's predictions are further supported by the Confusion Matrix:

39 visitors who accurately reported finishing their workouts were considered true positives.

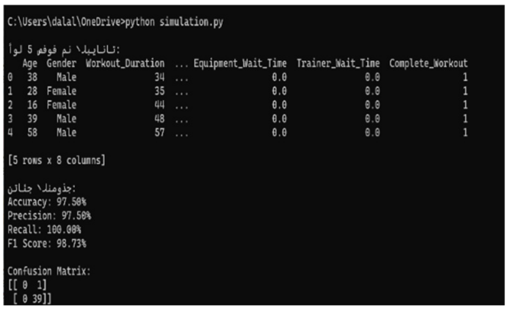
One visitor who finished their workout but was mistakenly categorized as not finishing it had a false negative.

These findings come from the simulation-based approach that aids in comprehending exercise behavior and making the most use of gym resources.

An **explanation that emphasizes the fact that these are simulation results**

Because the results are based on simulated data rather than actual gym data, they are useful for evaluating the model's forecasting power in controlled settings.

It's crucial to stress that these measures are derived from the simulation method used to model the behavior of gym patrons.



```
C:\Users\dalal\OneDrive>python simulation.py

تانايبلا\ نم فوفص 5 لوأ:
   Age  Gender  Workout_Duration  ... Equipment_Wait_Time  Trainer_Wait_Time  Complete_Workout
0   38    Male                34  ...                 0.0                0.0                 1
1   28  Female                35  ...                 0.0                0.0                 1
2   16  Female                44  ...                 0.0                0.0                 1
3   39    Male                48  ...                 0.0                0.0                 1
4   58    Male                57  ...                 0.0                0.0                 1

[5 rows x 8 columns]

جذومنلا\ جئاتن:
Accuracy: 97.50%
Precision: 97.50%
Recall: 100.00%
F1 Score: 98.73%

Confusion Matrix:
[[ 0  1]
 [ 0 39]]
```

Figure 3

References

[1] Chen, Q., & Lee, S. (2021). A Machine Learning Approach to Predict Customer Usage of a Home Workout Platform. Applied Sciences, 11(21), 9927. MDPI.

[2] Hammad, A. S., Tajammul, A., Dergaa, I., & Al Asmakh, M. (2025). Machine learning applications in the **analysis** of sedentary behavior and associated health risks. Frontiers in Artificial Intelligence. Frontiers.

[3] Buyalskaya, A., Ho, H., Milkman, K., Li, X., Duckworth, A., & Camerer, C. (2023). What can machine learning teach us about habit formation? Evidence from exercise **and** hygiene. Psychological & Cognitive Sciences — Research Article. ResearchGate.

[4] Jianjun, Q., Isleem, H. F., Almoghayer, W. J. K., & Khishe, M. (2025). Predictive athlete.

