ASSESSMENT 1: Rule-based AI scientific Research Paper

*ARTIFICIAL INTELLIGENCE [CPU5006-20,SEP-BU,SEM1,2024-2025]  
Marker: Andy Gray*

Contents

[Introduction 4](#_Toc182832157)

[Literature Review 5](#_Toc182832158)

[Rule-Based Expert Systems 5](#_Toc182832159)

[Decision Tree Classifiers 5](#_Toc182832160)

[Methodology 6](#_Toc182832161)

[Algorithms chosen 6](#_Toc182832162)

[Dataset Selection and Preprocessing 6](#_Toc182832163)

[Features: 6](#_Toc182832164)

[Results 8](#_Toc182832165)

[Conclusion 9](#_Toc182832166)

[References 10](#_Toc182832167)

**Abstract**

This study investigates the effectiveness of two rule-based AI algorithms in classifying gym members by their engagement level, specifically distinguishing between casual users, regular attendees, and high-intensity trainers. With a growing interest in personalized fitness and member retention strategies, categorizing gym members based on attributes such as age, body composition, workout preferences, and session intensity can provide gyms with critical insights for targeted engagement. This research compares a decision tree classifier and a rule-based expert system, evaluating their performance in classifying gym members based on attributes from the Gym Members Exercise Dataset.

The dataset includes detailed information about each member, such as age, gender, BMI (calculated from height and weight), maximum, average, and resting heart rates, and key metrics like workout type, duration, and calories burned per session. Additional attributes such as workout frequency, daily water intake, body fat percentage, and experience level provide further context about each member’s engagement and fitness level. Using these variables, the study assesses the accuracy, interpretability, and computational efficiency of each algorithm in identifying engagement categories.

Initial results indicate that each algorithm shows distinct strengths: the decision tree classifier identifies nuanced, data-driven patterns across multiple attributes, while the expert system offers greater interpretability and transparency through predefined rules. These findings contribute to understanding the practical applications of rule-based AI in fitness management, supporting gyms in optimizing member engagement strategies and improving personalization.

# Introduction

The fitness industry has seen a significant rise in participation in recent years, which has led businesses to use data to improve member engagement, provide individualised services, and increase retention rates. For fitness centres and gyms, gaining insights into members' engagement patterns—such as workout frequency, session intensity, and exercise preferences—is crucial for creating targeted strategies that address individual needs. By classifying members based on their behaviour and fitness levels, gyms can offer more tailored experiences while making data-driven decisions that positively impact each client’s journey.  
  
  
Using rule-based AI technologies, such as a decision tree classifier and a rule-based expert system, this study aims to boost gym member engagement. Using these algorithms, members will be paired with the best gym equipment according to their demographics, physical characteristics, and workout preferences. By using information from the Gym Members Exercise Dataset, which include factors like age, fitness levels, health indicators, and workout preferences, the study seeks to tailor equipment suggestions to meet the needs and goals of everyone.

In addition to optimising gym machine utilisation through personalisation, the study aims to spot underuse trends and make well-informed judgements on resource allocation, equipment placement, and upgrades. By use of customised suggestions and specific exercises, the study aims to provide a more interesting and rewarding experience for participants, ultimately increasing their engagement and retention.

The research question addressed in this paper is   
*How can rule-based AI algorithms match gym members with the most suitable equipment based on demographic, physical, and workout-related attributes, and how can these insights be used to optimise gym machine utilisation while increasing overall member engagement?*

The study’s objectives are to evaluate each algorithm’s performance in terms of accuracy, interpretability, and efficiency, and to determine which approach is more practical for real-world applications in gym management. By identifying the optimal algorithm for this classification task, the study aims to contribute to a better understanding of how AI can be effectively utilized in the fitness industry to enhance personalization and support member engagement.

# Literature Review

Rule-based systems are a type of AI that uses a series of predefined to make decisions and solve problems. These systems use an “if-then” structure, which mimics human decision making it a reliable and predictable system for various applications. This system stands out because it focuses and relies on the rules created by humans. Because of this it means every operation and decision it makes can is specify for the specific set of guidelines. Because of this it is commonly used in healthcare – for heart disease diagnosis, customer support – chatbots/smart assistants, and banking – fraud detection and risk management. These systems are efficient for decision-making, especially in well-structured domains where rules can be explicitly defined.

## Rule-Based Expert Systems

Rule-based systems are one of the earliest and most well-established types of AI. This system relies on a collection of predetermined rules to decide what to do next, for example if X then. This system is seen as transparent because the standards are clear and open to human inspection, which makes it simpler to comprehend how the system operates.

These systems are crucial in domains like **healthcare**, **engineering**, and **agriculture**, where human expertise is difficult to automate through direct calculations. However, rule-based expert systems are heavily dependent on the quality and completeness of the rules and often require ongoing maintenance by domain experts.

## Decision Tree Classifiers

Decision tree classifiers are a prominent example of rule-based algorithms. They function by recursively splitting data based on feature values, forming a tree-like structure. Each node in the tree represents a decision rule that partitions the data, and the leaf nodes represent the classification outcome. Decision trees are inherently rule-based because they break down complex decisions into a series of "if-then" conditions. For example:

Decision trees are widely used in applications such as **loan approvals**, **customer segmentation**, and **predictive maintenance**. However, they can suffer from overfitting, especially when the tree is too deep, which is often mitigated by pruning or using ensemble methods like **Random Forests**.

# Methodology

## Algorithms chosen

I’ve chosen two different methods to solve the question, decision tree classifier and rule-based expert system. These two systems fit into answering the question by using the dataset provided.

First Algorithm is a rule-based expert system. This algorithm works by using specific rules. These systems use if-else statements to test against rules. Examples of rules include Age >60, Gender = male. Using this system, I was able to find how many people workout 4 times a week and the most used type of workout. I also found the least used workout, from each gender.

The second algorithm I chose is a decision tree. I choose the method due to its ability to handle both categorical and numerical data that are present in the database. With the help of this technique, a tree-like structure is produced, with each node standing for a rule, such as Age < 30. The final classification, such as Workout Type = HIIT, is represented by the tree's leaves. This enables the algorithm to see trends and produce understandable classification rules, such suggesting exercises based on demographic variables, or seeing different age groups favourite workout type.

## Dataset Selection

The dataset that I have used to answer this question is **Gym Members Exercise Dataset** from Kaggle. Some uses for this dataset are for analysing fitness patterns and performance across diverse gym experience levels. This dataset includes a detailed overview of gym members routines, fitness metrics, and physical attributes. Contains 973 samples of age, gender, weight, height, max BPM, average BPM, resting BPM, session duration, calories burned, and workout type.  
I chose this dataset due to its quality data, and this dataset included all the basic information for this question. Not only this, but the number of samples will make the answer more accurate. All the features allow me to answer multiple different questions, such as what age group visits the gym the most, or how much time experienced people spend in the gym.

### Features:

* Age: Age of the gym member. (Years)
* Gender: Gender of the gym member (Male or Female).
* Weight (kg): Member’s weight in kilograms.
* Height (m): Member’s height in meters.
* Max\_BPM: Maximum heart rate (beats per minute) during workout sessions.
* Avg\_BPM: Average heart rate during workout sessions.
* Resting\_BPM: Heart rate at rest before workout.
* Session\_Duration (hours): Duration of each workout session in hours.
* Calories\_Burned: Total calories burned during each session.
* Workout\_Type: Type of workout performed (e.g., Cardio, Strength, Yoga, HIIT).
* Fat\_Percentage: Body fat percentage of the member.
* Water\_Intake (liters): Daily water intake during workouts.
* Workout\_Frequency (days/week): Number of workout sessions per week.
* Experience\_Level: Level of experience, from beginner (1) to expert (3).
* BMI: Body Mass Index, calculated from height and weight.

Gender (percentage of men vs. women): This could help differentiate equipment preferences or design tailored fitness plans (e.g., men vs. women might prefer different types of equipment or workout styles). Also, the gym could advertise for women to become a more inclusive gym.

Age Range: This will allow you to create rules for different age categories, like younger users might be suited to high-intensity workouts while older members may prefer low-impact exercises.

Workout Types Percentage: Knowing which types of workouts members prefer (e.g., cardio, strength training, HIIT) allows the system to recommend relevant equipment.

Workout Frequency: Members who work out more frequently may be suited to more intense equipment or a higher rotation of machines.

Unhealthy BMI: This can help classify members into beginner or advanced groups for appropriate equipment, e.g., for weight loss or strength training.

# Results

## Performance of Algorithm 1:

Information gathered:

* Gender distribution – This algorithm calculated the percentage of male and female members
* Workout preferences – Finds most common workout, for each gender, BMI category and age group
* Age group preferences – puts members into age-ranges and calculates their most common workout type
* BMI classification - categorizes members as underweight, obese, or healthy based on thresholds from NHS standards. Then shows what their most common workout type is.

Computational information:

Time Complexity – this algorithm uses pandas. So, the time complexity depends on the number of rows in the dataset, and how many values in the columns. For example, when the algorithm was calculating the percentage, it was going through every row in the dataset. The larger the database the more time it will take. Another example, is group by, using this means the algorithm will look at combinations of two or more columns, which make sit more complex than just looking at each row once.

Space complexity – This means how much memory the algorithm uses while it runs. For example, it needs to store the dataset in the memory, so the algorithm can process it quickly.

## Performance of Algorithm 2:

1. **Performance of Algorithm 1 (e.g., Decision Tree)**:
   * Present accuracy, precision, recall, F1-score, and any other chosen metric.
   * Provide insights into the computational efficiency and any unique strengths or weaknesses observed.
2. **Comparison and Analysis**:
   * Present a table or graph that compares both algorithms across all metrics.
   * Interpret differences in performance and explain why certain features may influence the outcomes in each model.

# Conclusion

# References

[*https://deepgram.com/ai-glossary/rule-based-ai*](https://deepgram.com/ai-glossary/rule-based-ai)

[*https://www.sciencedirect.com/science/article/pii/S2468227621003203*](https://www.sciencedirect.com/science/article/pii/S2468227621003203)

[*https://www.nhs.uk/conditions/obesity/#:~:text=For%20most%20adults%2C%20if%20your,re%20in%20the%20overweight%20range*](https://www.nhs.uk/conditions/obesity/#:~:text=For%20most%20adults%2C%20if%20your,re%20in%20the%20overweight%20range)

[*https://www.kaggle.com/datasets/valakhorasani/gym-members-exercise-dataset?resource=download*](https://www.kaggle.com/datasets/valakhorasani/gym-members-exercise-dataset?resource=download)

[*https://www.w3schools.com/python/python\_ml\_decision\_tree.asp*](https://www.w3schools.com/python/python_ml_decision_tree.asp)

[*https://www.geeksforgeeks.org/get-first-n-records-of-a-pandas-dataframe/*](https://www.geeksforgeeks.org/get-first-n-records-of-a-pandas-dataframe/)

[*https://www.datacamp.com/tutorial/decision-tree-classification-python*](https://www.datacamp.com/tutorial/decision-tree-classification-python)

[*https://www.datacamp.com/tutorial/decision-tree-classification-python?dc\_referrer=https%3A%2F%2Fwww.google.com%2F*](https://www.datacamp.com/tutorial/decision-tree-classification-python?dc_referrer=https%3A%2F%2Fwww.google.com%2F)

*ChatGPT*

*Question: how would i find the age range of a dataset*

*A screenshot of a computer

Description automatically generated*

*A screenshot of a computer program

Description automatically generated*

*A screenshot of a black screen

Description automatically generated*

*"How accurately can two different rule-based AI algorithms classify gym members by their engagement level, such as casual users, regular attendees, and high-intensity trainers, based on their exercise and demographic data?"*

**Results**

1. **Performance of Algorithm 1 (e.g., Decision Tree)**:
   * Present accuracy, precision, recall, F1-score, and any other chosen metric.
   * Provide insights into the computational efficiency and any unique strengths or weaknesses observed.
2. **Performance of Algorithm 2 (e.g., Expert System)**:
   * Present the same performance metrics.
   * Compare computational costs and describe scenarios where this algorithm performs best.
3. **Comparison and Analysis**:
   * Present a table or graph that compares both algorithms across all metrics.
   * Interpret differences in performance and explain why certain features may influence the outcomes in each model.

**Discussion**

1. **Evaluation of Results**: Discuss why one algorithm may have performed better or worse in specific areas, using evidence from the results.
2. **Strengths and Weaknesses**: Address the strengths and limitations of each algorithm based on the experiment.
3. **Application Potential**: Discuss the potential applications for each rule-based AI system, considering the dataset characteristics and algorithmic trade-offs.

**Conclusion**

1. **Summary of Findings**: Recap the key results and insights from the research question.
2. **Implications**: Address the broader implications, such as when each rule-based system might be preferable based on the characteristics of a given problem.
3. **Future Work**: Suggest further research directions, such as testing additional rule-based methods or hybrid models to improve performance.

**References**