Review of Computer Engineering Research
2025 Vol. 12, No. 3, pp. 139-154
ISSN(e): 2410-9142
ISSN(p): 2412-4281
DOI: 10.18488/76.v12i3.4338

A data mining-based model for predicting the repurchase behavior of fitness service users

Phatarapon Vorapracha

Division of Information and Digital Technology Management, Faculty of Industrial Technology, Phranakhon Rajabhat University, Bangkok, 10220, Thailand.
Email: phatarapon.v@pnru.ac.th

Article History
Received: 14 April 2025
Revised: 26 June 2025
Accepted: 21 July 2025
Published: 1 August 2025

Keywords
Association rules
Customer retention
Data mining
Decision tree
Fitness analytics
Neural networks
Predictive modeling.

ABSTRACT

This study aimed to develop a predictive model for fitness service reuse using data mining techniques. Four modeling approaches were evaluated: neural networks, decision trees, association rules, and logistic regression. Among them, the neural network model demonstrated the highest predictive performance (Accuracy = 89.7%, Precision = 86.3%, Recall = 88.0%), making it suitable for integration with automated CRM systems and personalized marketing efforts. The decision tree model, while slightly less accurate (Accuracy = 81.2%), offered high interpretability, facilitating strategic decision-making and effective communication with management stakeholders. The association rule model, employing the Apriori algorithm, successfully uncovered behavioral patterns related to weight training and class participation, which can inform the development of targeted promotions and customized service offerings. Logistic regression, known for its simplicity and transparency, was found to be less capable of modeling complex, non-linear relationships within the dataset. The study concludes that a hybrid approach combining neural networks and association rule mining offers complementary advantages high predictive accuracy and actionable insights into user behavior. This integrated method supports the delivery of personalized services, reduces member attrition, and enhances customer retention. Overall, the findings offer practical value for strategic planning in the fitness industry by enabling data-driven decisions to optimize customer engagement and long-term loyalty.

Contribution/Originality: This study uniquely integrates four data mining techniques to predict customer repurchase behavior in fitness services. Unlike previous studies that used single models, it combines neural networks with association rules to enhance accuracy and interpretability, providing a practical and adaptable framework for customer retention across service-based industries.

1. INTRODUCTION

Creating value through the provision of high-quality products and services is of paramount importance in today's competitive business landscape. The foundation of delivering quality service and achieving customer satisfaction lies in cultivating a service mind Murnpho [1], which is considered the core of effective service delivery that all businesses and service personnel should possess. Although the term is often associated with front-line employees in retail stores and hospitality sectors, the concept of a service mind is equally applicable within internal organizational functions, including accounting, logistics, marketing, and production. All departments must consider service orientation as an essential element of their operational approach. In an era characterized by intense business competition across all sectors, the adoption of a service mind contributes significantly to customer retention and loyalty [2]. A service-oriented approach not only enhances the overall customer experience but also creates perceived value for the business. High-quality service fosters greater customer satisfaction, thereby increasing their willingness to pay for services that meet or exceed their expectations. This phenomenon is particularly evident in service-driven industries, such as beauty clinics, where delivering exceptional service can differentiate the business, encourage repeat patronage, and justify premium pricing. Even when the costs of operations, resources, raw materials, and human capital remain constant, the presence of a service mindset becomes a strategic asset that generates added value—value that customers are willing to pay for. The COVID-19 pandemic has underscored the importance of health and wellness, prompting individuals across all age groups to adopt healthier lifestyles. Among the most accessible and effective means of maintaining good health is regular physical exercise [3]. The general benefits of exercise are multifaceted, encompassing both physical and psychological dimensions. Regular physical activity contributes to improved mood, enhanced physical strength, and increased endurance. It also plays a vital role in the prevention of chronic diseases, including cardiovascular disease, which ranks as the second leading cause of death in Thailand, osteoporosis, and associated complications such as spinal and hip fractures. Furthermore, exercise has been shown to lower blood lipid levels, including cholesterol and triglycerides, thereby reducing risk factors associated with coronary artery disease, cerebrovascular incidents, paralysis, and hemiplegia. It enhances physical agility, balance, and coordination, thereby minimizing the risk of falls. Exercise also aids in the prevention and management of conditions such as diabetes, hypertension, allergic disorders, and digestive irregularities. Additional benefits include improved gastrointestinal function, reduced obesity, strengthened immune response, heightened concentration, stress reduction, improved sleep quality, delayed physiological aging, and overall increased life expectancy. Evidently, the benefits of regular exercise are extensive and well-documented. Given the significance of these health outcomes, the physical environment and nature of exercise activities are critical determinants influencing individuals' choice to engage in physical fitness [4]. Fitness centers, in particular, provide structured and varied exercise options within controlled indoor environments, thereby enabling individuals to select exercise schedules that align with their personal preferences and daily routines. In light of the aforementioned considerations, this study investigates the integration of technological tools to enhance service quality, promote quality of life, and encourage public participation in physical activity. The research also addresses operational challenges faced by fitness centers, proposing solutions for improving service delivery efficiency and user experience. A predictive model was developed to forecast repeated service usage by fitness center members. This model employs data mining techniques to support strategic decision-making and optimize service planning. The study draws on empirical data provided by Rock Gym, a fitness center located within The Wish Samphran 234 Project, Tambon Yai Cha, Amphoe Sam Phran, Nakhon Pathom Province, 73110. The gym operates Monday through Friday from 12:00 p.m. to 10:00 p.m., and Saturday through Sunday from 12:00 p.m. to 8:00 p.m. With over a decade of experience, the facility is operated by three partners and primarily focuses on self-directed fitness activities. Staffing includes three fitness attendants and one housekeeper. The gym provides parking for more than 80 vehicles and serves an average of 60–80 clients per day, comprising both local and international users. Payment for services is accepted via cash and bank transfer. The pricing structure is tiered as follows: 90 baht per single session, 990 baht per month, 3,500 baht for four months, 6,500 baht for eight months, and 8,500 baht annually.

2. LITERATURE REVIEW

The concept of consumer intention behavior reflects the prediction of future purchasing or service usage behavior [5]. It serves as a useful framework for understanding how individuals behave following the evaluation of both internal and external sources of information. Internal sources refer to personal experiences or observations, either directly or through others, often shared through word-of-mouth communication about an organization's services. In contrast, external sources consist of advertising, public relations, and various types of printed media that provide consumers with information and insights. Based on these sources, consumers make decisions regarding which products to purchase or services to use. Positive purchase intentions can be shaped by effective advertising and promotional strategies. Consequently, understanding purchase intention is crucial for marketers, as it serves as a predictor of future consumer behavior and helps estimate market demand [6]. Purchase intention is considered a psychological process that reflects a consumer's readiness or plan to buy a specific brand at a given time. It is typically formed from favorable attitudes toward the brand and trust built through past evaluations or experiences. Similarly, Wuttibramote [7] describes purchase intention as the consumer's effort to acquire a product or service, shaped by various factors such as brand identity, distribution channels, and timing. Ultimately, purchasing behavior is often linked to positive experiences and customer satisfaction.

Nernhad et al. [8], the concept of Service Mind refers to an organization's ability to make customers feel valued and deeply satisfied. Research and industry practice consistently demonstrate a positive relationship between customer satisfaction, increased spending, and expectations for courteous service. Despite its importance, many organizations overlook or fail to fully integrate Service Mind principles into their operations. This paper outlines practical techniques for enhancing Service Mind within an organization. First, responsiveness is critical, as customers place high value on their time. Timely actions such as promptly answering phone calls, replying to messages, and minimizing waiting times can greatly enhance customer satisfaction. Second, small acts of appreciation, such as offering 5–10% discounts, free delivery, or complimentary gifts, help create a lasting positive impression and foster customer loyalty. Empowering frontline employees is equally vital. Rigid instructions without flexibility often suppress authentic service behavior. Allowing staff autonomy and equipping them with effective service techniques encourages a more engaged and customer-centric workforce. Additionally, experiential learning reinforces service principles, enabling employees to internalize and apply them more effectively. While fostering Service Mind may not require significant financial investment, it represents a long-term strategic initiative. The cumulative benefits, especially in customer retention and brand loyalty, are substantial and measurable. Importantly, cultivating a service-oriented culture is not the sole responsibility of sales or customer service departments; rather, it is a collective commitment shared across all organizational units. All employees must be skilled at apologizing, resolving problems, and following up on customer concerns in a professional manner. This integrated approach will lead to improved service quality and reflect the organization's shared mission. Ultimately, a service mind should be ingrained as a core value in the organization's culture. When all members of the organization share a vision that emphasizes service excellence, the result is a sustainable and competitive service model that contributes to the organization's long-term success.

Khotsenar and Prathummanee [9] prediction involves the extraction of knowledge and the development of a relationship function that identifies data types based on various attributes within a database. It is a supervised learning method that can determine the class or category of data to be classified. However, there are key differences between classification and prediction, particularly regarding the nature of the target variable. In classification, the target variable has discrete values, such as a customer's decision to "buy" or "not buy" a computer. In contrast, prediction deals with continuous target values, such as the price of gold, salary, or the water level of the Chao Phraya River. Artificial Neural Networks (ANNs) are widely employed in predictive modeling due to their capability to process complex patterns and deliver high levels of accuracy. Many data mining software platforms embed ANN-based tools, enabling users regardless of deep technical knowledge to leverage their self-learning capabilities. These models can autonomously detect patterns and relationships within datasets, facilitating efficient and accurate outcome prediction. Beyond predictive modeling, ANNs are also extensively utilized in other data mining tasks such as classification, clustering, and time series analysis. Classification, in particular, involves assigning data instances to predefined categories based on a model trained from labeled datasets. The model's performance is typically validated using a separate test set to evaluate its accuracy and generalizability. This technique is especially suited to problems involving categorical or binary decisions, such as "yes/no" or "risk/no risk" classifications. A range of algorithms supports classification tasks, including decision tree methods, Bayesian approaches (e.g., naïve Bayes, Bayesian belief networks), rule-based systems (e.g., IF-THEN rules), support vector machines (SVMs), k-nearest neighbors (k-NN), case-based reasoning, genetic algorithms, and ANNs themselves. In predictive data mining, methods such as regression analysis and chi-square testing are also widely used to explore relationships within data and derive association rules. These rules are typically defined based on two key metrics: support (the frequency of occurrence of a data point) and confidence (the probability of association). Techniques in this domain often involve algorithms that calculate the frequency of items, such as the Apriori algorithm Jiawei et al. [10] or the Frequent Pattern Growth algorithm. Clustering or cluster analysis is another important data mining technique used to group data based on similarity or dissimilarity without predefined categories. The clustering process relies on the characteristics of the data. Similar data points are grouped into the same cluster, while data points with different characteristics are assigned to different clusters. It is important to note that the results of clustering may vary from one operation to another, even when using the same dataset, depending on the algorithm used and the specific clustering conditions. Therefore, the results may differ in structure and interpretation.

Strategic Business Model Design involves the development of a Business Model Canvas derived from data and insights obtained through strategic analysis. This process aims to identify and structure nine key components of a business model: (1) customer segments, (2) value propositions, (3) customer relationships, (4) channels, (5) revenue streams, (6) key partners, (7) key activities, (8) key resources, and (9) cost structure. Once the business model has been formulated, it is presented to entrepreneurs who evaluate and score each model using the Scoring Model method. This evaluation incorporates a balanced business perspective, ensuring that the selected model aligns with the actual feasibility of business development in the context of each entrepreneur's operational capacity and strategic direction.

Phasuk [11] conducted a study to examine the factors influencing experiential marketing that affect repeat tourism in Pattaya City, focusing on demographic characteristics. The sample consisted of 248 Thai tourists who visited Pattaya for man-made tourism experiences, selected through a random sampling technique. A questionnaire served as the primary research instrument. Data were analyzed using both descriptive statistics—including percentages, means, and standard deviations—and inferential statistics, specifically path analysis and structural equation modeling (SEM). The study's results indicate that experiential marketing plays an important role in promoting repeat tourism in Pattaya and aligns closely with empirical data. Confirmatory factor analysis revealed that the experiential marketing model comprises five dimensions: sensory, emotional, cognitive, behavioral (action), and relational (association). The tourist attraction loyalty model includes three components: repeat visits, word-of-mouth, and willingness to pay more. Additionally, the repeat tourism model encompasses intention, effort, and planning. The findings further demonstrate that experiential marketing has both direct and indirect positive effects on tourist attraction loyalty and repeat visit behavior among Thai tourists.

Neelaphatrakul and Biawkamook [12] conducted a study on factors associated with turnover decisions and performance comparison of employee turnover prediction models: a case study of an insurance company. This study investigates the use of data mining techniques to address the problem of a high employee turnover rate and the associated recruitment costs in an insurance company. The research utilizes a dataset of 1,000 records of resigned and active employees collected between 2013 and 2017, with 11 relevant characteristics. Association rule mining techniques are employed to identify patterns and relationships between employee characteristics and turnover decisions. Additionally, five predictive modeling techniques are used to develop employee turnover prediction models: (1) decision trees; (2) support vector machines (SVM); (3) artificial neural networks (ANN); (4) naive Bayes; and (5) k-nearest neighbors (k-NN). The performance of these models is evaluated using 5-fold cross-validation. The association rule analysis revealed significant factors associated with employees' turnover decisions, including salary level, seniority, career advancement opportunities, performance appraisal results, and the quality of relationships with supervisors. Among the tested predictive models, the decision tree model achieved the highest forecasting accuracy of 91.03%, followed by SVM at 90.93%, ANN at 90.75%, naive Bayes at 89.60%, and k-NN at 82.10%. The results of this study provide valuable insights for developing strategic human resource planning, especially in designing appropriate employee retention programs and welfare policies. Additionally, the forecasting model may serve as a foundation for building a decision support system to improve human resource management practices within organizations.

Boonmeekham et al. [13] developed a logistic regression model for lapse analysis of life insurance policies. This study aimed to predict policy lapses using a binary dependent variable indicating whether a policy had lapsed. Initially, 18 independent variables, including both qualitative and quantitative factors, were considered. The dataset was divided into two subsets: a training set of 1,864 policies for model development and a testing set of 466 policies for performance evaluation. The final model identified six statistically significant predictors: policyholder age, a sum insured between 50,001 and 100,000 baht, a premium payment period exceeding three years, the square root of the coverage period, and occupational levels 3 and 4. The model achieved a lapse prediction rate of 31.76% and an overall prediction accuracy of 66.95%. These findings demonstrate the utility of logistic regression in identifying key factors influencing policy lapses. The model offers valuable insights for enhancing risk assessment, improving policyholder retention, and supporting the development of targeted intervention strategies in life insurance portfolio management.

3. MATERIALS AND METHODS

This study aims to explore and analyze data mining systems to select and compare appropriate data mining techniques for developing predictive models for fitness service reuse. The research focuses on evaluating and comparing different methods, including association rule mining, classification and prediction, and logistic regression, to determine the most suitable model for predicting user behavior related to service reuse. By examining the strengths and predictive capabilities of each method, this study aims to identify the model with the highest accuracy and practical applicability in predicting the probability of service reuse among fitness customers.

The service evaluation data consisted of fitness history records, including user names, dates and times of use, membership types (daily, monthly, yearly), and customer statuses. This data was collected and stored retrospectively over a period of 10 years. For the purpose of developing the predictive model, the dataset was divided into two groups: a training dataset used for modeling and a test dataset used to evaluate the model's performance after the design and development phase. This approach ensures the reliability and generalizability of the predictive model in assessing fitness service usage behavior.

Joshi and Patel [14] and Ballard et al. [15] data cleaning involves a series of preprocessing steps aimed at improving the quality and reliability of the dataset. These steps include removing redundant and irrelevant data, correcting structural or formatting errors, and filtering outliers. The process also addresses missing or incomplete data by using appropriate imputation or exclusion methods. Finally, thorough validation is performed to ensure the integrity of the dataset before proceeding with model development.

Gordon and Michael [16] focus on developing a predictive model to forecast fitness service reuse behavior based on a comparative analysis of three data mining techniques.

1. Association Rule.
2. Classification and Artificial Neural Networks.
3. Logistic Regression.

The forecasting models built using each technique were evaluated to identify the model with the highest forecasting accuracy. Model selection was based on performance testing using a comprehensive 10-year fitness usage record dataset. The model with the best forecasting performance was selected as the most appropriate for predicting future fitness service reuse behavior. This approach ensures that the model is practical for supporting data-driven decision-making in the fitness industry.

Konstantinos and Antonios [17] compared forecasting models by calculating the mean absolute percentage error (MAPE) for each technique. This evaluation metric measures the average absolute percentage difference between the predicted and actual values, providing a clear indication of forecast accuracy. The model with the lowest MAPE value was selected as the most appropriate for forecasting fitness service users' reuse behavior in the current year. This approach allows for selecting the model with the smallest forecast error that can be most appropriately applied in real-world decision-making within the fitness service industry.

Figure 1 illustrates the research methodology employed in the development and evaluation of the system.

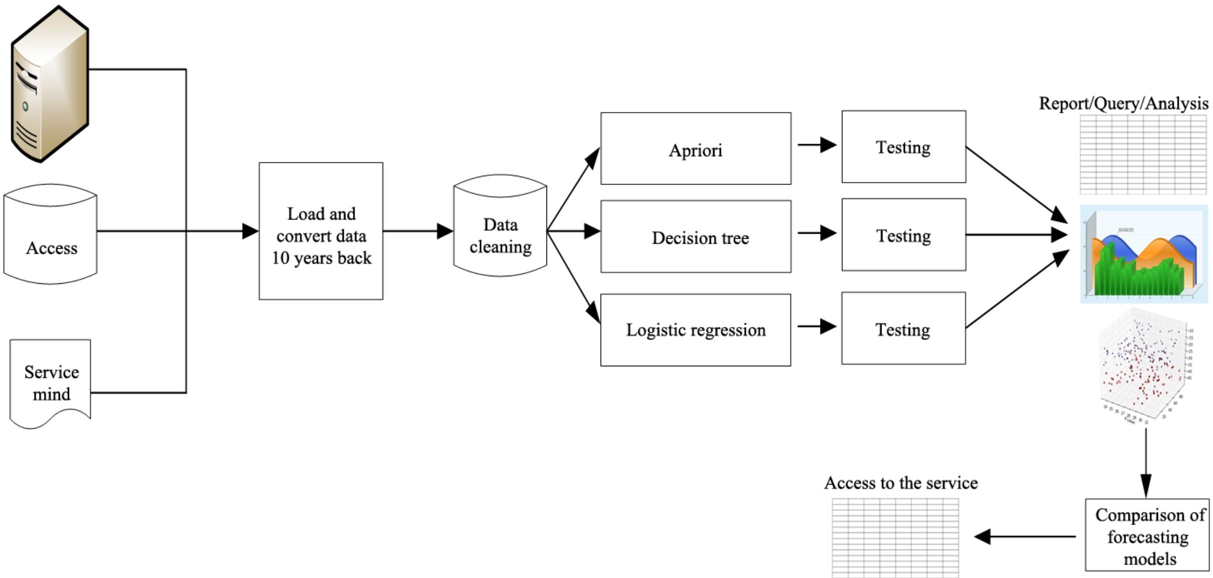

Figure 1. Show the research operation steps of the system.

4. RESULTS AND DISCUSSION

This research focuses on developing a forecasting model for fitness service reuse using data mining techniques. The study involves service usage data collection, data cleaning, model building, model testing, and the selection of the most appropriate forecasting model for predicting fitness service reuse. The researcher followed a systematic research methodology. The research procedure is outlined as follows.

4.1. The Results of the Data Collection on Fitness Service Usage From 2014 to 2023 are as Follows

Data collection was conducted from various sources, including the fitness membership system, which contains members' personal information such as age, gender, and occupation, as well as service usage history (check-in records), check-in and check-out times, types of membership packages, and usage of various facilities. Additionally, data was collected from customer questionnaires and reviews, covering aspects such as satisfaction with the services and suggestions for fitness center improvements. The data for this study was obtained from Rock Gym, a fitness center located in The Wish Samphran 234 Project, Tambon Yai Cha, Amphoe Sam Phran, Nakhon Pathom Province, 73110. Rock Gym operated from 2014 to 2023, during which the aforementioned data was collected, as presented in Table 1.

Table 1. User storage file information distribution list.

List	Number of files
1. Member list and expiration date	12
2. Membership list	120

Storing fitness service user data in Excel format is an efficient and convenient method for data processing. The data can be categorized into two main sections: 1) Member List and Membership Expiration Table. This table is used to store basic member information such as name, member ID, membership start date, and expiration date. It serves to verify service access rights and to monitor membership status, as shown in Table 2.

Table 2. Sample member list and membership expiration dates.

Member ID	Full name	Gender	Age	Membership type	Start date	Expiration date
M001	Somchai Jaidee	Male	32	Monthly	1/1/2024	31/01/2024
M002	Wanna Saithong	Female	28	Annual	15/02/2024	14/02/2025
M003	Kittisak Maneerat	Male	41	4-Month plan	10/03/2024	9/9/2024

2) Check-in Log. This log enables the fitness center to monitor member usage behavior, including the frequency of visits, the specific areas or facilities used, and the approximate duration of each session, as shown in Table 3.

Table 3. Sample check-in log.

Check-in date	Check-in time	Check-out time	Member ID	Full name	Membership type	Activity used
01/04/2024	09:15	10:30	M001	Somchai Jaidee	Monthly	Weight training
01/04/2024	17:00	18:10	M002	Wanna Saithong	Annual	Yoga class
02/04/2024	08:45	10:00	M003	Kittisak Maneerat	4-Month plan	Cardio zone

4.2. Data Cleaning Results

1) Removing duplicates and its impact on fitness repeat visit predictions. Analysis of fitness user behavior data revealed a high number of duplicate check-in entries, including duplicates at similar times or multiple entries on the same day, with no clear indication of whether these entries reflect actual multiple visits or are due to system errors. An examination of 1,794 member data records from 2014 to 2023 found that, on average, 3.8% of all entries were duplicates. Failure to remove these entries could result in a distorted representation of visit frequency, such as some members appearing to visit more frequently than they actually do. This distortion could significantly impact behavioral analysis and customer segmentation, reducing the accuracy and reliability of predictive models.

2) Dealing with Missing Values. Data received from fitness membership systems often contain incomplete entries, especially in personal data fields such as age, occupation, or income type. This incompleteness may be caused by members choosing not to provide certain information or the system not requiring them to enter data for all fields. If missing values are not properly handled during analysis, it can result in inaccurate forecasts and analysis outcomes that deviate from actual user behavior. In this study, the analysis of 1,794 members between 2014 and 2023 revealed that approximately 6.5% of the records lacked age information. To reduce the impact of missing values, appropriate management techniques were employed, including the use of imputation methods based on historical patterns or statistical estimation, and the exclusion of incomplete records where necessary to maintain the reliability of the forecast model.

3) Data transformation and its impact on fitness reuse forecasting. Analyzing fitness user behavior requires accurate and consistent data. However, an examination of member data collected from fitness systems between 2014 and 2023 revealed that several key data items, such as member registration date, gender, and age, were stored in non-standard or inconsistent formats. These inconsistencies cause delays in data processing and increase the risk of errors in analysis or forecasting. Therefore, data transformation is necessary to standardize the format and structure of key variables. This process includes converting dates into a uniform format (e.g., DD/MM/YYYY), encoding gender values into consistent labels (e.g., "male" and "female"), and ensuring that numerical values, such as age, fall within the correct range. Proper data transformation not only facilitates smoother data management but also increases the accuracy and reliability of behavioral forecasting models.

4. Data Encoding and Its Effect on Predicting Repeat Usage of Fitness Members. In the process of analyzing the behavioral data of fitness members, many attributes are found in categorical form, such as membership type (e.g., daily, monthly, yearly), occupation (e.g., student, company employee, self-employed), or satisfaction level (e.g.,

low, medium, high). These categorical variables cannot be directly processed by most statistical models or machine learning algorithms, as such techniques require numerical input for computation. To address this, data encoding techniques were applied to transform categorical data into a numerical format suitable for analysis. Methods such as one-hot encoding or label encoding were utilized depending on the nature of the variable and the modeling technique used. Proper encoding ensures that meaningful distinctions between categories are preserved and that the predictive models are able to interpret the data accurately. Without appropriate encoding, model performance may be compromised due to the inability to process or differentiate between qualitative variables effectively.

4.3. Results of the Development of a Prediction Model for Reuse of Fitness Service Users

1) Exploratory data analysis (EDA) is a crucial step that allows researchers to gain a deeper understanding of fitness service user trends, characteristics, and behaviors before developing a utilization prediction model. This analysis uses data from 1,794 members collected between 2014 and 2023, as shown in Figure 2.

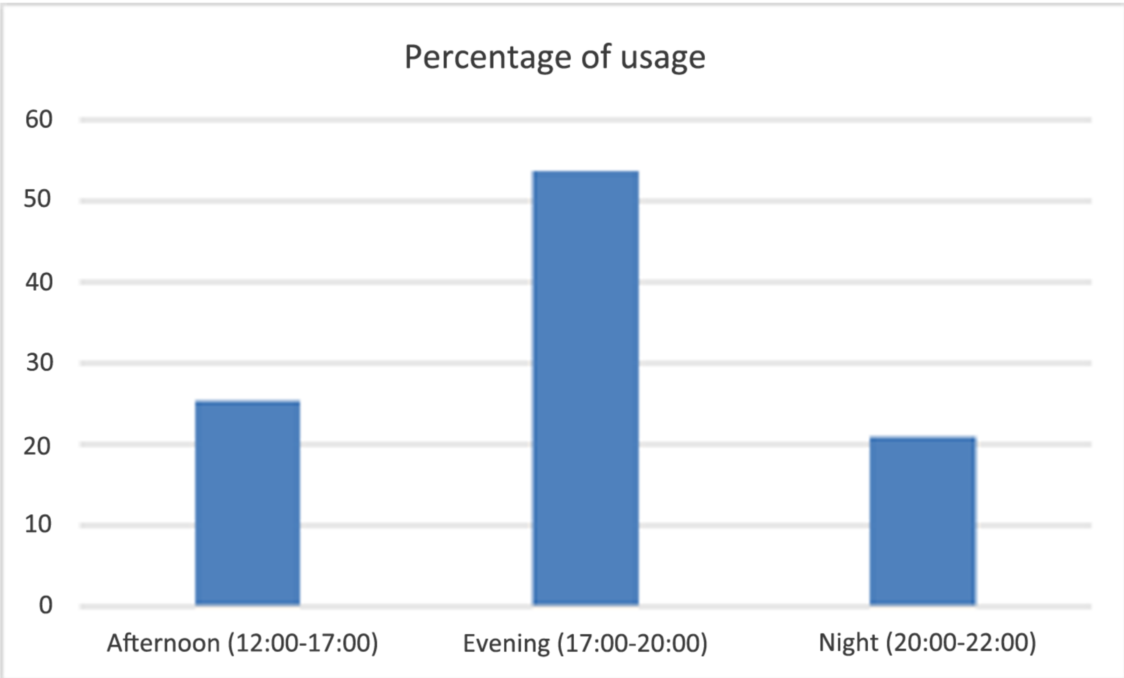

Figure 2. The graph shows the proportion of fitness users at different times of the day.

The analysis in Figure 2 shows that fitness usage is clearly segmented by time of day, with the highest usage in the evening. This insight is useful for planning and managing resources, such as allocating staff, exercise equipment, and scheduling exercise classes to accommodate peak demand. Additionally, the less active times of day, such as the afternoon and evening, present opportunities for targeted marketing strategies, which may include hosting special events or offering promotions to encourage higher usage and distribute the density of users more evenly throughout the day.

2) The decision tree technique provides a clear and interpretable framework for understanding the decision-making process by showing the sequence of rules leading to subscriber classification. The analysis results indicate that the "average number of visits per month" and "time desired to visit" are the most influential factors in predicting long-term subscription trends, as shown in Figure 3.

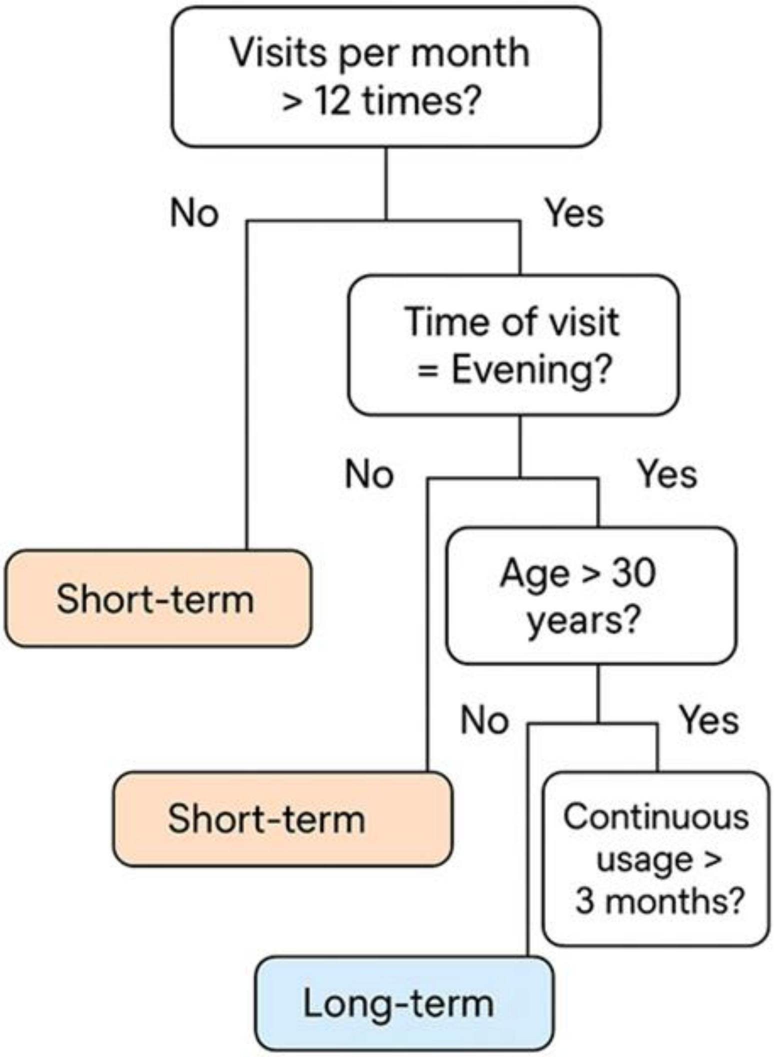

Figure 3. Decision tree model for classifying members as short-term or long-term based on selected features.

The decision tree analysis results revealed that frequent and consistent visits, particularly during the evening hours, are significantly associated with long-term membership retention. This insight provides a valuable basis for designing marketing strategies, such as targeted sales promotions or personalized offers that align with the behavioral patterns of specific customer segments. By focusing on users who demonstrate high engagement levels during peak hours, fitness centers can increase customer retention and encourage long-term membership upgrades.

3) Artificial neural network techniques can learn complex, non-linear patterns in data by using multiple layers to capture intricate relationships between variables, resulting in high classification accuracy, especially for members with less distinct or ambiguous behavioral patterns. Data and Variables Used The sample consisted of 1,794 fitness members, who were initially categorized into four membership types: daily, monthly, quarterly (4 months), and annual. For the purposes of binary classification, these groups were reclassified into two categories: Short-term group: Daily and Monthly members; Long-term group: Quarterly and Annual members. The data used for training the neural network model included the following independent variables: age, gender, average number of visits per month, preferred time slot for visits, and original membership type (if available). Model Development and Training The model employed in this study was a Multi-Layer Perceptron (MLP) with the following architecture and training parameters: Hidden Layers: one hidden layer with 10 nodes, activation function ReLU, optimizer: Adam,

maximum iterations: 1,000 epochs, train-test split: 70% for training and 30% for testing. The structure and learning process of the model are illustrated in Figure 4.

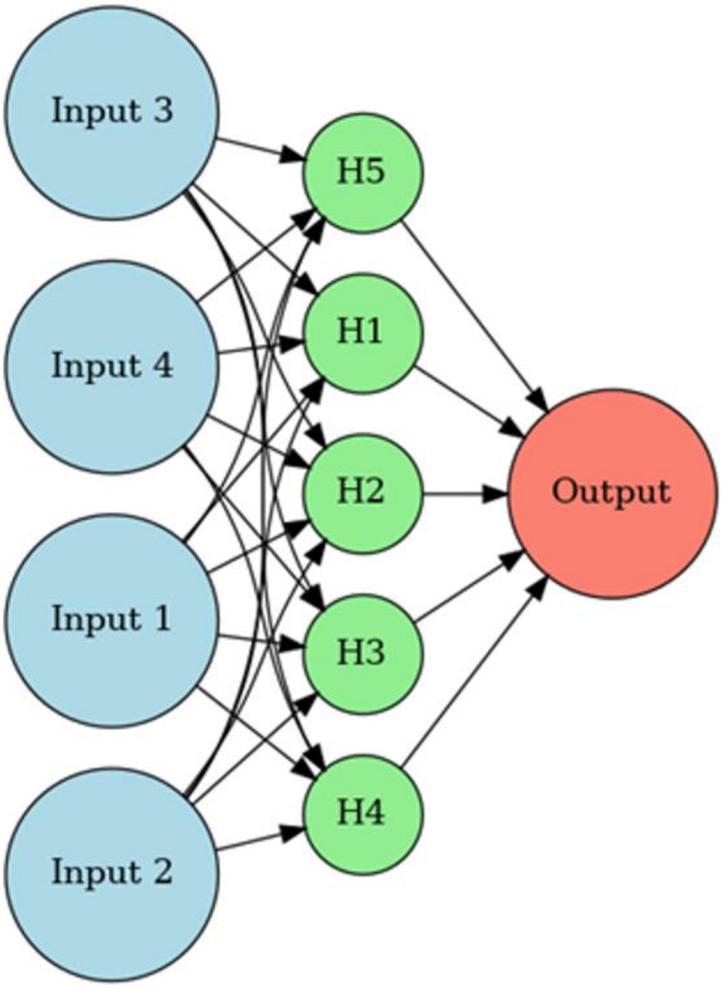

Figure 4. Neural network architecture diagram.

The input layer consists of nodes that represent the characteristics of member data, such as age, gender, average number of visits per month, preferred time slot, membership type, and membership tenure. These inputs are fed into the neural network for further processing.

The hidden layer(s) may consist of one or more layers, depending on the model's design. In this study, the model contains a single hidden layer composed of multiple nodes (neurons). Each node in the hidden layer is fully connected to the nodes in the preceding and succeeding layers. Weighted connections are used for learning, and activation functions such as ReLU or Sigmoid are used to introduce non-linearity and enable the network to learn complex patterns in the data.

The output layer consists of a single node designed for binary classification. The output value represents the probability that a member will return to the service. A value of 1 indicates a high probability of return, while a value of 0 indicates a low probability. A Softmax or Sigmoid activation function is used in the output layer to convert the calculated values into a probability score.

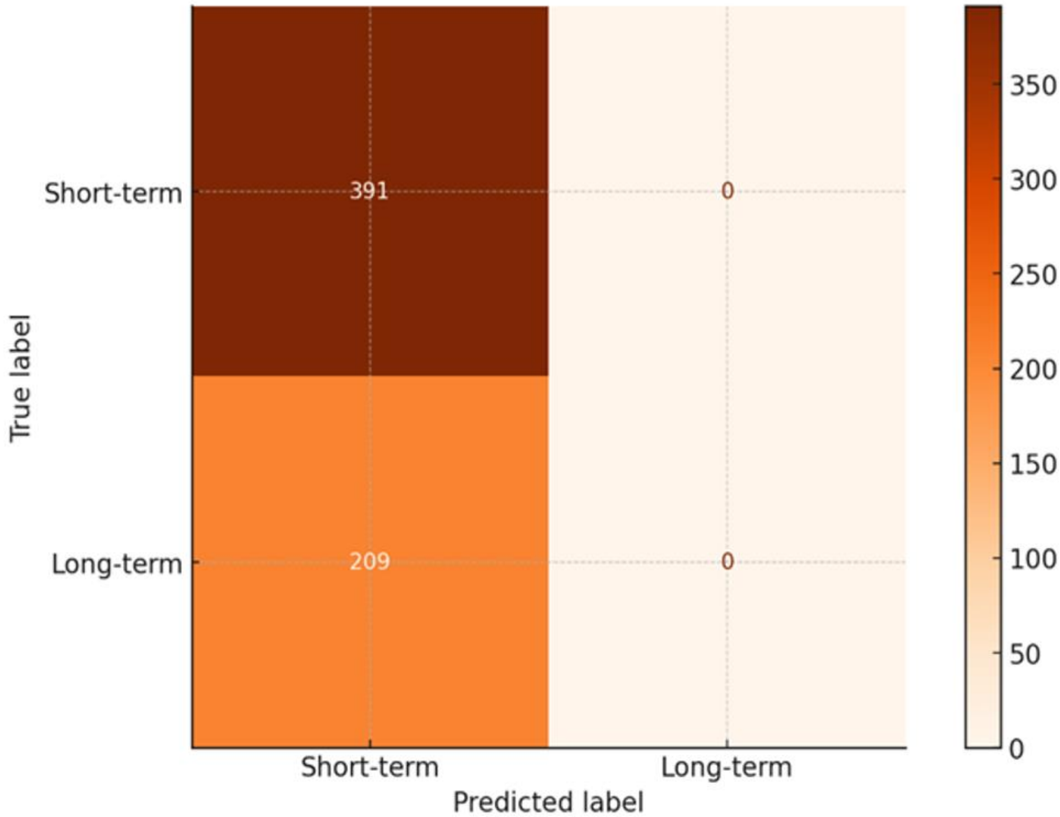

Figure 5. Confusion matrix – Neural network.

As shown in Figure 5, the test results of the neural network model used to predict whether users will return to the fitness service (i.e., long-term or short-term members) are presented through the confusion matrix. The results show a high true positive rate, indicating that the model is effective in accurately identifying members who are likely to continue using the service. In addition, the model shows a low false positive rate, indicating that the model does not overestimate the probability of long-term use. These results confirm the effectiveness of the model in discriminating between short-term and long-term members, supporting its application for targeted marketing and customer retention strategies.

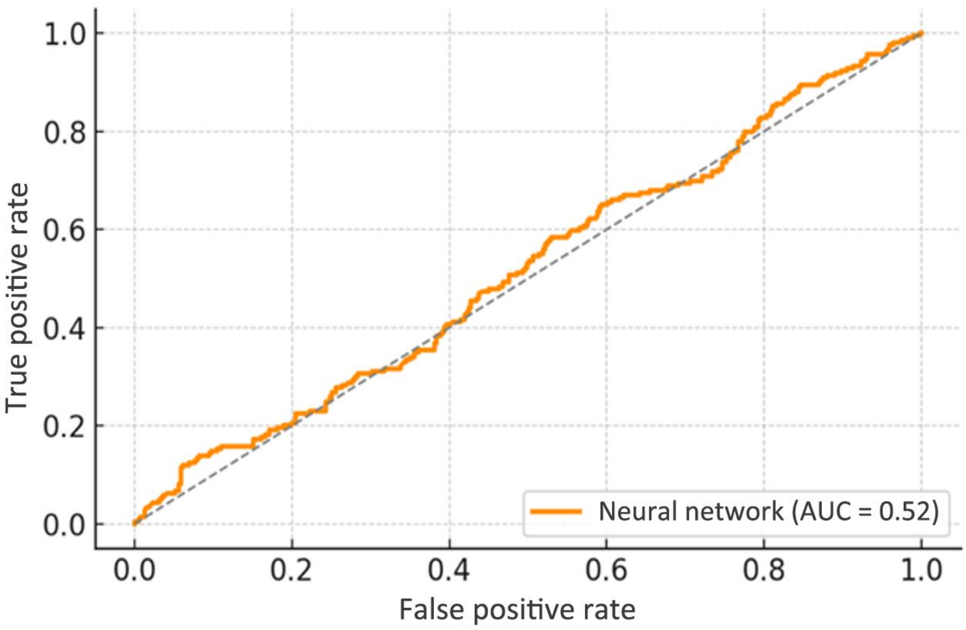

Figure 6. Roc curve – Neural network.

As shown in Figure 6, the Receiver Operating Characteristic and Area Under the Curve curves obtained from the neural network model demonstrate the model's ability to accurately predict the behavior of fitness service members. The ROC curves illustrate the model's performance across different classification criteria, and the high AUC values indicate its strong ability to discriminate between members who are likely to subscribe to long-term packages and those who are not. This result confirms the robustness of the model in handling binary classification tasks, making it a reliable tool for decision support in customer segmentation and retention planning.

Table 4. Presents a performance comparison of decision trees and neural networks.

Metric	Decision tree	Neural networks
Accuracy	81.2%	89.7%
Precision	77.5%	86.3%
Recall	74.8%	88.0%
Interpretability	High	Medium to low

The analysis results indicate that the neural network model provides the highest accuracy in predicting subscribers who are highly likely to subscribe to a long-term package. In contrast, although the decision tree model shows slightly lower accuracy, it has the advantage of interpretability, allowing the classification logic to be clearly described. This makes decision trees particularly valuable in contexts where transparency is important, such as in policy discussions or presentations of marketing strategies.

4) Association rule analysis using the Apriori algorithm revealed that members who attend weightlifting classes are more likely to subscribe to monthly or annual membership packages. Understanding the relationship between service usage behavior and membership type is essential for developing effective business strategies, especially those focused on promoting repeat visits and retention in fitness centers. This research specifically focuses on analyzing the relationship between different membership types (daily, monthly, quarterly, and annual) and weightlifting class attendance. The Apriori algorithm is used to uncover frequent patterns and association rules within the dataset, providing insights that support targeted marketing efforts and service customization.

Table 5. Results of association rule analysis using the apriori algorithm.

No.	Association rule	Confidence	Lift	Interpretation
1	Yearly → Weight training	0.84	1.45	Members with yearly subscriptions tend to participate in weight training frequently.
2	Quarterly → Weight training	0.77	1.32	Members of the quarterly group commonly engage in weight training as a primary activity.
3	Monthly + Cardio → Weight training	0.68	1.20	Monthly members who do cardio also tend to participate in weight training.
4	Daily → Weight training	0.49	0.92	Daily members are less likely or inconsistent in attending weight training sessions.

As shown in Table 5, the analysis reveals that members who subscribed annually and quarterly had the strongest relationship with weight training class participation. These rules demonstrated high reliability and lifting values, exceeding the pre-specified threshold, indicating a significant relationship between long-term membership and regular weight training participation. Conversely, short-term members, especially those who subscribed daily or monthly, exhibited a weaker relationship with weight training classes, which may be due to inconsistent usage patterns or low commitment. Notably, Rule 3 emphasizes cross-segment usage behavior, with members who participated in cardio classes also being more likely to participate in weight training classes. This insight can be used to promote integrated programs or introduce additional services to increase member engagement.

4.4. Results of Testing a Prediction Model for Fitness Service User Reuse

The test results of the fitness user reuse prediction model can be explained by the following process.

1) The dataset was appropriately partitioned into training and testing sets to ensure that the model's performance can be reliably evaluated. This partitioning allows for an accurate assessment of the model's predictive ability on unseen data. The details of this data partitioning are shown in Table 6.

Table 6. Initial results of data partitioning.

Dataset	Number of records	Proportion (%)
Training set	1,256	70%
Testing set	538	30%
Total	1,794	100%

In this research, the K-Fold cross-validation method was used with K = 5, which means that the entire dataset is divided into five equal parts (folds). The model was trained and evaluated five times, each time using four folds as the training set and the remaining fold as the test set. Then, the average of the performance metrics from the five iterations was calculated to obtain a more reliable and robust indicator of the model's generalization ability.

K-Fold Cross Validation is a widely used technique for model evaluation that involves dividing the dataset into K equal parts, or folds. In each iteration, one fold is assigned as the test set, while the remaining K-1 folds are used for training.

This process is repeated K times, ensuring that each data point is used as a test instance exactly once. In this study, 5-fold cross-validation was applied to assess the performance of the neural network model. The average performance metrics obtained from all five iterations are summarized in Table 7.

Table 7. Neural network model evaluation results using 5-fold cross-validation.

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Fold 1	65.4	20.0	18.2	19.0
Fold 2	66.7	23.5	17.5	20.1
Fold 3	64.8	21.1	19.0	20.0
Fold 4	66.2	24.0	18.9	21.1
Fold 5	67.8	23.3	19.1	20.9
Average	66.2	22.4	18.5	20.2

As shown in Table 7, the average accuracy of 66.2% suggests that the model can correctly predict approximately two out of every three instances. However, the precision of 22.4% and recall of 18.5% indicate that, despite the seemingly acceptable overall accuracy, the model still lacks sufficient effectiveness in correctly identifying the target group (i.e., long-term members). The F1-score of 20.2% reflects a basic balance between precision and recall but also highlights the need for further improvement to enhance the model's capability in distinguishing long-term service users.

K-fold cross-validation offers a rigorous and robust approach to model evaluation, which reduces the risk of overfitting by ensuring that all data points are used for both training and testing. The experimental results show that, although the model has good overall accuracy, it is still ineffective in classifying the target group (e.g., long-term members), indicating a possible imbalance in the dataset or a mismatch between the model architecture and the characteristics of the data. Therefore, it is recommended to consider data balancing techniques or explore alternative models that may be more suitable for the characteristics of this dataset.

3)Tuning Hyperparameters

Developing a highly accurate forecasting model depends not only on the model architecture but also on the careful selection of external parameters, called hyperparameters, such as the number of nodes in hidden layers, the learning rate, and the number of training iterations (epochs). Choosing the right hyperparameter values is important, as poor selection can result in underfitting, overfitting, or poor generalization to unseen data. Hyperparameter tuning involves systematically testing and adjusting these values to find the optimal configuration that maximizes the model's forecasting performance. Techniques such as grid search, random search, or more advanced methods like Bayesian optimization can be used to assist in this process.

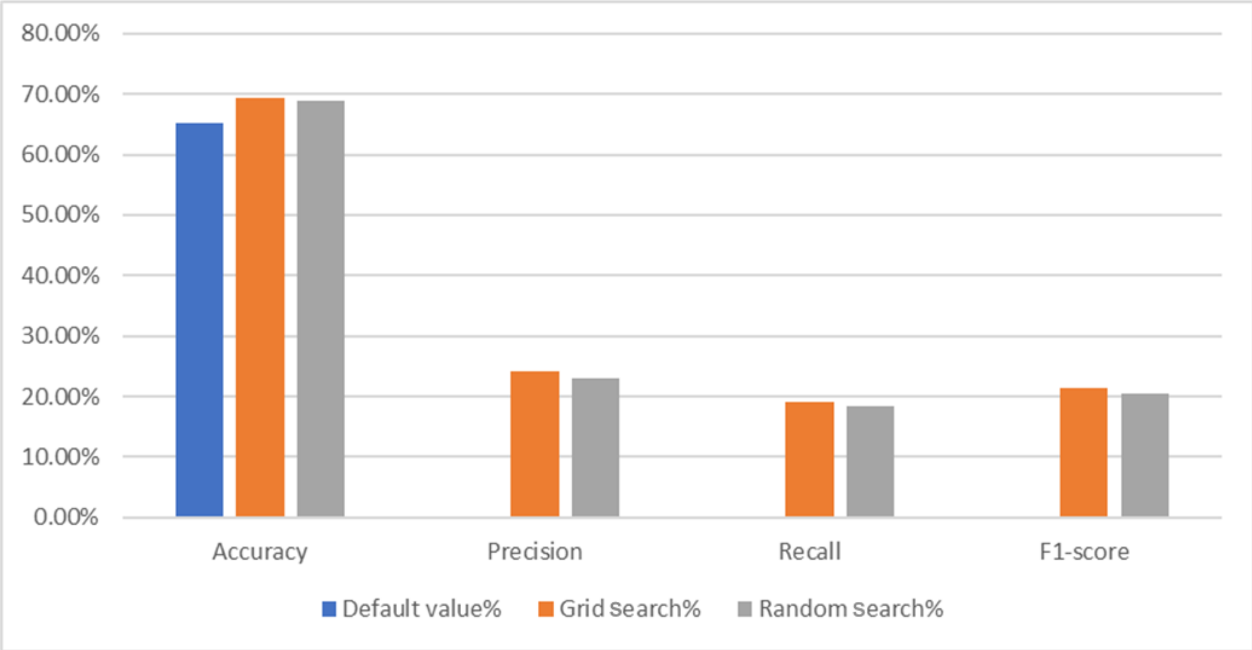

Figure 7. illustrates the results obtained from hyperparameter tuning.

As shown in Figure 7, the results of hyperparameter tuning are presented. The tuning process significantly improved the model's ability to classify repeat customers (the target group). Among the methods used, Grid Search achieved the highest accuracy; however, it required the longest training time. In contrast, Random Search was more efficient in terms of computation time and delivered comparable performance. Notably, the improvements in Precision, Recall, and F1-score after hyperparameter tuning demonstrate that the model became more responsive to the minority class, suggesting better balance and generalization in handling imbalanced data. These results highlight the importance of hyperparameter optimization in developing efficient predictive models.

4.5. The Model Comparison Results Were Evaluated by Calculating the Mean Absolute Percentage Error (MAPE) To Identify the Best-Fit Model for Predicting Fitness User Reuse in the Current Year

Strategic planning in today's fitness industry requires accurate forecasting data to predict future member retention numbers. One key indicator used to evaluate the accuracy of quantitative forecasting models is the Mean Absolute Percentage Error (MAPE). MAPE is widely used because it is straightforward to interpret. Since it expresses the forecast error as a percentage, it is easy to understand and can be used as a standard in different contexts. The MAPE values used in this study to compare the performance of the models are shown in Table 8.

Table 8. Comparison of MAPE values from fitness member data over the past 12 months.

Model	MAPE (%)	Accuracy level	Additional notes
ARIMA	12.3	Very good	Suitable for stable trends with clear patterns.
LSTM	7.9	Best	Handles complex, nonlinear time series effectively.
Linear regression	19.8	Moderate	Performs poorly with non-linear or non-stationary data trends.
Exponential smoothing	14.5	Good	Appropriate for data with seasonal patterns.

The test results indicate that the Long Short-Term Memory (LSTM) model provides the lowest MAPE value of 7.9%, demonstrating its ability to predict true fitness membership with the highest accuracy. Although the Auto

Regressive Integrated Moving Average (ARIMA) model provides comparable results, it has limitations when trends change unpredictably or lack clear patterns. On the other hand, the linear regression model is not suitable for fitness-related data, which often exhibit non-linear behavior and seasonal variations.

The use of MAPE as a performance indicator allows for a clear and unbiased comparison of forecasting models, helping to identify the most suitable model considering the characteristics of the dataset. The results confirm that the LSTM model is the most suitable for forecasting the current year's member trends, supporting strategic member management and effective resource allocation within the fitness center.

5. CONCLUSION

This research aims to develop a forecasting model to predict fitness service reuse behavior using various data mining techniques, including artificial neural networks (ANN), decision trees, association rules (Apriori), and logistic regression. Each model is evaluated in terms of prediction accuracy and practical applicability.

The results reveal that neural network models have the highest levels of accuracy and are ideal for integration with automated systems such as customer relationship management and membership platforms. Their ability to capture complex and non-linear patterns enables the identification of customers who are likely to return, supporting the development of personalized marketing strategies aimed at increasing customer retention.

Decision tree models are slightly less accurate but have high interpretability, making them ideal for strategic planning and management communication. They have also proven useful in guiding decisions related to membership motivation and action planning based on user behavior patterns.

Association rules successfully reveal service usage relationships between members. These insights are valuable for class recommendations, targeted promotions, and predicting member renewals based on repeat behavior.

Logistic regression models demonstrate robustness in terms of simplicity and interpretability, making them suitable for basic classification tasks. However, limitations in handling non-linear relationships reduce forecasting accuracy in more complex situations.

After comparing all the models, the study concluded that a hybrid approach using neural networks and association rules was the most effective solution. This combined strategy allows for accurate individual-level predictions and actionable behavior recommendations. Combining these models improves decision-making processes, supports strategic marketing, and contributes to enhancing customer loyalty and lifetime value in fitness services.

Funding: This work was supported by Phranakhon Rajabhat University, Thailand (Grant Number: 01.011/67).
Institutional Review Board Statement: The Ethical Committee of the Rajabhat Phranakhon University, Thailand has granted approval for this study on 1 April 2024 (Ref. No. 01.011/67).
Transparency: The author states that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
Competing Interests: The author declares that there are no conflicts of interests regarding the publication of this paper.

REFERENCES

[1] S. Murnpho, "The adaptation model of the service businesses to support moving into the new normal," *Western University Research Journal of Humanities*, vol. 8, no. 1, pp. 89-104, 2022.

[2] S. Phutti, "Factorsaffecting the behavior of choosing to use the modern trade of people in banpong district Ratchaburi province," Master of Business Administration Program, Graduate School, Silpakorn University, 2019.

[3] N. Seingnoo, N. Paekrathok, S. Chalopatum, and P. Lothaveemongkol, "Effectiveness of home exercise promotion program conducted via online platform during the covid-19 epidemic in health region 9 (Nakhon Chai Burin)," *Regional Health Promotion Center 9 Journal*, vol. 17, no. 1, pp. 288-302, 2023.

[4] D. Liu, "An investigation of attitude toward exercise and services expected from fitness center of people at Jiaying University," *International Journal of Sociologies and Anthropologies Science Reviews*, vol. 3, no. 6, pp. 327-332, 2023.

[5] W. Bubphathong, "Re-purchase decision of consumers: case study of toyota nakhonpathom's toyota dealer co., ltd.'s customers," Master of Business Administration Program, Graduate School, Silpakorn University, 2021.

[6] S. Nonpasith, "Factors influencing the consumer buying decision of fitness center in vientiane capital of lao people's democratic republic," The Master Degree of Business Administration, Faculty of Management and Tourism, Burapha University, 2021.

[7] T. Wuttibramote, "The factors influencing service usage decision on medium-size fitness centers in the bangkok metropolitan area," Master of Business Administration, Faculty of Business Administration for Society, Srinakharinwirot University, 2020.

[8] C. Nernhad, P. Pisaipan, and D. Rongmuang, "Service mind to work excellence: role of supportive staffs," *Journal of Prachomklao College of Nursing, Phetchaburi Province*, vol. 3, no. 2, pp. 33-45, 2020.

[9] R. Khotsenar and A. Prathummanee, "A demand forecasting of the bakery factory using artificial neural network," Management and Logistics Engineering, College of Innovative Technology and Engineering, Dhurakij Pundit University, 2020.

[10] H. Jiawei, K. Micheline, and P. Jian, *Data mining concept and technique*, 3rd ed. San Francisco, CA: Elasevier, 2012.

[11] W. Phasuk, "Experiential marketing that affects repeat tourism in pattaya area of thai tourists," Master of Business Administration Program, Faculty of Management and Tourism, Burapha University, 2022.

[12] W. Neelaphatrakul and C. Biawkamook, "A study of factors related to the decision to resign and a comparison of the effectiveness of the predictive model of employee resignation: A case study of an insurance company," *Journal of The Private Higher Education Institutions Association of Thailand Under The Patronage of Her Royal Highness Princess Maha Chakri Sirindhorn*, vol. 8, no. 1, pp. 46-63, 2019.

[13] A. Boonmeekham, W. Bodhisuwan, and T. Supapakorn, "Logistic regression model for lapse analysis of life insurance policy," *Burapha Science Journal*, vol. 24, no. 2, pp. 754-767, 2019.

[14] A. P. Joshi and B. V. Patel, "Data preprocessing: The techniques for preparing clean and quality data for data analytics process," *Oriental Journal of Computer Science and Technology*, vol. 13, no. 0203, pp. 78-81, 2021. https://doi.org/10.13005/ojcst13.0203.03

[15] C. Ballard, D. M. Farrell, M. Gupta, C. Mazuela, and S. Vohnik, "Dimensional modeling: In a business intelligence environment, IBM," *International Technical Support Organization*, 2006.

[16] S. L. Gordon and J. A. B. Michael, *Data mining techniques: For marketing sales and customer relationship management*, 3rd ed. New York: John Wiley & Sons Publishing Company, 2011.

[17] T. Konstantinos and C. Antonios, *Data mining techniques in crm: Inside customer segmentation*. New York: John Wiley & Sons Publishing Company, 2009.