



Predicting the Attraction and Retention of Customers in Sports Pools in Isfahan City Using a Decision Tree: Presenting a Data Mining-Based Model

Seyed Masoud Mirsaeidi¹, Davood Nasr Esfahani²

1. Department of physical Education and Sports Sciences, Isfahan (Khorasgan) Branch, Islamic Azad University Isfahan, Iran
da_n_e@yahoo.com
2. Department of Physical Education and Sports Sciences, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran
Davood_nasr_esfahani@yahoo.com (Corresponding Author)

Article Info

Abstract

Article type: Research Article

Received:
09 Oct 2024

Accepted:
20 Dec 2024

Published online:
25 Dec 2024

The purpose of this research is collective learning techniques and a combination of classification algorithms in data mining, evaluating the performance of these algorithms to predict whether customers of Isfahan sports pools would drop over a 6-year period. The collected data pertained to 54,000 customers and covered 22 primary characteristics. To evaluate the features, two algorithms—Grasshopper Optimization Algorithm (GOA) and Simulated Annealing (SA)—were used for feature selection, while classification algorithms, specifically Decision Tree (DT) and K-Nearest Neighbors (KNN), were employed to classify and recognize customer behavior. The results indicated that the combined GOA-DT algorithm, with an accuracy of 90.91%, outperformed the SA-DT algorithm, which achieved an accuracy of 87.95%. Furthermore, the combined GOA-DT algorithm selected only 7 out of the 22 features that were effective in detecting customer churn, whereas the SA-DT algorithm led to the selection of 8 features. In the subsequent step, the KNN algorithm was used instead of the DT algorithm. The results demonstrated that the KNN algorithm, when combined with GOA and SA, ultimately reached an accuracy of 83.86% for the GOA algorithm, with the selection of 9 features. Based on these findings, it can be concluded that the combined GOA-DT algorithm exhibits the best performance, identifying 7 key characteristics: average monthly recharge of the usage account, number of ticket purchases, customer satisfaction level, discounts received, service quality, monthly services, and free training. Therefore, it is recommended that managers make decisions based on these identified characteristics to retain customers and estimate customer behavior using such methods.

Keywords:

Customer Behavior, Data Mining, Decision Tree, Isfahan, Sports Pools

Cite this article:

Hosseini, S. E. , Khadem, A. , SoleymaniThehrani, M. , Kalashi, M. and Pourkiani, M. (2024). Requirements for Volunteers' Participation in Sports. *Archives in Sport Management and Leadership*, 2(2), 51-61. doi: [10.22108/asml.2025.143364.1056](https://doi.org/10.22108/asml.2025.143364.1056)



Introduction

In today's competitive world, customers are one of the most important assets for any organization and play a crucial role in enhancing market competitiveness and organizational performance (Bi, 2019). Amidst intense market competition, customers can easily make specific choices among multiple products or service providers. Studies show that the cost of acquiring a new customer is often higher than the cost of retaining an existing one. If a company maintains a good relationship with its customers over time, it will gain more profit from its current clientele. Therefore, to maintain market advantages, determining how to utilize existing customer resources and prevent the loss of current customers has become both important and necessary for companies (Xiahou & Harada, 2022).

Customers, in addition to their significance in physical markets, are also vital in electronic commercial markets. In other words, customers in e-commerce are the driving force behind the dynamics of online trade. Despite the considerable increase in online businesses compared to previous years, some users still lack the necessary trust in online shopping and prefer to purchase their required services through traditional means. Consequently, attracting positive customer feedback and motivating managers will lead to indirect marketing of their products. Thus, retaining and strengthening customer loyalty is a strategic challenge for organizations aiming to maintain and develop their competitive position in the market. Companies that focus not only on short-term sales but also on long-term customer satisfaction by providing valuable and differentiated products and services will naturally achieve greater market penetration and cultivate more loyal customers compared to their competitors (Mohammadi Javadi & Noorbakhsh, 2018). In this regard, there are two fundamental approaches that companies can use to attract and retain customers. The first approach is the "unfocused" approach. In this method, a company seeks to improve the quality of its product and relies on mass advertising to reduce customer churn. The second approach is the "focused" approach, in which the company targets its marketing campaigns toward customers who are more likely to churn. This approach can be further divided based on how customers are targeted. Marketing managers can establish favorable and long-term relationships with customers by identifying and predicting changes in customer behavior. Understanding these changes can help managers create effective advertising campaigns (Haghighatnia et al., 2018). Given the increasing competition in various industries, predicting customer behavior and understanding the factors influencing their attraction and retention is crucial for the success of businesses. This is also true in the sports and recreation industry, where sports complexes face the challenge of attracting and retaining customers. Therefore, companies need to evaluate the value of their customers, segment them based on their values, and develop strategies for each segment to acquire and retain profitable customers (Cheng & Lin, 2015).

In today's competitive and challenging markets, identifying potential customer churn and providing early warning indicators of problems that could lead to customer loss is essential (Cheng et al., 2019). By analyzing customer behavior, it has been established that an expert and relatively effective system for early detection of customer churn can assist businesses in addressing issues before they escalate. In the realm of predicting customer churn, Raeisi and Sajedi (2020) state that customer churn is a crucial criterion for evaluating a growing business, and companies need to predict it effectively to retain their customers. They also assert that the Gradient Boosted Trees method, with an accuracy of 90.86%, is highly accurate compared to other methods and can significantly benefit these types of businesses. Furthermore, some researchers believe that a customer's tendency to churn and switch in e-commerce depends on the following factors: the value paid for the first order, the number of items purchased, shipping costs, and categorization of purchased products, customer demographics, and customer location. However, the tendency of customers to switch is not influenced by factors such as population density in the customer's area and the division into rural and urban areas, as well as the quantitative analysis of the first purchase (Matuszelański & Kopczewska, 2022). Amajuoyi et al. (2024) in a study, pointed to the wider application of predictive analysis to increase customer loyalty and expand sports businesses. Suhandia et al. (2022) in their research on customer retention with increasing competition, they emphasized the direct relationship between customer satisfaction and retention rate. If a customer is not satisfied, the retention rate will automatically decrease. If a company fails to meet customer expectations, it will have a serious impact on the company, namely the transfer of customers to other services; such as service, price, value for money, satisfaction, and trust affect customer retention. The algorithm proposed by De Caigny et al. (2018) introduces a new combined classification method for predicting customer churn based on logistic

regression and decision trees. In this study, a new combined algorithm called the LogitLeaf model was suggested for improved data classification. The core idea behind the LogitLeaf model is that building different models on segments of the data, rather than on the entire dataset, results in better predictive performance while maintaining the interpretability of the models constructed on the leaves. The LogitLeaf model consists of two stages: the partitioning stage and the prediction stage. In the first stage, customer segments are identified using decision rules, and in the second stage, a model is created for each leaf of the tree. This new combined approach is analyzed alongside decision trees, logistic regression, random forests, and logistic model trees in terms of predictive performance and interpretability. The area under the receiver operating characteristic curve and lift charts are utilized to measure predictive performance, where the LogitLeaf model significantly outperforms logistic regression and decision trees, while performing at least as well as the other models.

In the study by Bamdad and Hatami (2017), it is stated that various types of neural networks can be employed for predicting customer churn, and data mining techniques are effective for this purpose. Additionally, the multilayer perceptron neural network model demonstrates higher accuracy compared to composite models. Moradzadeh and Khodayari (2017) also utilized decision tree and data mining methods to investigate customer churn in electronic banking services. They employed three algorithms—C&R TREE, QUEST TREE, and CHILD TREE—to predict customer churn, identify patterns leading to churn, and determine the most significant influencing factors. The researchers found that the C&R TREE algorithm outperformed the other algorithms in predicting customer churn. Furthermore, based on the decision trees formed and the percentage of churn at each node, rules leading to customer churn can be uncovered. According to the research findings, five crucial predictors of customer churn are occupation, branch level, education, current balance, and investment type. Banks should pay particular attention to these factors. The research results of Ali Beignejad and Haj Mohammadi (2019) on predicting customer churn in social networks using data mining analysis (including decision trees, neural networks, and regression) indicated that the proposed approach by the researchers demonstrated high usability, and their combined method achieved greater accuracy compared to other techniques. Regarding the predictive model of customer churn in online businesses, the research by Rostami Fard and Sajoudi Shijani (2021) utilized an aggregative classifier based on neural networks, with neural networks as the base learner. They incorporated classification algorithms such as decision trees, neural networks, naive Bayes, and support vector machines. Their proposed method showed superior performance, achieving an accuracy of 97.86% in predicting customer churn compared to other techniques. Aldosary & Alrashdan (2021) conducted a study on predicting gym membership churn using artificial neural networks, focusing on the concept of psychological habit formation in the fitness industry. The researchers aimed to develop a model that predicts gym membership churn using multilayer perceptron artificial neural networks with back propagation, emphasizing feature selection. Their results demonstrated that integrating the concept of psychological habit formation significantly enhanced the effectiveness of the neural network model in fitness maintenance strategies, achieving high predictive performance with accuracy, sensitivity, and specificity of 92.1%, 89.1%, and 93.8%, respectively. Xiahou et al. (2022) studied predicting customer churn in e-commerce using K-means clustering and support vector machine techniques. They divided customers into three groups and identified the main customer segments. The research findings revealed that each predictive indicator significantly improved after customer segmentation, highlighting the importance of using the K-means clustering algorithm. Furthermore, the results indicated that the accuracy of support vector machine predictions was higher than that of logistic regression predictions.

Techniques for predicting customer acquisition and retention can be utilized to identify customers who may be at risk of churn. Marketing strategies can be enhanced based on these predicted outcomes (Pondel, 2021). In the sports industry, customers play a vital role as the core of sports venues and as the driving force behind their success. Despite various challenges, national customer satisfaction indices have improved in most countries, and international satisfaction metrics have also developed globally. In Iran, providing quality services has consistently been a top priority for customers, reflecting their strong loyalty to high-quality offerings (Nasr Esfahani et al., 2024). Consequently, in an environment where customers are informed and possess the power of choice, neglecting their needs is no longer feasible. Currently, the market power balance is shifting toward customers; therefore, successful centers will be those that offer services that satisfy them. Retaining loyal customers over the long term is more beneficial than attracting new ones to replace those who have severed ties with sports facilities, particularly pools (Mohammadpoori et al., 2015).

In recent years, the water sports and recreation industry in Iran, especially in major cities like Isfahan, has witnessed significant growth. With the increase in competition among sports pools, attracting and retaining customers has become a fundamental challenge for the managers of these centers. Therefore, finding effective and scientific solutions to improve customer experience and increase their loyalty is essential in this field. Studies conducted in previous research indicate that despite the importance of customer retention and attraction, and the use of machine learning-based methods, educational methods based on various training and testing data that determine results in any condition without explicit planning (Patel & Prajapati, 2018), there is still a gap in the use of a precise approach to reduce feature dimensions and high accuracy in predicting customers of sports pools in Isfahan, which indicates a gap in scientific research. The present research, aimed at predicting the attraction and retention of customers of sports pools in Isfahan, is of particular importance. The use of data mining algorithms, especially decision trees, can significantly help in identifying patterns and factors influencing customer choice and loyalty. These tools can more accurately analyze customer needs and behaviors and, as a result, provide powerful models for predicting future customer behaviors. Additionally, the lack of sufficient research in this area and the scarcity of up-to-date and reliable data on the sports industry in Isfahan make this research even more necessary. On the other hand, the results of this research can help managers of sports pools design their marketing and service strategies in such a way that they not only attract new customers but also retain existing ones. Ultimately, this research can, in addition to helping improve the economic performance of sports pools, lead to increased customer satisfaction and the enhancement of sports services in society. Therefore, given the significant benefits in estimating the attraction and retention of customers of sports pools in Isfahan, this study aims to provide a fundamental tool for identifying the cognitive population of customers by utilizing a data mining approach, particularly decision trees, where each node represents a feature, each link represents a decision (rule), and each leaf represents a result (classification value or continuation); a model for evaluating the attraction and retention of customers of sports pools in Isfahan based on their behavior will be presented.

Research Methods

The current research method is based on data mining, utilizing the decision tree algorithm for customer feature classification. Additionally, the grasshopper optimization algorithm was employed to select optimal features. After reviewing the theoretical foundations and scientific research related to the topic, data preprocessing steps such as normalization, noise removal, and labeling were conducted to prepare the data for examination and implementation of the desired algorithms. Finally, various combined techniques derived from machine learning and meta-heuristic methods were used to evaluate the data and investigate the customer churn rate. In the data mining method, the term CRISP refers to standard industrial processes for data mining. There are various analytical methods for executing data mining projects. The CRISP analysis method is one of the most flexible and widely used approaches in this field, making it the preferred choice among other methods. In this study, the CRISP data mining method was carried out in six stages. Based on the steps of data mining, the following actions were taken to present a predictive model over 12 months in the studied organizations (sports pools in Isfahan City): Data Collection, Data Preparation and Cleaning, Data Encoding, Determining the Target Variable, Determining the Predictive Model Variables, Determining the Coefficients of Variables and Their Values, Data Initialization, Decision Tree Formation and Selection of the Optimal Conditions for Prediction Model Implementation, Evaluation.

Data Collection: The collected data was prepared for the data mining algorithm by filling in missing values and normalizing the data. This stage involves two basic steps: identifying the variables and sub-variables whose data can be collected. After interacting with the individuals from the organization in question (specifically, active customers of sports pools in Isfahan), the identified variables include:

Input Variables for the Model: The model includes several input variables (predictors) related to customer behavior and service usage at the pool. These variables are defined as follows: Pool Services Options (PO): Customer purchase methods, which can include POS, internet banking, mobile banking, interbank transfers, or payment gateways. These qualitative variables are labeled numerically. Average Monthly Recharge (MM): The average financial charge in the user panel for purchasing tickets via the internet gateway. Highest Amount Used (HP): The highest amount spent by the customer in a month. Amount of Ticket Purchase (BA): The total

amount spent on ticket purchases by the customer in a month. Number of Ticket Purchases (BN): The total number of tickets purchased by the customer in a month. Ticket Purchase Volume (BV): The total volume of all purchased tickets, including pool, sauna, Jacuzzi, etc., in a month. Customer Satisfaction Level (CS): A qualitative variable representing customer satisfaction, labeled numerically. Received Discount (DR): The amount of discount received by the customer on each purchase. Customer Income (CI): The average income reported by the customer. Number of Different Purchased Services (BD): The diversity of services purchased by the customer in each transaction. Ticket Purchase Repetition (BR): The number of times the customer has made similar purchases in one month. Customer Relationship (CT): The level of customer relationship, including feedback, offers, and purchase suggestions, labeled numerically. Service Level (SL): The level of service provided to the customer, which is qualitative and labeled numerically. Quality of Services (QL): The level of service quality, labeled numerically. Type of Service Used (TS): The type of service utilized by the customer, labeled numerically. Monumental Services (MS): Types of monthly services offered to customers, such as recommended packages, prizes, incentives, and follow-ups, labeled numerically. Time of Service Utilization (DT): The duration of service utilization by the customer after receiving the ticket. Number of Referrals to Friends (FD): The number of times the swimming pool has been referred by existing customers to friends and acquaintances. Duration of Presence in the Pool (TD): The total hours spent by the customer in the pool environment. Number of Times Using the Coffee Shop and Poolside Restaurant (CD): The frequency of customer use of poolside recreational services. Holding Swimming Competitions (SM): The occurrence of periodic competitions to motivate customers. Free Lessons (FL): Free lessons provided by pool lifeguards to customers. Target Variable, Churn (Target Variable): This variable indicates whether a customer is classified as a churn or non-churn customer ([Table 1](#)).

Table 1- Features used in the research

Row	Variable Name	Variable Type
1	Options for using pool services (PO)	Discrete
2	Average monthly account recharge charge for using the pool (MM)	Discrete
3	Highest amount used (HP)	Discrete
4	Amount (amount) of ticket purchase (BA)	Discrete
5	Number of ticket purchases (BN)	Discrete
6	Volume of ticket purchases (BV)	Discrete
7	Customer satisfaction level (CS)	Discrete
8	Discount received (DR)	Discrete
9	Customer income (CI)	Discrete
10	Number of different services purchased (BD)	Discrete
11	Repeat ticket purchase for one pool (BR)	Discrete
12	Customer communication (CT)	Continuous
13	Service level (SL)	Continuous
14	Service quality level (QL)	Continuous
15	Type of service used (TS)	Continuous
16	Monthly services (MS)	Continuous
17	Time of service use (DT)	Discrete
18	Number of referrals to friends (FD)	Discrete
19	Duration of presence in the pool (TD)	Discrete
20	Number of times using the pools café and restaurant (CD)	Discrete

Row	Variable Name	Variable Type
21	Swimming competitions (SM)	Discrete
22	Free lessons (FL)	Discrete
	Target variable/label/Target	Churn Binominal 0, 1

The process of data preparation and cleansing is vital for machine learning, ensuring data accuracy and consistency. It involves several steps to transform raw data into a usable format for model training. The process includes data cleansing, encoding, normalization, and feature selection, each addressing specific data issues and preparing it for analysis. This is crucial for building effective models and making accurate predictions:

Part 1: Preprocessing: When the values of dataset features are in different domains, the likelihood of errors in the results increases. Normalization is the process of adjusting the data of a statistical population to a similar domain. In the proposed model, normalization is performed using the following formula. Normalization Formula: The standard form of normalization places all data between the interval d1 to d2. The formula for normalization is given by:

$$\text{Equation (1)} \quad \bar{x} = \frac{(x - x_{min})(d2 - d1)}{x_{max} - x_{min}} + d1$$

According to the data, d1= 0 and d2=+1. In other words, using this relationship, all data fall within the range [0,1]. In a dataset, there may be missing values for some records. The data in a dataset must be complete and free of missing or incomplete values when entered into an algorithm. Additionally, cases where values are likely assigned incorrectly to the features of a record should be corrected; if they cannot be corrected, they should be removed from the dataset. Unfortunately, the dataset does contain missing values. In this study, the maximum possible value method has been used to handle these missing values. In the maximum possible value method, among the acceptable values for a specific feature, the maximum value is chosen for substitution (Farhang Far et al., 2008).

Part 2: Feature Selection: After reading the dataset related to the customers of sports pools, and performing preprocessing operations on the data, the grasshopper optimization algorithm is formed. In this algorithm, N is the number of grasshopper colonies and D is the number of decision variables or dimensions of the optimization problem. Therefore, the grasshopper optimization algorithm is simulated by an N*D matrix. Each row corresponds to a possible solution to the optimization problem. In the proposed model, N is the number of records in the dataset and D is the number of features. The population of grasshopper colonies, which consists of a large number of grasshoppers and are responsible for exploring the objective, is defined according to equation (2). In the proposed model, the working method for the dataset is such that the grasshopper algorithm consists of N colonies, and each colony is made up of a number of grasshoppers (features). Each hive is defined by D features present in the database.

(2) Equation

$$\text{Population of GOA} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1d} \\ x_{22} & x_{22} & \dots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nd} \end{bmatrix}$$

In the set $x_i = (x_{i1}, x_{i2}, \dots, x_{id}), i = 1, 2, \dots, n$, each x_i represents a possible solution in the solution space. Each category of grasshoppers consists of a group of attacker ants that are considered as elements of a solution. All attacker grasshoppers in a category are considered as a general unit that moves towards a suitable location with abundant resources. If a category of grasshoppers reaches an ideal position, an optimal solution is obtained. The evaluation of each category of grasshoppers is calculated based on the objective function according to equation (3).

$$fit_i = 1 - \frac{Obj_i - worst(Obj)}{best(Obj) - worst(Obj)} \quad (3)$$

In equation (3), fit_i is the fitness of the i th category. The Obj_i parameter is the value of the objective function for the i th category. Each category is calculated based on the distance criterion. The parameters worst and best are the worst and best ant categories relative to the prey. In the proposed model, the grasshopper algorithm should be transformed from a continuous to a discrete state. This is because the dataset values are discrete and in the range of 0 and 1. Therefore, each grasshopper is defined in the category x_{ij} of the floor of equation (4). The population of categories is encoded with D features. To convert numbers to binary, the sigmoid function in equation (5) is used. The output of the sigmoid function falls within a specific numerical range (usually between zero and one). In this function, the answer will not be 0 or 1, but a set of numbers between zero and one.

$$x_{ij} = \begin{cases} 0, & \text{if } g(x) < 0.5 \\ 1, & \text{otherwise} \end{cases} \quad (4)$$

$$g(x) = \frac{1}{1+e^{-x}} \quad (5)$$

The accuracy parameter is the percentage of accuracy and the values of the parameters δ and ρ are constant and their values are equal to 99 and 1, respectively.

In the proposed model, a subset of features is selected using an optimization algorithm to achieve the optimal value. The fitness function for feature selection from each category is defined according to equation (6). In equation (6), $|n|$ is the total number of features, and $|S|$ is the number of selected features. The parameter accuracy represents the percentage of correctness and the values of the parameters and constants are 99 and 1, respectively.

$$Fitness = \delta \cdot Accuracy + \rho \cdot \frac{|n| - |S|}{|n|} \quad (6)$$

Dataset sinter printability with high dimensions, despite the opportunities they bring, creating many computational challenges. One of the problems with high-dimensional data is that in most cases, not all data features are vital for finding the hidden knowledge in the data. The main idea in feature selection is to eliminate a subset of input features that have little information. Therefore, feature selection is used to reduce the feature space and increase the efficiency of classification. Feature selection not only improves the accuracy and efficiency of classification, but also enhances the interpret ability of the results.

The Grasshopper Optimization Algorithm: Grasshoppers, despite being solitary in nature, form large groups that can be destructive to agriculture. Their unique group behavior is exhibited in both larval and adult stages, with larvae swarming and feeding on plants, and adults migrating in groups. This behavior has inspired optimization algorithms, such as the grasshopper optimization algorithm, which mimics their search and movement strategies (7):

$$X_i = S_i + G_i + A_i \quad (7)$$

The equation X_i defines the position of the i th grasshopper, S_i is the social interaction, G_i is the gravitational force on the i th grasshopper, and A_i is the horizontal wind force. To provide random behavior, equation (7) can be rewritten as $X_i = r_1 S_i + r_2 G_i + r_3 A_i$ where r_1 , r_2 and r_3 are random numbers in the interval $[0,1]$. The component S_i is defined by equation (8):

$$S_i = \sum_{\substack{j=1 \\ j \neq i}}^N s(d_{ij}) \hat{d}_{ij} \quad (8)$$

That d_{ij} is the distance between i th and j th grasshopper, which is $d_{ij} = |x_j - x_i|$ is calculated, s is a function that defines the power of social forces as shown in equation (9), and $\hat{d}_{ij} = \frac{x_j - x_i}{d_{ij}}$ is the unit vector from grasshopper i to grasshopper j .

The S function that defines the social force is calculated by equation (9):

$$s(r) = f e^{-\frac{r}{l}} - e^{-r} \tag{9}$$

That r is the absorption intensity and l is the absorption length scale.

The parameters l and f in equation (9) influence the social behavior of artificial grasshoppers. Figure 1 illustrates the effects of varying these parameters, revealing significant changes in behavior across different zones. The attraction and repulsion zones exhibit particularly sensitive responses to specific parameter values (for example, $l = 1$ or $f = 1$).

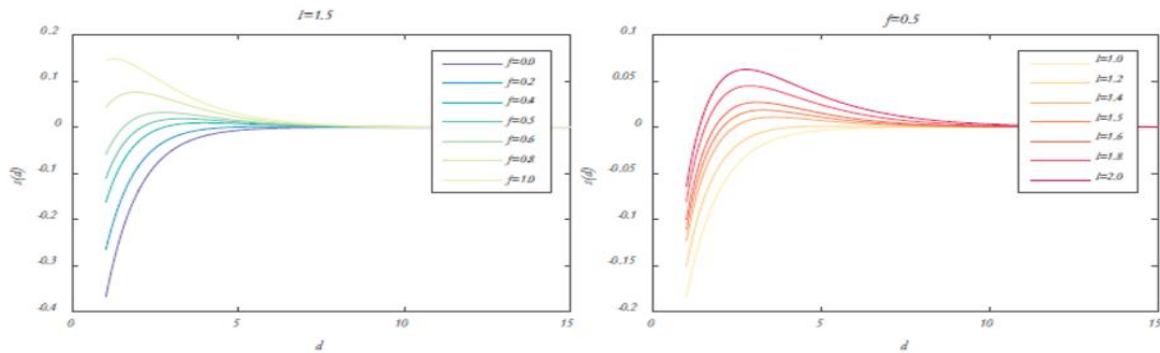


Figure 1. Behavior of the function s when changing l and f

The G component is calculated by equation (10):

$$G_i = -g \hat{e}_g \tag{10}$$

In which g is the gravitational constant and the \hat{e}_g is the unit vector is towards the center of the earth .

Part 3: Classification: For classification, it is necessary to first divide the dataset into two parts: training (80% of samples) and testing (20% of samples). The training data generates the evaluation model, and the testing data tests the model generated with the help of some records and determines the label of those records, thus identifying their class. For this purpose, the decision tree algorithm is used; which will be discussed further below.

Decision tree classifiers are a widely recognized and effective technique for data classification. This method is favored for its simplicity and accuracy, making it a popular choice in machine learning, image processing, and pattern recognition. (Charbuty & Abdulazeez, 2021).

Different types of approaches for decision tree exist, among the most important are methods such as Iterative Dichotomies 3 (ID3), Successor of the ID3 algorithm (C4.5), Classification And Regression Tree (CART), CHi-squared Automatic Interaction Detector (CHAID), Multivariate Adaptive Regression Splines (MARS), Generalized, Unbiased, Interaction Detection and Estimation (GUIDE), Conditional Inference Trees (CTREE), Classification Rule with Unbiased Interaction Selection and Estimation (CRUISE), Quick, Unbiased and Efficient Statistical Tree (QUEST).

The most important criterion in evaluating the performance of decision trees is the entropy criterion. Entropy is used to measure the impurity or randomness of a dataset. The value of entropy is always between 0 and 1. A value of 0 is better, while a value of 1 is worse, meaning the closer the value is to 0, the better. This index is calculated based on equation (11):

$$Entropy(S) = \sum_{i=1}^c P_i \log 2^{P_i} \tag{11}$$

That in this regard, P_i is equal to the rate of the number of subset samples and the value of the characteristic is (Charbuty & Abdulazeez, 2021).

Part 4: Evaluation Criteria: This study employs five metrics—accuracy, recall, measurement, precision, and error rate—to assess the effectiveness of a classification algorithm. These criteria provide insights into the

algorithm's performance, particularly its ability to correctly classify samples and its reliability in assigning labels (Kamel et al., 2019):

TP: Number of features correctly identified.

TN: Number of false positives detected.

FP: The number of correct features was incorrectly identified as wrong.

FN: Number of incorrect features mistakenly identified as correct.

- Accuracy: The ratio of correctly classified samples to all samples, which is calculated from the following relationship:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (12)$$

- Precision: It is the ratio of correctly classified positive samples to all available positive samples.

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

- Recall: It is the ratio of correctly classified positive samples to the total number of samples that have been diagnosed as positive. It should be noted that some of the samples that have been diagnosed as positive are wrong and are included in the FN collection.

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

This text introduces the need for a predictive model to forecast customer behavior in sports facilities, particularly in Isfahan's sports pools, to aid economic and management decisions. The research aims to develop and compare hybrid algorithms for accurate predictions until 2024, emphasizing the importance of such models in investment and management strategies.

Predicting the Attraction and Retention of Customers in Sports Pools that all results were obtained from programming in MATLAB 2021a software on a system with a Core i5 processor and 8GB of RAM.

Findings

Data Preprocessing: One of the most important stages of a data mining method is data preprocessing. In fact, preprocessing determines the results that will be achieved, and its significance is such that it can lead to either the best or the weakest outcomes. Therefore, in this study, preprocessing was conducted thoroughly according to established articles and in a principled manner, which includes the following steps:

1. Removing Noise and Outlier Data: During data collection, some columns may contain unreasonable data that need to be identified and corrected at this stage.

2. Data Sorting: Data should be sorted in a manner that is understandable and readable for MATLAB software. In this study, the number of rows represents the number of customers surveyed in Isfahan pools within a specified time period, while the number of columns represents the features related to the behaviors of these customers that lead to either slipping or not slipping.

3. Data Labeling: Since MATLAB software operates numerically, all variables need to be labeled numerically. Additionally, some qualitative variables under study must be evaluated alongside numerical variables. Therefore, all data is labeled numerically.

4. Data Normalization: Now that all variables have been converted to numerical formats, the data can easily be loaded into MATLAB. It is also important to note that since the scales of the data differ from one another, they should be transformed into a standard form. The standard form ensures that all data is placed within the

range of d1 to d2 using the following formula: Based on the selected data (d1=0 and d2=+1), it can be seen that all data within this range are standardized or normalized, so to speak. Consequently, all data with specified characteristics and labels is saved in a file called "Labeleddata.xls." It should be noted that based on the feature analysis, a column named "target" has been added to this Excel file, which contains the provided labels from the table above and is used for prediction.

Data Splitting: In this study, 70% of the data is used for training the network, while the remaining 30% is used to test the model. This splitting is completely random to ensure that all data is utilized in both tasks. The MATLAB software function can automatically create random indices and place the data corresponding to these indices in their respective matrices. Since the final number of data points is 54,000 rows, 70% of this amount equals 37,800. Therefore, 37,800 data points are used to create the classification model, while the remaining 16,200 data points (30% of the original data) are used to test the mode.

Creating Classification: The research employed decision tree and K-Nearest Neighbors classification methods, with feature selection via grasshopper and SA algorithms. The fusion of the Grasshopper Optimization Algorithm and decision tree (GOA-DT) achieved impressive results, with a high precision of 90.9091% and a low error rate (Figure 2). This hybrid method outperformed the SA-DT algorithm, showcasing its effectiveness in predicting customer churn with fewer features.

The KNN algorithm is introduced for feature classification, the results of this change are presented in Table 2, achieving an accuracy of 83.8636% when combined with the grasshopper algorithm (Figure 3). This is lower than the grasshopper-decision tree (GOA-DT) method, which has an accuracy of 90.9091%. The KNN-GOA combination, along with the SA algorithm, selects 9 features with the same accuracy. Tables showcase the selected features and comparative analysis of various algorithms' accuracy, solving time, and error rates (Tables 3-7). The GOA-DT algorithm is deemed superior for feature selection and customer behavior analysis.

Table 2- The results of applying the proposed algorithm (grasshopper-decision tree)

Amount	Parameter
284	TP
116	TN
20	FP
20	FN
%93.4211	Precision
%93.4211	Recall
%90.9091	Accuracy
%93.4211	F-Measure

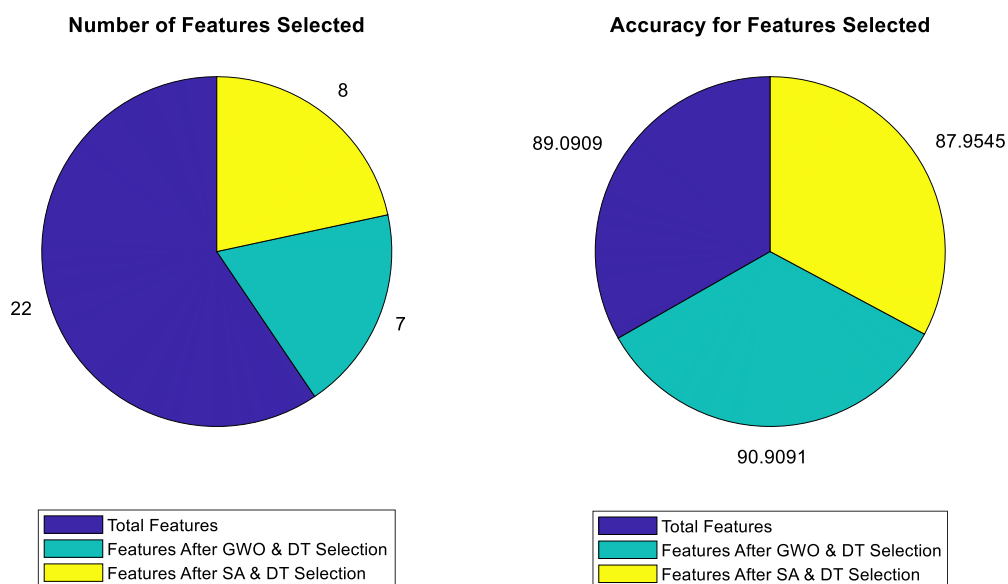


Figure 2. Comparison of the proposed algorithm's performance with the SA-DT algorithm

Table 3- Features selected by both GOA and SA algorithms in combination with the decision tree

SA Algorithm	GOA Algorithm
Monthly average recharge charge for pool account usage	Monthly average recharge charge for pool account usage
Amount of ticket purchase (BA)	Number of ticket purchases (BN)
Customer satisfaction level (CS)	Customer satisfaction level (CS)
Discount received (DR)	Discount received (DR)
Customer communication (CT)	Service quality level (QL)
Service level (SL)	Monthly services (MS)
Number of referrals to friends (FD)	Free tutorials (FL)
Free tutorials (FL)	-

Table 4- Results of applying the K-nearest neighbor algorithm to the grasshopper algorithm

Amount	Parameter
260	TP
109	TN
53	FP
18	FN
%83.0671	Precision
%92.5252	Recall
%83.8636	Accuracy
%87.9865	F-Measure

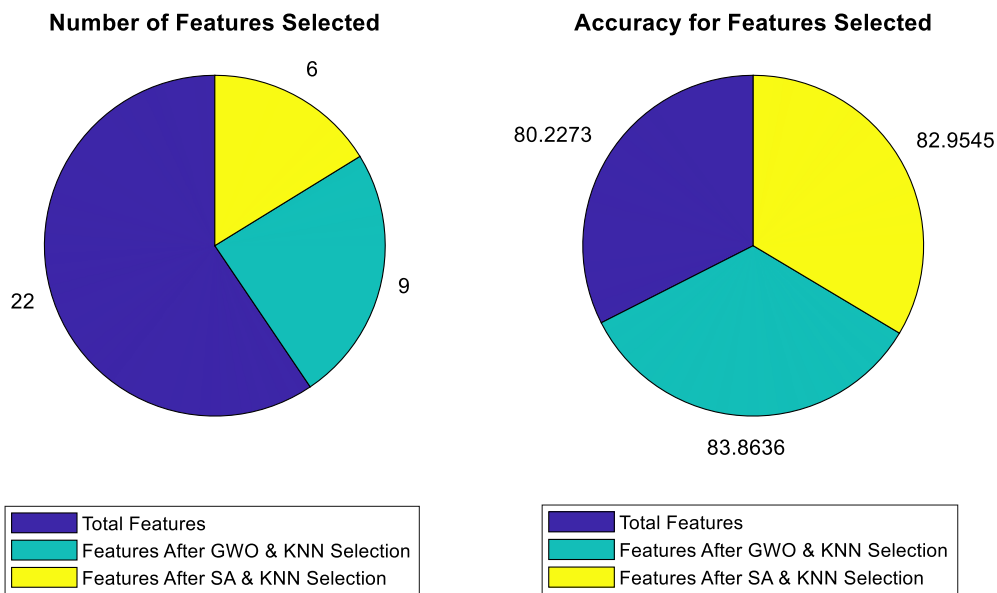


Figure 3. Comparison of the performance of the GOA-KNN algorithm with the SA-KNN algorithm

Table 5- Features selected by both GOA and SA algorithms in combination with K-nearest neighbors

Algorithm SA	Algorithm GOA
Monthly average recharge charge of pool account usage	Monthly average recharge charge of pool account usage
Customer satisfaction level (CS)	Number of ticket purchases (BN)
Discount received (DR)	Customer satisfaction level (CS)
Customer contact (CT)	Discount received (DR)
Service level (SL)	Service quality level (QL)
Free lessons (FL)	Monthly services (MS)
-	Time of service usage (DT)
-	Number of times using the café and pool restaurant (CD)
-	Free lessons (FL)

Table 6- Comparison of the accuracy of the proposed algorithm with the algorithms used

Number of features selected	Precision	Algorithm
7	90.9091	GOA – DT
9	83.8636	GOA – KNN
8	87.9545	SA – DT
6	82.9545	SA – KNN

Table 7- Error rates obtained from grasshopper-based algorithms

Error	Algorithm
0.0909	GOA – DT
0.1614	GOA – KNN

In addition, as shown in Table 8, the results of the study on non-churn customers in sports pools, aimed at investigating their churn in the years 2023 and 2024, are presented. It should be noted that a threshold of 0.6 has been established for estimating the likelihood of customer churn; this means that customers whose predicted target segment is equal to or greater than 0.6 are more likely to churn, while customers whose target value is less than 0.6 are considered less likely to churn. Consequently, values of 0 and 1 are assigned for non-churn and churn, respectively, as the target column in the original data also contained binary values of 1 and 0. Therefore, customers labeled with 1 are expected to churn by 2024.

Table 8- Customer Churn Forecast until 2024

Customer number	2023	2024	Churn/Non-churn	Customer number	2023	2024	Churn/Non-churn
1	0.1604	0.5562	0	66	1.0000	1.0000	1
2	0.3872	0.7346	0	67	1.0000	1.0000	1
3	0.8775	0.7786	0	68	0.3065	0.8307	0
4	0.3546	0.2641	0	69	0.3390	0.3728	0
5	0.3268	0.8677	0	70	0.4348	0.2883	0
6	0.2793	0.7018	0	71	0.0897	0.8551	0
7	0.7859	0.7311	0	72	0.3326	0.7494	0
8	0.9035	0.7230	0	73	0.6220	0.9614	0
9	0.5324	0.4963	0	74	0.1302	0.9586	0
10	0.8510	0.8276	0	75	0.3873	0.2544	0
11	0.9227	0.5211	0	76	0.8628	0.2063	0
12	0.1829	0.2342	0	77	0.6146	0.3018	0
13	0.6263	0.3391	0	78	0.0683	0.1866	0
14	0.2910	0.6451	0	79	0.8564	0.3754	0
15	0.1162	0.4851	0	80	0.4584	0.3669	0
16	0.5764	0.0465	0	81	0.1755	0.6384	0
17	1.0000	1.0000	1	82	0.6705	0.4601	0
18	0.8833	0.4219	0	83	0.4764	0.0503	0
19	0.6700	0.5678	0	84	0.1426	0.5441	0
20	0.3363	0.7636	0	85	0.3142	0.8035	0
21	0.2701	0.6063	0	86	0.3536	0.9538	0
22	0.3013	0.8621	0	87	0.0891	0.8649	0
23	1.0000	1.0000	1	88	0.9415	0.4259	0
24	1.0000	1.0000	1	89	0.6721	0.4429	0
25	0.4798	0.7985	0	90	0.6560	0.2851	0
26	0.6083	0.5813	0	91	0.5142	0.3472	0
27	0.0902	0.8847	0	92	0.1040	0.3086	0
28	0.6688	0.8260	0	93	0.2288	0.6348	0
29	0.2012	0.5130	0	94	0.6462	0.9975	0
30	0.7881	0.5670	0	95	0.3403	0.7781	0
31	0.4157	0.7573	0	96	0.2889	0.7639	0
32	0.4848	0.9752	0	97	0.1388	0.9550	0
33	0.8537	0.3275	0	98	0.3276	0.8500	0
34	0.9241	0.7061	0	99	0.0876	0.8165	0
35	0.2874	0.3274	0	100	0.2538	0.0553	0
36	0.9333	0.0894	0	101	0.4184	0.1916	0
37	0.8538	0.2192	0	102	0.8761	0.1558	0

Customer number	2023	2024	Churn/Non-churn	Customer number	2023	2024	Churn/Non-churn
38	1.0000	1.0000	1	103	0.6549	0.7151	0
39	0.5410	0.5196	0	104	0.6132	0.6031	0
40	0.9562	0.7111	0	105	0.3883	0.7802	0
41	0.0387	0.8336	0	106	0.8792	0.3254	0
42	0.1917	0.7742	0	107	0.5760	0.7127	0
43	0.7653	0.9505	0	108	1.0000	1.0000	1
44	0.3834	0.3014	0	109	0.7182	0.8477	0
45	0.5108	0.8054	0	110	0.4621	0.2626	0
46	0.3176	0.3221	0	111	1.0000	1.0000	1
47	0.9832	0.5790	0	112	1.0000	1.0000	1
48	0.9016	0.5813	0	113	0.9078	0.8909	0
49	0.9194	0.3573	0	114	0.5197	0.9784	0
50	0.6245	0.5157	0	115	0.2219	0.8588	0
51	0.6781	0.6426	0	116	0.3921	0.7270	0
52	0.1862	0.4231	0	117	0.1789	0.3416	0
53	0.2739	0.1485	0	118	0.1269	0.3603	0
54	0.3554	0.8720	0	119	0.9537	0.8123	0
55	1.0000	1.0000	1	120	0.1152	0.4318	0
56	0.5465	0.3009	0	121	1.0000	1.0000	1
57	0.6917	0.1154	0	122	1.0000	1.0000	1
58	1.0000	1.0000	1	123	0.7022	0.0430	0
59	0.3066	0.5487	0	124	0.2339	0.4792	0
60	0.1296	0.1793	0	125	0.9802	0.0711	0
61	0.8714	0.0795	0	126	0.1521	0.9311	0
62	0.6772	0.9549	0	127	0.2684	0.1322	0
63	0.5042	0.5721	0	128	0.7124	0.7155	0
64	0.8346	0.6538	0	129	0.7571	0.4292	0
65	1.0000	1.0000	1				

Discussion

In this study, a combination method based on the grasshopper Algorithm was used to enhance the accuracy of customer behavior estimation and prediction systems while reducing feature dimensions. First, the data obtained from the database was normalized following the preprocessing stages. In the next stage, using DT and KNN algorithms, combined with SA and GOA, the data was tested and evaluated. To assess the proposed method, criteria such as accuracy, detection rate, sensitivity, and performance rate were utilized, and each of these metrics was obtained for the desired algorithm. Given that the proposed method is based on the questions considered in this study, it can be briefly stated that: How can a decision tree-based model be presented to predict the level of attraction and retention of customers in sports pools in Isfahan City? The implementation of the decision tree method requires the collection of relevant data. Since the decision tree is one of the most common data mining methods, we collected data related to Isfahan sports pools over six years, from 2018 to 2023, for 54,000 customers. This allowed us to estimate customer behavior by implementing various stages of the algorithm, such as labeling, classification, and prediction.

On the other hand, since the decision tree is a powerful binary analysis method (0 and 1), the dataset was adjusted to reflect 0 (non-churn) and 1 (churn) customers, ensuring that the decision tree method could be implemented smoothly.

What features are effective in attracting and retaining customers of sports pools in Isfahan City? According to the study of articles and evaluations of experts' opinions, a total of 22 features were identified as significant for attracting and retaining customers of sports pools in Isfahan City. By applying the grasshopper Algorithm to select the most effective features, it was determined that ultimately 7 features are recognized as key indicators in this context: Average monthly account recharge, Pool usage, Number of, ticket purchases, Customer satisfaction level, received discounts, Service quality level, Monthly services Free training.

Can the proposed approach have better performance in terms of accuracy compared to previous methods? The results showed that the proposed method achieved a higher efficiency of 90.9091% accuracy compared to other methods.

Therefore, in today's economic world, having accurate and timely information is invaluable for owners, investors, creditors, and other stakeholders to make informed financial decisions. With the development of technology, it is now possible to use simple customer behavior prediction models for all sports centers and collections. The availability of straightforward yet powerful prediction tools can help owners prevent bankruptcy and take necessary actions to improve the situation regarding customer churn or retention. On the other hand, such tools can serve as a strong driver for selecting optimal investment portfolios for investors. Investors can better inform themselves about the past, present, and future of these centers. Predicting customer churn in sports centers is a crucial issue in financial decision-making within this sector. Given the effects and consequences of this phenomenon at both micro and macro levels in societies, various tools and models of significant importance have been developed, each differing in methods or variables for prediction, at both national and international levels.

In this study, the performance of algorithms in predicting customer churn or non-churn in Isfahan sports clubs over a 6-year period was evaluated using collective learning techniques and the combination of classification algorithms in data mining. The collected data pertains to 54,000 customers and includes 22 initial features. To evaluate the features, two optimization algorithms, GOA and SA, were employed for feature selection, while classification algorithms such as DT and KNN were utilized for classification and behavior recognition. The results indicated that the combined algorithm GOA-DT, with an accuracy of 90.9091%, outperformed the SA-DT algorithm, which achieved an accuracy of 87.9545%. Notably, the GOA-DT algorithm selected only 7 out of the 22 effective features for identifying customer churn, whereas the SA-DT algorithm resulted in the selection of 8 features. In the subsequent step, the KNN algorithm was used instead of the DT algorithm. The results showed that the KNN algorithm, in combination with GOA and SA, ultimately achieved an accuracy of 83.8636% for the GOA algorithm with the selection of 9 features.

Based on these findings, it can be concluded that the combined algorithm GOA-DT demonstrates the best performance, utilizing an average of 7 features: Monthly account recharge, Pool usage, Number of ticket purchases, Customer satisfaction level, Received discounts, Service quality level, Monthly services, Free training. These features can effectively be used to estimate customer behavior.

In Table 8, the results of investigating non-churning customers in sports pools to examine their churn in the years 2023 and 2024 are presented. It should be noted that a threshold of 0.6 has been set for estimating the likelihood of customer churn. This means that customers whose predicted target segment is equal to or greater than 0.6 are more likely to churn, while customers whose target value is less than 0.6 are assumed to have a lower likelihood of churning. Consequently, values of 0 and 1 are assigned for no churn and churn, respectively, as the target column in the original data also contained binary values of 1 (churn) and 0 (non-churn). Therefore, as indicated in Table 8, customers labeled 1 are predicted to churn by the year 2024. This information can be utilized by sports pool managers to make necessary decisions aimed at mitigating the negative impacts of customer churn. This issue is particularly significant given the current circumstances, which include specific restrictions such as epidemics, crises, and unforeseen events. Additionally, families are increasingly sensitive to being in sports environments like pools, and there has been a decline in the financial capacity of recreational sports families, resulting in a decreased tendency to use sports pools compared to the past. Consequently, this has led to increased losses for investors and managers of sports pools, with some facilities experiencing a significant drop in customers and increased churn rates.

Conclusion

In this context, practical suggestions such as encouraging monthly account recharges can be beneficial, as this often reflects customer satisfaction with pool services. Enhancing motivation for monthly account recharges by providing additional incentives and improving ticket rates in relation to the services offered could yield financial benefits. Another effective feature influencing customer behavior is the number of ticket purchases, particularly among customers who utilize pool services infrequently. Offering discount services, such as one free ticket for every five purchased, can attract more customers to the pool. Customer satisfaction is also a critical feature that all management systems strive to enhance. By employing specialized customer relationship management (CRM) teams, attractive solutions can be developed to boost customer satisfaction levels. Free training sessions, such as swimming techniques and diving lessons, can further increase customer motivation. Additionally, providing monthly services as part of the CRM system can enhance motivation and help introduce long-time users to new customers. Lastly, the quality of services is directly related to customer satisfaction; prioritizing service quality can help distinguish a pool from its competitors in the area. Thus, the

attention and consideration of these suggestions by managers and officials of recreational sports pools could be the key to the success of these venues. One of the main limitations in data mining-based research is the presence of incomplete or inconsistent data. In the case of sports pools in the city of Isfahan, it is possible that information related to customers, surveys, or their behavior may not be fully collected, which can affect the accuracy of predictive models. Therefore, to overcome the problem of incomplete data, a regular and systematic data collection system can be established, which includes periodic surveys of customers and recording their behaviors at different times. Additionally, data should be continuously updated to accurately and timely reflect changes.

Acknowledgments

The authors would like to appreciate the Isfahan (Khorasgan) Branch Islamic Azad University.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding

The authors declare that this research was done with the financial support of Isfahan (Khorasgan) Branch Islamic Azad University.

References

- Aldosary, M., & Alrashdan, A. (March 2021). Churn prediction for gym members using artificial neural networks assisted with the psychological concept of habit formation in the fitness industry. *Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management Singapore*. <http://www.ieomsociety.org/singapore2021/papers/720.pdf>
- Ali Beiginejad, F., & Haj Mohammadi, M. S. (2019). Improving the efficiency of predicting the amount of customer churn in social networks with Kavaneh data analysis (decision tree of neural networks and regression). *The First National Conference on Sustainable Development in Electrical and Computer Engineering, Isfahan*. <https://civilica.com/doc/1231685/> [In Persian].
- Amajuoyi, C. P., Nwobodo, L. K., & Adegbola, A. E. (2024). Utilizing predictive analytics to boost customer loyalty and drive business expansion. *GSC Advanced Research and Reviews*, 19(3), 191–202. <https://doi.org/10.30574/gscarr.2024.19.3.0210>
- Bamdad, Sh., & Hatami, R. (2017). Performance comparison of neural network techniques for predicting customer churn. *4th International Conference on Industrial and Systems Engineering, Mashhad*. <https://civilica.com/doc/811021> [In Persian].
- Bi, Q. (2019). Cultivating loyal customers through online customer communities: A psychological contract perspective. *Journal of Business Research*, 103, 34-44. <https://doi.org/10.1016/j.jbusres.2019.06.005>
- Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(01), 20-28. <https://doi.org/10.38094/jastt20165>
- Cheng, L. C., Wu, C. C., & Chen, C. Y. (2019). Behavior analysis of customer churn for a customer relationship system: an empirical case study. *Journal of Global Information Management (JGIM)*, 27(1), 111-127. <https://B2n.ir/w51041>
- Cheng, L., & Lin, C. (2015). Predicting customer churn in the sports lottery industry: A data mining approach. *Information Technology and Management*, 16(3), 201-209.
- De Caigny, A., Coussemont, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, 269(2), 760-772. <https://doi.org/10.1016/j.ejor.2018.02.009>
- Farhang far, A., Kurgan, L., & Dy, J. (2008). Impact of imputation of missing values on classification error for discrete data. *Pattern Recognition*, 41(12), 3692-3705. <https://doi.org/10.1016/j.patcog.2008.05.019>

- Haghighatnia, S., Abdolvand, N., & Rajae Harandi, S. (2018). Evaluating discounts as a dimension of customer behavior analysis. *Journal of Marketing Communications*, 24(4), 321-336. <https://doi.org/10.1080/13527266.2017.1410210>
- Kamel, S. R., Yaghoubzadeh, R., & Kheirabadi, M. (2019). Improving the performance of support-vector machine by selecting the best features by Gray Wolf algorithm to increase the accuracy of diagnosis of breast cancer. *Journal of Big Data*, 6, 1-15. <https://doi.org/10.1186/s40537-019-0247-7>
- Matuszelański, K., & Kopczewska, K. (2022). Customer Churn in Retail E-Commerce Business: Spatial and Machine Learning Approach. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(1), 165-198. <https://doi.org/10.3390/jtaer17010009>
- Mohammadi Javadi, P., & Noorbakhsh, S. K. (2018). Investigation of relational marketing implementation tactics and its effect on customer loyalty of airlines. *The Second International Conference on Modern Researches in Management, Economy and Development*. [In Persian].
- Mohammadpoori, M., Tojari, F., Esmaeli, M. R., & Nasr, D. (2014). Investigating differences between functions of brand association among consumers of sport bicycles according to demographic features. *Indian Journal of Fundamental and Applied Life Sciences*, 5(2), 2449-2459. <https://B2n.ir/r93946>
- Moradzadeh, Z., & Khodayari, B. (2017). Presentation of a prediction model of customer reversion to electronic banking services using decision tree and data mining method. *The 4th International Conference on Management, Entrepreneurship and Economic Development, Takestan, Takestan Institute of Higher Education*. <https://civilica.com/doc/836594/> [In Persian].
- Nasr Esfahani, D., Shirani, M., & Dalvi Esfahani, M. (2024). Investigating The Impact of Pricing of Sport Complex on Consumer's Choice in Isfahan. *Management and Entrepreneurship in Sport*, 3(1). <https://doi.org/10.48301/jmes.2024.462891.1064> [In Persian].
- Patel, H. H., & Prajapati, P. (2018). Study and analysis of decision tree-based classification algorithms. *International Journal of Computer Sciences and Engineering*, 6(10), 74-78. <https://B2n.ir/h11343>
- Pondel, M., Wuczyński, M., Gryniewicz, W., Łysik, L., Hernes, M., Rot, A., & Kozina, A. (2021, July). Deep learning for customer churn prediction in e-commerce decision support. *In Business Information Systems* (pp. 3-12). <https://doi.org/10.52825/bis.v1i.42>
- Raeisi, S., & Sajedi, H. (2020, October). E-Commerce Customer Churn Prediction by Gradient Boosted Trees. *In 2020 10th International Conference on Computer and Knowledge Engineering (ICCKE)* (pp. 055-059). IEEE. <https://doi.org/10.1109/ICCKE50421.2020.9303661>
- Rostami Fard, S., & Sajoudi Shijani, O. (2021). Presenting a model for predicting customer churn in online business using aggregated Bagging classifier based on neural networks. *4th International Conference on Electrical, Computer and Mechanical Engineering, Tehran*. <https://civilica.com/doc/1264704/> [In Persian].
- Suhanda, Y., Nurlaela, L. S., Kurniati, I., Dharmalau, A., & Rosita, I. (2022). Predictive Analysis of Customer Retention Using the Random Forest Algorithm. *TIERS Information Technology Journal*, 3(1), 35-47. <https://doi.org/10.38043/tiers.v3i1.3616>
- Xiahou, X., & Harada, Y. (2022). B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(2), 458-475. <https://doi.org/10.3390/jtaer17020024>