Extracting Actionable Knowledge to Increase Business Utility in Sport Services

Paulo Pinheiro[1] and Luís Cavique[2(✉)]

[1] Universidade Aberta and Cedis, Lisbon, Portugal
`ppinheiro@cedis.pt`
[2] MAS-BioISI, FCUL and Universidade Aberta, Lisbon, Portugal
`luis.cavique@uab.pt`

Abstract. The increase in retention of customer in gyms and health clubs is nowadays a challenge that requires concrete and personalized actions. Traditional data mining studies focused essentially on predictive analytics, neglecting the business domain. This work presents an actionable knowledge discovery system which uses the following pipeline (data collection, predictive model, loyalty actions). In the first step, it extracts and transforms existing real data from databases of the sports facilities. In a second step, predictive models are applied to identify user profiles more susceptible to dropout. Actionable rules are generated based on actionable attributes that should be avoided, in order to increase retention. Finally, in the third step, based on the previous actionable knowledge, experimental planning is carried out, with test and control groups, in order to find the best loyalty actions for customer retention. This document presents a simulation and the measure of the business utility of an actions sequence to avoid dropout.

Keywords: Sport services · Loyalty · Actionable knowledge · Business utility

1 Introduction

The promotion of physical activity as a means to prevent increasing rates of obesity and maintenance of well-being has provoked a proliferation of gyms and health clubs that compete with public sports facilities, and consequently there has been an increasing pressure on providers to maintain competitive advantage through services that provide higher levels of customer satisfaction (Howat and Assaker 2016). In a context of a great supply, the sports services sector is characterized by a high dropout rate (Avourdiadou and Theodorakis 2014), being the largest dropout in health clubs and gyms where the most promoted activity is fitness, than in sports facilities that have other facilities, namely swimming pools (Frota 2011).

Some gyms promote activities in which users subscribe continuously, becoming withdrawn when they fail to pay their monthly subscription or by the customer initiative. Other facilities work during sports seasons, and they need to encourage their users at the end of each season to renew their inscriptions for the following season. In both cases, with monthly subscription or seasonal subscription, the increase in retention

is nowadays a challenge that requires concrete and personalized actions, in addition to generic actions to improve the quality of services and facilities.

The widespread use of ERP and access control systems by these sports facilities nowadays enables a collection of quality data that allow us to find and measure user preferences and behavior from admission to abandonment. We intend to use existing data from those systems to get actionable knowledge that allows the identification of users of regular sports services at risk of moving out and those who keep being loyal to the service. The use of machine learning techniques, namely with decision trees, are able to generate rules that allow to predict the change of behavior of the dropout group.

In this context, the concept of actionable knowledge can be successfully applied, since the goal is not focused only on the predictive algorithms but in solving business problems (Cao 2010).

We reuse some concepts of Database Marketing (Cavique 2006) and IDIC model, <Identify, Differentiate, Interact, Customize>, (Peppers and Rogers 2004) to obtain Actionable Knowledge. The proposed model, which includes the data preparation, the profile discovery using predictive models and the loyalty actions with evaluation, can be presented in the following data pipeline:

$$\text{Data Collection} \rightarrow \text{Predictive Model} \rightarrow \text{Loyalty Actions}$$

This work extends previous works where this pipeline was presented (Pinheiro and Cavique 2018; Pinheiro and Cavique 2019). These works fill a gap in the study of retention through the use of machine learning techniques in regular sports services. The novelty of this work includes not only the creation of knowledge rules, but also the simulation and the measure of the business utility of a sequence of actions to avoid dropout.

This document has the following structure. In Sect. 2, the related work in sports service retention and in actionable knowledge is presented. In Sect. 3, we present the proposed method in three steps: the data collection, the predictive model and finally the application of loyalty actions. In Sect. 4, the loyalty actions process is detailed regarding the business utility measure. Finally, in Sect. 5 we draw the conclusions of this work.

2   Related Work

2.1   Sports Service Retention

Frota (Frota 2011) says that 25% of withdrawals are motivated by club-related issues, of which 45% are recoverable; 22% are related to money, of which 31% are recoverable; 29% are related to situational problems, of which 44% are recoverable; and finally 24% leave for personal reasons, and the recovery of these is extremely difficult. Doing the calculation we can conclude that about 30% of the dropouts are recoverable. These recoveries can be made through the implementation of loyalty actions that are preferably carried out before the withdrawal takes place. The solution to the problem is

to answer the following questions: who will be the target of these actions and what actions should be developed to avoid the dropout.

Gorgoglione (Gorgoglione 2011) identifies five possible approaches for the creation of personalized actions: the computational approach, the similarity approach, the bottom-up approach, the top-down approach and the personalized approach. The computational approach generates actions based on the profiles of clients, with no human intervention. The similarity based approach is used by recommendation systems and web content personalization methods; this type of approach assumes that actions are related to customer preferences, can be inferred through customer profiles, and it is assumed that similar customers behave similarly and similar actions cause similar reactions. The bottom-up approach includes the methods of knowledge discovery and is implemented in two separate steps: by creating customer profiles and deciding what actions are appropriate. The top-down approach consists in the same two steps of the bottom-up approach, but the decision on which actions to implement is taken before defining customer profiles. Finally, the personalized approach offers customers a number of different options, being at their discretion to choose which they prefer.

Once one tries to define profiles of behaviors that lead to abandonment, it is necessary to find characteristics or attributes that somehow allow to trace those profiles. Work related to retention in sports services (Avourdiadou and Theodorakis 2014; Howat and Assaker 2016; Surujlal and Dhurup 2012; Gonçalves 2012; Frota 2011) allows systematizing and identifying attributes necessary to characterize users and their behavior, both those who continue to use the services and those who leave.

In CRM applications, a decision tree can be built from a set of examples (customers) described by a rich set of attributes including customer personal information (such as name, gender, birthday), financial information (such as yearly income), family information (such as life style, number of children), and so on. Because decision trees can be converted to rules for explicit representation of the classification, one can easily obtain characteristics of customers belonging to a certain class (such as loyal customer or churner) (Yang et al. 2007).

2.2 Actionable Knowledge

Traditional data mining studies concentrated primarily on predictive mining, where the cause and effect scenario is described. But this information alone is not sufficient as it does not provide much benefit to the final user. What becomes more interesting and critical to organizations is to mine patterns in order to create knowledge actionable (Cao 2010).

As Yang et al. (2007) says a common problem in current applications of data mining, particularly in intelligent CRM, is that people tend to focus on, and be satisfied with, building up the models and interpreting them, but not to use them to get profit explicitly. More specifically, most data mining algorithms (predictive or supervised learning algorithms) only aim at the construction of customer profiles, which predicts the characteristics of customers of certain classes. This knowledge is useful but it does

not directly benefit the enterprise. To improve customer relationship, the enterprise must know what actions to take to change customers from an undesired status (such as churner) to a desired one (such as loyal customers).

Knowledge is considered actionable if users can take direct actions based on such knowledge to their advantage. Actionability should be a criterion which can measure the utility of the mined patterns. Among the most important and distinctive actionable knowledge are actionable behavioral rules that can directly and explicitly suggest specific actions to take to influence (retain and encourage) the behavior of customers (Su et al. 2014).

To face the increasingly complex challenges of data mining in real-life world problems, Cao (Cao et al. 2007; Cao 2010) presents a new approach, which opposes data-driven to domain-driven. Data-driven corresponds to the traditional data mining, while domain-driven is related to the business domain, or business area. The domain-driven data mining (D^3M) close the gap between researchers and practitioners, by generating actionable knowledge for real user needs.

The D^3M evaluates a pattern (p) using the utility measure U(p) from both technical and business perspectives. U(p) is measured in terms of technical significance (technicalU(p)) and business utility (businessU(p)), i.e., U(p) = f(technicalU(p), businessU(p)).

An example of technical utility can be given by: accuracy = 87.90% and precision = 91.54%. While an example of business utility is, for instance, the average frequency in days and the average customer value in euros.

Table 1 presents eight different aspects of data-driven and domain-driven extracted from (Cao et al. 2007; Cao 2010). In domain-driven data mining the object mined is not only the data but the business domain, where the goal is to develop effective problem-solving and discover actionable knowledge to satisfy real users, by using real data and information related. D^3M is a multiple-step, iterative and interactive process, where the human cooperates, in a customizable environment, to provide actionable knowledge, which are evaluated in a trade-off between technical significance and business utility.

Table 1. Data-driven versus domain-driven

Aspects	Data-driven	Domain-driven
Object	Data tells the story	Data and business domain tell the story
Objective	Effective algorithms, discover knowledge of research interest	Effective problem-solving, discover actionable knowledge to satisfy real users
Data	Abstract, synthetic data	Real-life data and information related
Process	One step	Multiple-step, iterative and interactive
Mechanism	Automated	Human mining cooperation
Usability	Predefined models and process	Customizable models and process
Deliverable	Patterns	Actionable knowledge
Evaluation	Technical metrics	Trade-off between technical significance and business utility

3   Proposed Method

To better explain the developed work, we introduce the concept in three steps: the data collection, where the data is prepared and the attributes selected, the selection of dropout and loyalty rules with decision tree algorithm, and the creation of loyalty actions to prevent dropout.

3.1   Data Collection

Following the first step of the pipeline, we need to create a data warehouse to support the predictive model. This data warehouse should present a fact table with a set of suitable attributes. The selection of attributes must take into account the existence of factors with a greater or lesser impact on the retention of sports facilities and the possibility of being able to be extracted from the data registered in the databases of the ERP systems used by the sports facilities.

In order to develop the experiment and test the results, we used data from a Lisbon sports facility to which we applied the Extraction, Transformation and Loading (ETL) processes as explained in previous works (Pinheiro and Cavique 2018; Pinheiro and Cavique 2019). Since the performance of Machine Learning techniques is obtained through the manipulation of values (Gama et al. 2017), in addition to the attributes directly mapped from the source database, some attributes have been transformed, discretized through numeric-symbolic conversions, or created new attributes that derive from transformations made on the original data.

As so the ETL process resulted in the construction of a fact table in the data warehouse with 51 relevant attributes, although only 45 have valid data. These 51 attributes fit into the four groups:

(1) Demographic: e.g. age, gender;
(2) Frequency: e.g. number of days without attendance; frequency of classes;
(3) Service agreement: e.g. number of months of enrollment, number of renewals;
(4) Service quality: e.g. indications of satisfaction, net promoter score;

Relevant attributes, such as those related to the quality of the service were not filled due to lack of data.

One of the attributes in the fact table, the attribute 'withdrawal' is what characterizes the state of the user to date and therefore the predicted attribute (the target attribute). For practical reasons it was decided to define the attribute as a binary value, corresponding to a value of 1 for a quitter user, and a value of 0 to an active member.

After executing all ETL processes the fact table was populated with 8381 users as shown in Table 2.

Table 2.   Number of users in data warehouse fact table

Users	In aquatic activities	In fitness activities	Total
Active	1226	803	1927
Dropouts	1697	4926	6454
Total	2923	5729	8381

3.2   Predictive Model

Since we intend to formulate a model or hypothesis capable of relating the values of the attributes in the fact table with the value of the target attribute, which is a nominal attribute, in practice we intend to construct a predictive model that is able to identify sets of characteristics that allow to rank a user against the level of their pre-dropout status.

To obtain the decision trees we use Microsoft Decision Trees algorithm available in Microsoft SQL Server Analysis Services Designer ver. 13.0.1701.8. This version uses a hybrid algorithm that incorporates different methods to create a tree. The algorithm offers three formulas for scoring information gain: Shannon's entropy, Bayesian network with K2 prior, and Bayesian network with a uniform Dirichlet distribution of priors (Microsoft 2017).

Since some attributes in the fact table result from different forms of classification or discretization of the same characteristic, the proposed models eventually use redundant attributes. Gama (Gama et al. 2017) states that since the process of constructing a tree selects the attributes to use, they result in models that tend to be quite robust in relation to the addition of irrelevant and redundant attributes. However, it is desirable to obtain models with significant predictive capacity and at the same time to obtain actionable profiles, so redundant attributes should be avoided.

After removing the redundant attributes, several adjustments were made to avoid overfitting. The resulting models correspond to shallow trees with leaves that always have a number of examples greater than 50. The evaluation metrics obtained with the Holdout method considering 70% of data for training and 30% for testing of the model are presented in Table 3.

Table 3. Evaluation metrics of Predictive Model with Holdout Method

#Nodes	Depth	Accuracy	Sensitivity	Specificity	Precision	F-Score
30	6	87.90%	92.69%	72.53%	91.54%	92.11%

In order to create loyalty actions specific attributes from the decision tree should be chosen to generate actionable knowledge. Some attributes cannot influence or be changed, such as the attribute 'age' or 'gender', denominated by 'non-actionable attributes'. On the other hand, customer retention strategies can change the content of some attributes that reflect user behavior. These attributes, that allows operational changes, are called 'actionable attributes'. An example of an actionable attribute is the number of 'days without frequency', since a strategy can be implemented that causes, at least to some users, to return to the sports facility, after some time without attending. Table 4 presents the actionable attributes selected by the decision tree whose metrics were presented in Table 3.

Table 4. Actionable attributes

Attribute	Description
X_1	Days without frequency
X_2	Aquatic activity attendance
X_3	Number of months of enrollment
X_4	Number of renewals
X_5	Weekly contracted frequency
X_6	Fitness activity attendance

Each branch of the decision tree forms a splitting rule, where each node includes an attribute used by the algorithm. In each leaf of the tree there are examples that correspond to users quitting and examples that correspond to active users. The relationship between these quantities defines the probability of withdrawal that we refer as dropout in Table 5.

Table 5. Actionable knowledge based on actionable rules

Actionable Rules	

Actionable rule A:
If
  X_1 (days without frequency) $\in$]30, 60] and
  X_4 (number of renewals) = 0
Then
  dropout = 99.29%
	Actionable rule B:
	If
	  X_1 (days without frequency) $\in$]60, ∞[and
	  X_3 (number of months of enrollment) $\notin$]12, ∞[
	Then
	  dropout = 98.15%

Actionable rule C:
If
  X_1 (days without frequency) $\in$]15, 30] and
  X_2 (aquatic activity attendance) = False
Then
  dropout = 96.47%
	Actionable rule D:
	If
	  X_1 (days without frequency) $\in$ [0, 7] and
	  X_2 (aquatic activity attendance) = False and
	  X_3 (number of months of enrollment) > 2 and
	  X_4 (number of renewals) = 0
	Then
	  dropout = 93.03%

Actionable rule E:
If
  X_1 (days without frequency) $\in$]30, 60] and
  X_4 (number of renewals) > 0
Then
  dropout = 92.84%	

Thus, it is possible to draw dropout profiles from the splitting rules on each leaf that shows a dropping rate above a considered threshold. These profiles allow us to segment users according to the criteria mentioned by Kotler (Kotler and Keller 2009) which indicates that segmentation is only useful if the segments meet five criteria: they are measurable, substantial, accessible, differentiable and actionable.

Actionable knowledge is supported by splitting the tree in actionable rules which include the actionable attributes. Table 5 presents several actionable rules above the

90% threshold that should be taken into account. Each actionable rule contains actionable attributes with values that should be avoided.

Actionable rule A shows a user who does not visit the facilities between 31 to 60 days and never renewed. Actionable rule B shows a user profile who does not visit the facilities for more than 60 days and whose enrolment is inferior to 12 months. Actionable rule C shows a user who does not visit the facilities between 16 to 30 days and does not attend aquatic activities. Actionable rule D shows a user who does not visit the facilities in the last 7 days, does not attend aquatic activities, whose enrolment has more than two months and has never renewed subscription. Finally, actionable rule E shows a user who does not visit the facilities between 31 to 60 days and has already renewed at least once.

3.3   Loyalty Actions with Evaluation

The retention strategy includes a communication process in three stages, starting with a personalized e-mail send to the user, where the cost is practically zero, followed by the use of a SMS and finally a personal contact, creating a funnel workflow. It is expected that after sending each communication the behavior of part of the target users will change, so for this subset it will no longer be necessary to send them the following communications.

After defining the retention strategy, it is necessary to evaluate its effectiveness. The experiments are constructed through the implementation of A/B tests and evaluated through the chi-square method, which will allow computing a statistical conclusion. A/B tests manipulate a causal variable in order to determine the impact of the variable in two different groups of individuals, the test group and the control group. The groups are created by splitting randomly the users into two groups with same number of elements. In the test group the loyalty actions are applied, and in the control group the loyalty actions will not be applied.

The users who received the communication and return to attend the facilities are called 'recovered'. The recovered users allow evaluating the loyalty campaign impact in euros. This last estimation is used to measure the business utility we were looking for.

3.4   Discussion of the Proposed Method

The proposed framework presented in (Cavique 2006; Pinheiro and Cavique 2018; Pinheiro and Cavique 2019) works in a multiple-step pipeline, shown in Table 6, where the columns phase, outcome, evaluation and groups are presented. The column 'groups' is presented in order to clarify and differentiate groups that have similar names but different meanings.

In the data collection phase the outcome is a data warehouse that will be used in the following steps. In the predictive model phase actionable rules are generated, the evaluation is given by technical metrics as accuracy and the performance of the model is measured using training and test groups. Finally, in the loyalty actions phase the business utility is the outcome, measured in euros and the significance of the actions is measured using test and control group in an A/B test.

Table 6. Pipeline phases, outcomes, evaluation and groups

Pipeline phase	Outcome	Evaluation	Groups
1. Data collection	Data warehouse	–	–
2. Predictive model	Actionable knowledge	Accuracy	Training and test
3. Loyalty actions	Business utility	Euros	Test and control

This method corresponds to the domain-driven approach presented by Cao (Cao et al. 2007; Cao 2010), where the predictive model delivers business-friendly decision support actions reified in actionable knowledge rules and the evaluation includes the trade-off between technical significance and business utility.

4   Detailed View of the Loyalty Actions

In this section we propose to detail and evaluate the effectiveness of the proposed model according to the following hypotheses:

H_0: After executing the loyalty actions, the number of dropouts is the same as if no loyalty actions had been taken;
H_1: After performing the loyalty actions, the number of dropouts is lower than if the loyalty actions had not been carried out;

For this purpose, an interface was developed in Visual Studio to select the users according to the rules of the model in which they fit and apply changes in their attributes and status, in order to simulate reactions to loyalty actions.

With this tool, we use the users defined by Rule A and work with this subset of users through the entire simulation of the campaign. This subset represents a possible real case in which a leaf of the tree may contain some users who, although yet active, presents the profile of dropout, and should therefore be the target of the loyalty actions proposed for this profile. The subset is divided into a test and control group, where the first group has 594 users and the second 593, and both contain at the beginning 3 users who have not given up.

The simulation of the communication strategy occurs in different instants: t_1, t_2 and t_3, using e-mail, SMS and personal contacts respectively. We expect a recovery of 5% of the users of the test group after the first action, a recovery of 10% after the second action and 15% after the third action, obtaining a final recovery of approximately 30%, as referred by Frota (Frota 2011). Table 7 illustrates the experiment, where it was avoided the dropout of 177 users.

Table 7. Loyalty actions simulation with the users of rule A

Time	Loyalty action	Control group		Test group		
		# potential dropout	# of dropouts	# potential dropout	# of avoided dropouts	# of dropouts
t_1	Email	593	590	594	35	559
t_2	SMS	593	590	559	57	502
t_3	Personal contact	593	590	502	85	417
Total					177	

To evaluate the significance of A/B test a matrix with observed (O) and expected (E) users is defined, where t indicates users in test group, c defines users in control group, n defines non-dropout (or retained) users and d defines dropout users, as shown in Table 8.

Table 8. Observed and expected users

	Observed		Total	Expected	
	$O_t = 594$	$O_c = 593$	$T = 1187$	$E_t = 594$	$E_c = 593$
Retained	$On_t = 177$	$On_c = 3$	$N = 180$	$En_t = 90.29$	$En_c = 90.14$
Dropout	$Od_t = 417$	$Od_c = 590$	$D = 1007$	$Ed_t = 503.71$	$Ed_c = 502.86$

With the result obtained for chi-square ($\chi^2 = 197.539$) we can conclude, through the query of the distribution table, with one degree of freedom and a confidence level of 99%, that the loyalty actions carried out with the proposed campaign had the effect wanted.

In addition to the significance obtained through the A/B test, we also calculate the business utility expressed in euros. To obtain this value, we consider the average monthly cost for a regular inscription (38.07€) indicated by AGAP (AGAP 2017). For the number of users recovered from the experience, the business utility is 6,624.18€ for each month in which the recovered users maintained their registration.

In this document the result of the simulation goes beyond the significance of A/B test and the value of the business utility, which could be easily calculate given the recovery rate of 30% (Frota 2011).

First of all, it is important to understand that 'triggered' users share behavioral similarities defined by sets of limited values in the actionable attributes given by the rules obtained from the predictive model, which allows directing perfectly personalized actions.

Secondly, actions were directed only to users of one of the rules with a threshold above 90% of withdrawals. Applying another actions to the other profiles defined by the other rules and according to their actionable attributes the business utility will be higher.

Finally, the simulated reaction to the actions carried out on the users in rule A was to consider that the users returned to attend the sports facilities. Thus, in the 174 users retrieved, where before the attribute $X_1 \in]30, 60]$, was changed to $X_1 \in [00, 07]$. By recycling the predictive model after this simulation a new set of actionable rules could be generated as presented in Table 9.

Table 9. Actionable knowledge in the second loyalty campaign

Actionable Rules	

Actionable rule A:
If
    X_1 (days without frequency) $\in]30, 60]$ and
    X_4 (number of renewals) $= 0$
Then
    dropout $= 99.71\%$
	Actionable rule B:
	If
	    X_1 (days without frequency) $\in]60, \infty[$ and
	    X_3 (number of months of enrollment) $\notin]12, \infty[$
	Then
	    dropout $= 98.42\%$

Actionable rule C:
If
    X_1 (days without frequency) $\in]30, 60]$ and
    X_2 (aquatic activity attendance) $=$ False and
    X_4 (number of renewals) > 0
Then
    dropout $= 97.18\%$
	Actionable rule D:
	If
	    X_1 (days without frequency) $\in]15, 30]$ and
	    X_6 (fitness activity attendance) $=$ False
	Then
	    dropout $= 96.34\%$

The observation of the Table 9 allows us to verify that the rules with a threshold of more than 90% have been reduced to 4, with two remaining, and two emerging rules that have a threshold of withdrawal above 90%. The first of the rules that remained, rule A, is the rule on which the loyalty actions were applied, and now presents a higher level of dropouts and a lower number of users, since the recovered ones have come to be considered by other rules.

With the recovery of users by changing the attribute X_1, changes would be expected at the level of the whole predictive model, which is why new rules with withdrawal limits higher than 90% appear, and new attributes, like X_6, 'fitness activities attendance', are considered. That is why the recycle of the predictive model should only take place after complete the cycle of the loyalty actions to avoid concept drift during the campaign. However, it is desirable that the predictive model be recycled after completing the campaign so that it adapts to the new data, discovering new actionable attributes and generating new rules if appropriate.

5   Conclusions

In a context of great supply in health clubs and gyms, the sports services sector is characterized by a high dropout rate. The increase in retention of customer is nowadays a challenge that requires concrete and personalized actions. Most of data mining techniques achieve interesting patterns for churning. However, the implementation in real scenarios is still in development.

We reuse the data pipeline presented in (Cavique 2006; Pinheiro and Cavique 2018; Pinheiro and Cavique 2019) with three steps: data collection, predictive model and loyalty actions. This data pipeline corresponds to the domain-driven approach presented by Cao (Cao et al. 2007; Cao 2010), where the predictive model delivers actionable knowledge and the loyalty actions return a business utility measure.

In the data collection phase, real data from databases of the sports facilities is used, containing information of more than 8,000 users with 51 attributes.

In the predictive model phase, actionable attributes are identified and actionable knowledge rules, extracted with the decision trees, with a dropout threshold above the 90%, are detailed.

In order to follow the simulation of the loyalty actions only one subset of users is created using the information of Actionable Rule A. The same subset was divided into test and control group, and A/B test was performed with success. This experiment allowed us to achieve a value for the business utility and to generate a set of actionable rules for the second loyalty campaign.

In this work data-driven approach was overcome and we believe domain-driven approach was reached with real actionable attributes, real actionable rules and a concrete business utility measure. It is not enough to have technical interesting and significant metrics. The real world requires a business utility, which can be easily justified and applied.

References

AGAP. Associação de empresas de ginásios e academias de Portugal, Barómetro 2017 (2017)

Avourdiadou, S., Theodorakis, N.D.: The development of loyalty among novice and experienced customers of sport and fitness centres. Sport Manag. Rev. **17**(4), 419–431 (2014). https://doi.org/10.1016/j.smr.2014.02.001

Cao, L., Zhang, C.: The evolution of KDD: towards domain-driven data mining. Int. J. Pattern Recognit. Artif. Intell. **21**, 677–692 (2007)

Cao, L.: Domain-driven data mining: challenges and prospects. IEEE Trans. Knowl. Data Eng. **22**(6), 755–769 (2010). https://doi.org/10.1109/TKDE.2010.32

Cavique, L.: Relatório da Unidade Curricular de Database Marketing, 2005–2006. Escola Superior de Comunicação Social. Instituto Politécnico de Lisboa (2006). Unpublished

Frota, M.: Gestão da Retenção. In: Manual de Gestão de Ginásios e Health Clubs - Excelência no sector do Health & Fitness, pp. 103–148 (2011)

Gama, J., de Leon Carvalho, A.P., Faceli, K., Lorean, A.C., Oliveira, M.: Extração de Conhecimento de Dados (3a edição; E. Silabo, Ed.) (2017)

Gonçalves, C.: Variáveis Internas e Externas ao Indivíduo que influenciam o Comportamento de Retenção de Sócios no Fitness. PODIUM Sport Leisure Tourism Rev. **1**(2), 28–58 (2012)

Gorgoglione, M., Panniello, U.: Beyond customer churn: generating personalized actions to retain customers in a retail bank by a recommender system approach. J. Intell. Learn. Syst. Appl. **03**(02), 90–102 (2011). https://doi.org/10.4236/jilsa.2011.32011

Howat, G., Assaker, G.: Outcome quality in participant sport and recreation service quality models: empirical results from public aquatic centres in Australia. Sport Manag. Rev. **19**(5), 520–535 (2016). https://doi.org/10.1016/j.smr.2016.04.002

Kotler, P., Keller, K.L.: Marketing management. In: Organization, 14th Editi, vol. 22 (2009). https://doi.org/10.1080/08911760903022556

Microsoft: Microsoft Decision Trees Algorithm Technical Reference (2017) https://docs.microsoft.com/en-us/sql/analysis-services/data-mining/microsoft-decision-trees-algorithm-technical-reference?view=sql-server-2014. Accessed 18 April 2019

Peppers, D., Rogers, M.: Managing Customer Relationships: A Strategic Framework. Wiley, Hoboken (2004). ISBN 047148590X

Pinheiro, P., Cavique, L.: Models for increasing retention in regular sports services: predictive analysis and loyalty actions (2018). https://doi.org/10.23919/CISTI.2018.8399160

Pinheiro, P., Cavique, L.: An actionable knowledge discovery system in regular sports services. In: Rocha, Á., Adeli, H., Reis, L., Costanzo, S. (eds.) New Knowledge in Information Systems and Technologies, vol. 931, pp. 461–471. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-16184-2_44

Su, P., Zhu, D., Zeng, D.: A new approach for resolving conflicts in actionable behavioral rules. Sci. World J. **2014**, 9 (2014). https://doi.org/10.1155/2014/530483. Article id 530483

Surujlal, J., Dhurup, M.: Establishing and maintaining customer relationships in commercial health and fitness centres in South Africa. Int. J. Trade Econ. Finance **3**(1), 14–18 (2012). https://doi.org/10.7763/IJTEF.2012.V3.165

Yang, Q., Yin, J., Ling, C., Pan, R.: Extracting actionable knowledge from decision trees. IEEE Trans. Knowl. Data Eng. **19**(1), 43–55 (2007). https://doi.org/10.1109/TKDE.2007.250584