

Modelos para incremento da retenção em serviços desportivos regulares

Análise preditiva e ações de fidelização

Models for increasing retention in regular sports services

Predictive analysis and loyalty actions

Paulo Pinheiro

Mestrando - Universidade Aberta
Lisboa, Portugal
ppinheiro@cedis.pt

Luís Cavique

Universidade Aberta
Lisboa, Portugal
Luis.Cavique@uab.pt

Resumo —As instalações desportivas que oferecem serviços desportivos regulares têm vindo a adotar sistemas ERP e CRM, existindo atualmente bases de dados com dados históricos de grande valia. Neste trabalho demonstramos que aplicando modelos preditivos a estes dados é possível identificar perfis de abandono. Com base nos perfis encontrados é realizado um planeamento de experiências, com grupos de teste e controlo, com vista a encontrar ações concretas de fidelização.

Palavras Chave - serviços desportivos; análise preditiva; planeamento de experiências; fidelização;

Abstract —*The sports facilities that offer regular sports services have been adopting ERP and CRM systems and there are now databases with historical data of great value. In this work, we demonstrate that by applying predictive models to these data it is possible to identify abandonment profiles. Based on the profiles found, experience planning is carried out, with test and control groups, in order to find concrete actions of loyalty.*

Keywords - sport services; predictive analysis; experience planning; loyalty.

I. INTRODUÇÃO

A promoção do exercício físico como um meio para prevenir os crescentes rácios de obesidade e manutenção do bem-estar tem provocado uma proliferação de ginásios e *health-clubs* que competem com instalações desportivas públicas, e consequentemente tem havido uma crescente pressão nos prestadores deste tipo de serviço para manterem vantagens competitivas através de serviços que providenciam maiores níveis de satisfação do cliente [1]. Num contexto de muita oferta, o setor da prestação de serviços desportivos é caracterizado por uma elevada taxa de desistência [2], sendo que em Portugal, e de acordo com o Barómetro do Fitness da Associação de Ginásios e Academias de Portugal [3] a taxa de cancelamentos geral em 2016 foi de 69% correspondendo a uma taxa de retenção geral de 31%.

Em *Database Marketing* [4] – modelo inspirado no IDIC (*Identify, Differentiate, Interact, Customize*) [5] – procura-se

partir dos dados para identificar, diferenciar e interagir através de ações concretas. Com este trabalho, seguindo os passos do *Database Marketing*, pretende-se utilizar os dados existentes para identificar os utentes de serviços desportivos regulares em risco de abandono, diferenciando-os dos restantes através de técnicas de *data mining* / *predictive analytics* / *machine learning* que encontrem perfis de utilização e comportamento, e que permitam interagir de forma personalizada e com ações de fidelização concretas com os utentes que apresentam esses perfis para evitar o seu abandono. O modelo proposto pode ser apresentado da seguinte forma esquemática:

dados -> modelos -> fidelização.

Este documento tem a seguinte estrutura. Na Secção II é feita uma abordagem breve a trabalhos relacionados que têm sido efetuados na área da retenção dos serviços desportivos e na aplicação da análise preditiva na retenção. Na Secção III apresentamos a metodologia de preparação dos dados, na Secção IV introduzimos o modelo preditivo aplicado aos dados apresentados na secção anterior e medimos os resultados obtidos. Na Secção V abordamos o planeamento das experiências e as ações de fidelização que podem ser introduzidas, e apresentamos uma proposta que permite medir os resultados obtidos. Finalmente, na Secção VI apresentamos a contribuições do trabalho na Secção VII indicamos os próximos passos a realizar.

II. TRABALHOS RELACIONADOS

Os estudos [1], [2] realizados na área da retenção nos serviços desportivos, concluem, de forma geral que a retenção e lealdade dos utentes está relacionada com a qualidade das instalações, do pessoal e com a qualidade global dos serviços prestados, resultados que têm vindo a ser utilizados pela gestão das instalações desportivas. Este tipo de trabalho é sempre realizado com base em inquéritos efetuados sobre uma amostra da população, o que sugere a procura por outros métodos que permitam aferir ou chegar a conclusões de carácter menos

genérico e que permita acionar mecanismos direcionados a utentes em fase de pré-abandono.

Para além de um estudo de Determinação de Padrões de Desistências em Ginásios [6] aplicado apenas aos serviços de *fitness*, não se constatou existirem outros estudos na área de *data mining* aplicados diretamente a dados obtidos em serviços desportivos. No entanto o problema das elevadas taxas de desistência / baixas taxas de retenção noutros tipos de serviço tem levado a realização deste tipo de estudos (*churn prediction*), nomeadamente na área das telecomunicações onde também se verificam elevadas taxas de abandono. Contudo, dada a grande dimensão das bases de dados e dos custos envolvidos, a maioria dos estudos existentes nesta área usam pequenas amostras dos registos de clientes, o que pode resultar na baixa confiabilidade e validade dos resultados obtidos [7].

Por outro lado, as empresas procuram cada vez mais concentrar-se nos clientes em que terão mais hipóteses de satisfazer e/ou nos que serão mais rentáveis em vez de dispersarem os seus esforços trabalhando indiscriminadamente todos os clientes. Por isso, procuram identificar padrões de semelhanças e necessidades em grupos de clientes (processo de segmentação).

Alguns exemplos de métodos de segmentação são o RFM e o RM [8] baseados em propriedades como a Recentidade (quando ocorreu a última visita do cliente), a Frequência das compras e o valor Monetário global. Estes tipos de segmentação são simples de determinar e não necessitam de modelos preditivos auxiliares.

Como métodos de segmentação que recorrem a modelos preditivos Siegel [9] refere os modelos *Lift* e *Uplift*. O *Lift*, identifica os clientes mais suscetíveis a uma determinada comunicação ou ação de marketing. O *Uplift*, para aprender a distinguir os clientes influenciáveis – aqueles a quem faz diferença efetuar algum tratamento – o modelo aprende com os clientes que foram contactados e os que não foram, pelo que é necessário utilizar dois conjuntos de dados para treino do modelo, constituindo-se um grupo de clientes que “é tratado” – grupo de tratamento - e um outro grupo de clientes que não o é – grupo de controlo. O método *Uplift* utiliza também técnicas de *data mining* para efetuar a segmentação através de árvores *Uplift* que, da mesma forma que as árvores de decisão, utilizam atributos para automaticamente identificarem sub-grupos, mas de forma distinta destas, tentam identificar segmentos extremos pela diferença de tratamentos, identificando segmentos que são particularmente influenciáveis.

Uma vez que se procura definir comportamentos ou perfis de comportamento que levam ao abandono, torna-se necessário encontrar características que de alguma forma permitam traçar esses perfis. A leitura de diversos trabalhos relacionados com a retenção em serviços desportivos ([1] [2] [10] [11] [12] [13]) permitiu identificar e agrupar os atributos passíveis de recolha e registo em base de dados e que se poderão encontrar nos sistemas informáticos utilizados atualmente pelas instalações desportivas: (1) Atributos demográficos como a idade e o género; (2) Atributos relacionados com a contratação do serviço como a frequência contratada, o número de meses da inscrição e o volume de negócios (*lifetime value* – LTV); (3) Atributos relacionados com a frequência nomeadamente frequência real,

frequência média e número de dias sem visitar as instalações; (4) Outros atributos relacionados com a qualidade do serviço como reclamações ou outras manifestações de insatisfação, abordagens e contatos pessoais ou por outras formas, avaliações da condição física ou outro tipo de avaliação comunicada regularmente ao utente;

III. PREPARAÇÃO DOS DADOS

Tendo em atenção a necessidade de obter registos com os atributos do tipo indicado nos grupos referidos, considerou-se uma base de dados utilizada de uma instalação desportiva de Lisboa que utiliza uma aplicação de mercado (e@sport) à qual foram aplicados processos *Extract, Transform, Load* (ETL) conforme descritos por Trujillo [14], tendo sido considerado todo o historial de utentes que estiveram (ou ainda estão) inscritos nas atividades aquáticas ou de *fitness* entre 01/Junho/2014 até 31/Outubro/2017.

De acordo com o primeiro passo do *Database Marketing* pretende-se criar um *Data Warehouse* com uma tabela de factos onde se irão concentrar os atributos relevantes que irão suportar a aplicação do modelo preditivo.

Uma vez que o desempenho de algumas técnicas de DM estão limitadas à manipulação de valores de determinado tipo ou o próprio desempenho é influenciado pelo intervalo de variação de valores [15], para além dos atributos diretamente mapeados a partir da base de dados origem, alguns atributos foram transformados, discretizados através de conversões numérico-simbólicas, ou ainda criados novos atributos que derivam de classificações e transformações efetuadas nos dados ou nos atributos originais.

Do processo ETL resultou a construção de uma tabela de factos no *data warehouse* com 51 atributos relevantes, embora apenas quarenta e cinco apresentem dados válidos. Atributos de relevo, como os relativos à qualidade do serviço, não foram preenchidos atributos por inexistência de dados. Os atributos considerados são apresentados na Tab. 1:

TABELA 1- ATRIBUTOS CONSIDERADOS

Grupo	Atributos
(1) Demográficos	Idade (2 atributos), Género, Referências (2 atributos), Distância à instalação
(2) Contratação do Serviço	Nº de meses da inscrição (3 atributos), Volume de negócios (2 atributos), Utilização Livre, Atividades que frequenta (10 atributos), Número de atividades que frequenta, Frequência contratada (2 atributos), Número de renovações
(3) Frequência	Nº de Dias sem frequentar (3 atributos), Frequência Média (3 atributos), Número de frequências total (2 atributos), Número de aulas (2 atributos), Frequência média de aulas (2 atributos), Frequência real, Ratio (Frequência real / frequência contratada) (2 atributos), Duração do treino (2 atributos)
(4) Qualidade do serviço	Número de contatos efetuados, Indicações de insatisfação (3 atributos), Número de indicações de insatisfação, Última resposta NPS, Número de avaliações efetuadas, Número de prescrições

Na Tab. 1 constata-se que alguns atributos apresentam variações que correspondem a derivações que visam discretizar o valor do atributo original, tendo por vezes sido usado mais do que um método. O método de Hughes (classificação A – atributos cujo nome inicia por *classe*) foi utilizado nos atributos *número de meses da inscrição*, *volume de negócios*, *número de dias sem frequentar*, *frequência média*, *número de frequências total*, *número de aulas* e *frequência média de aulas*.

Foram também adicionados atributos variantes numa segunda classificação (classificação B – atributos cujo nome inicia por *classe2*) a partir dos atributos base de acordo com indicações existentes na literatura referenciada, como é o caso do atributo *dias sem frequência* – foram criadas as classes “[00-07]”, “[07-15]”, “[15-30]”, “[30-60]” e “[60-inf]”, do atributo *idade* – foram criadas as classes “<20”, “<35”, “49”, “<65” e “>=68”, do atributo *número de meses da inscrição* – foram criadas as classes “[00-01]”, “[01-02]”, “[02-04]”, “[04-06]”, “[06-09]”, “[09-12]” e “[12-inf]” e do atributo *frequência média* – foram criadas as classes “<=0.1”, “<=0.2”, “<=0.5”, “<=1”, “<=2”, “<=3”, “>3”, do atributo *ratio frequência real / frequência contratada* – foram criadas as classes “Nunca”, “<=25”, “<=50”, “<=75”, “>75” - e do atributo *duração do treino* – foram criadas as classes “0”, “-60”, “-90” e “+90”.

Para além das operações referidas sobre os atributos, foram ainda corrigidas situações de valores em falta através da estratégia de remoção dos registos respetivos.

Uma vez que a maior parte dos utentes da instalação desportiva praticam atividades aquáticas ou de *fitness*, agrupou-se os utentes de acordo com as atividades que praticaram ao longo da sua frequência obtendo-se o número de registos final após os processos ETL indicado na Tab. 2.

TABELA 2 - Nº DE REGISTOS NA TABELA DE FACTOS DO DATA WAREHOUSE (DADOS OBTIDOS EM 31/OUT/2017)

Nº Utentes	Em Atividades Aquáticas	Em Atividades de Fitness	Total
Ativos	1226	803	1927
Desistentes	1697	4926	6454
Total	2923	5729	8381

IV. O MODELO PREDITIVO

Para construir e validar o modelo preditivo utilizou-se o *Microsoft SQL Server Analysis Services Designer Ver. 13.0.1701.8*. Este produto disponibiliza um algoritmo de classificação, o *Microsoft Decision Trees*, baseado em árvores de decisão que são, segundo vários autores [7] [9] [16], adequados e dos mais utilizados em estudos relacionados com a retenção onde se pretende prever a classe de um atributo nominal. Neste caso, o atributo *desistência* é o atributo que se pretende prever (o atributo alvo), uma vez que no caso dos ginásios e de outras instalações públicas e privadas que prestam serviços desportivos regulares, a situação de um cliente é avaliada por um atributo nominal que pode assumir os valores Ativo – é cliente – ou Desistente – deixou de ser cliente. Neste trabalho optou-se por definir o atributo como um valor binário, correspondendo o valor 1 a um Desistente, e o valor 0 a um Ativo.

Gama *et al.* [15] refere que uma vez que o processo de construção de uma árvore seleciona os atributos a usar resultam modelos que tendem a ser bastante robustos em relação à adição de atributos irrelevantes e redundantes. O produto utilizado para a construção do modelo preditivo através do algoritmo *decision trees* aplica automaticamente um processo de seleção de atributos (*Feature Selection*) podendo-se optar por um dos seguintes três métodos: Entropia, *Bayesian with K2 Prior* e *Bayesian Dirichlet Equivalent with Uniform Prior*. O *Bayesian Dirichlet Equivalent with Uniform Prior* foi o que resultou num modelo preditivo com menor número de nós e em que os nós terminais apresentam maior número de registos.

O modelo resultante da seleção automática de atributos resultou num conjunto de atributos redundante que foi simplificado para benefício e inteligibilidade do modelo a criar, sem grande prejuízo da precisão do modelo. Após a simplificação, e por ordem de importância, os atributos resultantes são: o *número de dias sem frequência* (classificação B), a *frequência contratada semanal*, o *número de renovações*, o *número de meses da inscrição* (classificação B), a *idade* (classificação B) e o *volume de negócios* (classificação A), dos quais se apresentam os histogramas na Fig. 1.

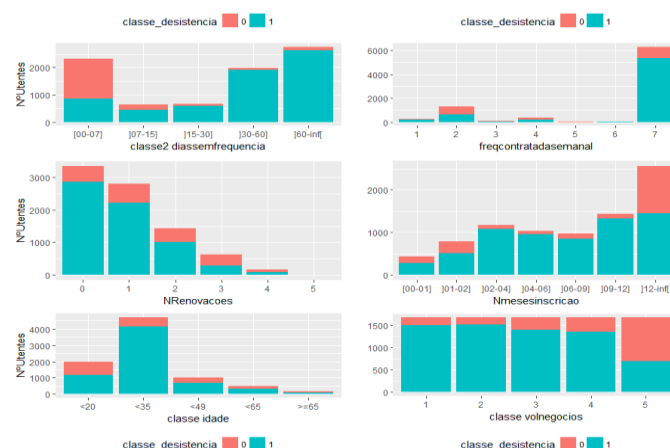


Figura 1- Histogramas dos atributos utilizados no modelo

O gráfico de correlação é apresentado na Fig. 2.

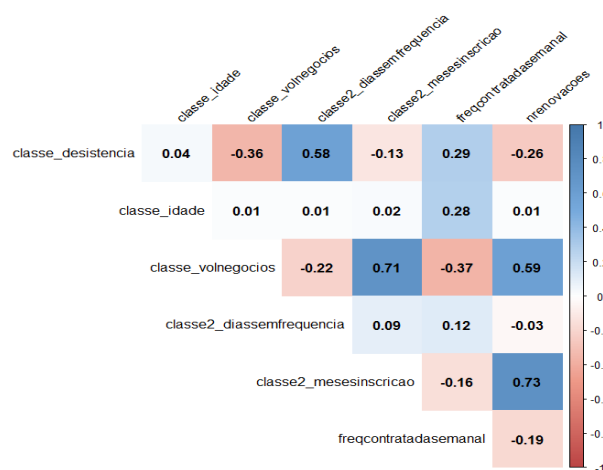


Figura 2 – Gráfico de correlação dos atributos usados no modelo

Tendo em atenção a necessidade de treinar e avaliar a precisão dos modelos construídos, consideraram-se os métodos *Holdout* e *Cross-Validation* [17].

No método *Holdout*, os dados foram particionados em dois conjuntos disjuntos designados por conjunto de treino, com 70% dos dados, e por conjunto de teste, que utiliza os restantes 30% para construir uma matriz de confusão que se apresenta na Tab. 3 a partir da qual é calculada a precisão do modelo e as métricas apresentadas na Tab. 4.

TABELA 3 - MATRIZ DE CONFUSÃO ENCONTRADA PELO MÉTODO *HOLDOUT*

<i>Real / Previsto</i>	<i>Não Desistente</i>	<i>Desistente</i>
Não Desistente	489	75
Desistente	198	1752

TABELA 4 - RESULTADO DAS MÉTRICAS OBTIDAS COM O MÉTODO *HOLDOUT*

<i>Métrica</i>	<i>Valor</i>
Accuracy rate	89.14%
Error rate	10.86%
Sensitivity / True Positive Rate	89.85%
Specificity / True Negative Rate	86.70%
False positive rate	13.30%
Precision rate	95.89%
F-Score	92.77%

No *Cross-Validation*, os registos foram segmentados em 10 partições de igual tamanho (586 ou 587 registos). Em cada execução, uma das partições é usada para testar enquanto as outras são usadas para treino do modelo, sendo o procedimento repetido 10 vezes sendo que cada partição é usada apenas uma vez para testar o modelo. Com este método de avaliação obtém-se a matriz de confusão e os desvios padrão apresentadas na Tab. 5 que permite concluir haver uma variação pequena, em função do coeficiente de variação.

TABELA 5 - MATRIZ DE CONFUSÃO E DESVIOS PADRÃO NO MÉTODO *CROSS-VALIDATION*

<i>Real/Previsto</i>	<i>Classificações médias</i>		<i>Desvios Padrão</i>	
	<i>Não Desistente</i>	<i>Desistente</i>	<i>Não Desistente</i>	<i>Desistente</i>
<i>Não Desistente</i>	107.597	40.6927	6.7004	9.1919
<i>Desistente</i>	29.6033	408.9076	6.921	9.5159

Os resultados das métricas obtidas e indicadas na Tab. 6 são ligeiramente inferiores ($\pm 1,12\%$) às obtidas com o método *Holdout*, exceto na métrica *Precision rate*, onde a diferença é de 2.64%, e nas métricas relacionadas com a classificação dos não desistentes (*Specificity* e *False Positive Rate*) onde a diferença é claramente superior (8.28%).

TABELA 6 - RESULTADO DAS MÉTRICAS OBTIDAS COM O MÉTODO *CROSS-VALIDATION*

<i>Métrica</i>	<i>Valor</i>
Accuracy rate	88.02%
Error rate	11.98%
Sensitivity / True Positive Rate	90.95%
Specificity / True Negative Rate	78.42%
False positive rate	21.58%
Precision rate	93.25%
F-Score	92.08%

A árvore resultante do modelo, onde estão expressas as regras que caracterizam os nós, é apresentada na Fig. 3.

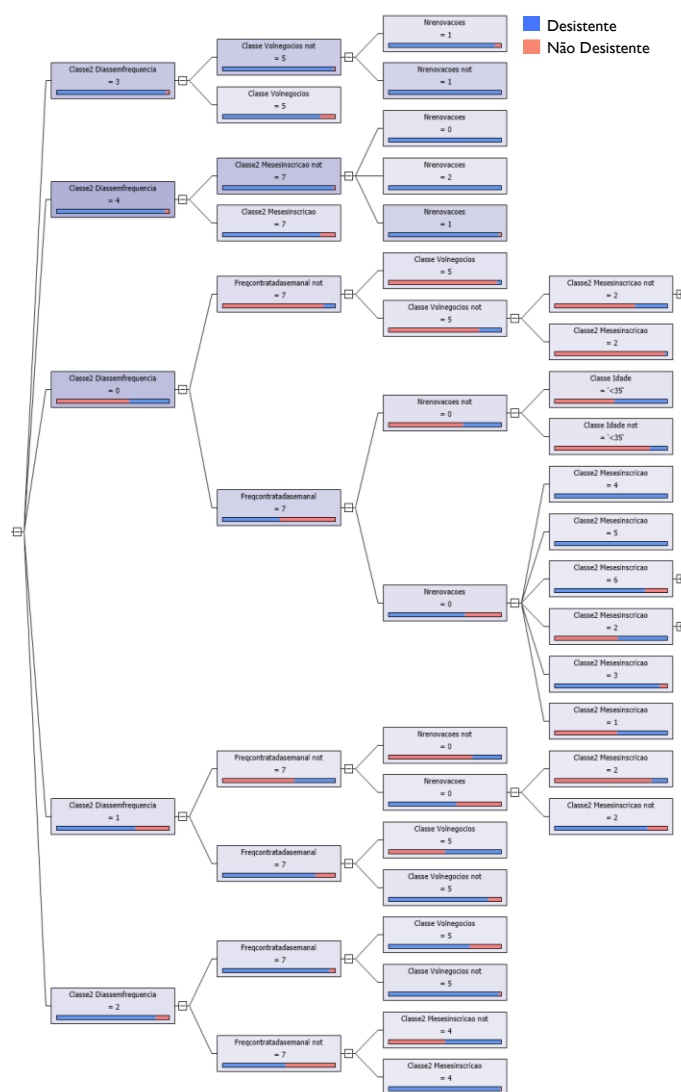


Figura 3 - Árvore de decisão obtida com o modelo indicado

V. FIDELIZAÇÃO: PLANEAMENTO DE AÇÕES

Para concentrar os esforços de aumento da taxa de retenção em grupos mais suscetíveis de abandono da prática desportiva

seguiu-se uma abordagem *bottom-up* conforme refere Gorgoglione [18] e propõe-se a aplicação de ações de fidelização sobre a segmentação obtida com o modelo preditivo referido no ponto anterior que obedece ao critério de utilidade dos segmentos segundo Kotler [19] que indica que uma segmentação só é útil se os segmentos obedecerem a cinco critérios: sejam mensuráveis, substanciais, acessíveis, diferenciáveis e acionáveis.

As ações de fidelização devem ser desenhadas de acordo com as características e nível de pré-desistência dos utentes que se encontram em cada folha da árvore de decisão criada pelo modelo preditivo escolhido e encadeadas num *workflow* cuja primeira ação que o define atinge inicialmente todos os utentes que se encontram nesse nó. No passo seguinte do *workflow*, que se sucede após um intervalo temporal definido de acordo com os critérios da organização, a ação de fidelização seguinte só atingirá os utentes cujo comportamento não foi alterado, ou seja, que se mantêm no mesmo nó da árvore de decisão.

Caso o comportamento ou características do utente se alterem, leia-se “as características e comportamento do utente deixam de o colocar no nó da árvore onde inicialmente foi colocado”, o utente deixa de ser alvo das ações de fidelização encadeadas no *workflow* em que se encontrava e caso não exista nenhum *workflow* aplicado ao novo nó da árvore em que o utente passou a enquadrar-se, ou a ser alvo de um *workflow* diferente aplicado ao nó da árvore em que agora se enquadra.

As ações possíveis de aplicar em cada *workflow* podem ser agrupadas de acordo com a sua finalidade em ações informativas - informação personalizada sobre os horários que pode frequentar e sobre benefícios incluídos no seu serviço de que não está a fazer uso -, de perceção da satisfação - Net Promoter Score (NPS), motivo das ausências, outros inquéritos de qualidade - e de oferta de benefícios - frequência gratuita de aulas de outras atividades não incluídas no seu serviço, aulas de substituição, participação em master-classes ou outros eventos, utilização gratuita de recintos, brindes ou vales.

Uma vez que é cada vez mais difícil escolher um meio eficaz para levar as mensagens aos clientes dada a fragmentação e desordem dos atuais canais de comunicação [19] propõe-se a utilização dos canais pessoais como o telefone (quer no que diz respeito às chamadas de voz como ao envio de SMS - *Short Message Service* e MMS - *Multimedia Message Service*) e o e-mail, estabelecidos na ordem e no *timing* adequado, para implementar as ações de fidelização, até porque se apresentam como potenciais canais para promover ações cujo objetivo é o de aumentar a lealdade e as taxas de retenção dada a sua capacidade de personalização e de obtenção de feedback relativo às ações efetuadas.

No que diz respeito aos e-mails em que a obtenção de feedback sobre entregas e leituras pode-se tornar uma tarefa mais complicada pelos diversos fatores que podem impedir que chegue ao seu destino, o processo implementado deve seguir algumas boas práticas que permita reduzir o risco de o mesmo não ser entregue ou de ser considerado SPAM. São exemplos dessas boas práticas a construção correta da versão HTML do e-mail, incluir uma versão *plain-text* do mesmo, personalizar cada mail enviado, evitar a utilização de palavras como “grátis” ou “promoção” e ser enviadas de um endereço pessoal, para além

de conterem uma possibilidade de o destinatário se remover definitivamente da lista de distribuição (*unsubscribe*).

Num nível mais aplicacional, os e-mails devem ser interativos e conter *links* com identificação que levem os utentes a seguir para uma página onde obtêm a informação ou benefício a que diz respeito a ação. Por outro lado, na página a que se acedeu pelo *link* devem constar mecanismos que permitam apreender quais as ações do utente, nomeadamente de que tomou conhecimento da informação ou benefício que lhe foi proposto, se tenciona utilizá-lo e quando.

Como este modelo pretende, a observação da Fig. 3 permite extrair perfis de abandono sobre os quais se podem fazer incidir *workflows* adequados. Por exemplo, 100% dos utentes que apresentam as seguintes características desistem: frequentaram as instalações nos últimos 7 dias da sua inscrição (*classe diassemfrecuencia=0*), contrataram uma frequência de 7 dias por semana - normalmente designado por livre trânsito - (*frequcontratadasemana=7*), nunca estiveram inscritos em mais do que uma época (*nrenovacoes=0*), e as suas inscrições têm uma duração entre 5 a 9 meses (*classe mesesinscricao in [4,5]*). Estando um utente nestas condições, a organização pode optar por inicialmente enviar um e-mail personalizado divulgando os horários e oferecendo a frequência de aulas incluídas (ou não) no serviço contratado pelo utente, seguido de um SMS e de um contato pessoal personalizado e efetuado por um técnico qualificado.

Como forma de medir a efetividade da aplicação das ações, propõe-se a criação de dois grupos de utentes: aqueles sobre quem se faz incidir as ações de fidelização - que constituem o grupo de teste -, e outro grupo sobre o qual não serão aplicadas as ações - que constituem o grupo de controlo.

Para que os grupos sejam criados de forma uniforme, propõe-se que a sua constituição se faça pela divisão em partes iguais dos utentes que se encontram em cada nó da árvore num determinado momento. Uma vez que em cada nó se encontram utentes com probabilidades de abandono diferentes, ordena-se os mesmos por ordem crescente dessa probabilidade, alternando-se a colocação dos mesmos em cada um dos grupos até se esgotarem os utentes desse grupo, repetindo-se a operação para os utentes que se encontram em cada nó da árvore.

No final de cada ciclo de *workflow* a observação da taxa de retenção dos dois grupos permitirá concluir pelo sucesso das ações de fidelização implementadas. Outras medidas suplementares para o sucesso das ações são as taxas de inativação (utilizadores que optaram por cancelar a subscrição), de abertura das páginas a partir dos *links* indicados nos mails e SMS, e da utilização dos benefícios oferecidos.

VI. CONCLUSÕES

Este trabalho apresenta um contributo substantivo no que concerne à implementação de metodologias que visam aumentar a fidelização e as taxas de retenção dos utentes de serviços desportivos regulares, uma vez que:

- A maior parte dos estudos de fidelização e retenção desta área envolvem dados de inquéritos. Contudo, as organizações supõem os dados existentes nas bases de

dados de utentes que são negligenciados. Neste trabalho são utilizados dados reais de ginásios.

- Aplica modelos preditivos sobre os dados reais, utilizando os atributos mais relevantes para a criação dos perfis, nomeadamente os relacionados com a frequência, como os principais fatores que influenciam o abandono, ao contrário do que refere Gonçalves [12] que indica o *género* como um dos principais fatores. Uma vez que a árvore resultante do algoritmo de *data mining* agrupa nos seus nós utentes que partilham comportamentos semelhantes é possível definir perfis de utentes fiéis e de utentes em pré-abandono e dirigir ações de fidelização personalizadas aos que se encontram acima de um determinado patamar de desistência.
- As ações de fidelização propostas são encadeadas num *workflow*, esperando-se que em cada patamar as ações incidam sobre um menor número de utentes do que no patamar anterior; um *workflow* com sequências do tipo “e-mail -> SMS -> contato telefónico pessoal” apresentar-se-á em forma de pirâmide havendo um menor número de SMS enviados do que e-mails, e um menor esforço nos contatos telefónicos pessoais, com uma redução significativa do custo e do esforço dedicado à retenção. A metodologia proposta é incremental no sentido de que periodicamente cada passo referenciado em *database marketing* pode ser revisto e melhorado.

Atingir estes objetivos constitui-se não só como uma poderosa ferramenta de suporte a decisão e de marketing, dada a otimização de recursos e redução de esforço que significa, resultando também num aumento da rentabilidade uma vez verificada uma maior taxa de retenção.

VII. TRABALHO FUTURO

Como áreas ainda a investigar refere-se a inclusão de atributos relativos à qualidade do serviço que pode melhorar substancialmente o resultado dos modelos e/ou ajudar a criar perfis mais significativos que poderão resultar em ações de fidelização diferentes das obtidas.

Por outro lado, deve-se proceder à avaliação do comportamento do modelo preditivo em contínuo, nos vários meses do ano, e especialmente em períodos festivos e de mudança de época em que se conhece haver maiores taxas de abandono, e ajustar o planeamento das experiências aos resultados obtidos pelos modelos preditivos nestes períodos.

Por fim, será também necessário avaliar o impacto que a entrada em vigor a 25 de Maio de 2018 do Regulamento 2016/679 do Parlamento Europeu e do Conselho (RGPD), de 27 de Abril de 2016, relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados. Uma vez que tanto as ações de fidelização, como a aplicação de técnicas de *machine learning* utilizam dados pessoais, será necessário analisar e propor formas de ultrapassar as limitações e/ou restrições que este regulamento possa colocar.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] G. Howat and G. Assaker, “Outcome quality in participant sport and recreation service quality models: Empirical results from public aquatic centres in Australia,” *Sport Management Review*, vol. 19, no. 5, pp. 520–535, 2016.
- [2] S. Avourdiadou and N. D. Theodorakis, “The development of loyalty among novice and experienced customers of sport and fitness centres,” *Sport Management Review*, vol. 17, no. 4, pp. 419–431, 2014.
- [3] AGAP, “Barómetro 2016.” 2016.
- [4] L. Cavique, “Relatório da Unidade Curricular de Database Marketing, 2005-2006,” *Esc. Super. Comun. Soc. Inst. Politécnico Lisboa*, no. unpublished, 2006.
- [5] D. Peppers and M. Rogers, *Managing Customer Relationships*, vol. 30097. Wiley, 2004.
- [6] P. Pinheiro and L. Cavique, “Determinação de Padrões de Desistência em Ginásios,” *Rev. Ciências da Comput.*, vol. 10, pp. 33–60, 2015.
- [7] V. Mahajan, R. Misra, and R. Mahajan, “Review of Data Mining Techniques for Churn Prediction in Telecom,” *Journal of information and organizational research*, vol. 39, no. 2, pp. 183–197, 2015.
- [8] L. Cavique, “Micro-Segmentação de Clientes com Base em Dados de Consumo : Modelo RM-Similis,” *Rev. Port. e Bras. Gestão*, vol. 2, no. n° 3, pp. 72–77, 2003.
- [9] E. Siegel, *PREDICTIVE ANALYTICS The power to predict who will click, buy, lie or die*. Wiley, 2013.
- [10] J. Surujlal and M. Dhurup, “Establishing and maintaining customer relationships in commercial health and fitness centres in South Africa,” *International Journal of Trade, Economics and Finance*, vol. 3, no. 1, pp. 14–18, 2012.
- [11] V. Marisa and D. Pedragosa, “O Processo de Fidelização no Health Club Bioritmo,” 2006.
- [12] C. Gonçalves, “Variáveis Internas e Externas ao Indivíduo que influenciam o Comportamento de Retenção de Sócios no Fitness,” *Podium Sport, Leisure and Tourism Review*, vol. 1, no. 2, pp. 28–58, 2012.
- [13] M. Frota, “Gestão da Retenção,” in *Manual de Gestão de Ginásios e Health Clubs - Excelência no sector do Health & Fitness*, 2011, pp. 103–148.
- [14] J. Trujillo and S. Luján-Mora, “A UML based approach for modeling ETL processes in data warehouses,” *Conceptual Modeling-ER 2003*, vol. 2813, pp. 307–320, 2003.
- [15] J. Gama, A. P. de L. Carvalho, K. Faceli, A. C. Lorean, and M. Oliveira, *Extração de Conhecimento de Dados*, 3ª edição. 2017.
- [16] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Pearson, 2006.
- [17] P.-N. Tan, M. Steinbach, and V. Kumar, *Classification : Basic Concepts , Decision Trees , and*, vol. 67, no. 17. 2006.
- [18] M. Gorgoglione, “Beyond Customer Churn: Generating Personalized Actions to Retain Customers in a Retail Bank by a Recommender System Approach,” *Journal of Intelligent Learning Systems and Applications*, vol. 3, no. 2, pp. 90–102, 2011.
- [19] P. Kotler and K. L. Keller, *Marketing Management*, vol. 22, no. 4. 2009.