

# MODELAGEM DE PROBABILIDADE DE CHURN

## CHURN PROBABILITY MODEL

### MODELADO DE PROBABILIDAD DE CHURN

#### RESUMO

Clientes podem abandonar uma organização ainda que altos investimentos em prospecção e retenção sejam realizados, o que requer diagnóstico e compreensão. Este artigo objetiva modelar a probabilidade de clientes abandonarem o relacionamento com uma organização, fenômeno conhecido como *churn*, utilizando dados do histórico de relacionamento cliente/empresa, validar o modelo em uma segunda amostra e descrever as possíveis variáveis que influenciam o abandono/permanência do cliente. Utilizou-se o modelo de regressão logística em uma amostra de calibração de 70.000 clientes que possuíam cartão de crédito próprio de uma grande rede varejista. Dezesesseis variáveis explicativas (14 características individuais e duas variáveis comportamentais) foram usadas e o modelo foi validado em uma amostra de 30.000 clientes, usando-se o teste de KS (Kolmogorov-Smirnov) e a curva ROC (*Receiver Operating Characteristic*), que demonstraram a boa adequação do modelo à amostra de validação. Implicações da pesquisa e sugestões para futuras investigações são discutidas à luz da gestão do relacionamento com o cliente.

**PALAVRAS-CHAVE** Abandono de cliente, *churn*, gestão do relacionamento com o cliente, valor do ciclo de vida do cliente, regressão logística.

**Delane Botelho** [delane.botelho@fgv.br](mailto:delane.botelho@fgv.br)

Professor da Escola de Administração de Empresas de São Paulo, Fundação Getúlio Vargas e pesquisador do Centro de Excelência em Varejo, Fundação Getúlio Vargas – São Paulo – SP, Brasil

**Frederico Damian Tostes** [fredt@ig.com.br](mailto:fredt@ig.com.br)

Mestre em Gestão Empresarial pela Escola Brasileira de Administração Pública e de Empresas, Fundação Getúlio Vargas – Rio de Janeiro – RJ, Brasil

Recebido em 29.04.2009. Aprovado em 08.11.2010

Avaliado pelo sistema *double blind review*

Editor Científico: José Mauro da Costa Hernandez

**ABSTRACT** Customer may abandon the organization, despite high investments made by the organization in their prospection and retention, which demands diagnose and understanding. The objectives of this article are threefold: i) to model the probability of churn, which refers to the rate of customers who leave a relationship with an organization during a given period; ii) to validate the model; and iii) to describe possible variables that explain the abandon of customers. We used historical data from a large retail chain's dataset, and applied logistic regression in a sample of 70.000 users of a private label credit card to calibrate the model. Sixteen variables (14 individual characteristics and two behavior variables) were used. The model was validated in a different sample of 30.000 customers from the same dataset, through Kolmogorov-Smirnov (KS) test and Receiver Operating Characteristic (ROC) curve, confirming the good prevision power of the model. Managerial implications and suggestion for future research are discussed based on customer relationship management field of research.

**KEYWORDS** Churn, data mining, customer relationship management, consumer lifetime value, logistic regression.

**RESUMEN** Los clientes pueden abandonar una organización a pesar de grandes inversiones en prospección y retención, lo que requiere diagnóstico y comprensión. Este artículo tiene el objetivo de modelar la probabilidad de que los clientes abandonen la relación con una organización, fenómeno conocido como *churn*, utilizando datos del historial de la relación cliente/empresa, validar el modelo en una segunda muestra y describir las posibles variables que influyen el abandono/permanencia del cliente. Se utilizó el modelo de regresión logística en una muestra de calibración de 70.000 clientes que poseían tarjeta de crédito propia de una gran red minorista. Fueron usadas dieciséis variables explicativas (14 características individuales y dos variables comportamentales) y se validó el modelo en una muestra de 30.000 clientes, usando el test de KS (Kolmogorov-Smirnov) y la curva ROC (*Receiver Operating Characteristic*), que demostraron la buena adecuación del modelo a la muestra de validación. Son discutidas, a la luz de la gestión de la relación con el cliente, las implicaciones de la investigación y sugerencias para futuras investigaciones.

**PALABRAS CLAVE** Abandono de cliente, *churn*, gestión de la relación con el cliente, valor del ciclo de vida del cliente, regresión logística.

## INTRODUÇÃO

É geralmente mais caro conquistar novos clientes do que manter os já existentes. Em muitos setores, o custo de aquisição de novos clientes pode ser cinco vezes superior ao seu custo de retenção (KURTZ e CLOW, 1998). Na indústria de telecomunicações, por exemplo, o custo de aquisição foi empiricamente auferido entre cinco e oito vezes superior ao custo de retenção do cliente (AU e outros., 2003). A compreensão do fenômeno *churn*, ou seja, o abandono ou a migração do cliente para o concorrente, possibilita que a organização atue sobre as variáveis-chave, e controláveis, que o influenciam, na tentativa de minimizá-lo. No setor bancário americano, por exemplo, estima-se que 30% da base de clientes seja vulnerável à migração (ACCENTURE, 2009) e algumas variáveis podem ajudar a prever tal vulnerabilidade, como características do indivíduo, atributos da organização e do produto e fatores externos (como variáveis macro econômicas e culturais).

A perda de clientes por atrito no relacionamento, como no caso de uma autorização não concedida em uma venda por cartão de crédito, ou o abandono do cliente, quando este troca de varejista/produto sem um motivo aparente, pode ocorrer em qualquer fase do ciclo de vida do cliente junto à organização. A gestão do relacionamento com o cliente, tema em que se enquadra este artigo, é responsável, entre outras atividades, por ações que previnam tal atrito ou abandono do cliente.

Atualmente, tanto o cliente tem se tornado mais exigente em termos de individualização do atendimento, quanto tem aumentado o número de clientes e transações em grandes empresas de varejo, seja pelo fenômeno de concentração de poucas grandes empresas, seja pelo maior acesso da população brasileira, na última década, ao consumo (KAKWANI e outros, 2010). Se, por um lado, o grande volume de dados é um obstáculo para a individualização do relacionamento, por outro esse volume produz uma massa de dados, capturada sistemicamente em forma digital, capaz de permitir respostas individualizadas aos clientes por meio de aplicação de técnicas de mineração de dados (*data mining*), associadas ao CRM (*Customer Relationship Management*).

No Brasil, ainda são escassos os artigos científicos que abordam conceitos de *data mining* e a questão do abandono de clientes, destacando-se Ferreira e outros (2004) e Cister e Shehata (2008). Este artigo tem como objetivos: i) modelar a probabilidade de clientes em (não) abandonar o relacionamento com uma organização, utilizando dados do histórico de relacionamento cliente/empresa, em uma amostra de calibração; ii) descrever as possíveis

variáveis que influenciam o abandono/permanência do cliente; verificar o desempenho do modelo em uma amostra de validação. Os dados de relacionamento do cliente referem-se, aqui, ao uso de cartão de crédito próprio de uma grande rede varejista.

Diante do grande número de emissores de cartão de crédito, é alta a competição entre as empresas, que utilizam criativas estratégias para reter clientes e maximizar o lucro. Uma vez conquistado o cliente, tem início o seu ciclo de vida na organização, e um dos principais objetivos do gestor de relacionamento é prolongar ao máximo esse ciclo, principalmente daqueles clientes mais rentáveis. Por meio do relacionamento com o cliente são, então, extraídas informações sobre suas necessidades, preferências e hábitos, gerando ações de desenvolvimento e modificação de produtos e serviços para que tal relacionamento perdure.

A escolha aqui pelo cartão de crédito se dá por duas razões. Primeiro, é um serviço que vem crescendo como meio de pagamento e relacionamento nos mercados brasileiro e mundial: segundo a Associação Brasileira de Cartões de Crédito (2009), a indústria brasileira de cartões movimentou em 2008 mais de R\$ 375 bilhões em transações e os cartões próprios de lojistas passaram de 99 milhões de unidades em circulação em 2005 para 167 milhões em 2008. Os demais cartões de crédito no mesmo período passaram de 67 milhões para 124 milhões de unidades emitidas. A tendência é de crescimento nos próximos anos, principalmente com sua penetração entre os indivíduos das classes C e D, que atualmente é de 63% e 40%, respectivamente (nas classes A e B, a penetração do uso de cartão de crédito é de 91% e 82%, respectivamente) (DATAFOLHA, 2010). Segundo, trata-se de um serviço que gera eletronicamente alta quantidade de dados nos bancos de dados de cada administradora. Essa fonte de dados torna o setor de cartões de crédito um importante usuário das técnicas de mineração de dados e provedor de dados para pesquisas. As decisões de aprovar um indivíduo candidato ao crédito, de aumento do limite de crédito aos clientes existentes e a análise de fraudes são exemplos de uso de modelos advindos dos grandes bancos de dados de administradoras de cartões de crédito.

A delimitação deste estudo é para uma amostra de clientes usuários de cartão próprio de uma grande rede varejista, não incluindo demais compradores da rede. Não se pretendeu, também, analisar índices de eficiência da administradora na gestão da fidelização de seus clientes. O artigo está assim estruturado: esta introdução apresentou o tema, os objetivos e a relevância da pesquisa. A seguir, são apresentados o referencial teórico, o método da pes-

quisa empírica e os resultados. O artigo é finalizado com as principais conclusões, implicações e sugestões para futuras pesquisas.

## REFERENCIAL TEÓRICO

### CRM e churn

A lealdade muitas vezes é interpretada como a retenção real do cliente, que é o fundamento da gestão de relacionamento com o cliente (GUSTAFSSON e outros, 2005). Apesar de não garantir a lealdade, a satisfação, definida como uma avaliação global do cliente ao desempenho de uma oferta ao longo do tempo (JOHNSON e FORNELL, 1991), tem um forte efeito positivo sobre as intenções de fidelização de clientes em uma ampla gama de categorias de produtos e serviços. O principal mecanismo para manter relacionamentos de longa duração com os clientes é aumentar sua satisfação por meio das interações com a empresa, tanto em qualidade quanto em quantidade (VAVRA, 1993).

Em setores altamente competitivos, como no caso do setor de cartões de crédito, a satisfação do cliente desempenha papel crítico para o seu relacionamento e, conseqüentemente, para sua manutenção. O conceito de satisfação comumente tratado na literatura caracteriza-se pela comparação (ou avaliação) subjetiva dos níveis esperados e recebidos da experiência com o produto ou serviço (OLIVER, 1999), que está relacionado ao paradigma da desconformidade. A resposta de satisfação ou insatisfação do indivíduo ocorre por meio de comparação entre expectativa e desempenho. Caso o desempenho do produto seja superior à expectativa do cliente, uma situação favorável é desencadeada, ou seja, a satisfação.

A compra consistente pode não ser um indicador de lealdade, devido a fatores como compra por impulso ou conveniência, portanto compras inconsistentes podem mascarar a lealdade de consumidores fiéis a várias marcas e não se pode inferir sobre lealdade ou deslealdade com base apenas em padrões de compras repetidas. Mas, a preocupação com a lealdade dos clientes torna-se mais relevante na medida em que a manutenção dos relacionamentos de longo prazo é essencial para a rentabilidade de uma empresa (OLIVER, 1999).

Uma das formas de aumentar as interações do consumidor com a empresa é a utilização de ferramentas de CRM que, de acordo com Neslin e outros (2006), é um processo interativo que transforma informação sobre os clientes em relacionamentos que beneficiam tanto a empresa quanto o cliente. O aprendizado contínuo, o conhecimento do

cliente e o incremento da lucratividade de longo prazo constituem-se nos dois aspectos mais significativos do processo de CRM. As estratégias que visam ganhar maior participação no hábito de consumo do cliente (*share of wallet*), em vez de maior participação de mercado, são o objetivo principal do CRM (BRETZKE, 2000).

O CRM envolve capturar os dados do cliente disponíveis por toda a empresa, consolidar todos os dados capturados interna e externamente em um banco de dados central, analisar os dados consolidados, distribuir os resultados dessa análise aos vários pontos de contato com o cliente e usar essa informação, ao interagir com o cliente por meio de qualquer ponto de contato com a empresa (PEPPERS e ROGERS, 2001).

Com a disponibilização de tecnologia, criaram-se condições para o desenvolvimento de modelos que permitem a distinção de clientes, em um processo de segmentação. Isto possibilitou ações de retenção daqueles mais interessantes para a organização, do ponto de vista de rentabilidade e lucratividade, além de possibilitar o desenvolvimento de produtos e serviços customizados. Somente a partir da década de 1990 esta prática se tornou acessível em grande escala, devido à disponibilidade de recursos tecnológicos (GUNNARSSON e outros, 2007). A tecnologia de hoje permite muitas empresas acompanharem seus clientes de forma quase individual. Tecnologias interativas pela internet (*e-marketing*), telemarketing, TV interativa e ferramentas de automação da força de vendas permitem a conexão automática dos clientes à empresa. O *feedback* é facilitado, ao se obter informação ao nível individual sobre as especificações e modificações dos produtos, do comportamento de compra e expectativas dos clientes. Também, a tecnologia de personalização em massa permitiu que a empresa configurasse as suas ofertas digitalmente, produzindo inúmeras variações do mesmo produto com reduções de custos típicas do mercado de massa, dando origem à customização em massa. As comunidades virtuais, *blogs* e *microblogs* atualmente influenciam o relacionamento de seus usuários com muitas empresas e a forma como avaliam produtos e serviços. *Blogs*, que são *sites* na internet que contêm informações pessoais com reflexões, comentários e *hyperlinks*, muitas vezes fornecidos pelo escritor, contêm fatos e opiniões que podem atender internautas que compartilham características comuns, inclusive influenciando-os ao abandono ou adoção de produtos ou organizações (SINGH e SINGH, 2008). Comunidades virtuais também têm sido objeto de estudo em marketing de relacionamento no Brasil (AÑAÑA e outros, 2008), apesar de nenhum estudo enfocar a sua

influência na taxa de abandono ou retenção de clientes de uma organização.

As aplicações da mineração de dados mostram que a decisão do cliente de interromper o relacionamento com a empresa acontece meses antes da real interrupção. Por isto, a detecção antecipada das mudanças no comportamento dos clientes que sinalizam um potencial abandono ou mudança de comportamento (como a redução do *ticket* médio de compra) é fundamental, ao contrário da detecção apenas durante o evento (BOSE e CHEN, 2009). A perda de clientes sofrida por uma empresa, ao longo de um período de tempo específico, também é conhecida como *churn*. É uma medida de clientes com atrito, relacionada à taxa anual de *turnover* da base de clientes (LEJEUNE, 2001). Os termos *churn* e *abandono* serão usados indistintamente neste artigo.

O valor econômico da retenção de clientes é amplamente reconhecido na literatura (NESLIN e outros, 2006) e apresentado das seguintes formas: i) o esforço de retenção de clientes diminui a necessidade de prospectar novos, permitindo que as organizações se concentrem nas necessidades dos atuais, buscando o relacionamento duradouro; ii) clientes de longo relacionamento tendem a comprar mais e, se satisfeitos, podem prover referências positivas pelo “boca a boca”; iii) atender e manter clientes de longo prazo é menos oneroso devido ao maior conhecimento sobre seu ciclo de vida ou de consumo, via informação dos bancos de dados, mesmo sabendo que clientes de longo prazo não são necessariamente fiéis; iv) tais clientes tendem a ser menos sensíveis às atividades de marketing da concorrência; e v) a perda de clientes não é somente um custo de oportunidade pela redução de vendas, mas também pela necessidade de atrair novos clientes para suprir essa perda.

O *churn* pode ser correlacionado com o ciclo de vida do produto ou da empresa. No estágio de introdução e crescimento, o número de novos consumidores excede fortemente o número de *churners*. No entanto, o impacto do *churn* torna-se nitidamente mais sensível em produtos na fase de maturidade e declínio (LEJUNE, 2001). Fatores como processos de desregulamentação e abertura de mercado são grandes influenciadores nas taxas de *churn* e, consequentemente, nas estratégias das empresas. A taxa de *churn* pode ser reduzida, além do uso de processos de CRM, pela realização de ações de pós-venda, via incremento do nível ampliado do produto. O *churn*, também, está diretamente associado ao tempo em que o indivíduo permanece como cliente de uma organização, portanto está diretamente relacionado ao conceito de *Consumer Life Time Value* (CLV ou valor do ciclo de vida do cliente)

(NESLIN e outros, 2006), que se refere à mensuração do lucro, presente e futuro, gerado por um cliente durante o seu ciclo de vida junto à empresa (GREENBERG, 2001). Glady e outros (2009), usando uma abordagem centrada no cliente, definiram *churner* como aquele cliente cujo CLV é decrescente ao longo do tempo.

### **Database marketing**

Desde os anos 1990, a evolução dos sistemas e tecnologias de informação e comunicação tem permitido uma crescente disseminação do *database marketing*, uma ferramenta que aplica sistemas de informação ao marketing direto, baseado em banco de dados com vistas ao gerenciamento do relacionamento com o cliente (CRM). O principal diferencial dessa metodologia é a extensiva atenção dada à análise dos dados, visando a agrupar toda informação disponível sobre clientes potenciais e reais em uma base de dados central. A proposta é a utilização de técnicas estatísticas e modelos que descrevam o comportamento e preferências dos clientes com o intuito de segmentá-los, comunicar-se de forma dirigida com eles e desenvolver novos produtos e serviços ou melhorar os atuais. Os profissionais que trabalham diretamente com o *database marketing* são usuários de *data warehouses* (armazéns de dados), pois lidam com uma grande quantidade de dados e informações sobre os clientes, o que possibilita a construção de modelos mais precisos (GUNNARSSON e outros, 2007).

Essa metodologia define toda uma classe de ferramentas que automatizam as funções de contato com o cliente e, fundamentalmente, significam uma mudança de atitude corporativa. Embora a ideia de armazenar eletronicamente dados dos clientes com o propósito de usá-los em *database marketing* venha sendo desenvolvida há décadas, apenas nos anos 1990 sistemas computacionais disponíveis tornaram possível, por exemplo, ter um completo histórico de um cliente na tela do computador no momento de uma interação (ligação, ida à loja etc.).

Atualmente, não existe uma definição de *database marketing* universalmente aceita e muitas confusões são feitas na utilização dessa expressão, devido principalmente a: i) *database marketing* vem de uma derivação dos princípios do marketing direto, o que confunde sua definição com o próprio marketing direto (PETRISON e outros, 1997); ii) *database marketing* dá suporte ao desenvolvimento da estratégia de marketing de relacionamento e, frequentemente, as duas expressões são usadas indistintamente (SHANI e CHALASANI, 1992); iii) rápidas mudanças no ambiente de negócio e na tecnologia têm permitido um debate sobre a definição de limites do marketing em

geral (se *database marketing* é primordialmente parte do escopo de marketing ou de tecnologia da informação (BAUER; MIGLAUTSCH, 1992); e iv) as aplicações de *database marketing* variam amplamente entre empresas e setores (STONE e SHAW, 1987). Por esses motivos, a definição de *database marketing* tem um amplo espectro, mas propriedades comuns aparecem na maioria delas: i) é um conjunto de dados informatizado; ii) que possui dados individuais de clientes reais e/ou potenciais, tais como informações demográficas, geográficas, psicográficas e comportamentais; iii) cujas técnicas estatísticas são instrumentos fundamentais para apoio à decisão; iv) com o objetivo de construir relacionamentos segmentados e melhorar o custo efetivo dos programas de marketing para aumentar a rentabilidade por cliente ou rentabilidade geral da empresa (RHEE e MCINTYRE, 2008).

Embora organizações de qualquer tamanho possam empregar *database marketing*, é nas empresas que possuem um número elevado de clientes que o seu emprego se torna mais apropriado. Isto porque uma grande massa de dados permite maiores oportunidades de encontrar segmentos de clientes ou *prospects* (possíveis clientes) que possam ser tratados de maneira customizada. O desafio principal do *database marketing* é garantir a informação que suporte decisões inteligentes e a utilização de medidas de desempenho e modelos de marketing adequados são fundamentais nesse processo. Medidas de desempenho permitem melhorar a efetividade dos programas de marketing: podem ser usadas para captar taxas de resposta das campanhas de marketing, *ticket* médio por cliente, dívida irrecuperável, aquisição de clientes, retenção e reativação. Tais medidas auxiliam na determinação da recenticidade, frequência e valor monetário (RFM) e CLV.

A modelagem de respostas do consumidor por meio de métodos convencionais de pesquisa, principalmente em marketing direto, tem sido predominantemente baseada em um enfoque *a priori* (a fonte de dados consiste no histórico de compra de cada cliente), como no caso do RFM. Nesse modelo, a probabilidade de os consumidores responderem a uma ação mercadológica é prevista pela recenticidade da última compra, pelo histórico da frequência de compra em determinado período e pelo valor monetário das compras de um cliente. O pressuposto básico é que os clientes que compraram mais recentemente, que fizeram compras em maior número e que realizaram as maiores despesas nas suas compras, responderão, provavelmente, mais positivamente a uma ação mercadológica, do que aqueles que compraram há mais tempo, que fizeram compras em menor número e que realizaram despesas menores (FADER e outros, 2005).

A análise de cada categoria (recenticidade, frequência e valor monetário) é feita separadamente, possibilitando segmentar a base de dados. É, portanto, um método utilizado na análise do comportamento do consumidor e na definição de segmentos de mercado.

CLV é um método utilizado para mensurar o valor do cliente, sintetizando as receitas e custos oriundos de transações futuras entre ele(a) e a empresa, e facilita o entendimento de quanto significa financeiramente perder um cliente. Embora seja um conceito simples, incorpora uma série de aspectos complexos, como a estimativa dos lucros que serão gerados pelo cliente no futuro, a probabilidade de o cliente deixar de comprar da empresa ao longo do tempo e o desconto progressivo dos ganhos futuros convertidos a valor presente.

Quando se trabalha com grandes volumes de dados, tanto para o RFM quanto para o CLV, usa-se a mineração de dados, que faz parte de um processo abrangente chamado extração de conhecimento a partir de bases de dados (KDD – *Knowledge Discovery from Databases*), que consiste em descobrir padrões nos dados minerados (ADDRIANS e ZANTINGE, 1996). A extração de conhecimento envolve a preparação da base de dados, a mineração propriamente dita e as consultas aos dados (OLAP – *Online Analytical Processing*) e ao conhecimento (OLAM – *On-line Analytical Mining*) (BERRY e LINOFF, 2004).

A mineração de dados usa tecnologias como redes neurais, modelos de regressão, árvores de decisão e diversas técnicas estatísticas para investigar grandes volumes de dados e construir modelos que visam a prever o comportamento dos clientes. As técnicas estatísticas multivariadas, juntamente com as diversas soluções de *softwares* existentes, tornam hoje essa tarefa mais acessível (GLADY e outros, 2009). Um método geralmente usado em *data mining* para a classificação de clientes, principalmente para a previsão de *churn*, é a regressão logística. Mozer e outros (2000) usaram tal modelo para a previsão da probabilidade de abandono do cliente, usando como variáveis independentes as características individuais de usuários de telefonia celular.

Para diminuir o risco de decisões errôneas na classificação dos clientes, métodos estatísticos têm sido empregados para descrever a habilidade dos modelos de classificação. A qualidade da previsão de um modelo pode ser avaliada a partir de medidas como sensibilidade, especificidade, valores preditivos e razão de verossimilhanças. Gráficos da sensibilidade/especificidade, comumente chamados de curva ROC (*Receiver Operating Characteristic*), permitem contemplar a capacidade preditiva de um modelo de classificação, bem como comparar modelos concorrentes

através de sua escala. A área sob a curva ROC consiste em uma métrica de desempenho do classificador, como explicado adiante nos resultados do presente estudo (SINHA e MAY, 2005).

A próxima seção trata do método estatístico empregado na pesquisa empírica.

## MÉTODO

A pesquisa aqui realizada é descritiva, com uso de dados secundários coletados longitudinalmente. O modelo usado para a análise dos dados é o da regressão logística, que assume a existência de uma variável resposta ( $Y$ ) binária (dicotômica), como no caso de um modelo de previsão de *churn* ( $Y = 1$ , se o cliente não abandonar o uso do cartão de crédito;  $Y = 0$ , caso contrário). Supõe-se  $x_1, \dots, x_k$  como as variáveis explicativas, e  $\pi(x)$  a proporção de clientes que “não abandonaram” o uso do cartão de crédito em função do perfil desse cliente, caracterizado por  $x$ . Portanto, o modelo logístico é adequado para definir uma relação entre a probabilidade de um cliente não abandonar o uso do cartão [ $\pi(X)$ ] e um vetor de características e comportamentos do cliente [ $X_1, \dots, X_k$ ], sendo definido pela função *logit* dada pela expressão:

$$\log\left\{\frac{\pi(x)}{1-\pi(x)}\right\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k,$$

interpretado como a probabilidade de o cliente não abandonar o uso do cartão de crédito de acordo com as características que possui. Toda essa análise pode ser feita também como  $Y = 1$  (se o cliente abandonar o uso do cartão de crédito;  $Y = 0$ , caso contrário), já que a interpretação é a mesma (MADDALA, 1993).

A estimação dos parâmetros para modelos com resposta binária usa geralmente o método de máxima verossimilhança, que consiste em encontrar os valores das estimativas do vetor de parâmetros  $\beta$  que maximizam a função de probabilidade  $L(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1-\pi(x_i)]^{1-y_i}$ . No entanto, matematicamente, é mais simples trabalhar com o logaritmo dessa equação conhecido como *função de log-verossimilhança*, dada por  $l(\beta) = \log L(\beta) = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1-y_i) \ln[1-\pi(x_i)]\}$ . Os valores estimados de  $\beta_0, \beta_1, \dots, \beta_k$  são aqueles que maximizam  $l(\beta)$  e são encontrados diferenciando-se  $l(\beta)$  em relação a cada um dos parâmetros e igualando as expressões resultantes a zero, como  $\{\partial l(\beta) / \partial \beta = 0\}$  (LIAO, 1994).

O produto aqui em estudo foi o cartão de crédito e a administradora desse cartão é uma grande rede varejista que

possuía, em 2008, 41 lojas nos estados do Rio de Janeiro, Espírito Santo, Minas Gerais, Sergipe, Pernambuco e Alagoas. Para a pesquisa empírica foi selecionada aleatoriamente uma amostra de 100.000 clientes que possuíam cartão de crédito próprio dessa rede, dos quais 70.000 registros foram utilizados para calibração do modelo (estimativas dos coeficientes) e 30.000 registros para sua validação. Todos os clientes foram identificados por um código único (CPF – Cadastro de Pessoas Físicas).

Muitas vezes, em uma administradora de cartão de crédito, o interesse está na permanência do cliente já conquistado, portanto, a partir de variáveis explicativas (características e comportamento do cliente), pode-se analisar o perfil de relacionamento e antecipar possíveis rupturas. As variáveis explicativas referem-se às informações fornecidas pelos clientes no momento do preenchimento da proposta de crédito (quando o cliente ainda é considerado um *prospect*). A variável dependente ( $Y$ ) foi construída por informações obtidas da base de dados referentes aos extratos (faturas mensais do cartão de crédito) ao longo de 18 meses, de setembro/2006 a fevereiro/2008. Os clientes que compuseram a amostra foram aqueles ativos (pelo menos uma compra) nos seis primeiros extratos mensais (setembro/2006 a fevereiro/2007). Então, baseado nas informações sobre o comportamento de compra do cliente nos seis primeiros meses e as informações cadastrais das características individuais, o modelo foi construído para prever a probabilidade de (não) abandono nos 12 meses seguintes (março/2007 a fevereiro/2008). Classificaram-se os clientes como ativos (não abandono, ou seja, pelo menos um extrato de utilização do cartão de crédito nos últimos 12 meses) e inativos (abandono, ou seja, nenhum extrato de utilização do cartão de crédito nos últimos 12 meses) ( $Y = 1$ , se cliente ativo ou não abandono;  $Y = 0$ , se cliente inativo, abandono ou *churn*). Esse critério temporal é semelhante ao usado por RADCLIFTE e SIMPSON (2008).

Um total de 16 variáveis explicativas (14 características individuais e duas de comportamento de uso do cartão) foram usadas no modelo, como mostra o Quadro 1. As variáveis inicialmente usadas foram aquelas disponíveis no cadastro dos clientes.

Para a validação do modelo foram usados o teste de Kolmogorov-Smirnov (KS) e a curva ROC (*Receiver Operating Characteristic*). O teste de Kolmogorov-Smirnov (KS) é uma estatística não paramétrica para testar se as funções de distribuição de probabilidades de dois grupos são iguais. O valor do KS do modelo é a maior diferença entre as distribuições acumuladas das probabilidades dos grupos de clientes ativos e não ativos para o uso do cartão de crédito. O valor da estatística pode variar entre 0 a 1, sendo que

quanto mais próximo de 1, maior o poder discriminatório do modelo ajustado (CONOVER, 1999). Sendo  $S(x)$  e  $F(x)$  a função de distribuição acumulada empírica dos clientes que não abandonaram e dos que abandonaram, respectivamente, então a estatística de Kolmogorov-Smirnov é dada por:  $KS = \max |S(x) - F(x)|$ . Portanto, a estatística KS é obtida pela distância máxima entre essas duas proporções acumuladas ao longo dos escores (probabilidades) obtidos pelos modelos. A curva ROC é uma métrica para avaliação de modelos, que permite estudar a variação para as medidas de sensibilidade e especificidade do modelo, para diferentes valores de ponto de corte (FAWCETT, 2003). Todos os dados foram analisados no software SAS versão 9.1. A seguir, são apresentados os principais resultados de calibração e validação do modelo.

## RESULTADOS

### Amostra de calibração

Na amostra de calibração do modelo, 55,5% dos clientes eram inativos (abandonaram o relacionamento nos últimos 12 meses do período de coleta de dados) e 44,5% ativos. Essa amostra apresenta as seguintes características: 68,6% dos clientes é do sexo feminino, 77,0%

não possuem outro tipo de cartão de crédito que não o usado pela rede varejista, 51,9% são solteiros e 34,9% casados; 94,9% residem no Estado do Rio de Janeiro, 63,6% possuem casa própria, 70,7% não informam o grau de escolaridade (dos 29,3% que informaram, 24,3% concluíram o segundo grau e 5% possuem curso superior completo). A idade média da amostra é de 33,7 anos e a renda média R\$ 2.533,00. Os resultados obtidos pelo modelo de Regressão Logística, após a verificação de não ocorrência de multicolinearidade e heteroscedasticidade, são apresentados na Tabela 1.

Os coeficientes da Tabela 1 são interpretados da seguinte forma: quanto maior sua magnitude, maior a probabilidade de o cliente ser ativo (ou menor a probabilidade de o cliente abandonar o relacionamento). Para obtenção do valor de probabilidade de o cliente ser ativo, atribui-se o valor individual para cada variável do modelo final de acordo com a equação 1:

$$\hat{\pi} = \frac{1}{1 + \exp\{-B\}} \quad (1)$$

Em que  $B = 5,742 - 0,318 \cdot X1 + 0,136 \cdot X2 + 1,013 \cdot X4 + 0,392 \cdot X5 - 0,074 \cdot X7 - 0,075 \cdot X8 - 0,326 \cdot X10 + 0,800 \cdot X12 - 0,182 \cdot X13 - 0,111 \cdot X14 - 1,591 \cdot X15 + 0,005 \cdot X16$ .

Quadro 1 – Variáveis utilizadas no modelo

VARIÁVEL	DESCRIÇÃO	MEDIÇÃO
X1	Sexo	1 = masculino; 0 = feminino
X2	Estado civil	1 = solteiro; 0 = outro
X3	Idade	Em anos
X4	Escolaridade	1 = curso superior; 0 = caso contrário
X5	Unidade da Federação	1 = RJ; 0 = outro
X6	Telefone residencial	1 = possui; 0 = não possui
X7	Tipo de residência	1 = própria; 0 = não própria
X8	Mora e trabalha na mesma cidade	1 = sim; 0 = não
X9	Mora e trabalha no mesmo estado	1 = sim; 0 = não
X10	Endereço de correspondência igual ao residencial	1 = sim; 0 = não
X11	Possui outro cartão de crédito	1 = sim; 0 = não
X12	Referência pessoal foi preenchida	1 = sim; 0 = não
X13	Número de cartões de crédito adicionais	Número de cartões
X14	Tempo de emprego do cliente	1 = até 6 meses; 0 = acima de 6 meses
X15	Extratos não pagos entre os meses 1 e 6	Número de extratos
X16	Valor médio dos extratos pagos entre os meses 1 e 6	Valor em R\$



O teste de Wald foi aqui usado por ser mais acurado em amostras grandes (KANZO, 2004). Os valores de  $p$  excedem 0,05 para as seguintes variáveis (não sendo, portanto, significativas a 95% de probabilidade): X3 (Idade), X6 (Telefone residencial), X9 (Mora e trabalha no mesmo estado), X11 (Possui outro cartão de crédito). Tais variáveis, portanto, foram excluídas do modelo final.

Observando os coeficientes das variáveis que representam as características dos clientes na Tabela 1, conclui-se que a probabilidade de ser um cliente ativo (ou de não abandonar o cartão de crédito) diminui se o cliente for do sexo masculino (coeficiente negativo significa menor probabilidade de ser um cliente ativo). Outras interpretações são:

- se o cliente é solteiro (X2), sua chance de ser um cliente ativo do cartão aumenta;
- se o cliente possui curso superior (X4), sua chance de ser um cliente ativo do cartão aumenta (essa informação deve ser interpretada com cuidado devido ao grande número de *missing values*, visto que apenas 29,3% dos clientes informaram a escolaridade);

- se o cliente mora no estado do Rio de Janeiro (X5), sede da rede varejista, sua chance de ser um cliente ativo do cartão aumenta;
- se o cliente preenche a referência pessoal (X12), sua chance de ser um cliente ativo do cartão aumenta;
- quanto mais cartões adicionais o cliente possui (X13), menor a chance de ser um cliente ativo do cartão;
- se o cliente possui menos que seis meses no mesmo emprego (X14), sua chance de ser um cliente ativo do cartão diminui.

As variáveis de comportamento do cliente refletem o uso dos cartões e, conseqüentemente, o relacionamento com a empresa. A interpretação de seus coeficientes é:

- quanto maior a quantidade de extratos não pagos (X15), menor a chance de ser um cliente ativo do cartão;
- quanto maior o valor médio dos extratos pagos (X16), maior a chance de ser um cliente ativo do cartão.

O passo seguinte foi validar o modelo obtido na amostra de calibração, usando a segunda amostra, a de validação (30.000 clientes).

Tabela 1 – Coeficientes das variáveis utilizadas no modelo

VAR.	DESCRIÇÃO	COEF.	ERRO PADRÃO	WALD	VALOR DE P
Intercep.		+ 5,742**	0,519	122,2	p<0,01
X1	Sexo	- 0,318**	0,025	158,0	p<0,01
X2	Estado civil	+ 0,136**	0,038	12,7	p<0,01
X3	Idade	+ 0,001	0,001	1,6	p=n.s.
X4	Escolaridade	+ 1,013**	0,049	435,9	p<0,01
X5	Unidade da Federação	+ 0,392*	0,175	5,0	p<0,05
X6	Telefone residencial	- 0,017	0,034	0,3	p=n.s.
X7	Tipo de residência	- 0,074*	0,029	6,5	p<0,05
X8	Mora e trabalha na mesma cidade	- 0,075**	0,024	9,8	p<0,01
X9	Mora e trabalha no mesmo estado	- 0,059	0,065	0,8	p=n.s.
X10	Endereço de correspondência igual ao residencial	- 0,326**	0,103	10,0	p<0,01
X11	Possui outro cartão de crédito	+ 0,015	0,028	0,283	p=n.s.
X12	Referência pessoal foi preenchida	+ 0,800**	0,079	101,4	p<0,01
X13	Número de cartões de crédito adicionais	- 0,182**	0,033	31,3	p<0,01
X14	Tempo de emprego do cliente	- 0,111**	0,031	13,0	p<0,01
X15	Extratos não pagos entre os meses 1 e 6	- 1,591**	0,021	6.012,7	p<0,01
X16	Valor médio dos extratos pagos entre os meses 1 e 6	+ 0,005**	0,001	11,4	p<0,01

\*Significativo a, no mínimo, 95% de probabilidade; \*\*Significativo a, no mínimo, 99% de probabilidade; n.s.= não significativo.



### Amostra de validação

O teste de Kolmogorov-Smirnov (KS) é uma medida de avaliação de desempenho do modelo e mede a sua capacidade de distinguir os clientes ativos dos inativos. Seu valor representa a diferença máxima entre as distribuições acumuladas dos clientes ativos e inativos (a linha tracejada no Gráfico 1 representa essa diferença). Pelo Gráfico 1 observa-se que, na amostra de validação, a diferença máxima é de cerca de 0,73. Segundo Corrêa e Vellasco (2008), uma diferença máxima de KS acima de 0,5 indica uma excelente discriminação entre grupos.

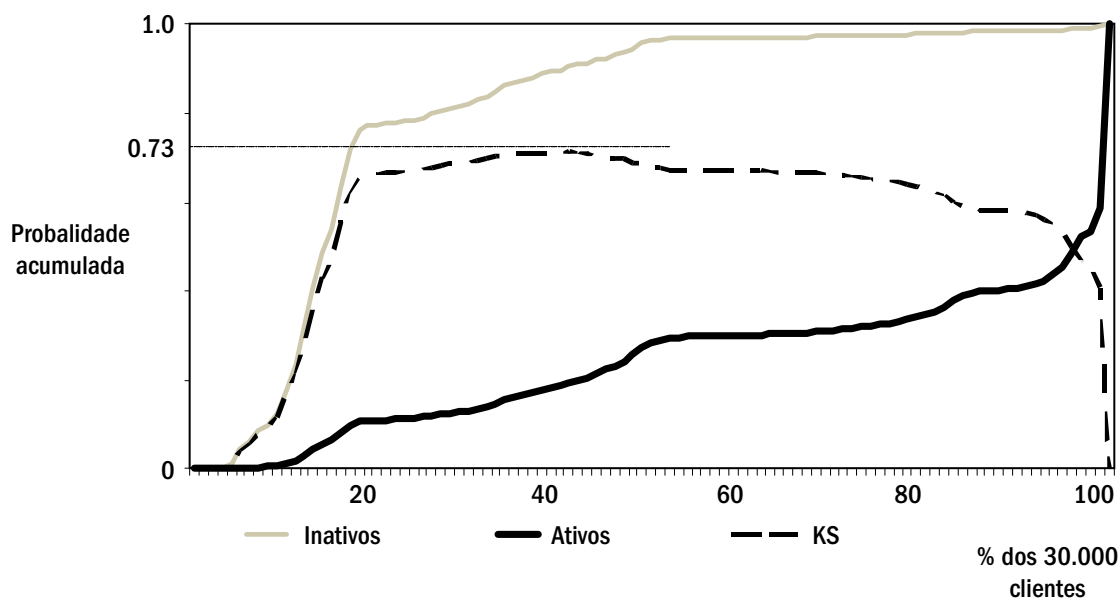
A Tabela 2 apresenta os clientes classificados por probabilidade de não abandono (ou de serem ativos). Observa-se que 47,9% estão classificados em uma faixa de 0 a 20% de probabilidade de serem ativos (alta probabilidade de serem inativos) e que 30,1% estão classificados em uma faixa de 80 a 100% de probabilidade de serem ativos (alta probabilidade de serem ativos). Isto significa que clientes ativos obtiveram altos escores pelo modelo, enquanto os inativos obtiveram baixos escores, o que demonstra a boa adequação do modelo à amostra de validação. Para obtenção do escore de cada cliente é necessário atribuir o valor para cada uma das variáveis do modelo da regressão logística de acordo com a equação 1.

A Tabela 3 apresenta a quantidade de clientes que a

empresa perde, obtida pelo modelo, por faixa de probabilidade  $P$ , dentro dos 30.000 clientes da amostra de validação, o que representa o risco de evasão por faixa de escore. Verifica-se que, quanto menor o escore, maior o risco de evasão, o que mostra o grande potencial do modelo na demonstração do risco dos clientes. Pela tabela observa-se a probabilidade média de o cliente abandonar o cartão de crédito por faixa de probabilidade ( $P$ ) de os clientes serem ativos. Por exemplo, os clientes classificados na faixa de  $0 \leq P < 10$  têm uma probabilidade média de 94,3% de abandonarem a empresa. Já para os clientes na faixa de  $90 \leq P < 100$ , a probabilidade de abandono é de 3,2%.

Os dados da Tabela 3 são usados para construir a Tabela 4 e, conseqüentemente, o Gráfico 2, que apresenta a probabilidade de abandono por ponto de corte (PC). Com intuito de manter clientes ativos na carteira de clientes, o gestor pode empregar ações mercadológicas para reduzir o abandono de um grupo específico de clientes que apresentam alta probabilidade de abandono, mesmo porque essas ações são de custo menor do que ações dirigidas a todos os clientes do banco de dados. O PC pode ser usado como um delimitador para apoiar na decisão de quais clientes comporão esse grupo específico e é definido pelo pesquisador e/ou gestor. Um  $PC = 30$ , por exemplo, significa que todos os clientes com proba-

Gráfico 1 – Distribuições acumuladas de probabilidade dos clientes ativos e inativos e a curva de Kolmogorov-Smirnov (KS)



bilidade (P) menor do que 30 de serem ativos poderiam ser tratados de forma diferenciada por possuírem maior probabilidade de abandono. Segundo a Tabela 4, o percentual de clientes que abandonam a empresa, definido pelo modelo, é de 55,4%, e estabelecendo-se um PC = 30, a probabilidade de abandono na amostra de validação cai para 20,3% (2.951 dos 14.547 clientes). Nesse caso, todos os clientes pontuados com escore menor do que 30 poderiam receber uma ação específica de marketing direto, por exemplo, na tentativa de reduzir sua probabilidade de abandono (em vez de essa ação ser destinada a toda a carteira de clientes da empresa).

O conjunto das características “especificidade” e “sensibilidade” indicam o grau de confiança do modelo quanto à sua capacidade de previsão em outra amostra de dados (aqui chamada de *amostra de validação*). Essa capacidade pode ser observada pelo Gráfico 3 como (curva ROC), que demonstra como o modelo se comporta com relação aos acertos e erros das classes previstas (ativo e inativo) em relação às observações reais. Esse comportamento também pode ser observado pelas matrizes de confusão: nos casos de classificação em que a variável dependente (Y) é binária, como neste artigo, a Matriz de Confusão (Quadro 2) apresenta a quantidade de casos

Tabela 2 – Quantidade e percentual de clientes por faixa de probabilidade de não abandono

PROBABILIDADE (P) DE NÃO ABANDONO (%)	QUANTIDADE DE CLIENTES	% DE CLIENTES
$0 \leq P < 10$	3.104	10,3
$10 \leq P < 20$	11.275	37,6
$20 \leq P < 30$	1.074	3,6
$30 \leq P < 40$	1.997	6,7
$40 \leq P < 50$	2.391	8,0
$50 \leq P < 60$	328	1,1
$60 \leq P < 70$	177	0,6
$70 \leq P < 80$	607	2,0
$80 \leq P < 90$	848	2,8
$90 \leq P \leq 100$	8.199	27,3
<b>Total</b>	<b>30.000</b>	<b>100,0</b>

Tabela 3 – Quantidade de clientes obtida pelo modelo que (não) abandonam a empresa na amostra de validação

PROBABILIDADE (P) DE NÃO ABANDONO (%)	NO. DE CLIENTES CONSIDERADOS	PROBABILIDADE MÉDIA DE ABANDONO (%)	NO. DE CLIENTES QUE ABANDONAM (PREVISTO)
$0 \leq P < 10$	3.104	94,3	2.927
$10 \leq P < 20$	11.275	88,4	9.967
$20 \leq P < 30$	1.074	73,4	788
$30 \leq P < 40$	1.997	63,1	1.260
$40 \leq P < 50$	2.391	45,7	1.093
$50 \leq P < 60$	328	37,5	123
$60 \leq P < 70$	177	22,0	39
$70 \leq P < 80$	607	13,0	79
$80 \leq P < 90$	848	11,2	95
$90 \leq P \leq 100$	8.199	3,2	262
<b>Total</b>	<b>30.000 (100,0%)</b>	<b>55,4</b>	<b>16.633 (55,4%)</b>

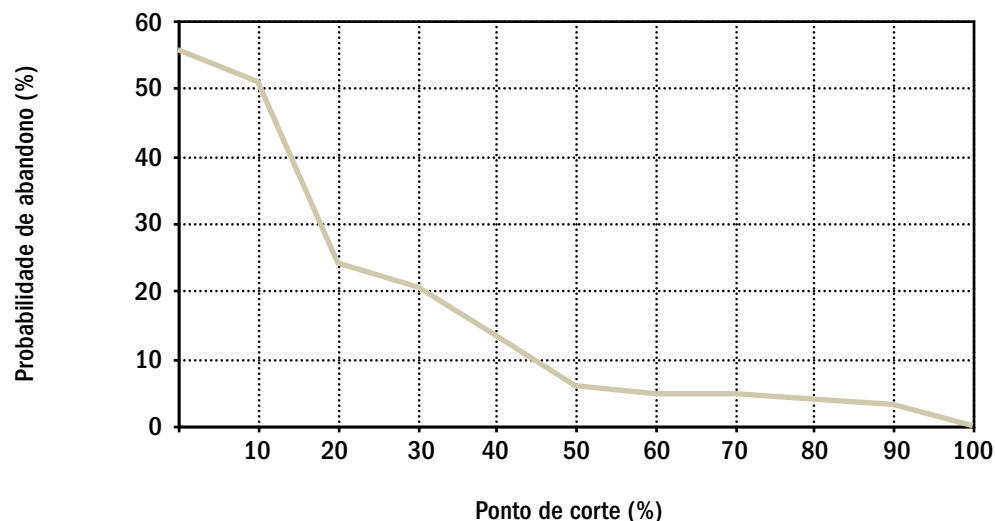
previstos como sendo da classe 0 e que realmente são da classe 0 (chamados de *verdadeiro negativo*), a quantidade de casos previstos como sendo da classe 0, mas que são na verdade da classe 1 (*falso negativo*), a quantidade de casos previstos como sendo da classe 1 e que realmente são da classe 1 (*verdadeiro positivo*) e os casos previstos como sendo da classe 1, mas que na verdade são da classe 0 (*falso positivo*). Sabendo que nessa pesquisa  $Y = 1$ , se cliente ativo, e  $Y = 0$ , caso contrário, a curva ROC é obtida tendo no seu eixo das ordenadas a “sensibilidade”

(taxa de clientes ativos que foram classificados como ativos pelo modelo, ou seja, a quantidade de clientes que foram classificados de forma correta pelo modelo, que é dada pela quantidade de verdadeiros positivos sobre a quantidade de positivos reais) e no eixo das abscissas o complemento da especificidade, ou  $1 - \text{especificidade}$  (“especificidade” é a taxa de clientes inativos que foram classificados como ativos pelo modelo, ou seja, a quantidade de clientes que foram classificados de forma incorreta pelo modelo, que é dada pela quantidade de

Tabela 4 – Probabilidade de abandono por ponto de corte

PONTO DE CORTE	PROBABILIDADE DE ABANDONO (%)	NO. DE CLIENTES CONSIDERADOS	NO. DE CLIENTES QUE ABANDONAM (PREVISTO)
0	55,4	30.000	16.633
10	51,0	26.896	13.706
20	23,9	15.621	3.739
30	20,3	14.547	2.951
40	13,5	12.550	1.691
50	5,9	10.159	598
60	4,8	9.831	475
70	4,5	9.654	436
80	3,9	9.047	357
90	3,2	8.199	262
100	-	0	0

Gráfico 2 – Probabilidade de abandono por ponto de corte



verdadeiros negativos sobre a quantidade de negativos reais) (GÖNEN, 2007).

A curva ROC é obtida calculando-se para cada ponto de corte (representados na diagonal principal) o valor da especificidade e sensibilidade. Essa curva confirma a boa capacidade de previsão do modelo (altas sensibilidade e especificidade), já que quanto maior a área entre a curva ROC e a diagonal principal, melhor o desempenho do modelo (GÖNEN, 2007).

### CONSIDERAÇÕES FINAIS

Considerando que as perdas de lucro devido ao abandono de clientes podem ser significativas, a previsão de *churn*

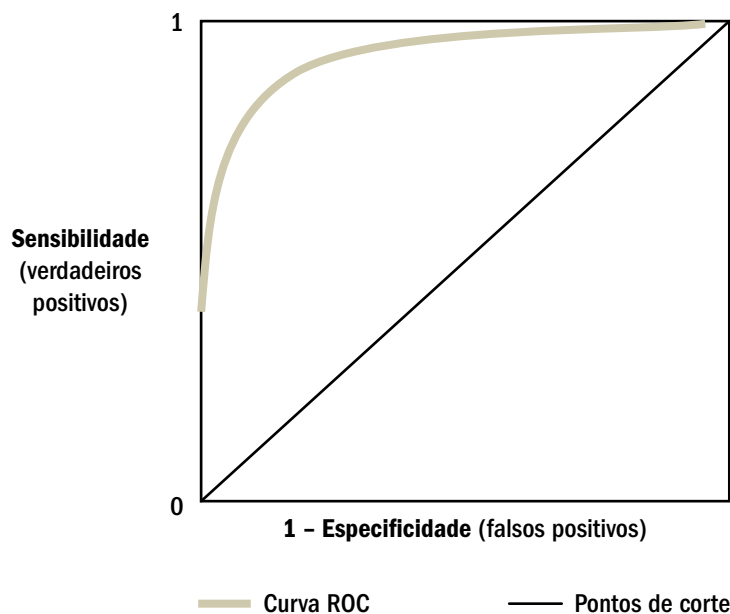
(abandono) representa informação relevante ao gestor, visto que clientes em perigo (altas probabilidades de *churn*) poderiam ser alvo de ações específicas de marketing, visando a sua retenção. Este artigo apresentou um modelo de estimação de probabilidades de abandono de determinado produto de uma organização em uma amostra de clientes, utilizando dados do histórico de relacionamento cliente/empresa, validou o modelo em outra amostra (amostra de validação) e descreveu as possíveis variáveis que influenciam seu abandono/permanência.

Aplicou-se o modelo de regressão logística em uma amostra de calibração de 70.000 usuários de cartão de crédito próprio de uma grande rede varejista e o validou em uma segunda amostra de 30.000 clientes. As possíveis variáveis (dentre aquelas disponíveis para análise neste

Quadro 2 - Matriz de confusão para um modelo de classificação

		PREVISTO PELO MODELO		
		0	1	
SITUAÇÃO REAL	0	Verdadeiro Negativo	Falso Positivo	<i>Negativo Real</i>
	1	Falso Negativo	Verdadeiro Positivo	<i>Positivo Real</i>
		<i>Negativo Previsto</i>	<i>Positivo Previsto</i>	

Gráfico 3 - Curva ROC



artigo) que influenciaram o abandono/permanência do cliente foram: sexo, estado civil, escolaridade, unidade da federação, tipo de residência, moradia e trabalho na mesma cidade, endereço de correspondência igual ao residencial, referência pessoal preenchida, número de cartões de crédito adicionais, tempo de emprego, extratos não pagos nos seis primeiros meses de análise, valor médio dos extratos pagos nos seis primeiros meses de análise. O modelo apresentou boa capacidade de previsão de abandono/permanência dos usuários/clientes na amostra de validação: a diferença máxima de 0,73 na curva de Kolmogorov-Smirnov (KS) indicou uma excelente discriminação entre grupos (clientes ativos e inativos) e a área abaixo da curva ROC confirmou a boa capacidade de previsão do modelo (altas sensibilidade e especificidade).

Os resultados deste estudo podem oferecer *insights* para que as empresas compreendam a taxa de abandono de seus clientes, relacionada à taxa de *turnover* da base de clientes. De posse dessa compreensão, podem-se personalizar ações de comunicação de marketing e tratamento diferenciado do segmento em risco de abandono, otimizando tempo e esforços mercadológicos. Se as empresas puderem identificar com mais precisão os potenciais *churners*, elas podem ser capazes de traduzir tal precisão em lucros reais, por meio de incentivos customizados, ainda mais se a previsão de *churn* for integrada ao *Business Intelligence*, incluindo gestão e análise eficiente de dados e decisões de marketing adequadas a tais *churners*. Obviamente, esses lucros não necessariamente precisam ser financeiros, como no caso de organizações não lucrativas que se preocupam com a perda do relacionamento de seus clientes. Os resultados aqui alcançados não são diretamente aplicáveis em qualquer organização, mas a forma como os dados foram operacionalizados no modelo aqui proposto pode ser útil para as áreas de interface entre tecnologia da informação e marketing de um grande número de organizações.

Alguns autores, entre eles Lovelock e Wirtz (2010), recomendam a construção da fidelidade dos clientes basicamente por: i) atingir o segmento desejado, assegurando elevados níveis de satisfação; ii) desenvolver vínculos estreitos com os clientes, visando a aprofundar o relacionamento (via *cross-selling*, recompensas etc.); iii) diagnosticar e eliminar os fatores que resultam em *churn*. Este artigo, portanto, contribui com *insights* para a literatura de lealdade na academia de marketing, bem como com a literatura da administração do relacionamento com o cliente.

Este artigo apresenta limitações que podem ser supe-

radas em trabalhos futuros no Brasil. Apenas um modelo foi proposto e validado. Outras abordagens poderiam ser propostas e comparadas, usando os métodos de validação aqui usados. Por exemplo, uma abordagem que utilize Modelos Aditivos Generalizados (GAM - *Generalized Additive Models*) relaxa a restrição de linearidade da regressão logística aqui usada, o que permite um ajuste não linear aos dados. Além disto, as variáveis usadas neste artigo restringiram-se àquelas disponíveis no banco de dados. O uso de outras variáveis explicativas pode ser testado em futuros estudos, como perfis de uso (volume, frequência de uso e tipo de produto dentre os oferecidos pelas administradoras de cartões de crédito), satisfação do cliente, obtida via *surveys*, ou informações sobre características do ponto de venda, como tamanho da loja, posicionamento e tipo de segmento de mercado em que atua. As interações pontuais no relacionamento, como contato via telemarketing, processos de autorizações de vendas ou pagamentos do extrato registradas na base de dados poderiam ser incorporadas em modelos futuros.

Um pressuposto deste artigo é que o atrito ocorre quando o cliente abandona a empresa para trocar de fornecedor. Mas, em alguns casos, o cliente pode abandonar porque deixa de adotar uma tecnologia ou um serviço (por exemplo, um cliente pode abandonar os serviços de uma locadora de carros não porque trocou de provedor, mas porque comprou um automóvel e não precisa mais do serviço). Tal diferenciação pode ser tratada em estudos futuros, o que seria inédito na literatura de marketing.

O produto aqui estudado é de característica contratual (cartão de crédito) e não se levou em conta efeitos intra-indivíduos (*cross-selling*) ou inter-indivíduos (*word-of-mouth*). Produtos não contratuais são mais usuais no mercado e tais efeitos podem ser investigados em estudos futuros por meio de um banco de dados mais abrangente. Além do mais, medidas estatísticas que considerem o lucro podem ser desejáveis, pois *churners* pouco lucrativos podem até ser interessantes para a empresa, sob a perspectiva financeira apenas.

A aplicação do modelo aqui apresentado não se dá isoladamente. A integração com modelos de CLV ou RFM, que possa oferecer maior adequação ao tipo de relacionamento cliente/empresa, e a definição de outras variáveis, como a aquisição de produtos financeiros no cartão ou o tipo de mercadoria comprada com o cartão, podem ser aplicadas em outras pesquisas e situações, para averiguar mais precisamente a retenção/abandono e rentabilidade dos clientes. Assim, o próprio conceito de *churn* pode ser revisto, como no caso de um cliente cujo CLV é de-

crecente no tempo (GLADY e outros, 2009) e que não necessariamente abandone a empresa.

Por fim, os dados secundários coletados longitudinalmente foram aqui tratados como em uma seção cruzada, logo outra possibilidade de extensão deste estudo seria o acompanhamento das variáveis longitudinalmente, com o objetivo de perceber a consistência dos resultados e as mudanças ocorridas ao longo do tempo, mais condizente com o conceito de CRM.

## REFERÊNCIAS

- ACCENTURE. The new customer imperative: retaining and acquiring customers in a changed banking landscape. *The Point*, v. 9, n. 4, p. 1-8, 2009.
- ADDRIANS, P; ZANTINGE, D. *Data mining*. London: Addison-Wesley, 1996.
- ANÇA, E. S; VIEIRA, L. M. M; PETROLL, M. M; WAGNER, R. P; COSTA, R. S. As comunidades virtuais e a segmentação de mercado: uma abordagem exploratória, utilizando redes neurais e dados da comunidade virtual Orkut. *Revista de Administração Contemporânea*, v.12, Edição Especial, p. 41-63, 2008.
- ASSOCIAÇÃO BRASILEIRA DE CARTÕES DE CRÉDITO. *Pagamento eletrônico alavanca o varejo*. Disponível em: <http://www.abecs.org.br>. Acesso em 24.04.2009.
- AU, W; CHAN, C. C; YAO, X. A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Transactions on Evolutionary Computation*, v. 7, n. 6, p. 532-545, 2003.
- BAUER, C; MIGLAUTSCH, J. A conceptual definition of direct marketing. *Journal of Direct Marketing*, v. 6, n. 2, p. 7-17, 1992.
- BERRY, M. J. A; LINOFF, G. *Data mining techniques: for marketing, sales, and customer relationship management*. New York: John Wiley & Sons, 2004.
- BOSE, I; CHEN, X. Hybrid models using unsupervised clustering for prediction of customer churn. *Journal of Organizational Computing & Electronic Commerce*, v. 19, n. 2, p. 133-151, 2009.
- BRETZKE, M. *Marketing de relacionamento e competição em tempo real*. São Paulo: Atlas, 2000.
- CISTER, A. M; SHEHATA, L. D. Definição das características do perfil do consumidor com a utilização de redes neurais artificiais. In: ENCONTRO NACIONAL DA ASSOCIAÇÃO NACIONAL DOS PROGRAMAS DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO, 32, 2008, Rio de Janeiro. *Anais*. Rio de Janeiro: ANPAD, 2008.
- CORRÊA, M. F; VELLASCO, M. Análise de risco de crédito em correspondentes bancários através de redes neurais. *Revista Inteligência Computacional Aplicada*, v. 1, n. 1, p. 23-37, 2008.
- DATAFOLHA. *Mercado de meios de pagamento eletrônico*. Disponível em: <http://www.abecs.org.br/cmep.asp>. Acesso em 14.2.2010.
- FADER, P. S; HARDIE, B. G. S; LEE, K. L. RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research*, v. 42, n. 4, p. 415-30, 2005.
- FERREIRA, J. B; SILVA, J. F; VELLASCO, M; PACHECO, M. A. Mining and understanding wireless churn. In: ENCONTRO NACIONAL DA ASSOCIAÇÃO NACIONAL DOS PROGRAMAS DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO, 28, 2004, Curitiba. *Anais*. Curitiba: ANPAD, 2004.
- GLADY, N; BAESENS, B; CROUX, C. Modeling churn using customer lifetime value. *European Journal of Operational Research*, v. 197, n. 1, p. 402-411, 2009.
- GÖNEN, M. *Analyzing Receiver Operating Characteristic curves with SAS* (Sas Press Series). Cary, NC: SAS Publishing, 2007.
- GREENBERG, P. *CRM Customer Relationship Management na velocidade da luz: conquista e lealdade de clientes em tempo real na Internet*. Rio de Janeiro: Campus, 2001.
- GUNNARSSON, C. L; WALKER, M. M; WALATKA, V; SWANN, K. Lessons learned: a case study using data mining in the newspaper industry. *Journal of Database Marketing & Customer Strategy Management*, v. 14, n. 4, p. 271-280, 2007.
- GUSTAFSSON, A; JOHNSON, M. D; ROOS, I. The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention. *Journal of Marketing*, v. 69, n. 4, p. 210-218, 2005.
- JOHNSON, M. D; FORNELL, C. A framework for comparing customer satisfaction across individuals and product categories. *Journal of Economic Psychology*, v. 12, n. 2, p. 267-286, 1991.
- KAKWANI, N; NERI, M. C; SON, H. Linkages between pro-poor growth, social programs and labor market: the recent Brazilian experience. *World Development*, v. 38, n. 6, p. 881-884, 2010.
- KURTZ, D. L; CLOW, K. E. *Services marketing*. New York: John Wiley & Sons, 1998.
- LEJEUNE, M. A. P. M. Measuring the impact of data mining on churn management. *Internet Research*, v. 11, n. 5, p. 375-387, 2001.
- LIAO, T. F. *Interpreting probability models: logit, probit, and other generalized linear models*. Thousand Oaks: Sage, 1994. (Series in quantitative applications in the social sciences, n. 07-101).
- LOVELOCK, C; WIRTZ, J. *Services marketing: people, technology, strategy*. 7th. ed. Upper Saddle River, NJ: Prentice Hall, 2010.

## ARTIGOS • MODELAGEM DE PROBABILIDADE DE CHURN

MADDALA, G. S. *Limited dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press, 1983.

MOZER, M. C.; WOLNIEWICZ, R.; GRIMES, D. B.; JOHNSON, E.; KAUSHANSKY, H. Churn reduction in the wireless industry. In: SOLLA S. A.; LEEN T. K.; MUELLER K. R. (Eds) *Advances in neural information processing systems*. Cambridge, MA: MIT Press, 2000. p. 935-941.

NESLIN, S. A.; GUPTA, S.; KAMAKURA, W.; LU, J.; MASON, C. H. Defection detection: measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research*, v. 43, n. 2, p. 204-211, 2006.

OLIVER, R. L. Whence consumer loyalty? *Journal of Marketing*, v. 63, Edição Especial, p. 33-44, 1999.

PEPPERS, D.; ROGERS, M. *Marketing 1 to 1: um guia executivo para entender e implantar estratégias de Customer Relationship Management (CRM Series)*. 2. ed. São Paulo: Makron Books, 2001.

PETRISON, L. A.; BLATTBERG, R. C.; WANG, P. Database marketing: past, present and future. *Journal of Direct Marketing*, v. 11, n. 4, p. 109-125, 1997.

RADCLIFTE, N. J.; SIMPSON, R. Identifying who can be saved and who will be driven away by retention activity. *Journal of Telecommunications Management*, v. 1, n. 2, p. 168-176, 2008.

RHEE, S.; MCINTYRE, S. Including the effects of prior and recent contact effort in a customer scoring model for database marketing. *Journal of the Academy of Marketing Science*, v. 36, n. 4, p. 538-551, 2008.

SHANI, D.; CHALASANI, S. Exploiting niches using relationship marketing. *Journal of Consumer Marketing*, v. 9, n. 3, p. 33-42, 1992.

SINGH, R. P.; SINGH, L. O. Blogs: emerging knowledge management tools for entrepreneurs to enhance marketing efforts. *Journal of Internet Commerce*, v. 7, n. 4, 2008.

SINHA, A.; MAY, J. H. Evaluating and timing predictive data mining models using receiver operating characteristic curves. *Journal of Management Information Systems*, v. 21, n. 3, p. 249-280, 2005.

STONE, M.; SHAW, R. Database marketing for competitive advantage. *Long Range Planning*, v. 20, n. 2, p. 12-20, 1987.

VAVRA, T. G. *Marketing de relacionamento*. São Paulo: Atlas, 1994.