

Regressão Logística: Um Modelo de Risco de Cancelamento de Clientes

Autoria: Karine de Almeida Karam, Jorge Ferreira da Silva, Flávia de Holanda Schmidt

RESUMO

O objetivo deste trabalho é desenvolver um modelo estatístico que relacione variáveis transacionais, demográficas e dados sobre o histórico de eventos com a probabilidade de cancelamento dos clientes assinantes de jornal, assim como definir o perfil dos clientes com maior risco de desligamento. Após a construção de um modelo teórico, a regressão logística foi utilizada como técnica estatística para desenvolver um modelo de previsão de cancelamento. Os resultados indicaram que o perfil do cliente que cancela a assinatura do jornal é o jovem de até 30 anos; com baixo nível sócio-demográfico; morador da baixada, subúrbio e outros estados que não o Rio de Janeiro; que tenha adquirido sua assinatura através do canal tele-marketing ativo; com a assinatura da modalidade anual e forma de pagamento em boleto ou débito em conta corrente; clientes que adquiriram sua assinatura mais recentemente; que comprem menos de 3 produtos da empresa e que não tenham feito reclamações através da central de atendimento. O modelo final de previsão de cancelamento contou com 11 variáveis e a tabela de classificação mostrou uma taxa de acerto geral de 75,3%.

INTRODUÇÃO

O transcorrer das últimas décadas, profundamente marcado pelo aumento da oferta de mercado, crescimento da competitividade entre empresas e alteração do comportamento do consumidor, impôs mudanças de estratégias aos gestores das organizações contemporâneas: mantidas as necessidades de crescimento e de lucratividade, intensificou-se a percepção de que cada cliente, de forma individual, pode ter seu valor maximizado de acordo com o seu potencial de negócios para com determinada organização.

De forma reflexa, acadêmicos e executivos de marketing vêm dedicando crescentes esforços para gerir o cliente como um ativo para a organização e lidar com os desafios de uma arena cada vez mais competitiva: a capacidade de captar e conservar clientes é provavelmente a chave do sucesso e um dos maiores desafios das organizações contemporâneas (KOTLER, 2002).

Dentro desta perspectiva, observou-se a evolução do marketing de um paradigma meramente transacional para o paradigma relacional: as empresas, cada vez mais, procuram desenvolver relacionamentos com os clientes (SINGH E SIRDESHMUKH, 2000). Dessa forma, o marketing de relacionamento, conjugado à estratégia de retenção de clientes, torna-se uma alternativa atraente. Dependendo do setor de atuação, as empresas podem aumentar seus lucros de 25% a 85%, ao reduzir em apenas 5% a perda de clientes (REICHHELD & SASSER JR., 1990).

A taxa de retenção de clientes é vista por Rust & Zahorik (1993) como o componente mais importante para que a empresa consolide sua participação no mercado, sendo direcionada pela satisfação do consumidor. Embora a satisfação não seja obrigatoriamente revertida em lealdade, as empresas devem persegui-la e reconhecer que ela é alcançada pela entrega consistente de valor superior ao cliente (REICHHELD, 1993). É imprescindível que as organizações, para a construção desse relacionamento, entendam profundamente os processos pelos quais os clientes mantêm trocas com as empresas, e como esses processos influenciam a lealdade

Os programas de relacionamento têm por objetivo aumentar o *life time value* dos clientes e, para alguns autores, inclusive, como Vavra (1993), quanto mais tempo se mantém um cliente, mais lucrativo ele poderá ser para o negócio, e quanto mais tempo o cliente comprar de uma mesma organização, mais dependente poderá tornar-se de seus produtos ou serviços e estará menos suscetível a ofertas da concorrência com preços mais baixos.

Na indústria jornalística, que retoma na atualidade um ciclo de crescimento após consecutivos declínios do IVC (Instituto Verificador de Circulação) entre 2001 e 2003, e especialmente no segmento *quality papers* (voltados para a classe AB), quase 80% das vendas são realizadas por meio de assinaturas, o que torna a retenção dos seus clientes atuais a melhor estratégia para manter a circulação.

Para diferenciar o grupo que cancela do que não cancela a assinatura do jornal, é preciso segmentar a base de clientes entre aqueles que possuem alto e baixo risco de abandono. A segmentação é baseada no conceito de que o mercado para um produto não é homogêneo em relação a seus desejos e necessidades. Isto explica por que uma parte do mercado vale mais que a outra e quais estratégias de marketing diferentes devem ser empregadas na venda e na retenção dos clientes dessas diferentes partes .

Faz-se necessário medir o risco de perda de cada cliente, calculando a probabilidade de cancelamento da assinatura, identificar os fatores que explicam o perfil do cliente que cancela e classificá-los em dois grupos: o dos que cancelam e o dos que não cancelam. Desta forma, este estudo desenvolve um modelo que relaciona variáveis transacionais, demográficas e dados sobre o histórico de eventos com a probabilidade de cancelamento do cliente.

Neste contexto, este trabalho pretende, usando a técnica da regressão logística binária, identificar os riscos de cancelamento na base de clientes, permitindo assim a construção e avaliação de regras de reconhecimento de padrões de cancelamento da assinatura para a classificação dos clientes de jornal em grupos. Ao identificar o perfil dos clientes com risco de cancelamento será possível atuar sobre eles através de ações de retenção, que revertam a perda desses clientes.

Para tanto, a questão central que norteia esta pesquisa é: *é possível definir o perfil do cliente que cancela uma assinatura de jornal?*

A seqüência do artigo está estruturada da seguinte forma: apresenta-se a revisão de literatura que deu origem às hipóteses de pesquisa, seguido de uma breve exposição dos aspectos metodológicos utilizados para alcançar esse objetivo. Por fim, apresentam-se os resultados encontrados nesta pesquisa, bem como suas possíveis implicações.

REFERENCIAL TEÓRICO

O erro mais crítico no desenvolvimento de modelos com base teórica é a omissão de uma ou mais variáveis independentes, um problema conhecido como erro de especificação (HAIR *et al.*, 1998). Ainda segundo o autor, o diagrama de definição das variáveis deve ser definido em termos de construtos, após o que devem ser levantadas as métricas que irão mensurar esses construtos. Assim, a estrutura deste trabalho contempla a elaboração de um modelo teórico *a priori*, para que, na etapa subsequente, sejam definidas as variáveis que devem compor o modelo matemático de previsão de cancelamento.

Para tanto, é fundamental revisar os diversos conceitos que se relacionam com a retenção de clientes, e, em seguida, definir as variáveis que influenciam o comportamento de compra do consumidor e, conseqüentemente, ajudam a descrever o grupo que cancela e o grupo que não cancela a assinatura do jornal.

O modelo teórico contará com três grupos de variáveis. O primeiro contempla as variáveis que avaliam a relação cliente-empresa com a retenção, são elas: lealdade, marketing de relacionamento e satisfação. O segundo é o grupo das variáveis transacionais que descrevem a interação operacional da empresa com o cliente. Por fim, contará com as variáveis relativas às características do cliente e que ajudam na segmentação do mercado consumidor, tais como: geográficas, demográficas, psicográficas e comportamentais.

Retenção

Diversos estudos indicam que reter consumidores é mais barato e competitivamente mais vantajoso do que atrair novos clientes (BERRY, 1995; PARVATIYAR e SHETH, 2000). Nash (1994) e Kotler e Armstrong (1995) afirmam que o custo para conservar um cliente atual satisfeito é menor do que a despesa para conquistar um novo. Vavra (1993) demonstra que a aquisição de novos clientes pode custar até 5 vezes mais do que o custo necessário para satisfazer e reter clientes já existentes.

Nesse sentido, outros estudos sugerem ainda que quanto maior o tempo que uma empresa retém seu cliente, maior é o lucro gerado por ele (REICHHELD, 1996). Reichheld e Sasser (1990) indicam que as empresas podem aumentar os lucros de 25% a 85%, reduzindo as fugas de clientes em apenas 5%. Um estudo feito por Oliver (1999) em 14 ramos de negócios diferentes mostrou que ao aumentar em 5% a retenção dos consumidores gera-se um incremento de 25% a 95% no valor presente líquido (VPL).

Retenção e Marketing de Relacionamento

Sheth e Parvatiyar (2000) argüem que o marketing de relacionamento é um processo de engajamento em atividades cooperativas e colaborativas e programas com consumidores finais para criar e intensificar valor econômico e redução de custos mútuos.

Alguns autores (KOTLER E ARMSTRONG, 1995; GRONROOS, 2003) enxergam o relacionamento com o cliente envolve um processo contínuo de interações, algumas vezes realizado pelo próprio vendedor e que resultam no desenvolvimento de laços mais fortes à medida que o gerenciamento do relacionamento é feito com esse propósito. Partindo-se do estágio inicial do relacionamento, ele normalmente progride de modo a alcançar um vínculo maior com o cliente, garantindo que suas necessidades sejam atendidas e que o cliente valorize esse relacionamento.

Retenção e Lealdade

Reichheld (1996) define que a lealdade de um cliente está na permanência dele com uma empresa durante muitos anos enquanto Gerpott, Rams e Schindler (2001) entendem lealdade como a atitude favorável de um cliente perante uma empresa e seus serviços. Segundo Lacoueille (1997) as primeiras definições de lealdade baseavam-se na observação do comportamento do consumidor fiel quando este comprava regularmente a mesma marca. Esta é uma abordagem comportamental ou behaviorista, que considera a lealdade uma resposta passiva baseada na análise do comportamento de compra.

Caruana (2004) também distingue a lealdade do cliente da retenção do cliente. A retenção do cliente, segundo ele, é a estratégia que a empresa implementa, enquanto que a lealdade do cliente é um estado psicológico que o cliente tem ou não tem.

Segundo Oliver (1999), a lealdade desenvolve-se em uma seqüência de fases, em que o consumidor, primeiramente, torna-se cognitivamente leal, depois passa a ser afetivamente leal, em seguida, conativamente leal, por fim, torna-se ativamente leal. O autor também relaciona a lealdade ao contexto sócio-emocional. Assim, Oliver propõe uma estrutura que relaciona o grau de determinismo e força pessoal e o suporte da sociedade/comunidade, podendo resultar em um estado de lealdade final (*ultimate loyalty*), quando o consumidor deseja ser leal e a sociedade também quer que ele seja leal. Esta definição de lealdade é utilizada como a definição teórica do construto neste estudo, privilegiando a concepção dinâmica de lealdade.

Retenção e Satisfação

Os clientes experimentam vários níveis de satisfação ou desapontamento após cada experiência de serviço à medida que suas expectativas são atingidas ou ultrapassadas (LOVELOCK e WRIGHT, 2001).

Yin (1990) conceitua satisfação como o julgamento atitudinal que se realiza após a compra ou após uma série de interações do consumidor com um produto. Oliver (1999) acrescenta que satisfação é um preenchimento prazeroso, ou seja, o consumidor sente que o consumo preenche alguma necessidade, desejo ou objetivo de forma prazerosa, gerando um desempenho favorável comparado a um padrão existente.

Como resultante da satisfação, estudos empíricos suportam a ligação da satisfação com a retenção de clientes (JONES, MOTHERSBAUGH e BEATTY, 2000) e, logo, com o desempenho organizacional (BANSAL e TAYLOR, 1999).

A satisfação do consumidor possui duas dimensões essenciais: a satisfação referente a uma transação específica e a satisfação acumulada (ROSSI E SLONGO, 1997). Há razoável consenso entre pesquisadores de marketing e de comportamento do consumidor concordam quanto ao fato da satisfação ser uma medida individual de uma transação específica ou como a avaliação de uma experiência particular com um produto ou serviço. Já a questão da definição da a qualidade percebida (ou desempenho) do serviço quanto a um antecedente da satisfação numa transação específica ainda permanece em questão. A visão de Johnson (1995) e Fornell (1996), que consideram a satisfação como um construto abstrato que descreve as experiências acumuladas de consumo com um produto ou serviço, ao longo do tempo, tem sido base de estudos mais recentes.

A Relação da Retenção com outros Conceitos

Para obter vantagem competitiva diante dos concorrentes, empresas tentam reter seus clientes através do aumento de sua satisfação e de sua lealdade (GERPOTT *et al.*, 2001). Kotler (2000) afirma que “a chave para a retenção de clientes é a satisfação”. Para Oliver (1999) a satisfação dos clientes não é suficiente para gerar lealdade, pois é apenas um determinante da lealdade, mas não é o único.

Zeithaml e Bitner (1996) sugerem que a retenção é uma das conseqüências da lealdade. Para Berry e Parasuraman (1995) a retenção de clientes é uma conseqüência desejável do marketing de relacionamento. Gerpott *et al.* (2001) sugerem que os conceitos de satisfação, lealdade e retenção estejam ligados em dois estágios: a satisfação influencia a lealdade e a lealdade influencia a retenção de clientes.

Rapp (1996) e Reichheld (1993) compartilham a opinião de que a satisfação do cliente não é um substituto para a retenção. Apesar de ser intuitivo que um aumento na satisfação do cliente terá como conseqüência um aumento na retenção e nos lucros, os fatos são controversos. Dos clientes que se desligam de uma empresa, entre 65% e 85% deles estavam satisfeitos ou muito satisfeitos com seus fornecedores. Para Reichheld (1993) os sistemas de medida de satisfação não foram desenvolvidos para oferecer um entendimento sobre quantos clientes permanecem leais à empresa e por quanto tempo.

A despeito das opiniões contraditórias sobre a relação entre lealdade e lucro, muitas empresas estão utilizando o marketing de relacionamento como ferramenta para aumentar a lealdade dos consumidores e garantir a retenção, considerando-o como instrumento essencial ao desenvolvimento de liderança, fidelidade do consumidor e rápida aceitação de novos produtos e serviços no mercado. A criação de relações sólidas e duradouras é uma tarefa árdua, de difícil manutenção. Mas acredita-se que, em um mundo onde o cliente tem tantas opções, mesmo em segmentos limitados, uma relação pessoal é a única forma de manter a fidelidade do cliente (MCKENNA, 1999).

Alguns autores afirmam que níveis maiores de satisfação do cliente levam a níveis maiores de lealdade do cliente (ANDERSON e SULLIVAN, 1993; OLIVER e SWAN, 1989; YI, 1991). Através da lealdade, alguns autores dizem que a satisfação ajuda a assegurar rendimentos futuros (FORNELL, 1992), reduzir custos de transações futuras (REICHHELD e SASSER, 1990),

diminuir a elasticidade de preços (ANDERSON, 1996) e minimizar a probabilidade de os clientes trocarem de fornecedor (ANDERSON e SULLIVAN, 1993).

A integração das teorias referentes à Satisfação do Consumidor, Marketing de Relacionamento e Lealdade é, portanto, relevante sob o ponto de vista teórico e um dos aspectos tratados neste artigo.

Segmentação do Mercado Consumidor

O mercado é composto por uma grande diversidade de clientes produtos e necessidades. Um papel importante do marketing é determinar quais segmentos oferecem maiores chances para a empresa alcançar seus objetivos.

Czinkota (2001) afirma que entender o consumidor na sua percepção individual, em suas decisões de compra, desde as questões culturais até as sociais, significa entender até que ponto estas variáveis interferem nas decisões do consumidor.

Diversas bases de segmentação já foram amplamente pesquisadas. Embora não se tenha encontrado consenso, a grande maioria destaca quatro tipos de segmentação: geográfica, sócio-demográfica, psicográfica e comportamental.

Variáveis	Definição
Variáveis Geográficas	Propõe dividir o mercado em unidades geográficas diferentes, como países, estados, regiões, cidades ou bairros.
Variáveis Demográficas	As variáveis demográficas como idade, tamanho da família, ciclo de vida da família, sexo, renda, ocupação, formação educacional, religião, raça, geração, nacionalidade ou classe social, são as bases mais comuns para se distinguir grupos de consumidores.
Variáveis Psicográficas	Compradores são divididos em grupos diferentes baseados no estilo de vida e na personalidade das pessoas
Variáveis Comportamentais	Variáveis como ocasiões, benefícios, status de usuário, taxa de uso, status de lealdade, estágios de aptidão de compra e atitude são variáveis importantes para dividir o mercado em grupos tomando como base o conhecimento do consumidor, sua atitude e o uso e a resposta para um produto.
Variáveis Transacionais	Os clientes não são iguais e expressam essa diversidade através de transações diferentes, que geram custos maiores ou menores para a empresa e impactam de maneira distinta na geração de caixa de uma companhia.

Fonte: Kotler, 1998

Quadro 1 –Variáveis de Segmentação

ASPECTOS METODOLÓGICOS

O universo da pesquisa é composto pelos clientes, ativos e inativos, registrados no banco de dados da empresa editora de jornal Infoglobo, que edita os jornais O Globo e Extra, é líder no Rio de Janeiro com 71% do *market share* carioca. O jornal O Globo é o principal jornal *quality paper* do Rio com 36% de *market share* de circulação e 81% das suas vendas são através de assinaturas. Um jornal como este sempre será alvo para o crescimento dos demais. A melhor estratégia para a manutenção da circulação do Globo é a retenção dos seus clientes atuais

A base de dados possui todas as assinaturas que estejam ativas, incluindo as vendidas desde o momento em que a empresa lançou o serviço de assinaturas em abril de 1976. Para as assinaturas

inativas existe registro das assinaturas canceladas desde julho de 1995. Foi definida uma janela de análise em que foram analisadas as assinaturas canceladas entre junho de 2004 e maio de 2005 e assinaturas ativas em maio de 2005. De um total de 230.858 assinaturas ativas e 105.524 assinaturas canceladas, extraiu-se duas amostras aleatórias simples, uma de treinamento com 35.549 casos e outra amostra para validação com 4.796 casos.

A coleta dos dados foi realizada da seguinte maneira:

- pesquisa bibliográfica em livros, dissertações, *papers* e periódicos, para conceituar marketing de relacionamento, satisfação, lealdade e retenção de clientes, variáveis transacionais, segmentação de mercado, variáveis geográficas, demográficas, psicográficas e comportamentais, além de métodos estatísticos para tratamento e modelagem dos dados do histórico de eventos dos assinantes;
- pesquisa documental na base de dados fornecida pela empresa editora de jornais observando três grupos de variáveis: histórico de eventos, dados transacionais e variáveis demográficas. A unidade de medida do tempo empregada é o mês que é uma variável contínua e a unidade de análise é o assinante pessoa física.

Tratamento dos Dados

A técnica utilizada para indicar o risco de cancelamento dos assinantes deste jornal é a regressão logística binária, que determina a probabilidade de uma assinatura pertencer ao grupo dos que cancelam ou ao grupo dos que não cancelam. Tanto a variável dependente quanto as variáveis independentes são categóricas. O tratamento estatístico dos dados e o desenvolvimento do modelo foram realizados com o apoio do SPSS 13.0.

O Método de Regressão Logística

Na literatura, existem duas técnicas com a capacidade de separar dois grupos ou alocar um novo elemento em um desses grupos. Ambas as técnicas, Análise Discriminante e Regressão Logística, se enquadram na classe de métodos estatísticos multivariados de dependência, pois relacionam um conjunto de variáveis independentes com uma variável dependente categórica (HAIR *et al.*, 1998).

O modelo logístico é mais comumente utilizado porque a análise discriminante impõe às variáveis independentes algumas condições, tais como serem normalmente distribuídas e terem suas matrizes de variância-covariância iguais entre os dois grupos de classificação. Outro ponto que é motivo de crítica quanto ao método da análise discriminante é que o resultado da expressão discriminante fornece um *score* que possui pouca interpretação intuitiva. Segundo Ohlson (1980), este *score* é basicamente um dispositivo (discriminatório) de classificação ordinal, não tendo embutido nenhum aspecto probabilístico. Além de não depender da exigência de normalidade das variáveis independentes e da igualdade de matrizes de covariância, a regressão logística é semelhante a uma regressão múltipla, pois possui o poder de incorporar efeitos não lineares (HAIR *et al.*, 1998). A possibilidade de interpretação direta dos coeficientes como medidas de associação é uma das grandes vantagens da regressão logística: a interpretação destes coeficientes pode ser estendida para qualquer problema prático.

Coleta e Análise dos Dados

A amostra inicial contava com 36.000 assinaturas. Após a análise da base de dados, foram excluídos os assinantes com as seguintes características:

- a. assinaturas com forma de pagamento “não informado”;
- b. assinaturas com tipo de assinatura “em branco”;

c. assinantes com idade inferior a 18 anos e superior a 90 anos.

Depois desse tratamento, a amostra ficou com 35.549 assinaturas, sendo 24.492 ativas e 11.057 canceladas. A base final para estudo contém as variáveis, sumariadas no Quadro 2: status (ativa ou inativa), tipo de assinatura, fonte de venda, forma de pagamento, região, tempo de permanência, LTV, SD&W, gênero, faixa etária, indicativo de participação em ações de fidelização, publicação de anúncio nos últimos 12 meses, histórico de reclamação, compra de produto agregado e quantidade de produtos agregados comprados.

Definição das Variáveis

Em uma primeira etapa foram consideradas as seguintes variáveis:

Variáveis Transacionais		
Descrição	Definição	Níveis
Status do cliente	Indica se o cliente continua assinando o jornal ou se cancelou o serviço.	Ativo e Inativo
Tipo de Assinatura	Período pelo qual o cliente contratou a assinatura	Anual; boleto mensal; débito automático; semestral e trimestral.
Fonte de Venda	Representa o canal de vendas pelo qual o cliente adquiriu uma assinatura.	Agente; antigo (na migração do sistema os dados foram convertidos como antigo); Internet; outros; sem fonte; terceiros; telemarketing ativo; telemarketing receptivo.
Forma de pagamento	Meio de pagamento utilizado pelo cliente	Boleto bancário; cartão de crédito no débito automático; cartão de crédito parcelado; débito em conta corrente automático; débito em conta corrente parcelado; não informada.
Tempo de permanência	Quantidade de meses que o cliente ficou ativo no caso dos cancelados; ou que ainda está ativo, no caso dos atuais clientes	De 0 até 3 meses; de 3 até 6 meses; de 6 meses até 1 ano; de 1 ano até 2 anos; de 2 anos até 3 anos; mais de 3 anos.
Quantidade de Produto Agregado	Indica o volume de produtos comprados pelos clientes	Até de 3 produtos; entre 4 e 7 produtos; entre 8 e 10 produtos e mais de 10 produtos
LTV – Life Time Value	Score de segmentação dos clientes por <i>life time value</i> que é igual ao tempo de permanência atual em meses mais o tempo de vida futuro em meses (estimado por análise de sobrevivência) x margem (R\$) da assinatura mensal do cliente.	Platina; Diamante; Ouro; Prata; Bronze; Lata; Indefinido.

Indicador de compra de produto agregado	Indica se o cliente já comprou ou não um produto agregado (livros, guias, CD e outros colecionáveis)	Sim ou Não
Indicador de compra de anúncio	Indica se o cliente já comprou um anúncio nos classificados ou não	Sim ou Não
Indicador de Reclamação	Indica se o cliente já fez uma reclamação através da central de atendimento ou pela internet	Sim ou Não
Variáveis Geográficas		
Descrição	Definição	Níveis
Região	Bairro ou cidade onde o assinante recebe o jornal agrupado por similaridade nos custos de distribuição	Baixada, subúrbio, Barra, Zona Norte, Zona Sul, Brasília, interior do RJ, Niterói, outros estados, São Paulo/Belo Horizonte
Variáveis Demográficas		
Descrição	Definição	Níveis
Gênero	Indica o sexo do assinante, mas alguns assinantes não têm o campo preenchido	Feminino, masculino e indefinido
Faixa Etária	Agrupamento que indica a idade do assinante titular	menos de 20 anos; de 20 a 25 anos; de 26 a 30 anos; de 31 a 35 anos; de 36 a 40 anos; de 41 a 50 anos; de 51 a 60 anos; de 61 a 80 anos; mais de 80 anos; sem preenchimento
SD&W*	Score de classificação sócio-demográfica utilizando variáveis do senso do IBGE para os setores censitários do Rio de Janeiro	Solidez; segurança; equilíbrio em conquista; estabilidade; sobrevivência; privação; penúria; subsistência e indefinido.
Variáveis de Fidelização		
Descrição	Definição	Níveis
Indicador de participação em ações de fidelização	Indica se alguma vez o cliente participou das ações de relacionamento tais como vantagens, descontos ou participação em eventos exclusivos.	Sim ou Não

Quadro 2 –Variáveis Definidas para a Pesquisa

*A SD&W é uma empresa de consultoria que desenvolve um modelo que utiliza variáveis sócio- demográficas do censo do Instituto Brasileiro de Geografia e Estatística para classificar a população em 9 segmentos distintos de acordo com as características destes indivíduos.

ANÁLISE DOS RESULTADOS

Construção do Modelo de Regressão Logística

No SPSS 13.0, foi aplicado o modelo de regressão logística binário, método *stepwise forward*, para definir o modelo final que minimiza o número de variáveis e maximiza a precisão do modelo.

Observados			Previsto		
			Status		% Correto
			cancelado	ativo	
Passo 0	STATUS	CANCELADO	11057	0	100
		ATIVO	24492	0	0
	PERCENTUAL GERAL				31.1

Tabela 1 – Resultado do Modelo Inicial : apenas uma constante

O resultado do modelo inicial apresenta a tabela de classificação considerando o modelo com apenas uma constante, ou seja, se arbitrariamente todas as assinaturas fossem consideradas canceladas, a taxa de acerto seria de 31%. O modelo de regressão logística que irá estimar o risco de cancelamento de clientes precisa ser mais assertivo na classificação dos clientes.

A primeira variável a ser incluída no modelo será aquela que tiver a estatística de pontuação mais alta, estatística Wald. No caso a variável tempo de permanência é selecionada a compor o modelo. Em segundo lugar, a variável forma de pagamento é incorporada ao modelo. E em seguida, o indicador de reclamação. Essas 3 variáveis contribuem com 92,1% do poder explanatório do modelo.

A análise direcionada a passos utilizando a estatística Wald consumiu 11 passos até se obter o modelo final. Observando-se as significâncias estatísticas do modelo, constatamos que o coeficiente é significativo a cada passo.

Após diversas interações, o modelo final selecionou 11 das 14 variáveis incluídas inicialmente no modelo. Excluiu as variáveis: gênero, indicador de compra de produto agregado e indicador de compra de um anúncio de publicidade. As variáveis que resultaram do modelo final são: “tempo de permanência”, “forma de pagamento”, “indicador de reclamação”, “tipo de assinatura”, “fonte de venda”, “faixa etária”, “SD&W”, “LTV”, “região”, “indicador de participação em ações de fidelização” e “quantidade de produtos agregados comprados”.

Avaliação do Ajuste Geral

A cada passo, quando uma nova variável é incluída no modelo, a estatística de probabilidade – $2\log$ diminui indicando uma melhora no modelo. Em contrapartida, as medidas pseudo R^2 aumentam à medida que previsores são adicionados. O pseudo R^2 de Nagelkerke no último passo aumentou em 30% o poder de explicação do modelo obtido no passo 1.

Passo	Probabilidade - 2log	Cox & Snell R- Square	Nagelkerke R- Square
1	35227.478 ^a	.220	.310
2	33727.998 ^a	.253	.355
3	33430.165 ^a	.259	.364
4	33157.816 ^b	.264	.372
5	32940.573 ^b	.269	.378
6	32711.454 ^c	.274	.385
7	32545.559 ^c	.277	.390
8	32397.244 ^c	.280	.394
9	32271.724 ^d	.283	.398
10	32221.754 ^d	.284	.399
11	32195.392 ^d	.284	.400

a. Estimaco encerrou-se na quinta interao porque os estimadores dos parmetros mudaram menos que 0.001.

b. Estimaco encerrou-se na sexta interao porque os estimadores dos parmetros mudaram menos que 0.001.

c. Estimaco encerrou-se na stima interao porque os estimadores dos parmetros mudaram menos que 0.001.

d. Estimaco encerrou-se na vigsima interao porque os estimadores dos parmetros mudaram menos que 0.001.

Tabela 2 – Sumrio do Modelo

Passo	Chi-quadrado	Graus de Liberdade	Sig.
1	0.000	4	1.000
2	139.350	7	0.000
3	135.151	7	0.000
4	187.903	7	0.000
5	129.408	8	0.000
6	134.869	8	0.000
7	134.234	8	0.000
8	120.247	8	0.000
9	144.911	8	0.000
10	142.347	8	0.000
11	135.804	8	0.000

Tabela 3 - Teste de Hosmer e Lemeshow

A medida Hosmer e Lemeshow de ajuste geral tem um teste estatstico que indica que no houve diferena estatisticamente significativa entre as classificaes observadas e previstas para todos os modelos com duas ou mais variveis. Um bom ajuste de modelo  indicado por um valor *chi-quadrado* no significativo (HAIR *et al.*, 1998), como o observado no modelo.

No modelo do ltimo passo, todas as medidas de ajuste melhoraram. O valor -2LL diminuiu para 32.195. Os valores R² variam de 0,310 para 0,400, indicando melhoria no modelo de 11 variveis, embora seja um valor distante dos valores R² geralmente encontrados em regresso mltipla. A medida Hosmer e Lemeshow indica a ausncia de diferena significativa na distribuo de valores dependentes efetivos e previstos.

Preciso da Estimativa

Problemas na utilizao de mtodos de regresso logstica stepwise podem ocorrer quando o objetivo da anlise  a preciso da estimativa. Os algoritmos direcionados a passos buscam um subconjunto de variveis que maximize a probabilidade, mas isto no  o mesmo que maximizar a preciso da estimativa.

As matrizes de classificao, idnticas em natureza s utilizadas na anlise discriminante (HAIR *et al.*, 1998), mostram taxas de acerto extremamente altas de casos corretamente classificados

para o modelo de 11 variáveis. A taxa de acerto geral é de 75,3% e, de forma adicional, as taxas de acerto de grupos individuais são consistentemente altas e não indicam um problema na previsão de qualquer um dos dois grupos. As taxas de acerto do grupo que cancela é maior que a taxa do grupo que não cancela, 77,4% contra 74,3%.

O modelo inicial que considerava apenas a constante tinha uma taxa geral de acerto de 31,1%. O modelo completo com 11 variáveis aumenta 2,5 vezes a taxa de acerto na previsão.

Observados		Previsto			
		Status		% Correto	
		cancelado	ativo		
Passo 1	STATUS	CANCELADO	8519	2538	77.0
		ATIVO	7140	17532	70.8
		PERCENTUAL GERAL			72.8
Passo 2	STATUS	CANCELADO	8528	2529	77.1
		ATIVO	6633	17859	72.9
		PERCENTUAL GERAL			74.2
Passo 3	STATUS	CANCELADO	8464	2593	76.5
		ATIVO	6477	18015	73.6
		PERCENTUAL GERAL			74.5
Passo 4	STATUS	CANCELADO	8462	2595	76.5
		ATIVO	6407	18085	73.8
		PERCENTUAL GERAL			74.7
Passo 5	STATUS	CANCELADO	8634	2423	78.1
		ATIVO	6590	17902	73.1
		PERCENTUAL GERAL			74.6
Passo 6	STATUS	CANCELADO	8509	2548	77.0
		ATIVO	6371	18121	74.0
		PERCENTUAL GERAL			74.9
Passo 7	STATUS	CANCELADO	8560	2197	77.4
		ATIVO	6438	18054	73.7
		PERCENTUAL GERAL			74.9
Passo 8	STATUS	CANCELADO	8553	2504	77.4
		ATIVO	6347	18145	74.1
		PERCENTUAL GERAL			75.1
Passo 9	STATUS	CANCELADO	8549	2598	77.3
		ATIVO	6286	18217	74.3
		PERCENTUAL GERAL			75.3
Passo 10	STATUS	CANCELADO	8558	2499	77.4
		ATIVO	6275	18217	74.4
		PERCENTUAL GERAL			75.3
Passo 11	STATUS	CANCELADO	8554	2503	77.4
		ATIVO	6291	18201	74.3
		PERCENTUAL GERAL			75.3

Tabela 4 - Tabela de Classificação

A partir do passo 9, a melhora no R^2 é pequena e a taxa de acerto geral do modelo não se altera. Isto indica que as variáveis 10 e 11 poderiam ser descartadas do modelo final porque quanto menos variáveis um modelo tiver, menor o tempo de processamento. No entanto, as variáveis “indicador de participação em ações de fidelização” e “quantidade de produtos agregados comprados” foram mantidas neste modelo não com o objetivo de aumentar a sua precisão, mas para ajudar na definição do perfil dos clientes que cancelam a assinatura do jornal.

Validação do Modelo (*Holdout Sample*)

Para conseguir a eficiência classificatória do modelo, a amostra foi separada em duas partes: uma utilizada para estimação do modelo, e outra para testar a eficiência da classificação – *holdout sample* (HAIR *et al.*, 1998). A amostra utilizada para estimação, também chamada de amostra de treinamento, contou com 35.549 assinantes. O processo de escolha foi realizado no software SPSS através da geração de números aleatórios.

Observados			Previsto		
			Status		% Correto
			cancelado	ativo	
Passo 11	STATUS	CANCELADO	1162	365	76,1
		ATIVO	809	2460	75,3
PERCENTUAL GERAL					75,5

Tabela 5 - Amostra de Validação –Tabela de Classificação

Observados			Previsto		
			Status		% Correto
			cancelado	ativo	
Passo 11	STATUS	CANCELADO	8554	2503	77,4
		ATIVO	6291	18201	74,3
PERCENTUAL GERAL					75,3

Tabela 6 - Amostra de Treinamento –Tabela de Classificação

Interpretação dos Resultados

O modelo de regressão logística selecionou as 11 variáveis que melhor explicam o cancelamento de uma assinatura de jornal. Gerado o modelo, atribui-se uma probabilidade de cancelamento a cada indivíduo da base em estudo. Todos os indivíduos foram classificados em dois grupos: o grupo que cancela e o grupo que não cancela.

O cliente pode se desligar da empresa motivado por diversos fatores. Alguns desses fatores podem ser descritos por meio de variáveis e utilizados para estabelecer uma relação entre cancelamento e a ocorrência ou não de situações relacionadas aos fatores em questão (BARROS, 2002).

Ao desenvolver o modelo de regressão logística sobre a base de dados, foi possível determinar um modelo que contempla as seguintes variáveis: “tempo de permanência”, “forma de pagamento”, “indicador de reclamação”, “tipo de assinatura”, “fonte de venda”, “faixa etária”, “SD&W”, “LTV”, “região”, “indicador de participação em ações de fidelização” e “quantidade de produtos agregados comprados”.

Essas variáveis aplicadas ao modelo de regressão logística geraram um *score* de probabilidade de cancelamento para cada um dos clientes da base de dados conforme a tabela a seguir.

Faixa Probabilidade	Ativo	Cancelado	Total Geral	% canc.
entre 0.00 e 0.20	15,249	1,601	16,85	9%
entre 0.20 e 0.40	4,763	1,656	6,419	18%
entre 0.40 e 0.60	2,828	2,635	5,463	32%
entre 0.60 e 0.80	1,512	3,848	5,36	54%
entre 0.80 e 1.00	140	1,317	1,457	61%

Tabela 7 – Probabilidade de Cancelamento para clientes da base de dados

Nota-se que o percentual de assinaturas canceladas cresce à medida que o *score* de probabilidade de cancelamento aumenta. Isto é coerente com a tabela de classificação que acerta em 75,3% dos casos.

Aplicado o *score* de probabilidade na base total de assinantes pode-se concluir que a maior parte dos clientes ativos apresenta baixa probabilidade de cancelar a assinatura do jornal: 62% da base de clientes tem até 20% de chance de cancelar a assinatura do jornal.

Na análise de regressão logística é possível identificar qual a mudança nas chances de um evento ocorrer dada a presença de um fator (variável categórica) ou alteração em uma variável contínua. É possível descrever ainda o perfil dos clientes que cancelam, definindo ações profiláticas que ajudem a reduzir a perda de clientes.

CONSIDERAÇÕES FINAIS

O objetivo deste estudo foi utilizar a base de dados de clientes de uma empresa para desenvolver um modelo de regressão logística que determinasse o risco de cancelamento de clientes que possuem uma assinatura de jornal, bem como traçar o perfil dos clientes com maior risco de desligamento, de forma que fosse possível propor ações de retenção que aumentem a lucratividade da carteira de assinantes desta empresa.

É possível afirmar que o modelo de regressão logística pode ser eficiente para a determinação do risco de cancelamento de clientes a partir de variáveis sócio-demográficas, transacionais e do histórico de eventos.

O modelo foi aplicado a uma amostra de dados com 35.549 clientes e as variáveis compostas no modelo final foram: “tempo de permanência”, “forma de pagamento”, “indicador de reclamação”, “tipo de assinatura”, “fonte de venda”, “faixa etária”, “SD&W”, “LTV”, “região”, “indicador de participação em ações de fidelização” e “quantidade de produtos agregados comprados”.

A tabela de classificação mostrou que a taxa de acerto geral do modelo de regressão logística é de 75,3% e que as taxas de acerto dos grupos individuais são altas e indicam uma consistência na previsão de qualquer um dos dois grupos. O grupo que cancela apresentou taxa de acerto de 77,4% enquanto o grupo que não cancela tem taxa de acerto de 74,3%.

O pseudo R^2 de Nagelkerke apresentou um poder de explicação de 0,4 e a medida Hosmer e Lemeshow de ajuste geral indica que não houve diferença estatisticamente significativa entre as classificações observadas e previstas para o modelo final. Além disso, o valor de -2LL aumentou a cada passo.

A combinação dessas medidas de avaliação do ajuste e da precisão do modelo indica a aceitação deste como um modelo de regressão logística significante. Além disso, o modelo foi aplicado a uma base de validação diferente da amostra utilizada na fase de estimação dos coeficientes do modelo. As taxas de acertos na amostra de validação são quase idênticas às taxas de acerto da amostra de treinamento, o que indica que o modelo de regressão logística possui suporte empírico tanto na amostra de validação quanto na de treinamento.

Assim, de acordo com as variáveis assumidas no modelo de regressão logística, conclui-se que o perfil do assinante com maior risco de cancelamento da sua assinatura é: o jovem de até 30 anos; dos segmentos sócio-demográfico de menor nível (sobrevivência, penúria, privação e subsistência); morador da Baixada, subúrbio ou de outros estados; que tenha adquirido sua assinatura através do telemarketing ativo; cuja assinatura seja da modalidade anual, com forma de pagamento boleto ou débito em conta corrente; que tenham comprado sua primeira assinatura recentemente; sem compra de produto agregado ao jornal e que nunca tenham feito reclamações.

Os resultados obtidos indicam que o modelo é bastante eficiente em distinguir assinantes que possivelmente cancelariam sua assinatura dos que não cancelariam. Assim, ao implementar uma ação de marketing com o objetivo de reter os clientes, as informações fornecidas pelo modelo

permitirão que se atinja o grupo com maior risco de cancelamento com mais eficiência do que sem a utilização do modelo.

Uma seleção aleatória para uma ação de retenção faria com que 31% dos clientes canceladores de sua assinatura fossem impactados pela ação. Estariam sendo atingidos 69% de clientes cujos riscos de cancelamento são pequenos, ou seja, haveria um desperdício de investimento nesta ação. A utilização do modelo permite que se alcance 75% de clientes canceladores. Com isso, há maior eficácia na ação de marketing, à medida que se diminui a dispersão do investimento e aumenta-se a probabilidade de o público alvo ser atingido.

Os resultados obtidos indicam algumas ações de marketing e de negócio que podem contribuir para aumentar a rentabilidade da empresa através da retenção de clientes ou da implantação de ações que melhorem a qualidade da venda, o que irá se reverter em clientes que permaneçam mais tempo na carteira.

A empresa precisa rever as ações de fidelização que estão sendo utilizadas atualmente, pois não estão se revertendo em menos cancelamentos para a empresa, uma vez que os clientes que participaram de ações de fidelização apresentam *score* de cancelamento quase idêntico aos assinantes que não participaram.

REFERÊNCIAS BIBLIOGRÁFICAS

- ANDERSON EW, Customer satisfaction and price tolerance. **Marketing Letters**, 1996
- ANDERSON EW, SULLIVAN MW. The Antecedents and Consequences of Customer Satisfaction for Firms - **Marketing Science**, vol.12, No. 2, 1993.
- BALDINGER A.L., RUBINSON J. Brand Loyalty: The Link between Attitude and Behavior. - **Journal of Advertising Research**, 1996
- BANDYOPADHYAY S.; MARTELL, M.. Does attitudinal influence behavioral loyalty? A theoretical and empirical study. **Journal of Retailing and Consumer Services**, v.14, n.1, p.35-44, 2007.
- BANSAL Harvir S., TAYLOR Shirley F. The Service Provider Switching Model (SPSM) A Model of Consumer Switching Behavior in the Services Industry, **Journal of Service Research**, Vol. 2, No. 2, 200-218, 1999.
- BARROS, José Eduardo G. **Análise de Sobrevivência: Modelo de Risco de Desligamento de Clientes**. 01/03/2002, 1v. 123p. Mestrado. Universidade Federal Do Rio de Janeiro – Administração. Orientador: Celso Funcia Lemme. Biblioteca Depositária: COPPEAD/UFRJ.
- BENNETT, R.; RUNDLE-THIELE, S.. A Comparison of Attitudinal Loyalty Measurement Approaches. **Journal Brand Management**, v.9, n.3, p.193-209, 2002.
- BERRY, L. Relationship Marketing of Services—Growing Interest, Emerging Perspectives, **Journal of the Academy of Marketing Science**, Vol. 23, No. 4, 236-245, 1995.
- CARUANA, Albert. The Impact of Switching Costs on Customer Loyalty: a study among corporate customers of mobile telephony. **Journal of Targeting, Measurement and Analysis for Marketing**, v. 12, n. 3, 2004.
- CHAUDHURI, A.; HOLBROOK, M.. The Chain of Effects from Brand Trust and Brand Affect to Brand Performance: The role of brand loyalty. **Journal of Marketing**, v.65, n.2, p.81-93, 2001.
- CZINKOTA MR. **Marketing: as melhores práticas**. Bookman, 2001.
- DAY, G.S. A two-Dimensional Concept of Brand Loyalty. **Journal of Advertising Research**, v. 9 n.3, p.29-35, 1969
- EVANSCHITZKY H., WUNDERLICH M. An Examination of Moderator Effects in the Four-Stage Loyalty Model , **Journal of Service Research**, 2006.

- EVANSCHITZKY, H.; IYER, G.R.; PLASSMANN, H.; NIESSING, J.; MEFFERT, H.. The relative strength of affective commitment in securing loyalty in service relationships. **Journal of Business Research**, v 59, n.12, p.1207-1213, 2006.
- FORNELL, Claes et all. The American Customer Satisfaction Index: Nature, Purpose, and Findings. **Journal of Marketing**, vol. 60, 7-18, 1996.
- FORNELL, Claes. A. National Customer Satisfaction Barometer: the Swedish experience. **Journal of Marketing**, v. 56, p. 6-21, January, 1992.
- GERPOTT T.J., RAMS W., SCHINDLER A. Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications.- **Telecommunications Policy**– Elsevier, 2001.
- GRÖNROOS, Christian. **Marketing: gerenciamento e serviços**. Rio de Janeiro: Elsevier, 2003, 481p.
- HAIR, J. F. et al. **Multivariate Data Analysis**. 4 ed. New Jersey: Prentice Hall, 1998
- JONES M.A., MOTHERSBAUGH D.L., BEATTY S.E., Switching barriers and repurchase intentions in services - **Journal of Retailing**, 2000
- JOHNSON, Michael D. et all. The evolution and future of national customer satisfaction index models. **Journal of Economic Psychology**, 22, p.217-245, 2001.
- KOTLER, P. **Administração de Marketing**.. Ed. São Paulo: Prentice Hall, 2000.
- KOTLER, Philp, ARMSTRONG, Gary. **Princípios de marketing**. 7.ed. Rio de Janeiro. Prentice-Hall, 1995.
- LACOEUILHE, J. Le concept d'attachement à la marque dans la formation du comportement de fidélité. **Revue Française du Marketing** . n° 165, p. 29-42, 1997.
- LOVERLOCK, Christopher H.; WRIGHT, Lauren. **Serviços: marketing e gestão**. Rio de Janeiro: Saraiva, 2001.
- McKENNA, Regis. **Marketing de relacionamento: estratégias bem-sucedidas para a era do cliente**. Rio de Janeiro: Campus, 1993.
- MELO, Alexandre Sérgio C. de. **O marketing de relacionamento como fator de influência na lealdade do cliente: um estudo de caso no mercado de assinatura de revistas**. 01/10/2002. 1v. 172p. Mestrado. UNIVERSIDADE DE SÃO PAULO – ADMINISTRAÇÃO. Orientador: Geraldo Luciano Toledo. Biblioteca Depositária: Biblioteca da FEA-USP.
- NASH, Edward L. **Database marketing: uma ferramenta atual e decisiva do marketing**. São Paulo. Makron, 1994
- NEAL, William D. Satisfaction is Nice, But Value Drives Loyalty. **Marketing Research**, v. 11, n. 1, p. 20-23, 1999.
- OLHSON, J. Financial Ratios and the Probabilistic Prediction of Bankruptcy. **Journal of Accounting Research**, Vol. 18, No. 1 , pp. 109-131, 1980.
- OLIVER, R.L. “Whence consumer loyalty”. **Journal of Marketing**, v. 63, p. 33-44, 1999.
- OLIVER, R.L.; SWAN, J.E. Equity and disconfirmation perceptions as influences on merchant and product satisfaction. **Journal of Consumer Research**, v. 15, n. 4, p. 374, dec., 1989.
- SIRDESHMUKH, Deepak, SINGH, Jagdip, SABOL, Barry. “Consumer trust, value, and loyalty in relational exchanges”. **Journal of Marketing**. Janeiro, 2002.
- REICHHELD, F. Loyalty-Based Management. In: **Harvard Business Review**. P. 64-73, March/April, 1993.
- REICHHELD F.F., SASSER W.E. Jr. Zero defections: quality comes to services, **Harvard Business Review**, vol 68(5):105-11, 1990.
- REICHHELD, Frederick F. **The loyalty effect: the hidden force behind growth, profits, and lasting value**. Boston: Harvard Business Scholl Press, 1996.

- ROSSI, Carlos Alberto V.; SLONGO, Luiz Antonio. **Pesquisa de Satisfação de Clientes: o Estado-da-Arte e Proposição de um Método Brasileiro.** Anais do XXI. ENANPAD/Marketing, 1997.
- RUST R.T., ZAHORIK A.J. Customer satisfaction, customer retention, and market share, **Journal of Retailing**, 1993.
- SHETH J.N., PARVATIYAR A. **Handbook of relationship marketing** - Sage Publications Thousand Oaks, 2000.
- VAVRA, Terry G. **Marketing de relacionamento: Aftermarketing.** São Paulo: Atlas, 1993.
- ZEITHAML VA, BITNER MJ. **Services marketing.** McGraw-Hill Boston, Massachusetts, 1996.
- YI, Y. A critical review of consumer satisfaction. In: ZEITHAML, V.A. **Review in marketing.** Ann Harbor: AMA, p. 68-113, 1990.