

Fidelidade em Telefonia Celular: Proposição e Validação de um Índice para Previsão da Fidelidade de Clientes

Autoria: Ana Denise Silva Veloso, José Marcos Carvalho de Mesquita

Resumo: No mercado de telecomunicações, fica evidenciado que a concorrência e a competição emergente no país, que vem ocorrendo em função da privatização do setor, deixam os consumidores rentáveis cada vez mais no alvo de políticas de fidelização por parte das operadoras de telefonia celular. Nesse contexto, a pesquisa apresentada foi elaborada a partir de informações de clientes de uma operadora de telefonia celular que atua no mercado de Minas Gerais, com o objetivo de avaliar como o perfil de utilização dos serviços por parte dos clientes e o relacionamento dos clientes com a operadora e os serviços prestados afeta seu comprometimento e sua lealdade, de modo a construir um índice de fidelização por cliente. O processo de análise utilizou dados de fonte secundária, disponíveis nos bancos de dados da operadora estudada, que foram explorados por técnicas de mineração de dados utilizando como ferramenta a técnica estatística de análise de regressão logística. Foi proposto um modelo para estimar a probabilidade de cancelamento do contrato por parte do cliente e os resultados confirmam a possibilidade de se detectar preventivamente a intenção de cancelamento.

Introdução

Acredita-se que a abertura de mercado (quebra de monopólio) tenha gerado grandes expectativas nos consumidores, os quais, por sua vez, começaram a apresentar maiores sinais de exigência com relação aos produtos e serviços ofertados, fazendo com que as empresas se empenhassem em atuar em consonância com a demanda dos clientes. Portanto, conhecer as oscilações do nível de satisfação dos consumidores com os serviços oferecidos e o grau de fidelidade dos clientes passou a constituir uma relevante fonte de informações, para que a empresa saiba em quais aspectos está apresentando um bom desempenho e em quais precisa melhorar, para garantir sua competitividade e rentabilidade em longo prazo.

Entender o relacionamento dos clientes com a operadora de telefonia celular é fundamental não apenas para que a empresa conquiste novos clientes, mas também para que mantenha seus assinantes em sua base. Segundo Vavra (1993), conquistar novos clientes pode até ser uma tarefa fácil, todavia a sua manutenção é um dos fatores-chave para o sucesso de uma empresa.

Um aspecto importante a ser considerado no cenário do setor de telecomunicações móveis diz respeito ao cancelamento do contrato de clientes, designado por *churn*, que significa perda de clientes de uma empresa.

Existem dois tipos de *churn*: o voluntário e o involuntário. O *churn* voluntário ocorre quando o cliente, por vontade própria, solicita à operadora o cancelamento da prestação dos serviços. Tal solicitação pode ser motivada por razões relacionadas diretamente à sua relação com a operadora (insatisfação, ofertas mais agressivas dos concorrentes, por exemplo, que o fazem optar por outra operadora) ou por razões que fogem ao controle tanto do cliente quanto da operadora (mudança de estado/país, falecimento, problemas financeiros, etc). O *churn* involuntário, por sua vez, é resultado de uma ação da própria operadora, que, por inadimplência do cliente ou por detecção de alguma fraude, decide cancelar o contrato, para que não sejam computadas perdas financeiras maiores.

É importante considerar ainda que a agilidade e o dinamismo do setor de telefonia celular não estão permitindo demora na identificação dos clientes infieis e na adoção de estratégias para mantê-los, razão pela qual acredita-se ser essencial a determinação de um índice de fidelidade, que possa ser utilizado para orientar a definição das ações de fidelização das empresas.

Segundo Reichheld (2000), se a empresa analisar o comportamento dos clientes antes de se afastarem, poderá identificar os padrões de deserção e, assim, evitar que se repitam em outros compradores e até mesmo reter quem estiver a ponto de partir.

Tendo em vista essas perspectivas, formulou-se o seguinte problema de pesquisa: ***Que variáveis interferem na determinação de um índice de fidelidade dos clientes de uma operadora mineira de telefonia móvel?***

O objetivo principal desse estudo é construir um índice de fidelização de clientes de uma operadora mineira de telefonia celular, quanto ao relacionamento empresa/clientes, com vistas a direcionar e aperfeiçoar a implantação de ações que permitam a fidelidade de sua base de clientes. Além disso, pretende-se identificar as variáveis que explicam a fidelização dos clientes no contexto estudado e validar a consistência do índice construído, por meio de dados internos.

Referencial Teórico

Foi feita uma revisão da literatura sobre o comportamento do consumidor, definindo-se e aprofundando-se conceitos referentes a fidelidade/lealdade, inteligência de *marketing* para tomada de decisão estratégica de manutenção de clientes, gestão de relacionamento com cliente (CRM - *Customer Relationship Management*) e exploração de dados referentes à base de clientes (*Data Mining*).

Comportamento do consumidor

Para a comercialização eficaz de serviços, é preciso entender os processos usados pelos consumidores durante o estágio de compra. É muito importante entender os consumidores, a maneira como escolhem os serviços alternativos que lhes são oferecidos e como avaliam esses serviços após recebê-los. (HOFFMAN e BATESON, 2003).

Segundo Engel et al. (2000) o comportamento do consumidor é a somatória das atividades diretamente envolvidas em obter, consumir e descartar produtos e serviços, incluindo os processos decisórios que antecedem e sucedem tais ações.

Kotler (2000) descreve o processo de avaliação de alternativas em três etapas: na primeira, o consumidor tenta satisfazer sua necessidade; na segunda, ele busca certos benefícios no produto ou serviço e, por último, observa em cada produto ou serviço um conjunto de atributos.

Uma vez tomada a decisão pela compra, o consumidor pode efetivá-la rapidamente (em especial se o produto ou serviço estiver em promoção), pode desistir, se considerar que suas necessidades não estão sendo atendidas, ou adiá-la em função de uma possível economia. O processo de modificar, adiar ou rejeitar a compra é altamente influenciado, segundo o autor, pelo risco percebido, que pode variar de acordo com o montante de dinheiro envolvido, o nível de incerteza quanto aos atributos e o nível de autoconfiança do consumidor.

Cada um desses estágios percorridos pelo consumidor tem suas características próprias e relevâncias para o entendimento do processo. Para Engel et al. (2000), todos os estágios podem ser explorados, a fim de facilitar o processo decisório do consumidor, podendo oferecer, dessa maneira ganhos substanciais na decisão de compra e re-compra de bens ou serviços.

Este estudo se restringe à etapa de pós-compra. Segundo Lovelock et al. (2001), os clientes continuam o processo iniciado na etapa de encontro do serviço, avaliando sua qualidade e a própria satisfação ou insatisfação com a experiência vivenciada. O resultado desse processo afetará suas intenções futuras, inclusive quanto a se permanecerão fiéis a seu fornecedor de serviço e se farão recomendações positivas a seu respeito para membros de sua família e conhecidos.

Conceito de fidelidade

O conceito de fidelidade tornou-se mais relevante ainda, a partir do momento em que as empresas começaram a perceber que a fidelidade dos seus clientes é que iria garantir retornos financeiros futuros. Conforme Oliver (1996), a fidelidade tem efeito sobre a lucratividade, na medida em que tem influência direta sobre o fluxo futuro de clientes. Atualmente, é destacada a forte tendência das empresas a migrarem de estratégias baseadas em satisfação para estratégias orientadas por fidelização, pois são evidentes os resultados financeiros positivos, quando se tem uma base de clientes fiéis.

Oliver (1999) observa também que a lealdade do consumidor é fortemente relacionada com sua retenção pela organização e, por conseguinte, com a lucratividade e o desempenho superior, ainda que mais estudos sejam necessários para confirmar a lucratividade da lealdade.

Reichheld (1996) desenvolveu uma importante contribuição para a administração estratégica, analisando a questão da retenção de clientes, com vistas a chegar a uma estratégia capaz de obter a lealdade (ou fidelidade) dos clientes. Segundo o autor, um consumidor verdadeiramente leal deve manter uma atitude favorável em relação ao produto/serviço, além de adquiri-lo repetidamente. Nesse sentido, a lealdade implicaria repetidas compras baseadas em fatores cognitivos, afetivos, avaliativos e disposicionais, que são os componentes clássicos de uma atitude.

Pensando na atitude dos consumidores, para competir em mercados instáveis, as empresas necessitam de um posicionamento dinâmico, voltado para a busca da sua integração com o cliente. Ao contrário do posicionamento tradicional, o posicionamento dinâmico é um processo que compreende três dimensões: produto, mercado e empresa (MCKENNA, 1999).

A primeira dimensão (produto) refere-se ao modo como a empresa deseja colocar seu produto no mercado competitivo, sendo aconselhável que dê atenção especial a fatores intangíveis do posicionamento, que constituem a chave para a construção de relações com os consumidores.

Na segunda dimensão (mercado), o produto tem de ganhar o reconhecimento do mercado e a credibilidade dos consumidores, tornando-se um “vencedor”. Para isso, as empresas necessitam conhecer e identificar os principais participantes da infra-estrutura do setor e trabalhar intimamente com eles.

Na terceira dimensão, estágio final do processo, as empresas têm de procurar um posicionamento para si mesmas, o qual depende de seu sucesso econômico, pois empresas com problemas financeiros são obrigadas a reconstruir a própria posição no mercado, já que os consumidores relutam em comprar produtos de empresas com imagem desgastada.

Mckenna (1999) entende que o posicionamento dinâmico “traça uma linha comum em todas as partes da empresa, conectando-as ao mercado”.

Assim, uma empresa que tenha boa imagem junto ao mercado poderá ter maior facilidade na seleção de seu pessoal e, com profissionais mais qualificados, poderá estar mais apta a enfrentar a concorrência. Além disso, terá maior facilidade para fazer negociações, pois as instituições financeiras apóiam empresas com bom posicionamento financeiro.

Inteligência de *marketing* para tomada de decisão estratégica de manutenção de clientes

Nos últimos anos, várias mudanças têm ocorrido com relação ao comportamento e às exigências dos consumidores, refletindo-se na administração de *marketing*. Impulsionadas pelo “poder e pela disseminação onipresente da tecnologia”, as organizações têm sido obrigadas a mudar sua forma de atuação, com base em estratégias voltadas para o cliente, para conseguir acompanhar essa evolução. Essa nova perspectiva tem determinado transformações fundamentais nas organizações, em termos de crenças e de valores (MCKENNA,1999).

A percepção pelas empresas da importância da retenção de seus clientes ativos, em face da concorrência cada vez mais acirrada, transforma em desafio a responsabilidade de reter o cliente ativo. Na verdade, novos clientes estão cada vez mais raros e difíceis. Surge, então, o conceito de *marketing* de relacionamento, como uma alternativa para o *marketing* de massa. O objetivo desse novo paradigma é possibilitar à empresa tratar seus clientes individualmente e desenvolver com eles um relacionamento duradouro. Forma-se, assim, uma rede de *marketing* composta pela empresa e os “*stakeholders*” (consumidores, funcionários, fornecedores, distribuidores, varejista, agências de propaganda, cientistas universitários e outros), cuja principal tarefa é criar forte lealdade nos consumidores (KOTLER, 2000).

A utilização das estratégias de *marketing* de relacionamento no setor de telecomunicações móveis tem ligação direta com o estabelecimento de uma relação de longo prazo com os clientes. As dificuldades advindas do aumento da concorrência têm levado os gestores a buscar novas estratégias de manutenção da base de clientes já existentes, que atualmente é considerada muito mais importante do que a aquisição de novos consumidores.

Rogers e Skinner (2005) explicam que nenhuma empresa pode prever exatamente qual será o comportamento de cada cliente. Empresas de serviços financeiros ou de telecomunicações, por exemplo, têm a felicidade de manter ligações diretas com seus clientes. Muitas dessas empresas usam a massa de dados dos clientes que possuem para criar modelos preditivos individuais. A maioria das outras empresas, entretanto, elabora modelos mais subjetivos e menos específicos para cada cliente. De qualquer forma, o futuro ainda é o futuro. Ele não pode ser previsto com 100% de certeza, mas ele pode ser modificado. Mudar o comportamento dos clientes é, de fato, a principal missão da maioria dos programas e campanhas de vendas e *marketing*.

Para Grönross (1995), a meta do *marketing* de transação é conquistar clientes, enquanto a meta do *marketing* de relacionamentos é manter clientes. É fundamental compreender essa

diferença, para se perceber que, no *marketing* de relacionamento, as ações devem ser implementadas visando à manutenção dos clientes no médio e no longo prazo.

As empresas vêm percebendo que o verdadeiro desafio está em desenvolver e gerenciar relações com os clientes, de modo a modificar seu comportamento, ao longo do tempo, e fortalecer o laço entre clientes e a empresa (NEWELL, 2000).

Administração do relacionamento com os clientes

A inteligência de *marketing* para a manutenção de clientes está intimamente ligada à filosofia do *Customer Relationship Management* (CRM), um sistema que ajuda a empresa a reunir todo o conhecimento coletivo sobre seus clientes, concentrando os dados em um lugar onde possam ser compartilhados por todos os funcionários, por meio do uso de um conjunto variado de ferramentas. Com esse sistema, a empresa consegue entender melhor os clientes, oferecer-lhes mais apoio, servi-los melhor e, em última análise, ganhar novos clientes (GASET, 2001).

Antes de ser um software, a filosofia CRM é o relacionamento da empresa com seus clientes. Por meio da tecnologia, a organização busca o resgate qualitativo das relações, estabelecendo um diálogo interativo com o cliente. A filosofia do CRM pode ser assim traduzida: “esqueça o produto, pense no cliente”. Para isso, é essencial retornar aos modelos de relacionamento dos tempos antigos e procurar compreender os métodos, os números e as restrições de tempo inerentes às organizações.

Peppers e Rogers (2000) têm um conceito mais simples: “CRM é a infra-estrutura para implementar-se a filosofia um-a-um de relacionamento com os clientes”.

A diferença entre o CRM e outros sistemas de computador é que os sistemas “tradicionais” são concebidos ao redor de processos e produtos; já o sistema de CRM é concebido ao redor dos clientes da empresa.

Como se pode perceber, embora os autores conceituem o CRM de maneiras diferentes, todos concordam em dois aspectos fundamentais: tecnologia e foco no cliente.

A chave para o CRM é identificar o que gera valor para o cliente e oferecer-lhe isso. Há muitos conceitos comuns de valor e cada cliente pode ter uma visão diferente, cabendo à empresa satisfazer esses conceitos, tendo em vista cada grupo de clientes.

Na visão de Newell (2000) são tarefas-chave do CRM:

- identificar os valores do consumidor, pertinentes a uma empresa específica;
- entender a importância relativa desses valores para cada segmento de cliente;
- determinar se a oferta desses valores afetará os resultados financeiros, de maneira positiva;
- comunicar e oferecer os valores adequados a cada segmento de clientes, de forma que queiram receber as informações;
- medir os resultados e fornecer retorno sobre o investimento.

A contribuição do *marketing* para o sucesso da empresa reside em sua análise quanto às futuras oportunidades da organização em atender a necessidades bem definidas do consumidor com produtos e serviços que lhe ofereçam melhores benefícios do que a concorrência.

Isso significa que o *marketing* deve sempre tentar descobrir o que os consumidores estão fazendo (comportamentos) e pensando (sentimentos) e quais os fatores externos (ambiente) que estão afetando suas decisões de compra. As empresas podem, então, usar essas informações para alterar o mix desses fatores, de modo a produzir o comportamento de compra desejado (NEWELL, 2000).

Identificação de padrões de comportamento de clientes (*Data Mining*)

A expressão *Data Mining* (mineração de dados) designa exploração e análise de grande quantidade de dados, mediante a utilização de técnicas estatísticas que possibilitem identificar padrões de comportamento dos clientes. Seu objetivo principal é permitir que as empresas aperfeiçoem suas estratégias de *marketing*, vendas e suporte a clientes através de seu melhor entendimento e conhecimento (BERRY e LINOFF, 1997).

Com a geração de um volume cada vez maior de informações, é essencial tentar aproveitar esse investimento ao máximo possível. Talvez a forma mais nobre de utilização desses vastos repositórios seja tentar descobrir os conhecimentos subjacentes às informações. Um banco de dados de transações comerciais pode, por exemplo, conter diversos registros indicando produtos que são comprados em conjunto. Quando se descobre isso, é possível estabelecer estratégias para otimizar os resultados financeiros da empresa. Essa já é uma vantagem suficientemente importante para justificar todo o processo; contudo, embora essa idéia básica seja facilmente compreensível, fica sempre uma dúvida sobre como um sistema é capaz de obter esse tipo de informação (NAVEGA, 2002).

Os dois objetivos de mais alto nível do *Data Mining* tendem a ser a predição e/ou a descrição. Os padrões preditivos são encontrados para resolver o problema de predizer o valor futuro ou desconhecido de um ou mais atributos do banco de dados, a partir do valor conhecido dos demais atributos. Os padrões descritivos (ou informativos) têm por objetivo encontrar informações interessantes, de forma interpretável pelo pesquisador, que descrevam os dados.

Metodologia

O estudo será desenvolvido a partir de dados de clientes, da operadora pesquisada, seu relacionamento com a empresa, obtidos em fontes internas de informações.

Embora alguns clientes permaneçam fiéis desde o início das atividades da operadora em estudo, esta pesquisa limitou-se a analisar os cancelamentos ocorridos em um período de nove meses (entre outubro/2004 e junho/2005), visto que diversos fatores podem influenciar um cliente a cancelar seu contrato ao longo do tempo.

Planejamento Amostral

Como delimitação da pesquisa e para caracterização da população com vistas a realizar um planejamento amostral com o rigor necessário para este estudo, foram considerados os seguintes aspectos:

- realização da análise por contrato (um contrato desativado independe de outros que o mesmo cliente possa ter);
- planos pós pagos de pessoas físicas;

- contratos ativos ou suspensos na data da extração dos dados, ou que foram cancelados voluntariamente nos últimos três meses antes da data da extração;
- período de análise: de 1º/outubro/2004 a 30/junho/2005.
- histórico de análise: de 1º/dezembro/1998 a 30/junho/2005.

A amostragem será selecionada em dois estágios: o primeiro por estratificação (contratos ativos e suspensos; contratos cancelados); o segundo, por amostragem sistemática. O processo amostral estratificado, segundo Scheaffer et al. (1996), é indicado quando os estratos são claramente identificados e cada unidade amostral da população pode ser alocada no estrato apropriado. Além disso, segundo os autores, deve ser certificado que a extração das amostras, em cada estrato, constitui processos independentes. Diferentes esquemas amostrais poderão ser adotados em cada estrato, ou seja, a escolha das observações de um estrato não pode depender da escolha feita no outro.

No segundo estágio será utilizada uma amostragem sistemática, que consiste, segundo Scheaffer et al. (1996), em obter um ponto de partida aleatório e, em seguida, toma-se cada *i*-ésimo elemento sucessivamente até percorrer toda a lista da população. Esse processo foi escolhido por ser significativamente numeroso o banco dados disponível.

Segundo Hair et al (2005), os testes estatísticos são muito sensíveis ao tamanho da amostra (poder estatístico) para atingir a significância estatística, tanto para amostras pequenas quanto grandes. O autor recomenda que sempre que tamanhos de amostras excederem 200 ou 400 casos, o pesquisador deverá examinar todos os resultados obtidos para garantir que tenham significância prática (e não apenas estatística), devido ao poder estatístico aumentado pelo tamanho da amostra.

Dada a aplicabilidade esperada do modelo desenvolvido, optou-se por uma amostra muito maior que a usual, garantindo uma margem de erro que seja a menor possível com o objetivo de otimizar os recursos destinados às ações de fidelização.

O fato das informações estarem disponíveis em bancos de dados, e o custo da amostragem e da coleta de dados não serem significativos, permite lançar mão de uma grande vantagem de se trabalhar com dados secundários. Portanto, serão extraídos dois conjuntos amostrais constituídos por 2000 registros cada um.

Scheaffer et al. (1996), recomendam que o cálculo do tamanho da amostra seja baseado em proporções:

$$n = \left(\frac{Z_{\alpha} \sqrt{p(1-p)}}{I_c} \right)^2$$

Cada amostra será constituída por 2000 clientes, sendo metade de clientes ativos e metade de clientes cancelados. Foram utilizados para este cálculo: probabilidade de cancelamento de cliente (*p*) igual a 5%, nível de significância (*α*) de 0,05. Sendo assim, essa amostra considera uma margem de erro (*I_c*) de 1,24%. Considerando-se a amostra total de 2000 clientes, em cada etapa, na definição e validação do índice, a margem de erro (*I_c*) é de 0,93%.

A primeira amostra (amostra de análise) será utilizada para a criação do modelo logístico, que dará origem à regra de classificação; a segunda amostra (amostra de validação) será utilizada para a etapa de validação do modelo proposto. Será utilizado na fase de validação o método

de colocação de elementos à parte, para classificação (*Holdout Method*) (MINGOTI, 2005). Além disso, o ajuste do modelo será testado para todo o universo.

Serão feitas duas validações (amostra de validação e universo), pois, de acordo com Churchill (1979), pelo menos dois estudos consecutivos são necessários para que novas escalas possam prover alguma evidência de validade e de confiabilidade.

As amostras serão balanceadas (50% clientes “não *churn*” e 50% ex-clientes “*churn*”). Para Hair et al (2005), os tamanhos das amostras também afetam os resultados quando a análise envolve grupos de elementos, como na análise discriminante ou regressão logística. Tamanhos de amostras diferentes entre grupos influenciam os resultados e exigem interpretação e/ou análise adicional. Esta posição corrobora a decisão de realização do procedimento de balanceamento amostral.

Para cada unidade amostral serão extraídas todas as variáveis disponíveis que possam caracterizar o perfil dos clientes de acordo com seu relacionamento com a empresa. Para a modelagem do índice de previsão da fidelidade de clientes, serão utilizadas, principalmente, as seguintes variáveis:

- Dados cadastrais: plano de serviço utilizado, tempo na base (tempo que a pessoa é cliente), perfil do usuário (opção por débito automático, quantitativo de contas pagas em atraso, valor médio faturado).
- Interação do cliente com a operadora: contatos com o setor de atendimento a clientes para solicitação de serviços; reclamações; solicitação de informações;
- Perfil de utilização de tráfego do cliente: taxa de crescimento do tráfego local e interurbano, recebido ou originado, tráfego com outras operadoras de telefonia móvel e/ou fixa (descritas no próximo capítulo).

Técnica de Regressão Logística

Em muitas situações práticas, pesquisadores desejam separar duas classes de objetos ou alocar um novo objeto em uma dessas classes. De todas as técnicas existentes para esta finalidade, a Análise Discriminante e a Regressão Logística são as mais encontradas na literatura. Ambas as técnicas se enquadram na classe de métodos estatísticos multivariados de dependência, pois relacionam um conjunto de variáveis independentes com uma variável dependente categórica (Sharma, 1996; Hair et al., 2005).

Segundo Johnson (1998), o modelo logístico de discriminação não faz suposições quanto à forma funcional das variáveis independentes, e o número de parâmetros envolvidos no processo de estimação pode ser menor.

Hair et al. (2005) apontam alguns motivos para se optar pela regressão logística: não é necessário supor normalidade multivariada; é uma técnica mais genérica e mais robusta, pois sua aplicação é apropriada no caso de grande variedade de situações; é uma técnica similar à regressão linear múltipla, com a seguinte equação:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

Em que

- Y é a variável resposta contínua;
- β_i são os coeficientes;
- X_i são as variáveis explicativas;
- ε é o erro aleatório.

Embora parecida com a regressão linear, na equação logística, a variável-resposta é dicotômica, e o interesse é estimar a probabilidade de cancelamento do cliente (*Churn*), $Y=1$.

Na regressão logística, a probabilidade de ocorrência de um evento pode ser estimada diretamente. No caso da variável dependente Y assumir apenas dois possíveis estados (1 ou 0) e haver um conjunto de p variáveis independentes X_1, X_2, \dots, X_p , o modelo de regressão logística pode ser registrado da seguinte forma:

$$P(Y=1) = \frac{1}{1 + e^{-g(x)}}$$

Em que $g(x) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$

Medidas de Ajuste do Modelo

O modelo ajustado tem como objetivo principal discriminar os indivíduos que cancelarão seus contratos dos indivíduos que permanecerão clientes. Existem várias medidas que permitem mensurar e comparar o desempenho de modelos na realização desse propósito. Nesta seção serão apresentadas três das medidas de desempenho mais populares e que serão utilizadas na comparação dos modelos ajustados: curva ROC, estatística de Kolmogorov-Smirnov e distância de Mahalanobis.

A curva ROC (*Receiver Operating Characteristics*) permite estudar a variação da sensibilidade e especificidade, para diferentes valores de corte. A área abaixo da curva ROC está associada ao poder discriminante de um modelo (HOSMER e LEMESHOW, 1989).

A estatística de Kolmogorov-Smirnov (KS) é usada na teoria estatística não paramétrica para testar se as funções de distribuição de uma variável são iguais em dois grupos (Conover, 1999). Para a regressão logística ajustada nesse estudo, a estatística é utilizada para comparar a distribuição da variável *score*, denotada por s , entre clientes e ex-clientes. O KS varia entre 0 e 1 e os valores mais altos indicam uma melhor performance.

A distância de Mahalanobis é a mais simples das três. Sabe-se que, quanto mais concentrados os clientes fiéis estiverem nos baixos *scores* e os infiéis nos altos *scores*, melhor é o desempenho do modelo. Assim como para as demais medidas, quanto maior o valor da distância de Mahalanobis, melhor é o desempenho do modelo. No entanto, essa medida tem uma desvantagem em relação às anteriores. Ela não tem um intervalo de variação limitado podendo, em princípio, variar de 0 a infinito. (HOSMER e LEMESHOW, 1989).

Coleta e Análise dos Dados

O estudo começou com o levantamento, nos sistemas da empresa, de todas as variáveis disponíveis, passíveis de relacionamento com o *churn* e/ou com o comportamento do cliente quanto à sua intenção de cancelar ou manter seu contrato ativo.

Foram levantadas inicialmente três variáveis de identificação dos contratos (número identificador, número do telefone celular e status do contrato: ativo, suspenso ou desativado) e mais 361 variáveis brutas. Essas variáveis foram analisadas e verificou-se a necessidade de agregação, cálculo de novas variáveis que fossem mais representativas, e por fim, o modelo

considerou apenas 18 variáveis. A TAB. 1 mostra a seqüência de filtros aplicados para a seleção das variáveis.

Tabela 1
Resumo dos filtros aplicados para seleção das variáveis, Minas Gerais
Out. 2004 – Jun. 2005

Tipos de variáveis	Variáveis brutas	Variáveis derivadas	Variáveis correlacionadas	Entraram no modelo	Ficaram no modelo
Tráfego	218	99	26	9	4
Fatura	29	20	9	1	1
Cadastro	17	16	12	10	4
Serviço	23	7	5	5	2
Relacionamento	74	68	51	10	7
Total	361	210	103	35	18

Fonte: Dados da pesquisa, Minas Gerais, Out. 2004 – Jun. 2005

Foram selecionadas duas amostras para a realização deste estudo: uma, denominada amostra de análise que será utilizada na proposição do índice de fidelização e outra, amostra de validação que será utilizada para verificar a consistência do índice.

A TAB. 2 apresenta o plano amostral utilizado e as margens de erro associadas a cada amostra.

Tabela 2
Plano amostral e margens de erro, Minas Gerais, Out. 2004 – Jun. 2005

Amostra/Universo	Clientes Ativos	Clientes Cancelados	Total
Amostra de análise	1.000	1.000	2.000
Erro amostral	1,32%	1,24%	0,93%
Amostra de validação	1.000	1.000	2.000
Erro amostral	1,32%	1,24%	0,93%
Universo	154.618	7.792	162.410
% clientes /Total	95%	5%	100%

Fonte: Dados da pesquisa, Minas Gerais, Out. 2004 – Jun. 2005

Ajuste dos modelos amostrais: Amostra de análise e amostra de validação

A construção do índice para prever a fidelidade de clientes envolveu, inicialmente, 35 variáveis e baseou-se em regressão logística. Como critério de seleção das variáveis componentes do modelo final da regressão logística utilizou-se os métodos *Enter*, *Forward* e *Backward*, procedimentos mediante os quais é possível encontrar o conjunto de 18 variáveis que geram o melhor ajuste do modelo.

A TAB. 3 apresenta as variáveis e seus respectivos coeficientes padronizados identificados pelo método Forward (método de máxima verossimilhança - *loglikelihood ratio method*), obtido automaticamente através de programação em SPSS.

Tabela 3

Amostra de análise: Modelo para previsão da fidelidade de clientes de telefonia celular

Código das Variáveis	Variáveis	β	Desvio padrão	Wald	Nível de Significância	Exp(β)	Intervalo de Confiança 95% para Exp(β)	
							Inferior	Superior
V1	Indicadora do serviço de Caixa Postal	0,69	0,15	20,02	0,00	1,99	1,47	2,70
V2	Tempo de base em dias	3,33	0,15	461,28	0,00	27,81	20,53	37,67
V3	Trafego Médio mensal	-0,65	0,21	9,41	0,00	0,52	0,34	0,79
V4	Quantidade de contas pagas em atraso	-0,36	0,08	19,29	0,00	0,70	0,60	0,82
V5	Média de informações	0,16	0,08	4,23	0,04	1,17	1,01	1,35
V6	Número de vezes que o nome do cliente foi enviado ao órgão de proteção ao crédito	0,34	0,12	7,44	0,01	1,40	1,10	1,78
V7	Razão entre o valor médio da assinatura e o valor médio da fatura	0,20	0,08	5,59	0,02	1,22	1,03	1,44
V8	Razão entre o tráfego total médio entrante (INCOMING) e o tráfego total médio saínte (OUTGOING)	0,52	0,11	23,45	0,00	1,69	1,36	2,08
V9	Indicadora de Tipo de pagamento -possui ou não débito automático	1,02	0,22	21,39	0,00	2,76	1,80	4,25
V10	Plano tarifário associado ao contrato	-0,28	0,12	5,20	0,02	0,76	0,60	0,96
V11	Razão entre tempo de tráfego e tempo apto a gerar tráfego	1,14	0,14	68,13	0,00	3,13	2,39	4,11
V12	Média de informação sobre cancelamento	1,29	0,10	183,03	0,00	3,65	3,02	4,40
V13	Média de contato ativo de cobrança	-0,57	0,12	21,05	0,00	0,56	0,44	0,72
V14	Média de solicitação de cancelamento retido sem custo	-2,33	0,17	195,95	0,00	0,10	0,07	0,14
V15	Média de solicitação de cancelamento retido com termo fidelidade	0,73	0,17	17,46	0,00	2,06	1,47	2,90
V16	Média de solicitação de cancelamento retido com custo	-0,32	0,16	3,84	0,05	0,73	0,53	1,00
V17	Média de solicitação de cancelamento	1,32	0,17	61,61	0,00	3,74	2,69	5,19
V18	Razão entre o tráfego TOTAL médio dos últimos 3 meses e o tráfego TOTAL médio	-0,46	0,14	10,93	0,00	0,63	0,48	0,83
	Constante	-0,25	0,23	1,16	0,28	0,78		

Fonte: Dados da pesquisa, Minas Gerais, Out. 2004 – Jun. 2005

Na regressão múltipla, o valor t é usado para avaliar a significância de cada coeficiente. A regressão logística usa uma estatística diferente, a estatística **Ward**. Ela fornece a significância estatística para cada coeficiente testado, de modo que o teste de hipóteses de que um coeficiente é diferente de zero pode ocorrer como acontece na regressão múltipla.

Ao avaliar o modelo ajustado para a amostra de análise, verifica-se, pela análise da estatística Ward, que todos os coeficientes são significativamente diferentes de zero, ao nível de 5%, exceto a constante do modelo. A estatística Ward apresentou um resultado de 1,164, com nível de significância de 28,1%, o que significa que a hipótese nula não foi rejeitada (constante é igual a zero).

Interpretação dos coeficientes dos modelos ajustados

- Se $\exp(\beta)$ tender a zero, um aumento na variável X provocará uma *diminuição* da probabilidade de *churn*, ou seja, a influência da variável X é inversamente proporcional à probabilidade de *churn*.
- Se $\exp(\beta)$ for muito maior que 1, um aumento na variável determinará um *aumento* da probabilidade de *churn*, ou seja, a influência da variável X é diretamente proporcional à probabilidade de *churn*.
- Se $\exp(\beta)$ tender a 1, significa que a variável X tem baixa influência na probabilidade de *churn*.

A partir dessas inferências, foi possível identificar as variáveis que mais oferecem risco e proteção para a base, conforme sintetiza o QUADRO 1, a seguir:

QUADRO 1
Fatores determinantes de risco e proteção para a base de clientes

Risco (coeficiente β grande e positivo)	Proteção (Coeficiente β grande e negativo)
Tempo de base	Retenção sem custo
Informação de cancelamento	Tráfego médio mensal
Solicitação de cancelamento	
Desativação de débito automático	
Razão do tempo de tráfego	

Fonte: Dados da pesquisa, Minas Gerais, Out. 2004 – Jun. 2005

Avaliação da consistência do modelo construído

A análise das porcentagens de classificação incorretas, apresentada na TAB. 4 mostra que os dois modelos têm uma melhor capacidade preditiva para os clientes ativos do que para os clientes cancelados.

Tabela 4
Acuracidade dos modelos propostos

Aplicação do modelo	Status do cliente		Total
	Clientes ativos	Clientes cancelados	
Amostra de Análise	90%	82%	86%
Amostra de Validação	92%	82%	87%
Universo	91%	82%	90%

Fonte: Dados da pesquisa, Minas Gerais, Out. 2004 – Jun. 2005

Foram posteriormente realizados testes para avaliar a qualidade do modelo, e os resultados foram positivos, confirmando sua acuracidade, conforme mostra a TAB. 5.

Tabela 5
Testes de verificação da qualidade dos modelos

Aplicação do modelo	Testes realizados		
	KS	ROC	MAHALANOBIS
Amostra de Análise	80%	0,95	5,05
Amostra de Validação	77%	0,94	4,37
Universo	78%	0,95	3,36

Fonte: Dados da pesquisa, Minas Gerais, Out. 2004 – Jun. 2005

O resultado do teste Kolmogorov-Smirnov (KS) confirma a qualidade do modelo, apresentando valores superiores a 75%, quando resultados entre 40 e 50% já seriam considerados como uma excelente discriminação.

Com relação à medida de Mahalanobis, o modelo também apresenta um bom ajuste, pois quanto maior é essa medida, melhor é a discriminação. A medida de Mahalanobis (k) representa a distância de um valor da variável X em relação à sua média, medida em unidades do desvio padrão. $K = |x - \mu| / \sigma$. Por exemplo, a $P(|x - \mu| \leq k\sigma) = 0,95$ para k igual a 2.

De acordo com os testes realizados sobre a capacidade de distinção (ROC Curve) do modelo nas duas amostras e no universo, o modelo apresenta capacidade de distinção em torno de 95%, valor considerado acima do normal. O teste ROC já revelaria uma boa aceitação, se o resultado obtido fosse um valor acima de 70%.

Através das estatísticas descritivas obtidas para a variável, a probabilidade de cancelamento de contratos, apresentadas na TAB. 6, verifica-se que a o percentual de clientes com probabilidade de cancelamento (índice de fidelidade) igual ou maior que 0,5 é de 12,31%. Além disso, ao analisar os percentis, conclui-se que 95% dos contratos apresentam uma probabilidade menor ou igual a 0,84, ou seja, a chance de cancelamento do contrato é muito alta (maior ou igual a 0,84) apenas para 5% do universo pesquisado. Isso significa que o foco da empresa, em termos de fidelização dos clientes, deve se concentrar inicialmente nessa parcela.

Tabela 6
Universo de clientes: estimativa do índice de fidelidade

Variável	Média	Variância	Percentis					% menor que 0,5	% maior ou igual a 0,5
			25	50	75	90	95		
Probabilidade de cancelamento do contrato	0,23	0,06	0,06	0,17	0,30	0,57	0,84	87,69%	12,31%

Fonte: Dados da pesquisa, Minas Gerais, Out. 2004 – Jun. 2005

Conclusões e Comentários Finais

Tendo em vista o objetivo geral deste estudo, qual seja, de construir um índice capaz de direcionar a implantação de ações que visem à fidelidade da base de clientes de uma empresa de telefonia celular em Minas Gerais, pode-se concluir que os resultados apresentados confirmam a possibilidade de se detectar preventivamente a intenção de cancelamento de contrato por parte de um cliente, com uma acuracidade de 90%. Este percentual é considerado alto, possibilitando a utilização dessa informação de forma confiável.

Foi possível gerar a probabilidade $P(Y=1)$, cuja utilização como índice de fidelidade apresenta relevante consistência. A partir desse índice estatístico, espera-se alcançar os benefícios descritos no referencial teórico, relacionados às ações de inteligência de *marketing* para tomada de decisão estratégica de manutenção de clientes e de administração do gerenciamento com os clientes - *Customer Relationship Management (CRM)*.

Além disso, a definição desse índice possibilitará não só o maior embasamento das premissas de potencial de mercado e das metas referentes à elaboração do orçamento destinado à manutenção de clientes, mas também a análise e priorização do contato ativo para a fidelização dos clientes de maior risco, bem como o planejamento de *marketing*, que poderá orientar estratégias de mídia, desenvolvimento de produtos e planejamento de venda cruzada e ainda a criação de um programa de relacionamento diferenciado para cada grupo identificado.

Quanto ao primeiro objetivo específico estabelecido (identificar as variáveis que explicam a fidelização no contexto estudado), foram inicialmente analisadas 361 variáveis, das quais apenas dezoito 18 integraram o modelo. O processo de eliminação de variáveis baseou-se tanto na transformação das 361 variáveis brutas em outras derivadas, como na realização de testes de hipóteses para avaliar a associação entre as variáveis selecionadas e a solicitação de cancelamento do contrato de adesão à operadora pelos clientes (*churn* voluntário) e ainda em análise de correlação.

Quanto ao segundo objetivo específico (avaliar a consistência do índice construído, através de dados internos) o modelo foi ajustado e validado mediante a utilização de duas amostras (análise e validação). Foram feitas duas validações (amostra de validação e universo), recomendadas por Churchill (1979), segundo o qual pelo menos dois estudos consecutivos são necessários, para que novas escalas possam prover alguma evidência de validade e de confiabilidade.

No decorrer dos trabalhos, foram identificados os seguintes pontos de atenção, que podem limitar sua aplicação:

- O modelo proposto para a construção do índice de fidelidade só é aplicável a clientes da empresa estudada (entendendo-se por modelo as variáveis e seus respectivos coeficientes β). Sua aplicação direta a outras operadoras pode não gerar os mesmos resultados de qualidade de ajuste. Entretanto, ao se replicar a mesma metodologia, é possível recriar o índice para outras operadoras.
- Outra limitação constatada é de que este modelo foi desenvolvido para operadoras de telefonia celular, não tendo como ser aplicado, no mesmo formato, a outras categorias de prestadores de serviços.
- Por fim, apesar de o índice criado levar em consideração um histórico mínimo de três meses, não se pode afirmar qual é seu prazo de validade, ou seja, o modelo é perecível, devendo ser revisto periodicamente.

Referências

- AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES – ANATEL. www.anatel.gov.br, acesso em: 13/10/2005, 22/03/2005.
- BERRY, M.; LINOFF, G. *Data Mining techniques: for marketing, sales, and customer support*. John Wiley & Sons, 1997.
- CHURCHILL, Jr., G. A. *A paradigm for developing better measures of marketing constructs*. *Journal of Marketing Research*, v.16, p.64-73, fev.1979.
- CONOVER, W. J. *Practical nonparametrics statistics*, 3 Ed. New York: John Wiley & Sons.
- ENGEL, J.F. BLACKWELL, R.D. MINIARD, PW. *Comportamento do consumidor*. Rio de Janeiro: LTC, 2000.
- GASET, P. J. C.. *O CRM não é sofrer: é uma nova filosofia*. HSM Management, São Paulo, n. 28, p. 116-121, set./out. 2001
- HAIR, JR., J. F. et al. *Análise multivariada de dados*. 5ª ed. Porto Alegre: Bookman, 2005.
- HOFFMAN, K. E.; BATESON, J. E.G.. *Princípios de marketing de serviços: conceitos, estratégias e casos*. São Paulo: Pioneira Thomson Learning, 2003.
- HOSMER, D.; LEMESHOW, S. *Applied logistic regression*. New York: John Wiley & Sons, 1989.
- JOHNSON, D. E., *Applied multivariate methods for data analysis*. New York: Duxbury Press, 1998.
- KOTLER, P. *Marketing para o Século XXI*. São Paulo: Ed. Futura, 2000.
- LOVELOCK, C.; WRIGHT, L. *Serviços, marketing e gestão*. São Paulo: Saraiva, 2001.
- MCKENNA, R. *Marketing de relacionamento*. Rio de Janeiro: Campos, 1999.
- MINGOTI, S. A. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Belo Horizonte: Editora: UFMG, 2005.
- NAVEGA, S. *Princípios essenciais do Data Mining*. Anais do Infoimagem, São Paulo: Cenadem, Novembro, 2002.
- NEWELL, F. *fidelidade.com – CRM: o gerenciamento das relações com o consumidor na era do marketing pela internet*. São Paulo: Makron, 2000.
- OLIVER, R. L. *Satisfaction: a behavioral perspective on the consumer*. Boston: Irwin/McGraw-Hill, 1996.
- OLIVER, R. L. *Whence consumer loyalty?* *Journal of Marketing*, v. 63, Special Issue, p. 33-44, 1999.

PEPPERS, D.; ROGERS, M. *CRM Series Marketing 1 to 1: Um guia executivo para entender e implantar estratégias de customer relationship management*. Peppers and Rogers Group do Brasil, São Paulo:2000.

PORTER, M. E. *Competição: Estratégias Competitivas Essenciais*. Rio de Janeiro: Campus, 1999.

REICHHELD, F. *A estratégia da lealdade: a força invisível que sustenta crescimento, lucros e valor*. Rio de Janeiro: Campus, 1996.

REICHHELD, F. *O valor da fidelidade*. Revista HSM Management. n.21, p.6 -10. Jul/Ago 2000.

ROGERS, M. SKINNER, S. *O valor dos clientes é a verdadeira medida do desempenho das empresas*.

<<http://www.agecefrj.com.br/paginas/materias/atualidades/atualidade4.html#valorclientes>>

acesso em: 23/10/2005.

SHARMA, S., *Applied multivariate techniques*. New York: John Wiley & Sons, 1996.

SCHEAFFER, R. L., MENDENHALL, W., OTT., L., *Elementary survey sampling*. 5th ed. Duxbury Press (1996)

VAVRA, T. *Marketing de relacionamento: como manter a fidelidade de clientes através do marketing de relacionamento*. São Paulo: Atlas, 1993.