



UMA APLICAÇÃO DE MINERAÇÃO DE DADOS NO GERENCIAMENTO DO CHURN EM SERVIÇOS DE BANDA LARGA DE TELECOMUNICAÇÕES

André Pedro Fernandes Neto (TELEMAR NORTE LESTE S/A)
andrepfneto@yahoo.com

Teófilo Camara Mattozo (TELEMAR NORTE LESTE S/A)
mattozo@oi.com.br

José Alfredo Ferreira Costa (UFRN)
alfredo@ufrnet.br

Resumo

O setor da economia associado ao segmento de telecomunicações é um dos mais promissores, estratégicos e dinâmicos. Atualmente o serviço de banda larga encontra-se em franca utilização em função de ter alavancado transformações na economia e na vida humana. Isso pode ser atribuído às inúmeras inovações tecnológicas que permitiram a proliferação de serviços como internet, comunicações móveis e conexões fixas de alta velocidade. Como resultado da grande competição no mercado, as empresas procuram encontrar formas diferenciadas de conseguir novos clientes, além de não perdê-los. No ano 2000, as empresas telefônicas brasileiras iniciaram um processo de comercialização baseado em um produto de grande valor agregado à telefonia fixa, o ADSL. Esse produto inicialmente surgiu como solução para o problema de saída dos clientes da telefonia fixa. A partir de 2005, com a elevação dos índices de abandono médio mensal acima de 2% levou ao questionamento de vários fatores, entre eles, de estar relacionado com causas voluntárias, por motivo deliberado ligado a qualidade dos serviços prestados. Esse fato gerou a necessidade crescente de gerenciar os parâmetros de desempenho de uma maneira mais eficiente e sistemática. Nesse cenário, foram realizados estudos utilizado-se vários métodos de seleção de variáveis, por meio de regressão linear múltipla, para relacionar os indicadores de qualidade de serviço com a taxa de abandono (Churn). Os resultados obtidos permitem sugerir que os indicadores de qualidade Tempo Médio de Reparo, Tempo Médio de Instalação, Preventivas e Taxa de Reparo podem influenciar substancialmente no abandono de clientes da planta das operadoras. Foram coletados dados mensais, de

uma grande empresa de telecomunicações que atua no Brasil, no período de janeiro a dezembro de 2006.

Abstract

The economy sector associated with the segment of telecommunications is one of the most promising, strategic and dynamic markets. Currently the service of broad band meets itself in large use in function to have started transformations in the economy and in the human life. This can be attributed to the lots of technological innovations that had allowed the increase of services as mobile internet, communications and fixed connections in high speed. As a result of the market competition, the companies look for to find differentiated ways to obtain new customers, and not losing them. In 2000, the Brazilian telephonic companies had initiated a process of commercialization based on a product of great aggregate value to the fixed telephony, the ADSL. This product initially appeared as a solution for the problem of the customers churn in the fixed telephony. From 2005, with the rise of monthly indices of average lost above 2 % took to the questions of some factors, among them to be related with voluntary causes, for deliberated reason on the quality of services. This fact conducted the increasing need to manage the parameters of performance in a more efficient and systematic way. In this way, studies had been carried using some methods of selection of variables, through a multiple linear regression, to relate the pointers of quality of services with churn rate. The obtained results suggest that the quality pointers Time Repair Average, Time Installation Average, Preventive Actions and Repair Tax can influence too much in the lost of customers in the operators plant. Monthly data had been collected from a great telecommunications company who acts in Brazil in the period from January up to December 2006.

Palavras-chaves: Medição de Desempenho; Qualidade de Serviço; Descoberta de Conhecimento em Bases de Dados; Taxa de Abandono (Churn).

1. INTRODUÇÃO

Nos últimos anos a tecnologia mudou de forma acelerada o mundo. Nesse cenário emergente, uma das alavancas de transformação da economia e da vida humana pode ser considerada a internet. Banda larga vem viabilizando alta velocidade nessa estrada de conhecimento e informação. No ano 2000 as empresas de telefonia brasileiras iniciaram um processo de comercialização baseado em um produto de grande valor agregado à telefonia fixa, o ADLS.

O aperfeiçoamento constante, em mercados competitivos é condição para o crescimento e desenvolvimento das empresas. O aumento da concorrência e dos custos associados à atração de novos clientes impõe o desafio de reter clientes cada vez mais exigentes, fazendo com que as empresas alterem seu comportamento tático e estratégico. Como resposta a esse cenário, uma das práticas de gestão que começou a ser empregada foi o gerenciamento da taxa de abandono (*Churn*). Os constantes avanços da área de tecnologia da informação têm viabilizado o armazenamento de grandes e múltiplas bases de dados. Os pesquisadores estão, ao longo das últimas décadas, tentando desenvolver formas inteligentes para analisar, interpretar e relacionar essas bases de maneira útil em benefício das organizações.

O relacionamento entre indicadores de desempenho e qualidade da gestão de serviços é relevante para o ambiente empresarial atual. As empresas prestadoras de serviços estão incluídas nesse contexto, já que a formulação das estratégias, o planejamento e aplicação das ações, programas e processos de melhoria e mudanças nas operações iniciam-se na elaboração de métricas que possam monitorar seus estágios, passados e atuais, permitindo previsões para o futuro (ARNETT, 2000).

O relacionamento entre a taxa de abandono e a correta gestão de qualidade dos serviços, caracteriza-se como um problema, pois se por um lado devem ser garantidos níveis de satisfação que fidelizem o cliente, por outro se gasta tempo e dinheiro na procura de novos clientes. Essa pesquisa foi direcionada para empresas operadoras de banda larga, tendo o objetivo de auxiliar na gestão das áreas operacionais das prestadoras de serviços, por meio da análise de performance da qualidade de serviços versus a taxa de abandono. Para isso, nesse artigo, procurou-se analisar os dados de banda larga armazenados pelas empresas de telefonia,

com o objetivo de encontrar relação entre os dados de qualidade de serviço com a taxa de abandono.

2. QUALIDADE DE SERVIÇO

O desenvolvimento desse trabalho segue as premissas defendidas como qualidade de serviço por Gronroos (1997) e Teboul (1995). Tradicionalmente, esses autores colocam três principais diferenças entre produto e serviço: serviços simultâneos, serviços intangíveis e serviços precisariam de presença do cliente para ser produzidos.

Gronroos (1993) e Giansi & Correa (1994) concordam que existem duas situações distintas em relação à visibilidade do cliente em função do serviço consumido. Uma das situações é que o cliente está presente e consome o serviço. A outra situação ocorre quando da entrega de bens, apenas parte do processo de produção do serviço é vivenciado e, portanto consumido simultaneamente pelo cliente. Entretanto é importante compreender que a parte de produção visível é a que vai influenciar a perspectiva do cliente.

Outro aspecto importante é a intangibilidade dos serviços. Correa (2002) salienta que uma implicação da intangibilidade maior ou menor dos serviços está na maior ou menor facilidade com que se avalia a qualidade do pacote de valor a ser recebido. Essa característica faz com que o serviço seja de difícil avaliação em sua qualidade pelo gestor da operação.

2.1 Engenharia de manutenção

Em qualquer setor onde a competição e a busca por cliente é questão de sobrevivência, a manutenção pode ser considerada estratégica. Na ausência da função requerida desses ativos, a do serviço, resulta em perdas e redução do lucro, afetando clientes, funcionários, investidores e a sociedade. Por esse motivo, dentro de uma nova visão estratégica, de influência na competitividade da empresa, Facina (1999) afirma: "... a manutenção agora faz parte do negócio. E não só faz parte como também influencia nos seus resultados".

As atividades da manutenção dos serviços não devem se limitar a apenas mantê-los. Devido a sua importância no setor de serviços, suas atividades e responsabilidades devem se revestir de uma maior abrangência. Nesse sentido, Xenos (1998) subdivide as atividades de manutenção em dois tipos: as atividades de manutenção propriamente ditas e as atividades de

melhoria. As atividades de manutenção, também chamadas de execução da manutenção, têm, num sentido mais restrito, a responsabilidade de "manter suas condições originais de operação e seu desempenho através do restabelecimento de eventuais deteriorações dessas condições". Essas atividades devem ser executadas pelo pessoal da manutenção e operação, com o auxílio de normas e padrões sistematizados. Essas atividades são chamadas de Engenharia da Manutenção.

2.1.1 Definição de manutenção

Como conceito adotado para essa pesquisa, a Engenharia de Manutenção, preocupa-se com a produção do serviço, atuando na capacitação do pessoal e no desenvolvimento de técnicas para atuação nos serviços e instalações em geral. Através de planejamento e gerenciamento de sistemas visa a auto-manutenção desses serviços e instalações, de modo a operar dentro dos requisitos requeridos de confiabilidade e disponibilidade, associadas ao menor custo, preservando a vida e a segurança das instalações. Busca-se ainda o aumento contínuo da produtividade e da qualidade dos serviços.

2.1.2 Tipos de manutenção

Existem duas maneiras principais de se classificar e entender os tipos ou formas usuais de manutenção. A primeira delas é classificar a manutenção de acordo com o tipo de intervenção que se faz no serviço ou instalação. Assim são identificadas manutenção corretiva, preventiva baseada no tempo, preventiva baseada em condição ou preditiva e manutenção de melhoria, com as suas definições descritas a seguir:

- Manutenção corretiva - intervenção decorrente de uma falha, quebra ou mau funcionamento. Um serviço exige manutenção corretiva quando é necessário intervir no mesmo porque interrompeu ou degradou sua função;
- Manutenção preventiva baseada em tempo - intervenção feita a intervalos regulares de tempo corrido (por exemplo, semanas) ou de funcionamento (por exemplo, horas trabalhadas). Em inglês conhecida como TBM (time based maintenance);
- Manutenção preventiva baseada em condição ou preditiva – intervenção feita de acordo com o acompanhamento de determinados parâmetros do serviço (por exemplo,

medição de desgaste ou elevado grau de degradação no serviço). Em inglês conhecida como CBM (condition based maintenance);

- Manutenção de melhoria - intervenção feita para alterar as condições de um serviço com o objetivo de aumentar o seu rendimento, a qualidade dos produtos processados ou melhorar algum parâmetro operacional.

3. DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS

Para Fayyad (1996), o conceito de descoberta de conhecimento em bases de dados pode ser resumido como o processo não-trivial de identificar padrões novos, válidos, potencialmente úteis e, principalmente, compreensíveis em meio às observações presentes em uma base de dados. Contudo o objetivo último da descoberta do conhecimento em bases de dados não é o de simplesmente encontrar padrões e relações em meio à imensa quantidade de informação disponível em bases de dados, e sim a extração de conhecimento inteligível e imediatamente utilizável para o apoio às decisões.

Os constantes avanços da área de tecnologia da informação têm viabilizado uma elevada quantidade de armazenamento de dados. As bases de dados podem ser tratadas seguindo um processo de descobertas do conhecimento em etapas. Como o conhecimento inicialmente está na forma de dados sendo passada para uma segunda fase denominada de informações e finalmente obtem-se, a partir dos dados, o conhecimento. O processo de descoberta de conhecimento em base de dados pode ser dividido em três etapas: Pré-processamento, Mineração de Dados e Pós-processamento (GOLDSCHMIDT et al, 2005). A etapa de pré-processamento compreende as funções relacionadas à captação, à organização e ao tratamento de dados. É na etapa de pré-processamento que se prepara os dados para a etapa de mineração onde é realizada a busca efetiva por conhecimentos úteis no contexto da pesquisa em questão. A etapa de pós-processamento abrange o tratamento do conhecimento obtido na mineração.

3.1 Modelando dados com uso de análise de regressão multivariada

Regressão é o termo utilizado para designar uma equação matemática que descreva as relações entre duas ou mais variáveis. Regressão linear é um método para se estimar o valor esperado de uma variável Y (variável dependente), dados os valores de algumas outras

variáveis X (variáveis independentes). Assim, dadas duas matrizes de dados, X e Y , a finalidade da regressão é construir um modelo $Y = f(X)$. Tal modelo tenta explicar, ou prever, as variações em Y dada as variações em X . A regressão multivariada leva em consideração as diversas variáveis preditivas simultaneamente, modelando a variável dependente com mais exatidão. Nesse trabalho, a variável dependente são as vendas efetivas e o grupo de variáveis independentes são os indicadores do desempenho de vendas. O modelo de regressão é representado pela equação 1.

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i. \quad (1)$$

Em que Y_i – representa a variável dependente, x_{ik} ($i = 1, \dots, n$) são as variáveis independentes ($k = 1, 2, \dots, p$); β_i 's são os coeficientes da regressão (parâmetros desconhecidos no modelo – a serem estimados); ε_i é o resíduo, variável aleatória que captura a parcela do comportamento da variável Y_i não explicada pela equação da regressão.

Os parâmetros de um modelo da regressão podem ser estimados de várias formas:

- 1) Mínimos quadrados, minimizando o erro quadrático médio dos resíduos;
- 2) Máxima verossemelhança;
- 3) Métodos Bayesianos;
- 4) Minimizando o desvio absoluto.

Os métodos 1 e 2 coincidem para um modelo com os erros normalmente distribuídos.

Estimativas dos mínimos quadrados, usados nesse trabalho, são dadas por (LAROSE, 2006)

$$\hat{\beta} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}. \quad (2-a)$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}. \quad (2-b)$$

O estimador de mínimos quadrados, na forma matricial, é dado por $\beta = (X'X)^{-1}(X'Y)$, onde o apóstrofo significa transposto. Cada observação tem seu próprio resíduo, que somados produzem a soma dos erros quadráticos, uma medida total dos erros da estimação. Três somas quadráticas (SSE, soma quadrática dos erros; SSR, a soma dos quadrados da regressão; SST, a soma total dos quadrados) podem ser calculadas como segue:

$$SSE = \sum (y_i - \hat{y}_i)^2. \quad (3-a)$$

$$SSR = \sum (\hat{y}_i - \bar{y})^2. \quad (3-b)$$

$$SST = \sum (y_i - \bar{y})^2. \quad (3-c)$$

A estatística da regressão pode ser apresentada sucintamente com uso de tabelas da análise de variância (ANOVA). Erros médios (por exemplo, MSE e o MSR) são derivados da equação 3. Um parâmetro importante é o coeficiente de determinação múltipla, que é definida como:

$$R^2 = \frac{SSR}{SST}. \quad (4)$$

Para a regressão múltipla, R^2 é interpretado como a proporção da variabilidade na variável alvo que é esclarecido no relacionamento linear com o conjunto de variáveis preditoras.

4. TAXA DE ABANDONO DE CLIENTES (*CHURN*)

A Taxa de Abandono ou *Churn* consiste no ato de um cliente abandonar uma empresa, geralmente em favor de um concorrente, terminando toda a sua relação com a antiga empresa (IKEDA, 2006). O termo ainda não tem uma tradução usual para o português, mas nas empresas de banda larga causa exatamente o que o verbo quer dizer: uma grande “agitação” de clientes no mercado, trocando de empresa a todo o momento, o que, por sua vez, leva as empresas a se “mexerem” em busca de novas formas de manter seus clientes no seu negócio, ao mesmo tempo em que buscam seduzir os clientes da concorrência (FERREIRA, 2005).

Estatísticas revelam que as empresas americanas perdem metade de seus clientes em cinco anos e que, nessa taxa de perda de clientes, as empresas têm seus desempenhos financeiros (performance) reduzidos de 25% a 50%. (Reichheld; Teal, 1996 apud Ganesh et al, 2000). *Churn* trata-se de um termo muito comum na indústria de telecomunicações (LINOFF & BERRY, 2000). Quando se diz que a taxa de *churn* de uma empresa varia entre 10% a 15% anualmente, deseja-se realmente dizer que essa empresa está perdendo 10% a 15% dos clientes que se encontram na sua base de dados, por fazerem ou já terem feito negócio com ela e que, por algum motivo, não ficaram satisfeitos e evadiram de sua base de clientes ativos.

Pesquisadores levantaram alguns fatores de influência correlacionados positivamente com a manutenção de clientes (VENETIS & GHOURI, 2004; GANESH et al, 2000):

- Qualidade de serviços prestados aos clientes;
- Satisfação do cliente;

- Lealdade do cliente;
- Comprometimento do cliente – definido como “a medida do quanto às partes em uma relação de troca deseja continuar o relacionamento de valor” (VERHOEF, 2003);
- Tratamento dados pelos funcionários.

O gerenciamento do *churn* é, justamente, o desenvolvimento de técnicas que permitam à empresa manter seus clientes mais lucrativos. Para Neslin et al, (2006), uma forma de gerenciar o *churn* é prever quais clientes com mais probabilidade de abandonar o relacionamento com a empresa e trabalhar com eles a fim de tentar evitar a ocorrência desse rompimento. Para isso, é preciso que a empresa seja capaz de realizar tal tipo de previsão e identificar, deter os clientes que provavelmente se desligarão, quais são os que, de fato, geram valor suficiente para a empresa e justifiquem o investimento de retenção.

5. METODOLOGIA

O processo discutido por Hair et al, (2005) consiste nos seguintes estágios:

a) Definição do problema da pesquisa, dos objetivos e da técnica multivariada a ser usada.

O ponto de partida para uma análise multivariada de dados é definir o problema da pesquisa e os objetivos de análise em termos conceituais, antes de especificar qualquer variável ou medida. Com o objetivo e o modelo conceitual especificados, deve-se escolher a técnica multivariada a ser utilizada. Após escolher entre um método de dependência ou independência, a última decisão é selecionar a técnica em particular com base nas características de medidas das variáveis dependentes e independentes.

b) Planejamento da pesquisa

Para cada técnica deve ser desenvolvido um plano de análise que aborde as questões particulares a seu propósito e projeto. As questões incluem considerações gerais, como tamanho mínimo da amostra, tipos permitidos ou exigidos de variáveis e métodos de estimação, além de aspectos específicos, como o tipo de medidas de associação de resultados agregados ou desagregados em análise conjunta ou uso de formulações especiais de variáveis

para representar efeitos não-lineares ou interativos em regressão. Em cada caso, essas questões resolvem detalhes específicos e exigências para a coleta dos dados.

c) Suposições em análise

Todas as técnicas multivariadas têm suposições inerentes, estatísticas e conceituais, que influenciam muito suas habilidades para representar relações multivariadas. As questões básicas a serem respondidas estão ligadas às suposições do modelo escolhido: linearidade do fenômeno medido, variância constante dos termos de erro, independência dos termos de erro e normalidade da distribuição dos termos de erro.

d) Estimação do modelo e avaliação do ajuste geral do modelo

Nesse estágio devem ser cumpridas três tarefas básicas: selecionar um método para especificar o modelo a ser estimado, avaliar a significância estatística do modelo geral na previsão da variável estatística e determinar se algumas das observações exercem uma influência indevida nos resultados.

e) Interpretação das variáveis estatística pesquisadas

Nesse estágio deve-se examinar a equação preditiva, e com isso avaliar a importância relativa que as variáveis individuais na previsão geral do produto. O pesquisador tem como função interpretar a variável estatística de regressão pela avaliação dos coeficientes de regressão estimados em termos de sua explicação da variável dependente, não se deve avaliar tão somente o modelo de regressão estimado, mas também as variáveis independentes potenciais que foram omitidos, se uma busca seqüencial ou abordagem combinatória foi empregada.

f) Validação dos resultados.

Essa fase consiste em generalizar a aplicação do modelo, demonstrando que ele não é específico as observações utilizadas na estimação. Deve-se garantir que ele represente a população geral e que sejam apropriadas as situações nas quais será usada.

6. IMPLEMENTAÇÃO E ANÁLISE DOS RESULTADOS

O processo de construção do modelo em estágios, como foi visto anteriormente, será utilizado como referência para discutir os fatores que impactam na criação, estimação, interpretação e validação da análise de regressão.

6.1 Objetivo

Considerando que o objetivo proposto é uma relação de dependência, há necessidade de se especificar e selecionar as variáveis dependentes e independentes. Descreve-se abaixo os indicadores de internet banda larga utilizados nas empresas de telecomunicações.

Tenta-se entender como o cliente percebe a qualidade dos serviços prestados e como isso pode afetar diretamente a taxa de abandono da planta. Isso pode ser feito por meio da análise de uma equação em que a taxa de abandono da planta possa ser relacionada com alguns indicadores de qualidade dos serviços.

Tabela 1: caracterização das variáveis dependente e independente

Indicador	Formula dos Indicadores	Valores	Tipo de Variável
TAXA PREVENTIVA	$\frac{\sum \text{dos Reparos Preventivos}}{\sum \text{dos Reparos Corretivos do Mês Anterior}}$	Percentual	Independente
TAXA DE ABANDONO	$\frac{\text{Quantidade de Retiradas} \times 100}{\text{Base em Serviço do Mês Anterior}}$	Percentual	Dependente
IGT	$\frac{\text{Reparos Aberto nos últimos 30 dias} \times 100}{\text{Total de Instalação Realizadas nos Últimos 30 dias}}$	Percentual	Independente
TEMPO MEDIO REPARO	$\frac{\sum \text{Tempos de Reparo no mês}}{\text{Base de Terminais Reclamados no mês}}$	Percentual	Independente
IPGC	$\frac{\text{Instal. Executados em até 3 dias corridas} \times 100}{\text{Total de Instalação no Período}}$	Percentual	Independente
RPAV	$\frac{\text{Reparos Executados no Prazo (4 h)} \times 100}{\text{Total de Reparos no Mês}}$	Percentual	Independente

RPDC	<u>Reparos Executados no Prazo (8 h) X 100</u> Total de Reparos no Mês	Percentual	Independente
RPT	<u>Reparos Executados no Prazo X 100</u> Total de Reparos no Mês	Percentual	Independente
RRAV	<u>Reparos Reincidente < 90 dias Alto V. X 100</u> Total de Reparos no Mês	Percentual	Independente
RRDC	<u>Reparos Reincidente < 90 dias D. C. X 100</u> Total de Reparos no Mês	Percentual	Independente
RRT	<u>Reparos Reincidente < 90 dias X 100</u> Total de Reparos no Mês	Percentual	Independente
TAXA DE REPARO	<u>Quantidade Reparos em 30 dias X 100</u> Base em Serviço Mês	Percentual	Independente
TEMPO DE INSTALAÇÃO	<u>∑ Tempos de Instalação no mês</u> Base de Terminais instalados no mês	Percentual	Independente

Fonte: Elaborada pelo autor

A regressão múltipla foi selecionada como a técnica multivariada a ser utilizada porque fornece um meio de avaliar objetivamente o grau e caráter da relação entre as variáveis dependentes e independentes, pela formação da variável estatística a partir da variável independente. Nesse artigo escolheu-se a ferramenta SPSS pela sua facilidade de utilização, pela sua vasta literatura e por ser bastante usada na análise de regressão múltipla.

6.2. Planejamento da pesquisa

Foram obtidos um total de 336 observações ocorridas em uma grande empresa de telecomunicações existente no Brasil, com atuação nas regiões nordeste, norte e sudeste. A fim de comparar diversos modelos da regressão, é feito inicialmente um estudo com o método de busca por meio de uma abordagem combinatória que é um processo de busca generalizada em todas as possíveis combinações das variáveis independentes pelo método Combinatório (*Enter*). Por último, é utilizado o método de busca sequencial para estimar a equação de regressão com um conjunto de variáveis sendo então acrescentado seletivamente ou eliminado

variáveis até que uma medida de critério geral seja alcançada, sendo denominado Sequencial (*Stepwise*).

No método *Enter* um bloco (ou todas) são selecionadas e incorporadas ao modelo em uma única etapa. No método *Stepwise*, em cada etapa, a variável independente ainda ausente da equação são incorporadas caso possuam valores baixos da estatística *F*. As variáveis já na equação da regressão são removidas se sua probabilidade de *F* se tornar suficientemente grande. O método conclui quando não mais há variável elegível para a inclusão ou a remoção. O nível de tolerância usado foi 0,05.

6.3 Suposições em análise de regressão linear múltipla

Fazendo-se análise da matriz de correlação entre as variáveis do modelo, pode ser verificado que, entre as variáveis independentes, não existem valores superiores a 0,5, conforme dados obtidos a partir do *SPSS* (NORUSIS, 2004). Ou seja, não existindo nenhum valor $r_{x_i, x_j} \geq 0,5$, pode ser concluído pela inexistência de multicolinearidade. Os dados obtidos mostram também que os valores do Fator de Inflação de Variância são bastante próximos de um, pelo que se pode concluir pela não existência de Multicolinearidade.

No histograma apresentado na figura 1, com o ajustamento à distribuição normal, pode ser verificado que existe uma distribuição consistente dos resíduos em relação à distribuição teórica, ou seja, segue aproximadamente a forma da curva normal. Existe apenas uma elevação um pouco maior, nomeadamente na zona central da distribuição, sendo aceitável perto da curva normal, porém com a existência de pontos residuais positivos.

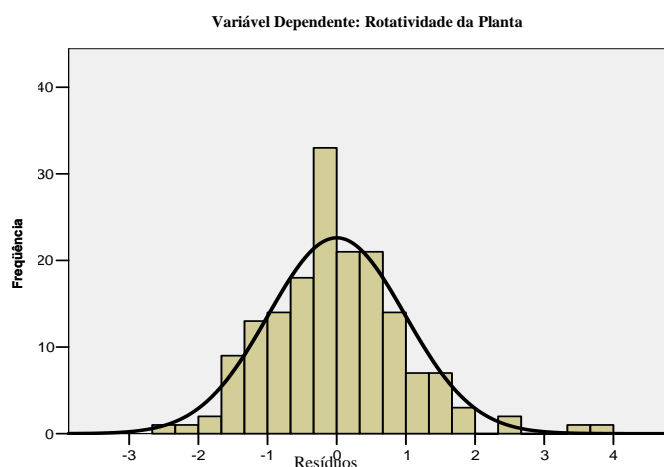


Figura 1 – Histograma de resíduos da variável dependente do método sequencial.

Fonte: Relatório do SPSS.

Na figura 2 nota-se uma tendência de distribuição uniforme, ou seja, os pontos encontram-se bastante próximos de uma reta, sem desvios substanciais ou sistemáticos. Não há razão para duvidar da normalidade dos erros, sendo os resíduos considerados representativos de uma distribuição normal. Nem o histograma nem o gráfico *PP-plot* indicam que a suposição da Normalidade foi violada.

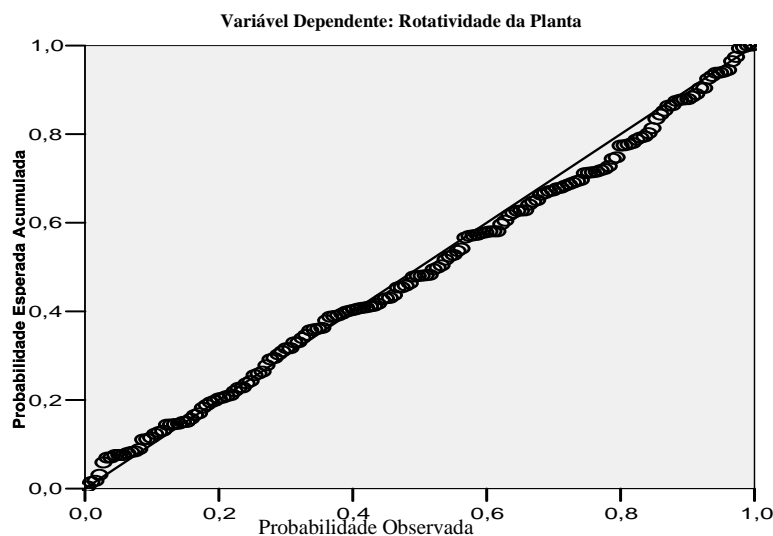


Figura 2– Probabilidade normal PP-plot dos resíduos do método seqüencial.

Fonte: Relatório do SPSS.

6. 4 Estimação do modelo de regressão, avaliação do ajuste geral e da significância estatística do modelo

Detalha-se o método seqüencial *Stepwise* pois foi o que melhor se adaptou aos objetivos dessa pesquisa. A principal vantagem em relação aos outros métodos seqüenciais é a sua habilidade em acrescentar ou eliminar variáveis em cada estágio.

6.4.1 Coeficiente de determinação

São mostrados os dados obtidos através do método *Stepwise*. A partir de seus respectivos valores, obtidos através dos dados do SPSS (sumário do modelo), foi considerado adequado em desconsiderar o modelo de especificação combinatório e a utilizar o modelo de busca seqüencial *stepwise*. O método seqüencial mostrou-se mais adequado devido ser mais

objetivo para selecionar as variáveis que maximizam o poder de explicação com o menor número de variáveis empregadas.

6.4.2 Equação estimada

O principal objetivo dos modelos de regressão é a estimação dos parâmetros desconhecidos β . A esse processo é comum chamar-se de “ajuste do modelo aos dados”. Os coeficientes do modelo *stepwise* contém as estimativas dos parâmetros e respectivas estimativas do erro padrão, as estimativas dos coeficientes padronizados e o valor da estatística *t-Student*. Por meio dos dados do SPSS (coeficientes do modelo), pode ser caracterizada a equação de regressão que pode ser vista abaixo:

$$\text{Taxa de Abandono} = 0,53(\text{TMI}) + 0,41(\text{TRP}) + 0,60(\text{TMR}) - 0,39(\text{QRPR}) \quad (6)$$

Onde :

TMI – Representa o tempo médio gasto na instalação do circuito, ou seja, é o tempo desde abertura da ordem de serviço até o seu efetivo fechamento.

TRP – Representa a taxa de reparo executado no prazo acordado com o cliente esse prazo varia de 8 horas para clientes Alto Valor e 24 horas para os Demais Clientes geralmente são os clientes varejo.

QRPR – Esse valor representa a quantidade de reparos preventivos efetuados nas empresas.

TMR – Representa o tempo médio gasto no reparo do circuito, ou seja, é o tempo desde abertura do Bilhete de Defeito até o seu efetivo fechamento.

6.4.3 Teste de hipótese para a significância do modelo

Esse teste é usado para estabelecer se existe ou não alguma relação entre a variável dependente e o conjunto de variáveis independentes. Consiste em testar as seguintes hipóteses:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \beta_k \neq 0 \text{ para algum } k$$

A estatística do teste tem, sob H_0 (a hipótese H_0 significa que a regressão não é significativa), a distribuição F com p e (n-p) graus de liberdade. A rejeição da hipótese H_0

indica a significância de regressão. As quantidades necessárias para calcular o valor observado dessa estatística estão dispostas nos dados do SPSS (análise de variância – ANOVA).

Como $F_{obs} = 43.367$ é claramente superior a $F_c = F_{(95\% ; 4 ; 163)} = 2,40$ (valor percentual de uma distribuição F com 4 graus de liberdade no numerador e 163 no denominador) deve ser rejeitada H_0 em favor da hipótese alternativa de que a regressão é estatisticamente significativa. Em outras palavras, pode-se dizer que ao nível de significância de 0,05, rejeita-se a hipótese da regressão não ser significativa. Então há evidência para afirmar que existe um relacionamento linear entre o conjunto de variáveis explicativas e a taxa de abandono.

6.4.4 Teste de hipótese para o parâmetro β_k

Após a verificação de que pelo menos um dos parâmetros β_k é significativo, deve-se testar a significância de cada um deles, isto é, para cada parâmetro β_k ($k=1, \dots, p$), testam-se as hipóteses:

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

O teste para cada parâmetro é feito utilizando a estatística t de Student com $(n-p)$ graus de liberdade, o desvio padrão amostral $S(b_k)$ e o estimador b_k . A estatística do teste é dada por:

$$t^* = \frac{b_i}{s(b_i)} \cap t_{(n-p)} \quad i = 1, \dots, p$$

Se $|t^*| \leq t(1-\alpha/2; n-p)$, o teste não rejeita H_0 ; caso contrário o teste rejeita H_0 em favor de H_1 . A rejeição de H_0 indica uma contribuição significativa da variável independente X_k no modelo. Através dos resultados apresentados na ANOVA se obtém os dados da estatística t para os 4 parâmetros a serem testados. Para verificar que a variável TMR não é muito útil, isto é, que sua influência sobre a variável dependente não é significativa testa-se a hipótese nula do coeficiente para essa variável cujo valor crítico é $t_c \cong 2,258$ (para os graus de liberdade são 4 e 163).

Como $|t_c| \cong 2,258 < |t| = 9,051$ conclui-se que, ao nível de significância de 0,05, há evidência para dizer que a variável TRM é útil na explicação da taxa de abandono, ou seja, a variável TMR contribui para a explicação da taxa de abandono. Por meio dos dados

observados a partir da referida tabela e fazendo-se comparação, similares ao que foi feito anteriormente, com o valor tabelado cujo valor crítico de $t_c \cong 2,258$ pode ser concluído que todas as estatísticas t correspondentes aos parâmetros β_1 , β_2 , β_3 e β_4 podem ser validadas, ou seja, as variáveis TMI, TMP, TMR e QRPR são estatisticamente significativas. Nota-se que no modelo escolhido, as variáveis têm pesos bem distribuídos em relação à variável estatística, não tendo nenhuma um peso consideravelmente muito maior do que um outro.

6.5 Interpretação das variáveis estatísticas

Pode-se interpretar os coeficientes do modelo *Stepwise* da seguinte forma: o valor do coeficiente para a variável Tempo Médio de Instalação TMI é de 0,53, o que significa que, isolando-se as outras variáveis da equação, para um aumento de 10% normatizado na variável TMI ocorre um aumento normatizado de 5% na taxa de abandono, que em valores não normatizados se refere a um aumento em média de 2500 clientes na taxa de abandono e um aumento em média de 10 horas no tempo de instalação. De modo similar os outros indicadores podem ser interpretados.

7. CONCLUSÕES E RECOMENDAÇÕES

O objetivo desse trabalho foi tentar explicar o comportamento do cliente no setor de telecomunicações, usuário de internet de banda larga, por meio de uma equação que representa-se a taxa de abandono em função dos indicadores de qualidade de serviço. Para isso foi necessário caracterizar a taxa de abandono, bem como suas causas. Há dois tipos de causas associadas à taxa de abandono: causas voluntárias e causas involuntárias (FERNANDES, A.P., 2007). Essa pesquisa procurou se deter apenas nas causas voluntárias, ou seja, desligamento por motivo deliberado ligado à qualidade. Podem-se destacar como principais contribuições desse trabalho:

a) Relacionamento dos indicadores de qualidade de serviços de banda larga com a taxa de abandono:

Um passo inicial nesse sentido foi realizar um estudo sobre o processo de descoberta de conhecimento em banco de dados para, a partir daí, analisar os dados da pesquisa,

normalizá-los e selecionar o método para especificação do modelo da equação final. Para o equacionamento houve preocupação com a avaliação da significância estatística e a interpretação das variáveis. O resultado alcançado desse relacionamento, que está caracterizado por meio da equação 6, sinalizou que alguns indicadores de qualidade devem ter, por partes dos gestores de manutenção das empresas de telecomunicações, uma maior atenção, pois são esses indicadores que efetivamente podem influenciar na taxa de abandono.

Após o entendimento da aplicabilidade da equação estatística, foi realizada uma análise gerencial dos dados encontrados, em que se verificou a existência de uma correlação entre os indicadores de qualidade de serviço e a taxa de abandono, podendo estar relacionada a causas voluntárias por motivo deliberado ligado à qualidade. Essas causas foram associadas a quatro indicadores de qualidade: Tempo Médio de Reparo, Tempo Médio de Instalação, Preventivas e Taxa de Reparo.

b) Análise do conjunto de variáveis que exercem maior relevância para o problema abordado:

Por meio da análise das 12 variáveis identificou-se que Tempo Médio de Reparo, Tempo Médio de Instalação, Preventivas e Taxa de Reparo possuem uma maior relevância sobre a taxa de abandono. Consequentemente existem indicativos que devem ser focados que permitem diminuir a taxa de abandono, os quais são comentadas em seguida:

Diminuição da Taxa de Reparo, ou seja, o cliente que utiliza banda larga é exigente com o serviço prestado, não aceitando paralisações. Para isso deve-se agir na redundância de sistemas, procurando evitar as causas comuns de reparos como quedas de *backbone*, falta de conectividade, etc.;

O tempo de restabelecimento do serviço deve ser o menor possível. Esse fator foi medido pelo tempo médio de reparo, e foi demonstrado que quanto maior for o tempo para remover a falha, maior é a taxa de abandono. Nesse caso devem ser procuradas ferramentas de recuperação remota de falhas no intuito de diminuir o Tempo de Reparo;

Em relação aos Reparos Preventivos, a pesquisa sinalizou que a execução sistemática de manutenção preventiva é uma boa prática de retenção, pois a antecipação das possíveis falhas conta positivamente. Sendo assim, surge a necessidade de se implementar sistemas de gerência e análise da degradação do sinal, para que, a partir dessas análises, sejam gerados reparos preditivos com a finalidade de antever as falhas;

Por fim, o Tempo Médio de Instalação, que afeta a percepção de qualidade do cliente antes mesmo que ele seja cliente. A análise mostra que quanto menor for esse tempo, menor serão as taxas de abandono. Caso ocorra algum problema nos processos de instalação ou que as suas expectativas tenham valores diferentes das possibilidades de atendimento das operadoras de serviço, fica como sugestão manter o cliente bem informado e que os atrasos e novos prazos sejam sempre negociados com o mesmo.

8 - REFERÊNCIAS

ARNETT, Dennis B; Menon, Anil, Wilcox, James B. “*Using Competitive Intelligence: Antecedents and Consequences*”, *Competitive Intelligence Review*, Vol. 11(3), 2000.

CORREA, H. L. *The Links Between Uncertainty, Variability of Outputs and Flexibility in Manufacturing Systems*. University of Warwick, Warwick, 2002.

FACINA, Taís. *Manutenção Sem Fronteiras*. Revista Manutenção. Julho e Agosto, 1999.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R. “*Advances in Knowledge Discovery & Data Mining*”, Cambridge, MA (The AAAI Press/The MIT Press), 1996.

FERNANDES, A. P. *Análise dos Indicadores de Qualidade versus Taxa de Abandono Utilizando Método de Regressão Múltipla para Serviço de Banda Larga* (Dissertação de Mestrado). Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal do Rio Grande do Norte, Natal, 2007.

FERREIRA, J. B. *Mineração de Dados na Retenção de Clientes em Telefonia Celular*. Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro - PUC, Rio de Janeiro, 2005.

GANESH, J. ARNOLD, M. J.; REYNOLDS, K. E. *Understanding the Customer Base of Service Providers: Na Examination of the Differences Between Switchers and Stayers*. *Journal of Marketing*, Vol. 64, 2000.

GIANESI, I. G. N., CORREA, H. L. *Administração estratégica de serviços – operações para a satisfação do cliente*. São Paulo: Atlas, 1994.

GOLDSCHMIDT, Ronald; PASSOS, Emmanuel. *Data Mining – Um Guia Prático, Conceitos, Técnicas, Ferramentas, Orientações e Aplicações*. São Paulo: Campus, 2005.

GRONROOS, Christian. *Value-driven relational marketing: from products to resources and competencies*. *Journal of Marketing Management*, Vol.13, n.5, 1997.

_____. *Marketing Gerenciamento e Serviço A Competição por Serviço na Hora da Verdade*. 4º edição, São Paulo: Campus, 1993.

HAIR, J.; ANDERSIN, R. E.; TATHAM, R. L. *Análise Multivariada de Dados*. São Paulo: Bookman, Vol.1, 2005.

IKEDA, Ana Akemi. *Segure o Churn! In: LOVELOCK, Christopher; WIRTZ, Jochen. Marketing de Serviços: pessoas, tecnologia e resultados*. 5o ed. São Paulo: Editora Prentice Hall, 2006

LAROSE, Daniel T. *Data Mining Methods and Models*. a John Wiley & Sons, inc; 2006.
LAPIERRE, J. *Customer - Perceived Value in Industrial Contexts*. Journal of Business & Industrial Marketing, Vol.15, n. 2/3, 2000.

LINOFF, Gordon.; BERRY, Michael; “*Mastering Data Mining: The Art and Science of Customer Relationships Management*”, John Wiley & Sons Inc, USA, 2000.

NEELY, A. et. al. *Designing Performance Measures: A Structured Approach*. International Journal of Operations & Production Management, Vol. 17, n.11, 1997.

NESLIN, Scott A.; GUPTA, Sunil; KAMAKURA, Wagner, LU, Junxiang; MASON, Charlott H. *Defection detection: measuring and understanding the predictive accuracy of customer churn models*. Journal of Marketing Research, Chicago, Vol. 43, n.2, 2006.

NORUSIS, M. *SPSS 13.0 Statistical Procedures Companion*. Upper Saddle-River, N.J.: Prentice Hall, Inc.. 2004.

PINTO, Alan Kardec; XAVIER, Júlio Nassif. *Manutenção: Função Estratégica*. Rio de Janeiro: Qualitymark, 1998.

PIATETSKY-SHAPIRO, “*Measuring Lift Quality In Database Marketing*”. Journal Of SIGKDD Exploration Newsletter, Vol. 2, New York; 2000.

REICHHELD, Frederick. *A estratégia da Lealdade: a força invisível que mantém clientes e funcionários e sustenta crescimento, lucros e valor*. Rio de Janeiro: Campus, 1996.

SPSS (2007). *Statistical Package for Social Sciences*. SPSS v. 13.0. URL: <http://www.spss.com>. Acesso em novembro, 2007.

TÉBOUL, James. *A Era dos Serviços – Uma Abordagem de Gerenciamento*. Rio de Janeiro: Qualitymark, 1995.

VENETIS, K. A.; GHAURI, P. N. *Service quality and customer retention: building long-term relationships*. European Journal of Marketing. Vol. 38, n. 11/12, 2004.

VERHOEF, P. *Understanding the Effect of Customer Relationship Managment Efforts on Customer Retention and Customer Share Development*. Journal of Marketing. Vol. 67, n. 10, 2003.

XENOS, Harilaus Georgius d'Philippos. *Gerenciando a Manutenção Produtiva.* Belo Horizonte: Desenvolvimento Industrial, 1998.