

APLICABILIDADE DO PROCEDIMENTO *CHI-SQUARE AUTOMATIC INTERACTION DETECTION (CHAID)* PARA SEGMENTAÇÃO DO MERCADO DE CONSUMIDORES RESIDENCIAIS DE ENERGIA ELÉTRICA

Resumo:

Este trabalho teve como objetivo propor a aplicação do procedimento *Chi-Square Automatic Interaction Detection (CHAID)* para segmentação do mercado de consumidores residenciais de energia elétrica. Esse procedimento explora e analisa a classificação de variáveis categóricas e institui o maior valor da estatística Qui-quadrado (χ^2), de modo a estabelecer todas as partições em subgrupos possíveis, para cada variável independente. Utilizou-se uma base de dados de 628.339 consumidores obtidos junto ao cadastro de clientes de uma distribuidora de energia elétrica considerada modelo no mercado nacional. O estudo é de natureza quantitativa e caracteriza-se como exploratório descritivo. Embora a segmentação clássica do setor elétrico tenha sua relevância para a distribuição de energia, observa-se que a aplicação do *CHAID* para segmentação apresenta resultados mais favoráveis para as decisões de marketing, principalmente para aquelas distribuidoras que procuram investir em novos negócios não associados à energia. Para melhorar as características dos grupos encontrados, sugere-se a aplicação de bases psicográficas em cada um dos agrupamentos.

Palavras-chaves:

Segmentação, *CHAID*, distribuidoras de energia.

Abstract:

This project aims to propose the application of the Chi-Square Automatic Interaction Detection (CHAID) procedure to segment the home consumer market of electricity. This procedure explores and analyses the classification of categorical varieties and institutes a greater statistic value Qui-quadrado (χ^2), so that it establishes all the partitions in possible sub-groups, for each independent variable. It was used a Data Base of 628.339 consumers from a well know Electricity Company which is a model of distributor nationwide. This is a quantity study and is characterized as exploring and describing. Although the classic segmentation of the Electric Sector has its relevance to the Electrical Distribution, the application of the CHAID to the segmentation present more favorable results to marketing decisions, mainly for those distributors which intend to invest in new different businesses. To improve the characteristics of the found groups, we suggest the application of psychographic bases in each of the groups.

Keywords:

Segmentation, CHAID, Electricity Companies.

Francisco Vicente Sales Melo

Mestre em Administração e Controladoria pela UFC com concentração em marketing, especialista em Pesquisa Científica pela UECE, bacharel em Administração de Empresas pela UNIFOR e tem formação superior em Gestão Estratégica de Empresas pela FIC. Atualmente atua como coordenador e professor dos cursos de Administração e Processos Gerenciais da Faculdade Ateneu.

vicsmelo@gmail.com

Luiz Carlos Murakami

Doutor e mestre em Administração de Empresas pela EAESP/FGV com concentração em Marketing e graduado em Economia pela USP. Realizou curso de extensão de Pós-Graduação em Marketing de Entretenimento na Universidade de Southern Califórnia. Atualmente é professor adjunto da FEAAC/UFC.

murakami@ufc.com

1 – INTRODUÇÃO

O setor elétrico brasileiro passou por intensas transformações em 1998 devido às privatizações. Tais transformações visavam reduzir custos e os impactos ambientais incorridos na produção de energia elétrica. Outros objetivos dessas mudanças eram estimular a competição na geração e na comercialização de energia, além da introdução de mecanismos para incentivar a regulação destes segmentos que se caracterizam como monopólio natural (transmissão e distribuição), procurando ainda criar mecanismos para garantir o funcionamento eficiente do setor elétrico (PIRES, 1999).

Segundo Pires (1999), o modelo proposto pelo governo para as privatizações adotou uma estratégia gradual, pretendendo, de forma gradativa, reduzir a dívida pública, melhorar a eficiência produtiva e resgatar a capacidade de investimento das empresas. O governo priorizou a venda das empresas distribuidoras, pois entendia que dificilmente conseguiriam atrair interessados para os ativos de geração. Desta forma, além de privatizar as distribuidoras federais Light e Escelsa, o governo estimulou a venda de distribuidoras estaduais, com ajuda do Banco Nacional de Desenvolvimento Econômico e Social (BNDES). Do ponto de vista estratégico do governo o processo foi tão eficiente que hoje praticamente todas as distribuidoras são geridas pela iniciativa privada.

Passados esses anos, depois de tantas mudanças, atualmente observa-se que as empresas do setor elétrico brasileiro, mais especificamente as distribuidoras, estão bem consolidadas e procuram se aperfeiçoar para melhorar a prestação dos serviços ao cliente. Não é simples desenvolver estratégias atraentes aos consumidores de energia elétrica, já que este serviço é considerado como essencial e muitas vezes surpreendê-los não é tarefa fácil. Acredita-se que o primeiro passo é conhecer suas características e comportamentos para que as estratégias de marketing sejam, melhor, direcionadas. Feito isso, fica menos difícil aperfeiçoar a prestação de serviços de energia elétrica, além de possibilitar meios para melhorar a prestação dos serviços de distribuição de energia elétrica, objetivando atender às reais necessidades dos clientes.

Nesse momento, surge o papel do órgão regulador como agente promotor da competição e da mudança de postura das distribuidoras de energia em termos de orientação para o mercado. Uma vez por ano as distribuidoras têm acesso às pesquisas de satisfação realizadas pela agência reguladora. Com essas informações estas procuram identificar seus pontos de melhorias e verificar os que já conseguiram bons resultados. Além disso, esse novo contexto abre as portas para o desenvolvimento de estratégias de marketing almejando garantir e dar continuidade ao direito de concessão.

Na perspectiva de buscar melhorias, as distribuidoras precisam acompanhar as tendências no marketing. No século XXI essas tendências aparecem dando uma crescente ênfase na qualidade, valor e satisfação dos consumidores, no desenvolvimento de relacionamento, em administrar processos gerenciais integrando funções administrativas, em marketing direto e pela internet, em serviços e tecnologia industrial, na ética do marketing e na

psicologia do consumidor que busca entender o comportamento de consumo, dentre outras. Dentre estas tendências observa-se que o consumidor é o responsável pelas mudanças no marketing, pois o que faz os profissionais de marketing perceber em novos rumos para o marketing são as mudanças de comportamento das pessoas. Blackwell; Miniard; Engel (2005) asseguram que as alterações no comportamento dos consumidores são, muitas vezes, o resultado das adaptações a mudanças ambientais. Afirmando também que a empresa precisa acompanhar constantemente seus clientes, indo ao encontro das suas expectativas, pois é o consumidor quem acaba mostrando os possíveis rumos da empresa.

Frente a essas tendências, a segmentação de mercado, que é a busca do isolamento de fatores que distinguem certos grupos de consumidores do mercado global, é bastante utilizada nas estratégias de marketing (SHETH; MITTAL; NEWMAN, 2001).

Aparentemente é uma ferramenta antiga e pouco utilizada, no entanto observa-se que esta é fundamental para direcionar as ações de marketing nas organizações. Não se tem muitos registros de estudos de segmentação de consumidores de energia elétrica no Brasil, assim acredita-se que este estudo pode preencher essa lacuna na área tanto na academia quanto para o mercado em si.

Dentre os poucos trabalhos, pode-se citar o de Silva e Jannuzzi (2009). Os autores apresentam a segmentação clássica de grandes clientes de energia elétrica utilizadas no setor e apresentam doze potenciais formas de segmentar esse consumidores. As formas propostas são: localização geográfica, valor da fatura, setor de atividade, composição acionária e gestão empresarial, formas e características da utilização da energia nas unidades consumidoras, grau de sofisticação tecnológica, relação custo-benefício do cliente, fatores comuns de compra de energia, marca associada ao cliente, fatores situacionais (como urgência, aplicação e volume de energia), características pessoais do comprador, ex-clientes que migraram do mercado cativo para o livre ou vice-versa (SILVA; JANNUZZI, 2009, p. 9).

As contribuições são relevantes para os estudos de segmentação de consumidores de energia, porém os autores trabalharam apenas os clientes de alta tensão, ou seja, o grupo A dos consumidores. Sabendo-se que o mercado de clientes de baixa tensão, ou grupo B, também é possível de ser segmentado, este estudo procurará identificar nichos específicos dessa massa de consumidores, procurando identificar grupos com características semelhantes, através da modelagem de dados categóricos.

Procura-se identificar segmentos que caracterizam os grupos de consumidores de energia elétrica residenciais. A ideia é proporcionar meios para que as empresas desse setor de atividade identifiquem grupos de consumidores homogêneos e possam identificar novas oportunidades para o seu negócio, com base nos dados internos de seus clientes, não ficando reféns da segmentação clássica utilizada pelas distribuidoras no Brasil. Do ponto de vista mercadológico e social, o estudo torna-se relevante porque as distribuidoras de energia vêm investindo em melhorias de distribuição e atendimento ao cliente, além de novos negócios que

não sejam simplesmente distribuição de energia elétrica.

Não se pretende propor nenhum modelo de segmentação específico para o setor elétrico brasileiro, mas sim propor a aplicação de uma metodologia que possa facilitar a identificação de grupos de consumidores através da modelagem de dados categóricos, por meio da utilização de bases de clientes existentes nas distribuidoras de energia elétrica. Assim, este trabalho tem como objetivo propor o uso da técnica *Chi-Square Automatic Interaction Detection (CHAID)* na segmentação do mercado de consumidores residenciais de energia elétrica.

2. SEGMENTAÇÃO DE MERCADO

Aparentemente, o termo segmentação vem sendo utilizado em relação a uma multiplicidade tão grande de situações que é prudente empregá-lo de modo mais limitado e acertado (VEIGA-NETO, 2002). De acordo com Richers e Lima (1991), a segmentação não deve referir-se ao setor de atividade, canais de distribuição ou produtos, mas sim ao mercado. Os autores afirmam ainda que como técnica ou como estratégia a segmentação só faz sentido se puder contribuir para que uma instituição ou empresa comercial possa introduzir seus produtos ou serviços de forma a aumentar sua participação nos mercados que escolheu como principais. É notório perceber que as pessoas estão constantemente em mudança, no entanto a única certeza de mudança que os profissionais de marketing sabem, é que os consumidores mudam naturalmente em diversas fases da vida.

Se não fosse a diversidade humana, possivelmente não seria necessária a segmentação de mercado. O motivo que faz esta ferramenta de marketing ser utilizada e ter um papel de suma importância como estratégia competitiva são as diferentes origens, estilos de vida, valores, interesses e desejos das pessoas, assim como, seus traços comportamentais, culturais e sociais. Schiffman e Kanuk (2000, p. 31) concordam que “a diversidade no mercado global torna a segmentação de mercado uma estratégia atraente, viável e potencialmente muito lucrativa”. Com isso, o método de segmentação de mercado permite formular ações estratégicas e táticas, visando atender de forma mais ampla e dirigida às expectativas, necessidades e desejos dos consumidores de cada segmento identificado.

Quando uma empresa decide trabalhar o seu mercado com segmentos específicos, existe uma grande possibilidade de que seus clientes fiquem mais satisfeitos. Isso acontece devido à facilidade de identificar os reais desejos e necessidades dos clientes, e supri-los com produtos ou serviços que vão estar de acordo com suas aspirações, possibilitando assim, uma possível fidelização dos mesmos. As empresas, em contrapartida, ganham quando identificam que todos os seus clientes estão satisfeitos, e isto, quando ocorre, beneficia também os colaboradores que se motivam pelo fato de ser um dos principais contribuintes no sucesso da empresa junto ao seu mercado. Segundo Solomon (2002), a segmentação de mercado versa na capacidade de se observar em

um mercado heterogêneo diversos grupos homogêneos menores que demonstram preferências frente a produtos ou serviços com características similares.

Para Lamb Jr., Hair Jr. e McDaniel (2004, p. 206), “o termo mercado significa diferentes coisas para diferentes pessoas”. Ainda completam afirmando que um mercado é composto por pessoas ou organizações com desejos e necessidades e com recursos e disposição para consumir algo. Um grupo de pessoas ou organização que não possua qualquer uma dessas características não é um mercado. Richers e Lima (1991) apresenta quatro opções distintas que o administrador deve focalizar em suas opções de segmentação: a opção do ramo de atividade, que está relacionada com as considerações relativas ao seu mercado concorrente; a opção dos segmentos de mercado, que está relacionada às oportunidades; a opção relacionada com a distribuição, cuja estratégia se baseia em adaptações; e a opção da comunicação, que se está amparada na escolha da mídia (RICHERS; LIMA, 1991).

O trabalho de seleção dos segmentos para atingir os mercados-alvo é tarefa do marketing, que busca encontrar os nichos do mercado em que a empresa atua. Simpson (2001) acredita que o marketing de nicho é o processo de direcionar-se para um segmento relativamente pequeno com um composto de marketing especializado. Embora se tenha definições semelhantes a essa, ainda se vê profissionais dessa área que procuram identificar nichos somente através de divisões de sub-grupos dos segmentos ou meramente definem um grupo por determinadas semelhanças estereotipadas. Isto pode gerar distorções nas decisões de marketing, já que as variações e as mudanças das pessoas no ambiente são dinâmicas e um dos fatores que são mais importantes na identificação de nichos é a identificação de grupos com características semelhantes.

Em relação aos objetivos, Veiga-Neto (2002) considera que seu objetivo é evitar o desperdício de energia e de trabalho em mercados pouco interessantes para a empresa. A segmentação permite o aprofundamento no conhecimento das necessidades e desejos do consumidor que são, gradualmente, aumentadas na medida em que novas variáveis, como, por exemplo, os perfis comportamentais, são combinados. Desta maneira, proporciona aos administradores a possibilidade de traçar um perfil aproximado do tipo de preferências de consumo dos grupos homogêneos de consumidores.

Os grandes agregados do mercado podem ser substituídos por subdivisões em segmentos menores do mercado-alvo. Para ser dado o primeiro passo no desenvolvimento de uma estratégia de segmentação deve-se fazer a seleção da base ou das bases mais adequadas nas quais serão segmentadas (SCHIFFMAN; KANUK, 2000). As bases da segmentação de mercado são entendidas de distintas formas, cada autor apresenta análises diferenciadas para essa ferramenta de marketing.

Os pontos em comum que podem ser encontrados na maioria das obras que tratam do respectivo assunto referem-se a variáveis geográficas, demográficas, psicográficas (valores e estilos de vida) e comportamentais (ETZEL, WALKER, STANTON,

2001; SCHIFFMAN; KANUK, 2000; RICHERS; LIMA, 1991; SIMPSON, 2001; SHETH; MITTAL; NEWMAN, 2001; LAMB Jr., HAIR Jr.; McDANIEL, 2004; BLACKWELL; MINIARD; ENGEL, 2005; ENGEL; BLACKWELL; MINIARD, 2000; MOWEN; MINOR, 2003). Alguns autores apresentam outras bases para segmentação diferentes das quatro que foram apresentadas anteriormente. Como exemplo de obras que apresentam outras variáveis, Schiffman e Kanuk (2000) citam mais quatro bases: sociocultural, relacionada com o uso, por uso-situação e híbrida.

3. SEGMENTAÇÃO CLÁSSICA UTILIZADA NO SETOR ELÉTRICO

A segmentação empregada de forma clássica no mercado de distribuição de energia elétrica é habitualmente realizada por níveis de tensão, classe tarifária e setores da economia (ANTUNES, 2006). “Essa segmentação tem como principais objetivos a diferenciação de tarifas e também o planejamento de expansão do mercado, ou seja, o foco na oferta contínua de energia elétrica, visando eliminar possíveis desequilíbrios de mercado” (SILVA; JANNUZZI, 2009, p. 3).

As principais divisões realizadas são grupo A e grupo B. De acordo com a Resolução nº 456/2000 da ANEEL, o grupo A (média e alta tensão) é composto de unidades consumidoras com fornecimento em tensão igual ou superior a 2,3 KV, ou, ainda atendidas em tensão inferior a 2,3 KV a partir de sistema subterrâneo de distribuição. Conforme a Aneel (2000), é subdividido nos seguintes subgrupos: Subgrupo A1 – tensão de fornecimento igual ou superior a 230 kV; Subgrupo A2 – tensão de fornecimento de 88 kV a 138 kV; Subgrupo A3 – tensão de fornecimento de 69 kV; Subgrupo A3a – tensão de fornecimento de 30 kV a 44 kV; Subgrupo A4 – tensão de fornecimento de 2,3 kV a 25 kV e Subgrupo AS – tensão de fornecimento inferior a 2,3 kV, atendidas a partir de sistema subterrâneo de distribuição e faturadas neste Grupo em caráter opcional.

Já os clientes de menor porte (baixa tensão) conhecidos como grupo B é composto de unidades consumidoras com fornecimento em tensão inferior a 2,3 KV. Conforme a Aneel (2000), é subdividido nos seguintes subgrupos: Subgrupo B1 – residencial; Subgrupo B1 – residencial baixa renda; Subgrupo B2 – rural; Subgrupo B2 – cooperativa de eletrificação rural; Subgrupo B2 – serviço público de irrigação; Subgrupo B3 – demais classes e Subgrupo B4 – iluminação pública. A Resolução nº 456/2000 da ANEEL ainda estabelece quais as classes e subclasses de consumo de energia para efeito de aplicação de tarifas. As divisões são feitas de acordo com o tipo de unidade de consumo que receberá o fornecimento (sugere-se ver as classes de consumo disponíveis na resolução nº 456/2000, ANEEL).

As divisões realizadas caracterizam segmentos de mercado considerando apenas aspectos de fornecimento. Com as alterações advindas das mudanças do setor elétrico, observa-se que existem limitações nesse tipo de segmentação. No trabalho de Silva e Jannuzzi (2009), os autores apresentam a segmentação clássica de grandes clientes de energia elétrica utilizadas no setor e apresentam

doze potenciais formas de segmentar os grandes clientes. As contribuições são relevantes para os estudos de segmentação de consumidores de energia, porém os autores trabalharam apenas os grandes clientes, ou seja, o grupo A dos consumidores.

4. PROCEDIMENTOS METODOLÓGICOS

Este estudo é de natureza quantitativa e caracteriza-se como exploratório descritivo. A pesquisa exploratória descritiva procura esclarecer a natureza de um problema descrevendo as características de uma população ou de um fenômeno (ZIKMUND, 2006). Quanto à natureza, as pesquisas quantitativas são as que empregam a estatística e a matemática, utilizando os números e cálculos como principal recurso para análise dos dados (LEITE, 2004; MARCONI; LAKATOS, 2006). A estrutura metodológica foi construída a partir da *Knowledge Discovery in Databases (KDD)*, termo que é traduzido por Descoberta de Conhecimento em Bases de Dados (DCBD). A KDD consiste numa série de passos sistematizados e bem definidos que vão desde a preparação dos dados até a extração de padrões que possam ser analisados posteriormente. O método pode ser utilizado em diversos ramos da ciência, como saúde, finanças, vendas, varejo ou qualquer outro campo, onde as análises de dados são realizadas classicamente por um ou mais analistas que procura fazer uma interface entre os dados de usuários e produtos (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). Cabe ressaltar que a validade do processo depende diretamente dos analistas, visto que não existe ainda nenhuma solução tecnológica que seja capaz de determinar autonomamente se os padrões encontrados nas bases de dados são de fato novos ou válidos para fins de tomada de decisão (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

O estudo foi realizado com base nos dados de uma concessionária de energia elétrica, considerada modelo, no mercado de distribuição de energia. A distribuidora está localizada no estado do Ceará e tem suas atividades regulamentadas e fiscalizadas pela ANEEL. A companhia é responsável pelo atendimento de 2,71 milhões de clientes, correspondendo à universalização do serviço de mais de 95% de clientes, distribuídos nos 184 municípios, que totalizam uma área de atuação de 146.817 km². É uma sociedade anônima de capital aberto, regulamentada pela Lei nº 6.404/76.

De posse da base de dados para o processo KDD, verificou-se que a empresa dispunha de 2,71 milhões de clientes. Para o estudo, consideraram-se apenas os clientes residenciais, visto que os não residenciais apresentavam variáveis diferentes dos residenciais, totalizando 2,44 milhões de clientes. Além dos não residenciais, foram excluídos os grandes clientes (grupo A), que são aqueles que têm sua unidade consumidora ligada em média e alta tensão e são caracterizados em outro segmento no mercado de distribuição de energia.

Continuando o processo de preparação dos dados, foram excluídos todos os clientes que não tinham todas as variáveis independentes disponíveis no banco de dados. Assim, após esse procedimento, obtiveram-se como amostra 628.339 consumidores da distribuidora para o tratamento estatístico o que corresponde a

25,8% do total de clientes residenciais. Os dados foram extraídos no mês de junho de 2009.

A variável dependente escolhida para o modelo foi faturamento médio de energia, sendo esta utilizada como “proxy para o nível de renda e/ou desenvolvimento do município, visto que existe correlação muita alta entre estas duas variáveis” (LOUREIRO, 2008, p. 6). A variável está distribuída em cinco categorias, divididas nos seguintes códigos: 1. até R\$ 25,00; 2. de R\$ 25,01 a R\$ 50,00; 3. de R\$ 50,01 a R\$ 100,00; 4. de R\$ 100,01 a R\$ 300,00; e, 5. mais de R\$ 300,00. As variáveis preditoras utilizadas foram: estado civil, idade, escolaridade e profissão do titular da conta de energia, número de pessoas no domicílio, sub-regiões, bairro, possui produto, tipo de produtos denominados bens de consumo, cobrança fácil, serviços elétricos, kit de energia e seguros, classe de consumo e segmentação transacional.

4.1 MODELAGEM: TÉCNICA DE ANÁLISE DE DADOS

Após a preparação dos dados, seguiram-se as fases necessárias para a recuperação das informações ou mineração/garimpagem e, em seguida, a seleção do conjunto de dados utilizados no processo descrito anteriormente. A preparação e seleção dos dados foram realizadas através da utilização dos softwares: Excel 2007 e SPSS 17.0. Para a modelagem, optou-se pela técnica *Chi-Square Automatic Interaction Detection (CHAID)* para segmentação de mercado conforme proposto por Freitas e Heineck (2008).

4.1.1 A TÉCNICA CHAID

Como mencionado anteriormente, a técnica *CHAID* explora e analisa a classificação de variáveis categóricas e institui o maior valor da estatística Qui-quadrado (χ^2), de modo a estabelecer todas as partições em subgrupos possíveis, para cada variável independente. Os agrupamentos formados são homogêneos internamente, apesar de apresentarem heterogeneidade entre si. Para verificar a influência de K-1 variáveis independentes, cada subgrupo é reanalisado utilizando o critério de seleção estatística de Qui-quadrado de Pearson ou a Razão de Verossimilhança (LOPES, 2003).

Em suma, o algoritmo definido pela técnica *CHAID* agrupa os valores estatisticamente semelhantes (homogêneos) à variável independente e mantém os outros valores heterogêneos. De outra maneira, pode-se dizer que o algoritmo seleciona a melhor variável preditora para formar o primeiro ramo da árvore de decisão, de forma que cada nó tenha em sua constituição valores semelhantes da variável selecionada. Este processo se dá recursivamente até obter-se a árvore por completo (FOWDAR; BARDAR; CROCKETT, 2002). Assim, de acordo com Fowdar, Bardar e Crockett (2002), os procedimentos estatísticos do *CHAID* na construção da árvore de decisões são descritos da seguinte forma:

• **Passo 1:** Para cada variável preditora ou variável independente X, busca-se o par de categorias de X que apresente a menor

diferença significativa (ou seja, que apresente o maior *p-valor*) com relação à variável dependente Y. O cálculo deste *p-valor* depende do nível de medida de Y – se Y é contínuo, usa-se o teste de F e se Y é nominal tem-se duas formas de tabulação que são: as categorias de X como linhas e categorias de Y como colunas. Usa-se o teste de Qui-quadrado de Pearson ou o teste de relação de probabilidade.

• **Passo 2:** Para o par de categorias da variável independente X, com o *p-valor* maior, compara-se seu valor para um nível de alfa pré-especificado, α_{merge} (igual a 0,05, com 95% de significância). Em caso do *p-valor* ser maior que α_{merge} , agrupa-se este par em uma única combinação de categoria. Um novo bloco de categorias de X é formado e o algoritmo reexamina este preditor e procede novamente o passo 1. Por outro lado, se o *p-valor* for menos que α_{merge} , segue-se o passo 3.

• **Passo 3:** Calcula-se o *p-valor* ajustado que usa um multiplicador formal de Bonferroni.

• **Passo 4:** Seleciona-se a variável independente X, que tem o *p-valor* ajustado menor, isto é, nomeia-se aquela variável que é mais significativa. Compara-se este valor para um alfa com nível pré-especificado, α_{split} (igual a 0,05, com 95% de significância). Se o *p-valor* for menor que ou igual a α_{split} , divide o nó, fundando outro conjunto de dados das categorias de X. Já se o *p-valor* é maior que α_{split} , então este é um nó terminal; não se divide.

• **Passo 5:** Este processo será contínuo e crescente até que todas as “amarrasções” da árvore sejam conhecidas.

Vale salientar que este processo cessará se quaisquer das condições seguintes forem estabelecidas. As condições segundo Fowdar, Bardar e Crockett (2002), são: (1) todos os casos em um nó têm valores idênticos para todos os preditores; (2) o nó fica puro. Isto significa que todos os casos dentro do nó tenham o mesmo valor designado; ou (3) se uma profundidade foi especificada, o crescimento da árvore cessará quando a profundidade (nível) da árvore alcançar seu limite.

Corroborando com os passos descritos anteriormente, Lopes (2003, p. 47-48) argumenta que “o procedimento desenvolvido pelo módulo *Answer@Tree* do pacote SPSS (algoritmo) segue as etapas propostas por Kass em 1980” descritos pelos passos 1 e 2. Complementa dizendo que este “processo de geração de novos nós termina na ocorrência de uma das seguintes situações: (1) após a análise de todas as variáveis independentes; (2) na ausência de significância estatística nas associações; ou (3) quando o número de observações for pequeno demais por subgrupo”.

Considerando o processo de Data Mining, Freitas e Heineck (2008, p. 41) descrevem como se dá a aplicação prática deste método estatístico para utilização no processo de segmentação, mediante a realização dos seguintes passos lógicos:

Seleção da variável dependente categórica (passo 1), criação de categorias para esta variável (passo 2), definição das variáveis independentes (passo 3), estabelecimento das opções de segmentação (passo 4), onde se define as regras de finalização da divisão e o tamanho mínimo dos grupos formados. É também nesta etapa que se estabelece o nível de significância estatística que irá

impactar na junção dos níveis das variáveis independentes e na sua elegibilidade para serem incluídas no processo. A cada passo podem ser visualizados testes estatísticos de associação entre as variáveis predictoras e a variável dependente (passo 5) (FREITAS; HEINECK, 2008, p. 41).

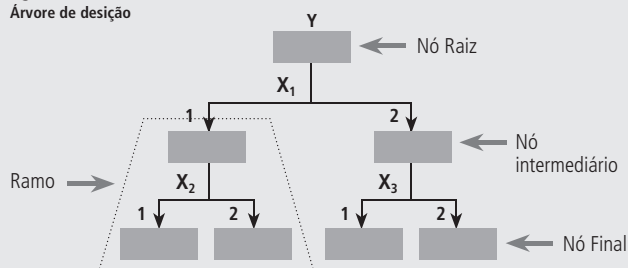
4.1.2 A ÁRVORE DE DECISÃO

Conforme Lopes (2003, p. 48), a técnica *CHAID* apresenta seus resultados em forma de uma árvore, em que seus últimos ramos são chamados “terminais e definem um subgrupo de indivíduos classificados em um dos níveis da variável dependente: o nível (ou categoria) da variável dependente (Y) que apresentar a maior proporção de indivíduos define sua classificação”.

Em uma árvore, tem-se a representação em nós, ramos e “nós-folhas”. Cada nó representa o teste ou decisão sobre o item considerado, cada ramo indica o resultado do teste e os nós folha ilustram a classe ou classes de distribuição. No desenvolvimento de uma árvore de decisão, inicialmente, separam-se os dados em um grande número de conjuntos que se tornam menores à medida que ficam mais específicos. Quanto maior é o número de diferentes casos a ser analisado, menor será cada subgrupo que se segue, reduzindo a exatidão da classificação adicional correta. Quanto maior for o número dos pequenos ramos de uma árvore, menor a representatividade de cada um dos nós (MONTEIRO, 2008). Pode-se observar a segmentação dos resultados do *CHAID* ilustrado na Figura 1.

A árvore tem seu início no chamado nó raiz da árvore (em Y), de onde parte o caminho para os ramos, baseados em atributos de novos casos, antes de chegar ao nó-folha (ou nó terminal, que pode ser intermediário ou final). O novo caminho fundamenta-se em valores que estabelece a divisão (MONTEIRO, 2008). A árvore de decisão ilustrada na Figura 1 apresenta-se em três níveis, com sete nós terminais (nós folhas), distribuídos em 3 ramos.

Figura 1
Árvore de decisão



Fonte: Adaptado de Monteiro (2008).

Para definição do tamanho mínimo de elementos, verificou-se que no manual de utilização do SPSS 2001 é sugerido que para amostras com mais de 1.000 casos sejam fixados no mínimo 100 casos para os nós pais e 50 de nós filhos. Para o número de máximo de níveis gerados para árvore é de 5. Devido ao tamanho da amostra, adotou-se 30.000 para os nós pais e 15.000 para nós filhos com 5 níveis para a formação dos agrupamento do estado.

5. RESULTADOS

Conforme visto na fase de preparação dos dados, a variável dependente é categórica e está distribuída em cinco categorias. Na formação dos segmentos, o teste de Person Qui-quadrado das variáveis predictoras elegíveis no modelo apresentou valores entre 197,300 a 298.966,237, todos significativos com p-valor próximo de zero (0,000). Esses dados são gerados pelo próprio modelo do *CHAID*, permitindo uma melhor análise das divisões. Analisando o dendograma extraído, verifica-se que grande parte das variáveis independentes foi escolhida para explicar o faturamento médio de energia. As variáveis eleitas no modelo foram: classe de consumo, possui produto, número de pessoas no domicílio, escolaridade, idade, cobrança fácil, profissão, sexo e segmentação transacional. Nesta formação, não foram contempladas as variáveis estado civil, município, bens de consumo, serviços elétricos, kit energia e seguros. Em cada nó da árvore, a frequência das faixas de faturamento médio de energia é apresentada em forma de gráfico, onde a legenda está ao lado do nó raiz.

Com relação ao gênero, viu-se que a maior parte da amostra compõe-se de pessoas do sexo feminino, 75,0% para o estado e 73,0% na capital. Considerando a amostra do total do estado, 63,3% são de pessoas casadas, enquanto que 27,6% são solteiras. Apenas 9,1% dos consumidores são divorciados, desquitados ou viúvos (categoria denomina outros). A idade média dos consumidores no estado é de 44 anos e em Fortaleza é de 43 anos. Com esses dados, acredita-se que ações de vendas e marketing que atinjam o público feminino possam ser mais eficientes, já que a maioria do público é desse sexo. Ações que expressem mensagens para a família também podem ser eficientes, já que existe um percentual significativo de pessoas casadas.

A variável que mais explica o faturamento médio de energia é a classe de consumo. Observa-se que modelo apresenta os grupos a partir dessas classes, onde cada grupo apresenta segmentos de acordo com as características dos domicílios. Com isso, verificou-se que os resultados seriam melhor apresentados de acordo com esses grupos: **Grupo 1:** Cliente residencial baixa renda com consumo de até R\$ 25,00; **Grupo 2:** Cliente residencial normal com consumo de R\$ 50,01 a R\$ 100,00; e **Grupo 3:** Cliente residencial baixa renda pelo NIS. Nota-se que nenhuma variável independente explica a variável residencial baixa renda (NIS), que são aqueles consumidores que ganham o benefício por meio do Número de Inscrição Social do Governo Federal e consomem de 80 KWh até no máximo 220 KWh, considerando a média dos últimos 12 meses. Assim, o Grupo 3 isola-se por essa característica. Com relação aos consumidores de baixa renda pelo consumo (Grupo 1), nota-se que estes foram subdivididos pelos que têm produto e não têm produtos ofertados pela distribuidora. Já os residenciais normais (Grupo 2), foram melhor explicados pelo grau de escolaridade.

Percebe-se que o consumidor de baixa renda por consumo (Grupo 1) tem um faturamento médio de no máximo R\$ 25,00 e, provavelmente, não tem uma renda familiar elevada. No entanto, os mesmos costumam comprar produtos ofertados pela distribuidora

(seguros), incluindo produtos e serviços de empresas parceiras que fazem sua arrecadação na fatura de energia e, geralmente, apresenta até duas pessoas no imóvel. No perfil do cliente que consome produtos da distribuidora, mas não consome produtos e serviços de empresas parceiras que fazem sua arrecadação na fatura de energia, destacam-se os subgrupos diferenciados por número de pessoas que residem no imóvel e nível de escolaridade do titular da conta; sendo que esta última característica não é relevante para pequenas famílias (com até duas pessoas).

Vale observar que os clientes que não consomem produtos estão subdivididos, principalmente, pela quantidade de pessoas que residem no imóvel e pela categoria profissional que exerce. Destacam-se ainda, as características de sexo, nível de escolaridade e segmentação transacional a que pertence o cliente (Ver Tabela 1 no site abep.org/novo/Content.aspx?SectionID=96). Já para os residenciais normais (Grupo 2), que naturalmente têm um maior valor médio de conta de energia (de R\$ 50,01 a R\$ 100,00), o nível de escolaridade é o que melhor explica esse perfil de consumo. Nesse caso, observa-se que o fator idade é relevante para segmentar este tipo de consumidor com nível de escolaridade superior ao ensino médio incompleto, ou seja, aqueles clientes que apresentam um melhor nível de escolaridade devem ser considerados na determinação de estratégias gerenciais. Além disso, é importante perceber que existe uma relação entre o nível de escolaridade e o consumo, pois a maior frequência de faturamento médio de energia que vai de R\$ 100,01 a R\$ 300,00 está no nó que explica os consumidores residenciais normais.

Nota-se que nos domicílios em que o titular tem um maior nível de escolaridade há maior possibilidade de aumentar o consumo de energia, demonstrando ser um segmento de mercado potencial. Percebe-se que o faturamento médio de energia nos domicílios que têm o benefício de baixa renda pelo NIS é superior ao residencial normal e baixa renda pelo consumo. Acredita-se que isso se dá pelo fato de que o limite de KWh seja elevado e, mesmo que a unidade consumidora consiga consumir até este nível, a conta de energia torna-se mais barata devido ao desconto do benefício.

Neste sentido, é possível considerar que os domicílios que têm o benefício de baixa renda pelo NIS podem ser considerados como consumidores residenciais normais, visto que seu consumo aproxima-se muito mais ao nível de faturamento dos residenciais normais do que os de baixa renda pelo consumo. A partir dos segmentos gerados pelo *CHAID* para análise dos grupos formados, viu-se que dos 22 segmentos, 12 são caracterizados por domicílios que não compraram produtos e os demais que já compraram. As categorias de profissões foram bem distribuídas no nó dos consumidores de baixa renda e não apareceram nos residenciais normais. Isso confirma a hipótese popular de que é comum que as pessoas mais pobres têm menor grau de escolaridade.

O nível de faturamento é praticamente o mesmo para ambos os segmentos, demonstrando que esse perfil de consumidores tem um baixo consumo. Embora esse público, aparentemente, tenha renda familiar baixa, apenas um dos segmentos apresenta características de domicílio que costuma ficar inadimplente ou fazer irregularidades

na unidade consumidora. Esse grupo é formado por famílias que chegam até 4 pessoas no domicílio e o titular nunca comprou nenhum produto ou serviço da distribuidora.

6. CONCLUSÕES

Nas discussões realizadas no tocante a segmentação de mercado, viu-se que os segmentos de mercado são formados por agrupamentos de consumidores que tenham as mesmas características e sejam bastante heterogêneos em relação aos outros. A ideia é que os grupos homogêneos sejam facilmente identificados e tenham características semelhantes e com poucas variações internas em seus perfis. Adotando-se como técnica para segmentar os consumidores de energia elétrica, constatou-se que o *CHAID* apresenta-se como uma alternativa relevante para os estudos de segmentação, principalmente, quando se utiliza grandes bases de dados, além das vantagens comentadas ao longo da revisão literária.

Antes da aplicação da técnica, foi fundamental a utilização do processo de KDD, devido ao grande número de dados utilizados, complexidade e particularidade das variáveis, como também, dificuldades de extração dos dados. Em seguida, o processo de *data mining* (DA CRUZ, 2007; HAN; KAMBER, 2001; WAH, 2006) foi importante para o processamento e preparação dos dados, sendo essas etapas antecessoras a aplicação da técnica de modelagem. O processo de avaliação e desenvolvimento é contemplado nos resultados do trabalho, porém verificou-se que o monitoramento seria utilizado após a aplicação e utilização prática do modelo, ficando essa fase fora do processo deste estudo.

Além de não ser comum no setor de distribuição de energia elétrica, de uma forma geral, a segmentação de mercado utilizada, exclusivamente, para o direcionamento e identificação de novas estratégias para prestação dos serviços ao cliente, ainda é considerada como complexa e de difícil aplicação. Sabendo disso, faz-se necessário trabalhar com técnicas menos complexas, confiáveis e de fácil aplicação como é a técnica *CHAID*, para que as organizações possam utilizar a segmentação com maior frequência, passando a considerá-las como uma ferramenta de apoio à decisão.

Para explicar o faturamento médio de energia dos consumidores residenciais, a classe de consumo é a principal variável em todas as análises realizadas. Após essa variável, as outras duas mais significantes são grau de escolaridade e aquela que indica se o cliente tem produto ou não, isso considerando a análise de todo o estado. Na formação dos segmentos da capital, as duas variáveis mais significantes são número de pessoas no domicílio e faixa de idade, variáveis essas também significantes na segmentação dos consumidores que não compram outros produtos ofertados pela distribuidora. Já para os que compram, as duas outras mais significantes são grau de escolaridade e a variável cobrança fácil, na qual indica que o cliente comprou produtos de empresas parceiras da distribuidora e solicitou arrecadação por meio da conta de energia.

Quanto às limitações do estudo, entende-se que seria interessante terem sido considerados os tipos de produtos de cada categoria de

produtos ofertados pela distribuidora. O intuito seria analisar o efeito de cada tipo na formação dos agrupamentos do CHAID, para os consumidores que compram com relação aos que não compram, tais produtos e serviços. Além disso, seria interessante examinar se os clientes que consomem outros produtos e pagam através da fatura de energia tendem a ficar inadimplentes com a distribuidora. A base da segmentação utilizada foi o faturamento médio de energia e por uso.

Verificou-se que no estado 14,4% são clientes residenciais normal, 81,0% residenciais de baixa renda pelo consumo e 4,1% residenciais de baixa renda pelo NIS. Isso mostra que na região existem características de pobreza, mesmo com seu crescimento. Dessa forma, ainda que prevaleça esse perfil de pessoas na região, sabe-se que este mercado é emergente e esses consumidores gostam de consumir da mesma maneira que os que apresentam maior poder de compra. Isso pode ser confirmado, pois os consumidores de baixa renda por consumo foram os que apresentaram maior constância de compra de produtos da distribuidora, principalmente, o produto seguro de vida, ou seja, 28,8% dos consumidores de energia têm pelo menos um seguro de vida, o que equivale a 180.886 clientes.

Embora 79,8% dos consumidores que têm produto sejam baixa renda por consumo, observa-se que ainda existe uma camada do mercado que pode ser explorada, tanto com os produtos já existentes, quanto com a criação de novos produtos e serviços. Por exemplo, criar um produto voltado para consumidores de baixa renda, com incentivos voltados para educação, já que esses detêm de baixo nível de escolaridade. Com relação aos consumidores residenciais normais, é interessante investigar os segmentos identificados através de entrevistas com questionário ou grupos de foco com objetivo de identificar características dos estilos de vida dos consumidores, a fim de que possam ser identificadas mais oportunidades de negócio,

especificamente para esse perfil de cliente, uma vez que esses se apresentam como um grupo de maior poder aquisitivo e elevado nível de escolaridade. É importante lembrar que as famílias que têm maior quantidade de pessoas residindo nos domicílios apresentam maior faturamento de energia, porém há apenas 5,9% de famílias com mais de 6 pessoas no domicílio.

Nos segmentos, verifica-se que nos domicílios de baixa renda existe uma maior incidência de famílias maiores, no entanto, existe uma boa distribuição das faixas de números de pessoas no domicílio que explicam de forma significativa os subgrupos de consumidores que não compram produtos no estado. Ressalta-se que haja também incidência de famílias maiores na capital, porém a maior frequência é de famílias que têm de 3 a 4 pessoas no domicílio.

Conclui-se que os segmentos de mercado identificados para o mercado de consumidores residenciais de energia elétrica podem ser úteis para o desenvolvimento de estratégias de marketing mais direcionadas, sendo essas possíveis de adequação para identificação de novas estratégias para as distribuidoras. No caso deste estudo, ressalta-se a importância de se criarem mais produtos e serviços voltados para a classe baixa renda, pois embora os mesmos tenham baixo poder aquisitivo, é um mercado emergente e está em crescimento. Sugere-se a aplicação de bases psicográficas em cada um dos grupos identificados, com objetivo de ressaltar os estilos de vida desses consumidores. Embora o atributo atraso de pagamento esteja contemplado na variável segmentação transacional, pode-se analisar o nível de inadimplência nas camadas de consumidores de baixa renda. Isso poderia ajudar na tomada de decisão com relação a ações promocionais de vendas em locais com maior risco de inadimplemento. Para estudos futuros, vê-se a necessidade de realizar novos agrupamentos entre os grupos, objetivando reduzir e caracterizar melhor os segmentos.

Referências

ANEEL (Agência Nacional De Energia Elétrica). Resolução n.º 456, 2000.

ANTUNES, Flávia Mesquita. A institucionalização do modelo regulatório do setor elétrico brasileiro: o caso das distribuidoras de energia elétrica. Dissertação (mestrado). Departamento de Administração, Pontifícia Universidade Católica do Rio de Janeiro. Rio de Janeiro, 2006.

BLACKWELL, Roger D.; MINIARD, Paul W; ENGEL, James F. *Comportamento do consumidor*. Trad. Eduardo Teixeira Ayrosa. São Paulo: Pioneira Thomson Learning, 2005.

DA CRUZ, Armando Jorge Ribeiro. *Data Mining via Redes Neurais Artificiais e Máquinas de Vectores de Suporte*. Dissertação (mestrado). Escola de Engenharia da Universidade de Minho. Portugal, 2007.

Disponível em: http://74.125.93.132/search?q=cache:_J9UPhZG90AJ:https://repositorium.sdum.uminho.pt/bitstream/1822/81811/1/Tese%2520de%2520Mestrado.pdf+Data+Mining+metodo+CRISP-DM+e+SEMMA&cd=14&hl=pt-BR&ct=clnk&gl=br. Acesso em 21.09.2009.

ENGEL, James F.; BLACKWELL, Roger D.; MINIARD, Paul W. *Comportamento do consumidor*. Rio de Janeiro, LTC – Livros Técnicos e Científicos S.A., 2000.

ETZEL, Michael J.; WALKER, Bruce J.; STANTON, William J. *Marketing*. São Paulo: Makron, 2001.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From *Data Mining* to knowledge discovery in databases. American Association for Artificial Intelligence. All rights reserved, 1996.

FOWDAR, Jay; BANDAR, Zuhair; CROCKETT, Keeley. Inducing Fuzzy Decision Trees in Non-Deterministic Domains using CHAID. American Association for Artificial

Intelligence. All rights reserved, 2002.

FREITAS, Ana Augusta Ferreira de; HEINECK, Luiz Fernando Mähmann. Segmentação de mercado: proposta de uma metodologia de associação entre clientes e produtos no contexto do mercado imobiliário. *Ambiente Construído*. Porto Alegre, v. 8, n. 3, p. 37-49, jul./set. 2008.

HAN, Jiawey.; KAMBER, Micheline. *Data Mining: Concepts and Techniques*. Academic Press Morgan Kaufmann Publishers. San Francisco: USA, 2001.

LAMB Jr., Charles W.; HAIR Jr, Joseph F.; McDANIEL, Carl. *Princípios de Marketing*. Trad. Luciana Penteado Miquelino. São Paulo: Pioneira Thomson Learning, 2004.

LEITE, Francisco Taciso. *Metodologia científica: iniciação a pesquisa científica, métodos e técnicas de pesquisa, metodologia da pesquisa e do trabalho científico (monografias, dissertações, teses e livros)*. Fortaleza: Universidade de Fortaleza, 2004.

LOPES, Fernanda. *Mulheres negras e não negras vivendo com HIV/AIDS do Estado de São Paulo: Um estudo sobre suas vulnerabilidades*. (Tese de doutorado). Faculdade de Saúde Pública da Universidade de São Paulo. São Paulo, 2003.

LOUREIRO, André Oliveira Ferreira. *Avaliando o efeito do policiamento sobre a criminalidade no Ceará*. Textos para discussão do instituto de pesquisa e estratégia econômica do Ceará (IPECE). n. 53. Fortaleza, 2008.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. *Técnicas de pesquisa: planejamento e execução de pesquisas, amostragens e técnicas de pesquisa, elaboração, análise e interpretação dos dados*. 6. ed. São Paulo: Atlas, 2006.

MONTEIRO, Lúcia Pedro. 2008. *Algoritmo Plug-in LVQ para Servidor SQL*. (Dissertação de mestrado). Instituto Superior Técnico. Universidade Técnica de Lisboa. Lisboa, 2008.

MOWEN, J. C.; MINOR, M. S. *Comportamento do Consumidor*. São Paulo: Pearson Prentice Hall, 2003.

PIRES, José Cláudio Linhares. O Processo de Reformas do Setor Elétrico Brasileiro. *Revista do BNEDES*, Rio de Janeiro, v. 6, n 12, p. 138, 1999.

RICHERS, R.; LIMA, C. P. *Segmentação: Opções estratégicas para o mercado brasileiro*. São Paulo: Nobel, 1991.

SCHIFFMAN, Leon G.; KANUK, Leslie Lazar. *Comportamento do consumidor*. Rio de Janeiro: LTC – Livros Técnicos e Científicos Editora S.A., 2000.

SHETH, Jagdish N.; MITTAL, Banwari; NEWMAN, Bruce I. *Comportamento do cliente: indo além do comportamento do consumidor*. São Paulo: Atlas, 2001.

SILVA, Ana Lúcia Rodrigues da; JANNUZZI, Gilberto de Martino. Segmentação de clientes de energia elétrica por nichos especiais. *Espaço Energia*. Campinas, 2009.

SIMPSON, Penny M. *Segmentação de mercado e mercados-alvo*. IN: CZINKOTA, Michael R. et al. *Marketing: as melhores práticas*. Porto Alegre: Bookman, 2001.

SOLOMON, Michel R. *Comportamento do consumidor: comprando, possuindo e sendo*. Trad. Lene Belon Ribeiro. 5ª ed. Porto Alegre: Bookman, 2002.

VEIGA-NETO, Alípio Ramos. *Segmentação psicográfica de consumidores de ensino superior: a visão da psicologia*. Tese de doutorado. Campinas: Pontifícia Universidade Católica, 2002.

WAH, Yap Bee. Some applications of data mining. Paper Presented at the National Statistics Conference. Putrajaya International Convention Centre. Putrajaya, Malaysia, 4-5th Sept 2006.

ZIKMUND, William G. *Princípios da pesquisa de marketing*. Trad. Cristina Bacellar. São Paulo: Pioneira Thomson Learning, 2006.