

DEFINIÇÃO DAS CARACTERÍSTICAS DO PERFIL DO CONSUMIDOR COM A UTILIZAÇÃO DE REDES NEURAIS ARTIFICIAIS

Autoria: Angelo Maia Cister, Lucy Domingues Shehata

O presente trabalho foi desenvolvido dentro de uma empresa fornecedora de energia elétrica que possui, atualmente, grande interesse em modificar sua estratégia de relacionamento de seus clientes. Embora os clientes não tenham a opção de troca de empresa (*churn*), a empresa fornecedora de energia entende que seu relacionamento vai além do simples fornecimento. Assim sendo, foi contratada a consultoria para delinear e traçar as novas diretrizes do perfil de seus clientes. Para tanto, foi exigência da empresa fornecedora de energia que técnicas modernas e avançadas fossem utilizadas na prospecção e perfil de seus clientes. Assim sendo, como estratégia de análise muitos conceitos de *data mining* foram utilizados e testados tendo-se a escolha natural a partir de uma explicação razoável da modelagem matemática utilizada na comparação com a análise de cluster convencional. Desta forma a escolha residiu nas técnicas de amostragem *bootstrap* e a utilização de redes neurais artificiais com a calibragem de seus parâmetros. Conceitos como análise de confusão, análise de sensibilidade, amostragem *bootstrap*, análise multivariada são citadas e explicadas sua utilização no presente trabalho assim como alguns conceitos de marketing de relacionamentos e CRM.

1 Introdução

Devido à rápida difusão de tecnologias, a diferenciação entre os produtos ou serviços oferecidos por empresas concorrentes é mínima. Com isso, as empresas têm que encontrar novas formas de atrair e reter seus clientes. Disponibilizar atendimento “personalizado”, que permita que os clientes se sintam únicos, tendo suas particularidades respeitadas e evidenciadas, é uma maneira de compensar a pouca diferenciação do que é oferecido por uma empresa em relação às concorrentes. É nesse contexto que entra a idéia de marketing de *Customer Relationship Management* (CRM).

A aplicação do conceito de CRM exige que a empresa possua um banco de dados central com informações sobre seus clientes. É a partir desses dados que se torna possível extrair conhecimento sobre os clientes utilizando técnicas de Estatística Multivariada e *Data Mining* (DM).

Alguns autores definem o *Data Mining* (mineração de dados) como uma técnica que visa encontrar relacionamentos e padrões globais, ainda não descobertos em um banco de dados, que possam gerar respostas corretas para novos casos. É possível dizer, então, que esse processo transforma informações em conhecimento. Desse modo, é possível, utilizando essa ferramenta, caracterizar e segmentar os clientes de uma empresa.

O presente trabalho aborda duas áreas distintas, marketing e mineração de dados.

O **tema** é *Customer Relationship Management* (CRM) analítico; o **sujeito de estudo**, análise de agrupamentos; e o **objeto de estudo**, consumidores de energia elétrica. Em adição, a análise de agrupamentos para os consumidores de energia elétrica de uma determinada empresa (que guarda o direito de não ser identificada) consiste no **sujeito e o objeto de estudo delimitado**.

O **objetivo** do trabalho foi traçar o perfil dos consumidores de energia elétrica com o intuito de direcionar a estratégia de CRM da empresa fornecedora de energia elétrica.

Ao término do trabalho, foi possível classificar os clientes que podem ser alvo prioritário de ação de marketing e os que devem ter prioridade inferior.

É importante ressaltar que o presente trabalho se aproximou ao máximo da realidade, por meio de referências bibliográficas especializadas, base de dados verídica, análise de dados e extração de conhecimentos.

Por meio da bibliografia especializada selecionada, foram desenvolvidas as seções teóricas sobre a evolução do marketing, marketing de relacionamento, CRM e análise multivariada.

A parte empírica do trabalho refere-se à aplicação ao banco de dados (devidamente limpo) dos conceitos apresentados. Dessa forma, a abordagem metodológica é teórico-empírica.

1.1 A Evolução do Marketing

De acordo com Churchill e Peter (2005), existem quatro orientações tradicionais de marketing, a saber:

- I. Orientação para produção: Centra-se nos produtos e em como fabricá-los com eficiência. Entregam-se produtos em locais onde eles possam ser comprados.
- II. Orientação para vendas: Centra-se nas atividades de marketing para a venda de produtos disponíveis. É normalmente utilizada quando a oferta de produtos e serviços é maior que a demanda.
- III. Orientação para marketing: Concentra-se em compreender as necessidades e desejos dos clientes e construir produtos para satisfazê-los. Com isso, é possível criar lealdade e competir de forma eficiente.
- IV. Orientação voltada para valor: Concentra-se em desenvolver e entregar um valor superior para os clientes como modo de alcançar os objetivos da organização.

É fato que os avanços tecnológicos e sociais acarretados pela revolução industrial possibilitaram a produção em massa com eficiência. Além disso, devido à abundância de demanda não atendida, o foco era na produção. A idéia era que se os produtos fabricados em escala, de maneira eficiente, fossem melhores, os clientes os comprariam.

Entretanto, no período de 1900 a 1950, por causa da eficiência adquirida, a produção acabou superando a demanda. Com isso, a orientação mudou para vendas (LANINI, 2005).

Em meados dos anos 1960, foi difundida a orientação de marketing. Em adição, segundo Lanini (2005), “os executivos de vendas passaram a se denominar gerentes de marketing e as escolas de negócios passaram a ensinar os 4P’s” (Produto, Ponto de venda, Preço e Promoção).

Mais tarde, surgiu a orientação voltada para o valor, explicitada por muitos autores como sendo uma orientação de marketing aprimorada. É com essa orientação que “surge” o marketing de relacionamento e, conseqüentemente, marketing direto e *Customer Relationship Management* (CRM).

O passado recente e o cenário atual caracterizam-se pela intensa competição global que tende a diminuir a lucratividade das empresas. Com intuito de suprir essa desvantagem, as empresas optaram por focar na redução de custos, desencadeando a competição baseada em custo. É importante ressaltar que esse tipo de competição

propiciou o efeito de *comoditização* de produtos e serviços juntamente com a inexistência de interação entre cliente e empresa. Com os produtos e serviços “idênticos”, os clientes não tinham nenhuma razão, além do preço, para serem fiéis.

Com a nítida escassez de demanda não atendida, é imprescindível manter os clientes por meio de relacionamentos intensos que visem a *fidelização*. O único meio de aumentar a carteira de clientes é atraindo os dos concorrentes. Sendo assim, as empresas não podem mais focar somente em redução de custos. O novo tipo de competição baseia-se no marketing de relacionamento.

1.2 Marketing de Relacionamento

Grönoss (1997) define o marketing de relacionamento como um:

“Processo de identificar e estabelecer, manter, aprimorar e, quando necessário, encerrar relacionamentos com clientes e outros interessados, com lucro, de modo que os objetivos de todas as partes envolvidas sejam alcançados e que isso seja feito pela oferta e cumprimento mútuo de promessas.”

Mas há outras definições:

“Todos os esforços de marketing dirigidos ao estabelecimento e manutenção de trocas relacionais bem sucedidas” (MORGAN e HUNT, 1994).

“Marketing de relacionamento trata-se de entender, criar e gerenciar a troca de relacionamentos entre parceiros, fabricantes, provedores de serviços, vários membros do canal e consumidores finais” (MÖLLER e WILSON, 1995).

É possível perceber a partir das definições acima que o conceito de marketing de relacionamento envolve todos os tipos de interações que a empresa tem com o ambiente externo. Essa nova visão de marketing surgiu como uma alternativa para tentar enfrentar a cada vez mais acirrada concorrência que oferece produtos e serviços similares (efeito de *comoditização*). A idéia é criar vínculo com os clientes para que eles sejam fiéis por se sentirem importantes e únicos.

Na década de 1990, os holofotes foram todos voltados para o marketing de relacionamento. Alguns autores consideram a existência de duas linhas distintas do marketing de relacionamento: para o consumidor e relacionamento e para as relações interorganizacionais (fornecedores, prestadores de serviços, parceiros, governos).

1.3 Customer Relationship Management

Payne et al. (1999) salientam que o CRM é uma estratégia que dá ênfase aos efeitos da retenção de clientes na rentabilidade da empresa. Sendo assim, prega que é mais eficiente conservar o relacionamento com um cliente do que criar um novo.

Por outro lado, para Rese (2003), o CRM é explicado como uma “ferramenta administrativa para entender e influenciar o comportamento dos clientes por meio de comunicações significativas para melhorar além da retenção, as compras, lealdade e a lucratividade”.

Outra forma de entendê-lo é encontrada em Greenberg (2002): uma estratégia que tem como base a utilização de ferramentas tecnológicas para sua viabilização. Em adição, Plata (apud GARRAFONI et al, 2005) ressalta que CRM é uma “estratégia cujo propósito é transformar os processos de negócios para conservar e conseguir clientes”.

Por último, Cister (2005) destaca que “CRM é a infra-estrutura para implementar-se a filosofia *one-to-one* (um a um) de relacionamento com os clientes”.

Este trabalho pauta-se no CRM como uma estratégia que tem como objetivo reter, ganhar clientes e maximizar a contratação de serviços ou produtos oferecidos pela empresa. A forma de se atingir esse objetivo é antecipando e satisfazendo as expectativas dos clientes por meio da análise dos perfis de seu comportamento. Destaque-se que os termos “*Customer Relationship Management*” e “*one-to-one marketing*” não são considerados sinônimos. Isso por que nem todas as interações com os clientes precisam ser personalizadas. Ao contrário, as interações só serão personalizadas se os clientes forem ou tiverem potencial de ser tornarem valiosos. Dessa forma, para a autora deste trabalho, o *one-to-one* marketing é uma ferramenta utilizada pelo CRM para reter os clientes que trazem ou podem trazer retorno financeiro para a empresa.

Para se conseguir segmentar e caracterizar os clientes pelo perfil de comportamento, é de extrema relevância utilizar ferramentas de Tecnologia da Informação.

Do ponto de vista tecnológico, CRM envolve capturar os dados do cliente ao longo de toda a empresa, consolidar todos os dados capturados interna e externamente em um banco de dados central, analisar os dados consolidados, distribuir os resultados dessa análise aos vários pontos de contato com o cliente e usar essa informação ao interagir com o cliente através de qualquer ponto de contato com a empresa (CISTER, 2005).

1.3.2 CRM nas Atividades da Empresa

Segundo Greenberg (2002), o CRM deve estar presente em três segmentos de atividades da empresa, a saber:

- I. CRM Analítico: Refere-se a coleta, processamento, organização, modelagem e análise das informações sobre o cliente (*customer intelligence*). Isso torna possível determinar os clientes que devem ser tratados de forma personalizada (*one-to-one*) e os que merecem prioridade inferior.
- II. CRM Operacional: Cister (2005) afirma que esse segmento é o foco da maioria das empresas. Está relacionado à automatização da força de vendas, centros de atendimento a clientes (*call-centers*), *sites* de comércio eletrônico e sistemas automatizados de pedido. É relevante lembrar que essas atividades devem estar alinhadas com as necessidades e conveniências dos clientes.
- III. CRM Colaborativo: Foca na interação entre cliente e empresa. “Sendo assim, é a aplicação da tecnologia de informação que permite a automação e a integração entre todos os pontos de contato do cliente com a empresa” (Cister, 2005). É de extrema relevância repassar as informações levantadas para os sistemas do CRM Operacional, possibilitando o enriquecimento de dados para compor o perfil de comportamento dos clientes.

1.3.3 As etapas do CRM

Peppers e Rogers (1999) destacam que existem quatro etapas estratégicas básicas para a implementação do CRM: identificar, diferenciar, interagir e personalizar.

1.3.3.1 Identificar

Para iniciar uma relação com um cliente é necessário conhecer sua história: “sua identidade, forma de contato preferida, todas as transações e interações realizadas com a empresa, todas as reclamações feitas e quais foram as providências tomadas” (Cister, 2005). É a partir dessas informações que se pode estabelecer quais os clientes merecedores de relação personalizada. Desse modo, são atividades desenvolvidas nesse estágio:

- Coleta e expansão do banco de dados dos clientes;
- Coleta de informações adicionais sobre os clientes;
- Atualização de informações sobre o cliente.

1.3.3.2 Diferenciar

É fato que os clientes possuem necessidades distintas e valores diferentes para a empresa. O principal objetivo dessa etapa é encontrar os clientes de maior valor, os de maior potencial, os intermediários e os *below zeros*. Em adição, traçar as necessidades dos clientes com os quais valha à pena intensificar a relação. Segundo Cister (2005), é preciso “acumular e estudar os dados transacionais dos clientes para aprender com eles”. Sendo assim, são atividades dessa etapa:

- Identificar os clientes mais valiosos, os que têm grande potencial e suas respectivas necessidades;
- Identificar os clientes intermediários e suas necessidades;
- Identificar os clientes *below zeros*;
- A frequência de reclamações e quais foram feitas pelos clientes que interessam;
- A frequência de compra e quais foram os serviços ou produtos adquiridos pelos clientes que interessam;
- Criação de um *ranking* de clientes.

1.3.3.3 Interagir

Depois de conhecer os clientes que interessam, é preciso incentivá-los a interagir com a empresa. Quanto mais freqüente for essa interação, mais a empresa irá saber sobre os clientes podendo, assim, satisfazer suas necessidades e expectativas. De acordo com Cister (2005), “o objetivo final é desenvolver uma relação de aprendizado, onde cada vez mais os serviços e produtos são personalizados para o cliente que verá conveniência em continuar fiel à empresa”. Em complemento, para Lanini (2005), é importante levar em consideração todos os contatos feitos anteriormente e dispor das informações em tempo real para que o atendimento possa ser o mais personalizado possível. Desse modo, as atividades desta etapa consistem em:

- Utilizar os contatos feitos pelos clientes como oportunidade de venda;
- Avaliar o sistema automatizado de *call-center*;
- Aperfeiçoar o fluxo de informações do cliente dentro da empresa;

- Intensificar a relação com os clientes mais valiosos e os que têm grande potencial;
- Repassar as informações adquiridas para o banco de dados.

1.3.3.4 Personalizar

A personalização do serviço é a última e decisiva etapa. Com base nos conhecimentos advindos das atividades anteriores, é possível adequar os serviços e produtos às necessidades dos clientes. São atividades desta etapa:

- Dar prioridade de atendimento aos clientes mais valiosos e aos que têm grande potencial;
- Perguntar aos clientes como e quando desejam ser abordados;
- Identificar o que os clientes desejam;
- Preencher formulários e questionários específicos para os diferentes grupos de clientes.

2 ANÁLISE MULTIVARIADA

Há três tipos de análises estatísticas:

- I. Análise univariada: análise da distribuição de uma única variável;
- II. Análise bivariada: análise de duas variáveis utilizando classificação cruzada, correlação, variância e regressão simples;
- III. Análise multivariada: análise de mais de duas variáveis.

Hair et al (1998) afirmam que qualquer análise simultânea de mais de duas variáveis pode ser considerada análise multivariada. Entretanto, para uma análise ser considerada verdadeiramente multivariada, é preciso que todas as variáveis sejam randômicas e inter-relacionadas de forma que seus diferentes efeitos não possam ser considerados separadamente.

Em muitas situações, as técnicas multivariadas são meios de elaborar uma única análise que, se feita pela técnica univariada, resultaria em diversas análises, já que ter-se-ia que analisar uma variável por vez. Por outro lado, existem técnicas multivariadas específicas às questões multivariadas, que são listadas a seguir:

- **Componente Principal e Análise de Fator Comum (Análise de Fator):** Método estatístico que pode ser usado para analisar as inter-relações entre um grande número de variáveis e suas correlações. O objetivo é encontrar uma forma de condensar a informação contida no número de variáveis originais em um grupo menor de variáveis (fator) com perda mínima de informação.
- **Regressão Múltipla:** Método de análise apropriado quando o problema envolve uma única variável numérica dependente que se presume estar relacionada a duas ou mais variáveis numéricas independentes. O objetivo é prever mudanças na variável dependente em resposta a mudanças ocorridas nas variáveis independentes.
- **Análise Discriminante Múltipla:** Método apropriado para situações onde as variáveis são categóricas, podendo ser dicotômicas (sim ou não, por exemplo) ou multicotômicas (alto, médio ou baixo, por exemplo). É importante ressaltar que é assumido que as variáveis independentes são numéricas. Aplicável em

- situações nas quais a amostra pode ser dividida em grupos conhecidos com base em uma variável categórica dependente. O objetivo é entender as diferenças de grupos e prever a qual grupo um objeto ou indivíduo irá pertencer a partir de diversas variáveis numéricas independentes.
- **Análise Multivariada da Variância e Covariância:** A análise multivariada da variância (MANOVA) é uma técnica estatística que pode ser usada para, simultaneamente, explorar a relação entre diversas variáveis categóricas independentes e duas ou mais variáveis numéricas dependentes. Já a análise multivariada da covariância (MANCOVA) pode ser utilizada em conjunto com a MANOVA para retirar um possível efeito de uma variável numérica independente não controlada.
 - **Análise Conjunta:** Específica para um novo produto ou serviço em desenvolvimento, permitindo uma decisão realística ao longo da evolução complexa de produtos ou serviços. É possível que o pesquisador de marketing tenha acesso à importância de atributos, assim como o nível deles, enquanto os consumidores avaliam apenas alguns aspectos que consistem na combinação desses atributos.
 - **Correlação Canônica:** Pode ser entendida como a extensão lógica da regressão múltipla. Enquanto a análise de regressão múltipla envolve uma única variável numérica dependente e várias independentes, na correlação canônica é possível correlacionar ao mesmo tempo diversas variáveis numéricas dependentes e independentes. O objetivo é obter pesos para as variáveis dependentes e independentes que propiciem a correlação simples máxima entre os dois grupos de variáveis.
 - **Análise de Cluster:** É uma técnica analítica para criar subgrupos de indivíduos ou objetos. O objetivo é classificar uma amostra entidades (indivíduos ou objetos) em grupos menores tendo por base semelhanças entre elas. Ao contrário da análise discriminante múltipla, os grupos não são pré-definidos. Esta técnica é utilizada exatamente para definir os grupos.
 - **Escala Multidimensional:** O objetivo é transformar os julgamentos de consumidores de produtos e serviços em relação às similaridades e preferências em distâncias representadas em espaço multidimensional. Os mapas de percepção resultantes mostram a posição de todos os objetos, mas são necessárias análises adicionais para descrever quais atributos influenciam a posição de cada objeto.
 - **Modelos Lineares de Probabilidade:** Esta técnica é a combinação de regressão múltipla e análise discriminante múltipla. É a análise de regressão múltipla em que uma ou mais variáveis numéricas independentes são usadas para prever uma variável categórica.
 - **Modelagem de Equação Estrutural:** Permite a separação das relações para cada grupo de variáveis dependentes. Esta modelagem propicia a mais apropriada e eficiente técnica de estimação para uma série de equações de regressão múltipla diferentes estimadas simultaneamente.
 - **Data Mining (mineração de dados):** Conjunto de técnicas específicas para extrair relacionamentos e padrões existentes em banco de dados e *Datawarehouse* que não podem ser vistos facilmente devido ao grande

número de informações. O objetivo é conseguir gerar respostas corretas para novos casos. É um processo que transforma informações em conhecimento.

- **Bootstrapping (re-amostragem):** Técnica que elimina a necessidade de preceitos estatísticos de distribuição de amostragem (como a normalização) por meio do computador para fazer a reamostragem da amostra original, reorganizando e gerando uma estimativa empírica da distribuição da amostra.

Dado que o presente trabalho tem como meta dividir a carteira de clientes em subgrupos que apresentem características semelhantes, as técnicas multivariadas utilizadas são: Componente Principal e Análise de Fator Comum, Análise de Cluster e *Data Mining*. Deste modo, é importante fazer algumas considerações a mais sobre essas técnicas.

2.1 Componente Principal e Análise de Fator Comum (Análise de Fator)

Como dito anteriormente, a análise do componente principal e do fator comum tem como principal objetivo reduzir o número de variáveis que expliquem um comportamento com o mínimo de informação perdida. Scremin (2003) afirma que as componentes principais são constituídas de combinações lineares das variáveis originais, são não correlacionadas entre si e retêm o máximo da informação contida nos dados originais. Além disso, as variâncias são ordenadas decrescentemente e as componentes não são correlacionadas. O mesmo autor define a Análise Fatorial como “uma técnica que tem por objetivo definir o número e a natureza das variáveis subjacentes a um grande número de variáveis. Ela procura descrever, quando possível, a covariância relacionada entre muitas variáveis observáveis em função de poucas variáveis latentes (fatores).”

A seleção das variáveis, nesse caso, é feita pelo método Kaiser. Este método sugere que se consegue o número de fatores (grupo de variáveis) ideal utilizando a matriz de correlação, por meio do número de fatores cujos autovalores são maiores ou iguais a 1.

Em seguida, é feita a análise de comunalidade, que consiste em estipular uma porcentagem mínima de explicação que cada variável deve apresentar. Sendo assim, são retiradas todas as variáveis que não apresentarem a porcentagem mínima e se faz a análise de comunalidade novamente a fim de verificar se ainda existe alguma variável que não cumpre a exigência. Esse processo é repetido até que as variáveis apresentem a porcentagem mínima estabelecida.

O próximo passo é gerar uma análise da variância (ANOVA) com o intuito de explorar quanto da variação do modelo as variáveis conseguem explicar.

Por último, são analisadas as variáveis de cada fator com o objetivo de separar as mais significativas para agrupamento (análise de *cluster*)

2.2 Análise de Cluster

Após definir, na análise de componente principal e de fator, quais são as variáveis que mais explicam o comportamento do consumidor, estipula-se o número de subgrupos a serem utilizados, que geralmente é de 2 a 5. Neste trabalho, para *clusterizar*, utilizou-se a técnica K-means que agrupa registros (dados) com base no centróide do grupo de variáveis escolhidas (Figura 1). Desse modo, o agrupamento é feito a partir da distância dos registros aos centróides dos clusters.

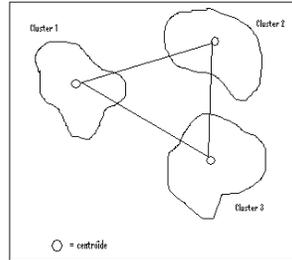


Figura 1: Agrupamento K-means.

2.3 Data Mining

Segundo Coutinho (2003), *data mining* é um:

“ ... processo para extrair informação válida, previamente desconhecida e de máxima abrangência a partir de grandes bases de dados e/ou armazéns de dados, denominados *Datawarehouse* (DW); conjuntos de dados agrupados a partir de uma base de dados qualquer, usando-as para um processo decisório. O DM vai muito além da simples consulta a um banco de dados. Permite aos usuários explorar e inferir informações úteis a partir dos dados, descobrindo relacionamentos escondidos no banco de dados e buscando padrões válidos. É sem dúvida, um conjunto de técnicas utilizada para descobrimento de conhecimento em base de dados robustas.”

São técnicas de mineração de dados:

- Rede Neural.
- Árvores de Decisão.
- Mapas Auto-Organizáveis.

No presente trabalho optou-se por utilizar a Rede Neural.

2.3.1 Rede Neural

A rede neural é uma técnica que, ao invés de resolver um problema de forma matemática, utiliza a modelagem do cérebro humano e sua estrutura para desenvolver uma estratégia de processo. A questão chave da rede neural é o aprendizado por meio do qual os erros de resposta voltam para o sistema para serem reajustados de forma correta. Esse processo é repetido inúmeras vezes para que se consiga aprender com os erros e obter o melhor modelo possível, podendo ser supervisionado ou não-supervisionado. É um processo seqüencial que trabalha com um caso de cada vez. Existem quatro conceitos utilizados pela rede neural:

- I. Tipos de modelo da rede neural: *Multilayer Perception* (MLP), *Radial Basis Function* (RBF) e *Kohonen*, apropriado apenas para problemas de agrupamento (clusterização).
- II. As unidades individuais do processo (nós): são similares aos neurônios do cérebro humano, que recebe entradas (inputs) para depois criar saídas (outputs). Todas as conexões entre nós têm um peso. Sendo assim, o primeiro passo da rede neural é processar os dados de entrada para criar um valor (peso) pelo qual cada saída será multiplicada. Esses pesos são então processados por uma função matemática dentro

- dos nós que transforma todos os pesos de um nó de entrada em um único valor de nó de saída (função de ativação).
- III. O sistema de nós arrumados para transferir sinais dos nós de entrada para os de saída, com alguns nós intermediários: nós de entrada (inputs) que recebem os dados iniciais de cada caso e os transmitem para a rede neural, os nós escondidos que abrigam a camada escondida e a função de ativação, quando existem relações não lineares, e nós de saída que recebem os dados dos nós de entrada e calculam um valor de saída (Figura 2).
 - IV. A função de aprendizado por meio dos erros: os pesos das conexões são a memória, o melhor “chute” de como fazer as previsões dos nós de saída. Uma vez que o valor de entrada é processado pelo sistema, ele pode ser comparado com o valor de saída. Se houver alguma diferença entre o valor de entrada e o de saída (resíduo), existe a necessidade de melhorar o modelo com o intuito de minimizar o erro (HAIR et al, 1998).

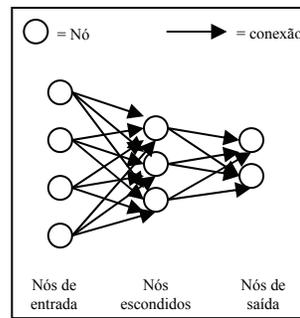


Figura 2: Rede Neural

3 ESTUDO DE CASO

O estudo de caso do presente trabalho envolve a segunda etapa da estratégia de CRM, que consiste em diferenciar os clientes. Sendo assim, foram utilizadas as técnicas de análise de fator e análise de cluster e, posteriormente, a rede neural. A utilização da rede neural teve como objetivo criar um código capaz de classificar todos os clientes da base de dados da empresa.

O banco de dados da empresa em questão abriga dois milhões de registros. Entretanto, para diferenciar os clientes foi usada uma amostra representativa da base com 4934 registros, resultado da utilização da técnica *bootstrapping*.

Em adição, foi utilizado o intervalo de confiança de 95,5% e, conseqüentemente, um erro amostral de 1,42% (Equação 1).

$$\text{Equação 1: } \varepsilon = z \sqrt{\frac{p(1-p)}{n}} = 2 \sqrt{\frac{0.5 * 0.5}{4934}} = 0.0142 = 1.42\%$$

Onde:

- p é a proporção (desconhecida) de elementos em uma população apresentando certa característica de interesse contida numa amostra.
- Z é o valor correspondente do intervalo de confiança na distribuição N(0,1) que, nesse caso, é aproximadamente 2.

- n é o número de registros do banco de dados que, nesse caso, é 4934.

Essa base contém as informações de 22 variáveis entre os anos de 2003 e 2005:

Numéricas: consumo lido, quantidade de consumo em 30 dias, valor do importe, valor do ICMS, valor do corte de reposição, valor total da fatura, valor de débitos, valor de interesses, valor do parcelamento, valor manual e valor de aquisição.

Catégoricas: produtos adquiridos, reclamações, situação do débito automático, situação baixa renda, situação consulta diária de consumo dia-a-dia, subclasse, segmentação, situação pré-pago, tipo de rede, tipo de conexão e tarifa.

Ressalte-se que a empresa não informou as unidades e as definições das variáveis.

É importante destacar que para conseguir trabalhar de forma mais simples, para cada cliente, em vez dos valores de cada variável durante os três anos considerou-se a média deles. Desse modo foi possível trabalhar apenas com um único dado de uma variável para cada cliente. Além disso, os valores das variáveis foram normalizadas de três formas (as mais utilizadas):

- **Por padronização:** aplicou-se a média ponderada com intuito de trabalhar com números menores.
- **Por Log_{10} :** aplicou-se o logaritmo de base 10 nos valores das variáveis com intuito de trabalhar com números menores.
- **Por Log_N :** aplicou-se o logaritmo neperiano nos valores das variáveis com intuito de trabalhar com números menores.

Outro fato relevante é que foram retirados da análise todos os clientes com valores das variáveis consumo lido, consumo em 30 dias, importe e fatura total iguais a zero. Com isso, o banco de dados analisado passou a ter 4181 registros.

3.1 Análise do Componente Principal e Análise de Fatores

Para encontrar uma forma de condensar a informação contida no número de variáveis originais em um grupo menor de variáveis, com perda mínima de informação, utilizou-se o método de extração análise do componente principal (comunalidade) e o método Kaiser (autovalor igual a 1). Foi estipulado que cada uma das variáveis a serem usadas para agrupar os clientes deveria explicar no mínimo 70% da variação de comportamento dos clientes. É importante ressaltar que a análise do componente principal foi aplicada às variáveis padronizadas, com logaritmo de base decimal e com logaritmo neperiano. O tipo de normalização das variáveis que apresentou a maior variância total explicada foi usado para o agrupamento dos clientes.

Os resultados da análise do perfil dos consumidores estão resumidos no Quadro 1.

Quadro 1: Resumo da segmentação dos clientes.

CLUSTER	Porcentagem de Clientes	Classe	Consumo Lido	Valor Total da Fatura	Tarifa	Tipo de Conexão	Tipo de clientes	Tipo de Reclamações	Produtos Adquiridos
1	3,56%	Residencial E Demais Classes	984,9	491,38	90	Trifásico	INTERMEDIÁRIOS	Atendimento a clientes e qualidade no fornecimento	Seguro energia premiada
2	0,19%	Demais Classes	5485,47	5671,56	92	Trifásico	HEAVY USERS	Qualidade no atendimento	Nenhum
3	95,74%	Residencial	179,2	83,48	90	Monofásico	CLIENTES POUCO RENTÁVEIS E BELOW ZEROS	Faturamento e qualidade do fornecimento	Seguro energia premiada
4	0,5%	Demais Classes	2792,38	1341,89	92	Trifásico	CLIENTES COM GRANDE POTENCIAL	Faturamento	Nenhum

3.3 Rede Neural

Foram retirados, de forma aleatória, 119 registros do banco de dados base deste trabalho, que apresenta 4181 registros. Os registros retirados foram utilizados para testar o *screening tool* criado a partir da rede neural que apresentou o melhor desempenho.

É importante ressaltar que a melhor rede neural corresponde à que mais se aproximou da classificação feita pelo agrupamento utilizando a análise de *cluster*. Sendo assim, foi usada a base de dados com 4181 registros e os *clusters* correspondentes. O objetivo da utilização da rede neural foi desenvolver um código (função) capaz de explicar a classificação feita pela técnica *K-means*. Em adição, para gerar a rede neural foram usados os dados originais e não os normalizados.

3.3.1 Resultados

A matriz de confusão que mostra os casos onde os dados foram classificados corretamente e incorretamente, de acordo com a classificação previamente existente, do Quadro 2 mostra que a rede neural que melhor classifica os registros em relação à da técnica K-means é a MLP 6: 6-9-4: 1. O Quadro 3 de desempenho das redes neurais resume a informação da matriz de confusão.

Quadro 2: Matriz de Confusão

	Cluster	1	2	3	4
MLP 6: 6-9-7-4: 1	1	133	0	0	2
	2	0	3	0	0
	3	0	0	3903	0
	4	0	2	0	19
RBF 6: 6-127-4: 1	1	132	0	1	2
	2	0	3	0	1
	3	1	2	3092	0
	4	0	0	0	18
MLP 6: 6-7-4: 1	1	133	0	0	2
	2	0	5	0	0
	3	0	0	3093	0
	4	0	0	0	19
MLP 6: 6-9-4: 1	1	133	0	0	0
	2	0	5	0	0
	3	0	0	3093	0
	4	0	0	0	21
MLP 6: 6-7-7-4: 1	1	132	0	1	0
	2	0	5	0	0
	3	0	0	3092	0
	4	1	0	0	21

Quadro 3: Desempenho das Redes Neurais

	MLP 6: 6-9-7-4: 1				RBF 6: 6-127-4: 1				MLP 6: 6-7-4: 1			
	Cluster				Cluster				Cluster			
	1	2	3	4	1	2	3	4	1	2	3	4
Total	133	5	3903	21	133	5	3903	21	133	5	3903	21
Acertos	133	3	3903	19	132	3	3902	18	133	5	3903	19
Erros	0	2	0	2	1	2	1	3	0	0	0	2
Desconhecido	0	0	0	0	0	0	0	0	0	0	0	0
Acertos (%)	100	60	100	90,5	99,3	60	100	85,7	100	100	100	90,5
Erros (%)	0	40	0	9,25	0,75	40	0,03	14,3	0	0	0	9,52
Desconhecido (%)	0	0	0	0	0	0	0	0	0	0	0	0
	MLP 6: 6-9-4: 1				MLP 6: 6-7-7-4: 1							
	Cluster				Cluster							
	1	2	3	4	1	2	3	4				
Total	133	5	3903	21	133	5	3903	21				
Acertos	133	5	3903	21	132	5	3902	21				
Erros	0	0	0	0	1	0	1	0				
Desconhecido	0	0	0	0	0	0	0	0				
Acertos (%)	100	100	100	100	99,3	100	100	100				
Erros (%)	0	0	0	0	0,75	0	0,03	0				
Desconhecido (%)	0	0	0	0	0	0	0	0				

A análise de sensibilidade das redes neurais do Quadro 4 permite avaliar o grau de confiança dos resultados em situações de decisões incertas ou suposições sobre os dados e resultados utilizados. Sendo assim, ela é fundamental para investigar a heterogeneidade entre os dados. Esta análise é semelhante à análise de comunalidade, onde se verifica a importância de cada variável e sua contribuição no modelo proposto. Nesse caso, a variável que mais contribui para a rede neural MLP 6: 6-9-4: 1 é o valor do ICMS.

Quadro 4: Análise de Sensibilidade das Redes Neurais

	Consumo Lido	Consumo em 30 dias	Valor do Importe	Valor do ICMS	Valor do Corte de Reposição	Valor Total da Fatura
MLP 6: 6-9-7-4: 1						
Razão	7,43	6,8	8,95	7,37	5,96	7,92
Importância	3	5	1	4	6	2
RBF 6: 6-127-4: 1						
Razão	2,29	2,12	2,07	2,11	2,74	1,64
Importância	2	3	5	4	1	6
MLP 6: 6-7-4: 1						
Razão	4,1	6,11	7,63	7,78	5,14	6,12
Importância	6	4	2	1	5	3
MLP 6: 6-9-4: 1						

Razão	10,64	6,08	9,21	12,89	8,25	7,94
Importância	2	6	3	1	4	5
MLP 6: 6-7-7-4: 1						
Razão	5,8	6,82	11,03	10,58	9,1	17,32
Importância	6	5	2	3	4	1

4 CONSIDERAÇÕES FINAIS

Como ressaltado por Hair (1998), na grande maioria dos mercados não é possível se guiar pelo pensamento simplista que considera os consumidores homogêneos e caracterizados por um pequeno número de variáveis demográficas. As empresas estão atendendo cada vez mais segmentos de consumidores com diferentes características demográficas e comportamentais e inúmeras implicações legais, econômicas e tecnológicas. As técnicas de análise multivariada são a única forma de entender e conhecer as múltiplas relações desse cenário atual.

Além disso, a tendência a partir de agora é que a eficiência e eficácia da estratégia de CRM definam a nível competitividade das empresas. Desse modo, é imprescindível que as etapas da estratégia de CRM sejam muito bem planejadas e executadas. A primeira etapa, que consiste na identificação dos clientes, é crucial. Um banco de dados ou *Datawarehouse* contendo poucas variáveis demográficas e, principalmente, comportamentais afeta diretamente a etapa de diferenciação dos clientes e, possivelmente, a eficácia da estratégia de CRM. Entretanto, muitas empresas ainda não tratam a questão da identificação dos clientes com a seriedade que deviam.

Este trabalho mostrou como algumas das técnicas de análise multivariada podem e devem auxiliar administradores da área de *marketing*. Conhecer e usar essas técnicas são imprescindíveis para uma empresa que almeja ser competitiva.

5 BIBLIOGRAFIA

- CHURCHILL, Gilbert A. Jr; PETER, J. Paul. **Marketing: criando valor para os clientes**. Tradução BARTALOTTI, Cecília Camargo; MOREIRA, Cidd Knipel. São Paulo: Saraiva, 2005. 626 p.
- CISTER, Angelo Maia. **Mineração de dados para a análise de atrito em telefonia móvel**. 2005. 150 f. Tese (Doutorado em Engenharia) – Universidade Federal do Rio de Janeiro, Instituto de Pós-Graduação em Engenharia – COPPE-UFRJ, Rio de Janeiro, 2005.
- COUTINHO, Fernando V.; Datamining; www.dwbrasil.com.br/2003.
- GARRAFONI, Atílio Júnior; SANTADE, Hélio Oliva; PIZZINATTO, Nadia Kassouf; FARAH, Osvaldo Elias. CRM: conceitos e métodos de aplicação no marketing de relacionamento. **Revista Gestão Industrial**, Paraná, v. 01, n. 03, p.13-23, 2005.
- GREENBERG, P. **CRM at the Speed of Light: Capturing and Keeping Customers in Internet Real Time**. New York: McGraw-Hill, 2002.
- GRÖNOSS, C. Value-driven relational marketing: from products to resources and competencies. **Journal of Marketing Management**, U.K., v .13, p. 407-426, 1997.
- HAIR, Joseph F. Jr.; ANDERSON, Rolph E.; TATHAM, Ronald L.; BLACK, William C. **Multivariate data analysis**. New Jersey: Prentice-Hall, 1998. 730 p.
- LANINI, Fabiano Durão. **A Sofisticação da Gestão e o CRM em Agências de Viagens e Turismo Operando no Brasil**. 2005. 158 f. Dissertação (Mestrado em Administração) – Universidade Federal do Rio de Janeiro, Instituto de Pós-Graduação em Administração, Rio de Janeiro, 2005.
- MÖLLER, K.; WILSON, D.T. **Business Marketing: an interaction and network perspective**. Kluwer Academic, Norwell, MA, 1995.
- MORGAN, R.M.; HUNT, S.D. The commitment-trust theory of relationship marketing. **Journal of Marketing**, ESTADO, v .58, p. 20-38, jul., 1994.
- PAYNE, A.; CRISTOPHER, M.; CLARK, M.; PECK, H. **Relationship marketing for competitive advantage**. Oxford: Butterworth Heinemann, 1999.
- PEPPERS, Don; ROGERS, Martha. **One to One Manager: Real World Lessons in Customer Relationship Management**. The New York: Currency/Doubleday, 1999.
- PLATA, Sandra. De acordo com a vontade do cliente. Disponível em: <http://www.intermanagers.com.br/archivo/nota.jsp?id=707>. Acesso em: 10 out. 2003.
- RESE, Mario. Relationship marketing and customer satisfaction: an information economics perspective. **Marketing Theory**, v .3, n. 1, p. 97-117, 2003.
- SCREMIN, Marcos Antônio Antonello. **Método para a seleção do número de componentes principais com base na lógica difusa**. 2003 124 f. Tese (Doutorado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Engenharia de Produção, Santa Catarina, 2003.