**Universidade Presbiteriana Mackenzie**
**Faculdade de Ciências Econômicas, Contábeis e Administrativas**
**Programa de Pós-Graduação em Administração de Empresas**

# Descoberta de conhecimento em bases de dados e estratégias de relacionamento com clientes: Um estudo no setor de serviços

**Marcelo Pires Fernandes**

**São Paulo**
**2007**

**Marcelo Pires Fernandes**

# Descoberta de conhecimento em bases de dados e estratégias de relacionamento com clientes: Um estudo no setor de serviços

**Dissertação apresentada ao Programa de Pós Graduação em Administração de Empresa da Universidade Presbiteriana Mackenzie para a obtenção do título de Mestre em Administração de Empresas**

**Orientador: Prof. Dr. Sílvio Popadiuk**

**São Paulo**

**2007**

**Reitor da Universidade Presbiteriana Mackenzie**
**Professor Dr. Manasses Claudino Fonteles**

**Coordenadora Geral da Pós-Graduação**
**Professora Dra. Sandra Maria Dotto Stump**

**Diretor da Faculdade de Ciências Econômicas Contábeis e Administrativas**
**Professor Dr. Reynaldo Cavalheiro Marcondes**

**Coordenadora do Programa de Pós-Graduação em Administração de Empresas**
**Professora Dra. Eliane Pereira Zamith Brito**

***Dedico esta dissertação às pessoas especiais da minha vida: meus pais Dilvo e Sônia, meus irmãos Fábio e Daniel e minha esposa Érica, pelos ensinamentos e pelo apoio incondicional, à minha filha Marina, que está prestes a chegar e, sobretudo, a Deus, razão da minha existência.***

**Agradeço, inicialmente, ao professor e orientador, Prof. Dr. Sílvio Popadiuk, por ter me orientado e apoiado no desenvolvimento deste trabalho e por todo o carinho e paciência demonstrados ao longo dos últimos 18 meses.**

**Aos Professores, Prof. Dr. Alberto Luiz Albertin e à Prof[a] Dr[a] Maria Virginia Llatas, pelos valorosos comentários acerca deste trabalho e pelas grandes realizações para a realização deste estudo.**

**Agradeço à Universidade Presbiteriana Mackenzie, por fornecer acesso amplo e irrestrito aos materiais de pesquisa necessários para o desenvolvimento desta dissertação.**

**Agradeço à Redecard S.A. por ter possibilitado a realização do mestrado, em especial a Patrício Laguna e Edson Santos.**

**Agradeço, especialmente, ao professor e ex-colega de trabalho Johan Hendrik Poker Jr., pelo constante incentivo e estímulo no desenvolvimento de uma carreira acadêmica.**

**Agradeço principalmente a Deus, pela força, vontade de viver, entusiasmo e desejo incessante de ser melhor a cada dia.**

**Resumo**

O problema de pesquisa a ser investigado está associado ao modo como empresas do setor de serviços utilizam bases de dados para descobrir conhecimento sobre o cliente e embasar o desenvolvimento de estratégias de relacionamento. Este tema é importante, visto que em função do aumento da concorrência e da exigência dos clientes, as empresas precisam tratar seus clientes de forma diferenciada, de forma a manter em sua carteira aqueles mais rentáveis. Neste sentido, a literatura tem sugerido uma integração cada vez mais intensa entre disciplinas como Marketing de Relacionamento, CRM e Mineração de Dados. O presente trabalho estudou o modo como a literatura apresenta e descreve processos de análise de bases de dados e algumas propostas foram encontradas, propostas que segmentam o processo de descoberta de conhecimento em bases de dados em etapas como entendimento do problema, entendimento e preparação dos dados, modelagem dos dados, avaliação do modelo e implementação da solução desenvolvida. O universo estudado foi o de empresas do setor de serviços que atuam nas cidades de São Paulo e do Rio de Janeiro e uma pesquisa quantitativa foi realizada por meio da aplicação de um questionário a 67 respondentes. Nesta pesquisa, foi investigado o nível de utilização das etapas dos processos de descoberta de conhecimento em bases de dados, as técnicas de mineração utilizadas, bem como as estratégias de relacionamento adotadas com clientes. Constatou-se que as empresas pesquisadas possuem um alto nível de utilização das etapas de descoberta de conhecimento identificadas na literatura, que elas utilizam de forma uniforme apenas algumas das técnicas de mineração de dados identificadas na literatura e que, do ponto de vista de estratégias de relacionamento com clientes, as estratégias de aquisição de novos clientes e identificação dos melhores clientes possuem um nível de utilização superior ao de estratégias de retenção de clientes (considerando resultados da amostra). Esta última constatação, de certo modo, contraria o pensamento de algumas correntes teóricas, que defendem que as empresas devem focar suas estratégias de relacionamento na retenção de clientes. Estes resultados pode servir de apoio aos gestores das empresas, no que se refere aos processos de desenvolvimento de estratégias de relacionamento com clientes, sustentados em análise integrada dos aspectos de negócio envolvidos, informações sobre o cliente, bem como modelos quantitativos de análise destas informações, de forma a transformá-las em conhecimento útil para a tomada de decisão.

Palavras-chave: Estratégias de Relacionamento, Descoberta de Conhecimento em Bases, Mineração de Dados

**Abstract**

The research problem to be studied is related to the way companies from the services industry use customer databases to discover useful knowledge about their customers, in order to improve the development of relationship strategies with them. This issue is important mainly because due to the increasing of concurrence and customer demand, the company needs to relate differently with their customers, so that thy can keep in its portfolio the most profitable ones. In this way, the theory has suggested a deeper integration among distinct disciplines as Relationship Marketing, CRM and Data Mining. In this current study, it was investigated the way the theory presents and describes database analysis processes and, as a result, some proposals were found out, that segment the processes of discovering knowledge in databases in stages like problem understanding, data understanding, data preparation, data modeling data, model evaluation and deployment. The target population was composed by companies from the services industry from São Paulo and Rio de Janeiro cities and a quantitative research was made by applying a questionnaire to 67 professionals from the target population. In this research, themes as utilization level from stages of process of discovering knowledge in databases, utilization level of data mining techniques and utilization level of relationship strategies were investigated. It was discovered that the companies researched have a high utilization level of the stages of knowledge discovery identified in the theory, just only a small part of the data mining techniques are uniformly used by the companies researched and, at last, the strategies with the highest utilization levels are that related to the acquisition of new customers and identification of profitable ones. This last discover was a little bit surprising, because it is opposed to the way of thinking of some authors who defend companies should focus on their relationship strategies in the customer retention. These results can be used to support companies, in subjects related to the development of customer relationship strategies, based in an integrated analysis of business issues, customer information, as well quantitative models of analysis from this information, in order to turn it into useful knowledge to the making decision.

Keywords: Relationship Strategies, Knowledge Discovery in Databases, Data Mining.

# SUMÁRIO

## LISTA DE TABELAS

## LISTA DE QUADROS:

## LISTA DE FIGURAS:

## SIGLAS UTILIZADAS

- **AMA –** *American Marketing Association*
- **CART ou C&RT –** *Classification and Regression* Trees – Árvores de classificação e regressão
- **CHAID –** *Chi-Square Automatic Interaction Detection* – Detecção automática de interações via qui-quadrado.
- **CMP –** Clientes de maior valor potencial
- **CMV –** Clientes de maior valor atual
- **CRISP-DM** – *Cross Industry Standard Process for Data Mining* – Processo padrão da indústria para mineração de dados
- **CRM –** *Customer Relationship Management* – Gestão do relacionamento com clientes
- **GLTV –** *Generalized Lifetime Value* – Valor vitalício generalizado do cliente
- **GRI –** *Generalized Rule Induction* – Regra generalizada de indução
- **KDD –** *Knowledge Discovery in Databases* – Descoberta de conhecimento em bases de dados
- **LTV –** *LifetimeVvalue* – Valor Vitalício do Cliente
- **OLAP –** On-line Analytical Processing – Processamento analítico online
- **SPSS –** *Statistical Package for the Social Sciences*
- **SAS –** *Statistical Analysis System*

# 1. INTRODUÇÃO

Níveis crescentes de concorrência e de exigência dos clientes têm obrigado as empresas a se adequarem a este novo ambiente econômico. Isso tem provocado uma intensa preocupação com melhoria da qualidade, oferta de novos serviços e redução dos custos, com o objetivo de manterem-se competitivas em um cenário turbulento e em constante mudança.

Barney (2002) defende que para as empresas manterem-se competitivas e obterem retornos financeiros acima da média de seu setor, devem ser capazes de utilizar recursos que sejam valiosos, raros, difíceis de serem imitados e que possam ser implementados. Kogut e Zander (1992) defendem que o conhecimento é o mais importante recurso estratégico e a habilidade de adquirir, integrar, armazenar, compartilhar e aplicá-lo é a capacidade mais importante para construir e sustentar a vantagem competitiva, visão esta também compartilhada por Nonaka e Takeuchi (1997).

Tapscott (1998) conceitua o conhecimento como uma informação que foi contextualizada e analisada de forma a tornar-se significativo e, portanto, apresentar valor para a empresa. Liebowitz e Beckman (1998) consideram que o conhecimento diz respeito à informação aplicada que leva ativamente à execução de tarefas, resolução de problemas e à tomada de decisões.

Davenport (2001) argumenta que um dos conhecimentos mais importantes que a empresa precisa ter é o conhecimento a respeito de seus clientes, de modo a ter condições de oferecer produtos e serviços que estejam adequados às suas necessidades. Alavi e Leidner (2001) defendem que muito mais que a geração e posse de conhecimento, a utilização do conhecimento adquirido é que torna a empresa competitiva no mercado em que atua. Com o desenvolvimento tecnológico e uma capacidade cada vez maior de armazenar dados, muitas empresas têm gerado e armazenado grandes quantidades de dados sobre seus clientes. Porém, de acordo com Davenport (2006), o que diferencia as empresas não é sua capacidade de armazenamento de dados, mas sim sua capacidade de transformar grandes volumes de dados sobre os clientes em conhecimento que pode ser utilizado para a tomada de decisão da empresa, conhecimento este que é fruto de interpretação humana. O autor ainda destaca a importância de se criar uma cultura analítica na empresa, de modo a torná-la capaz de se diferenciar em um mercado em que empresas oferecem produtos e serviços similares e utilizam tecnologias comparáveis.

Parvatiyar e Sheth (2001) destacam que se tornou conhecimento comum que o valor de todos os clientes não é igual, indicando que por volta de 20% dos clientes representam mais de 80% da receita para a maioria das empresas e que, em algumas delas, um percentual ainda menor de clientes pode gerar até 90% da receita das mesmas. Os autores ressaltam que as empresas precisam direcionar esforços de marketing para aqueles clientes com maior potencial de geração de receita.

Muito do conhecimento sobre os clientes é proveniente de seu histórico comportamental com relação à utilização de produtos e serviços oferecidos pela empresa, dados estes que a maioria delas dispõe em seus *datawarehouses* ou em bases de dados espalhadas pela empresa (BERRY e LINOFF, 2000) . O desafio é como utilizar o volume cada vez maior de dados sobre clientes e transformá-los em conhecimento útil para a geração de estratégias de relacionamento com clientes (BERSON, SMITH e THEARLING, 1999). Autores do campo do marketing de relacionamento (BERRY, 1983; GRÖNROOS, 1996; DAY, 2003) defendem que os clientes devem ser tratados de forma diferente e que as informações que a empresa dispõe sobre eles são importantes para realizar esta diferenciação.

Uma possível abordagem para avaliar como as empresas analisam dados sobre clientes está no estudo de processos de descoberta de padrões comportamentais escondidos entre grandes volumes de dados armazenados pelas empresas (FAYYAD *et al.*,1996). Este estudo dos padrões de comportamento presentes nos históricos transacionais dos clientes permite às empresas identificar quais são os clientes atualmente mais rentáveis, quais os mais propensos a cancelar um relacionamento com a empresa ou mesmo quais são aqueles com maior potencial de receita nos próximos meses (BERRY e LINOFF, 2004).

O tema de estudo desta dissertação é o processo pelo qual as empresas analisam bases de dados para extrair conhecimento e usá-lo como subsídio à criação de estratégias de relacionamento com clientes.

## 1.1 Justificativas

Os elementos que justificam o estudo deste tema são os seguintes:

a) Este é um tema discutido na literatura internacional nos últimos quinze anos, de exploração relativamente recente, mas que tem sido abordado com muita freqüência nos últimos cinco anos, sobretudo fora do Brasil.

b) Há relativamente poucos registros sobre estudos realizados no Brasil, que tenham analisado a importância de processos estruturados de análise de bases de dados para a descoberta de conhecimento sobre o cliente, sendo esta uma lacuna existente na literatura nacional.

c) Motivações pessoais e profissionais quanto aos aspectos da contribuição do processo de descoberta de conhecimento em bases de dados para a potencialização do relacionamento com clientes e disposição em agregar conhecimento relevante sobre o tema à literatura acadêmica nacional e à amplitude do assunto em empresas do setor de serviços.

## 1.2 Contribuições do estudo

A partir desta dissertação, tem-se a pretensão de contribuir academicamente com o estudo de processos de análises de bases de dados para a descoberta de conhecimento sobre o cliente, bem como sua relevância estratégica para a empresa. Também serve como fonte de pesquisa sobre o tema para profissionais que desenvolvam estratégias de relacionamento com clientes e que pretendam obter um panorama mais amplo a respeito de como as empresas fazem uso das bases de dados para chegar a estas estratégias.

Contudo, a maior contribuição deste estudo é o fornecimento de uma visão abrangente da utilização de ferramentas de análise de bases de dados para suportar o desenvolvimento de estratégias de relacionamento com clientes.

## 1.3 Problema de pesquisa

*Como as empresas utilizam bases de dados para descobrir conhecimento sobre o cliente e subsidiar a criação de estratégias de relacionamento ?*

## 1.4 Objetivos da dissertação

**Objetivo geral**: O objetivo geral desta dissertação é identificar como as empresas utilizam e analisam bases de dados para subsidiar a criação de estratégias de relacionamento com clientes.

O estudo será realizado considerando empresas do setor de serviços que atuam nas cidades de São Paulo e do Rio de Janeiro, pois é um setor em que se identificam alguns segmentos (finanças, telecomunicações e seguros) com aplicações da mineração de dados ao relacionamento com clientes, tanto em empresas que adotam CRM (do inglês, *Customer Relationship Management*), quanto naquelas que não possuem programas formais de relacionamento com seus clientes (PEACOCK, 1998; DAVENPORT, 2001; DREW *et al*., 2001). Estas cidades foram escolhidas por sua representatividade no cenário econômico nacional e pela concentração de empresas de serviço nestes locais. Os **objetivos específicos** que auxiliam a responder o objetivo geral são os seguintes:

a) Avaliar o nível de utilização, por parte das empresas de serviços que atuam na cidade de São Paulo e do Rio de Janeiro, das etapas dos processos de análise e descoberta de conhecimento em bases de dados, identificadas na fundamentação teórica (item 2), de modo a subsidiar a criação de estratégias de relacionamento;

b) Avaliar o nível de utilização, nas empresas de serviços que atuam nas cidades de São Paulo e do Rio de Janeiro, das técnicas de mineração de dados identificadas na fundamentação teórica (item 2), de modo a subsidiar a criação de estratégias de relacionamento;

c) Avaliar o nível de utilização, por parte das empresas de serviços que atuam nas cidades de São Paulo e do Rio de Janeiro, das estratégias de relacionamento com clientes identificadas na fundamentação teórica (item 2);

d) Avaliar se o nível de utilização das técnicas de mineração de dados para geração de estratégias de relacionamento com clientes tem relação com a existência de um CRM analítico na empresa;

e) Verificar se o nível de utilização das etapas dos processos de análise e descoberta de conhecimento em bases de dados para geração de estratégias de relacionamento com clientes tem relação com variáveis intrínsecas à empresa, como segmento de atuação na área de serviços, faturamento anual, quantidade de clientes e nacionalidade da empresa.

## 1.5 Estrutura desta dissertação

Esta dissertação de mestrado foi estruturada em seis itens principais:

O item 1, introdução da dissertação, destacou aspectos a serem considerados pela empresa para adaptarem-se às mudanças do cenário econômico, em especial a importância do conhecimento sobre o cliente para a criação e diferenciação de estratégias de relacionamento. Também descreveu o tema a ser estudado, bem como o problema de pesquisa, os objetivos gerais e específicos, além das justificativas e motivações para a escolha deste tema.

O item 2, fundamentação teórica, resgata a visão da literatura pesquisada acerca de estratégias de relacionamento com clientes, processos de análise de dados e descoberta de conhecimento em bases de dados e, dentro desses processos, as técnicas de análise de dados mais comumente relatadas na literatura, referentes a aplicações em marketing e relacionamento com clientes. Também aponta algumas aplicações encontradas na literatura sobre a mineração de dados para o desenvolvimento de estratégias de relacionamento com clientes no setor de serviços.

O item 3, procedimentos metodológicos, retrata o processo utilizado para responder ao problema de pesquisa e aos objetivos (geral e específicos) desta dessa dissertação.

O item 4, análise e interpretação dos resultados da pesquisa, traz a leitura do pesquisador obtida após a compilação dos resultados do estudo realizado.

O item 5, conclusões, limitações e propostas de estudos futuros, aborda as conclusões fornecidas por esta dissertação, as limitações do estudo, bem como propostas de análises e estudos futuros, que possam enriquecer o universo de pesquisa relacionado aos processos de descoberta de conhecimento contido em bases de dados.

O item 6, referências, traz toda a fonte bibliográfica (artigos, livros, revistas e outras referências) utilizada para o desenvolvimento do conteúdo apresentado nesta dissertação.

# 2. FUNDAMENTAÇÃO TEÓRICA

## 2.1 Definição e classificação de serviços

Lovelock e Wright (2006) conceituam serviço como um ato ou desempenho oferecido por uma parte a outra e que, embora possa estar relacionado a um produto concreto, o desempenho é essencialmente intangível e normalmente não resulta na propriedade de nada (KOTLER, 2000). Etzel, Walker e Stanton (2001) segmentam serviços em duas classes. Na primeira, estão serviços que são o propósito ou objeto principal de uma transação entre empresa e cliente e, na segunda classe, estão serviços que apóiam ou facilitam a venda de um outro produto ou serviço.

Kotler e Armstrong (2003) consideram que os serviços apresentam quatro características especiais, conforme mostra a Figura 1:

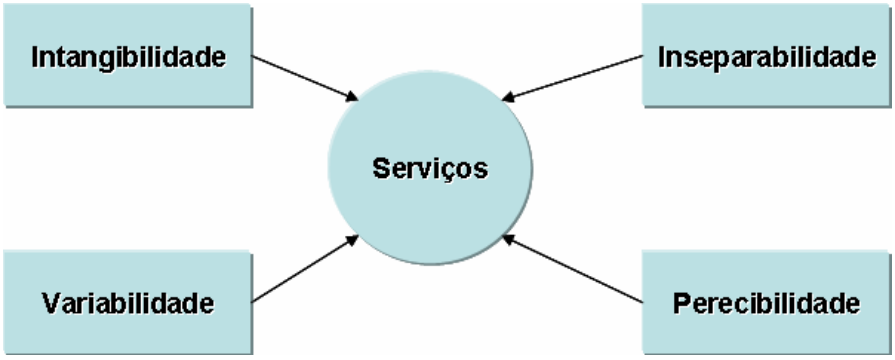


**Figura 1** - Características dos serviços
**Fonte:** Kotler e Armstrong (2003) – Adaptado pelo autor

A Figura 1 mostra que o serviço possui quatro características, associadas à intangibilidade, inseparabilidade, variabilidade e perecibilidade. A intangibilidade caracteriza que serviços não podem ser tocados, vistos, sentidos, cheirados ou ouvidos antes da compra. A inseparabilidade posiciona o serviço como não separável daquele que o provém. A variabilidade do serviço significa que a qualidade dos serviços depende de quem os executa e de quando, onde e como são executados. A perecibilidade significa que os serviços não podem ser armazenados para venda ou uso posterior. Kotler e Armstrong (2003) colocam que, como os serviços são diferentes de produtos tangíveis, muitas vezes eles exigem abordagens de marketing diferentes das utilizadas na venda de produtos. No caso de uma empresa que presta serviços, na maioria das vezes, o cliente e o empregado da linha de frente da empresa

interagem no momento da execução do serviço. Por essa razão, empresas que prestam serviço precisam interagir efetivamente com seus clientes, para criar valor superior durante a realização destes serviços.

Lovelock e Wright (2006), ao considerarem serviços a partir de uma perspectiva operacional, classificam serviços em quatro grandes grupos, com base em ações tangíveis, seja no corpo das pessoas ou nos bens dos clientes, e ações intangíveis nas mentes das pessoas ou em seus bens intangíveis, conforme retrata o Quadro 1:

**Quadro 1** – Natureza dos atos do serviço

| Qual a natureza do ato do serviço ? | Quem ou o quê é destinatário direto do serviço ? | |
|---|---|---|
| | Pessoas | Bens |
| Ações tangíveis | (Processamento com pessoas)<br>**Serviços dirigidos aos corpos das pessoas:**<br>Transporte de passageiros<br>Assistência médica<br>Hospedagem<br>Salões de beleza<br>Fisioterapia<br>Academias de ginástica<br>Restaurantes/bares<br>Barbearias<br>Serviços funerários | (Processamento com bens)<br>**Serviços dirigidos a posses físicas:**<br>Transporte de cargas<br>Reparo e manutenção<br>Armazenamento/estocagem<br>Zeladoria de edifícios<br>Distribuição de varejo<br>Lavanderias<br>Abastecimento de combustíveis<br>Paisagismo/jardinagem<br>Remoção e reciclagem de lixo |
| Ações intangíveis | (Processamento com estímulo mental)<br>**Serviços dirigidos às mentes das pessoas**:<br>Propaganda<br>Artes e entretenimento<br>Transmissões de rádio e TV<br>Consultoria administrativa<br>Educação<br>Serviços e informação<br>Concertos de música<br>Psicoterapia<br>Religião<br>Telefone | (Processamento com informações)<br>**Serviços dirigidos a bens intangíveis:**<br>Contabilidade<br>Finanças<br>Processamento de dados<br>Transmissão de dados<br>Seguros<br>Serviços jurídicos<br>Programação<br>Pesquisa<br>Investimentos<br>Consultoria de *software* |

**Fonte**: Lovelock e Wright (2006, p. 35)

O Quadro 1 apresenta os ramos de atividade classificados em quatro categorias básicas que, embora pareçam muito diferentes, apresentam características comuns relacionadas aos processos (LOVELOCK e WRIGHT, 2006). De acordo com os autores, os serviços, como processos, podem ser classificados em quatro categorias: processamento com pessoas, processamento com bens, processamento com estímulo mental e processamento com informações.

A categoria de serviços relacionada ao processamento com pessoas pressupõe que os clientes precisam estar fisicamente presentes durante a entrega do serviço para receber seus benefícios desejados. A categoria de processamento com bens considera que o objeto a ser processado precisa estar presente, mas o cliente não necessariamente precisa estar. A categoria de processamento com estímulo mental considera que os clientes devem estar mentalmente presentes, mas podem estar em uma instalação de serviço específica ou em um local remoto conectado por sinais de radiodifusão ou ligações de telecomunicação.A categoria de processamento com informações exige pouco envolvimento direto com o cliente, uma vez que o pedido de serviço foi iniciado.

Dentro do setor de serviços, alguns segmentos têm demonstrado maior preocupação com o armazenamento de dados e utilização dos mesmos para tomada de decisão, especialmente no que se refere ao relacionamento com seus clientes. Segmentos de serviços financeiros (bancos e financeiras), seguros, telecomunicações e hotéis são alguns dos segmentos que têm concentrado empresas preocupadas com o armazenamento de dados sobre seus clientes e a utilização do conhecimento obtido a partir destes dados para estabelecerem um relacionamento mais estreito com os clientes e manterem sua posição competitiva no mercado em que atuam (SMITH, WILLIS e BROOKS, 2000 ; DREW *et al.*, 2001 ; CHYE e GERRY, 2002; HORMAZI e GILES, 2004).

Em função do volume de dados armazenado e da diversidade de serviços oferecidos pelas empresas, a complexidade de análise dos dados tem se tornado crescente, exigindo que estas desenvolvam mecanismos de tratamento e análise dos dados que suportem o desenvolvimento das estratégias de relacionamento com seus clientes.

## 2.2 Estratégias de relacionamento com clientes

### 2.2.1 Do marketing transacional ao marketing de relacionamento

O foco do marketing transacional é garantir a realização de um negócio entre as partes envolvidas (empresa x cliente). O objetivo central das empresas com a utilização do marketing transacional é garantir a venda e para isso praticam o marketing de massa, oferecem produtos e serviços idênticos a todos os clientes, ou então segmentam estes clientes em grupos com perfis semelhantes e adaptam seus produtos a cada um destes grupos (VALENTE, 2002). A medida de sucesso do negócio é a participação de mercado (*market share*), o que faz com que as empresas realizem grandes investimentos em propaganda, treinamentos em vendas e consolidem na mente do cliente a imagem da empresa de modo a convencê-lo a comprar o produto que oferecem (KOTLER, 2000).

No marketing transacional, as empresas têm seus produtos padronizados e dificilmente conseguem diferenciar seus clientes, o que faz com que se tornem intercambiáveis, ou seja, se a empresa perde um cliente e ganha outro, seu "mercado" permanece estável (VALENTE, 2002). Conforme destacam Peppers e Rogers (2004), por meio do marketing transacional, a empresa desenvolve uma comunicação unidirecional, através da mídia de massa, já que seu objetivo é o aumento da quantidade de clientes e da participação de mercado, mostrando que o foco desta estratégia está na aquisição de clientes.

Segundo Keefe (2004), a primeira definição de marketing, datada de 1935, conceituava o marketing como relacionado ao desempenho de atividades de negócio que direcionam o fluxo de bens e serviços dos produtores até os consumidores.

Esta primeira definição oficial de marketing foi adotada em 1935 pela Associação Nacional de Professores de Marketing, antecessora da AMA (*American Marketing Association*). Foi adotada pela AMA em 1948, e novamente em 1960, quando ela revisitou esta definição e decidiu não alterá-la. Essa definição durou por 50 anos, até que foi revista em 1985. A nova definição propunha o Marketing como o seguinte:

> O processo de planejar e executar a concepção, precificação, promoção e distribuição de idéias, bens e serviços para criar trocas que satisfaçam objetivos individuais e organizacionais (AMA, 1985).

Em 2004, a AMA novamente revisitou sua visão sobre o conceito de marketing, enfatizando o conceito de relacionamento com clientes. Esta evolução da tradicional conceituação de marketing, atribuindo especial destaque ao relacionamento com o cliente, é a evidência mais recente de que as empresas têm passado por uma profunda mudança na forma de fazer negócios. Segundo a definição mais recente:

> Marketing é uma função organizacional e uma série de processos para a criação, comunicação e entrega de valor para clientes, e para a gerência de relacionamentos com eles de forma que beneficie a empresa e seus *stakeholders*". (AMA, 2004).

Segundo Keefe (2004), esta definição traz conseqüências importantes para os negócios, pois coloca o cliente (e não o produto) como elemento central na razão de ser da empresa, estabelece uma visão de processos para atendê-lo, atribui papel fundamental e integrador ao marketing, e valoriza os relacionamentos com os clientes como estratégia de sobrevivência. Leite (2004, p. 65) apresenta as motivações para a adoção do marketing de relacionamento com os clientes, como estratégia de negócios. A autora defende que as novas estratégias de gestão baseadas no marketing de relacionamento com os clientes usam como pressuposto a criação de vínculos com os clientes com o objetivo de melhor conhecê-los para melhor atendê-los.

Apesar da clara tendência para a adoção do marketing de relacionamento com os clientes, Kotler (2000) lembra que a mudança para o marketing de relacionamento não significa que as empresas devam abandonar totalmente o tradicional marketing de transações. Para o autor, é importante considerar o perfil do mercado em que a empresa atua, para então definir a melhor estratégia de negócios. De acordo com Day (2001), há importantes razões para que as empresas passem de uma cultura de transações para uma cultura de relacionamentos com clientes, quais sejam: menores custos de atendimento, compras maiores, menor sensibilidade a preços e divulgação boca a boca favorável.

Observando o conceito de marketing de relacionamento, tem-se que o mesmo, segundo a visão de Berry (1983), está relacionado ao processo de atração, manutenção e relacionamento com clientes. Grönroos (1996, p. 11) apresenta uma definição convergente com a definição de Berry (1983), apontando o marketing de relacionamento como o processo de identificação, estabelecimento, manutenção e aperfeiçoamento do relacionamento com clientes e outras partes interessadas (empregados, fornecedores, acionistas), de forma que os objetivos de todas as partes interessadas sejam atendidos. Aijo (1996, p. 15) resume de forma objetiva a amplitude do marketing de relacionamento, ressaltando que há um consenso

crescente na definição de marketing de relacionamento como envolvendo uma relação de longo prazo entre participantes envolvidos em processos de troca de valor. No conceito estabelecido por Vavra (1993), o marketing de relacionamento é considerado essencialmente como a retenção de clientes e ele ainda apresenta táticas denominadas *aftermarketing*, com o objetivo de manter o cliente em contato com a empresa depois da realização de uma compra.

Conforme destaca Swift (2001), o marketing de relacionamento não permite argumentos que tenham como objetivo enganar os clientes buscando apenas a realização de uma venda, pois, neste momento, mais do que nunca, os clientes tornaram-se importantes, tanto do ponto de vista interno das empresas, quanto do ponto de vista do mercado. O marketing de relacionamento busca o aumento da participação em longo prazo, porém de forma duradoura. O marketing transacional, por outro lado, deseja resultados imediatos, nem sempre duradouros.

A questão “tempo” é importante neste caso. Jackson (1985) destaca que o marketing de relacionamento não é eficaz em todas as situações. Ela percebe o marketing transacional como mais adequado com clientes que têm um horizonte curto de tempo e baixos custos de mudança. Por outro lado, a autora destaca que investimentos em marketing de relacionamento oferecem retorno com clientes que possuem horizontes longos de tempo e altos custos de mudança.

### 2.2.2 A gestão do relacionamento com clientes (CRM)

O marketing de relacionamento está fundamentado não somente na relação empresa x cliente (GUMMESSON, 2002), mas principalmente na relação da empresa com todas as partes envolvidas no negócio da mesma, fornecedores, empregados, clientes e até mesmo o governo (MORGAN e HUNT, 1994). O CRM (*Customer Relationship Management*) ou gestão do relacionamento com clientes, tem uma preocupação mais acentuada com a gestão da relação que a empresa tem com seus clientes finais. Conforme aponta Brown (2001), o CRM já é um conceito antigo se for considerado que era a maneira como os comerciantes costumavam tratar seus clientes. A novidade é que, agora, este conceito pode ser massificado, podendo ser aplicado para uma grande quantidade de clientes.

A visão de Sin, Tse e Yim (2005) situa o CRM como um processo e estratégia detalhados que capacitam a empresa a identificar, adquirir, reter e nutrir clientes rentáveis através da construção e manutenção de relacionamentos de longo prazo com eles. De acordo

com Kotler (2000), um cliente rentável ou lucrativo é aquele que, ao longo do relacionamento com a empresa, rende um fluxo de receita que supera os custos de atração, venda e atendimento da empresa relativo a ele.

Estas definições indicam que o foco central do CRM e das perspectivas de marketing de relacionamento está na relação entre comprador e vendedor, que estas relações são longitudinais por natureza e que ambos se beneficiam desta relação. Da perspectiva da empresa, tanto os conceitos de CRM quanto de marketing de relacionamento podem ser vistos como associados a uma cultura organizacional que coloca a relação comprador-vendedor no centro do pensamento operacional e estratégico da empresa (SIN, TSE e YIM, 2005). Conforme apontam Ryals e Payne (2001), o marketing de relacionamento é relativamente mais estratégico, enquanto o CRM é utilizado em um sentido mais tático. Yau *et al.* (2000) destacam que o marketing de relacionamento é relativamente mais emocional e comportamental, centrando-se em variáveis como empatia, reciprocidade e confiança. Por outro lado, o CRM está focado em avaliar como o gerenciamento pode ser feito para a atração, manutenção e aperfeiçoamento do relacionamento com clientes. Parvatiyar e Sheth (2001) fornecem uma definição abrangente do que significa o CRM. Segundo os autores,

> CRM é uma estratégia detalhada e processo de aquisição, retenção e parceria com clientes selecionados, de forma a criar valor para a empresa e para o cliente. Isto envolve a integração de marketing, vendas, serviços a clientes, e as funções da cadeia de fornecimento da empresa para atingir maior eficiência e efetividade na entrega de valor ao cliente (PARVATIYAR e SHETH, 2001).

Fuller e Lewis (2002) caracterizam uma estratégia de relacionamento como um conceito utilizado para explicar como a empresa planeja e procura colocar em prática um particular método para lidar com suas partes interessadas, mais especificamente os clientes.

Brown (2001) posiciona o CRM como uma estratégia de negócios que visa entender, antecipar e administrar as necessidades dos clientes potenciais de uma empresa. Ele ainda ressalta que o CRM se apresenta como uma jornada de estratégias, processos, mudanças organizacionais e técnicas pelas quais a empresa pretende administrar melhor seu próprio empreendimento com relação ao comportamento de seus clientes (BROWN, 2001). Mais importante, acarreta em adquirir e distribuir conhecimento sobre os clientes e usar estas informações por meio de vários pontos de contato para equilibrar rendimentos e lucros com o máximo de satisfação para os mesmos (BROWN, 2001).

### 2.2.3 Estratégias de relacionamento à luz do ciclo de vida do cliente

Kotler (2000) destaca o seguinte ponto no que se refere ao modo como as empresas utilizam o marketing de relacionamento:

> Infelizmente, a maior parte da teoria e prática de marketing concentra-se na arte de atrair novos clientes, em vez de focar na retenção dos existentes (KOTLER, 2000, p. 69).

Segundo Kotler (2000), no passado, muitas empresas consideravam que o vínculo com os clientes estava garantido. Em função de talvez não possuírem muitas alternativas, todos os fornecedores possuíam deficiências em termos de atendimento e o mercado estava crescendo tão rapidamente que a empresa não se preocupava com a satisfação de seus clientes (KOTLER, 2000). Conforme destaca o autor, houve mudanças no cenário competitivo e os clientes de hoje são muito mais difíceis de serem agradados, são mais inteligentes, mais conscientes em relação aos preços, mais exigentes, perdoam menos as falhas da empresa e são abordados por mais concorrentes com ofertas iguais ou melhores. Desta forma, o desafio não é deixar os clientes satisfeitos, pois vários concorrentes podem fazer isso. O desafio é conquistar a fidelidade dos clientes (KOTLER, 2000).

Oliver (1997, p. 6) define fidelidade como “um compromisso forte em recomprar ou repatrocinar um produto ou serviço oferecido consistentemente no futuro, apesar das influências circunstanciais e tentativas de marketing, que podem acarretar um comportamento de troca”. Mckenna (1992, p.12) destaca que “o marketing moderno é uma batalha para obter a fidelidade dos clientes”. Segundo o autor, o interesse da empresa em construir e manter a fidelidade do cliente denota que ela não está apenas interessada em vender a qualquer custo, mas está concentrada em alcançar rentabilidade a longo prazo através da repetição de compra de produtos e/ou serviços e da retenção dos clientes.

As empresas têm percebido que, para manterem-se competitivas no mercado em que estão posicionadas, devem estar atentas à taxa de consumidores perdidos e desenvolver ações para reduzi-la (VALENTE, 2002). Brown (2001) defende que é mais lucrativo manter os clientes atuais do que adquirir novos clientes. Segundo ele, durante o desenvolvimento normal de um relacionamento com um cliente, o custo com marketing e vendas declina gradualmente, e o potencial para a melhora da margem bruta aumenta (BROWN, 2001). Swift (2001) argumenta que o CRM está baseado na premissa que custa menos manter os clientes atuais do que obter clientes novos

Outros autores como Day (2003), Greenberg (2001) e Reichheld (1996) também defendem que a empresa deve focar suas estratégias de relacionamento na retenção de clientes. Brown (2001, p. 14) traz um estudo da consultoria *PricewaterhouseCoopers* que mostra a evolução do lucro da empresa com o cliente à medida que o tempo de relacionamento evolui, conforme ilustra a Figura 2:

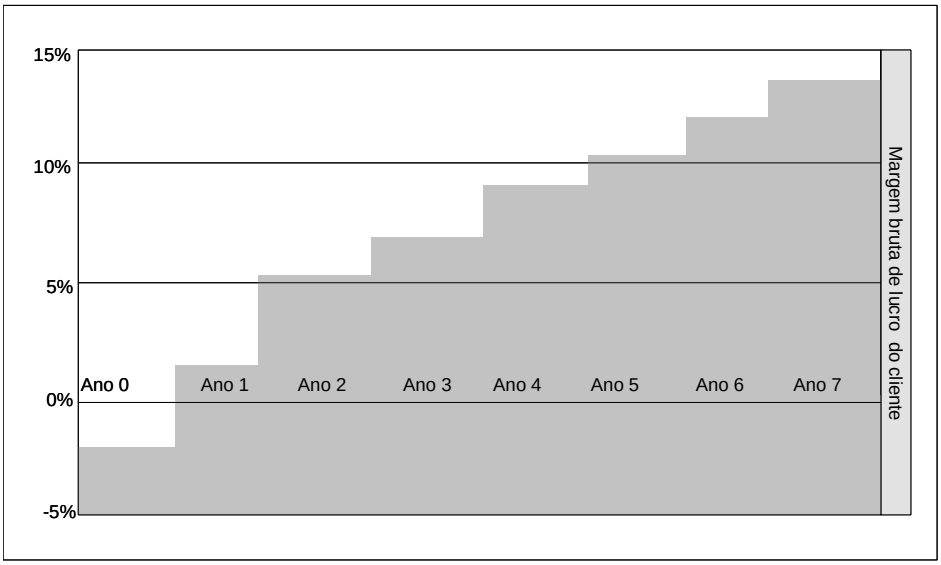


**Figura 2** - Evolução da margem bruta do cliente ao longo do tempo

**Fonte**: Brown (2001, p. 14)

A Figura 2 traz a evolução do valor da margem bruta do cliente ao longo de seu relacionamento com a empresa. Segundo relata Brown (2001), no início do relacionamento, o cliente reduz a margem bruta de lucro da empresa em 3% (em função dos custos de aquisição) e em três anos, esse cliente aumenta a margem bruta de lucro da empresa para 7%. Este estudo reforça as afirmações de Day (2003) e Reichheld (1996) quanto à importância da retenção de clientes e mostra que é mais lucrativo manter clientes atuais do que adquirir clientes novos e que o desenvolvimento gradual do relacionamento da empresa com o cliente faz com que sua margem de lucro bruta aumente.

Segundo Brown (2001), o relacionamento com o cliente nasce antes do momento da aquisição do mesmo, pois ocorre com a oferta de produtos e serviços que o futuro cliente, ainda na condição de *prospect* (ou potencial cliente) pode aceitar ou não. Quando ele aceita a oferta deste produto ou serviço, passa a tornar-se cliente da empresa.

A utilização de estratégias de relacionamento por meio do CRM se estende por todo o ciclo de vida do cliente (KAMAKURA *et al.*, 2005; KAMAKURA, 2002), incluindo estratégias de aquisição, desenvolvimento e retenção de clientes, conforme ilustra a Figura 3:

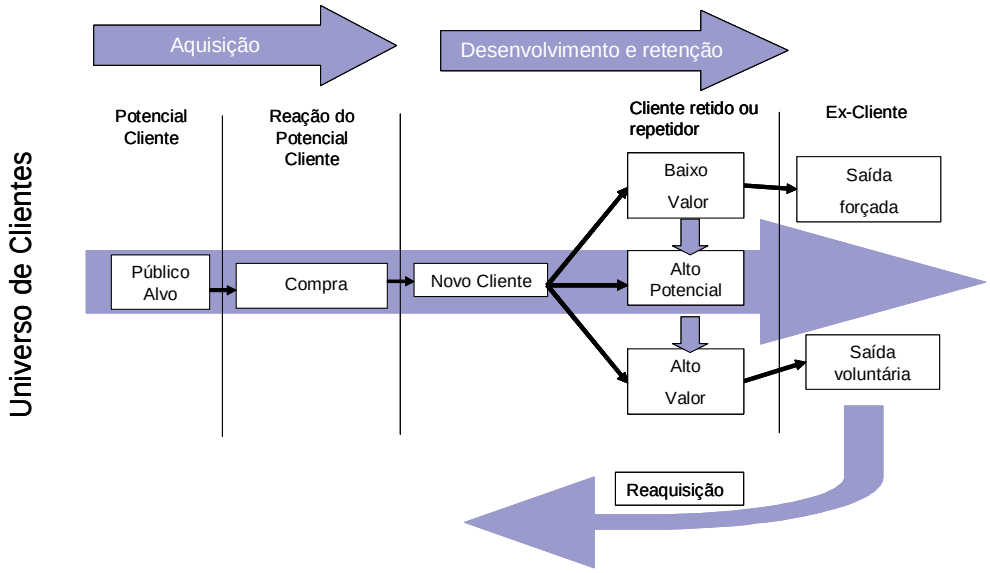


**Figura 3** - Ciclo de vida do cliente

**Fonte**: Adaptado de Kamakura (2002)

Conforme destaca Kamakura (2002), o ciclo de vida dos clientes descrito na Figura 3 é composto pelas etapas de aquisição, desenvolvimento e retenção. A empresa focaliza esforços em campanhas direcionadas ao seu público alvo, ou aqueles clientes potenciais cujas necessidades estão ajustadas ao produto ou serviço que deseja oferecer e adquire novos clientes que aderem a estas campanhas. Então, desenvolve estratégias para elevar o valor das receitas geradas pelos clientes, oferecendo produtos adequados às suas necessidades e gerenciando os canais de comunicação com o cliente, reduzindo custos com contato (KAMAKURA *et al.*, 2005; KAMAKURA, 2002). Ao longo do tempo, muitos clientes, de forma voluntária, passam a não mais desejar os produtos e serviços da empresa e rompem o relacionamento, muitas vezes mudando para um concorrente que oferece um produto ou serviço similar.

O desafio da empresa é identificar quem são os melhores clientes, em função do valor que trazem para a empresa, e retê-los, de forma a evitar que mudem para o concorrente, assim como precisa também deixar de gastar esforços com clientes que trazem baixo retorno à empresa (PARVATIYAR e SHETH, 2001). Segundo Kotler (2000), de forma a fortalecer a retenção de clientes, a empresa precisa construir barreiras à mudança, obtendo altos níveis de satisfação de clientes e buscando continuamente atender às necessidades por eles levantadas .

Greenberg (2001) comenta que, cada vez que um cliente se aproxima de seu negócio, ele possui uma expectativa, que pode ser constituída por uma necessidade de serviço ou interesse por um novo produto, mas, em qualquer caso, ele tem uma expectativa que

acompanha o interesse dele no negócio. Segundo o autor, o que acontece na seqüência criará uma experiência que modelará seu comportamento. Uma boa experiência pode elevar sua fidelidade e sua tendência a comprar novamente, enquanto que uma experiência ruim poderá transferi-lo para outro competidor. De acordo com Greenberg (2001), a habilidade da empresa em reconhecer este processo e ativamente gerenciá-lo representa a base para o relacionamento com clientes.

### 2.2.4 CRM sob o enfoque estratégico

Do ponto de vista de utilização do CRM como estratégia de relacionamento com clientes, Brown (2001) aponta a existência de estratégias, cuja utilização está relacionada ao momento do ciclo de vida do cliente no relacionamento com a empresa. São elas:

- **Busca de clientes em potencial**: Representa a busca e conquista de novos clientes, por meio de técnicas de segmentação (com base nas necessidades dos potenciais clientes) e de seletividade (buscando identificar quão lucrativos são os *prospects* ou potenciais clientes para a empresa).

- **Reconquista ou salvamento de clientes**: Tem o propósito de convencer o cliente a manter seu vínculo com a empresa no momento em que ele está abandonando o serviço, ou convencê-lo a voltar, quando este já tiver abandonado o vínculo com a empresa. Brown (2001) aponta que este tipo de ação é a que tem a maior sensibilidade com relação ao tempo, significando que uma campanha de reconquista tem quatro vezes mais chances de ser bem-sucedida se o contato com o cliente for realizado na primeira semana após o abandono do que ser for feito na quarta semana.

- ***Cross-selling* / *Up-selling***: O objetivo deste tipo de ação é elevar o valor que o cliente gasta com a empresa, identificando ofertas complementares que o cliente gostaria de receber, elevando, com isso, a participação no cliente (*wallet share*). Kotler e Armstrong (2003) defendem que as empresas precisam aumentar constantemente sua participação no cliente, seja transformando-se no único fornecedor de produtos/serviços que o cliente adquire, seja convencendo-o a comprar produtos/serviços adicionais da empresa. Eles ainda defendem que uma das melhores maneiras de elevar a participação no cliente é por meio do *cross-selling* (ou venda cruzada), que significa conseguir mais preferência dos clientes existentes para um produto ou serviço por meio da oferta de vendas adicionais ou complementares.

A diferença entre o *cross-selling* e o *up-selling* é que no caso do *cross-selling*, a empresa faz ofertas de produtos complementares e no caso do *up-selling*, faz a oferta do mesmo produto, só que aperfeiçoado (BROWN,2001). Brown (2001) acredita na importância deste tipo de ação, visto que como a oferta é dirigida a clientes que já têm relacionamento com a empresa, eles têm menos probabilidade de ver a oferta como uma *commodity* e são mais propensos a pagar um preço por elas.

**- Fidelização do cliente:** A empresa tenta evitar que o cliente a abandone e faz uso de três elementos essenciais: a segmentação com base no valor, a segmentação com base na necessidade dos clientes e mecanismos de previsão de desistência. A segmentação com base no valor possibilita à empresa determinar o quanto ela está disposta a investir na retenção do cliente (BROWN, 2001). Considerando aqueles clientes que a empresa está disposta a investir, pode ser utilizada a segmentação baseada nas necessidades, que pode ser constituída por programas de fidelidade adequados às necessidades dos clientes ou linhas especiais de serviços, adequados ao público em questão. Por fim, a empresa pode fazer uso de modelos preditivos de desistência, que têm como objetivo identificar clientes com maior tendência a romper o relacionamento com a empresa, de modo a tentar reverter o processo e mantê-los em sua base de clientes.

Peppers e Rogers (2004) concentram o CRM na aplicação de quatro estratégias essenciais no tocante ao relacionamento com clientes: Identificação, diferenciação, interação e personalização.

**- Identificação de clientes:** É preciso conhecer os clientes individualmente, com o maior número de detalhes possível, e ser capaz de reconhecê-los em todos os pontos de contato e de venda. Peppers e Rogers (2004) consideram que o marketing *one to one* deve ser aplicado para os melhores clientes e, para isso, torna-se necessário identificá-los, conhecendo suas características e sua história individual.

O conhecimento destas características não é um processo simples. São necessárias ferramentas e sistemáticas para sua viabilização, e deve-se trabalhar sempre com uma perspectiva de longo prazo para o refinamento das informações. Toda interação fornece informações que devem ser captadas, armazenadas e processadas para formar uma base de conhecimento sobre o cliente (PEPPERS e ROGERS, 2004).

**- Diferenciação de clientes**: Os clientes podem ser diferenciados de duas maneiras: pelo nível de valor para a empresa e pelas necessidades que têm de bens e serviços da empresa. Para Peppers e Rogers (2004, p. 40), “o objetivo da diferenciação de clientes é

encontrar os clientes de maior valor atual (CMV) e os clientes de maior valor potencial (CMP)". Estes são os clientes prioritários para se desenvolver uma relação de aprendizado.

**- Interação com os clientes:** Toda e qualquer interação com o cliente deve ser estabelecida dentro do contexto das outras interações com aquele cliente. A nova conversa deve começar do ponto em que a última conversa terminou. Sobretudo as interações iniciadas pelo cliente têm um potencial maior que as interações iniciadas pela empresa (PEPPERS e ROGERS, 2004). Quando o cliente entra em contato com a empresa é porque tem uma necessidade clara e está em busca de uma solução para ela. As interações pós-venda também constituem uma importante fonte de informações para o aprendizado em um processo de relacionamento. Day (2001) percebe a assistência ao cliente como uma rica e vasta fonte de informações para descobrir problemas em potencial e necessidades latentes para alimentar o processo de desenvolvimento da relação com o cliente.

**- Personalização**: A empresa precisa adaptar-se às necessidades individuais expressas pelo cliente. Para Peppers e Rogers (2004):

> Devemos criar um ciclo de personalização e retroalimentação para que cada vez mais possamos, entendendo o cliente, fornecer o que ele espera e na forma que ele espera. Devemos ainda adequar nossa mensagem e nosso diálogo à forma preferida pelo cliente. A personalização é algo bem simples quando se conhecem as necessidades e as preferências do cliente, mas exige muita flexibilidade da empresa e treinamento adequado das pessoas que têm contato com ele (PEPPERS e ROGERS, 2004, p. 52).

Os autores destacam que é importante, porém, que uma empresa com um grande número de clientes não busque a personalização de produtos e de atendimento a todos estes clientes, utilizando a estratégia *one-to-one* apenas para os melhores (PEPPERS e ROGERS, 2004). Uma boa forma de compreender os princípios do CRM é comparar a dimensão de competição desta forma e da forma tradicional de fazer negócios, por meio do marketing transacional.

Peppers e Rogers (2004) defendem que quando a empresa compete na dimensão horizontal, que é o que grande parte das empresas fazem, ela acaba ganhando mais participação de mercado, porém à custa de uma redução da margem unitária. Por outro lado, quando a empresa atua na dimensão vertical, passa a participar cada vez mais no cliente, que vê valor em continuar fazendo mais negócios com ela. Conseqüentemente, ela vende mais ao mesmo cliente, com margens melhores (PEPPERS e ROGERS, 2004, p. 32). A Figura 4 demonstra a diferença entre as estratégias do marketing de massa, ou marketing transacional (dimensão horizontal) e o marketing de relacionamento (dimensão vertical):


**Figura 4 -** Dimensões de competição nas diferentes estratégias de Marketing
**Fonte**: Vicente (2005)

A identificação das características dos clientes, de modo a avaliar suas necessidades e classificar seu potencial de saída, bem como seu nível de lucratividade, depende da qualidade da informação à disposição das pessoas na empresa. Neste contexto, a informação assume um papel estratégico, já que ela é o centro de todo o processo de relacionamento com o cliente (SWIFT, 2001).

No conceito estabelecido por Brown (2001), a evolução das iniciativas de CRM passa por três estágios distintos, que é a aquisição de clientes, a retenção de clientes e o atendimento estratégico ao cliente. O estágio em que a empresa opera no momento tem um impacto significativo na forma como trata os clientes.

No estágio I, a aquisição de clientes é o foco principal. A atenção da empresa é direcionada para a construção de uma base de clientes por meio do uso da tecnologia e treinamento específico para elevar a eficácia dos vendedores. No estágio II, a empresa foca a retenção de clientes conquistados durante o primeiro estágio, direcionando esforços para a maximização do relacionamento com o cliente. Neste estágio, ela busca segmentar clientes em grupos com características similares, de modo a servir cada grupo de clientes de maneira

diferente. No estágio III, a empresa atinge o estágio de atendimento estratégico ao cliente, quando percebe a necessidade de prever quem são os clientes mais valiosos e é neste estágio que, segundo Brown (2001), estes passam a depender da empresa ou produto.

Ainda segundo Brown (2001), para a empresa evoluir de um estágio a outro, ela necessita incorporar determinadas práticas que suportam a construção de um processo de atendimento estratégico ao cliente. A forma como estas práticas são desenvolvidas na empresa determina em que estágio de desenvolvimento do processo ela se encontra. A Figura 5 mostra os cinco pilares que representam a base para o desenvolvimento do processo:

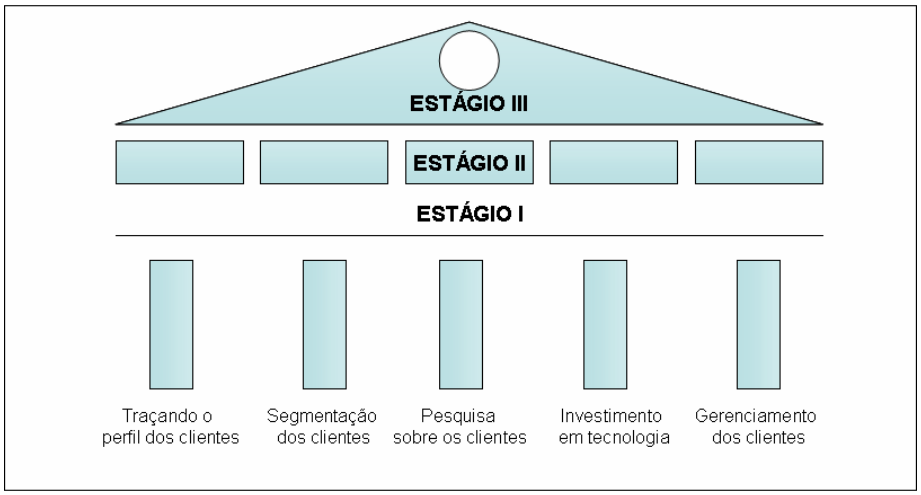


**Figura 5** - Os cinco pilares do atendimento estratégico ao cliente

**Fonte**: Brown (2001, p. 67) – Adaptado pelo autor

A Figura 5 aponta as diferentes práticas citadas por Brown (2001) como base de sustentação do atendimento estratégico aos clientes. Segundo o autor, estas práticas criam condições para a evolução do projeto de um estágio mais básico (estágio I) até um estágio mais avançado (estágio III).

O primeiro pilar, “traçando o perfil dos clientes”, parte da idéia de que nem todos os clientes têm o mesmo valor para a empresa e que alguns clientes são mais valiosos que outros. Aqui entra o aspecto destacado por Parvatiyar e Sheth (2001), que, na maioria das empresas, 20% dos clientes são responsáveis por mais de 80% do lucro da empresa. As empresas que estão no estágio III, traçaram o perfil dos clientes de tal forma que se concentraram naqueles que demonstraram ser mais promissores. Empresas que estão no estágio II ainda não aprenderam a dominar os sistemas de gerenciamento pelos quais as informações são coletadas. As empresas que estão no estágio I ainda se preocupam basicamente com a aquisição de clientes.

O segundo pilar, “segmentação dos clientes”, destina-se a identificar aqueles clientes que têm merecido uma atenção especial por parte da empresa. Segundo Brown (2001), empresas que estão no estágio II perceberam a importância de segmentar clientes em categorias distintas de atendimento, cujas categorias mais importantes são denominadas de “jóias da coroa”. Na visão do autor, clientes classificados nesta categoria são diferentes dos demais, pois contribuem com a maior parte do lucro da empresa.

O terceiro pilar, “pesquisas sobre clientes” refere-se àquele em que a empresa gera e administra o conhecimento sobre seus clientes. Brown (2001) afirma que empresas que atingem o estágio III fazem uso da tecnologia para descobrir, corresponder continuamente e até antecipar as necessidades de seus clientes. Conseqüentemente, o cliente não tem a necessidade de procurar o serviço ou produto em outro lugar e, desta forma, uma situação de benefício bilateral foi atingida.

De acordo com Brown (2001), o quarto pilar, “investimentos em tecnologia”, torna-se necessário para sustentar processos de atendimento estratégico ao cliente, como gravação de ligações ou automação de vendas. Empresas que pertencem ao estágio III fazem uso de processos de modelagem de dados e de *database marketing* para identificar os clientes mais propensos à aquisição de determinados produtos e serviços e que aumentam as fontes de informação da empresa, sistemas de informações executivas que selecionam dados e os apresentam de uma forma mais simplificada e robusta do que nas aplicações do estágio I, permitindo que a empresa tenha mais agilidade para reagir às mudanças e trocas tanto nos setores de mercado quanto no desempenho dos clientes (BROWN, 2001).

O quinto e último pilar, “gerenciamento dos clientes”, consiste em estabelecer planos de ação individualizados e específicos para cada segmento de clientes, baseados em informações reunidas sobre eles. Empresas no estágio III criam equipes para lidar com as diferentes necessidades de serviços dos clientes, assegurando que as pessoas certas atendam às necessidades dos clientes com a informação adequada e com o apoio correto.

### 2.2.5 Geração de conhecimento sobre o cliente a partir de tecnologia

Rowley (2002) considera que o conhecimento sobre o cliente abrange o conhecimento da empresa sobre potenciais clientes, segmentos de clientes e clientes individuais. As estratégias e ações de relacionamento com clientes mencionadas até então partem da premissa de que a empresa conhece, pelo menos em parte, os clientes com os quais está se relacionando. Muito deste conhecimento é proveniente dos dados que a empresa armazena sobre estes clientes, bem como seu comportamento na utilização de produtos e serviços oferecidos pela empresa. Com o crescimento da base de clientes, o uso de tecnologia, seja para estabelecer canais de relacionamento com clientes, seja para armazenar as informações relativas à interação entre empresa e cliente, torna-se importante para viabilizar o relacionamento com os mesmos (ROWLEY, 2002). Greenberg (2001) considera que a tecnologia é fundamental para a operacionalização da estratégia, permitindo acompanhar os clientes de forma individual.

De acordo com o *Meta Group*, a prática de CRM nas empresas pode ser dividida em três tipos de processos que envolvem o uso da tecnologia: CRM operacional, colaborativo e analítico (PEPPERS E ROGERS, 2004, p.69), conforme destaca a Figura 6:

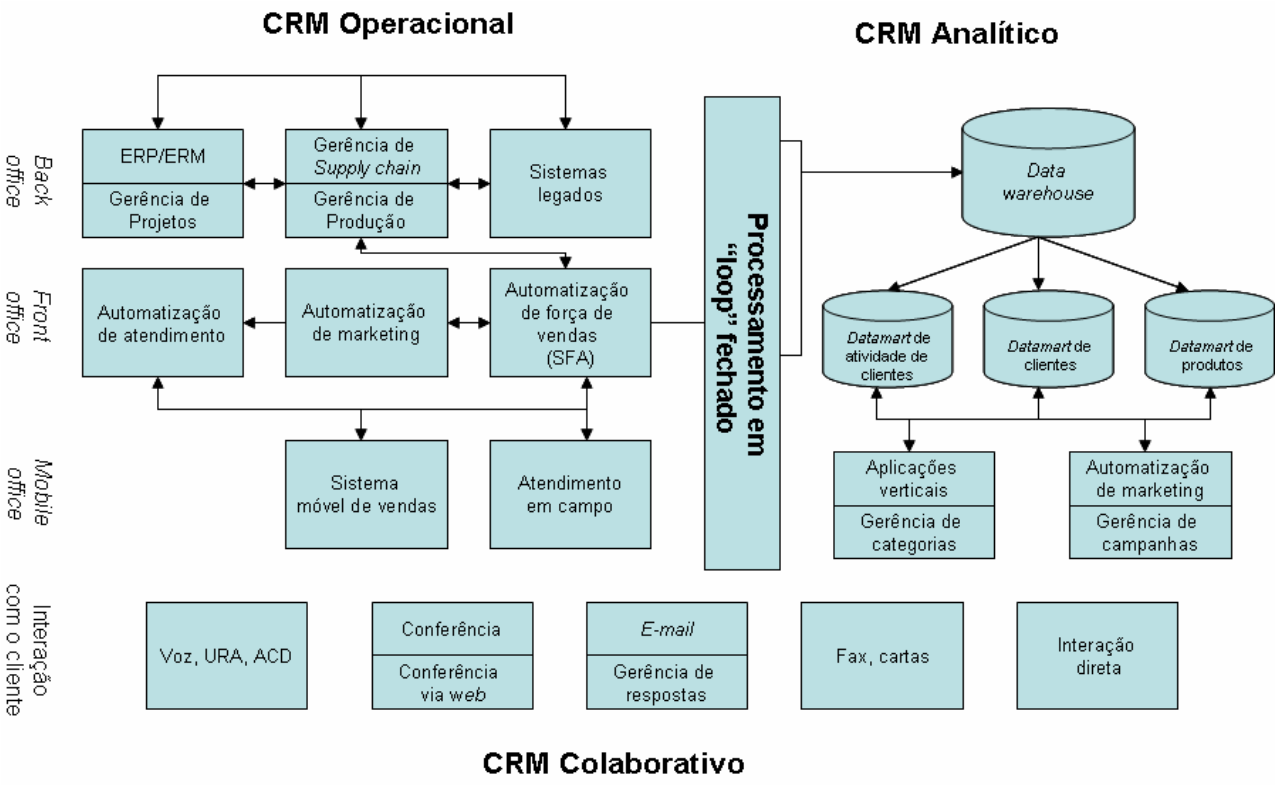


**Figura 6 -** Processos de tecnologia para aplicação de CRM

**Fonte:** Peppers e Rogers (2004, p. 69)

A Figura 6 apresenta os diversos processos relacionados ao CRM que podem ser desenvolvidas dentro da empresa por meio do uso da tecnologia. Eles são basicamente divididos em três partes:

- **CRM operacional**: Permite a integração do *back office*, *front office* e *mobile office*. Dados de clientes são coletados por meio de uma ampla gama de pontos de contato entre o cliente e a empresa, como *contact centers*, *e-mail*, fax, força de vendas, internet, etc. (XU e WALTON, 2005).
- **CRM colaborativo:** Envolve todos os pontos de contato com o cliente, em que ocorre algum tipo de interação entre ele e a empresa. Os vários canais de contato (voz, e-mail, fax, cartas, internet) precisam estar preparados não só para permitir esta interação, como também para garantir o fluxo adequado dos dados resultantes desta interação para o resto da empresa (XU e WALTON, 2005).
- **CRM analítico**: Conforme consideram Peppers e Rogers (2004), trata-se da fonte de inteligência do processo, pois serve para o ajuste e manutenção das estratégias de diferenciação de clientes, bem como o acompanhamento de seus hábitos e necessidades e os eventos que podem ocorrer na história do relacionamento entre eles e a empresa. Os dados armazenados nas bases de dados da empresa são analisados por meio de ferramentas analíticas, de modo a gerar perfis de clientes, identificar padrões de comportamento, determinar níveis de satisfação e dar suporte à segmentação de clientes. A informação e o conhecimento adquiridos a partir do CRM analítico auxiliam no desenvolvimento de estratégias adequadas de promoção e marketing. (PEPPERS e ROGERS, 2004).

O CRM analítico é classificado por Kotorov (2003) como uma visão de 360 graus do cliente. Tecnologias que reforçam e suportam o CRM analítico incluem *datawarehouses*, ferramentas analíticas e preditivas (ECKERSON e WATSON, 2001), regras de associação e descoberta de padrões, classificação e avaliação do valor do cliente (AHN, KIM e HAN, 2003). Como resultado das análises que o CRM analítico possibilita à empresa, clientes são segmentados de forma mais precisa e recebem ofertas de produtos e serviços que melhor ajustam-se às suas necessidades e perfis de consumo (XU e WALTON, 2005).

Segundo Carvalho (2003), é a partir do CRM analítico que se determina quais são os clientes que devem ser priorizados pela empresa. Os clientes são diferentes em seu valor para a empresa, de modo que o CRM analítico permite que sejam identificados e selecionados os clientes de maior valor e os clientes de maior potencial para a empresa, permitindo formas diferenciadas de relacionamento com eles (PARVATIYAR e SHETH, 2001).

## 2.3 Processos de descoberta de conhecimento em bases de dados

A receita gerada pelos clientes, a partir da utilização de produtos e serviços da empresa não é uniforme, significando que alguns clientes geram mais receita que outros. As empresas possuem grande interesse em avaliar os dados disponíveis sobre seus clientes, transformando-os em conhecimento, de modo a utilizá-lo no desenvolvimento de estratégias de relacionamento com eles, direcionando esforços para aqueles clientes mais rentáveis (XU e WALTON, 2005).

De acordo com Harrington (1993, p. 10), processo é qualquer atividade que recebe uma entrada, agrega-lhe valor e gera uma saída para um cliente interno ou externo, fazendo uso dos recursos da organização para gerar resultados concretos. A literatura descreve processos que auxiliam na análise de dados e na transformação de dados brutos em conhecimento, que mais tarde pode ser utilizado para diversos fins estratégicos da empresa, como, por exemplo, para o desenvolvimento de estratégias de relacionamento com clientes. Segundo Fayyad, Piatetsky-Shapiro e Smith (1996) estes processos são chamados de descoberta de conhecimento em bases de dados (do inglês KDD – *knowledge discovery in databases*) e o conceituam como o mecanismo de descoberta de conhecimento útil contido em bases de dados. Segundo Fayyad *et al.* (1996), este termo, “descoberta de conhecimento em bases de dados”, surgiu no primeiro workshop de KDD, realizado em 1989, para enfatizar que o produto final do processo de descoberta em bases de dados era o “conhecimento” relevante para tomada de decisão.

## 2.4 Metodologia de Fayyad, Piatetsky-Shapiro e Smith

Segundo Fayyad, Piatetsky-Shapiro e Smith (1996), o processo de identificação de padrões em bases de dados pode ser disciplinado, considerando que este processo inteiro seja segmentado em um conjunto de tarefas menores, com o intuito de, ao final, algum tipo de conhecimento útil ser obtido a partir de uma base de dados. Os autores argumentam que este processo é necessário, pois à medida que aumentam os volumes de dados das empresas, torna-se custoso e lento analisar manualmente os dados, abrindo a possibilidade de tornar objetivo e sistemático o processo de análise dos dados e identificação de padrões.

A Figura 7 mostra a proposta de Fayyad, Piatetsky-Shapiro e Smith (1996) para disciplinar este processo de descoberta de conhecimento em bases de dados:

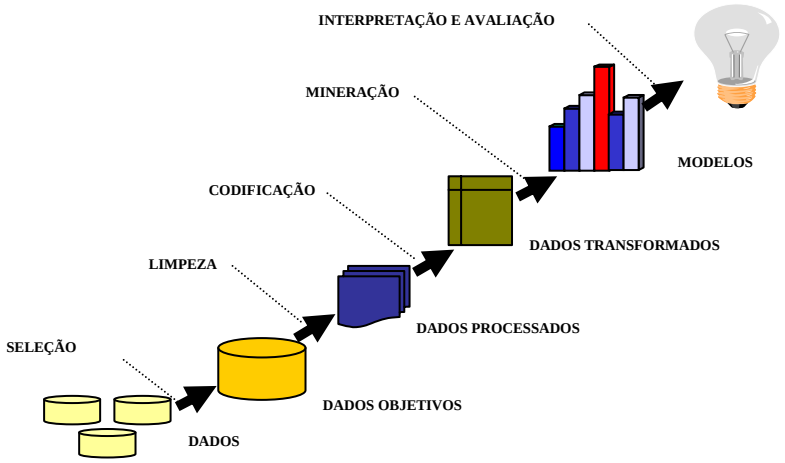


**Figura 7 -** Fases do processo de KDD
**Fonte:** Adaptado de Fayyad, Piatetsky-Shapiro e Smith (1996)

A definição proposta por Fayyad, Piatetsky-Shapiro e Smith (1996) é convergente com a de Frawley *et al*. (1992), que consideram este processo de descoberta de conhecimento como a extração, a partir de dados brutos, de informações implícitas, anteriormente desconhecidas e potencialmente úteis em determinado contexto. O processo de descoberta de conhecimento em bases de dados (KDD), descrito pela Figura 7, pressupõe seu início a partir de bases de dados que, na seqüência, são manipuladas, modificadas e interpretadas por meio de modelos matemáticos ou estatísticos que são utilizados para identificar padrões presentes nestas bases (FAYYAD *et al.*, 1996). Por fim, estes modelos geram conhecimento que pode ser útil para a resolução de algum problema da empresa.

Segundo Fayyad, Piatetsky-Shapiro e Smith (1996), a etapa do KDD referente à utilização de técnicas quantitativas para a criação dos modelos é chamada de mineração de dados (do inglês, *data mining*) e é definida como a aplicação de técnicas estatísticas e matemáticas para identificar e extrair padrões a partir dos dados. Berry e Linoff (2004) convergem para este conceito, denotando a mineração de dados como a exploração e análise de grandes volumes de dados com o intuito de descobrir regras e padrões significativos para o negócio da empresa.

O conceito de mineração de dados faz uma analogia ao processo de garimpagem, realizado pelas mineradoras. Neste contexto, os trabalhadores que exercem a atividade de garimpagem (também chamados de mineiros) escavam grandes volumes de terra em busca de

metais e pedras preciosas que possuam valor comercial. Trazendo este significado para o universo dos dados disponíveis na empresa, a mineração de dados busca encontrar conhecimento útil a partir de grandes volumes de dados, conhecimento que pode ter grande valor comercial para a empresa, quando utilizado, entre outros fins, para a formulação de estratégias de relacionamento com os clientes.

## 2.5 Metodologia de Brachman e Anand

Brachman e Anand (1996) criticam a maneira como são descritos os processos de descoberta de conhecimento em bases de dados, em função de acreditarem que a etapa de mineração de dados é apenas uma pequena parte do processo completo de descoberta de conhecimento, e que etapas anteriores à modelagem e posteriores a ela precisam ser consideradas como partes fundamentais deste processo de descoberta. Eles consideram, principalmente, que o ser humano, no papel de usuário, exerce um papel fundamental durante todo o processo e, baseado nisso, conceituam o processo de descoberta de conhecimento como um conjunto de tarefas intensivas em conhecimento, composta de interações complexas, desenvolvida ao longo de um determinado período de tempo, entre um ser humano e uma base de dados, possivelmente auxiliadas por um conjunto heterogêneo de ferramentas. Considerando as etapas do processo de descoberta de conhecimento, o fluxo proposto por Brachman e Anand (1996) é similar ao considerado por Fayyad, Piatetsky-Shapiro e Smith (1996), conforme mostra a Figura 8:

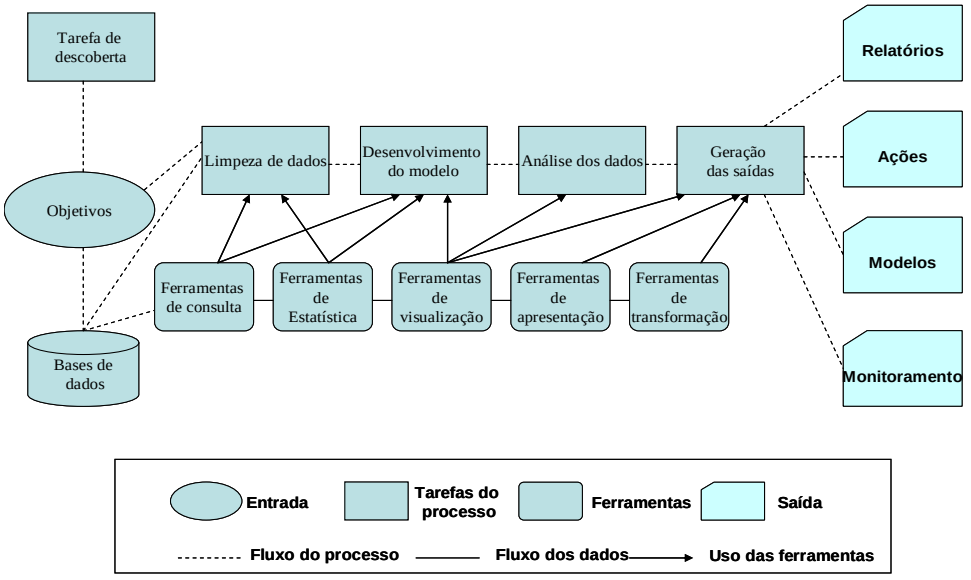


**Figura 8 –** Processo completo de KDD

**Fonte:** Adaptado de Fayyad *et al.* (1996)

Brachman e Anand (1996) consideram que a primeira etapa do processo de descoberta de conhecimento está relacionada aos propósitos da tarefa de descoberta e aos objetivos a serem atingidos. Nesta etapa, o papel humano é ainda mais essencial, visto que a realização de todas as etapas subseqüentes dependerá daquilo que se considera como problema a ser resolvido.

A partir da definição do problema, os autores propõem a realização de um conjunto de tarefas muito similares às propostas por Fayyad *et al.* (1996), relacionadas à construção de uma base de dados, limpeza destes dados e análise, que é baseada em um conjunto de técnicas estatísticas e de inteligência artificial (BRACHMAN e ANAND, 1996), que modelam padrões existentes na base de dados.

Esta modelagem, segundo Brachman e Anand (1996) inclui a especificação do modelo, escolha das técnicas mais adequadas aos dados existentes e o ajuste do modelo com base em uma ou mais técnicas escolhidas. A tarefa de análise de dados, proposta pelos autores, inclui desde a avaliação do modelo (baseando-se em critérios técnicos e de negócio) até sua apresentação para os membros da empresa ou para os que demandaram a solução de determinado problema. Na visão dos autores, todas estas etapas não podem ser realizadas de forma otimizada, sem a participação de pessoas que interajam com os dados e forneçam sentido às informações geradas.

Por fim, os autores consideram que a tarefa de geração de saídas pressupõe que alguma resposta ao problema inicial deve ser fornecida, tanto em termos de relatórios, modelos que devem ser aplicados à resolução do problema, ações baseadas no conhecimento adquirido por meio dos modelos construídos, bem como monitoramento do resultado das ações decorrentes do resultado dos modelos.

## 2.6 Metodologia CRISP-DM

Esta visão do processo de KDD centrada no usuário também é partilhada por outras metodologias. Shearer (2000) descreve a metodologia de livre distribuição denominada CRISP-DM (do inglês, *Cross Industry Standard Process for Data Mining*), desenvolvida em 1996 por um importante consórcio de empresas européias (RODRIGUEZ *et al.*, 2006), composto por Daimler-Benz (atual DaimlerChrysler), *Integral Solutions Ltd*. (ISL), NCR e OHRA, época em que as empresas, sobretudo do setor de indústria na Europa, não possuíam uma metodologia oficial para a condução de projetos de mineração de dados e que a necessidade de descoberta de conhecimento em bases de dados era crescente, em função do

aumento da concorrência e da quantidade de clientes. Segundo Silva (2002), CRISP-DM é uma metodologia validada, tendo abrangência, detalhamento de passos, busca de padronização, rotinas e etapas genéricas para o desenvolvimento de aplicações de KDD.

A metodologia CRISP-DM posiciona a mineração de dados dentro de um contexto de negócio mais amplo, de descoberta de conhecimento em bases de dados, estabelecendo seis etapas principais, conforme ilustra a Figura 9:

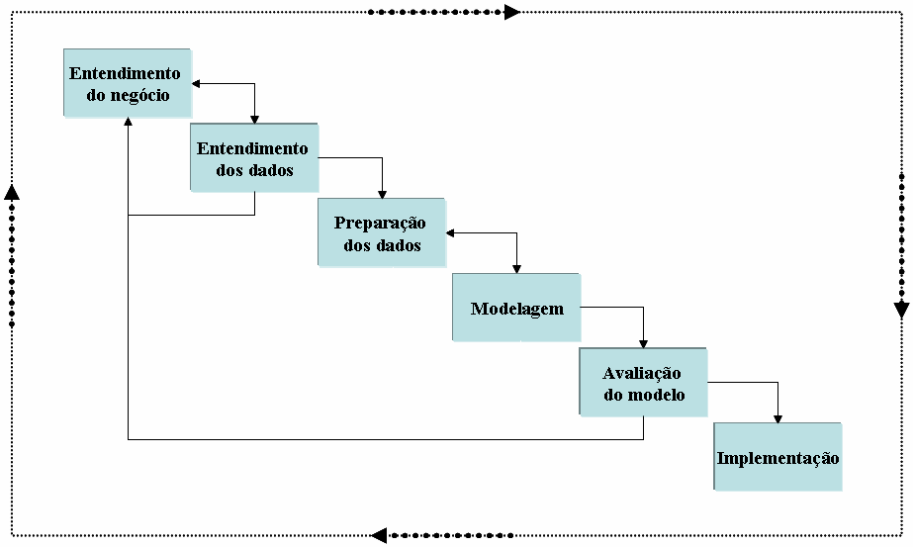


**Figura 9 -** Fases da metodologia CRISP-DM

**Fonte:** Adaptado de Chapman *et al.* (1999)

Com base na Figura 9, verifica-se que o conceito explicitado pela metodologia CRISP-DM converge com as definições de KDD destacadas por Fayyad, Piatetsky-Shapiro e Smith (1996) e Brachman e Anand (1996), que posicionam a mineração de dados como uma das etapas do processo de KDD. Considerando o processo como um todo, o modelo proposto Brachman e Anand (1996) assemelha-se ainda mais ao CRISP-DM, sobretudo no que se refere ao entendimento do problema de negócio e à fase de implementação, em que o conhecimento gerado por meio do processo de KDD é transformado em ações para a resolução do problema proposto no início do processo. Cabe apenas considerar que Chapman *et al.*(1999) propõem ainda o CRISP-DM como um processo cíclico, que pode ser repetido inúmeras vezes, até a resolução do problema de negócio.

Shearer (2000) considera que a etapa inicial de entendimento do problema de negócio é essencial, pois transforma o problema, descrito em uma linguagem de negócio, em um problema de mineração de dados, de forma a estabelecer um plano para resolver o problema de negócio da empresa. Cada uma das seis etapas apresentadas na Figura 9 possui tarefas que

precisam ser avaliadas antes da passagem para a fase seguinte. De acordo com Chapman *et al.* (1999), esta subdivisão ocorre da seguinte maneira, descrita por meio da Figura 10:

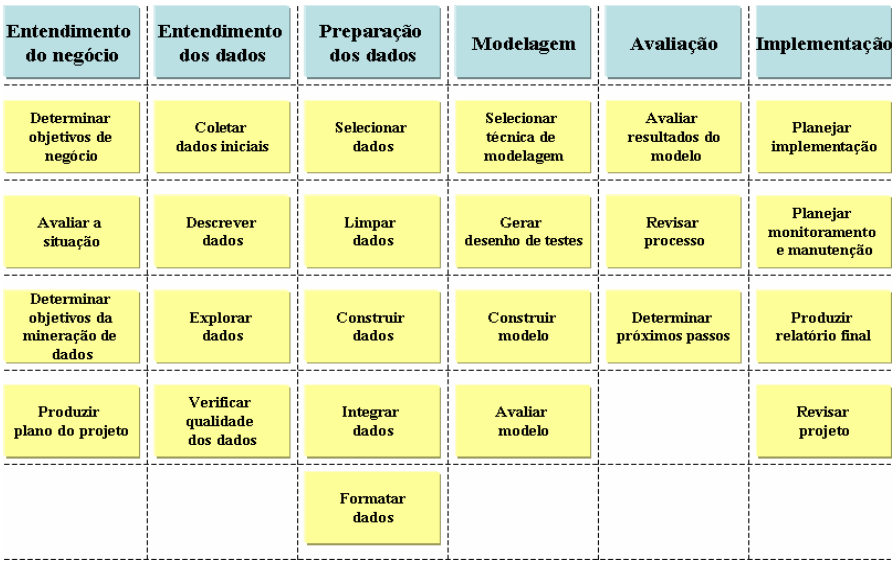


**Figura 10 -** Tarefas chave de cada uma das etapas da metodologia CRISP-DM

**Fonte**: Adaptado de Chapman *et al.* (1999)

A Figura 10 mostra que cada uma das seis etapas que compõem a metodologia CRISP-DM é subdividida em um determinado número de tarefas chave, que precisam ser realizadas, de modo que seja recomendável prosseguir para a etapa seguinte. Na seção 2.6.1, Shearer (2000) descreve a necessidade de cada uma das tarefas, associadas às etapas da metodologia CRISP-DM.

### 2.6.1 Etapa 1 - Entendimento do negócio

Berry e Linoff (2000) atribuem importância à etapa de entendimento do negócio e consideram que esta etapa envolve comunicação entre pessoas que entendem o negócio da empresa e pessoas com habilidades quantitativas para transformar dados em conhecimento útil para a tomada de decisão. Eles argumentam que o trabalho desenvolvido por meio da interação com pessoas envolvidas no negócio permite responder a questões do tipo: Esforços de mineração de dados são necessários para resolver o problema de negócio da empresa ? O quê a intuição e a experiência dizem que é importante ser considerado ? Com isto, Berry e Linoff (2004) defendem que o conhecimento técnico, bem como das tecnologias disponíveis são importantes, porém precisam ser combinados a uma visão clara do negócio e dos

objetivos da empresa, de modo a obter modelos ou ferramentas, cujos resultados sejam coerentes, aplicáveis, possam ser implementados e, portanto, utilizados para a tomada de decisão dentro da empresa.

De acordo com a metodologia CRISP-DM, a primeira fase de um processo de descoberta de conhecimento em bases de dados, o entendimento do negócio, possui quatro tarefas principais, que se referem à determinação dos objetivos de negócio, avaliação da situação, determinação dos objetivos da mineração de dados e produção do plano do projeto.

Shearer (2000) destaca que a compreensão dos reais objetivos do negócio da empresa é vital para descobrir fatores que devem ser envolvidos e considerados no projeto, de forma que o resultado apresente as respostas certas para as questões certas. A avaliação da situação ou do cenário permite determinar se os recursos disponíveis, desde pessoas, dados, até *software*, são suficientes para responder ao problema de negócio da empresa. Chapman *et al.* (1999) destacam que, nesta fase de entendimento do negócio, é essencial registrar a informação que é conhecida sobre a situação do negócio da empresa no início do projeto. Berson, Smith e Thearling (1999) consideram essencial esta etapa para a identificação do modelo de negócio da empresa e de como os resultados da mineração de dados serão implementados, de modo a resolver seu problema.

A definição dos objetivos da mineração de dados torna-se importante à medida que guiará a seleção das técnicas de mineração de dados a serem utilizadas na etapa de modelagem. Chapman *et al.* (1999) salientam que é importante ter claro os objetivos do negócio para transportá-los para uma linguagem técnica, de modo a representar os objetivos da mineração de dados. Esta linguagem técnica diz respeito, por exemplo, à capacidade mínima preditiva, como % de acerto na classificação de clientes, que se espera dos modelos a serem desenvolvidos.

Por fim, a produção do plano do projeto refere-se à elaboração de um planejamento para se atingir os objetivos da mineração de dados, bem como riscos potenciais, tempo para cada tarefa e uma avaliação inicial das técnicas e ferramentas necessárias para dar suporte ao projeto. Para a realização de cada uma das tarefas relacionadas ao entendimento do problema de negócio da empresa, é essencial a interação entre pessoas da área técnica e pessoas da área de negócio, reforçando que este processo de descoberta de conhecimento em bases de dados é centrado no ser humano, conforme defendem Brachman e Anand (1996).

### 2.6.2 Etapa 2 - Entendimento dos dados

A compreensão dos dados é importante para que se obtenha um maior conhecimento e familiaridade a respeito de sua natureza, da qualidade presente nestes dados, bem como a necessidade de dados adicionais. De acordo com Shearer (2000), a etapa de entendimento de dados está subdividida em coleta inicial, descrição dos dados, exploração e verificação da qualidade dos dados.

Shearer (2000) destaca que esta etapa é essencial para que se descubram visões ou particularidades iniciais sobre os dados, de maneira a formular hipóteses sobre possíveis padrões que possam ser encontrados. A coleta inicial de dados engloba a busca de conteúdo que possa ter relação com o problema de negócio e com os objetivos da mineração de dados, formulados na etapa de entendimento do problema de negócio. Nesta tarefa, podem ser encontrados problemas com dados provenientes de várias fontes e que possam ter um conteúdo desatualizado ou errôneo.

A descrição dos dados permite que se tenha uma dimensão do volume das informações, em termos de quantidade de registros e de campos presentes nas bases de dados, bem como responder se os dados disponíveis são suficientes para responder aos problemas de negócio e da mineração de dados. A exploração dos dados permite a identificação de padrões iniciais, que servirão como hipóteses a serem testadas na etapa de modelagem. A tarefa de verificação da qualidade dos dados permite identificar problemas no preenchimento dos campos, conteúdo errôneo e incoerência em seu conteúdo, fatores estes que, se não identificados, podem impactar a qualidade dos modelos a serem desenvolvidos na etapa de modelagem.

### 2.6.3 Preparação dos dados

A etapa de preparação dos dados engloba a construção da base de dados final a ser utilizada na etapa de modelagem. As tarefas associadas a esta etapa incluem a seleção, a limpeza, a construção, a integração e a formatação dos dados (CHAPMAN *et al.*, 1999).

A seleção dos dados é essencial, tanto em termos de tipos de dados a serem considerados, quanto da quantidade de registros, de modo que o conteúdo da base de dados final seja relevante e possibilite responder aos objetivos da mineração, definidos na etapa de entendimento dos dados. A limpeza dos dados é importante para garantir que o conteúdo da

base de dados final seja verdadeiro e possa ser utilizado para o desenvolvimento da modelagem. Shearer (2000) recomenda que se teste se a ausência de determinada informação pode ou não ter influência na etapa de modelagem, de modo que um possível impacto seja identificado e eliminado, antes que a etapa de modelagem tenha início. A tarefa de construção da base de dados envolve a criação de campos ou variáveis, cujo conteúdo representa a combinação de campos já existentes na base de dados, de modo a agregar valor na relação das variáveis importantes para atingir os objetivos da mineração de dados (SHEARER, 2000). A integração de dados está relacionada à junção de dados provenientes de múltiplas bases de dados, de modo a criar novos campos ou novos registros, ou ao resumo de campos existentes na base de dados (SHEARER, 2000). A tarefa de formatação refere-se a deixar a base de dados com um formato adequado para a aplicação das técnicas a serem utilizadas na etapa de modelagem (SHEARER, 2000).

### 2.6.4 Etapa 4 – Modelagem de dados

Esta é a etapa em que a base de dados construída nas etapas de entendimento e preparação dos dados será utilizada para o desenvolvimento de modelos que responderão aos objetivos da mineração de dados. Para tal, esta etapa é composta por tarefas como seleção da técnica a ser utilizada, geração de desenho de testes, construção e avaliação do modelo (CHAPMAN *et al.*, 1999).

A seleção da técnica, segundo Shearer (2000), será feita de acordo com os objetivos definidos na etapa 1 (entendimento do problema de negócio) para a mineração de dados e cujo conteúdo dos campos da base de dados foi definido e trabalhado nas etapas 2 e 3 (entendimento e preparação dos dados). A tarefa de geração do desenho de testes tem a finalidade de estabelecer quais critérios serão utilizados para definir se os modelos a serem desenvolvidos serão aceitos ou se será necessário mudar a técnica a ser utilizada ou voltar à etapa anterior e redefinir novos dados para a modelagem. Possíveis critérios incluem o nível de precisão das estimativas dos modelos, bem como a capacidade de generalização de seus resultados. Berson, Smith e Thearling (1999) citam a acurácia, ou percentagem total de casos cujas predições fornecidas pelos modelos foram corretas, como uma das medidas de avaliação dos modelos.

Shearer (2000) recomenda a separação da base de dados em base de treinamento e base de testes, separação esta a ser realizada antes do desenvolvimento dos modelos. Após

definir os critérios que serão utilizados para avaliar um ou mais modelos desenvolvidos, surge a tarefa de construção, em que a técnica definida no início desta etapa será utilizada para o desenvolvimento de um ou mais modelos. Por fim, a tarefa de avaliação dos modelos existe para julgar o sucesso dos modelos desenvolvidos. Shearer (2000) sugere que, tanto nesta tarefa, quanto na tarefa de construção dos modelos, ocorra a interação entre pessoas que desenvolvem os modelos e pessoas com conhecimento do negócio, de modo que potenciais problemas com os dados ou com resultados incoerentes gerados pelos modelos possam ser identificados. Aqui novamente o autor destaca a importância da participação do ser humano na identificação de resultados incoerentes gerados pelo modelo, reforçando o modelo proposto por Brachman e Anand (1996).

Berry e Linoff (2000) consideram que o processo de modelagem tem três suposições chave: (1) o passado é um bom preditor do futuro, ou seja, para desenvolver modelos, o comportamento passado do cliente é utilizado para prever o comportamento futuro, o que nem sempre é verdadeiro, visto que fatores externos não observados no passado podem influenciar o comportamento futuro do cliente; (2) Os dados estão disponíveis, ou seja, todos os dados necessários para a análise do problema de negócio da empresa estão disponíveis. Conforme defendem os autores, muitas vezes, restrições são impostas, como dados importantes, mas que não foram registrados, dados residentes em outros departamentos e dados no formato errado, que limitam a capacidade de desenvolvimento dos modelos, ainda que se tenha compreendido o problema de negócio da empresa; (3) Aquilo que se deseja como *output* dos modelos desenvolvidos foi contemplado no levantamento dos dados. Conforme colocam Berry e Linoff (2000), muitas empresas geram expectativas sobre o resultado da modelagem de dados, sem contudo considerá-las no levantamento dos dados necessários para o desenvolvimento dos modelos.

### 2.6.5 Etapa 5 – Avaliação dos modelos

Esta é a etapa em que se avaliará se os modelos desenvolvidos atendem ou não às expectativas do negócio e respondem aos objetivos da mineração de dados, definidos na etapa inicial de entendimento do problema de negócio. Shearer (2000) adverte que é crítico verificar se nenhum aspecto relevante do negócio foi omitido ou não considerado suficientemente durante as etapas de desenvolvimento da base de dados e modelagem. Considerando que o modelo tenha sido julgado adequado, será decidido como os resultados dos modelos serão

utilizados para a elaboração de estratégias de negócio. As tarefas chave desta etapa referem-se à avaliação dos resultados, à revisão do processo e à determinação dos próximos passos.

Segundo Chapman *et al.* (1999), a avaliação do modelo realizada nesta etapa difere da avaliação técnica do modelo, realizada na etapa de modelagem. A finalidade desta tarefa é verificar o quanto os modelos desenvolvidos atendem aos objetivos do negócio. A tarefa de revisão do processo pretende avaliar se algum fator foi negligenciado, se os modelos foram corretamente construídos e se o conteúdo e o formato do modelo são passíveis de implementação. A tarefa de próximos passos decide se os modelos construídos podem ser implementados ou se alguma alteração em sua composição, bem como no problema de negócio se fazem necessários.

### 2.6.6 Etapa 6 – Implementação dos modelos

Shearer (2000) argumenta que o conhecimento gerado pelos modelos precisa ser disponibilizado para a empresa e seu conteúdo precisa ser organizado e apresentado de tal forma que as pessoas possam utilizá-lo. As tarefas chave associadas à etapa de implementação são o plano de implementação, o plano de monitoração e manutenção, a produção do relatório final e a revisão do projeto.

O plano de implementação refere-se à estratégia utilizada pela empresa para implementar os modelos e disponibilizar seus resultados para tomada de decisão. O plano de monitoração e manutenção prevê os critérios de acompanhamento do funcionamento dos modelos, de modo que problemas gerados pelo resultado fornecido por eles sejam identificados rapidamente e não impactem o negócio da empresa. A tarefa referente ao relatório final reflete a realização que envolve desde a construção da documentação oficial do projeto, com todos os resultados obtidos nas etapas anteriores até a apresentação final à área da empresa que demandou os esforços de mineração de dados. Por fim, a revisão do projeto é o momento em que se reflete a respeito dos sucessos e falhas de todas as etapas, bem como potenciais fontes de melhoria para uso em projetos futuros (SHEARER, 2000).

Um documento pode ser criado para registrar as falhas, dificuldades, acertos, dicas, critérios de seleção de técnicas de mineração de dados, bem como todos os dados considerados pertinentes para análise. Shearer (2000) coloca que em projetos ideais, a documentação da experiência sobre o projeto também cobre todos os relatórios escritos por membros do projeto durante todas as fases e tarefas envolvidas. Esta etapa é fundamental, por

buscar a explicitação do conhecimento gerado durante todo o processo de descoberta de conhecimento, transferindo-o para um nível organizacional (NONAKA e TAKEUCHI, 1997).

Shearer (2000) destaca a importância do CRISP-DM para a realização de projetos de gestão de relacionamento com clientes (CRM). Segundo ele, a DaimlerChrysler adaptou CRISP-DM para desenvolver sua própria ferramenta especializada de gestão de relacionamento, de modo a aperfeiçoar o marketing direcionado ao cliente.

## 2.7 A mineração de dados

A mineração de dados, como uma das etapas do processo de descoberta de conhecimento (FAYYAD, PIATESTKY-SHAPIRO e SMITH, 1996) e presente na metodologia CRISP-DM (CHAPMAN *et al.*, 1999) na fase de modelagem de dados, focaliza principalmente a aplicação de técnicas quantitativas para a identificação de padrões que configurem conhecimento útil sobre os clientes, de forma que possa ser utilizado para a elaboração de estratégias de aquisição, desenvolvimento e retenção de clientes. Como já especificado na seção 2.4, Berry e Linoff (2004) conceituam a mineração de dados como um processo de exploração e análise de grandes quantidades de dados, de modo a descobrir regras e padrões significativos para o negócio da empresa. Outras definições encontradas na literatura convergem para a definição de Berry e Linoff (2004), conforme mostra o Quadro 2:

**Quadro 2**: Definições de mineração de dados na literatura

| Autor | Definições propostas para mineração de dados |
|---|---|
| Fayyad *et al.* (1996) | Mineração de dados é uma etapa da descoberta de conhecimento em bases de dados (KDD) e refere-se a técnicas que são aplicadas para extrair padrões dos dados. |
| Cabena *et al.* (1998) | Mineração de dados é definida como o processo de extração, a partir de bases de dados, de informação anteriormente desconhecida, com conteúdo válido, para então utilizá-la para a tomada de decisões de negócio. |
| Hui e Jha (2000) | Mineração de dados é o processo de descoberta de conhecimento em grandes volumes de dados, que pode ser usado para ajudar empresas a tomar melhores decisões e permanecerem competitivas no mercado em que atuam. |

**Fonte:** Adaptado de Hormazi e Giles (2004)

O Quadro 2 mostra três definições de mineração de dados que ressaltam sua importância para o processo de descoberta de conhecimento, conhecimento este que a

empresa busca para desenvolver suas estratégias de negócio e tomada de decisão. Samli, Pohlen e Bozovic (2002) destacam que a mineração de dados difere de técnicas estatísticas tradicionais, pois exigem uma menor intervenção humana. A mineração de dados também difere de ferramentas de processamento analítico *online*, ou OLAP (*online analytical processing*), pois tem maior autonomia na descoberta de padrões nos dados, enquanto ferramentas OLAP confiam no analista para seguir uma abordagem tradicional de testes de hipóteses por meio da geração de padrões ou relacionamentos hipotéticos e utilização de consultas para confirmá-las ou refutá-las (EDELSTEIN, 1999). Isto, porém, não significa que a mineração de dados seja uma atividade totalmente autônoma, pois conforme defendem Brachman e Anand (1996) e Chapman *et al.* (1999), o ser humano exerce um papel fundamental, no que se refere a atribuir significado aos resultados fornecidos pelos modelos gerados.

A maioria das técnicas de mineração de dados, pelo menos como algoritmos acadêmicos, existe há anos ou décadas. Entretanto, segundo Berry e Linoff (2004), somente a partir da década de 90, aplicações comerciais envolvendo mineração de dados passaram a ser utilizadas de maneira intensiva nas empresas. Segundo os autores, isto é fruto da convergência de diversos fatores:

- **Produção de dados**: A mineração de dados faz mais sentido quando há grandes volumes de dados, já que a maioria dos algoritmos de mineração de dados exige grandes volumes para construir e treinar modelos que serão utilizados para realizar tarefas de classificação, predição ou estimação. De acordo com Berry e Linoff (2004), segmentos do setor de serviços como telecomunicações e cartões de crédito há algum tempo têm mantido um relacionamento interativo com clientes e gerado registros de transação. Contudo, recentemente a maioria dos setores têm armazenado, em um ritmo cada vez mais acelerado, dados a respeito da interação com clientes, possibilitando o uso de bases de dados para a avaliação do comportamento dos clientes. Han e Kamber (2000) percebem algo similar quando apontam que a crescente atenção que a mineração de dados tem recebido nos últimos anos por parte da indústria de informação é resposta à ampla disponibilidade de grandes volumes de dados e à iminente necessidade de se transformar dados em informação e em conhecimento, necessidade esta, em parte, motivada pelo aumento da concorrência entre as empresas, aumento de oferta de produtos e serviços, maior exigência por parte dos clientes, o que faz com que as empresas tenham a necessidade de conhecer melhor seus clientes para

poderem oferecer produtos e serviços mais adequados às suas exigências e necessidades e consigam, a partir disso, manter os melhores clientes (PEPPERS e ROGERS, 2004).

- **Armazenamento de dados**: Berry e Linoff (2004) declaram que não só a quantidade de dados tem aumentado, mas como também a variedade de dados armazenada. Dados provenientes de todo tipo de interação com o cliente têm sido armazenados em *datawarehouses* e se tornado parte da memória da empresa.

- **Recursos computacionais acessíveis:** A queda no preço de discos rígidos, memória e processadores tem possibilitado o uso de técnicas de mineração de dados que antes só poderiam ser realizadas por computadores com grande poder de processamento. Berry e Linoff (2004) comentam que a introdução de *softwares* de gerenciamento de bases de dados relacionais paralelos como Oracle, Teradata e IBM têm disponibilizado, pela primeira vez, o poder do processamento paralelo para as empresas. Estas plataformas com servidores de bases de dados com processamento paralelo fornecem um excelente ambiente para a mineração de dados em larga escala. Além disso, muitos programas computacionais diretamente associados à modelagem de dados têm sido desenvolvidos e amplamente utilizados no mercado. Programas como o *Enterprise Miner* (SAS *Institute Inc.*), *Clementine* (SPSS *Inc*.), *Intelligent Miner* (IBM *Corporation*), KXEN (KXEN *Inc.*) e *Statistica Data Miner* (*Statsoft Inc.*) têm sido largamente utilizados para a aplicação de técnicas de mineração de dados (BERSON, SMITH e THEARLING, 1999 ; PEACOCK, 1998 ; CHYE e GERRY, 2002; BERRY e LINOFF, 2004).

- **Forte interesse na gestão do relacionamento com clientes (CRM**) : Em vários setores, as empresas têm percebido que seus clientes são essenciais para o sucesso de seu negócio e que informação sobre cliente é um dos ativos chave para este sucesso (DAVENPORT, 2001). Berry e Linoff (2004) destacam que o conhecimento sobre clientes confere vantagem competitiva às empresas. Alavi e Leidner (2001) são mais incisivas e destacam que mais do que a posse de conhecimento, o que forma a base para o atingimento de uma vantagem competitiva é a habilidade da empresa em efetivamente aplicar o conhecimento existente para criar novos conhecimentos e tomar decisão com base nisso. Conforme destacam Chapman *et al.* (1999), a metodologia CRISP-DM se propõe a estruturar e integrar as etapas associadas à mineração de dados, como a identificação e entendimento dos problemas de negócio, o entendimento e a preparação dos dados, bem como a disponibilização do conhecimento obtido, do nível individual ou grupal (analista de mineração de dados ou equipes de mineração de dados) para o nível organizacional, de forma

a subsidiar a criação e aplicação de estratégias de relacionamento da empresa com seus clientes.

A amplitude de aplicações da mineração de dados transcende os limites da organização e dos clientes, já que em qualquer área do conhecimento humano, a mineração de dados terá espaço quando existirem dados reunidos acerca de um determinado fenômeno e o processo de descoberta de conhecimento a partir deles se fizer necessário (BOZDOGAN, 2004). Bozdogan (2004) também apresenta a mineração de dados em confluência com outras áreas como a Estatística, Ciência da Computação, Aprendizado de Máquina, Inteligência Artificial e Reconhecimento de Padrões.

Considerando o contexto desta dissertação, que envolve o processo de relacionamento com clientes, o objetivo da mineração de dados é permitir à empresa a melhoria das áreas de vendas, marketing e operações de suporte a clientes, por meio de um melhor entendimento de seus clientes (BERRY e LINOFF, 2004). A mineração de dados tem grande preocupação com a construção de modelos que revelem conhecimentos sobre o cliente. Segundo Berry e Linoff (2004), um modelo representa um algoritmo ou conjunto de regras que conecta uma coleção de *inputs* (na forma de campos ou variáveis de uma base de dados da empresa) a um particular alvo ou resultado, que, em uma linguagem estatística, pode ser chamada também de variável dependente.

### 2.7.1 Mineração de texto e mineração na internet

Além da mineração de dados tradicional, que utiliza técnicas quantitativas a partir de bases de dados relacionais (conforme defende a metodologia CRISP-DM), há duas outras vertentes da mineração de dados com crescente aplicação na área de marketing e gestão de relacionamento com clientes, que fazem uso de muitos dos métodos da mineração tradicional, mas têm como ponto de partida dados com nível menor de estruturação: São a mineração de texto (do inglês, *text mining*) e a mineração na internet (do inglês, *web mining*).

Segundo Witten e Frank (2005), a mineração de texto parte de dados não-estruturados e mais difíceis se serem compilados do que os dados armazenados em bases relacionais. Dados não-estruturados referem-se a dados textuais, como documentos corporativos, conteúdos de e-mail e reclamações de clientes. Os autores colocam que apesar de tanto a mineração tradicional quanto a mineração de texto avaliarem grandes volumes de dados, a mineração de dados tradicional tem uma preocupação em descobrir padrões existentes nos

dados, enquanto a mineração de texto tem o propósito de extrair a informação, claramente contida e explicitada no texto e, que, por razões de limitação de tempo, é inviável de ser realizado por pessoas.

Com a evolução dos computadores e do advento e expansão da internet, muitas empresas passaram a relacionar-se com seus clientes também de maneira virtual, de forma que muitos dos dados sobre eles ficam armazenados nas páginas quando pesquisam um determinado produto ou serviço ou mesmo quando realizam uma transação virtual (CHAKRABARTI, 2003). As empresas têm utilizado a mineração de dados na internet não somente para analisar a estrutura de suas páginas, mas principalmente as características das pessoas que as visitam, de forma a descobrir interesses e poder oferecer produtos e serviços adequados às necessidades de seus clientes. Chakrabarti (2003) destaca que a mineração de dados na internet analisa documentos contidos nas páginas, escritos de tal forma que seja possível citá-los mutuamente por meio de *hyperlinks*, já que eles são uma espécie de hipertexto. Lau *et al.* (2004) definem mineração de dados na internet como um processo de recuperação e conversão de informação de texto contido em páginas em uma base de dados organizada contendo variáveis chave de interesse para melhor entender clientes. Eles ainda comentam que a mineração de dados na internet utiliza de forma intensiva técnicas de mineração de texto para extrair conhecimento de documentos não-estruturados ou semi-estruturados contidos nas páginas. Segundo os autores, em 2004, aproximadamente 10% dos dados armazenados eletronicamente nas páginas da internet eram estruturados, enquanto 90% eram não-estruturados, com tendência à manutenção deste ritmo.

### 2.7.2 Categorias de mineração de dados

As técnicas de mineração de dados, com aplicação mais direcionada ao marketing e às estratégias de relacionamento com clientes, podem ser agrupadas em seis grandes categorias de análise (DAVIDSON E SOUKUP, 2002; BERRY E LINOFF, 2004), conforme mostra a Figura 11:

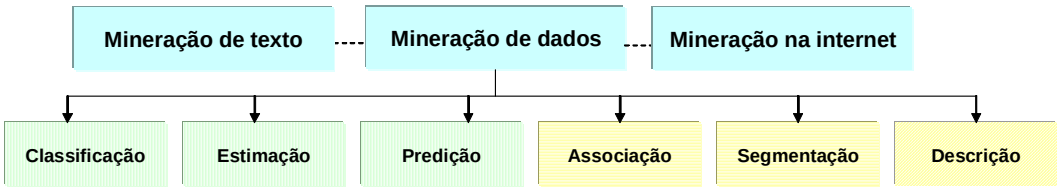


**Figura 11 -** Categorias de mineração de dados

**Fonte**: Elaborada pelo autor

Conforme destaca a Figura 11, as três primeiras categorias (classificação, estimação e predição, destacadas em linhas verticais) são referidas por Berry e Linoff (2004) como mineração de dados direcionada, pois seu objetivo é obter um valor para uma determinada variável alvo. As duas categorias seguintes (associação e agrupamento, destacadas em linhas horizontais) são classificadas como mineração de dados não direcionada, pois o objetivo é igualmente descobrir padrões nos dados, porém sem se remeter a uma determinada variável alvo. A última categoria assinalada pelos autores, a descrição (destacada em linhas diagonais), pode ser tanto direcionada, quanto não direcionada. O termo "direcionada" é também encontrado na literatura como "supervisionada". Pyle (2003) explica que o termo "supervisionada" é uma terminologia americana, enquanto que o termo "direcionada" é uma terminologia européia para referir-se ao modo como a mineração de dados é conduzida. Quando a mineração de dados é "direcionada" ou "supervisionada", pressupõe-se a realização da modelagem do comportamento de uma determinada variável resposta, enquanto para a mineração de dados "não direcionada" ou "não supervisionada", não se pressupõe a existência de tal variável resposta (PYLE, 2003). Cada uma das categorias de análise podem ser assim descritas:

- **Classificação:** Consiste em examinar as características de um objeto e atribuí-lo a um conjunto pré-definido de classes. De acordo com Berry e Linoff (2004), objetos a serem classificados normalmente são representados por registros em uma base de dados. O processo de classificação se caracteriza por uma especificação das classes e um conjunto de registros já classificados. O propósito deste tipo de análise é construir um modelo que possa ser aplicado a dados não classificados, de modo a classificá-los. Por exemplo, para a aplicação de uma estratégia de *cross-selling* com o objetivo de elevar o valor que cada cliente traz para a empresa, pode ser necessário o conhecimento sobre quem são os clientes com maior potencial a adquirir um determinado produto ou serviço. A empresa pode utilizar este conhecimento criado a partir da aplicação de técnicas de mineração de dados para direcionar as campanhas que suportem suas estratégia de vendas (BERRY e LINOFF, 2004).

Davidson e Soukup (2002) comentam que a utilização de atividades de classificação de clientes em empresas do setor de serviços é intensa, destacando que empresas de serviços financeiros freqüentemente categorizam pedidos de empréstimo como aceitável ou não aceitável (em função de uma série de critérios como risco e rentabilidade), empresas de seguros decidem se um cliente têm ou não propensão a adquirir um determinado tipo de

seguro e empresas de internet determinam que tipo de propaganda é mais provável de obter uma resposta de um determinado tipo de cliente.

As tarefas de classificação de clientes, baseadas em dados a seu respeito, tornam-se possíveis com a aplicação de técnicas de mineração de dados que exploram as bases de dados de clientes e procuram descobrir padrões e similaridades no comportamento dos clientes, de forma a classificá-los em determinadas categorias de interesse da empresa.

- **Estimação:** o objetivo das técnicas alocadas a esta categoria é atribuir um valor a cada cliente, valor este que é contínuo. Davidson e Soukup (2002) destacam a semelhança entre técnicas de estimação e técnicas de classificação, diferença que recai basicamente ao formato da variável alvo ou variável dependente, que no caso da classificação é discreta (formada por um conjunto limitado de categorias) e no caso da estimação, tende a ser uma variável de natureza contínua (com uma quantidade ilimitada de possíveis valores) . Uma estratégia de aquisição de clientes pode necessitar do conhecimento de quem são os *prospects* (ou potenciais clientes) com maior propensão de gerarem receita no primeiro ano de relacionamento com a empresa. Este conhecimento pode orientar a empresa na escolha dos clientes potencialmente mais rentáveis e alavancar a receita da empresa.

- **Predição:** Há uma grande semelhança entre as três categorias de análise, classificação, estimação e predição. Todas utilizam um conjunto de variáveis ou *inputs* (na linguagem estatística, variáveis independentes) para gerar o resultado de um *output* (na linguagem estatística, uma variável dependente). Conforme destacam Berry e Linoff (2004), a diferença é que na predição, o objetivo é estimar o comportamento futuro de uma determinada variável. Muitas aplicações envolvendo relacionamento com clientes utilizam técnicas de predição para estimar o valor vitalício do cliente (do inglês, LTV - *lifetime value*), de modo a gerar conhecimento sobre quais clientes gerarão receita para a empresa nos próximos períodos de tempo. Este conhecimento subsidia as estratégias de retenção de clientes, focalizando esforços nos clientes com maior potencial de receita, baseado em seu histórico comportamental de utilização de produtos e serviços oferecidos pela empresa.

- **Associação:** Diferentemente das três categorias anteriores, o propósito das técnicas de associação não é estimar nem gerar qualquer *output* específico, mas sim estabelecer o grau de correlação ou afinidade entre duas ocorrências. Conforme mencionam Davidson e Soukup (2002), a associação está preocupada em identificar quais entidades ou itens são mais prováveis de coexistirem na mesma situação. As técnicas de associação têm o potencial de gerar regras a partir de dados e subsidiar estratégias de relacionamento com clientes quando

geram o conhecimento de qual a tendência de um cliente adquirir um produto ou serviço, dado que possuem outro.

- **Segmentação:** Envolve o processo de dividir uma população heterogênea em um número de subgrupos ou *clusters* mais homogêneos ou semelhantes entre si. A segmentação pode ser prelúdio de outras formas de mineração de dados (DAVIDSON e SOUKUP, 2002; BERRY e LINOFF, 2004) por fornecer subsídios que permitam a utilização de técnicas pertencentes a outras categorias de mineração de dados. A base de dados de clientes pode ser segmentada com base em seu perfil demográfico e os *clusters* ou subgrupos encontrados podem ser utilizados para o desenvolvimento de modelos de propensão ao cancelamento, específicos para cada *cluster*, que subsidiem o desenvolvimento de estratégias segmentadas de retenção de clientes.

- **Descrição**: Conforme apresentado na Figura 11, pode ser utilizada tanto como técnica de mineração direcionada, quanto não direcionada. O objetivo do negócio, bem como do estudo em questão, indicará o tipo de técnica a ser utilizada. Berry e Linoff (2004) citam que determinadas técnicas podem ser utilizadas para obter o perfil de clientes, com relação a uma determinada variável alvo (ou variável dependente), assim como outras podem ser utilizadas para a obtenção de perfis descritivos não diretamente relacionados a uma variável específica. Peacock (1998) considera úteis as técnicas de descrição na tarefa de compreender a estrutura dos dados, para identificar problemas de conteúdo dos dados e interpretações erradas que possam surgir em decorrência da análise.A Figura 12 mostra as principais técnicas de mineração de dados observadas na literatura, aplicadas ao relacionamento com clientes:


**Figura 12 -** Técnicas de mineração de dados identificadas na literatura

**Fonte**: Elaborada pelo autor

A Figura 12 apresenta as técnicas identificadas na literatura sobre mineração de dados, sobretudo aquelas com aplicação específica na área de marketing e gestão de relacionamento com clientes. Algumas técnicas como árvores de decisão e redes neurais estão presentes em mais de uma categoria de mineração, em função de sua amplitude de uso a partir dos dados. A seção 2.7.3 descreve as técnicas de mineração de dados identificadas na literatura para o subsídio à criação de estratégias de relacionamento com clientes.

### 2.7.3 Técnicas de mineração de dados identificadas na literatura

#### 2.7.3.1 Árvores de Decisão

Conforme destacam Berry e Linoff (2004), uma árvore de decisão diz respeito a uma estrutura que pode ser utilizada para dividir uma grande coleção de registros em conjuntos sucessivamente menores de registros, por meio da aplicação de regras simples de decisão. A partir destas sucessivas divisões, os membros dos conjuntos resultantes tornam-se cada vez mais similares entre si. Um modelo de árvore de decisão representa um conjunto de regras para dividir uma grande população heterogênea em grupos menores e mais homogêneos com respeito a uma determinada variável alvo (ou variável dependente). Hand, Mannila e Smyth (2001) destacam que o trabalho de Breiman *et al.* (1984) como altamente influente nesta área, pois descrevem o algoritmo CART (*Classification and Regression Trees*) na adoção deste tipo de técnica para classificação em mineração de dados.

Segundo Gehrke (2003), o objetivo da classificação é construir um modelo da distribuição da variável alvo (ou variável dependente) em função das variáveis preditoras (ou variáveis independentes), de modo a utilizar o modelo resultante para atribuir valores a uma base de dados na qual os valores das variáveis preditoras são conhecidos, mas os valores da variável dependente não são. Breiman *et al.* (1984) defendem que, por inúmeras razões, árvores de decisão são especialmente atrativas em ambientes de mineração de dados: (1) Sua representação visual e intuitiva facilita o entendimento dos resultados do modelo ; (2) São especialmente adequadas para descoberta exploratória de conhecimento, sem a necessidade de atender a pressupostos na distribuição dos dados e (3) podem ser construídas relativamente rápido, se comparadas a outras técnicas de mineração (LIM, LOH & SHIH, 2000).

#### 2.7.3.2 Análise Discriminante

Constitui outra importante técnica que pode ser utilizada para propósitos de classificação. Hand, Mannila e Smyth (2001) comentam que o princípio por trás desta técnica está baseado no conceito da busca de uma combinação linear das variáveis independentes que melhor separa as classes da variável alvo. O termo “discriminante” reflete este princípio, que é discriminar ou separar as classes, em função das diferenças que apresentem em termos de variáveis independentes (HAND, MANNILA e SMYTH, 2001).

Malhotra (2001) conceitua a análise discriminante como uma técnica de análise de dados em que a variável alvo possui categorias ou classes e as variáveis independentes possuem natureza intervalar ou contínua. Destaca ainda que os objetivos desta técnica circundam em torno de cinco pontos principais: (1) Conforme destacam Hand, Mannila e Smyth (2001), esta técnica busca obter combinações lineares das variáveis independentes, que melhor discriminem as categorias da variável alvo; (2) Verificar se existem diferenças estatisticamente significativas entre os grupos, em termos das variáveis independentes. (3) Determinar as variáveis independentes que mais contribuem para as diferenças entre os grupos. (4) Classificar os registros em um dos grupos, com base nos valores das variáveis independentes. (5) Avaliar a precisão da classificação.

#### 2.7.3.3 Redes Neurais

Berson, Smith e Thearling (1999) e Berry e Linoff (2004) atribuem a esta técnica o nome de “redes neurais artificiais” por fazerem uso de neurônios artificiais para a identificação de padrões nos dados. Redes neurais verdadeiras, segundo os autores, são sistemas biológicos (cérebro) que detectam padrões, fazem predições e aprendem. As redes neurais artificiais são programas computacionais que implementam em um computador sofisticados algoritmos de detecção de padrões e aprendizado de máquina, de modo a desenvolver modelos preditivos a partir de bases de dados históricas.

De acordo com Berson, Smith e Thearling (1999), redes neurais artificiais derivam seu nome de seu desenvolvimento histórico, que se iniciou com a premissa de que máquinas deveriam ser construídas para pensar e que a ciência desejava emular no computador a estrutura e funcionamento do cérebro humano.

Berry e Linoff (2004) e Johnson e Wichern (2002) apontam que as redes neurais artificiais têm tido ampla e crescente utilização em diversas categorias da mineração de dados, como classificação, estimação, predição e segmentação, aplicados aos diversos campos do conhecimento, inclusive marketing. O sucesso de sua aplicação para a descoberta de padrões nos dados está relacionado à sua capacidade de detectar padrões complexos, muitas vezes obtidos por relações não-lineares entre os dados, tarefa esta que boa parte das outras técnicas de mineração apresenta uma eficiência menor (HAIR *et al.*, 2005). Segundo Hair *et al.* (2005), uma das aplicações mais comuns de redes neurais é a classificação e eles defendem sua utilização sobretudo quando a ênfase é sobre a precisão da classificação e não sobre a interpretação das variáveis que geraram o resultado fornecido pela técnica.

Conceitualmente, Johnson e Wichern (2002) consideram que redes neurais artificiais são algoritmos computacionalmente intensivos para transformar *inputs* (variáveis e atributos contidos nas bases de dados) em *outputs* desejados (resultados atribuídos à variável alvo) por meio de redes altamente conectadas de unidades de processamento relativamente simples, denominados “nós” ou “neurônios”.

#### 2.7.3.4 Regressão Linear Múltipla

Considerada uma das principais e mais utilizadas técnicas estatísticas dentro da mineração de dados (BERRY e LINOFF, 2000). Parte da definição de *inputs* ou variáveis independentes, variáveis que buscam explicar o comportamento de uma variável dependente, ou *output*. A regressão linear múltipla é uma generalização da regressão linear simples, que considera apenas uma variável independente e uma variável dependente (BERSON, SMITH e THEARLING, 1999). Hair *et al.* (2005) consideram que o objetivo da regressão linear múltipla é usar as variáveis independentes, cujos valores são conhecidos, para estimar os valores da variável dependente, que, em quase todas as situações, é desconhecido. Cada uma das variáveis independentes recebe um peso, que denota sua contribuição relativa para a previsão da variável dependente.

Esta técnica parte do princípio que existe uma relação linear entre as variáveis independentes e a variável dependente, de forma que esta relação possa ser ajustada por meio de uma reta (MONTGOMERY, PECK e VINING, 2001). Ela também possui alguns pressupostos, como a necessidade de a variável dependente possuir escala intervalar e a

distribuição dos erros de predição seguir uma distribuição normal (JOHNSON e WICHERN, 2002).

#### 2.7.3.5 Regressão Logística

Modelos de regressão logística são denominados desta forma, pois, assim como em modelos de regressão em geral, procura-se uma relação entre uma ou variáveis independentes e uma variável dependente, porém, ao contrário de modelos de regressão linear, em que o ajuste é feito por meio de uma reta, modelos de regressão logística buscam este ajuste por meio de uma função logística (MONTGOMERY, PECK e VINING, 2001).

Os modelos de regressão logística apresentam a mesma finalidade dos modelos de regressão múltipla, ou seja, estimar o valor da variável dependente, em função dos valores conhecidos das variáveis independentes, bem como estabelecer o grau de relação entre estas variáveis e a variável dependente. Contudo, ela torna-se mais adequada quando a variável dependente é categórica. Quando esta apresenta apenas duas categorias de resposta, o método é denominado regressão logística binária e quando a quantidade de categorias da variável dependente é superior a dois, passa a ser chamado de regressão logística multinomial. Igualmente, é uma técnica mais adequada que a regressão linear múltipla, quando a distribuição dos erros não segue uma distribuição normal (MONTGOMERY, PECK e VINING, 2001). Pelas mesmas razões, também é uma alternativa à análise discriminante, para tarefas de classificação (JOHNSON e WICHERN, 2002 ; HAIR *et al.*, 2005).

#### 2.7.3.6 Séries Temporais

O interesse da empresa em avaliar o comportamento futuro de determinados fenômenos exige a utilização de ferramentas de análise que capturem os padrões de variação destes fenômenos ao longo do tempo. Modelos de séries temporais são como modelos de regressão, em que uma ou mais variáveis independentes procuram explicar o comportamento de uma variável dependente (fenômeno). A diferença é que modelos de séries temporais capturam a correlação serial entre cada unidade de tempo, modelando esta dependência para prever valores futuros da variável dependente (MONTGOMERY, PECK e VINING, 2001). Por conta disso, modelos de séries temporais também são capazes de capturar sazonalidade

(comportamentos cíclicos da variável) e tendência nos dados, de forma a prever quando eles podem voltar a ocorrer no futuro (BROCKWELL e DAVIS, 2002).

#### 2.7.3.7 Análise de Sobrevivência

Especialmente utilizada para a análise de eventos que possuam dados censurados, ou seja, depois de determinado período de tempo, a empresa passa a não mais ter dados acerca de um determinado evento (HOSMER e LEMESHOW, 1999), por exemplo, o comportamento de compra de um cliente, em função deste ter deixado de utilizar os serviços da empresa. A análise de sobrevivência (do inglês, *survival analysis*), segundo Berry e Linoff (2004) é valiosa no entendimento do comportamento dos clientes, sinalizando quando o cliente passa a ter risco de "censura", ou cancelamento do relacionamento com a empresa, identificando também quais fatores têm mais relação com a ocorrência desta censura. Os autores ainda apontam que esta técnica, além de identificar o potencial de clientes cancelarem o relacionamento com empresa, também pode ser útil na identificação do potencial de migração do cliente para outro produto ou serviço, bem como na identificação de quando um cliente retornará após ter cancelado seu relacionamento com a empresa. Berry e Linoff (2004) também consideram que a informação sobre o tempo estimado em que o cliente manterá seu relacionamento com a empresa é um ingrediente fundamental para o cálculo do valor vitalício do cliente (do inglês LTV – *lifetime value*), que fornecerá subsídios à empresa sobre quem são os clientes mais rentáveis.

#### 2.7.3.8 Análise de Agrupamentos

Conforme destaca Kantardzic (2003), análise de agrupamentos (do inglês, *cluster analysis*) refere-se a um conjunto de técnicas de segmentação para classificação automática de dados em um determinado número de grupos utilizando medidas de associação e similaridade, de forma que os registros pertencentes a um grupo sejam similares e registros pertencentes a outros grupos não sejam similares. O autor ainda ressalta que seres humanos estão habituados a fazer este tipo de agrupamento ao trabalhar com uma ou duas variáveis ou características. Contudo, a análise de agrupamentos torna-se mais relevante e útil quando a quantidade de variáveis a serem analisadas conjuntamente aumenta, limitando a capacidade humana de

análise (KANTARDZIC, 2003). Berry e Linoff (2000) apontam a existência de basicamente três conjuntos de técnicas de agrupamento de dados: (1) técnicas divisivas, em que inicialmente todos os registros fazem parte de um grande agrupamento, para então serem particionados em grupos menores, de acordo com os níveis de similaridade detectados entre os registros; (2) técnicas aglomerativas, em que inicialmente cada registro ocupa um grupo separado e, de forma iterativa, combinam-se até que todos os registros tenham sido alocados a um determinado número de grupos; (3) mapas auto-organizáveis ou redes de *Kohonen* (em função de seu criador, o pesquisador finlandês Tuevo *Kohonen*), uma forma especializada de rede neural que pode ser utilizada para a identificação de grupos.

#### 2.7.3.9 Análise de Componentes Principais

Da mesma forma que é possível encontrar agrupamentos de registros similares, também é possível encontrar variáveis com comportamentos similares ou cujo conteúdo seja correlacionado com o de outras variáveis. Sobretudo quando se trabalha com grandes quantidades de dados e muitas variáveis, a questão da dimensionalidade adquire complexidade no processo de mineração de dados (JOHNSON e WICHERN, 2002). A análise de componentes principais busca analisar a estrutura de correlação das variáveis, de forma a encontrar fatores (denominados “componentes principais”) que expliquem, pelo menos em parte, o comportamento das variáveis originais, reduzindo, assim, a dimensionalidade dos dados e facilitando o processo de análise (JOHNSON e WICHERN, 2002). Estes componentes principais representam combinações lineares das variáveis originais e são não-correlacionados. Esta técnica é também bastante utilizada quando se utilizam modelos de regressão linear múltipla ou logística para o desenvolvimento de um modelo, mas as variáveis em análise são altamente correlacionadas, muitas vezes gerando multicolinearidade nos dados e distorcendo as estimativas dos parâmetros dos modelos gerados (MONTGOMERY, PECK, VINING, 2001). Segundo os autores, uma das principais estratégias para a redução da multicolinearidade em modelos de regressão é a utilização da técnica de análise de componentes principais, de forma que os fatores obtidos não apresentem correlação entre si e possam ser utilizados como variáveis independentes no modelo de regressão.

#### 2.7.3.10 Análise de Cestas de Mercado

Esta técnica tem o propósito básico de encontrar associações entre os dados, ou seja, padrões nos dados, sem a necessidade de estimar ou predizer um alvo específico (variável dependente). De acordo com Berry e Linoff (2004), a análise de cestas de mercado (do inglês, *market basket analysis*) identifica associações nos dados e a coerência destas associações fica a cargo da interpretação humana. Seu nome deriva de sua utilização inicial ter sido para a identificação da associação entre a compra de produtos em um supermercado.

Webb (2003) descreve que o objetivo da análise de cestas de mercado é identificar combinações de itens com afinidades com outros itens, de forma a estabelecer o nível de propensão a adquirir um produto, dado que adquiriu um outro qualquer. Em outras palavras, esta análise busca identificar combinações de itens, cuja presença em uma transação afeta a probabilidade da presença de um outro item específico ou de uma combinação de itens (WEBB, 2003). Dependendo do número de itens a serem avaliados, a quantidade de possíveis regras de associação a serem geradas torna-se muito grande, dispendendo um grande esforço computacional. Agrawal e Srikant (1994) apresentam o algoritmo *Apriori*, para geração de regras de associação entre itens. Este algoritmo lida com este problema, utilizando o conceito de suporte, que expressa a representatividade da regra. Desta forma, ocorre uma restrição no espaço de procura das regras, de forma que apenas regras com nível mínimo de suporte sejam geradas, tornando mais ágil o processo de geração das regras de associação.

#### 2.7.3.11 Análise Exploratória de Dados

Este tipo de análise é utilizado nos estágios iniciais da etapa de modelagem de dados, para explorá-los e identificar padrões, tendências e relacionamentos iniciais, que recomendem o uso de determinadas técnicas de mineração de dados. Segundo Chye e Gerry (2002), a descrição e exploração auxilia no entendimento e compreensão dos dados, sugerindo possíveis relações entre variáveis ou mesmo identificando características anormais em sua distribuição, que pode exigir um retorno à etapa de preparação dos dados, de forma a corrigi-los ou evitando que determinados dados sejam utilizados na análise. Davidson e Soukup (2002) consideram que a análise exploratória de dados não é utilizada apenas na etapa de modelagem de dados, mas em todas as etapas do processo de descoberta de conhecimento em bases de dados, por exemplo, nas etapas de preparação e entendimento dos dados, em que são

utilizadas tabelas cruzadas e medidas descritivas como média e desvio padrão para compreender o relacionamento entre variáveis presentes na base de dados e identificar possíveis relações ou mesmo inconsistências em seu conteúdo (DAVIDSON e SOUKUP, 2002).

#### 2.7.3.12 Visualização de Dados

Todo começo de um projeto de mineração de dados envolve coletar e preparar dados, enquanto no final, torna-se necessário explicar os resultados encontrados, não importando se foram utilizadas técnicas de classificação, estimação ou predição (DAVIDSON e SOUKUP, 2002). Pyle(2003) considera que o uso de técnicas de visualização de dados, como histogramas, gráficos de dispersão, gráficos de teia, mapas e gráficos tridimensionais torna a apresentação mais relevante e os padrões encontrados nos dados são mais facilmente explicados, sobretudo quando relações complexas são encontradas. Davidson e Soukup (2002) também defendem a utilização de ferramentas de visualização de dados para auxiliar na tarefa de identificação de padrões não observados por meio do uso de outras técnicas, bem como identificar anomalias nos dados que necessitem ser corrigidas antes do término da etapa de modelagem de dados.

## 2.8 Mineração de dados sob um contexto geral de utilização

Esta dissertação focaliza a mineração de dados, assim como os processos de descoberta de conhecimento em bases de dados, aplicados ao desenvolvimento de estratégias de relacionamento com clientes no setor de serviços. Contudo, a mineração de dados possui aplicação em diversos campos da ciência em que seja necessário qualquer tipo de investigação relacionado à descoberta de padrões ou tendências (FAYYAD *et al.*, 1996), de forma que muitos setores se beneficiam destes conceitos.

Pinheiro (2005) destaca a utilização de redes neurais e mapas auto-ajustáveis (redes de *Kohonen*) para a prevenção de inadimplência em operadoras de telefonia. Nesta aplicação, o autor desenvolve um modelo de propensão ao não-pagamento, identificando os clientes com maior tendência a não pagar suas faturas de telefone, possibilitando à empresa de telefonia

realizar ações preventivas e de acompanhamento junto aos clientes, de forma a reduzir o volume de perdas em função da inadimplência.

Rejesus, Little e Lovell (2004) apresentam um exemplo da mineração de dados em agricultura, quando descrevem a utilização de técnicas de análise exploratória de dados e conceitos de séries temporais para a identificação de padrões fraudulentos na utilização de seguro de grãos nos Estados Unidos. Baseando-se nestas técnicas, os autores mostram ser possível identificar padrões no comportamento de utilização deste seguro, auxiliando no planejamento de ações anti-fraude nesta área.

Stolzer e Halford (2007) descrevem uma aplicação da mineração de dados ao setor aéreo americano, por meio de um comparativo entre as técnicas de análise de regressão múltipla, árvores de decisão e redes neurais, com o propósito de analisar dados de aeronaves *Boeing* 757, de forma a desenvolver programas de garantia da qualidade em operações de vôo destes equipamentos, por meio da identificação de aeronaves com perdas de combustível acima do esperado. Os autores mostram que estas ferramentas podem ser utilizadas para a previsão de consumo de combustível e programas preventivos podem ser desenvolvidos para aperfeiçoar o desempenho das aeronaves, reduzindo custos para as companhias aéreas.

Solomon *et al.* (2006) propõem a utilização de mineração de dados para avaliar como a utilização de câmeras que monitoram sinais vermelhos podem aperfeiçoar segurança de tráfego, reduzindo o volume de fatalidades no trânsito. Por meio do uso de técnicas como árvores de decisão, redes neurais e técnicas de segmentação, os gerentes da companhia de engenharia de trânsito conseguem identificar padrões úteis nos dados de ocorrências de trânsito, permitindo ações preventivas para o funcionamento do tráfego nos cruzamentos controlados por sinais, de forma a reduzir o volume de fatalidades no trânsito.

Carpenter e Lachtermacher (2005) utilizam técnicas de mineração de dados para a busca de padrões implícitos nos dados de alunos de pós-graduação *lato sensu* de uma instituição de ensino superior, de forma a identificar quais características do aluno impactam positiva ou negativamente em seu desempenho no curso. Por meio do uso de técnicas de associação, os autores determinam os fatores que mais afetam o desempenho dos alunos, possibilitando ações preventivas da instituição de ensino.

Rokach e Maimom (2006) apresentam a mineração de dados aplicada à descoberta de padrões existentes em processos de manufatura. Segundo os autores, estes padrões podem ser usados, entre outras coisas, para a melhoria da qualidade no setor de manufatura. Os autores utilizam um novo algoritmo baseado em combinações de testes F de *Snedecor* e os aplicam nos setores de processamento de alimentos e de fabricação de circuitos integrados. Segundo

os autores, este algoritmo pode ser usado para a descoberta da estrutura adequada de decomposição dos atributos dos componentes utilizados nos produtos, de forma a elevar a qualidade da manufatura.

Francisco, Petrielli e Reina (2006) apresentam uma aplicação de técnicas de segmentação a dados do setor elétrico brasileiro, com o propósito de formar e caracterizar segmentos de clientes de baixa tensão da AES Eletropaulo, distribuidora de energia elétrica da grande São Paulo. Com o auxílio da técnica *Two-Step Clusters*, foram identificados quatorze segmentos de clientes, o que permitiu à companhia de energia uma identificação mais precisa dos hábitos e padrões de consumo de cada segmento de clientes e um contínuo aperfeiçoamento do conhecimento do perfil de consumo de clientes-chave, de forma a viabilizar a adoção de ações comerciais específicas e/ou oferta de serviços diferenciados a cada segmento de clientes.

## 2.9 Mineração de dados para criação de estratégias de relacionamento com clientes em empresas de serviços

Min, Min e Eman (2002) destacam que um dos fatores para a manutenção da competitividade dos hotéis no segmento em que atuam é o desenvolvimento de uma estratégia de retenção que possa ser implementada. Segundo eles, o sucesso desta estratégia está associado à gestão realizada pelos hotéis com seus clientes, identificando os meios mais rentáveis para desenvolver uma relação leal com o cliente. Para tal, os autores defendem ser necessário conhecer os clientes e uma das maneiras de atingir este conhecimento é entender suas preferências, de modo a poder interagir com eles. Por meio da análise de questionários respondidos por 281 hóspedes que permaneceram em pelo menos um entre onze hotéis de luxo localizados em Seul (Coréia do Sul), os autores levantaram o perfil dos clientes que, segundo eles, constitui uma importante base para a gestão do relacionamento com clientes e o posterior desenvolvimento de uma estratégia de retenção com os mesmos. Destacam que um dos propósitos mais importantes deste levantamento de perfil dos clientes é o foco naqueles mais rentáveis, de modo a oferecer tratamento especial àqueles clientes que trazem mais retorno para o hotel. Por meio da análise da base de dados, os autores procuram responder às seguintes questões:

- Quais clientes têm maior probabilidade de hospedar-se novamente no hotel ?
- Quais clientes têm maiores riscos de hospedarem-se em um hotel concorrente?
- Como segmentar a população de clientes em rentáveis e não rentáveis?

Fazendo uma analogia à metodologia CRISP-DM, descrita em Chapman *et al.* (1999), esta etapa é a relacionada ao entendimento do problema de negócio e ao objetivo da mineração de dados. Depois desta etapa, Min, Min e Eman (2002) relatam o desenvolvimento das etapas de coleta de dados (similar à etapa de entendimento dos dados da metodologia CRISP-DM) e formatação de dados, de modo a preparar os dados para a aplicação de técnicas de mineração de dados.

A partir da base de dados preparada para a análise, os autores relatam a utilização de árvores de decisão, por meio do algoritmo C5.0, presente no *software* de mineração de dados SPSS *Clementine* (versão 6.0), de modo a estabelecer o perfil dos clientes e responder às questões propostas pela mineração de dados. Em uma etapa anterior de análise, fizeram uso de estatísticas descritivas para construir o perfil demográfico dos clientes e auxiliar no desenvolvimento da árvore de decisão.

O resultado das análises foi resumido em 47 regras geradas pela árvore de decisão, regras estas que podem ser utilizadas pelos hotéis para compreender o perfil de seus clientes, elevar sua retenção e fazer com que retornem ao hotel em algum momento do futuro. Uma das regras desenvolvidas a partir da análise dos questionários, segundo Min, Min e Eman (2002), aponta que se um cliente permanece hospedado em um quarto para não-fumantes do hotel Hilton e ele acredita que a variedade de instalações esportivas e de lazer é extremamente importante para a qualidade do serviço do hotel, então é provável que o cliente revisite o hotel pelo menos mais quatro vezes. Os autores concluem que ao formular uma estratégia de retenção bem-sucedida, deveria ser considerada uma grande variedade de atributos relacionados ao perfil demográfico do cliente, propósitos da viagem, experiências anteriores com o serviço do hotel e a disponibilidade de determinados benefícios ao cliente.

Smith, Willlis e Brooks (2000) apresentam uma aplicação da mineração de dados no segmento de seguros, com o objetivo de compreender os padrões de retenção de clientes nesta área, estimar quem são os clientes mais prováveis a renovar suas apólices de seguro, compreender padrões de reclamação de clientes e estimar quais são os clientes com maior probabilidade de cancelar a apólice antes de seu término.

Os autores relatam que a indústria de seguros é extremamente competitiva e que uma combinação de crescimento de participação de mercado e rentabilidade são vistos como

imperativos de sucesso. Segundo os autores, técnicas de mineração de dados têm sido de enorme utilidade para o mundo dos negócios, em termos de sua capacidade de identificação de padrões complexos nos dados, auxiliando a prever o comportamento futuro dos clientes. A empresa de seguros utilizada para o estudo possui um grande repositório de dados (*datawarehouse*) que registra toda e qualquer transação financeira ou reclamação realizada pelo cliente. Por meio de uma base de dados de 20.914 clientes com apólices de veículos automotivos, cujo vencimento ocorreria em abril de 1998, foram desenvolvidas análises cujos propósitos básicos requeriam descobrir quais eram os clientes mais prováveis a renovar a apólice e aqueles com maior probabilidade de cancelar a mesma antes de seu vencimento.

Por meio do *software* de mineração de dados SAS *Enterprise Miner*, os autores descrevem a utilização de redes neurais para a predição da probabilidade de uma apólice ser renovada ou cancelada e de técnicas de segmentação para a avaliação do padrão das reclamações dos clientes. Os resultados destas análises foram combinados para determinar o preço ótimo de uma apólice individual, um preço que equilibrasse a oportunidade de lucro com a necessidade de retenção do cliente.

Ryals (2003) destaca que um banco inglês utilizou redes neurais para identificar que produtos ele espera que seus clientes adquiram em determinados estágios de seu ciclo de vida ao longo do tempo. O banco então associou o resultado desta análise com os produtos ou serviços que ele sabe que o cliente já possuía e, segundo a autora, abriu-se uma oportunidade para a realização de vendas cruzadas (*cross-selling*), de modo que o cliente adquira um produto ou serviço que necessite, mas não possua, antes que isso seja feito pela concorrência.

Chye e Gerry (2002) apresentam uma aplicação da mineração de dados em um banco. O problema de negócio do banco está relacionado com as altas taxas de cancelamento de seus clientes, de forma que ele deseja descobrir se existe algum padrão no comportamento de clientes que cancelam relacionamento com o banco, se comparado a clientes que mantém seu relacionamento. O propósito é desenvolver um modelo de propensão ao cancelamento, de modo a manter os clientes mais rentáveis e reduzir as taxas de cancelamento. Para desenvolver esta análise, os autores consideraram três técnicas principais: árvores de decisão, regressão logística e redes neurais, por meio da utilização do *software* de mineração de dados SPSS *Clementine*.

Antes da realização das tarefas de modelagem, os autores descrevem a utilização de técnicas descritivas e de visualização dos dados, como média, desvio padrão, mediana e histograma, de modo a compreender o comportamento dos dados, principalmente a relação das variáveis preditoras com a variável alvo (ou variável independente). Esta avaliação auxilia

na compreensão de quais variáveis têm mais relação com o objeto do estudo, que é o cancelamento de clientes. Esta fase tem relação direta com a etapa de entendimento dos dados, da metodologia CRISP-DM (CHAPMAN *et al.*, 1999).

Os autores também mencionam a utilização de técnicas de associação como regras de indução generalizada ou GRI (do inglês, *Generalized Rule Induction*) para estabelecer relações entre variáveis comportamentais associadas ao modo como os clientes utilizam os produtos e o cancelamento voluntário. Por meio de redes neurais, descobrem as variáveis que têm mais importância para explicar o cancelamento de clientes. Utilizam também árvores de decisão e modelos de regressão logística para comparar as similaridades nos resultados, bem como o nível de acerto fornecido por cada técnica, no que diz respeito ao cancelamento de clientes. De acordo com Chye e Gerry (2002), as regras obtidas pela técnica de árvore de decisão, foram as que tiveram o melhor poder de discriminação de clientes que cancelaram, sendo possivelmente este o modelo a ser utilizado pelo banco para identificar clientes com potencial de cancelamento e oferecer a eles pacotes de incentivo ou tomar outras ações preventivas.

Peacock (1998) relata a utilização de modelos de redes neurais pela American Express para examinar centenas de milhões de registros em suas bases de dados, que revelam como e onde seus clientes utilizam o cartão de crédito para realizar transações. O objetivo é um modelo de propensão à compra, que gera um *score* para cada cliente. Baseado neste resultado, a empresa relaciona as ofertas de estabelecimentos afiliados com o histórico transacional dos clientes, de modo a gerar valor para os clientes por meio de ofertas associadas ao seu perfil de compra. Peacock (1998) também aponta a utilização de técnicas de mineração de dados na seleção de *prospects*, ou potenciais clientes. Segundo o autor, áreas de marketing direto das empresas aplicam mineração de dados para descobrir atributos que predigam respostas do cliente a ofertas e programas de comunicação dirigidos pela empresa. Em um próximo estágio, os atributos apontados pelo modelo como associados a potenciais clientes, são utilizados para selecionar clientes potenciais a partir de uma listagem de nomes de clientes compradas no mercado, de modo a elevar a taxa de resposta à oferta da empresa, reduzir custos com aquisição e adquirir clientes com potencial de receita para a companhia.

Drew *et al.* (2001) utilizam um exemplo do setor de telecomunicações e descrevem a utilização de modelos de análise de sobrevivência para a avaliação do valor vitalício do cliente – LTV (*lifetime value*), de forma a descobrir o potencial de receita de cada cliente ao longo do tempo e estabelecer estratégias de retenção em função do nível de lucratividade individual de cada cliente. Contudo, segundo os autores, a utilização do cálculo do LTV para

avaliar o valor do cliente, ignora potenciais efeitos de ações da empresa, como ações de retenção. A partir disso, eles utilizam uma outra medida, baseada em modelos de análise de sobrevivência, denominada valor vitalício generalizado do cliente – GLTV (*generalized lifetime value*), sendo possível, desta maneira, quantificar o efeito que ações da empresa tiveram no valor do cliente ao longo do tempo. Também consideram que o GLTV pode ser utilizado para fins de segmentação de clientes.

Mena e Pettit (2001) apresentam uma aplicação de mineração na internet (*web mining*), utilizando técnicas de mineração de dados para descobrir o perfil de visitantes de uma página de estações de rádio na internet. O objetivo da página é manter os visitantes o maior tempo possível utilizando seus serviços. Os autores descrevem a utilização de análises exploratórias nos dados, por meio do *software* estatístico SPSS, de modo a compreender, de forma descritiva, o comportamento dos visitantes na utilização dos recursos da página e verificar, por meio da aplicação de árvores de classificação, quais variáveis ou características dos visitantes têm relação com a quantidade de minutos em que permanecem conectados aos serviços da página.

Os autores destacaram a importância das etapas anteriores à modelagem e análise de dados, sobretudo a definição de um problema de negócio que seja claramente compreendido, além da qualidade dos dados, que interferem diretamente na qualidade e na lógica dos resultados fornecidos pelas técnicas de mineração de dados.

# 3. PROCEDIMENTOS METODOLÓGICOS

De acordo com Sekaran (1992), a natureza do estudo está relacionada aos estágios de avanço do conhecimento na área de pesquisa. Segundo ela, existem três tipos de estudo: o exploratório, o descritivo e o experimental (conduzido a se testar hipóteses). Estudos descritivos têm a pretensão de descrever e compreender características de variáveis de um estudo. Conforme aponta Sekaran (1992), estudos descritivos também auxiliam a entender as características de um grupo em uma situação de interesse. Estudos deste tipo apresentam as características de determinada amostra ou de determinado fenômeno, porém não se propõem a explicar os fenômenos que descrevem, embora sirvam de base para tal explicação

Esta dissertação teve caráter **descritivo**, pois se pretendeu descrever os aspectos do processo de análise de bases de dados para a descoberta de conhecimento, conhecimento este que subsidia a criação de estratégias de relacionamento com clientes.

O estudo em questão foi **correlacional**, pois se pretendeu investigar a relação existente entre a utilização das etapas do processo de análises de bases de dados para a descoberta de conhecimento e variáveis associadas à empresa, como tamanho, origem, faturamento e tipo de serviço prestado, sem contudo estabelecer uma relação de causa e efeito entre essas variáveis. Cavana, Delahaye e Sekaran (2000) defendem o uso de investigações correlacionais quando se deseja encontrar as variáveis mais importantes associadas ao problema, sem contudo estabelecer uma relação de causa e efeito entre elas.

O conceito de população, segundo Cavana, Delahaye e Sekaran (2000) é referente ao grupo inteiro de pessoas, eventos ou objetos de interesse que o pesquisador deseja investigar. Nesta dissertação, a **população de interesse** foi composta pelas empresas do setor de serviços que atuam nas cidades de São Paulo e do Rio de Janeiro e que utilizam modelagem de dados para o desenvolvimento de estratégias de relacionamento com clientes.

De acordo com Malhotra (2001), amostra é um subgrupo de uma população, selecionado para participação do estudo. A **amostra utilizada** para pesquisa foi um subconjunto das empresas do setor de serviços que atuam nas cidades de São Paulo e do Rio de Janeiro, que utilizam processos de modelagem de dados para o desenvolvimento de estratégias de relacionamento com clientes.

A unidade de análise refere-se ao nível de agregação dos dados coletados durante análises posteriores. Segundo Cavana, Delahaye e Sekaran (2000), as unidades de análise podem ser indivíduos, grupos, empresas, setores ou países.

Nesta dissertação, a **unidade de análise** foi composta por gerentes e analistas de CRM analítico, *database marketing* ou de modelagem de dados de marketing, que trabalhem em empresas de serviços que atuam nas cidades de São Paulo ou do Rio de Janeiro e que façam uso de métodos quantitativos para o desenvolvimento de estratégias de relacionamento com clientes.

Richardson (1999) define o método científico como um tipo específico de método de pesquisa que consiste na delimitação de um problema, realização de observações e interpretação dos dados com base nas relações encontradas, fundamentando-se, quando possível, nas teorias existentes. Ele propõe dois grandes grupos de métodos: O método quantitativo, caracterizado pelo uso da quantificação e uso de métodos estatísticos na coleta e análise dos dados, e o qualitativo, que não emprega instrumentos estatísticos de análise, pois sua pretensão não é estabelecer medidas, mas sim compreender a natureza de um dado fenômeno.

Esta dissertação propôs a quantificação do nível de utilização de cada uma das etapas dos processos de análise de dados para a descoberta de conhecimento sobre o cliente em empresas de serviços que atuam nas cidades de São Paulo e do Rio de Janeiro, de tal forma que o estudo realizado foi **quantitativo**, tendo utilizado como instrumento de pesquisa, questionários com questões fechadas e questões semi-abertas, cuja aplicação foi feita por meio do envio, via *e-mail*, a profissionais que atuam nestas empresas.

O acesso ao público alvo da pesquisa (gerentes ou analistas de CRM analítico, *database marketing* ou de modelagem de dados em marketing) é restrito e não foi encontrado qualquer cadastro público que contivesse uma listagem dos profissionais destas áreas, até porque não possuem formação específica. Portanto, uma amostragem probabilística tornou-se difícil neste caso, tanto por causa do acesso restrito a todos os profissionais, quanto pela limitação financeira e temporal para a realização da pesquisa.

A proposta alternativa foi a utilização de uma técnica de amostragem **não-probabilística denominada "bola-de-neve"**. Conforme descreve Malhotra (2001), na amostragem tipo "bola-de-neve", um grupo inicial de entrevistados é selecionado e, a partir deles, selecionam-se entrevistados subseqüentes, com base nas informações fornecidas pelos entrevistados iniciais. Cavana, Delahaye e Sekaran (2000) comentam que uma amostragem tipo "bola-de-neve" é utilizada quando os elementos da população tem características específicas e são muito difíceis de serem localizados ou contatados.

No que se refere ao tamanho da amostra, foram entrevistados 67 profissionais que atuam em empresas de serviços das cidades de São Paulo e do Rio de Janeiro, desenvolvendo

atividades nas áreas de CRM analítico, database marketing ou modelagem de dados em marketing.

Para a avaliação dos níveis de utilização e dos níveis de importância das etapas de análise de processos de descoberta de conhecimento, bem como das técnicas de mineração de dados, foi utilizada uma **escala numérica** (CAVANA, DELAHAYE e SEKARAN, 2000), **variando entre 0 e 10** (0 – nenhuma utilização e 10 – utilização total, ou 0 – nenhuma importância e 10 – importância total).

Um teste piloto pode ser conduzido para detectar pontos fracos no planejamento da pesquisa (COOPER e SCHINDLER, 2003). Alguns elementos da população alvo são selecionados para a simulação dos procedimentos e protocolos designados para a coleta de dados (COOPER e SCHINDLER, 2003), sendo que estes elementos devem guardar semelhanças com os entrevistados da pesquisa real, em termos de familiaridade com o assunto e atitudes e comportamentos de interesse (MALHOTRA, 2001). Malhotra (2001) também considera que o teste piloto deve ser abrangente e todos os aspectos do questionário devem ser testados, como enunciado, seqüência, *layout* e instruções, de forma que as questões formuladas representem efetivamente aquilo que se deseja medir. Foi realizado um teste piloto com **dez entrevistados**, de modo a validar o instrumento de pesquisa utilizado nesta dissertação.Estes dez entrevistados posteriormente responderam novamente o questionário e fizeram parte da amostra final

Do ponto de vista da análise dos resultados, como o estudo teve características descritivas, muito em função de ter sido utilizada uma amostra não-probabilística (COCHRAN, 1976) e seu tamanho não ter sido determinado por meio de um cálculo de tamanho mínimo de amostra (COCHRAN, 1976), foi necessário ter cuidado e precaução com relação à aplicação de técnicas estatísticas inferenciais, visto que são extremamente úteis para detectar determinados padrões em uma amostra e generalizar estes resultados para a população (COCHRAN, 1976). Considerando que a amostra coletada não foi aleatória e não houve evidências acerca de sua representatividade, estas técnicas perdem sua função científica, de generalização do resultado.

Contudo, somente para efeito de apoio à comparação dos resultados, foram aplicados alguns testes não paramétricos de comparação de amostras, como testes de *Mann-Whitney* (comparação de dois grupos independentes) e *Wilcoxon* (comparação de dois grupos relacionados), como alternativas à realização do teste t, bem como os testes de *Kruskal-Wallis* (comparação de mais de dois grupos independentes) e de *Friedman* (comparação de mais de dois grupos dependentes ou relacionados) (SIEGEL, 1975), como alternativas à Análise de

Variância (ANOVA). A escolha de testes não-paramétricos, em detrimento de testes paramétricos como o teste t, deveu-se ao fato de o tamanho da amostra ser relativamente pequeno e de não terem existido evidências quanto à simetria da distribuição do valor de cada um dos itens pesquisados. Nestes casos, a aplicação de testes não-paramétricos é recomendada (HOGG e TANIS, 1997).

A aplicação dos testes não-paramétricos pode ou não levar à identificação de diferenças significativas entre os grupos. Vale lembrar que, como o tamanho da amostra é reduzido, muitas diferenças que numericamente existem, podem não ser confirmadas do ponto de vista estatístico. Contudo, com o aumento do tamanho da amostra e manutenção das diferenças observadas na amostra original, há mais possibilidades de se encontrar diferenças estatisticamente significativas.

Para a avaliação do nível de consistência das medidas (níveis de utilização e níveis de importância de cada uma das etapas do processo de KDD), foi utilizado o indicador *Alpha de Cronbach* (MALHOTRA, 2001), que varia entre 0 e 1 e que quanto mais próximo de 1, maior a confiabilidade do item avaliado. O *software* utilizado para a compilação do questionário, bem como para a construção das tabelas e análises foi o SPSS 13.0.

## 3.1 Relação de objetivos específicos x questões da pesquisa quantitativa

O Quadro 3 interliga os objetivos específicos da dissertação às questões exploradas no questionário, de forma que os resultados da pesquisa possam responder aos propósitos dessa dissertação.

**Quadro 3**: Relação de objetivos específicos x questões

| **Objetivos Específicos** | **Conceitos-chave: fundamentação teórica** | **Perguntas do Questionário** |
|---|---|---|
| **Objetivo específico 1:** Avaliar o nível de utilização, nas empresas de serviços que atuam nas cidades de São Paulo e do Rio de Janeiro, das etapas do processo de descoberta de conhecimento em bases de dados (KDD), representadas na metodologia | Shearer (2000) descreve a metodologia de livre distribuição chamada CRISP-DM (do inglês *Cross Industry Standard Process for Data Mining*), desenvolvida em 1996 por um importante consórcio de empresas européias (RODRIGUEZ *et al.*, 2006) composto por Daimler-Benz (atual DaimlerChrysler), *Integral Solutions Ltd.* | - Assinale um valor (entre 0 e 10) para o grau de intensidade com que sua empresa utiliza as seguintes etapas de análise para dar suporte à criação de estratégias de relacionamento com clientes, considerando que 0 (zero) significa nenhuma |

| | | |
|---|---|---|
| CRISP-DM ? | (ISL), NCR e OHRA, época em que as empresas, sobretudo do setor da indústria na Europa, não possuíam uma metodologia oficial para a condução de projetos de mineração de dados e que a necessidade de descoberta de conhecimento em bases de dados era crescente, em função do aumento da concorrência e da quantidade de clientes. A metodologia CRISP-DM posiciona a mineração de dados dentro de um contexto de negócio mais amplo, de descoberta de conhecimento em bases de dados, estabelecendo seis fases principais, desde o entendimento do problema de negócio da empresa, passando pela mineração de dados, até a implementação do modelo desenvolvido | utilização da respectiva etapa de análise e 10 (dez) significa utilização total da respectiva etapa de análise.<br><br>- Assinale um valor (entre 0 e 10) para o grau de importância com que sua empresa utiliza as seguintes etapas de análise para dar suporte à criação de estratégias de relacionamento com clientes, considerando que 0 (zero) significa nenhuma utilização da respectiva etapa de análise e 10 (dez) significa utilização total da respectiva etapa de análise. |
| **Objetivo específico 2**: Avaliar o nível de utilização, nas empresas de serviços que atuam nas cidades de São Paulo e do Rio de Janeiro, das técnicas de mineração de dados identificadas na literatura. | As técnicas de mineração de dados, com aplicação mais direcionada ao marketing e às estratégias de relacionamento com clientes, podem ser agrupadas em seis grandes categorias de análise (DAVIDSON E SOUKUP, 2002; BERRY E LINOFF, 2004): Classificação, estimação, predição, agrupamento, associação e descrição. Cada uma destas categorias possui técnicas associadas (Quadro 2). | - Assinale um valor (entre 0 e 10) para o grau de intensidade com que sua empresa utiliza as seguintes técnicas de análise de modelagem de dados para dar suporte à criação de estratégias de relacionamento com clientes, considerando que 0 (zero) significa nenhuma utilização da respectiva técnica de análise e 10 (dez) significa utilização total da respectiva técnica de análise. |
| **Objetivo específico 3:** Avaliar o nível de utilização, nas empresas de serviços que atuam nas cidades de São Paulo e do Rio de Janeiro, das estratégias de relacionamento com clientes identificadas na literatura. | Brown (2001) considera a existência de quatro tipos de estratégias a serem utilizadas em função do momento do ciclo de vida do cliente no relacionamento com a empresa: Busca de clientes em potencial, *cross-selling/up-selling* e fidelização de clientes. | - Assinale um valor (entre 0 e 10) para o grau de intensidade com que sua empresa utiliza as seguintes estratégias de relacionamento com clientes, considerando que 0 (zero) significa nenhuma utilização da estratégia e 10 (dez) significa utilização total da estratégia. |

| | | |
|---|---|---|
| **Objetivo específico 4:** Verificar se o nível de utilização das técnicas de mineração de dados para geração de estratégias de relacionamento com clientes tem relação com a existência de processos de CRM analítico | - **CRM analítico**: Conforme consideram Peppers e Rogers (2004), trata-se da fonte de inteligência do processo, pois serve para o ajuste e manutenção das estratégias de diferenciação de clientes, bem como o acompanhamento de seus hábitos e necessidades e os eventos que podem ocorrer na história do relacionamento entre eles e a empresa. Os dados armazenados nas bases de dados da empresa são analisados por meio de ferramentas analíticas, de modo a gerar perfis de clientes, identificar padrões de comportamento, determinar níveis de satisfação e dar suporte à segmentação de clientes. A informação e o conhecimento adquiridos a partir do CRM analítico auxiliam no desenvolvimento de estratégias adequadas de promoção e marketing. Este tipo de CRM é classificado por Kotorov (2003) como uma visão de 360 graus do cliente. Tecnologias que reforçam o CRM analítico incluem datawarehouses, ferramentas analíticas e preditivas (ECKERSON e WATSON, 2001), regras de associação e descoberta de padrões, classificação e avaliação do valor do cliente (AHN, KIM e HAN, 2003). | - Assinale um valor (entre 0 e 10) para o grau de intensidade com que sua empresa utiliza as seguintes técnicas de análise de modelagem de dados para dar suporte à criação de estratégias de relacionamento com clientes, considerando que 0 (zero) significa nenhuma utilização da respectiva técnica de análise e 10 (dez) significa utilização total da respectiva técnica de análise.<br><br>- A empresa em que você trabalha possui uma área de CRM analítico ? |
| **Objetivo específico 5:** Verificar se o nível de utilização das etapas dos processos de descoberta de conhecimento em bases de dados para geração de estratégias de relacionamento com clientes tem relação com variáveis intrínsecas à empresa, como segmento de atuação no setor de serviços, faturamento anual, quantidade de | Segundo Berry e Linoff (2004), a mineração de dados faz mais sentido quando há grandes volumes de dados, já que a maioria dos algoritmos de mineração de dados exige grandes volumes para construir e treinar modelos que serão utilizados para realizar tarefas de classificação, predição ou estimação. De acordo com Berry e Linoff (2004), segmentos do setor de serviços como | - Em que segmento do setor de serviços atua a empresa em que você trabalha ?<br>- Qual o faturamento em 2006 da empresa em que você trabalha ?<br>- Qual a quantidade de clientes da empresa em que você trabalha ?<br>- Qual a origem da empresa em |

| | | |
|---|---|---|
| clientes, tipo de serviço prestado e nacionalidade da empresa. | telecomunicações e cartões de crédito há algum tempo têm mantido um relacionamento interativo com clientes e gerado registros de transação. Contudo, recentemente a maioria dos setores têm armazenado, em um ritmo cada vez mais acelerado, dados a respeito da interação com clientes, possibilitando o uso de bases de dados para a avaliação do comportamento dos clientes. Han e Kamber (2000) percebem algo similar quando apontam que a crescente atenção que a mineração de dados tem recebido nos últimos anos por parte da indústria de informação é resposta à ampla disponibilidade de grandes volumes de dados e a iminente necessidade de se transformar dados em informação e em conhecimento, necessidade esta, em parte, motivada pelo aumento da concorrência entre as empresas, aumento de oferta de produtos e serviços, maior exigência por parte dos clientes, o que faz com que as empresas tenham a necessidade de conhecer melhor seus clientes para poder oferecer produtos e serviços mais adequados às suas exigências e necessidades e consigam, a partir disso, manter os melhores clientes. | que você trabalha ?<br>- Com que freqüência a empresa em que você trabalha utiliza as seguintes etapas de análise para subsidiar as estratégias de relacionamento com clientes ? |

**Fonte**: Elaborado pelo autor

# 4. ANÁLISE E INTERPRETAÇÃO DOS RESULTADOS DA PESQUISA

## 4.1 Descrição da amostra coletada

**Tabela 1 -** Distribuição por segmento do setor de serviços

| Segmento | Freqüência | % |
|---|---|---|
| Finanças | 27 | 40,30 |
| Telecom | 8 | 11,94 |
| Seguros | 7 | 10,45 |
| Consultoria | 5 | 7,46 |
| Comunicação | 15 | 22,39 |
| Outros | 5 | 7,46 |
| Total | 67 | 100,00 |

**Fonte:** Elaborado pelo autor

O tamanho final da amostra coletada para o estudo foi de 67 profissionais (correspondente a 60 empresas diferentes), distribuídos nos segmentos mostrados na Tabela 1, referente a empresas do setor de serviços que atuam nas cidades de São Paulo e do Rio de Janeiro. Houve uma concentração nos segmentos de finanças, de telecomunicações e de seguros, totalizando quase 63% do total de profissionais. O segmento de comunicação, responsável por mais de 22% do total da amostra foi composto pela junção dos segmentos de TV por assinatura (4 profissionais), editorial (5 profissionais), publicidade e propaganda (3 profissionais) e portal de internet (3 profissionais). A concentração mencionada, de certa forma é explicada por Berry e Linoff (2004), que mencionam que o setor financeiro e de telecomunicações têm um registro mais antigo de armazenamento de dados sobre clientes, de forma que, nestes setores, tem-se observado uma quantidade maior de aplicações das ferramentas de mineração de dados, não somente para relacionamento com clientes, como também para análise de risco de crédito (MADEIRA e OLIVEIRA, 2003; ANTUNES, KATO e CORRAR, 2002) e detecção de fraude (HORMAZI e GILES, 2004).

Vale também comentar que o conteúdo informado pelos 67 questionários representa a percepção dos profissionais que aceitaram participar da pesquisa. Outros 22 profissionais contatados não aceitaram participar da pesquisa (por questões de confidencialidade das informações) ou não responderam o questionário após cinco tentativas de contato por e-mail (cada contato ocorrido ao fim de cada semana).

**Tabela 2 -** Distribuição por faturamento anual

| Faturamento anual | Freqüência | % | % acumulado |
|---|---|---|---|
| Menos de R$ 1 milhão | 2 | 2,99 | 2,99 |
| Entre R$ 1 milhão e R$ 5 milhões | 4 | 5,97 | 8,96 |
| Entre R$ 5 milhões e R$ 10 milhões | 3 | 4,48 | 13,43 |
| Entre R$ 10 milhões e R$ 25 milhões | 2 | 2,99 | 16,42 |
| Entre R$ 25 milhões e R$ 50 milhões | 2 | 2,99 | 19,40 |
| Entre R$ 50 milhões e R$ 100 milhões | 4 | 5,97 | 25,37 |
| Entre R$ 100 milhões e R$ 250 milhões | 2 | 2,99 | 28,36 |
| Entre R$ 250 milhões e R$ 500 milhões | 8 | 11,94 | 40,30 |
| Acima de R$ 500 milhões | 40 | 59,70 | 100,00 |
| Total | 67 | 100,00 | - |

**Fonte:** Elaborado pelo autor

**Tabela 3 -** Distribuição por quantidade de clientes

| Quantidade de clientes | Freqüência | % | % acumulado |
|---|---|---|---|
| Menos de 1.000 clientes | 12 | 17,91 | 17,91 |
| Entre 1000 e 10000 clientes | 6 | 8,96 | 26,87 |
| Entre 10.001 e 25.000 clientes | 0 | 0,00 | 26,87 |
| Entre 25.001 e 50.000 clientes | 2 | 2,99 | 29,85 |
| Entre 50.001 e 100.000 clientes | 2 | 2,99 | 32,84 |
| Entre 100.001 e 250.000 clientes | 1 | 1,49 | 34,33 |
| Entre 250.001 e 500.000 clientes | 5 | 7,46 | 41,79 |
| Entre 500.001 e 1.000.000 de clientes | 4 | 5,97 | 47,76 |
| Mais de 1.000.000 de clientes | 35 | 52,24 | 100,00 |
| Total | 67 | 100,00 | - |

**Fonte:** Elaborado pelo autor

**Tabela 4-** Distribuição por segmento do setor de serviços (para empresas com quantidade de clientes <=50.000)

| Segmento | Freqüência | % |
|---|---|---|
| Finanças | 3 | 15,00 |
| Telecom | 2 | 10,00 |
| Seguros | 1 | 5,00 |
| Consultoria | 5 | 25,00 |
| Comunicação | 8 | 40,00 |
| Outros | 1 | 5,00 |
| Total | 20 | 100,00 |

**Fonte:** Elaborado pelo autor

De acordo com as Tabelas 2 e 3, mais de 70% dos profissionais declararam que suas empresas possuem faturamento superior a R$ 250 milhões (referência: 2006) e mais de 65% declararam que suas empresas possuem mais de 250.000 clientes. Considerando estas informações, percebe-se que a maior parte da amostra é composta por profissionais que trabalham em médias e grandes empresas. A Tabela 4 aponta que, das empresas com menos de 50.000 clientes, 65% delas são do segmento de consultoria e comunicação, cujos clientes são outras empresas, muitas delas de grande porte.

Segundo Berry e Linoff (2004), em países desenvolvidos, o uso de técnicas de mineração de dados tende a ser mais comum em empresas com grandes volumes de dados, o que justificaria a concentração observada na amostra coletada. Contudo, como a amostra coletada foi não-probabilística, não é possível afirmar que, para o cenário brasileiro, empresas pequenas não utilizam modelagem de dados para o desenvolvimento de estratégias de relacionamento com clientes.

**Tabela 5 -** Distribuição por nacionalidade da empresa

| Nacionalidade da empresa | Freqüência | % |
|---|---|---|
| Brasileira | 34 | 50,75 |
| Americana | 19 | 28,36 |
| Inglesa | 2 | 2,99 |
| Espanhola | 5 | 7,46 |
| Holandesa | 3 | 4,48 |
| Outras | 4 | 5,97 |
| Total | 67 | 100,00 |

**Fonte:** Elaborado pelo autor

**Tabela 6 -** Distribuição por *software* utilizado (questão com respostas múltiplas)

| *Software* utilizado para análise | Sim | Não | % sim |
|---|---|---|---|
| SPSS | 45 | 22 | 67,20 |
| SAS | 42 | 25 | 62,70 |
| *Statistica* | 1 | 66 | 1,50 |
| Minitab | 2 | 65 | 3,00 |
| SPSS *Clementine* | 15 | 52 | 22,40 |
| SAS *Enterprise Miner* | 25 | 42 | 37,30 |
| *Statistica Dataminer* | 0 | 67 | 0,00 |
| IBM *Intelligence Miner* | 1 | 66 | 1,50 |
| Outros | 14 | 53 | 20,90 |

**Fonte:** Elaborado pelo autor

**Tabela 7 –** Utilização de *softwares* de Estatística

| Nacionalidade da empresa | Freqüência | % |
|---|---|---|
| SPSS ou SAS | 64 | 95,52 |
| Outros | 19 | 4,48 |

**Fonte:** Elaborado pelo autor

**Tabela 8 –** Utilização de *softwares* específicos de mineração de dados

| Nacionalidade da empresa | Freqüência | % |
|---|---|---|
| *Clementine* e *Enterprise Miner* | 34 | 50,75 |
| Outros | 0 | 0,00 |
| Não utilizam este tipo de ferramenta | 33 | 49,25 |

**Fonte:** Elaborado pelo autor

A Tabela 5 mostra que quase 80% da amostra é composta por profissionais que trabalham em empresas brasileiras ou americanas. A partir da Tabela 6, percebe-se que as ferramentas utilizadas para o desenvolvimento de modelos e análises que subsidiam a construção de estratégias de relacionamento com clientes são provenientes de dois grandes fornecedores de software: SAS *Institute* e SPSS *Inc.*, duas empresas americanas que comercializam soluções analíticas para estatística e mineração de dados. Considerando *softwares* específicos de estatística, os *softwares* SPSS e SAS foram os mais mencionados pelos 67 profissionais que responderam à pesquisa, com respectivamente 67,20% e 62,70% de respostas positivas quanto à sua utilização. Um número importante registrado na Tabela 7 é que aproximadamente 96% dos respondentes declararam utilizar SPSS ou SAS, o que comprova a grande penetração destas duas marcas na amostra pesquisada.

Quanto a *softwares* específicos de mineração de dados, a Tabela 6 mostra que mais de 37% declararam utilizar SAS *Enterprise Miner* e aproximadamente 22% declararam utilizar SPSS Clementine. Segundo informa a Tabela 8, pouco mais da metade dos respondentes declararam utilizar pelo menos uma ferramenta de mineração de dados, sendo que destes, 100% declararam utilizar *Clementine* ou *Enterprise Miner*. Uma explicação para estes percentuais serem inferiores aos da utilização de SPSS e SAS pode estar no preço e na difusão da cultura de mineração de dados nas empresas. Primeiro, o preço de um *software* de mineração de dados como SPSS *Clementine* e SAS *Enterprise Miner* é substancialmente maior que o de *softwares* como SPSS e SAS. Segundo, as empresas que atuam no Brasil têm passado por uma fase de adaptação à cultura analítica e de modelagem de dados, de forma que estes indicadores tendem a se elevar em alguns anos, até mesmo em função da redução de

custo de ferramentas analíticas de mineração de dados e o aumento de investimento nos setores de inteligência analítica das empresas.

## 4.2 Análise dos objetivos específicos

**Objetivo específico 1:** Avaliar o nível de utilização, por parte das empresas de serviços que atuam na cidade de São Paulo, das etapas (e de cada uma de suas tarefas) do processo de descoberta de conhecimento em bases de dados identificadas na literatura, de modo a subsidiar a criação de estratégias de relacionamento.

**Tabela 9 –** Nível de utilização e de importância das tarefas do processo de KDD

| Item | Etapa de análise | Intensidade | | Importância | | Sig. |
|---|---|---|---|---|---|---|
| | | Média | Desvio padrão | Média | Desvio padrão | |
| 1 | Problema de negócio da empresa | 8,51 | 1,31 | 8,88 | 1,29 | **0,01** |
| 2 | Metas e objetivos do projeto de modelagem de dados | 8,33 | 1,46 | 8,57 | 1,34 | 0,13 |
| 3 | Cronograma com fases do projeto | 8,01 | 1,97 | 8,01 | 1,89 | 0,80 |
| 4 | Seleção das variáveis mais relevantes | 8,87 | 1,39 | 8,55 | 1,69 | 0,11 |
| 5 | Identificação de inconsistência nos dados | 8,78 | 1,51 | 8,60 | 1,61 | 0,17 |
| 6 | Limpeza dos dados | 8,76 | 1,46 | 8,58 | 1,58 | 0,09 |
| 7 | Criação de novas variáveis a partir das existentes | 8,48 | 1,88 | 8,31 | 1,81 | **0,03** |
| 8 | Reunião com áreas de negócio para discussão das variáveis | 8,19 | 1,64 | 8,27 | 1,63 | 0,10 |
| 9 | Montagem de uma base única de análise | 8,63 | 1,54 | 8,39 | 1,82 | **0,03** |
| 10 | Estudo de técnicas de modelagem a serem usadas | 8,16 | 1,83 | 7,46 | 2,18 | **0,00** |
| 11 | Comparação de resultados por mais de uma técnica | 7,18 | 2,64 | 6,88 | 2,50 | 0,13 |
| 12 | Criação de amostra de desenvolvimento x validação | 8,85 | 1,59 | 8,22 | 2,02 | **0,01** |
| 13 | Avaliação do modelo utilizando critérios técnicos | 8,57 | 1,45 | 8,07 | 1,95 | **0,02** |
| 14 | Avaliação do modelo utilizando critérios de negócio | 8,46 | 1,67 | 8,45 | 1,81 | 0,97 |
| 15 | Reunião com áreas de negócio para discussão do modelo | 8,03 | 1,76 | 8,19 | 1,83 | 0,31 |
| 16 | Especificação de um plano de implementação do modelo | 7,99 | 1,71 | 8,22 | 1,81 | 0,10 |
| 17 | Cronograma de especificação do modelo desenvolvido | 7,88 | 1,93 | 7,93 | 2,07 | 0,79 |
| 18 | Montagem de documentação oficial do projeto | 7,66 | 2,20 | 7,79 | 2,16 | 0,44 |
| 19 | Apresentação oficial dos resultados do projeto | 8,22 | 1,82 | 8,21 | 1,72 | 0,99 |
| 20 | Simulação dos modelos em sistemas | 7,57 | 2,56 | 7,58 | 2,46 | 0,73 |
| 21 | Acompanhamento do modelo após a implementação | 8,42 | 1,51 | 8,30 | 1,76 | 0,48 |
| 22 | Revisões periódicas dos modelos desenvolvidos | 8,00 | 1,83 | 8,10 | 1,84 | 0,53 |

**Fonte:** Elaborado pelo autor

**Tabela 10 –** Nível de utilização e de importância das etapas do processo de KDD

| Etapa do KDD | Intensidade | | Importância | | Sig. |
|---|---|---|---|---|---|
| | Média | Desvio padrão | Média | Desvio padrão | |
| Entendimento do problema de negócio | 8,28 | 1,28 | 8,49 | 1,27 | **0,03** |
| Entendimento dos dados | 8,80 | 1,26 | 8,58 | 1,48 | 0,08 |
| Preparação dos dados | 8,43 | 1,23 | 8,32 | 1,37 | 0,29 |
| Modelagem dos dados | 8,06 | 1,68 | 7,52 | 1,93 | **0,01** |
| Avaliação do modelo | 8,35 | 1,37 | 8,24 | 1,62 | 0,45 |
| Implementação do modelo | 7,96 | 1,49 | 8,02 | 1,58 | 0,57 |

**Fonte:** Elaborado pelo autor

Na Tabela 9, os itens 1 a 22 são tarefas de análise que correspondem às etapas dos processos de descoberta de conhecimento em bases de dados (CHAPMAN *et al.*, 1999; BRACHMAN e ANAND, 1996; FAYYAD, PIATESTKY-SHAPIRO e SMITH, 1996). Os itens 1 a 3 representam a etapa de entendimento do problema de negócio, enquanto que os itens 4 a 6 representam a etapa de entendimento dos dados, os itens 7 a 9 representam a etapa de preparação dos dados, os itens 10 a 12 representam a etapa de modelagem dos dados, os itens 13 a 15 representam a etapa de avaliação do modelo e os itens 16 a 22 representam a etapa de implementação do modelo.

Os níveis de intensidade na utilização de etapas do processo de descoberta de conhecimento a partir de bases de dados, mostrados na Tabela 9 foram relativamente altos, considerando que a escala utilizada para a medição dos 22 itens pesquisados varia entre 0 e 10. Observou-se uma aderência entre o nível de utilização destas etapas nas empresas de serviços pesquisadas na amostra e as metodologias observadas na literatura com relação às etapas dos processos de descoberta de conhecimento (CHAPMAN *et al.*, 1999; BRACHMAN e ANAND, 1996; FAYYAD, PIATESTKY-SHAPIRO e SMITH, 1996). Ainda conforme mostra a Tabela 9, dos 22 itens avaliados, 17 apresentaram média de utilização igual ou superior a 8, sendo que dos 5 itens restantes, cuja média é inferior a 8, quatro deles referem-se a etapas do processo de implementação dos modelos construídos em etapas anteriores. O item 11, referente à comparação de resultados por mais de uma técnica, apresentou média de 7,18, menor média de utilização entre todos os itens avaliados, porém com o maior nível de dispersão em torno da média (desvio padrão = 2,64), indicando que algumas empresas têm utilizado este item de forma intensiva, enquanto outras não o tem considerado de forma tão freqüente em seus processos de descoberta de conhecimento em bases de dados.

Traçando um paralelo entre os níveis médios de intensidade de utilização e de importância, dos 22 itens associados ao processo de descoberta de conhecimento em bases de dados, nota-se que os respondentes, de maneira geral, apenas considerando as diferenças observadas, têm a percepção de que a empresa considera muito importante as etapas de entendimento de seu problema de negócio (item 1), definição de metas e objetivos de um projeto de modelagem de dados (item 2) e discussões realizadas com áreas de negócio (tanto para a definição de variáveis, quanto para o debate sobre os resultados do modelo – itens 8 e 15, respectivamente) e que o nível de importância atribuído a estas etapas é maior que o realizado por eles na prática dos projetos de mineração de dados.

Apenas para efeito de comparação e partindo de uma premissa de aleatoriedade e representatividade da amostra (conforme discutido no capítulo 3 – Procedimentos Metodológicos), é possível testar estatisticamente estas diferenças entre os níveis de utilização e de importância, por meio de um teste estatístico não-paramétrico de comparação de 2 amostras dependentes (o mesmo profissional respondeu quais eram os níveis de utilização e de importância) denominado teste de *Wilcoxon* (SIEGEL, 1975). A significância do teste aparece na coluna “sig”., da Tabela 9. Quanto menor o valor da significância, (que varia entre 0 e 1), mais estatisticamente significativa é a diferença entre o nível de utilização e o nível de importância de determinado item pesquisado.

No caso, o resultado do teste de *Wilcoxon* para os itens 1, 7, 9, 10, 12 e 13 foram destacados em negrito, pois para estes, o nível médio de utilização foi estatisticamente maior que o nível de importância (considerando um nível de significância de 5%). Exceto para a tarefa 1, todas as outras cinco referem-se a tarefas de preparação de dados, modelagem e avaliação do modelo.

Dos itens associados à questão que envolvem discussões entre a área de negócio e a área técnica (itens 1, 2, 8 e 15), apenas para o item 1 – entendimento do problema de negócio da empresa – foi detectada uma diferença significativa entre o nível de utilização e o nível de importância atribuído pela empresa à tarefa.

A Tabela 10 mostra que, considerando as seis etapas do processo de KDD, o nível de importância atribuído à etapa de entendimento do problema de negócio foi estatisticamente superior ao nível de utilização desta etapa. Por outro lado, o nível de utilização da etapa de modelagem de dados foi estatisticamente superior ao nível de importância atribuído a ela. Isto mostra que a percepção dos respondentes é que o nível de importância que a empresa atribui à etapa inicial de planejamento é maior do que o nível de utilização desta etapa na prática dos projetos de mineração de dados, situação que se inverte quando se refere à etapa de

modelagem de dados, em que os respondentes declararam que o nível médio de utilização desta etapa, na prática, é superior ao nível de importância atribuída a ela.

**Objetivo específico 2:** Avaliar o nível de utilização, nas empresas de serviços que atuam na cidade de São Paulo, das técnicas de mineração de dados identificadas na literatura, de modo a subsidiar a criação de estratégias de relacionamento.

**Tabela 11 –** Nível de utilização das técnicas de mineração de dados

| Técnica de modelagem | Média | Desvio padrão | Coeficiente de variação (%) |
|---|---|---|---|
| Árvores de decisão | 7,25 | 3,12 | 43,02 |
| Redes neurais | 3,40 | 3,43 | 100,90 |
| Análise discriminante | 5,46 | 3,39 | 62,06 |
| Regressão linear múltipla | 5,79 | 3,41 | 58,87 |
| Regressão logística | 8,69 | 2,33 | 26,83 |
| Análise de componentes principais | 4,61 | 3,27 | 70,81 |
| Análise de sobrevivência | 3,97 | 3,31 | 83,42 |
| Séries temporais | 4,49 | 3,49 | 77,73 |
| Técnicas de segmentação | 8,37 | 1,95 | 23,32 |
| Análise de cestas de mercado | 4,75 | 3,66 | 77,14 |
| Algoritmos de associação | 2,39 | 3,12 | 130,59 |
| Análise exploratória de dados | 9,09 | 1,80 | 19,79 |
| Visualização dos dados (análises visuais dos dados) | 8,51 | 2,47 | 29,04 |
| Mineração de texto (*text mining*) | 3,31 | 3,56 | 107,31 |
| Mineração na internet (*web mining*) | 2,78 | 3,26 | 117,28 |

**Fonte:** Elaborado pelo autor

A Tabela 11 aponta o nível médio de utilização das técnicas de mineração de dados, segundo a avaliação dos 67 respondentes da pesquisa. Ela destaca que a utilização de algumas técnicas parece estar mais uniformemente difundida entre os profissionais que utilizam modelagem de dados para o desenvolvimento de estratégias de relacionamento com clientes. Técnicas como árvores de decisão, regressão logística, técnicas de segmentação, análise exploratória de dados e análises visuais dos dados, apresentaram as médias de utilização mais altas e também os menores níveis de dispersão em torno da média (exceto árvores de decisão),

mostrado pelo coeficiente de variação, todos inferiores a 30%. Mendenhall, Reinmuth e Beaver (1993) definem o coeficiente de variação como uma medida relativa de variabilidade, que é sempre maior ou igual a zero e é expresso em termos relativos. Quanto menor for o coeficiente de variação, menor a variação em torno da média.

Todas as outras técnicas avaliadas apresentaram altos níveis de dispersão e médias de utilização inferiores às das quatro técnicas citadas anteriormente, indicando que seu uso tende a ser maior em algumas empresas e menor em outras. Nenhuma das técnicas avaliadas teve uma baixa média de utilização, acompanhada de baixa variação, o que significa que, mesmo técnicas como mineração de texto e mineração na internet, apresentaram altos níveis de utilização em determinadas empresas.

O teste não-paramétrico de *Friedman* (alternativa ao procedimento paramétrico de análise de variância) pode ser aplicado para verificar se o nível médio de utilização das técnicas de regressão logística, de segmentação, análise exploratória de dados e visualização de dados é estatisticamente maior que o das técnicas restantes. A Tabela 12 mostra os resultados:

**Tabela 12 – Teste de *Friedman* para diferença entre níveis de utilização das técnicas de mineração**

| Comparação | Qui-quadrado | Significância |
|---|---|---|
| Regressão logística x demais* | 220,48 | 0,00 |
| Segmentação x demais* | 198,12 | 0,00 |
| Análise exploratória x demais* | 205,41 | 0,00 |
| Visualização de dados x demais* | 188,11 | 0,00 |
| Árvores de decisão x demais | 145,35 | 0,00 |

**Fonte:** Elaborado pelo autor

* Comparação das técnicas com as demais, exceto com as citadas na tabela

A Tabela 12 mostra que, na comparação das cinco técnicas ali mencionadas com as demais técnicas pesquisadas, há evidências claras de que estas técnicas têm um nível de utilização estatisticamente maior que as demais. Para constatar se existem diferenças entre estas e todas as demais, é possível aplicar o teste de *Wilcoxon*, que compara cada uma das cinco técnicas com cada uma das restantes, duas a duas, de modo a identificar as diferenças significativas. Os resultados são destacados na Tabela 13:

**Tabela 13 – Teste de *Wilcoxon* para diferença de níveis de utilização entre técnicas de mineração de dados**

| Técnica de modelagem | Níveis de significância das comparações | | | | |
|---|---|---|---|---|---|
| | **Árvores de decisão** | **Regressão logística** | **Técnicas de segmentação** | **Análise exploratória de dados** | **Análises visuais dos dados** |
| Redes neurais | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Análise discriminante | 0,01 | 0,00 | 0,00 | 0,00 | 0,00 |
| Regressão linear múltipla | 0,01 | 0,00 | 0,00 | 0,00 | 0,00 |
| Análise de componentes principais | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Análise de sobrevivência | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Séries temporais | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Análise de cestas de mercado | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Algoritmos de associação | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Mineração de texto (*text mining*) | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Mineração na internet (*web mining*) | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |

**Fonte:** Elaborado pelo autor

Os resultados do teste de *Wilcoxon*, expressos na Tabela 13 mostram que há uma diferença estatisticamente significativa entre os níveis de utilização das cinco técnicas (árvores de decisão, regressão logística, técnicas de segmentação, análise exploratória de dados e análises visuais dos dados) e todas as demais técnicas pesquisadas. Uma possível explicação para o destaque apresentado por estas técnicas é sua simplicidade de uso e o poder dos resultados por elas fornecidos, o que torna mais rápida sua utilização para profissionais menos habituados ao uso de ferramentas de análise de dados.

Ainda que as diferenças entre os níveis de utilização das técnicas, observados na amostra, tenham sido expressivas e os testes de *Friedman* e *Wilcoxon* tenham mostrado diferenças estatisticamente significativas, não é possível generalizar estes resultados para a população de interesse, sendo as conclusões aqui obtidas válidas somente para a amostra em questão.

**Objetivo específico 3:** Avaliar o nível de utilização, por parte das empresas de serviços que atuam na cidade de São Paulo, das estratégias de relacionamento com clientes identificadas na literatura.

**Tabela 14 –** Níveis de utilização das estratégias de relacionamento com clientes

| Estratégias de relacionamento com clientes | Média | Desvio padrão | Coeficiente de variação ( %) |
|---|---|---|---|
| Aquisição de novos clientes | 8,28 | 1,88 | 22,71 |
| Identificação dos melhores clientes | 8,09 | 1,82 | 22,54 |
| Valor do cliente no tempo | 6,03 | 2,89 | 47,96 |
| *Cross-selling* | 7,36 | 2,21 | 30,08 |
| *Up-selling* | 7,12 | 2,51 | 35,22 |
| Diferenciação de clientes | 7,63 | 2,11 | 27,66 |
| Reconquista de clientes | 6,07 | 2,62 | 43,11 |
| Fidelização de clientes | 7,21 | 2,03 | 28,22 |
| Salvamento de clientes | 6,63 | 2,60 | 39,21 |

**Fonte:** Elaborado pelo autor

A Tabela 14 mostra que, aparentemente, as empresas têm utilizado de forma mais intensiva as estratégias de aquisição, identificação e diferenciação dos melhores clientes. Estas três estratégias, além de terem apresentado as maiores médias de utilização, também apresentaram os menores níveis de dispersão em torno da média, o que indica que estas práticas têm se apresentado, de certa forma, uniforme, nas empresas. Por outro lado, ela também indica que estratégias de cálculo do valor vitalício do cliente (*lifetime value*) e de reconquista de clientes, foram as que apresentaram os menores níveis médios de utilização entre as estratégias pesquisadas, porém com os maiores níveis de dispersão em torno da média, o que sinaliza que o uso destas estratégias não têm sido prática comum nas empresas observadas na amostra. Brown (2001) classifica as empresas de acordo com o estágio em que se encontram com relação ao desenvolvimento da estratégia de clientes. Observando os resultados da Tabela 14, o foco das empresas, em termos médios, aparentemente, ainda tem sido maior na aquisição de clientes do que em estratégias de retenção como *cross-selling*, *up-selling*, fidelização e salvamento de clientes, o que contraria o posicionamento colocado por autores como Day (2003), Greenberg (2001) e Reichheld (1996), que defendem que a empresa deve focar suas estratégias de relacionamento mais na retenção do que na aquisição

de clientes, em função de ser mais barato manter um cliente existente do que conquistar um novo.

Para verificar se as diferenças observadas entre os níveis de utilização das estratégias podem ser consideradas estatisticamente significativas, é possível aplicar o teste de *Friedman* para comparação do nível de utilização das estratégias de aquisição, identificação e diferenciação de clientes com as demais, conforme mostra a Tabela 15:

**Tabela 15 –** Teste de *Friedman* para diferença entre níveis de utilização das estratégias de relacionamento

| Comparação | Qui-quadrado | Significância |
|---|---|---|
| Estratégia de aquisição x demais* | 59,73 | 0,00 |
| Estratégia de identificação x demais* | 52,70 | 0,00 |
| Estratégia de diferenciação x demais* | 34,63 | 0,00 |

**Fonte:** Elaborado pelo autor

* Comparação das estratégias com as demais, exceto com as citadas na tabela

A Tabela 15 mostra que as três estratégias mais utilizadas (de aquisição, de identificação e de diferenciação de clientes) têm um nível de utilização estatisticamente superior ao das demais estratégias de relacionamento com clientes. Para constatar se existem diferenças entre estas três e todas as demais, é possível aplicar o teste de *Wilcoxon*, que compara cada uma das três estratégias com cada uma das restantes, duas a duas, de modo a identificar as diferenças significativas. Os resultados são destacados na Tabela 16:

**Tabela 16 –** Teste de *Wilcoxon* para diferença entre níveis de utilização das estratégias de relacionamento

| Estratégia de relacionamento | Níveis de significância das comparações | | |
|---|---|---|---|
| | Aquisição de novos clientes | Identificação dos melhores clientes | Diferenciação de clientes |
| Valor do cliente no tempo | 0,00 | 0,00 | 0,00 |
| *Cross-selling* | 0,00 | 0,01 | 0,41 |
| *Up-selling* | 0,00 | 0,00 | 0,09 |
| Reconquista de clientes | 0,00 | 0,00 | 0,00 |
| Fidelização de clientes | 0,00 | 0,00 | 0,13 |
| Salvamento de clientes | 0,00 | 0,00 | 0,00 |

**Fonte:** Elaborado pelo autor

A Tabela 16 indica que as estratégias de aquisição e identificação de clientes possuem um nível de utilização estatisticamente superior às demais estratégias de relacionamento com clientes (exceto entre elas mesmas). A princípio, a estratégia de diferenciação de clientes também parecia apresentar níveis de utilização estatisticamente superiores aos demais, mas isso não foi confirmado após a realização do teste de *Wilcoxon*, que apontou que, do ponto de vista estatístico, não há diferença significativa entre as estratégias de diferenciação de clientes, *cross-selling*, *up-selling* e de fidelização de clientes, considerando um nível de significância de 5%.

**Objetivo específico 4:** Verificar se o nível de utilização das técnicas de mineração de dados para geração de estratégias de relacionamento com clientes tem relação com a existência de um CRM analítico na empresa.

**Tabela 17 –** Comparação entre empresas com e sem áreas de CRM analítico, quanto à utilização de mineração de dados

| Técnica de modelagem | Empresa possui área de CRM analítico ? | | | | |
|---|---|---|---|---|---|
| | Sim | | Não | | Significância |
| | Média | Desvio padrão | Média | Desvio padrão | |
| Árvores de decisão | 7,15 | 3,22 | 7,50 | 2,95 | 0,81 |
| Redes neurais | 3,28 | 3,24 | 3,70 | 3,93 | 0,71 |
| Análise discriminante | 5,60 | 3,27 | 5,15 | 3,72 | 0,76 |
| Regressão linear múltipla | 5,87 | 3,51 | 5,60 | 3,23 | 0,71 |
| Regressão logística | 8,79 | 2,44 | 8,45 | 2,09 | 0,05 |
| Análise de componentes principais | 5,00 | 3,18 | 3,70 | 3,37 | 0,14 |
| Análise de sobrevivência | 4,15 | 3,18 | 3,55 | 3,66 | 0,47 |
| Séries temporais | 4,55 | 3,33 | 4,35 | 3,94 | 0,85 |
| Técnicas de segmentação | 8,40 | 1,84 | 8,30 | 2,25 | 0,87 |
| Análise de cestas de mercado | 4,94 | 3,63 | 4,30 | 3,79 | 0,52 |
| Algoritmos de associação | 2,45 | 3,35 | 2,25 | 2,57 | 0,66 |
| Análise exploratória de dados | 9,19 | 1,50 | 8,85 | 2,39 | 0,45 |
| Visualização dos dados | 8,40 | 2,52 | 8,75 | 2,40 | 0,55 |
| Mineração de texto (*text mining*) | 3,19 | 3,53 | 3,60 | 3,69 | 0,58 |
| Mineração na internet (*web mining*) | 2,87 | 3,33 | 2,55 | 3,14 | 0,92 |

**Fonte:** Elaborado pelo autor

Autores como Parvatiyar e Sheth (2001) defendem a existência de áreas de CRM analítico nas empresas, de forma a utilizar melhor os dados disponíveis sobre os clientes, por meio de modelagem estatística. A Tabela 17 mostra que, visualmente, em 11 das 15 técnicas de mineração de dados pesquisadas, o nível médio de utilização é superior em empresas que possuem áreas de CRM analítico, o que justificaria a opinião dos autores a respeito da função de áreas de CRM analítico.

Estatisticamente, estas diferenças podem ser verificadas por meio da aplicação de um teste não-paramétrico. Neste caso, em função da comparação ser feita entre dois grupos independentes (as pessoas que trabalham em empresas com área de CRM analítico não são as mesmas que trabalham em empresas que não possuem esta área), o teste não-paramétrico a ser aplicado será o de *Mann-Whitney* (comparação de dois grupos independentes). Os resultados são mostrados na coluna "significância" da Tabela 17. Eles mostram que as diferenças observadas, para quase todas as técnicas, exceto para regressão logística, numericamente existem, mas não podem ser consideradas estatisticamente significativas, levando-se em consideração a amostra em questão.

**Objetivo específico 5:** Verificar se o nível de utilização das etapas dos processos de descoberta de conhecimento em bases de dados para geração de estratégias de relacionamento com clientes tem relação com variáveis intrínsecas à empresa, como segmento de atuação na área de serviços, faturamento anual, quantidade de clientes e nacionalidade da empresa.

**Tabela 18 –** Nível de utilização das etapas do processo de KDD, por classes de faturamento da empresa

| **Etapa do KDD** | **Faturamento anual da empresa** | | | | | |
|---|---|---|---|---|---|---|
| | Até R$ 50 milhões | | Entre R$ 50 milhões e R$ 500 milhões | | Acima de R$ 500 milhões | |
| | **Média** | **Desvio padrão** | **Média** | **Desvio padrão** | **Média** | **Desvio padrão** |
| Entendimento do problema de negócio | 8,67 | 0,92 | 8,62 | 1,25 | 8,04 | 1,35 |
| Entendimento dos dados | 9,28 | 0,81 | 8,60 | 1,29 | 8,72 | 1,36 |
| Preparação dos dados | 8,85 | 1,26 | 8,38 | 1,50 | 8,32 | 1,12 |
| Modelagem dos dados | 7,79 | 1,79 | 8,12 | 1,70 | 8,13 | 1,68 |
| Avaliação do modelo | 8,56 | 1,30 | 8,21 | 1,63 | 8,33 | 1,32 |
| Implementação do modelo | 8,29 | 1,06 | 8,47 | 1,56 | 7,68 | 1,54 |
| **Base amostral** | 13 | | 14 | | 40 | |

**Fonte:** Elaborado pelo autor

A tabela 18 mostra que, aparentemente, as empresas que compõem a amostra com faturamento de até R$ 50 milhões ao ano, quando comparadas às empresas com faturamento superior a este valor, possuem níveis mais elevados das etapas de entendimento do problema de negócio, entendimento dos dados relacionados a este problema e preparação dos dados para a etapa de modelagem. De alguma forma, isso poderia ser justificado pelo fato de, em empresas menores, existir uma distância menor entre áreas técnicas e áreas de negócio, o que poderia facilitar a comunicação quanto aos objetivos e problemas de negócio da empresa.

No caso da etapa de modelagem estatística, ocorreu o inverso: Nas empresas com faturamento de até R$ 50 milhões ao ano, aparentemente, a média de utilização das tarefas referentes a esta etapa foi menor do que nas empresas com faturamento superior a este valor. Isto poderia estar associado à cultura de desenvolvimento comparativo de modelos, presentes nas empresas maiores, que trabalham a etapa de modelagem utilizando diversas técnicas e comparando seus resultados, de forma a aperfeiçoar a capacidade de identificação de padrões que, por meio do uso de uma determinada técnica, pode ser mais acurada do que por outra.

Considerando as empresas com faturamento superior a R$ 250 milhões, ocorre uma situação interessante: A média de utilização das etapas de entendimento do problema de negócio e de implementação dos modelos é aparentemente menor do que em empresas menores. Uma possível explicação para isso seria que, em empresas maiores, a pressão por entrega de resultados de curto prazo e acúmulo de atividades tende a ser maior, o que tende a fazer com que os profissionais que desenvolvem os modelos, priorizem tarefas mais técnicas e abreviem determinadas tarefas de planejamento.

Estatisticamente, estas diferenças podem ser testadas por meio da aplicação do teste de *Mann-Whitney*, conforme mostra a Tabela 19:

**Tabela 19 –** Teste de comparação dos níveis de utilização por classe de faturamento

| **Etapa do KDD** | Teste de *Mann-Whitney* | | |
|---|---|---|---|
| | Significância | | |
| | *Grupo 1 x **Grupo 2 | *Grupo 1 x ***Grupo 3 | **Grupo 2 x ***Grupo 3 |
| Entendimento do problema de negócio | 0,79 | 0,13 | 0,13 |
| Entendimento dos dados | 0,16 | 0,25 | 0,60 |
| Preparação dos dados | 0,40 | 0,10 | 0,65 |
| Modelagem dos dados | 0,76 | 0,58 | 0,98 |
| Avaliação do modelo | 0,65 | 0,49 | 0,95 |
| Implementação do modelo | 0,55 | 0,21 | 0,10 |

**Fonte:** Elaborado pelo autor

* Empresas com faturamento inferior a R$ 50 milhões, ** empresas com faturamento entre R$ 50 milhões e R$ 500 milhões, *** empresas com faturamento superior a R$ 500 milhões.

Ainda que visualmente tenham sido identificadas diferenças entre os grupos, conforme mostra a Tabela 19, não foi possível confirmar estatisticamente estas diferenças, muito em função do tamanho da amostra dos grupos e da variação do nível de utilização de cada uma das etapas.

**Tabela 20 –** Nível de utilização das etapas do processo de KDD, por quantidade de clientes da empresa

| **Etapa do KDD** | **Quantidade de clientes da empresa** | | | | | |
|---|---|---|---|---|---|---|
| | Até 50 mil clientes | | Entre 50 mil e 1 milhão de clientes | | Acima de 1 milhão de clientes | |
| | **Média** | **Desvio padrão** | **Média** | **Desvio padrão** | **Média** | **Desvio padrão** |
| Entendimento do problema de negócio | 8,87 | 0,80 | 8,33 | 1,32 | 7,93 | 1,38 |
| Entendimento dos dados | 9,25 | 0,96 | 8,97 | 0,94 | 8,49 | 1,44 |
| Preparação dos dados | 8,98 | 1,13 | 8,03 | 1,47 | 8,26 | 1,12 |
| Modelagem dos dados | 8,42 | 1,54 | 7,58 | 1,96 | 8,03 | 1,66 |
| Avaliação do modelo | 8,77 | 1,23 | 8,19 | 1,23 | 8,17 | 1,47 |
| Implementação do modelo | 8,64 | 0,92 | 8,36 | 1,15 | 7,44 | 1,67 |
| **Base amostral** | 20 | | 12 | | 35 | |

**Fonte:** Elaborado pelo autor

Ao observar a Tabela 20, identifica-se que as empresas com até 50.000 clientes aparentemente apresentaram níveis de utilização mais elevados das etapas de KDD, quando comparadas a empresas com mais de 1.000.000 de clientes. Para verificar se estas diferenças são estatisticamente significativas, a Tabela 21 mostra os resultados do teste de *Mann-Whitney*:

**Tabela 21 –** Teste de comparação dos níveis de utilização por classe de quantidade de clientes

| **Etapa do KDD** | Teste de *Mann-Whitney* | | |
|---|---|---|---|
| | Significância | | |
| | *Grupo 1 x **Grupo 2 | *Grupo 1 x ***Grupo 3 | **Grupo 2 x ***Grupo 3 |
| Entendimento do problema de negócio | 0,24 | **0,01** | 0,57 |
| Entendimento dos dados | 0,43 | **0,05** | 0,40 |
| Preparação dos dados | **0,05** | **0,01** | 0,58 |
| Modelagem dos dados | 0,20 | 0,35 | 0,43 |
| Avaliação do modelo | 0,20 | 0,10 | 0,82 |
| Implementação do modelo | 0,41 | **0,01** | 0,15 |

**Fonte:** Elaborado pelo autor

* Empresas com até 50.000 clientes, ** empresas entre 50 mil e 1 milhão de clientes, *** empresas com mais de 1 milhão de clientes.

A Tabela 21 aponta que as empresas da amostra com menos de 50.000 clientes apresentaram níveis de utilização estatisticamente superiores aos verificados para empresas com mais de 1.000.000 de clientes, exceto nas etapas de modelagem e avaliação de modelos. Estas diferenças podem ser explicadas pelo fato de 60% das empresas de consultoria e de comunicação da amostra (12 das 20 empresas) estarem concentradas no grupo de empresas com menos de 50.000 clientes. As empresas destes segmentos tendem a possuir processos mais estruturados e disciplinados de gerenciamento de projetos e de análise de dados, o que faz com suas médias de utilização das etapas do processo de KDD sejam mais elevadas, quando comparadas a empresas maiores de outros segmentos do setor de serviços.

**Tabela 22 –** Nível de utilização das etapas do processo de KDD, por segmento da empresa

| **Etapa do KDD** | **Segmento da empresa** | | | | | |
|---|---|---|---|---|---|---|
| | Finanças | Telecom | Seguros | Consultoria | Comunicação | Outros |
| | **Média** | | | | | |
| Entendimento do problema de negócio | 7,88 | 7,75 | 8,57 | 9,00 | 8,93 | 8,27 |
| Entendimento dos dados | 8,64 | 7,96 | 9,00 | 9,80 | 9,20 | 8,53 |
| Preparação dos dados | 8,37 | 7,79 | 8,38 | 9,47 | 8,64 | 8,20 |
| Modelagem dos dados | 8,00 | 8,21 | 7,57 | 8,60 | 8,16 | 8,07 |
| Avaliação do modelo | 8,33 | 8,04 | 7,71 | 8,93 | 8,56 | 8,67 |
| Implementação do modelo | 7,81 | 7,34 | 7,37 | 9,00 | 8,44 | 8,14 |
| **Base amostral** | 27 | 8 | 7 | 5 | 15 | 5 |

**Fonte:** Elaborado pelo autor

A Tabela 22 mostra que, aparentemente, os segmentos de consultoria e de comunicação apresentaram as maiores médias de utilização em todas as etapas do processo de descoberta de conhecimento em bases de dados. Isso já era esperado, pois até por conta da exigência de seus clientes, assim como pela própria necessidade de padronização de seus procedimentos, empresas dos segmentos de consultoria e de comunicação tendem a possuir processos de gerenciamento de projetos mais disciplinados e com uma preocupação intensa em superar expectativas de seus clientes (na maior parte, outras empresas). O teste de *Mann-Whitney* pode ser utilizado para comparar o nível médio de utilização das etapas de KDD das empresas de consultoria, comunicação e os demais setores, conforme mostram as Tabelas 23 (consultoria x demais segmentos) e 24 (comunicação x demais segmentos):

**Tabela 23 –** Teste de comparação dos níveis de utilização (consultoria x demais segmentos)

| **Etapa do KDD** | Teste de *Mann-Whitney* | | | |
|---|---|---|---|---|
| | Significância | | | |
| | Finanças | Telecom | Seguros | Outros |
| Entendimento do problema de negócio | 0,14 | **0,04** | 0,34 | 0,22 |
| Entendimento dos dados | 0,14 | **0,01** | 0,11 | **0,02** |
| Preparação dos dados | **0,04** | **0,01** | 0,15 | **0,06** |
| Modelagem dos dados | 0,69 | 0,35 | 0,43 | 0,42 |
| Avaliação do modelo | 0,42 | 0,35 | 0,21 | 0,55 |
| Implementação do modelo | 0,10 | **0,01** | **0,05** | 0,15 |

**Fonte:** Elaborado pelo autor

**Tabela 24 –** Teste de comparação dos níveis de utilização (comunicação x demais segmentos)

| **Etapa do KDD** | Teste de *Mann-Whitney* | | | |
|---|---|---|---|---|
| | Significância | | | |
| | Finanças | Telecom | Seguros | Outros |
| Entendimento do problema de negócio | **0,04** | **0,01** | 0,24 | 0,12 |
| Entendimento dos dados | 0,52 | **0,01** | 0,24 | 0,12 |
| Preparação dos dados | 0,26 | **0,04** | 0,54 | 0,20 |
| Modelagem dos dados | 0,72 | 0,73 | 0,41 | 0,55 |
| Avaliação do modelo | 0,65 | 0,29 | 0,37 | 0,93 |
| Implementação do modelo | 0,38 | 0,08 | 0,08 | 0,55 |

**Fonte:** Elaborado pelo autor

As Tabelas 23 e 24 mostram o resultado do teste de *Mann-Whitney* relativos à comparação entre os níveis de utilização das etapas de KDD para o segmento de consultoria x demais segmentos (Tabela 23) e para o segmento de comunicação x demais segmentos (Tabela 24). Apesar de, aparentemente, as médias de utilização destes dois segmentos terem sido superiores aos demais segmentos, para a maioria dos casos,isto não foi se comprovou após a realização dos testes. Apenas com a comparação com o segmento de seguros, tanto o segmento de consultoria, quanto o de comunicação, apresentaram níveis superiores de utilização nas três primeiras etapas do processo de KDD.

**Tabela 25 –** Nível de utilização das etapas de KDD, por nacionalidade da empresa

| **Etapa do KDD** | **Nacionalidade da empresa** | | | | |
|---|---|---|---|---|---|
| | Brasileira | Americana | Espanhola | Holandesa | Outras |
| | **Média** | | | | |
| Entendimento do problema de negócio | 8,27 | 8,56 | 8,20 | 6,22 | 8,56 |
| Entendimento dos dados | 9,03 | 8,88 | 7,33 | 7,67 | 9,06 |
| Preparação dos dados | 8,43 | 8,56 | 8,13 | 8,33 | 8,33 |
| Modelagem dos dados | 8,02 | 8,14 | 8,60 | 7,11 | 8,11 |
| Avaliação do modelo | 8,20 | 8,54 | 8,13 | 9,00 | 8,50 |
| Implementação do modelo | 7,89 | 8,11 | 7,49 | 8,24 | 8,14 |
| **Base amostral** | 34 | 19 | 5 | 3 | 6 |

**Fonte:** Elaborado pelo autor

A Tabela 25 mostra que, aparentemente, empresas americanas e brasileiras apresentaram médias mais elevadas de utilização das etapas de entendimento do problema de negócio, entendimento e preparação dos dados, enquanto que empresas espanholas apresentaram a maior média de utilização na etapa de modelagem e empresas holandesas as maiores médias de utilização nas etapas de avaliação e implementação. Em função da baixa quantidade amostral de empresas holandesas, estes padrões verificados podem estar distorcidos, pois, por exemplo, não parece fazer sentido uma empresa valorizar mais as etapas de avaliação e implementação do modelo, do que etapas de planejamento construção do projeto de mineração de dados.

O teste de *Kruskal-Wallis* pode ser utilizado para a comparação de todos os grupos entre si, para a identificação de diferenças significativas, quanto à utilização das etapas do processo de KDD. A Tabela 26 mostra os resultados destas comparações:

**Tabela 26 –** Teste de comparação dos níveis de utilização, por nacionalidade da empresa

| **Etapa do KDD** | Teste de *Kruskal-Wallis* | |
|---|---|---|
| | Qui-quadrado | Significância |
| Entendimento do problema de negócio | 5,36 | 0,25 |
| Entendimento dos dados | 6,66 | 0,16 |
| Preparação dos dados | 1,06 | 0,90 |
| Modelagem dos dados | 2,44 | 0,67 |
| Avaliação do modelo | 1,44 | 0,84 |
| Implementação do modelo | 1,85 | 0,73 |

**Fonte:** Elaborado pelo autor

Os resultados expressos na Tabela 26 mostram que não foram detectadas diferenças estatisticamente significativas entre os níveis de utilização das etapas do processo de KDD para empresas de diferentes nacionalidades. Isto mostra que, apesar de o estágio da pesquisa sobre mineração de dados, bem como sua utilização em empresas ser mais avançado em países desenvolvidos, as empresas brasileiras têm evoluído e, baseado nos dados da amostra coletada, não é possível afirmar, considerando um nível de significância de 5%, que o nível de utilização de etapas do processo de KDD seja mais intenso em empresas multinacionais do que em empresas brasileiras.

## 4.3 Análise do nível de confiabilidade das respostas

**Tabela 27 –** Nível de consistência interna das etapas do processo de KDD, segundo nível de utilização e importância das etapas

| Itens de análise | Etapa do KDD | Nível de utilização | | Nível de importância | |
|---|---|---|---|---|---|
| | | Quantidade de itens | Alpha de Cronbach | Quantidade de itens | Alpha de Cronbach |
| 1 a 3 | Entendimento do problema de negócio | 3 | 0,71 | 3 | 0,77 |
| 4 a 6 | Entendimento dos dados | 3 | 0,84 | 3 | 0,90 |
| 7 a 9 | Preparação dos dados | 3 | 0,70 | 3 | 0,74 |
| 10 a 12 | Modelagem dos dados | 3 | 0,74 | 3 | 0,82 |
| 13 a 15 | Avaliação do modelo | 3 | 0,79 | 3 | 0,83 |
| 16 a 22 | Implementação do modelo | 7 | 0,88 | 7 | 0,90 |

**Fonte:** Elaborado pelo autor

Para verificação da consistência interna das medidas avaliadas no questionário (nível de utilização e de importância de cada uma das etapas do processo de descoberta de conhecimento em bases de dados), foi calculado, por meio do SPSS 13.0, o índice *alpha de Cronbach,* cujos valores possíveis variam entre 0 e 1. Malhotra (2001) considera que um *alpha de Cronbach* acima de 0,6 já revela um nível satisfatório de consistência interna para as medidas utilizadas, destacando que quanto mais próximo de 1, maior é a confiabilidade dos resultados da aplicação do teste. Considerando a resposta fornecida pelos 67 respondentes a cada um dos 22 itens pesquisados, todas as seis etapas do processo de KDD tiveram *alpha de Cronbach* igual ou superior a 0,7 (conforme mostra a Tabela 27) denotando um nível satisfatório de confiabilidade dos itens mensurados.

# 5. CONCLUSÕES, LIMITAÇÕES E PROPOSTAS DE ESTUDOS FUTUROS

## 5.1 Limitações do estudo

Há que se destacar nesta seção que não houve possibilidade de se desenvolver um planejamento amostral adequado para que os resultados obtidos a partir das análises pudessem ser mais conclusivos e generalizáveis. Algumas limitações cercam esta dissertação. A primeira questão é justamente a amostral: A amostragem utilizada para coleta de dados foi não- probabilística, tipo “bola de neve”, o que impossibilitou que os resultados obtidos pudessem ser expandidos para o universo de interesse, empresas do setor de serviços que atuam nas cidades de São Paulo e Rio de Janeiro e que utilizam bases de dados para o desenvolvimento de estratégias de relacionamento com clientes.

A segunda limitação, diretamente relacionada à primeira, foi a concentração da amostra em basicamente quatro segmentos: finanças, telecomunicações, seguros e comunicação.Ainda que pese o fato destes segmentos representarem também a maior fatia de aplicações de mineração de dados observadas na literatura, não foi possível afirmar se um estudo baseado em amostragem probabilística obteria esta mesma concentração de segmentos e de níveis de faturamento observados no estudo.

A terceira limitação considerada foi que as medidas pesquisadas no questionário referiram-se a percepções dos profissionais que trabalham com modelagem de dados. Estas percepções tiveram um componente subjetivo importante, que pode, de alguma forma, ter distorcido os resultados obtidos. Inclusive, as percepções coletadas podem ter sido influenciadas pelo tempo de entrega do projeto, política corporativa e características pessoais no trato com a informação e com  a análise de dados.

A quarta limitação observada foi, de certa forma, decorrente do modo como foi realizado o planejamento amostral da pesquisa. Os resultados apresentados refletiram as percepções e informações fornecidas pelos 67 profissionais que aceitaram responder à pesquisa. Um determinado número de potenciais respondentes, 22 ao todo, não aceitou responder a pesquisa ou não respondeu após cinco contatos (um contato a cada semana) via e-mail. Estes que não responderam poderiam ter apresentado percepções diversas das coletadas, o que poderia ter enriquecido ainda mais os resultados apresentados.

## 5.2 Conclusões do estudo

O objetivo geral do estudo foi o de verificar como as empresas do setor de serviços utilizam bases de dados para subsidiar a criação de estratégias de relacionamento com clientes. Foi possível constatar que as empresas pesquisadas fazem uso de um processo que tem como *input* um determinado problema de negócio que a empresa precisa solucionar e um *output* que é o conhecimento necessário para a construção de uma estratégia de relacionamento com clientes. Constatou-se também que, de uma maneira geral, as empresas pesquisadas estão aderentes com relação à utilização de processos de descoberta de conhecimento em bases de dados (KDD) identificados na literatura, ou seja, tais processos fazem parte da realidade das empresas participantes do estudo.

As empresas do setor de serviços pesquisadas estiveram concentradas basicamente nos setores de finanças, seguros, telecomunicações e comunicação, que representaram mais de 85% do total de 67 respondentes. Essa concentração está em conformidade com o que foi observado na literatura, pois grande parte das aplicações concentraram-se nestes segmentos do setor de serviços. Outra característica importante foi que a maior parte dos respondentes declararam trabalham em empresas de grande porte, com faturamento anual superior a R$ 250 milhões. Houve dificuldade em encontrar pequenas empresas (exceto consultorias) que tivessem como prática a utilização de processos de análise de dados para subsidiar a criação de estratégias de relacionamento com clientes. Por um lado, isso também está de acordo com a literatura que descreve que a mineração de dados faz mais sentido quando há grandes volumes de dados, já que a maioria dos algoritmos de mineração de dados exige grandes volumes para construir e treinar modelos que serão utilizados para realizar tarefas de classificação, predição ou estimação (BERRY e LINOFF, 2004). Por outro lado, percebe-se que o que restringe o uso de processos de análise de dados em uma empresa não é somente o tamanho das bases de dados ou a carteira de clientes, mas também a falta de uma cultura analítica (DAVENPORT, 2006), que leve a empresa, mesmo sendo pequena, a refletir acerca da importância de se armazenar dados sobre clientes e de utilizar estes dados para tornar racionais os processos de tomada de decisão, no que se refere ao modo de se relacionar com o cliente e de oferecer produtos e serviços ao mesmo.

Estudou-se também a comparação entre níveis de utilização x importância das etapas do processo de KDD. A percepção dos profissionais que responderam à pesquisa mostrou que eles acreditam que a empresa atribui um maior nível de importância (do que eles realmente utilizam na prática) à etapa de entendimento de seu problema de negócio, enquanto, na etapa

de modelagem, a percepção dos respondentes é de que o nível de utilização destas etapas é superior ao nível de importância que eles consideram que a empresa atribui a elas. Outro ponto importante, apesar de não ter sido considerado estatisticamente diferente do nível de utilização, foi o nível de importância atribuído a etapas de planejamento associando recursos da área técnica x área de negócio (itens 2, 8, 15 e 16 da Tabela 9). Para estes itens, os respondentes consideraram que a empresa dá mais importância a eles, do que eles de fato utilizam na prática.

No que se refere ao uso de técnicas de mineração para gerar conhecimento existente em bases de dados, a literatura apresentou uma quantidade bastante extensa de algoritmos e técnicas que, em sua maioria, não tem sido usadas com intensidade pelas empresas de serviço pesquisadas. Grande parte destas empresas têm concentrado mais atenção a um grupo menor de técnicas, notadamente regressão logística, análise de agrupamentos, análise exploratória de dados e visualização de dados, técnicas que apresentaram níveis médios de utilização significativamente superiores ao das outras técnicas pesquisadas. Outro ponto é que o item referente à comparação de resultados por mais de uma técnica não pareceu ser considerada uma atividade realizada de forma uniforme nas empresas, de forma que algumas empresas a realizam com mais intensidade e outras não.

Um outro ponto importante é que, considerando as empresas da amostra, aquelas que possuem áreas de CRM analítico apresentaram níveis de utilização de técnicas de mineração muito próximos aos verificados em empresas sem áreas de CRM analítico. Descritivamente, em 11 das 15 técnicas de mineração avaliadas, o nível médio de utilização foi maior em empresas com áreas de CRM analítico, porém, talvez em função do tamanho da amostra, estas diferenças não foram consideradas estatisticamente significativas. Mesmo assim, as diferenças observadas quanto ao nível de utilização das técnicas de mineração de dados entre empresas com e sem áreas de CRM analítico, reforçam a opinião de Parvatiyar e Sheth (2001), que defendem a existência de áreas de CRM analítico, por terem foco no uso de técnicas para melhor analisar as informações sobre clientes.

Quanto às estratégias de relacionamento utilizadas pelas empresas, a amostra coletada mostrou uma característica divergente daquela destacada na literatura a respeito de estratégias de relacionamento. Autores como Reichheld (1996), Greenberg (2001) e Day (2003) defendem que as empresas devem focar seus esforços em estratégias de retenção de clientes, por ser mais barato, para a empresa, reter os melhores clientes, do que adquirir novos. O comportamento observado na amostra de empresas coletadas mostrou que as empresas ainda realizam com mais intensidade estratégias de aquisição do que estratégias de retenção como

*cross-selling*, *up-selling*, reconquista e salvamento de clientes. Isto mostra que pode ser necessária uma evolução no que se refere à capacidade das empresas em utilizar estratégias de relacionamento com clientes, considerando cada cliente ou grupo de clientes de forma diferente e estabelecendo formas diferenciadas de se relacionar com os mesmos, com o objetivo de manter na carteira os clientes mais rentáveis (mais participação no cliente) e adquirindo novos clientes com potencial de gerarem receitas futuras.

Um padrão interessante encontrado, ainda que não tenha sido estatisticamente diferente dos demais segmentos, foi o nível de utilização das etapas do processo de KDD verificado nos segmentos de consultoria e de comunicação. Estes segmentos, em sua maioria, foram compostos por empresas que prestam serviços a outras empresas, de maneira que, sobretudo consultorias, aprenderam a disciplinar seus processos de análise, dando ênfase a cada detalhe de um projeto de mineração de dados, sendo este tipo de empresa que o mais se aproxima do nível de excelência registrado na literatura a respeito de processos de descoberta de conhecimento em bases de dados (FAYYAD *et al.*, 1996).

A conclusão final é que as empresas do setor de serviços, sobretudo as grandes empresas, utilizam processos e ferramentas sofisticados de análise de bases de dados, de forma a extrair delas o conhecimento necessário para o desenvolvimento de novas formas de se relacionar com seus clientes e de realizar ofertas de produtos e serviços que estejam em linha com suas necessidades. Este processo de análise é amplamente segmentado em tarefas menores, desde o entendimento do problema a ser resolvido, passando pelo levantamento das variáveis e bases necessárias, até a modelagem e implementação do modelo. De um ponto de vista geral da etapa de modelagem de dados, há uma concentração importante no uso de algumas técnicas e relativamente pouca ênfase na comparação dos resultados dos modelos por mais de uma técnica. Com relação às estratégias de relacionamento utilizadas, as empresas pesquisadas, de uma forma geral, ainda têm atribuído mais ênfase a estratégias de aquisição do que a estratégias de retenção, sendo este um ponto de evolução na utilização de estratégias de relacionamento com clientes.

## 5.3 Proposta de estudos futuros

A primeira sugestão a ser feita com relação a estudos futuros diz respeito aos procedimentos de amostragem. Quaisquer conclusões não podem ser cientificamente embasadas sem uma amostra probabilística e com um tamanho adequado, de forma que represente o universo a ser estudado. Assim, para que o estudo seja robusto com relação a seus resultados e conclusões, a sugestão é a realização de um planejamento de uma amostra probabilística e representativa do universo a ser estudado.

Uma segunda sugestão é realizar um estudo similar para outros segmentos, expandindo o escopo de atuação, considerando principalmente empresas do setor da indústria e do varejo, em que também são identificadas inúmeras aplicações de mineração de dados para o desenvolvimento de estratégias de relacionamento.

Uma terceira sugestão é estudar a influência da utilização de múltiplas técnicas de mineração de dados no resultado de estratégias de relacionamento com clientes, de forma a avaliar se a combinação do resultado fornecido por diversas técnicas, em vez da utilização do resultado isolado de uma delas, faz alguma diferença para o resultado do negócio e/ou se é a forma como as estratégias são priorizadas (aquisição e retenção) e implementadas que tornam o resultado mais ou menos significante.

Por fim, uma última sugestão é a de estudar o impacto que técnicas de análise ainda emergentes na área de marketing como algoritmos genéticos, lógica difusa, cadeias de markov e redes bayesianas, possuem para a melhoria do resultado de uma estratégia de relacionamento com clientes.

# 6. REFERÊNCIAS

AGRAWAL, R.; SRIKANT, R. Fast algorithms for mining association rules. **Proceedings of the 20th International Conference on Very Large Databases.** P. 487–499. Santiago, 1994.

AIJO, T.S. The theoretical and philosophical underpinnings of relationship marketing. **European Journal of Marketing**. Vol. 30, n.2, p.8-18,1996.

ALAVI, M.; LEIDNER, D.E. Review; knowledge management and knowledge management systems: conceptual foundations and research issues. **MIS Quarterly**, Vol. 25, n. 1 p. 107-136, 2001.

AMA - **American Marketing Association**. [*online*]. Disponível em: <http://www.marketingpower.com/content4620.php>. Acesso em: 22/03/2007.

AHN, J.Y.; KIM, S.K.; HAN, K.S. On the design concepts for CRM systems. Industrial **Management & Data Systems**. Vol. 103, n. 5, p. 324 – 331, 2003.

ANTUNES, M.T.P.; KATO, H.T.; CORRAR, L.J. **A eficiência das informações divulgadas em "melhores & maiores" da revista exame para a previsão de desempenho das empresas**. Trabalho apresentado no XXVI ENANPAD – 26º Encontro da Associação Nacional de Pós-graduação e Pesquisa em Administração, 2002.

BARNEY, J. **Gaining and sustaining competitive advantage**. New Jersey: Prentice- Hall, 2002.

BERSON, A.; SMITH, S.; THEARLING, K. **Building data mining applications for crm**. London: McGraw-Hill, 1999.

BERRY, M. J. A.; LINOFF, G.S. **Mastering data mining**. Indianopolis: Wiley Publishing, 2000.

BERRY, M. J. A.; LINOFF, G.S. **Data mining techniques for marketing, sales, and customer relationship management**. Indianopolis: Wiley Publishing, 2004.

BERRY, L. L. Relationship marketing. In: BERRY, L. L.; SHOSTACK, G. L.; UPAH, G. D. **Emerging perspectives on services marketing**. Chicago: American Marketing Association, p. 25-28, 1983.

BOZDOGAN, H. **Statistical data mining and knowledge discovery**. Boca Raton: Chapman & Hall/CRC, 2004.

BRACHMAN, R.J.; ANAND, T. The process of knowledge discovery in databases: A human-centered approach. In FAYYAD, U.M., PIATESTSKY-SHAPIRO, G., SMYTH, P., & UTHURASAMY, R. **Advances in knowledge discovery and data mining**. Cambridge: AAAI /MIT Press, 1996.

BREIMAN, L.; *FRIEDMAN*, J. H. ; OLSHEN, R.A.; STONE, C. J. **Classification and regression trees**. Belmont: Wadsworth Statistical Press, 1984.

BROCKWELL, P.J.; DAVIS, R.A. **Introduction to time series and forecasting**. New York: Springer-Verlag, 2002.

BROWN, S. **CRM – Customer relationship management – uma ferramenta estratégica para o mundo do e-business**. São Paulo: Makron Books do Brasil, 2001.

CABENA, P. ; HADJINIAN, P. ; STADLER, R. ; VERHEES, J. ; ZANASI, A. **Discovering data mining: from concept to implementation**. New Jersey: Prentice Hall, 1998.

CARPENTER, E.O.; LACHTERMACHER, G. **Determinação dos fatores críticos na análise de desempenho de alunos de pós-graduação utilizando metodologia de mineração de dados**. Trabalho apresentado no XXIX ENANPAD – 29º Encontro da Associação Nacional de Pós-graduação e Pesquisa em Administração, 2005.

CARVALHO, R.B. **Tecnologia da informação aplicada à gestão do conhecimento**. Belo Horizonte: Com Arte, 2003.

CAVANA, R Y.; DELAHAYE, B. L.;SEKARAN, U. **Applied business research – qualitative and quantitative methods**. New York: John Wiley & Sons, 2000.

CHAKRABARTI, S. **Mining the web: discovering knowledge from hypertext data**. San Francisco: Morgan Kaufmann Publishers, 2003.

CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHABAZA, T., REINARTZ, T., SHEARER, C.;WIRTH, R. **CRISP-DM 1.0 - Step-by-step data mining guide**, CRISP-DM Consortium, 1999. Disponível em <http://www.crisp-dm.org>. Acesso em 10/04/2007, 17:39:15

CHYE, K.H.; GERRY, C. K. L. Data mining and customer relationship marketing in the banking industry. **Singapore Management Review**. Vol. 24, n. 2, p. 1-27, 2002.

COCHRAN, W. G. **Sampling techniques**. New York: John Wiley & Sons, 1977.

COOPER, D.R.; SCHINDLER, P.S. **Métodos de pesquisa em administração**. Porto Alegre: Bookman, 2003.

DAVENPORT, T. H. Competing on analytics. **Harvard Business Review**. Vol. 84, n. 1, p. 98-107, 2006.

DAVENPORT, T. H.; HARRIS, J.G.; KOHLI, A.K. How do they know their customers so well ? **MIT Sloan Management Review**. Vol. 42, n. 2, p. 63-73, 2001.

DAVIDSON, I.; SOUKUP, T. **Visual data mining: techniques and tools for data visualization and mining**. New York :Wiley Publishing, 2002.

DAY, G. S. Creating a superior customer-relating capability. **MIT Sloan Management Review**. Vol. 44. n. 3, p. 77-82, 2003.

DAY, G. S. **A empresa orientada para o mercado: compreender, atrair e manter clientes valiosos**. Porto Alegre: Bookman, 2001.

DREW, J. H. ; MANI, D. R. ; BETZ, A. L. ; DATTA, P. Targeting customers with statistical and data mining techniques. **Journal of Service Research**. Vol. 3, n. 3, p. 205-219, 2001.

ECKERSON, W.; WATSON, H. Harnessing customer information for strategic advantage: technical challenges and business solutions. **Industry Study**, The Datawarehousing Institute, Seattle, WA, p. 6, 2001.

EDELSTEIN, H., **Introduction to Data Mining and Knowledge Discovery**. Edelstein Corporation, Potomac, MD USA, 1999.

ETZEL, M. J. ; WALKER, B. J. ; STANTON, W. J. **Marketing**. São Paulo: Makron Books, 2001.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**. Vol. 17, n. 3, p. 37-54, 1996.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHURASAMY, R. **Advances in knowledge discovery and data mining**. Cambridge: AAAI/MIT Press, 1996.

FRANCISCO, E.R.; PETRIELLI, A.; REINA, C.S. Segmentação comportamental de clientes para o setor elétrico. **Congresso Anual de Tecnologia de Informação**. Fundação Getúlio Vargas. Escola de Admnistração de Empresas de São Paulo, 2006.

FULLER, T.; LEWIS, J. “Relationships” mean everything.; A typology of small-business relationship strategies in a reflexive context. **British Journal of Management**. Vol. 13, n. 4, p. 317-336, 2002.

GEHRKE, J. Methodologies of data mining. In: **The handbook of data mining**. New Jersey: Lawrence Erlbaum Associates, 2003.

GILLIES, C.; RIGBY, D.;REICHHELD, F. The story behind successful customer relations management. **European Business Journal**. Vol. 14. n. 2, p. 73, 2002.

GREENBERG, P. **CRM at the speed of light: capturing and keeping customers in internet real time**. Berkeley: Osborne/McGraw-Hill, 2001.

GRÖNROOS, C. Relationship marketing: Strategic and tactical implications. **Management Decision**. Vol. 34, n. 3, p. 5-14, 1996.

GUMMESSON, E. **Total relationship marketing: marketing management, relationship strategy and CRM approaches for the network economy**. 2. ed. London: Butterworth-Heinemann, 2002.

HAIR, J,F.; ANDERSON, R.E.; TATHAM, R. L.; BLACK, W.C. **Análise multivariada de dados**. Porto Alegre: Bookman, 2005.

HAN, J.; KAMBER, M. **Data mining: concepts and techniques**. San Francisco: Morgan Kaufmann Publishers, 2000.

HAND, D., MANNILA ,H.;SMYTH, P. **Principles of data mining: adaptive computation and machine learning**. Cambridge: MIT Press, 2001.

HARRINGTON, H. J.. **Aperfeiçoando processos empresariais**. São Paulo: Makron Books, 1993.

HOGG, R.V.; TANIS, E. **Probability and Statistical Inference**. New York: Prentice-Hall, 1997

HORMAZI, A. M. ; GILES, S. Data mining: a competitive weapon for banking and retail industries. **Information Systems Management**. Vol. 21, n. 2, p. 62-71, 2004.

HOSMER, D.W.; LEMESHOW, S. Applied survival analysis: **Regression modeling of time to event data**. New York: John Wiley & Sons, 1999.

HUI, S.C.; JHA, G. Data mining for customer service support. **Information and Management**. Vol. 38, n. 1, p. 1-13, 2000.

JACKSON, B. B. **Winning and keeping industrial customers: the dynamics of customer relationships**. Lexington: D.C. Heath, 1985.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. New Jersey: Prentice Hall, 2002.

KAMAKURA, W. ; MELA, C. ; ANSARI, A. ; BODAPATI, A. ; FADER, P. ; IYENGAR, R. ; NAIK, P. ; NESLIN, S. ; SUN, B. ; VERHOEF, P. C. ; WEDEL, M. ; WILCOX, R. Choice models and customer relationship management. **Marketing Letters**. Vol. 16, n. 3/4 , p. 279-291, 2005.

KAMAKURA, W. **Tendências da pesquisa acadêmica na área de marketing**. Palestra realizada no XXVI ENANPAD – 26º Encontro da Associação Nacional de Pós-graduação e Pesquisa em Administração, junho/2002.

KANTARDZIC, M. **Data mining: concepts, models, methods and algorithms**. New York: Wiley-Interscience, 2003.

KEEFE, L.M. What is the meaning of marketing ? **Marketing News**. Vol. 38. n. 15, p. 17-18, 2004.

KOGUT, B.; ZANDER, U. Knowledge of the firm, combinative capabilities, and the replication of technology. **Organization Science**, Vol. 3, n. 3, p. 383-397, 1992.

KOTLER, P.; ARMSTRONG, G. **Princípios de marketing**. São Paulo: Prentice Hall, 2003.

KOTLER, P. **Administração de marketing: a edição do novo milênio**. São Paulo: Prentice Hall, 2000.

KOTOROV, R. Customer relationship management: strategic lessons and future directions. **Business Process Management Journal**. Vol. 9, n. 5, p. 566-571, 2003.

LAU, K. ; LEE, K. ; HO, Y. ; LAM, P. Mining the web for business intelligence: Homepage analysis in the internet era. **Journal of Database Marketing and Customer Strategy Management**. Vol. 12, n. 1, p. 32-54, 2004.

LEITE, M.M. **Pressupostos para implantação de estratégias de relacionamento com os clientes em pequenas e médias organizações**: **uma abordagem baseada em gerenciamento de projetos**, 2003, 324 p. Tese (Doutorado em Engenharia de Produção). Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina. Florianópolis.

LIEBOWITZ, J.; BECKMAN, T. J. **Knowledge organizations: what every manager should know**. Boca Raton: St Lucie Press, 1998.

LIM, T.S.; LOH, W.Y.; SHIH, Y.S. A comparison of prediction accuracy, complexity and training time of thirty-three old and new classification algorithms. **Machine Learning**. Vol. 48, n. 2, p. 203–228, 2000.

LOVELOCK, C. ; WRIGHT., L. **Serviços: marketing e gestão**. São Paulo: Saraiva, 2006.

MADEIRA, S.C.; OLIVEIRA, A.L. A data mining approach to risk assessment in telecommunications. **Gauss 2003: 2° encontro do grupo acadêmico de utilizadores SAS**. Lisboa, 2003.

MALHOTRA, N. K. **Pesquisa de marketing: Uma orientação aplicada**. Porto Alegre: Bookman, 2001.

MCKENNA. R. **Marketing de relacionamento: estratégias bem-sucedidas para a era do cliente**. Rio de Janeiro: Campus, 1992.

MENA, J. ; PETTIT, R. Web mining case study: An internet radio website. **Interactive Marketing**. Vol. 3, n. 1, p. 46-52, 2001.

MENDENHALL, W.; REINMUTH, J.E.; BEAVER, R.J. **Statistics for management and economics.** Belmont: Duxbury Press, 1993.

MIN, H. ; MIN, H. ; EMAN, A. A data mining approach for developing the profile of hotel customers. **International Journal of Contemporary Hospitality Management**. Vol. 14, n. 6, p. 274-285, 2002.

MONTGOMERY, C. D.; PECK, A. E. ; VINING, C. G. **Introduction to linear regression analysis**. New York: Wiley-Interscience, 2001.

MORGAN, R. M.; HUNT, S. D. The commitment–trust theory of relationship marketing. **Journal of Marketing**, v.58, n. 3, p. 20-38, 1994.

NONAKA, I.; TAKEUCHI, H. **Criação de conhecimento na empresa**. Rio de Janeiro: Campus, 1997.

OLIVER, R.L. **Satisfaction: a behavioral perspective on the consumer**. New York, McGraw Hill, 1997.

PARVATIYAR, A.; SHETH, J. N. Customer relationship management: emerging practice, process and discipline. **Journal of Economic and Social Research**. Vol. 3, n. 2, p. 1-34, 2001.

PEACOCK, P. R. Data mining in marketing: part 1. **Marketing Management**. Vol. 6, n. 4, p. 8-18, 1998.

PEPPERS, D.; ROGERS, M. **CRM series – Marketing 1 to 1**. Peppers and Rogers Group do Brasil, 2004.

PINHEIRO, C.A.R. **Redes neurais para a prevenção de inadimplência em operadoras de telefonia**, 2005, 245 p. Tese (Doutorado em Engenharia). Programa de Pós-Graduação em Engenharia, Universidade Federal do Rio de Janeiro. Rio de Janeiro.

PYLE, D. Management of data mining. In: YE, N. **The handbook of data mining**. New Jersey: Lawrence Erlbaum Associates, 2003.

REICHHELD, F. **The loyalty effect: The hidden force behind growth, profits and lasting value**. Boston: Harvard Business School Press, 1996.

REJESUS, R.M.; LITTLE, B.B.; LOVELL, A.C. Using data mining to detect crop insurance fraud: Is there a role for social scientists ? **Journal of Financial Crime**. Vol. 12, n.1, p.24, 2004.

RICHARDSON, R. J. **Pesquisa social: métodos e técnicas**.São Paulo : Atlas, 1999.

ROCACK, L.; MAIMOM, O. Data mining for improving the quality of manufacturing: A feature set decomposition approach. **Journal of Intelligent Manufacturing**. Vol. 17, n.3, p. 285, 2006.

RODRIGUEZ, M. T.; ÁLVAREZ, J. V.; MESA, J. M.; GONZÁLEZ, A. Metodologías para la realización de proyectos de data mining. **CD del VII Congreso Internacional de Ingeniería de Proyectos**. Pamplona AEIPRO, 2003

ROWLEY, J. Reflections on customer knowledge management in e-business. **Qualitative Market Research**. Vol. 5, n. 4. p. 268-280, 2002.

RYALS, L. Creating profitable customers through the magic of data mining. **Journal of Targeting, Measurement and Analysis for Marketing**. Vol. 11, n. 4, p. 343-349, 2003.

RYALS, L.; PAYNE A. Customer relationship management in financial services: towards information-enabled relationship marketing. **Journal of Strategic Marketing**. Vol. 9, n. 1, p. 3-27, 2001.

SAMLI, A.C.; POHLEN, T. L.; BOZOVIC, N. A review of data mining techniques as they apply to marketing: Generating strategic information to develop market segments. **The Marketing Review**. Vol. 3, n.2, p. 211-227, 2002.

SEKARAN, U. **Research methods for business: a skill building approach**. New York: John Wiley & Sons, 1992.

SHEARER, C. The CRISP-DM model: The new blueprint for data mining. **Journal of Datawarehousing**. Vol. 5, n. 4, p. 13-22, 2000.

SIEGEL, S. **Estatística não-paramétrica para as ciências do comportamento**. São Paulo:McGraw-Hill do Brasil, 1975.

SILVA, W., LADEIRA, M. Mineração de Dados em Redes Bayesianas. **Congresso da Sociedade Brasileira de Computação**:19° Jornada de Atualização em Informática. Rio de Janeiro, 2002..

SIN, L. Y. M.; TSE, A. C. B.; YIM, F. H. K. CRM: Conceptualization and scale development. **European Journal of Marketing**. Vol. 39, n. 11/12, p. 1264-1290, 2005.

SMITH, K. A.; WILLIS, R. J.; BROOKS, M. An analysis of customer retention and insurance claim patterns using data mining: a case study. **Journal of the Operational Research Society**. Vol. 51, n. 5, p. 532-541, 2000.

SOLOMON, S.; NGUYEN, H.; LIEBOWITZ, J.; AGRESTI, W. Using data mining to improve traffic safety programs. **Industrial Management + Data Systems**. Vol. 106, n. 5, p. 621, 2006.

STOLZER, A.J.; HALFORD, C. Data mining methods applied to flight operations quality assurance data: A comparison to standard statistical methods. **Journal of Air Transportation**. Vol. 1, n.1, p.6, 2007.

SWIFT, R. **CRM – Customer relationship management – O revolucionário marketing de relacionamento com o Cliente**. Rio de Janeiro: Campus, 2001.

TAPSCOTT, Don. Make knowledge an asset for the whole company. **Computerworld.** Vol. 32, n. 51, p. 32, 1998.

VALENTE, T.R.G. **Marketing de relacionamento e CRM: Uma análise da gestão de clientes no setor financeiro**, 2002, 190 p. Trabalho de conclusão de curso. Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo. São Paulo.

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

VICENTE, C.R. **Gerenciamento do relacionamento com o cliente em insitituição científica e tecnológica para melhoria da interação com a indústria**, 2005, 192 p. Dissertação (Mestrado em Engenharia de Produção). Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina. Florianópolis.

WEBB, G. I. Association rules. In: **The handbook of data mining**. New Jersey: Lawrence Erlbaum Associates, 2003.

WITTEN, I.H.; FRANK, E. **Data mining: Practical machine learning tools and techniques**. San Francisco: Morgan Kaufmann Publishers, 2005.

YAU, O.; LEE, J.; CHOW, R. ; SIN, L. ; TSE, A. Relationship marketing: the chinese way. **Business Horizon**. Vol. 43, n. 1, p. 16-24, 2000.

XU, M.; WALTON, J. Gaining customer knowledge through analytical CRM. **Industrial Management + Data Systems**. Vol. 105, n. 7, p. 955-971, 2005.

## Apêndice A- Pesquisa sobre uso de modelagem de dados

O propósito deste questionário é avaliar como empresas do setor de serviços nas cidades de São Paulo e do Rio de Janeiro utilizam bases de dados para criar conhecimento sobre seus clientes e dar suporte à criação de estratégias de relacionamento.O preenchimento deve ser feito de acordo com sua percepção sobre a realidade da empresa em que trabalha.

Todos os dados coletados nesta pesquisa serão mantidos em absoluto sigilo, sendo tratados estatisticamente e apresentados percentualmente dentro do contexto do estudo, servindo de base para a elaboração de uma dissertação de mestrado. Faremos absoluta questão de divulgar os resultados deste estudo para todos aqueles que se interessarem, com as devidas análises e conclusões.

Gostaríamos de ressaltar a importância e a necessidade de sua contribuição para a consistência das conclusões dos estudos.

Desde já, agradecemos sua colaboração:

Marcelo Pires Fernandes
Mestrando em Administração de Empresas
Rua da Consolação, 896 – 3º andar
São Paulo – SP
Tel. (011) 8304-5680

Prof. Dr. Silvio Popadiuk
Orientador
Rua da Consolação, 896 – 2º andar
São Paulo – SP
Tel. (011) 2114-8275

**01 – Em qual segmento do setor de serviços atua a empresa em que você trabalha ?**

( ) Serviços financeiros (bancos e financeiras)

( ) Telecomunicações

( ) Seguros

( ) Hotelaria

( ) Consultoria

( ) Saúde

( ) Informática

( ) Outros Informar setor:_____

**02 – Qual foi faturamento em 2006 da empresa em que você trabalha ?**

( ) Menos de R$ 1 milhão

( ) Entre R$ 1 milhão e 5 milhões

( ) Entre R$ 5 milhões e R$ 10 milhões

( ) Entre R$ 10 milhões e R$ 25 milhões

( ) Entre R$ 25 milhões e R$ 50 milhões

( ) Entre R$ 50 milhões e R$ 100 milhões

( ) Entre R$ 100 milhões e R$ 250 milhões

( ) Entre R$ 250 milhões e R$ 500 milhões

( ) Acima de R$ 500 milhões

**03 – Qual a quantidade atual de clientes da empresa em que você trabalha ?**

( ) Menos de 1.000 clientes
( ) Entre 1.001 e 10.000 clientes
( ) Entre 10.001 e 25.000 clientes
( ) Entre 25.001 e 50.000 clientes
( ) Entre 50.001 e 100.000 clientes
( ) Entre 100.001 e 250.000 clientes
( ) Entre 250.001 e 500.000 clientes
( ) Entre 500.001 e 1.000.000 clientes
( ) Mais de 1.000.000 de clientes

**04 – Qual a origem da empresa em que você trabalha ?**

( ) Brasileira
( ) Americana
( ) Inglesa
( ) Francesa
( ) Alemã
( ) Outra. Informar _____
( ) Mais de uma origem Informar os países de origem _____

**05 – A empresa em que você trabalha utiliza algum tipo de estratégia para se relacionar com seus clientes ?**

( ) Sim
( ) Não

**06 – Assinale um valor (entre 0 e 10) para o grau de intensidade com que sua empresa utiliza as seguintes estratégias de relacionamento com clientes, considerando que 0 (zero) significa nenhuma utilização da estratégia e 10 (dez) significa utilização total da estratégia.**


| Estratégia de relacionamento com clientes | Grau de utilização (0 a 10) |
|---|---|
| Aquisição de novos clientes | |
| Identificação dos melhores clients | |
| Valor vitalício do cliente (LTV – *lifetime value*) | |
| *Cross-selling* (venda cruzada) | |
| *Up-selling* (oferta de atualizações, complementos do mesmo produto ou serviço) | |
| Diferenciação de clientes | |
| Reconquista de clientes | |
| Fidelização de clientes | |
| Salvamento de clientes (políticas anti-attrition ou anti-churn) | |
| Outras (informar nome e grau de utilização) | |

**07 – A empresa em que você trabalha utiliza algum tipo de modelagem (estatística ou matemática) para subsidiar a criação de estratégias de relacionamento com clientes ?**

( ) Sim

( ) Não

**08 –Assinale um valor (entre 0 e 10) para o GRAU DE INTENSIDADE com que sua empresa utiliza as seguintes etapas de análise para dar suporte à criação de estratégias de relacionamento com clientes, considerando que 0 (zero) significa nenhuma utilização da respectiva etapa de análise e 10 (dez) significa utilização total da respectiva etapa de análise.**


| Etapa de análise | Grau de utilização (0 a 10) |
|---|---|
| Entendimento do problema de negócio da empresa | |
| Definição de metas e objetivos do projeto de modelagem de dados | |
| Utilização de um cronograma com todas as fases do projeto de modelagem | |
| Seleção das variáveis mais relevantes para a modelagem dos dados | |
| Identificação de inconsistência nos dados | |
| Limpeza dos dados (exclusão de dados inválidos, duplicados ou incoerentes) | |
| Criação de novas variáveis a partir das existentes | |
| Reunião com áreas de negócio para discussão da importância das variáveis coletadas | |
| Montagem de uma base única para análise de dados (integração de dados) | |
| Estudo de possíveis técnicas de modelagem a serem usadas | |
| Comparação de resultados do modelo por mais de uma técnica | |
| Criação de amostra de desenvolvimento x validação | |
| Avaliação do modelo utilizando critérios técnicos | |
| Avaliação do modelo utilizando critérios de negócio | |
| Reunião com áreas de negócio para discussão do modelo | |
| Especificação de um plano de implementação do modelo | |
| Cronograma de especificação do modelo desenvolvido | |
| Montagem de documentação oficial do projeto | |
| Apresentação oficial dos resultados do projeto | |
| Simulação dos modelos em sistemas (ambiente de simulação ou teste) | |
| Acompanhamento do modelo após a implementação | |
| Revisões periódicas dos modelos desenvolvidos | |

**09 – Assinale um valor (entre 0 e 10) para o GRAU DE IMPORTÂNCIA que sua empresa dá às seguintes etapas de análise para fornecer suporte à criação de estratégias de relacionamento com clientes, considerando que 0 (zero) significa nenhuma importância dada à respectiva etapa de análise e 10 (dez) significa importância total dada à respectiva etapa de análise.**


| Etapa de análise | Grau de importância (0 a 10) |
|---|---|
| Entendimento do problema de negócio da empresa | |
| Definição de metas e objetivos do projeto de modelagem de dados | |
| Utilização de um cronograma com todas as fases do projeto de modelagem | |
| Seleção das variáveis mais relevantes para a modelagem dos dados | |
| Identificação de inconsistência nos dados | |
| Limpeza dos dados (exclusão de dados inválidos, duplicados ou incoerentes) | |
| Criação de novas variáveis a partir das existentes | |
| Reunião com áreas de negócio para discussão da importância das variáveis coletadas | |
| Montagem de uma base única para análise de dados (integração de dados) | |
| Estudo de possíveis técnicas de modelagem a serem usadas | |
| Comparação de resultados do modelo por mais de uma técnica | |
| Criação de amostra de desenvolvimento x validação | |
| Avaliação do modelo utilizando critérios técnicos | |
| Avaliação do modelo utilizando critérios de negócio | |
| Reunião com áreas de negócio para discussão do modelo | |
| Especificação de um plano de implementação do modelo | |
| Cronograma de especificação do modelo desenvolvido | |
| Montagem de documentação oficial do projeto | |
| Apresentação oficial dos resultados do projeto | |
| Simulação dos modelos em sistemas (ambiente de testes ou simulaçao) | |
| Acompanhamento do modelo após a implementação | |
| Revisões periódicas dos modelos desenvolvidos | |

**10 – Assinale um valor (entre 0 e 10) para o grau de intensidade com que sua empresa utiliza as seguintes técnicas de análise de modelagem de dados para dar suporte à criação de estratégias de relacionamento com clientes, considerando que 0 (zero) significa nenhuma utilização da respectiva técnica de análise e 10 (dez) significa utilização total da respectiva técnica de análise.**


| Técnica de modelagem | Grau de utilização (0 a 10) |
|---|---|
| Árvores de classificação ou regressão (CHAID, C&RT, C5.0, etc.) | |
| Redes neurais | |
| Análise discriminante | |
| Regressão linear múltipla | |
| Regressão logística | |
| Análise de componentes principais | |
| Análise de sobrevivência | |
| Séries temporais | |
| Técnicas de segmentação (*cluster analysis*) | |
| Análise de cestas de mercado (*market basket analysis*) | |
| Algoritmos de associação (APRIORI ou GRI) | |
| Análise exploratória de dados | |
| Visualização dos dados (análises visuais dos dados) | |
| Mineração de texto (*text mining*) | |
| Mineração na internet (*web mining*) | |
| Outras técnicas (informar nome e freqüência) | |

**11 – Quais *softwares* estatísticos ou de mineração de dados a empresa utiliza para o desenvolvimento de modelos estatísticos/matemáticos que subsidiem a criação de estratégias de relacionamento com clientes ?**

( ) SPSS
( ) SAS
( ) Statistica
( ) Minitab
( ) SPSS Clementine
( ) SAS Enterprise Miner
( ) Statistica Data Miner
( ) IBM Intelligent Miner
( ) KXEN
( ) Outros. Informar o nome do *software*: _____

**12 – A empresa em que você trabalha possui algum programa ou processo de CRM implementado ?**

( ) Não
( ) Sim Qual? _____
( ) Em fase de implementação

**13 – A empresa em que você trabalha possui uma área de CRM analítico ?**

( ) Sim
( ) Não
( ) Em fase de criação

# Apêndice B – Cronograma de trabalho pós-qualificação

| Etapa | Período | Comentários |
|---|---|---|
| Incorporação das sugestões da banca à dissertação | 13-08-2007 a 15-09-2007 | Incluir comentários e sugestões da banca no conteúdo da dissertação |
| Levantamento de base inicial de potenciais respondentes | 13-08-2007 a 15-08-2007 | Contato inicial com profissionais da área, interessados em responderem a pesquisa |
| Início do envio de questionários por e-mail (preenchimento em arquivo word) | 15-08-2007 | Foi considerado um plano alternativo de coleta (envio do questionário por e-mail), no caso da recusa de preenchimento via internet. |
| Acompanhamento das respostas | 21-08-2007 a 05-11-2007 | Foram realizados acompanhamentos semanais, lembrando os potenciais respondentes sobre o preenchimento do questionário. Questionários com respostas inválidas foram devolvidos para reanálise. |
| Fim do envio de questionários | 15-11-2007 | Período de envio dos questionários: 3 meses |
| Compilação dos dados | 20-08-2007 a 20-11-2007 | Digitação e compilação dos resultados em *software* estatístico SPSS 13.0 |
| Análise dos dados | 05-11-2007 a 03-12-2007 | Utilização de tabelas de freqüência, tabelas de contingência, medidas descritivas e testes não-paramétricos para análise e interpretação dos resultados |
| Escrita da etapa de análise da dissertação | 20-11-2007 a 06-12-2007 | Transposição da interpretação escrita para o documento da dissertação, bem como limitações do estudo e possíveis sugestões. |
| Envio para o orientador para correção | 06-12-2007 | Silvio lê a dissertação e devolve com comentários |
| Revisão final da dissertação | 09-12-2007 a 12-12-2007 | Revisão final do documento inteiro, buscando incorreções a serem corrigidas antes da entrega do documento final. |
| Impressão do documento | 13-12-2007 a 15-12-2007 | Entrega do documento para impressão e encadernação |
| Entrega do documento final | 18-12-2007 | Depósito da dissertação |
| Apresentação | 15-12-2007 a 21-12-2007 | Confecção da apresentação final para a defesa da dissertação. |