

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE TECNOLOGIA
PROGRAMA DE ENGENHARIA DE PRODUÇÃO

**ANÁLISE DE DESEMPENHO DE VENDAS EM TELECOMUNICAÇÕES
UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS**

por

TEÓFILO CAMARA MATTOZO

ENGENHEIRO ELETRICISTA, UFRN, 1977

DISSERTAÇÃO SUBMETIDA AO PROGRAMA DE ENGENHARIA DE PRODUÇÃO DA
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE

MESTRE EM CIÊNCIAS EM ENGENHARIA DE PRODUÇÃO

NOVEMBRO, 2007

© 2007 TEÓFILO CAMARA MATTOZO
TODOS DIREITOS RESERVADOS.

O autor aqui designado concede ao Programa de Engenharia de Produção da Universidade Federal do Rio Grande do Norte permissão para reproduzir, distribuir, comunicar ao público, em papel ou meio eletrônico, esta obra, no todo ou em parte, nos termos da Lei.

Assinatura do Autor: _____

APROVADO POR:

Prof. José Alfredo Ferreira Costa, D.Sc - Orientador, Presidente

Prof. Manoel Veras de Sousa Neto, D.Sc - Membro Examinador

Prof. Paulo César Formiga Ramos, D.Sc - Membro Examinador

Prof. Marcelo Sampaio de Alencar, D.Sc - Examinador Externo

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Divisão de Serviços Técnicos

Catálogo da Publicação na Fonte. UFRN / Biblioteca Central Zila Mamede

Mattozo, Teófilo Camara.

Análise de desempenho de vendas em telecomunicações utilizando técnicas de mineração de dados / Teófilo Camara Mattozo. – Natal, RN, 2007.

115 f.

Orientador: José Alfredo Ferreira Costa.

Dissertação (Mestrado) – Universidade Federal do Rio Grande do Norte. Centro de Tecnologia. Programa de Pós-Graduação em Engenharia de Produção.

1. Indicadores de desempenho – Dissertação. 2. Bases de Dados – Conhecimento – Dissertação. 3. Telecomunicações – Gestão de Negócios – Dissertação. 4. Sistema de Apoio à Decisão – Dissertação. I. Costa, José Alfredo Ferreira. II. Universidade Federal do Rio Grande do Norte. III. Título.

RN/UF/BCZM

CDU 658.817(043.3)

CURRICULUM VITAE RESUMO



Experiência Profissional:

- Atuação e domínio do portfólio de soluções de voz básica, voz avançada, comunicação de dados, mobilidade (voz e dados) e das tecnologias envolvidas;
- Conhecimentos técnicos de soluções de comunicação convergente e *networking*;
- Gerenciamento de grandes contas, em empresas Públicas e Privadas;
- Integração, implantação e expansão de sistema de comutação, *backbone* e redes de acesso;
- Gerência, identificação e desenvolvimento de agentes autorizados, parceiros e integradores de negócios;
- *Marketing* de Serviços e Planejamento Estratégico de *Marketing*;
- Gestão de grandes contas (principalmente públicas) com participações em licitações, da concepção à adjudicação;
- Elaboração de planos de compensação e incentivo para Equipes de Vendas e Agentes Autorizados.

Formação Acadêmica:

- Graduado em Engenharia Elétrica pela Universidade Federal do Rio Grande do Norte, no período de Jan/73 a Jul/77;
- Curso de pós-graduação em Engenharia da Elétrica pela Universidade Federal do Rio Grande do Norte, no período de Jan/86 a Dez/87 (dissertação não defendida);
- Curso de especialização em Telecomunicações pela Universidade Federal da Paraíba, no período de Nov/94 a Fev/96;
- Curso de pós-graduação em Marketing pela Escola de Pós Graduação em Economia da Fundação Getúlio Vargas – FGV/RJ, no período de Fev/98 a Mai/99;
- Curso de pós-graduação de Finanças Empresariais pela Escola de Pós Graduação em Economia da Fundação Getúlio Vargas – FGV/RJ, no período de Ago/99 a Abr/00 (não concluído);
- Cursos de aperfeiçoamento profissional e gerencial, com destaque para Marketing, Vendas, Liderança, Finanças Corporativas, Orçamento Empresarial, Sistema de

Comutação Local/Trânsito, Tráfego Telefônico, Sistema Móvel Celular, Sistema Rádio, Comunicação de Dados, Redes de Telecomunicações e Consultoria de Negócios em Telecomunicações;

Experiência Acadêmica:

- Trabalho publicado no XIV SIMPEP (Nov/07) Simpósio de Engenharia de Produção com o título “Sistema de Análise de Desempenho de Vendas em Telecomunicações Utilizando Técnicas de Mineração de Dados”.
- Palestrante no SUCESU 86 Congresso Nacional de Informática – RJ, com o tema “Sistema Automático de Medição de Tráfego”;
- Projeto de desenvolvimento do curso Técnico em Telecomunicações do SENAI – CTI/RN;
- Coordenador do curso Técnico em Telecomunicações do SENAI – CTI/RN;
- Consultoria e coordenação do programa de Formação em Telecomunicações da ACCESSUS – Consultores Associados;
- Professor de cursos de Comunicação de Dados, Atualização em Telecomunicações e Comutação Digital na TELEBRÁS;
- Professor de cursos de RDSI (Rede Digital de Serviços Integrados) e Marketing de Serviços na TELEMAR;
- Professor de cursos de Sistemas de Comutação na ACCESSUS – Consultores Associados;
- Professor de cursos de Princípios de Telecomunicações, e Sistemas de Comutação no SENAI – CTI/RN;
- Professor de cursos de Consultor de Negócios de Telecomunicações na TELEMAR;
- Supervisor de vários estágios acadêmicos de Engenharia Elétrica da UFRN.

DEDICATÓRIA

A meu pai, **Deusdedith Mattozo** (*in memorian*), minha mãe, **Gelza Mattozo**, minha mulher, **Margareth Mattozo**, e meus filhos, **Diogo**, **Renato** e **Thuiza**, pela grande paciência que sempre têm comigo, nessa minha incessante e inquietante sede de saber.

AGRADECIMENTOS

Realizar um trabalho como esse implica em empenho, perseverança, disciplina, e no acompanhamento e estímulo das pessoas que estão mais próximas. Na conclusão desta etapa da minha formação acadêmica fico muito grato e expresso o meu reconhecimento e gratidão:

Aos professores Dr. José Alfredo Ferreira Costa e Dr. Manoel Veras de Sousa Neto, pelas aulas que tantas informações e aprendizagem nos trouxeram, pela orientação e incentivo no desenvolvimento desse trabalho.

Aos professores Dr. Paulo César Formiga Ramos e Dr. Marcelo Sampaio de Alencar pela paciência e dedicação que nos concederam, por terem aceitado participar da banca examinadora e, principalmente, pelos conhecimentos que compartilharam conosco.

A todos os professores, alunos e funcionários do Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Norte, pela oportunidade de realização do mestrado e pelo profissionalismo com que conduziram o desenvolvimento do curso. Às empresas que colaboraram para a sua realização pela amizade que nos dedicaram ao longo do mestrado.

Aos meus colegas de turma, cujo trabalho de equipe contribuiu, sem dúvida, para o bom êxito desse trabalho.

A meus pais Deusdedith (*in memoriam*) e Gelza por tudo, desde o primeiro minuto da minha existência; por terem me tornado a pessoa que sou hoje, pelos meus estudos, pela minha educação, pela paciência, pela tolerância, enfim, por todo o amor.

A meus irmãos Sidney e Liane, pelo companheirismo em todos os momentos da minha vida e pela incansável torcida.

A minha esposa Margareth (minha amiga e companheira), pela paciência e apoio nos momentos mais difíceis que passei e que entendeu que a execução desse trabalho exigiria sacrifícios, tendo sido tolerante ao abrir mão de vários momentos de nosso convívio para que eu pudesse estudar.

A meus filhos Dinho, Tato e Tuca, todos, sempre no meu pensamento. Obrigado por fazerem parte da minha história. São vocês a razão desse sucesso.

E a todos aqueles que, de maneira direta ou indireta, contribuíram para a realização desse trabalho. Por fim, esta vitória é para todas as pessoas que fizeram com que isto acontecesse.

Resumo da Dissertação apresentada à UFRN/PEP como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências em Engenharia de Produção.

ANÁLISE DE DESEMPENHO DE VENDAS EM TELECOMUNICAÇÕES UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS

TEÓFILO CAMARA MATTOZO

Novembro/2007

Orientador : José Alfredo Ferreira Costa

Curso: Mestrado em Ciências em Engenharia de Produção

Telecomunicações é uma das mais dinâmicas e estratégicas áreas no mundo atual. Há constante necessidade das organizações buscarem novas formas de gerenciamento, em um ambiente cada vez mais competitivo e com recursos cada vez menores. A existência de bases de dados nas empresas passou a ter maior importância. Na grande maioria dos casos, dados não são ainda explorados adequadamente. Técnicas de Descoberta de Conhecimento em Bases de Dados (DCBD) surgem como alternativas, permitindo o estudo de problemas complexos, sendo cada vez mais utilizadas nas diferentes áreas de gestão.

O presente trabalho apresenta uma proposta para a sistematização das atividades de DCBD a qual integra as metodologias CRISP-DM, SEMMA, FAYYAD, em um ambiente interativo, bem como um estudo de viabilidade do uso de análise de regressão linear múltipla para explicação de vendas, no setor corporativo de telecomunicações, utilizando indicadores de desempenho. Foi delineado um método estatístico para analisar o efeito que os indicadores de desempenho têm sobre o comportamento da produtividade de venda.

Mediante análises estatísticas e comerciais criteriosas, as equações foram definidas, sendo ajustados os seus respectivos coeficientes de determinação. Foram também realizados testes de hipóteses de seus parâmetros, visando à validação ou não dos modelos de regressão e análise da qualidade de seus ajustamentos. Ficou evidenciada a existência de relacionamento entre as características desses indicadores de desempenho com o volume de vendas realizado.

Palavras-chave: Indicadores de Desempenho. Descoberta de Conhecimento em Bases de Dados. Gestão de Negócios em Telecomunicações. Sistema de Apoio à Decisão.

Abstract of Master Thesis presented to UFRN/PEP as fulfillment of requirements to the degree of Master of Science in Production Engineering

ANALYSIS OF BUSINESS DEVELOPMENT IN TELECOMMUNICATION USING DATA MINING TECHNIQUES

TEÓFILO CAMARA MATTOZO

November/2007

Thesis Advisor : José Alfredo Ferreira Costa

Program: Master of Science in Production Engineering

Nowadays, telecommunications is one of the most dynamic and strategic areas in the world. Organizations are always seeking to find new management practices within an ever increasing competitive environment where resources are getting scarce. In this scenario, data obtained from business and corporate processes have even greater importance, although this data is not yet adequately explored. Knowledge Discovery in Databases (KDD) appears then, as an option to allow the study of complex problems in different areas of management.

This work proposes both a systematization of KDD activities using concepts from different methodologies, such as CRISP-DM, SEMMA and FAYYAD approaches and a study concerning the viability of multivariate regression analysis models to explain corporate telecommunications sales using performance indicators. Thus, statistical methods were outlined to analyze the effects of such indicators on the behavior of business productivity.

According to business and standard statistical analysis, equations were defined and fit to their respective determination coefficients. Tests of hypotheses were also conducted on parameters with the purpose of validating the regression models. The results show that there is a relationship between these development indicators and the amount of sales.

Key words: Performance Indicators. Knowledge Discovery in Database. Business Management in Telecommunication. Decision Support Systems.

SUMÁRIO

Capítulo 1 Introdução.....	1
Capítulo 2 Pesquisa Bibliográfica.....	12
2.1 Descoberta de Conhecimento em Bases de Dados.....	12
2.2 Mineração de Dados.....	38
2.3 Medição de Desempenho.....	63
2.4 Vendas Industriais.....	82
2.5 Relação entre Produtividade de Vendas e Medição de Desempenho.....	102
Capítulo 3 Metodologia da Pesquisa.....	126
3.1 Tipo de Pesquisa.....	126
3.2 Problema de Pesquisa.....	127
3.3 Detalhamento dos Objetivos da Pesquisa e das Hipóteses Definidas.....	128
3.4 Estabelecimento de Relacionamentos entre as Medidas de Desempenho.....	129
3.5 Detalhamento da Metodologia Desenvolvida.....	130
3.6 A Metodologia Proposta MDAV versus CRISP-DM.....	167
3.7 Ferramenta SPSS.....	170
Capítulo 4 Resultados da Pesquisa.....	172
4.1 Metodologia Utilizada para Estimação e Avaliação do Modelo de Regressão.....	172
4.2 Métodos Testados no Modelo de Regressão Linear Múltipla.....	174
4.3 Teste F Generalizado do Modelo <i>Stepwise</i> com Relação ao Modelo Combinatório.....	186
4.4 Interpretação dos Coeficientes do Modelo.....	187
4.5 Análise aos Pressupostos do Modelo.....	189
4.6 Verificação dos Testes de Hipóteses.....	199
Capítulo 5 Conclusões e Recomendações.....	202
5.1 Conclusões.....	202
5.2 Sugestões e Recomendações.....	205
Referências Bibliográficas.....	206
Anexo.....	214

LISTA DE TABELAS E QUADROS

- Tabela 4.1 – Coeficientes de regressão do método confirmatório
- Tabela 4.2 – Sumário do modelo da regressão do método confirmatório
- Tabela 4.3 – Variância (ANOVA) do método confirmatório
- Tabela 4.4 – Coeficientes de regressão do método combinatório
- Tabela 4.5 – Sumário do modelo da regressão do método combinatório
- Tabela 4.6 – Variância (ANOVA) do método combinatório
- Tabela 4.7 – Coeficientes de regressão do método seqüencial
- Tabela 4.8 – Sumário do modelo da regressão do método seqüencial
- Tabela 4.9 – Variância (ANOVA) do método seqüencial
- Tabela 4.10 – Produtividade de vendas em função das correlações do método seqüencial
- Tabela 4.11 – Correlações de Pearson do método seqüencial
- Tabela 4.12 – Estatísticas de colinearidade do método seqüencial
- Tabela 4.13 – Diagnóstico de colinearidade do método seqüencial
- Tabela 4.14 – Diagnóstico de *Casewise* da regressão do método seqüencial
- Tabela 4.15 – Comparativo dos métodos testados
- Quadro 2.1 – Comparação entre as metodologias FAYYAD, ADRIAANS e CRISP-DM
- Quadro 2.2 – Diferenças entre o mercado empresarial e o de bens de consumo
- Quadro 2.3 – Estágios percorridos no processo de compra empresarial
- Quadro 2.4 – Variáveis utilizadas na análise na compra empresarial
- Quadro 2.5 – As bases de segmentação em nível macro
- Quadro 2.6 – As bases de segmentação em nível micro
- Quadro 2.7 – Efeito da experiência do cliente no setor de plásticos de engenharia
- Quadro 2.8 – Estratégias de marketing empresarial x experiência do cliente
- Quadro 2.9 – Segmentação da base de clientes x potencial e vendas correntes
- Quadro 2.10 – Atratividade do cliente x tempo de fechar a venda
- Quadro 2.11 – Matriz de dispersão dos clientes x preço líquido e custo de servir
- Quadro 2.12 – Informações dos clientes
- Quadro 2.13 – Informações dos serviços
- Quadro 2.14 – Informações da gestão de vendas
- Quadro 2.15 – Informações da concorrência
- Quadro 2.16 – Sugestão de indicadores de desempenho de vendas
- Quadro 2.17 – Medidas complementares para avaliação de desempenho

LISTA DE FIGURAS, GRÁFICOS E ILUSTRAÇÕES

- Figura 1.1 – Estrutura da dissertação
- Figura 2.1 – Dado, informação e conhecimento
- Figura 2.2 – Infra-estrutura tecnológica de apoio ao *Business Intelligence*
- Figura 2.3 – Fases do processo de KDD
- Figura 2.4 – Percentagem de tempo despendido em cada fase do processo de KDD
- Figura 2.5 – Relação entre o número de mailings e o número de respostas
- Figura 2.6 – Metodologia CRISP-DM
- Figura 2.7 – Metodologia SEMMA
- Figura 2.8 – Metodologia ADRIAANS
- Figura 2.9 – Metodologia ADRIAANS
- Figura 2.10 – Metodologia KLEMETTINEN
- Figura 2.11 – Um modelo para o *design* da medição de desempenho.
- Figura 2.12 – Posicionamento do sistema de medição de desempenho em uma organização
- Figura 2.13 – Estrutura do modelo de medição de desempenho *Performance Pyramid*
- Figura 2.14 – Mapa estratégico no *Balanced Scorecard*
- Figura 2.15 – Analisando o desempenho de entrega no IDPMS
- Figura 2.16 – Integração do sistema de medição de desempenho no IPMS
- Figura 2.17 – Cinco atividades fundamentais da medição de desempenho.
- Figura 2.18 – Segmentação por níveis
- Figura 3.1 – Modelo esquemático do teste de hipóteses
- Figura 3.2 – Proposta da metodologia de Mineração de Dados (MDAV)
- Figura 3.3 – Obtenção de informação de diversos sistemas
- Figura 3.4 – Organização das atividades de pré-processamento e transformação de dados
- Figura 3.5 – Paralelismo entre a metodologia proposta e CRISP-DM
- Gráfico 4.1 – Resíduos padronizados & resíduos REFX do método seqüencial
- Gráfico 4.2 – Regressões parciais das variáveis independentes do método seqüencial
- Gráfico 4.3 – Histograma de resíduos da variável dependente do método seqüencial
- Gráfico 4.4 – Probabilidade normal PP-plot dos resíduos do método seqüencial
- Gráfico 4.5 – *Boxplot* do método seqüencial
- Ilustração 2.1 – *The continuum of industrial buying situations.*

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS.

ABNT	– Associação Brasileira de Normas Técnicas
AM	– Aprendizado de Máquina
ASE	– Avaliação de Vendas de Serviços
B2B	– Mercado Empresarial
B2C	– Mercado de Consumo
BOVESPA	– Bolsa de Valores de São Paulo
BD	– Bases de Dados
BI	– <i>Business Intelligence</i>
BSC	– <i>Balanced Scorecard</i>
BZ	– Clientes que dão Prejuízo
CMP	– Clientes de Maior Potencial
CMW	– Clientes de Maior Valor
CO1	– Indicador Vendas versus Mapeamento da Concorrência
CP	– Estatística de Mallows
CRISP-DM	– <i>CROSS-Industry Standard Process for Data Mining</i>
DM	– <i>Data Mining</i>
DSS	– <i>Decision Support System</i>
DW	– <i>Data Warehouse</i>
EO1	– Indicador Assertividade de Entrega e Cadastro de Contratos
EO4	– Indicador Retorno de Contratos
EO5	– Indicador Abertura de Ordem de Serviços
EO6	– Indicador Conversão de Contratos
EO7	– Indicador Cancelamento de Vendas
FC1	– Indicador Média de Visitas por Consultor
FC2	– Indicador Assertividade de Agendamento
FC3	– Indicador Visitação versus Agendamento
FIV	– Fator de Inflação de Variância
GLE	– Grau de Liberdade Associado ao Erro
GP1	– Indicador Produtividade de Visitas
GP2	– Indicador Eficácia de Visitas
GP3	– Indicador Cesta de Produtos
GP4	– Indicador Distribuição de Vendas

GP5	– Indicador Produtividade de Oportunidades
IA1	– Indicador Volume de Vendas Com Descontos
IA2	– Indicador Volume Aprovação de Viabilidade
IA3	– Indicador Assertividade de Análise de Viabilidade
IDPMS	– <i>Integrated Dynamic Performance Measurement System</i>
IPMS	– <i>Integrated Performance Measurement System</i>
KDD	– <i>Knowledge Discovery in Data Bases</i>
KM	– <i>Knowledge Management</i>
MAD	– Desvio Médio Absoluto
MADV	– Mineração de Dados para Administração de Vendas
MIS	– <i>Management Information System</i>
MSE	– Média da Soma dos Quadrados dos Erros
MSR	– Média da Soma dos Quadrados da Regressão
OLAP	– <i>On-Line Analytical Processing</i>
P&D	– Pesquisa e Desenvolvimento
PIB	– Produto Interno Bruto
PP	– <i>Performance Prism</i>
RMSE	– Raiz do Erro Médio Quadrático
SA3	– Sistema de Vendas
SAS	– <i>SAS Institute Inc</i>
SCI	– Sistema de Cadastro de Vendas
SEMMA	– <i>Sample, Explore, Modify Model, Assessment</i>
SGBD	– Sistemas Gerenciadores de Bancos de Dados
SMART	– <i>Strategic, Measurement, Analysis and Reporting Technique</i>
SMD	– Sistema de Medição de Desempenho
SPSS	– <i>Statistical Package for Social Sciences</i>
SSE	– Soma dos Quadrados do Erro
SSR	– Soma dos Quadrados da Regressão
SST	– Soma dos Quadrados Totais
TI	– Tecnologia de Informação
UDT	– Unidade Tomadora de Decisão

Capítulo 1

Introdução

O mercado brasileiro de telecomunicações vem passando por grandes mudanças nos últimos anos. Essas mudanças têm sido impulsionadas principalmente pela evolução da tecnologia e pela chamada desregulamentação do mercado, com a delegação da prestação dos serviços de telecomunicações à iniciativa privada. O que era antes um monopólio estatal foi transformado em um mercado cuja competição tende a aumentar cada vez mais, por meio do incentivo do governo à entrada de novos concorrentes e do aumento da variedade de serviços e facilidades proporcionados pelas novas tecnologias. Com mercados e seus participantes mudando constantemente, para que uma empresa consiga estabelecer uma vantagem competitiva sustentável é necessário que ela procure continuamente inovar para competir. A história da evolução dos negócios é repleta de empresas, incluindo grandes companhias multinacionais, que falharam porque não acompanharam as mudanças e tendências de mercado ou porque não entenderam ou souberam atender às necessidades de seus clientes.

O setor de telecomunicações, por sua vez, foi escolhido para ser objeto desse estudo por que:

- É um dos mais proeminentes setores no Brasil: responde, atualmente, por 7% do PIB nacional. O faturamento das empresas de telecomunicações em 2005 somou 46 bilhões de dólares, o que significa um crescimento de 32,69% em relação ao ano anterior, enquanto o PIB se expandiu apenas 2,3% no mesmo período (ANUÁRIO TELECOM, 2007). Em 2007 espera-se que o mercado de transmissão de dados no mercado corporativo cresça mais que o de voz. (está previsto o

crescimento de 35% ao ano contra 5% de aumento ao ano da telefonia privada e doméstica);

- É um setor intensivo em tecnologia, portanto, espera-se que as empresas atuantes no setor sejam inovadoras no uso de novas ferramentas para obter e manter clientes;
- É um setor que está assumindo um perfil essencialmente global, com a maioria das empresas atuantes no mercado sendo multinacionais e tendo também em seu portfolho de clientes grandes corporações multinacionais, algumas delas com alcance global.
- Finalmente, é um setor que apenas recentemente abriu suas portas para a concorrência (até bem pouco tempo atrás o setor era monopólio do governo brasileiro), pois as empresas entrantes parecem ter focado totalmente seus esforços em nichos específicos de mercado, mais especificamente no mercado corporativo e nos serviços de maior retorno. Em suma, com o aumento da competitividade no setor de telecomunicações, a importância a longo prazo de se prever a produtividade de vendas e entender as dimensões dos relacionamentos com base nos seus indicadores de desempenho é um aspecto crítico para o futuro das empresas que pode ser alavancado por uma melhoria na sua gestão e, por isso, seu uso é o foco desse estudo.

A observação da realidade é consequência da manifestação de eventos complexos e incertos no decorrer do tempo. Embora esses eventos não sejam exatamente os mesmos, eles também não são completamente diferentes. Há uma linha de continuidade, similaridade e preditibilidade nesses eventos, que permite generalizar eventos futuros, freqüentemente de modo correto, a partir de experiências passadas. A antecipação do comportamento futuro dos acontecimentos sempre despertou interesse nas mais diversas áreas do conhecimento humano. Em telecomunicações, especificamente, a possibilidade de se realizar algum prognóstico que permita a tomada de decisão antecipada de um gestor em relação às expectativas do mercado para que uma empresa consiga estabelecer uma vantagem competitiva sustentável.

O ambiente empresarial, dentro do contexto da economia globalizada tem sido fortemente marcado pela necessidade de se buscar e aplicar novas técnicas e ferramentas de gestão. Essas técnicas e ferramentas têm o objetivo de traduzir, em linguagem organizacional corrente, o conceito de excelência empresarial, em uma perspectiva prática, que proporcione soluções aos desafios organizacionais. Um dos desafios que as organizações necessitam superar é o de descobrir como aperfeiçoar a sua produtividade de negócios realizados, priorizando os processos que devam ser gerenciados. É preciso planejar estrategicamente o futuro da organização e ater-se a processos que criem valor para o cliente, bem como ter um

sistema de medição capaz de demonstrar como a organização está se comportando perante as turbulências do mercado.

O processo de globalização da economia fez com que alterasse as perspectivas de negócios e mercados caracterizados por relações internacionais. Ao mesmo tempo, com modificações nos aspectos de rentabilidade e ativos. Os países da América do Sul vão delineando uma economia instável, características de economias com forte dependência de capitais externos e de acirrada competitividade, impostas pelos países desenvolvidos. Nesse contexto, as organizações, buscam modelos e alternativas de gestão, alicerçadas no conhecimento e educação, de mecanismos que forneçam informações, visando minimizar os efeitos negativos do processo de uma economia em crescentes dificuldades. Entre os diversos mecanismos, um sistema de medição de desempenho eficiente pode ser uma ferramenta de fundamental importância para as organizações.

A ampliação da concorrência entre os mercados, proporcionada pela globalização somada a mudanças sociais e culturais contemporâneas, tornou obsoletos vários sistemas tradicionais de controle de desempenho. O processo de medição objetiva alavancar vantagens internas e externas para que as organizações possam alcançar a excelência empresarial. Assim, insistir num processo de melhoria contínua e no aperfeiçoamento dos processos empresariais é ponto fundamental para a diferenciação e conquista de qualidade e superioridade.

Atualmente, os sistemas são implementados com a finalidade de auxiliar as tarefas humanas em qualquer área de atuação. Esses atuam gerando e coletando dados operacionais, ou seja, do dia-a-dia transacional. Grandes empresas, organizações e instituições adotam sistemas computacionais com armazenamento de informações em bancos de dados para os mais diversos fins: sistemas financeiros, controle de estoque, sistemas cadastrais, entre outros.

Os avanços no armazenamento de informações, disponibilizados pelos Sistemas Gerenciadores de Bancos de Dados (SGBD), tais como: velocidade, facilidade de acesso e baixo custo, impulsionaram a geração e aumento no volume de dados armazenados. Nessas grandes bases, com o passar dos anos, torna-se impraticável à análise dos dados de forma manual, tornando-se uma tarefa difícil de ser realizada utilizando-se métodos tradicionais. Essas informações, de extrema relevância para a organização, devem ser utilizadas e manipuladas da melhor forma possível a fim de prover informações importantes para auxiliar na tomada de decisões.

A Descoberta de Conhecimento em Bases de Dados (KDD) é a descoberta de conhecimento interessante, porém escondido, em grandes bases de dados. Bases de dados corporativas frequentemente contêm tendências desconhecidas, que são de importância

estratégica para a organização. KDD é uma tecnologia baseada numa nova geração de *hardware* e *software* que inclui análises estatísticas, exploração visual, árvores de decisão, redes neurais, entre outras, para explorar grandes bases de dados e descobrir relações e padrões existentes nessas informações. Difere de técnicas estatísticas porque, ao invés de verificar padrões hipotéticos, utilizam os próprios dados para descobrir tais padrões (HAIR, J. F. et al., 2005).

As várias tarefas desenvolvidas em Descoberta de Conhecimento em Bases de Dados (KDD) têm como objetivo primário a predição e a descrição, ou seja, a produção de um modelo ou a produção de informação, respectivamente. A predição usa atributos para predizer o desconhecido ou os valores futuros de outras variáveis. Descoberta de Conhecimento em Bases de Dados utiliza técnicas estatísticas e de aprendizado de máquinas para construir modelos capazes de predizer o comportamento de determinado atributo. Como descrição, diversas técnicas podem trazer percepções diferentes das apresentadas em tabelas ou relatórios. A descrição procura por padrões que descrevem os dados e são interpretáveis facilmente pelos seres humanos.

Para que esse grande volume de dados não seja analisado manualmente, existem ferramentas e técnicas de KDD que são fundamentadas na idéia de adquirir conhecimento e descobrir tendências, baseadas na análise em busca de padrões nos dados armazenados em bancos de dados. Com a proliferação da utilização da tecnologia de armazenamento de dados, a utilização de ferramentas baseadas em técnicas de mineração de dados vem crescendo na mesma proporção. O objetivo dessas ferramentas é a extração de informações ocultas de grandes quantidades de dados armazenados para a descoberta de conhecimento. A área de descoberta de conhecimento tem despertado interesse tanto no meio comercial quanto científico e está sendo amplamente difundida e aperfeiçoada pelo fato de fornecer resultados promissores para qualquer domínio de aplicação.

Os sistemas tradicionais de informação para a gestão padecem, segundo alguns autores, de várias debilidades e insuficiências. O fraco relacionamento com os objetivos estratégicos, não permite fazer uma clarificação, comunicação e avaliação das estratégias, fator preponderante para o êxito das organizações da atualidade. Os indicadores financeiros e de curto prazo são insuficientes para medir o desempenho das organizações, pois relevam demasiadas ações de natureza histórica (NORREKLIT, 2000 e KAPLAN & NORTON, 2001), acrescentando ainda que esses indicadores não contemplam a análise de vários ativos, que cada vez mais contribuem para a manutenção de uma vantagem competitiva. Por outro lado, o fato das organizações terem de sobreviver num mercado cada vez mais competitivo e globalizado,

com novas características do meio operacional, como são, por exemplo, a inovação, a qualificação dos funcionários e as relações com os clientes, obrigam a que as mesmas revejam os seus sistemas de informação para a gestão.

Considerando o sistema de medição de desempenho como um dos elementos centrais do sistema de gestão, pesquisas acadêmicas têm procurado propor novos modelos de sistemas de medição de desempenho que contribuam para o aumento ou restabelecimento da coerência entre sistema de gestão e organização da produção. Entretanto, por vezes, o tema de pesquisa sistema de medição de desempenho é tratado de forma descolada da problemática da falta de coerência entre sistema de gestão e organização da produção.

A literatura revisada sobre medição de desempenho demonstrou a existência de uma quantidade muito grande de trabalhos com propostas de novos modelos e uma quantidade muito menor de trabalhos sobre o processo de relacionamento entre indicadores e sobre como utilizar as informações geradas pelos sistemas propostos. No caso do uso das informações geradas pelos novos sistemas de medição de desempenho, existe a falta de um modelo de referência para guiar a estruturação do uso por todas as pessoas que tomam decisões numa organização. Apesar das considerações acerca dos vários aspectos relativos ao uso, esse problema ainda merece atenção dos pesquisadores.

O acompanhamento do desempenho do processo de planejamento deve ser realizado a partir da definição de indicadores de desempenho, que facilitem a análise das causas e efeitos dos desvios entre o planejado e o realizado, de forma que os gestores possam corrigir distorções na execução do plano. Nesse contexto, passa a ser muito importante a utilização de indicadores que realmente possam verificar se a missão da empresa está sendo atingida. Para isso, a amplitude dos sistemas de controle de desempenho tem sido modificada a fim de incluir os ativos intangíveis em seu conjunto de indicadores, além dos resultados financeiros. A evolução da tecnologia e dos processos de gestão empresarial tem provocado importantes mudanças no ambiente organizacional. A globalização e os avanços da tecnologia da informação aumentaram significativamente a competitividade do mercado, fazendo com que as empresas necessitem cada vez mais de ferramentas que permitam avaliar seu desempenho frente aos concorrentes.

Diante disso, o problema é saber se as medidas de desempenho escolhidas são as mais adequadas, se o objetivo está sendo alcançado e saber se as melhorias implantadas estão surtindo efeito. Normalmente, não há informações suficientes ou necessárias para responder a essas questões. Desse modo, a definição dos indicadores de desempenho se torna ponto crucial para o sucesso de uma empresa, já que eles podem ser usados como ferramentas para traçar

estratégias em diferentes níveis, departamentos e até localidades de uma mesma organização.

Um outro problema é como definir indicadores precisos para avaliar o cumprimento das rotinas e a melhoria das instituições nos diversos níveis de gerenciamento e como disseminar seu uso por toda a organização. Assim como os indicadores de desempenho servem para medir o desempenho de uma organização é necessária a sua utilização para realizar comparações em níveis estratégicos para poder permitir ao gestor mensurar o desempenho de organizações adotando a lógica de comparações, pois, um dos fundamentos das medidas de desempenho é a própria possibilidade de fazer comparações.

1.1 Motivação

Durante os últimos anos tem sido verificado um crescimento substancial da quantidade de dados armazenados pelas organizações. A análise desses dados, produzidos e armazenados em larga escala, realizada por especialistas com métodos manuais tradicionais, é impraticável. Por outro lado, a informação vem desempenhando um papel fundamental em todos os setores da sociedade e a grande quantidade de dados equivale a um maior potencial de informação. Entretanto, algumas informações não estão caracterizadas explicitamente em um banco de dados, uma vez que, sendo dados operacionais, não interessam quando estudados individualmente. Diante desse cenário, surge a necessidade de explorar esses dados para extração de conhecimento implícito para utilizá-lo no âmbito do problema.

A extração de conhecimento é uma área em constante desenvolvimento, quer no que se refere às técnicas e ferramentas de Mineração de Dados, quer nas tecnologias complementares que contribuem para a persecução de projetos nesse âmbito. Essa área encontra-se também em franca expansão, sendo usada em inúmeros domínios de aplicação e na resolução de problemas de decisão, cada vez mais sofisticados. Apesar dos sucessivos progressos tecnológicos e do aprofundamento de competências dos especialistas, a concretização de projetos de descoberta de conhecimento continua a ser muito complicada e a exigir um conjunto bastante diverso de conhecimentos, ações e decisões.

Um dos muitos desafios que emerge nesse processo consiste na seleção dos métodos de Mineração de Dados mais apropriados, a aplicar na resolução de um determinado problema de análise de dados, de modo a se obter resultados com utilidade para um fim específico. Essa escolha tem naturalmente grande impacto nos resultados alcançados, não existindo critérios simples e gerais que suportem sistematicamente tal decisão, já que a adequação das técnicas de

Mineração de Dados varia em função de diferentes tipos de fatores, entre os quais, alguns complexos (em geral pressupostos dos métodos em termos de características dos dados) e mesmo subjetivos (em geral simplicidade na interpretação de resultados) e, por vezes, conflituosos (em geral nível de precisão versus facilidade de interpretação de resultados).

Conforme levantamento realizado, as empresas de telecomunicações analisadas não possível identificar sistema que utilize de técnicas de descoberta de conhecimento sobre as bases de dados comerciais. Estima-se que tal estudo seja importante na obtenção de informações anteriormente desconhecidas que se encontra em dados históricos das empresas, bem como informações relevantes na tomada de decisão provendo agilidade a diversos procedimentos e ações da administração de vendas.

Com base no exposto, a motivação do trabalho desenvolvido e da presente dissertação, tem, como ponto de partida, a reconhecida dificuldade em estabelecer relacionamentos entre os parâmetros de desempenho e a relevância de assistir esse tipo de exercícios, desenrolando-se em torno de uma abordagem, entendida como adequada, para lidar com tal dificuldade e especificamente devotada a esse âmbito, para colocar a área de administração de vendas com mais informações para que seja realizada uma melhor gestão comercial.

1.2 Objetivo

Como qualquer iniciativa de Descoberta de Conhecimento em Bases de Dados, esse trabalho parte da hipótese de que é possível descobrir conhecimento novo "escondido" no grande volume de dados de vendas e ainda, a partir do comportamento conhecido dos atributos dos indicadores de desempenho de vendas, aumentarem as chances de se descobrir padrões que podem ou não correlacionar os mesmos com o desempenho das vendas para explicar e ajudar a prever o comportamento futuro dos mesmos. Essas hipóteses são reforçadas pela percepção de especialistas em telecomunicações de que é possível extrair informação nova e útil e prever o comportamento futuro dos parâmetros de vendas a partir dos parâmetros associados ao desempenho de vendas. A regressão linear múltipla foi escolhida como método para ser aplicado na questão abaixo citada por fornecer um modelo de fácil interpretação e, principalmente, de simples aplicabilidade.

Dentro desse contexto, a questão de pesquisa da presente dissertação é: *“Existe relação de dependência entre os indicadores de desempenho e a produtividade de vendas? É possível*

determinar a importância desses indicadores, os tipos de relações eventualmente encontradas ou as inter-relações entre eles?”.

Para responder essas questões, a Descoberta de Conhecimento em Bases de Dados por meio da técnica de Regressão Linear Múltipla se enquadra como o método de análise mais apropriado, pois o problema de pesquisa envolve uma única variável dependente que é a produtividade de vendas e um conjunto de variáveis independentes, quais sejam: os indicadores de desempenho. Essa técnica também fornece um meio de avaliar objetivamente o grau e o caráter da relação entre as variáveis dependente e independentes, pela formação da variável estatística através das variáveis independentes. Essas, além de sua previsão coletiva da variável dependente, também podem ser consideradas por sua contribuição individual à variável estatística e suas previsões (HAIR, J. F. et al., 2005).

Essa pesquisa tem como objetivo identificar relações entre os parâmetros de venda e de indicadores de desempenho, visando descobrir eventuais novos conhecimentos, por intermédio de técnicas de Descoberta de Conhecimento em Bases de Dados, possibilitando a previsão da receita gerada pelas vendas de telefonia fixa, com maior grau de confiança e num intervalo de tempo adequado, com a intenção de auxiliar os gestores na tomada de decisões. Na seção seguinte, é analisando a hipótese que existe relação de dependência entre o volume de vendas (variável dependente) com os indicadores de desempenho (variáveis independentes). Entende-se como variável dependente como o efeito presumido ou resposta a uma mudança nas variáveis independentes. Essas, por sua vez, podem ser definidas como sendo a causa presumida de qualquer mudança da variável dependente, prevendo as mudanças da variável dependente como resposta nas variáveis independentes por meio de uma técnica de dependência a qual pode ser definida como aquela em que uma variável é identificada como a variável dependente a ser conhecidas como variáveis independentes preditiva ou explicada por outras variáveis. O detalhamento dos objetivos e das hipóteses é elucidado no capítulo três.

1.3 Contribuição Esperada

Como consequência do estudo do processo de descoberta de conhecimento aplicado nos dados, espera-se contribuir para as empresas de telecomunicações analisadas. Com os estudos realizados, espera-se fornecer a oportunidade de se obter informações relevantes que auxiliem no processo de tomada de decisão. Com isso, o planejamento estratégico e operacional das empresas poderá encontrar alternativas para solucionar problemas de seu planejamento e

organização. Assim, tanto as empresas como a comunidade científica, poderão ser beneficiadas com o estudo das técnicas de Mineração de Dados incorporadas no auxílio à tomada de decisões.

Esse trabalho propõe a aplicação de técnicas de Descoberta de Conhecimento em Base de Dados por meio de um estudo realizado sobre os dados comerciais de telecomunicações das empresas analisadas. Esse estudo poderá contribuir com o meio empresarial e acadêmico por intermédio de um exemplo de aplicação real do processo de KDD, possibilitando a futura ampliação do estudo dando continuidade a essa pesquisa.

Os dois pontos principais dessa contribuição são: primeiro a metodologia para construção de um modelo para Descoberta de Conhecimento em Base de Dados utilizando a técnica de regressão linear múltipla para estimar as vendas baseada nos parâmetros de desempenho com um bom ajuste aos dados observados e que forneça uma estimativa calibrada, não distante da realidade e o segundo ponto consiste na metodologia de análise e diagnóstico do modelo construído. Espera-se que esta pesquisa venha contribuir para o aprimoramento do sistema de gestão em telecomunicações e que seja um instrumento útil para a área acadêmica bem como melhoria do gerenciamento das empresas das áreas afins, atenuando uma eventual carência de bibliografia especializada no assunto, na língua portuguesa.

1.4 Organização da Dissertação

A dissertação, além desse capítulo introdutório, apresenta mais quatro capítulos e um conjunto de anexos considerados de interesse, quer para o trabalho desenvolvido, quer para trabalhos futuros que sobre ele possam vir a ser desenvolvidos. Uma visão geral pode ser elucidada pela Figura 1.1 a seguir mostrada.

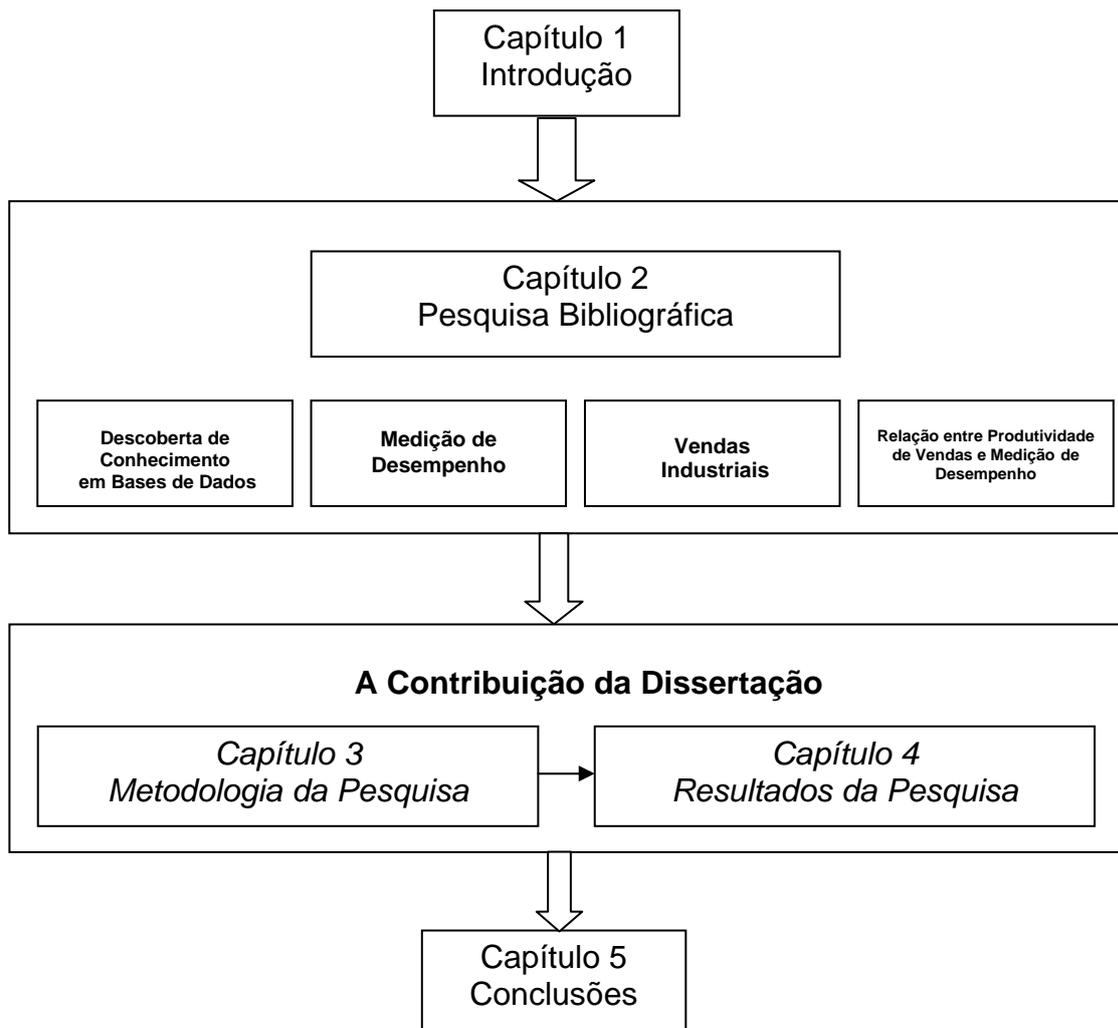


Figura 1.1 - Estrutura da dissertação.

Fonte: Elaborada pelo autor.

Num primeiro grupo (Figura 1.1), intitulado Pesquisa Bibliográfica, existe a preocupação de contextualizar o trabalho desenvolvido, quer analisando a posição dos diferentes pesquisadores da área quer fazendo uma exploração tecnológica da lacuna existente. São apresentados os objetivos, conceitos, tipos de abordagem, áreas relacionadas, metodologias e especificações, para a Descoberta de Conhecimento em Bases de Dados, Medição de Desempenho, a Relação entre Produtividade de Vendas e Medição de Desempenho. Por último, as Vendas Industriais com particular ênfase naquelas que foram usadas para a aquisição de conhecimento na aplicação prática na verificação da existência ou não de relacionamento de dependência entre os indicadores de desempenho de vendas e a sua produtividade bem como determinar a importância desses indicadores, os tipos de relações eventualmente encontradas ou as inter-relações entre eles.

São apresentadas diferentes abordagens fazendo uma comparação entre as visões dos autores dos processos, para a criação de um modelo de processo para ser utilizado no decorrer desse estudo. Nesse capítulo, também são brevemente mostradas as técnicas mais utilizadas para Mineração de Dados.

O capítulo 3 denominado Metodologia da Pesquisa serve para a apresentação de outra contribuição da dissertação que consiste em estruturar a utilização de BD num contexto comercial. O trabalho de pesquisa é desenvolvido no sentido de propor um conjunto de atividades que, de um modo sistematizado, suportem o desenvolvimento de projetos de Database Comercial baseados na Descoberta de Conhecimento em Bases de Dados.

A utilização prática do sistema apresentado é concretizada no capítulo 4 por intermédio da sua aplicação na área de desempenho comercial em telecomunicações em um caso de demonstração real e discussão dos resultados obtidos baseado em técnicas de Mineração de Dados. São também implementados modelos para a análise de relacionamento de dependência entre os indicadores de desempenho e a produtividade de vendas, da determinação da importância desses indicadores, os tipos de relações eventualmente encontradas ou as inter-relações entre os mesmos.

Por último, no capítulo 5 são apresentadas as conclusões ao trabalho desenvolvido, identificando-se as principais contribuições para a área de Gestão de Negócios em Telecomunicações, nomeadamente no âmbito da utilização da Descoberta de Conhecimento em Bases de Dados, sendo lançadas linhas orientadoras para o trabalho a desenvolver no futuro.

Pesquisa Bibliográfica

São apresentados os objetivos, conceitos, tipos de abordagem, áreas relacionadas, metodologias e especificações, para a Descoberta de Conhecimento em Bases de Dados, Medição de Desempenho, o relacionamento entre Produtividade de Vendas e Medição de Desempenho e, por último, as Vendas Industriais com particular ênfase naquelas que foram usadas para a aquisição de conhecimento na aplicação prática na verificação da existência ou não de relacionamento de dependência entre os indicadores de desempenho de vendas e a sua produtividade bem como determinar a importância desses indicadores, os tipos de relações eventualmente encontradas ou as inter-relações entre eles.

2.1 Descoberta de Conhecimento em Bases de Dados

2.1.1 Introdução

O avanço das *Tecnologias de Informação* (TI) bem como a sua penetração nas organizações, proporcionou o armazenamento e acesso a grandes volumes de dados, criando uma excelente oportunidade para a obtenção de conhecimento. Contudo, a transformação dos dados em conhecimento útil é um passo moroso e difícil.

As primeiras abordagens na aplicação de técnicas para a extração de conhecimento em *Bases de Dados (BD)* enfrentaram bastantes dificuldades devidas, principalmente, pelo fato dos algoritmos existentes terem sido desenhados para aplicações de laboratório, onde, em geral, a qualidade dos dados era garantida e a quantidade de dados bastante reduzida.

Tornaram ainda evidente a necessidade de seguir uma abordagem sistemática e fortemente centrada no processo de preparação dos dados, que permitisse aumentar a confiança no resultado final. Essa abordagem sistemática, integrando fases de pré-processamento dos

dados e pós-processamento dos resultados, foi denominada *Knowledge Discovery in Data Bases (KDD)*, termo que é traduzido para *Descoberta de Conhecimento de Bases de Dados*. KDD consiste numa série bem definida de passos que vão desde a preparação dos dados até à extração de padrões sobre esses dados e à avaliação dos mesmos.

A técnica de KDD vai buscar conhecimento em outras áreas, como a Estatística, os BD, a Inteligência Artificial, a Visualização de Dados e o Reconhecimento de Padrões. As técnicas desenvolvidas nessas áreas de estudo são utilizadas em KDD com o objetivo de extrair conhecimento dos BD.

Como não poderia deixar de ser, dadas as diferentes perspectivas com as quais a área de KDD pode ser abordada, na literatura podem ser encontradas diversas caracterizações da área. No artigo (ZHOU, 2003), o autor evidencia a caracterização da área sob as perspectivas tratadas em três livros sobre KDD avaliados por ele, sendo um de cada uma das três áreas. Sob a perspectiva da área de Base de Dados, citada em (HAN & KAMBER, 2001) KDD é “o processo de descoberta de conhecimento interessante em grandes quantidades de dados armazenados em Bases de Dados, *Data Warehouses* ou outros arquivos de dados”; sob a perspectiva da área de Aprendizado de Máquina, conforme apontada em (WITTEN & FRANK, 1999) é caracterizada como a “extração de conhecimento implícito, previamente desconhecido e potencialmente útil a partir de dados”; e sob a perspectiva da área de Estatística, conforme citado em (MANNILA, 1997), é “a análise de conjuntos de dados supervisionados, normalmente em grandes quantidades, para encontrar relacionamentos inesperados e resumir os dados em novas formas que são compreensíveis e úteis para o proprietário dos dados”.

A literatura sobre ‘Mineração de Dados’ e ‘Descoberta de Conhecimento em Bancos de Dados’, propõe diferentes abordagens na definição desses termos. A procura por padrões nos dados é definida de diferentes formas de acordo com o autor. Essas diferentes abordagens são apresentadas a seguir na seção 2.2. As abordagens mais voltadas a estatísticos, profissionais de tecnologia da informação, analistas de dados e de negócios, chamam o processo de KDD de ‘Mineração de Dados’. Na visão dos profissionais da inteligência artificial o processo é denominado de Descoberta de Conhecimento em Bancos de Dados e Mineração de Dados é apenas a etapa onde são utilizados os algoritmos para extração de conhecimento. Na caracterização da área de Mineração de Dados é importante, também, discutir a caracterização do que na literatura é chamado de KDD. De acordo com (W. FRAWLEY & PIATETSKY-SHAPIRO, 1992), KDD é a “extração de conhecimento previamente desconhecida, implícita e potencialmente útil, a partir de dados”. Na literatura existente atualmente as opiniões divergem

a respeito dos termos Mineração de Dados e KDD. Existem autores que consideram os termos sinônimos (FAYYAD et al., 1996) enquanto outros consideram a Mineração de Dados apenas um dos passos do processo de KDD, embora seja o passo principal de todo o processo (S. MITRA & PAL, 2002) e (I. SARAFIS et al., 2002).

KDD é uma área multidisciplinar que incorpora técnicas utilizadas em diversas áreas como Inteligência Artificial, especialmente Aprendizado de Máquina (AM), Base de Dados e Estatística. Por isso, as técnicas utilizadas em KDD não devem ser vistas como substitutas de outras formas de análises (por exemplo, OLAP), mas, como práticas para melhorar os resultados das explorações feitas com as ferramentas atualmente utilizadas.

O foco central de KDD é o de como transformar dados armazenados em conhecimento, expresso em termos de formalismos de representação, tal como regras e relações entre dados. Existe conhecimento que pode ser extraído diretamente de dados sem o uso de qualquer técnica, entretanto, existe também muito conhecimento que está de certa forma “embutido” na Base de Dados, na forma de relações existentes entre itens de dados que, para ser extraído, é necessário o desenvolvimento de técnicas especiais. Assim na próxima seção são definidos os conceitos básicos dados, informação e conhecimento para em seguida abordar o processo de KDD como um todo.

2.1.2. Dados, Informação e Conhecimento

Os conceitos de dados, informação e conhecimento estão interligados. Na Figura 2.1 é mostrada uma representação gráfica do relacionamento entre dados, informação e conhecimento, em função da capacidade de entendimento e da independência de contexto que cada um desses conceitos implica (KOCK Jr. et al., 1996).

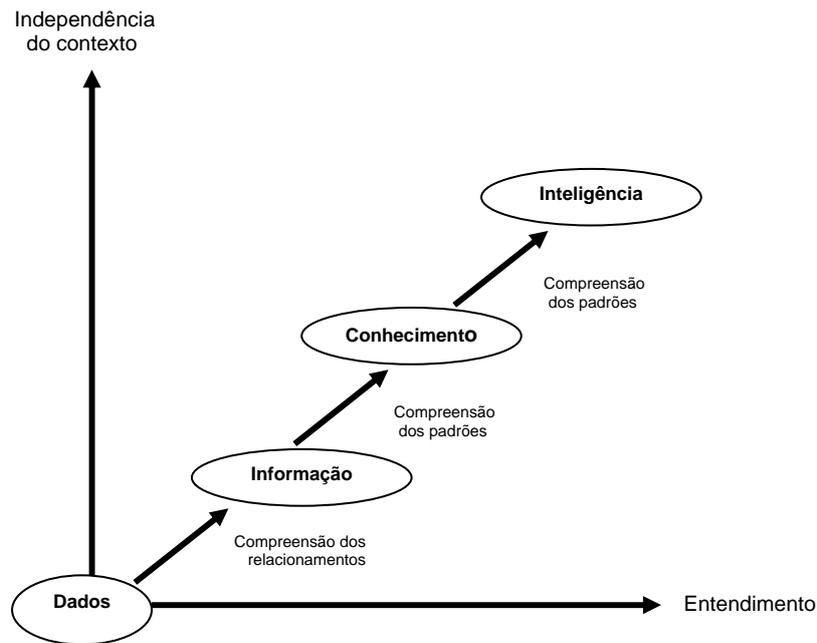


Figura 2.1 - Dado, informação e conhecimento.

Fonte: Adaptada pelo autor a partir de (KOCK Jr. et al., 1996).

Antes de estabelecer qualquer ligação desses conceitos com as diferentes tecnologias para seu registro e processamento, faz-se necessária uma breve discussão sobre a distinção entre dado, informação e conhecimento.

O *dado* é um elemento puro, quantificável sobre um determinado evento. Dados são fatos, números, texto ou qualquer mídia que possa ser processada pelo computador. Hoje em dia, as organizações estão acumulando vastas e crescentes quantidades de dados em diferentes formatos e em diferentes tipos de arquivos. Ressalta-se que o dado, por si só, não oferece embasamento para o entendimento da situação. Entre os dados armazenados atualmente estão:

- Dados operacionais ou transacionais como vendas, custos, inventários, folhas de pagamento ou contas correntes;
- Dados não operacionais como previsões de mercado, vendas a nível industrial e dados macro-econômicos;
- Meta dados ou dados descrevendo a estrutura dos dados armazenados, tais como projetos lógicos de Bases de Dados ou dicionários de dados;
- Mídia contendo imagens, sons ou uma combinação de ambos, que além de ser armazenada, pode ser catalogada eletronicamente.

A *informação* é o dado analisado e contextualizado. Envolve a interpretação de um conjunto de dados, ou seja, a informação é constituída por padrões, associações ou relações que todos aqueles dados acumulados podem proporcionar. Por exemplo, a análise de pontos de

equilíbrio no mercado pode fornecer informação acerca de quais produtos estão sendo vendidos e a frequência de tais operações.

A informação pode gerar conhecimento que ajude na análise de padrões históricos para conseguir uma previsão dos fatos futuros (pelo menos no contexto das variáveis que estão sendo envolvidas na análise). Por exemplo, a informação dos dados sumarizados nas vendas de serviços de telecomunicações pode ser analisada com a finalidade de fornecer informações relacionadas com a natureza dos clientes.

Enquanto que a informação é descritiva, o *conhecimento* é utilizado fundamentalmente para fornecer uma base de previsão com um determinado grau de certeza. O conhecimento refere-se à habilidade de criar um modelo mental que descreva o objeto e indique as ações a serem implementadas e as decisões a tomar. Uma decisão é o uso explícito de um conhecimento. O conhecimento pode ser representado como uma combinação de estruturas de dados e procedimentos interpretáveis que levam a um comportamento conhecido. Esse comportamento fornece informações que podem, então, serem utilizadas para planejar e decidir.

A compreensão, análise e síntese, necessárias para a tomada de decisões inteligentes, são realizadas a partir do nível do conhecimento. Assim, é fundamental que se mantenha a coerência dos dados que estão armazenados nos diferentes arquivos e das informações nos diferentes níveis.

Assim, nesse contexto, o desafio dos anos de 1980 foi migrar os dados para as informações, por meio do desenvolvimento dos Sistemas de Informação, que tinham por finalidade analisar os dados e organizar a informação para melhorar o processo decisório empresarial. A partir da década de 1990, o desafio era criar sistemas que fossem capazes de representar e processar conhecimento, em resposta as diferentes necessidades de indivíduos, grupos e culturas.

2.1.3 Conceitos Correlatos

Atualmente, os analistas de negócios precisam usar ferramentas capazes de responder as perguntas complexas como: “qual produto de alta lucratividade venderia mais com a promoção de um item de baixa lucratividade, analisando os dados dos últimos dez anos de vendas?”. Esse tipo de informação pode ser fundamental para oferecer vantagens competitivas às empresas. Dessa maneira, ferramentas que apóiam o processo de obtenção e análise das informações e extração de conhecimento devem ser usadas no processo decisório.

A-) Gestão do Conhecimento (*Knowledge Management* – KM)

A sociedade do conhecimento está impondo uma competitividade cada vez maior entre países e entre empresas, o que leva a uma necessidade de mudança e reflexão contínuas de que é preciso inovar e adquirir sucessivamente novos conhecimentos.

A gestão do conhecimento (*Knowledge Management* – KM) tem o objetivo de estabelecer meios, de maneira integrada e colaborativa, para capturar, criar, organizar e usar todos os ativos de informação de uma corporação. A gestão do conhecimento é o primeiro passo na criação de uma estrutura lógica para manipular as informações que uma determinada entidade possui e gerenciar tanto as entradas quanto os resultados da mesma. Em outras palavras, a gestão do conhecimento é responsável pela recuperação e organização das práticas, documentos, políticas, experiências de funcionários, entre outras fontes, de onde é possível obter conhecimento explícito de uma organização.

O foco principal é o conhecimento organizacional que é inerente a todas as empresas e é definido como “a capacidade de executar coletivamente tarefas que as pessoas não conseguem fazer atuando de forma isolada, tarefas essas projetadas para criar valor para as partes interessas na organização”, (GARVINL et al., 1998). Há uma diferença entre o conhecimento estar embutido em estruturas, regras e processos de trabalho em grupo – conhecimento explícito – e estar embutido em trabalhos individuais – conhecimento tácito. (GARVINL et al., 1998) acreditam que o conhecimento organizacional deve ser explícito e tácito, pois o conhecimento tácito, que inclui o discernimento, o instinto e a compreensão individual, é fundamental para tornar o conhecimento explícito útil.

A utilização da Tecnologia de Informação (TI) para a gestão do conhecimento tem seus primórdios nos anos 70, quando esta passa de um foco computacional voltado ao processamento de dados para um foco mais voltado ao processamento da informação, como nos sistema de suporte a decisão gerencial (DSS – *Decision Support System*) e nos sistemas de informação gerencial (MIS – *Management Information System*).

Nos anos 80, o processamento do conhecimento passa a estar cada vez mais presente nos recursos oferecidos, com os Sistemas Baseados em Conhecimento (KBS's – *Knowledge Based Systems*). Talvez a face mais visível e conhecida desses sistemas sejam os sistemas especialistas (*Expert Systems*). Atualmente os DSS e os MIS desdobram-se em inúmeras linhas de atuação, desde a utilização de sistemas baseados em Inteligência Artificial, modelos matemáticos e estatísticos, passando pela criação de conhecimento a partir de dados e informações presentes em Base de Dados (Mineração de Dados e *Data Warehousing*), até a

representação do conhecimento em sistemas especialistas e redes neurais que procuram automatizar a tomada de decisões.

Entretanto, é necessário destacar que, antes de investimentos ou decisões em TI, a empresa já deveria saber identificar, desdobrar e atribuir o devido valor aos conhecimentos disponíveis, respeitar e motivar o trabalho em rede dos grupos voluntariamente formados dentro da organização e desses com os parceiros da empresa, e saber aprender com a experiência (no sentido de refletir e aprender a incorporar as mudanças e atualizações, conseqüência da natureza dinâmica do conhecimento). Além disso, a organização precisa estar ciente de que muitos dos recursos de TI requerem significativos e continuados esforços para se construir, manter e atualizar, tanto em seus aspectos funcionais quanto nos conteúdos envolvidos, bem como é necessário conquistar a aceitação e confiança do usuário corporativo.

B-) Inteligência de Negócios (*Business Intelligence* – BI)

Os sistemas de Inteligência de Negócios utilizam os dados disponíveis nas organizações para disponibilizar informações relevantes para a tomada de decisão. Combinam um conjunto de tecnologias que permitem a geração de relatórios, para produzir informações que são utilizadas pela gestão das organizações, no suporte à tomada de decisão. A informação que essas tecnologias colocam a disposição das organizações irá tornar-se um componente essencial do entendimento que os agentes organizacionais partilham sobre o negócio e a possibilidade que têm de agir para criar condições internas e externas favoráveis ao sucesso da organização.

B.1) Definição

O conceito de Inteligência de Negócios (*Business Intelligence* – BI), de forma mais ampla, pode ser entendido como a utilização de variadas fontes de informação para definir estratégias de competitividade nos negócios. Os sistemas de *Business Intelligence* combinam dados com ferramentas analíticas de forma a disponibilizar informação relevante para a tomada de decisão. O objetivo desses sistemas é melhorar a disponibilidade e qualidade dessa informação (CODY et al., 2002). Os sistemas de *Business Intelligence* combinam a obtenção de dados, o armazenamento dos mesmos e a gestão de conhecimento com diversas ferramentas de análise que permitem extrair informação útil, a partir dos dados armazenados. As tarefas normalmente associadas ao *Business Intelligence* são:

- Elaborar previsões baseadas em dados históricos, nos desempenhos passados e atuais da organização;
- Criar cenários que evidenciam o impacto da alteração de diversas variáveis;
- Permitir o acesso *ad-hoc* aos dados para responder a questões não pré-definidas;
- Analisar detalhadamente a organização, obtendo um conhecimento mais profundo da mesma.

Os sistemas de *Business Intelligence* estão tradicionalmente associados a três tecnologias: *Data Warehouse*–DW (Armazém de Dados), *On-Line Analytical Processing*–OLAP (Processamento Analítico On-Line) e *Data Mining*–DM (Mineração de Dados). Esses sistemas têm aplicado à funcionalidade, escalabilidade e segurança dos atuais sistemas gestores de bases de dados para construir DW que são analisados com técnicas de OLAP e de DM. Essas análises visam à obtenção de informação relevante para o negócio pela exploração de grandes quantidades de dados armazenados nas bases de dados organizacionais.

Numa perspectiva organizacional, salienta-se ainda contribuições que indicam o BI como utilizador da informação que diz respeito ao ambiente do negócio onde a organização opera. Quando desenvolvido, o BI permite à empresa sustentar a tomada de decisões, conferindo-lhe vantagem competitiva (PRIOR, 1998). A *Society of Competitive Intelligence Professionals* coloca BI entre a Inteligência Competitiva (*Competitive Intelligence*) e a Gestão do Conhecimento, abordando BI como uma combinação de dados, informação e conhecimento acerca do ambiente de negócio, capaz de conferir uma significativa vantagem competitiva ou facilitar as decisões a tomar (ARNETT et al., 2000).

Em síntese, é possível definir BI como um processo que combina atividades de obtenção, armazenamento de dados (DW ou OLAP) e extração de conhecimento (DM), tornando a organização mais competitiva. Como ilustração descritiva de BI, num sistema ideal, todos os funcionários teriam no monitor do seu computador, a informação de que necessitam, em tempo real (relativa aos mercados, aos consumidores, aos produtos ou mesmo financeira), de modo a poderem analisar, partilhar com outros gestores e tomar as decisões necessárias.

B.2) Tecnologias de Suporte ao *Business Intelligence*

Os sistemas de *Business Intelligence* podem ser suportados por diversas tecnologias. As mais comuns, e como já referidos anteriormente, estão associadas aos sistemas de DW, OLAP e DM, os quais são em seguida brevemente descritos. A Figura 2.2 apresenta a arquitetura de

uma infra-estrutura tecnológica de apoio ao *Business Intelligence*. Esta arquitetura, proposta por Han e Kamber (HAN & KAMBER, 2001), apresenta três níveis: o nível do servidor de DW, o nível do servidor OLAP e o nível das ferramentas de análise, no qual pela utilização de gráficos, relatórios ou modelos (identificados, por exemplo, com técnicas de DM) é disponibilizada informação relevante para a tomada de decisão nas organizações.

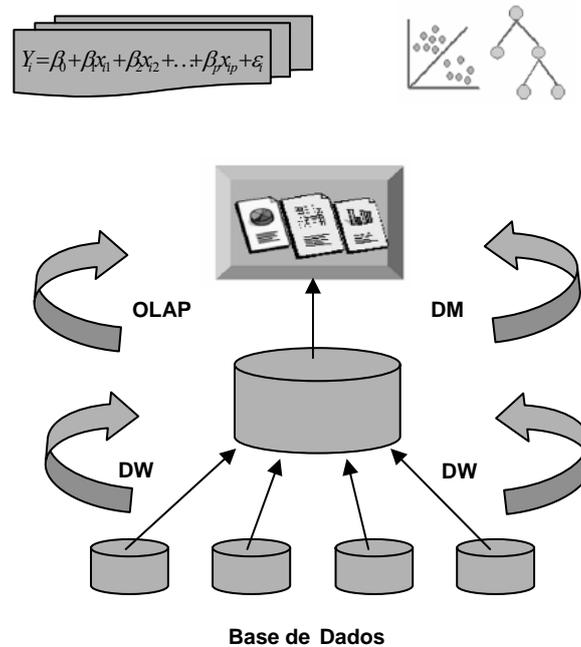


Figura 2.2: Infra-estrutura tecnológica de apoio ao Business Inteligente.

Fonte: Adaptada pelo autor a partir de (HAN & KAMBER, 2001).

DW: Data Warehouse (Armazém de Dados)

Um DW é um depósito construído especificamente para a consolidação da informação da organização num formato válido e consistente, permitindo aos seus gestores a análise de dados de uma forma seletiva. O conceito de *Business Intelligence* integra a atividade de exploração do DW, incluindo consultas pré-definidas, consultas *ad-hoc* e elaboração de relatórios, que normalmente permitem a monitorização da evolução ocorrida nos principais indicadores de negócio ao longo da vida da organização.

Um DW é uma base de dados que é mantida de uma forma autônoma em relação às bases de dados operacionais da organização. Toda a informação armazenada num DW é etiquetada temporalmente e deve permitir armazenar dados de vários anos. Os gestores, analistas de informação, podem colocar questões complexas no DW, libertando as bases de

dados operacionais para as tarefas que ditaram a sua implementação, isto é, a obtenção, o armazenamento e a manipulação da informação do dia a dia da organização (INMON, 2003).

De forma resumida, um DW integra um conjunto de dados orientado por assunto, catalogado temporalmente e não volátil que suporta os gestores no processo de tomada de decisão. As principais características desses depósitos são de serem orientados para os principais assuntos da organização (compras, vendas, encomendas, entre outros), integrarem dados provenientes de diversas fontes (bases de dados organizacionais), catalogarem temporalmente os dados permitindo a análise da informação sob uma perspectiva histórica e serem não voláteis, permitindo o carregamento de dados e a sua posterior análise (operações de eliminação de registros, por exemplo, não são permitidas).

Além dos DW existem os *Data Marts*. Esses depósitos apresentam as mesmas características de um DW, com a diferença de que são mais restritos, no que diz respeito aos dados que armazenam, retratando apenas parte da organização, como por exemplo, um departamento específico ou um determinado processo de negócio. A existência desses dois tipos de depósitos de dados pressupõe que a organização deve identificar qual é a arquitetura que melhor satisfaz as suas necessidades. Assim, uma organização pode optar pela implementação de um DW organizacional, pela implementação de *Data Marts* independentes ou pela implementação de *Data Marts* dependentes do DW organizacional.

OLAP: On-Line Analytical Processing (Processamento Analítico On-line)

O OLAP faz análises multidimensionais dos dados depositados em um armazém de dados. Essas análises buscam padrões em diferentes níveis de abstração, ou seja, uma visão lógica dos dados. OLAP é uma análise interativa, permitindo ilimitadas visões por meio de agregações em todas as dimensões possíveis. Permite obter informações e mostrá-las em tabelas de 2D e 3D, mapas e gráficos. Além disso, derivam-se análises estatísticas (razões, médias, variâncias) envolvendo quaisquer medidas ou dados numéricos entre muitas dimensões. Uma consulta OLAP é executada com um tempo de resposta pequeno, pois é manipulada em um Sistema Gerenciador de Bancos de Dados com otimização para DW.

Os sistemas DW e OLAP são tecnologias complementares. Um DW armazena e gerencia os dados enquanto que o OLAP transforma os dados de um DW em informação estratégica. OLAP abrange desde a navegação básica e cálculos até análises de séries temporais e modelagem complexa. Os sistemas OLAP são utilizados, entre outros, para explorar DW. Permitem criar cubos para analisar a informação sob diferentes perspectivas. Os cubos

permitem analisar os fatos disponíveis, indicadores de negócio, pelas diferentes dimensões consideradas na modelação do negócio realizada (clientes, produtos, tempo, etc.).

DM: Data Mining (Mineração de Dados)

Os princípios associados à Descoberta de Conhecimento em Bases de Dados (KDD) conjugam fundamentos provenientes de diversas áreas, tais como a estatística, a inteligência artificial, a aprendizagem automática, as bases de dados, os sistemas de informação, entre outras. As aplicações disponíveis têm como objetivo a extração de conhecimento partir de grandes bases de dados (FAYYAD, et al., 1996).

Os algoritmos utilizados para procurar padrões nos dados são denominados de algoritmos de DM. O processo global de KDD, que se desenvolve em várias fases, inclui a utilização dos algoritmos de DM a interpretação dos padrões encontrados pelos mesmos, os quais são posteriormente utilizados no suporte à tomada de decisão. As técnicas de análise de dados do DW geralmente não extrapolam a realização de consultas SQL (*Structured Query Language*) simples, a utilização de ferramentas OLAP ou os mecanismos de Visualização de Dados. Por meio desta forma de análise de dados, algumas questões importantes para tomada de decisão podem ser inviáveis de serem expressas, como:

- Quais são os usuários potenciais para praticar fraude?
- Quais clientes gostariam de comprar um novo produto de telecomunicações?

As ferramentas utilizadas para analisar um DW, normalmente, são orientadas as consultas, ou seja, são dirigidas pelos usuários, os quais possuem hipóteses que gostariam de comprovar, ou simplesmente, executam consultas aleatórias. Essa abordagem dependente do usuário pode impedir que padrões escondidos nos dados sejam encontrados de forma “inteligente”, uma vez que o usuário não terá condições de imaginar todas as possíveis relações e associações existentes em um grande volume de dados. Por isso, faz-se necessária a utilização de técnicas de análise dirigidas por computador que possibilitem a extração automática (ou semi-automática) de novos conhecimentos a partir de um grande arquivo de dados (BRADLEY et al., 1998). A extração automática de conhecimento a partir de dados, denominada Mineração de Dados, é abordada, com o nível de detalhe necessário, no item 2.1.6.D.

B.3) Potencialidade de *Business Intelligence*

Os sistemas de *Business Intelligence* permitem às empresas construir conhecimento sobre o comportamento dos seus clientes e outras entidades externas com que interagem. Esses sistemas permitem ainda construir conhecimento relativo ao funcionamento interno da empresa.

Desde que exista informação histórica sobre as atividades da empresa, ela poderá ser sujeita às análises possibilitadas pelas tecnologias OLAP e DM. Essas análises podem constituir-se num apoio fundamental para a formulação de estratégias de fidelização dos bons clientes da empresa, maximizando os lucros que daí advêm e evitando os custos com campanhas mal direcionadas. Por outro lado, as análises à informação sobre os processos internos da empresa poderão ajudar os decisores a descobrir ineficiências e oportunidades de inovação, apoiando a redefinição de práticas de trabalho de forma a assegurar que as metas e objetivos da empresa sejam atingidos. Na chamada economia do conhecimento, a vantagem competitiva de uma empresa está principalmente ligada à sua capacidade para ir ao encontro das necessidades e preferências dos consumidores, à sua flexibilidade para se ajustar às alterações dos mercados, e à sua capacidade para influenciar o comportamento do mercado pela inovação de processos, produtos e serviços.

Os sistemas de *Business Intelligence* são imprescindíveis a decisores dinâmicos e motivados para aproveitar as oportunidades que uma sociedade em transformação rápida pode oferecer, sendo particularmente úteis numa óptica positiva e empreendedora em que se procura explorar as janelas de oportunidade que se abrem ao negócio para maximizar os lucros da empresa.

B.4) Aplicações Organizacionais de BI

As aplicações de BI são vastas e têm aplicação em diferentes níveis da estrutura organizacional. Segundo a *Society of Competitive Intelligence Professionals*, o envolvimento nas operações de BI permite à organização (ARNETT et al., 2000):

- *Antecipar e gerir o risco*, analisando as potenciais evoluções que estejam ocorrendo no meio envolvente, por meio de uma análise criativa dos pontos fortes e fracos do ambiente econômico, extraindo assim potenciais ameaças e minimizando os potenciais riscos, ligados à incerteza;

- *Organizar* as informações dispersas, extraindo somente as que possam ser relevantes para a empresa, e construir cenários para ajudar os gestores na tomada das decisões;
- *Inovar* pela detecção de oportunidades e de novos mercados, analisando de forma criativa os dados obtidos durante a pesquisa e que se tenham tornado pertinentes para a organização;
- *Agir oportunamente* face aos concorrentes, conhecer o mercado, analisando quais as potencialidades de crescimento, quais os produtos que mais se adaptem às necessidades dos clientes;
- *Posicionar-se estrategicamente no mercado* adquirindo um conhecimento aprofundado do meio envolvente à organização. É necessário conhecer bem os atores em todas as suas dimensões: nível de competitividade, o número de concorrentes e respectiva quota de mercado, capacidade financeira, produtos concorrentes, procedimentos, tecnologia de que dispõem, e, se possível, estratégia; sendo viável desse modo explorar as fraquezas dos concorrentes, por meio do seu acompanhamento contínuo;
- *Criação de BD* com toda a informação referente às atividades interna e externa da organização, relativo a concorrentes, a histórico de ações e resultados anteriores;
- *Antecipar* as mudanças nos mercados assegurando as informações relevantes sobre as mudanças em curso e determinar sobre elas, quais representam ameaças ou oportunidades.

Em resumo, as organizações onde se desenvolvem e aplicam as atividades de BI se beneficiam de um sistema capaz de integrá-las no meio envolvente onde operam (obtenção de informações do exterior) e antecipar o comportamento dos mercados permitindo a adoção de estratégias competitivas (processos de detecção de oportunidades e análise da concorrência).

É importante realçar que esses sistemas devem ser vistos apenas como ferramentas de apoio que dentro de circunstâncias organizacionais e de gestão adequadas favorecem a inteligência, aprendizagem e criatividade organizacional. Sem políticas adequadas de gestão de recursos humanos, gestão estratégica e operacional, sem a existência de ambiente de trabalho que favoreçam a colaboração, a comunicação, a aceitação do risco e a tolerância ao erro, esses sistemas podem nunca cumprir os objetivos para os quais foram adotados e tornarem-se encargos financeiros demasiadamente pesados para a maioria das organizações.

2.1.4 Princípios da Descoberta de Conhecimento em Base Dados

Descobrir conhecimento significa extrair, de grandes volumes de dados, informações relevantes e até então desconhecidas, que se revelam úteis e válidas para processos de tomada de decisão. Recorrendo à definição elaborada por Usama Fayyad, KDD pode ser definida como “*um processo iterativo não trivial de identificar novos padrões nos dados que sejam válidos, potencialmente úteis e interpretáveis*” (FAYYAD et al. 1996):

- O termo *processo não trivial* encontra-se associado à execução de diversos passos interativos (requer a intervenção do analista em cada uma das fases do processo) e iterativos (em cada fase do processo existe sempre a possibilidade de retrocesso para fases anteriores);
- Os *dados* representam um conjunto de fatos F , casos de um BD, na qual subconjuntos do mesmo são responsáveis pela caracterização de diversos padrões.
- Um *padrão* pode ser caracterizado por modelos, relações ou estruturas existentes nos dados, que se revelam perceptíveis após o processamento. Um padrão é uma expressão E numa linguagem L que descreve um subconjunto de fatos FE do conjunto F . Por exemplo, em relação aos dados sobre empréstimos bancários, o padrão $EI = “Se Salário < T$ Então a pessoa faltou ao pagamento” poderia ser um padrão para uma escolha apropriada de T ;
- Os padrões encontrados devem manter-se *válidos* quando aplicados sobre novos dados, com algum grau de confiança (C) ou probabilidade. O grau de confiança associado a um padrão E pode ser definido como a função $C(E, F)$ que pode ser lógica ou real;
- A *novidade* (N) dos padrões é definida sempre que o conhecimento obtido ainda não foi detectado por nenhuma outra abordagem, podendo ser representado como função $N(E, F)$;
- A *utilidade* dos padrões representa o seu grau de utilização (U), isto é, até que ponto o padrão contribui para os objetivos inerentes ao processo como, por exemplo, o esperado aumento de lucro de uma base por aplicação da regra de decisão EI . A utilidade pode ser definida pela função $U(E, F)$;
- Um dos objetivos de KDD é gerar padrões que sejam compreendidos pelos humanos na perspectiva de contribuir para uma melhor compreensão dos dados. Assume-se que o grau de *interpretação* (S) de um padrão é definido pela função $S(E, F)$.

O conhecimento descoberto pode também ser quantificado, seja $i=I(E,F,C,N,U,S)$ o grau de interesse (I) num dado padrão E , diz-se que o padrão E é conhecimento se para um valor dado i , $I(E,F,C,N,U,S) > i$.

Em que pese às funções de avaliação apresentadas, a validade do processo de KDD depende diretamente da ação do analista, dado o fato de não existir ainda nenhuma solução tecnológica capaz determinar autonomamente se os padrões encontrados são de fato novos ou válidos.

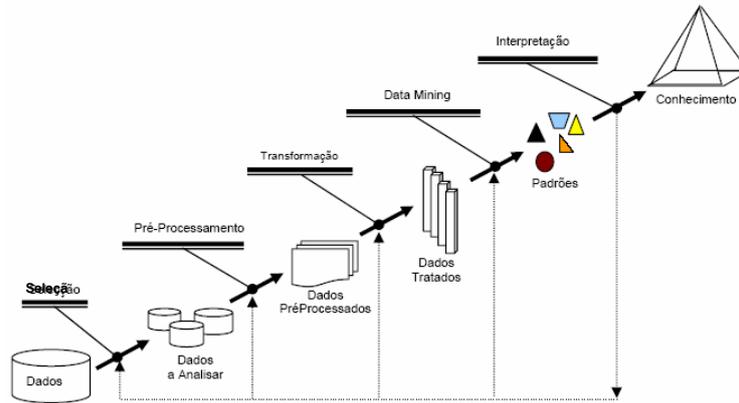


Figura 2.3 - Fases do processo de KDD.

Fonte: Adaptada pelo autor a partir de (FAYYAD et al., 1996).

A Figura 2.3 expõe uma representação do processo de KDD a partir do momento em que os objetivos estão definidos. O processo normalmente não é linear, e envolve uma forte interação com o analista e várias iterações entre as suas fases constituintes (em geral, ao analisar os padrões obtidos, o analista pode concluir de que não são válidos e fazer com que o processo volte à fase de pré-processamento). Essas iterações estão representadas, nesta figura, pelas setas tracejadas e podem inclusivamente fazer o processo voltar à fase de especificação dos objetivos.

O tempo gasto num processo de KDD não é distribuído equitativamente entre as várias fases, conforme pode ser ilustrado com a Figura 2.4. Ao contrário do que seria de esperar, a fase onde é despendido menos tempo é normalmente na fase de Mineração de Dados (DM), sendo, contudo uma fase que exige um bom conhecimento dos algoritmos disponíveis. Esse conhecimento é determinante para a afinação dos algoritmos, ajustando os parâmetros desses com vista ao sucesso do processo de extração de padrões (CABENA et al., 1998).

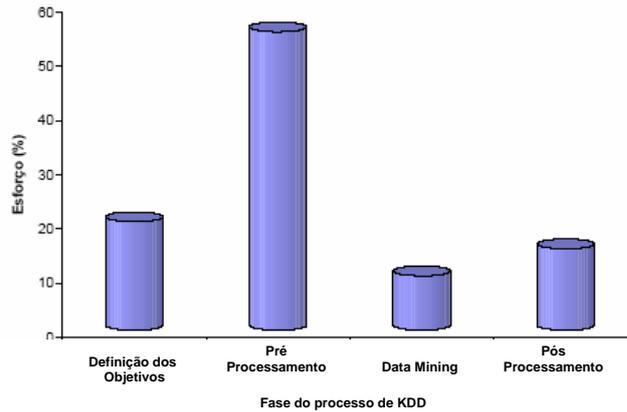


Figura 2.4 - Percentagem de tempo despendido em cada fase do processo de KDD.

Fonte: Adaptada pelo autor a partir de (CABENA et al., 1998).

KDD fornece métodos e ferramentas que permitem descobrir o conhecimento contido em grandes BD. Um exemplo típico da utilização da KDD encontra-se nas ações de envio de correio publicitário (ADIAANS et al., 1996).

O problema pode, genericamente, ser descrito do seguinte modo: uma empresa pretende desenvolver uma ação de marketing direto procedendo ao envio de correio publicitário para casa de cada um dos seus clientes, mas devido ao custo unitário, deseja realizar o contato apenas com aqueles indivíduos mais receptivos, maximizando a taxa de respondentes (relação entre cartas enviadas e respostas obtidas). Partindo então das BD onde se encontram registrados os comportamentos em campanhas publicitárias realizadas no passado, é possível por meio desses elaborar um modelo descritivo dos indivíduos com maior propensão para reagir positivamente; isto é, recorrendo à indução de regras, é possível extrair das BD regras que indiquem as características dos indivíduos que responderam à publicidade remetida no passado e com as mesmas selecionar indivíduos com características semelhantes (supostamente mais susceptíveis de responder positivamente a novas campanhas). Na Figura 2.5 encontra-se uma representação gráfica que relaciona o número de cartas enviadas e o número de respostas obtidas (com e sem KDD). Normalmente a percentagem de respostas a campanhas desse gênero é extremamente baixa, rondando 1%. São citados casos na literatura em que pelo recurso à KDD, foi possível obter mais de 2% de respostas. Para uma maior facilidade de se compreender o impacto dos ganhos obtidos com esses métodos, considere-se o cenário de existir um orçamento global de R\$ 6.000,00 para a remessa de correio publicitário, pelo que com o custo de cerca de R\$ 0.15 por carta enviada, seria possível contatar aproximadamente 40.000 indivíduos. Seguindo o exemplo bibliográfico referido, não existindo recurso a métodos de KDD é previsível que se obtenham cerca de 400 respostas. Pelo contrário, no caso de serem

utilizados métodos de KDD para selecionar quem vai receber a publicidade, é provável que as mesmas 40.000 cartas dêem origem a cerca de 900 respostas (Figura 2.5).

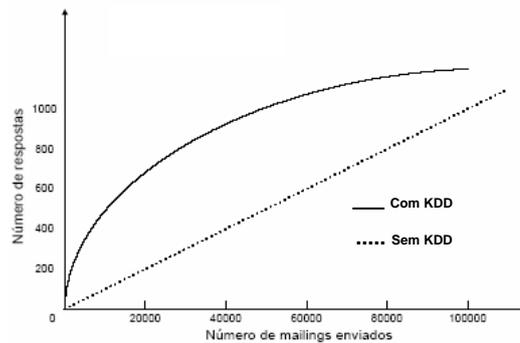


Figura 2.5 - Relação entre o número de mailings e o número de respostas.

Fonte: Adaptada pelo autor a partir de (ADRIAANS et al., 1996).

2.1.5 Fases do Processo de KDD

Conforme já ilustrado na Figura 2.3 as fases do processo de KDD incluem:

- Exploração e Seleção de Dados;
- Pré – Processamento
- Transformação dos Dados;
- Mineração de Dados;
- Interpretação dos Resultados Alcançados.

A-) Exploração e Seleção de Dados

A fase de seleção compreende duas componentes: estudo e compreensão do domínio da aplicação, e seleção dos dados a analisar. No estudo e compreensão do domínio, deve-se enquadrar o trabalho a desenvolver com a área de negócio da organização, pela aquisição de conceitos fundamentais e da definição clara dos objetivos para o projeto. O conhecimento do domínio é determinante em qualquer processo de KDD servindo como elemento condutor, podendo o conhecimento existente ser complementado com o conhecimento obtido no processo de descoberta. Por esse fato, torna-se necessária a presença, na equipe de desenvolvimento do processo de KDD, de especialistas na área de aplicação. Essas equipes são, por norma multidisciplinar, integrando especialistas, ou seja, da área de negócio, técnicos de BD ou especialistas em técnicas de DM.

Antes de proceder à seleção dos dados é necessário explorá-los e interpretá-los. A exploração e interpretação visam o conhecimento dos dados disponíveis para a realização do

processo de KDD. O acesso aos registros das tabelas não garante que os dados sejam acessíveis sem que se compreenda o contexto das tabelas e o significado de cada atributo, advindo daí a necessidade compreensão quer das estruturas (significado da designação atribuída a tabelas e atributos) quer dos conteúdos (em geral, significado dos valores de cada atributo). A origem dos dados pode ser interna (sistemas da empresa ou de outra da mesma organização) ou externa (BD adquiridos a empresas especializadas), podendo a fonte ser documental (formulários, fichas de cliente) ou digital (sistemas informáticos de vendas).

Após uma compreensão dos dados, é então possível proceder à seleção dos que irão ser utilizados no processo de KDD, tendo em vista a limitação do espaço de pesquisa, direcionando o foco para subconjuntos de variáveis ou de dados. A seleção de dados incorpora ainda a função de filtragem de dados duplicados (normalmente ocorrem sempre que existem diversas fontes de dados envolvidas), ou seja, um indivíduo cujo registro se encontra duplicado porque havia participado em duas ou mais campanhas de marketing distintas ou porque o seu contato proveio de BD distintas.

B-) Pré-Processamento

Como acontece na maioria dos processos informáticos (e não só), o resultado final é bastante dependente do que é fornecido à entrada. A conhecida regra GIGO (*Garbage In, Garbage Out*) (FEELDERS, 2002) aplica-se na sua plenitude ao processo de KDD. O sucesso desse depende diretamente da qualidade dos dados sobre os quais se analisa.

Os dados, tal como a sua qualidade, são conceitos multidimensionais. Uma das perspectivas mais interessante acerca da qualidade dos dados adotada pela literatura evidencia o carácter de utilização dos dados “*data that is fit for use*”, (BROWN, 2002), ou em outras palavras, a elevada qualidade dos dados depende da sua utilização operacional, suporte a ações de tomada de decisão ou de planeamento, para as quais foram pensados. Os dados estão preparados para serem utilizados se estiverem livres de defeitos e possuírem as características desejadas. A qualidade dos dados no âmbito do processo de KDD é determinante, havendo autores que defendem apenas ser possível prosseguir com o processo de KDD se existirem dados limpos de erros e prontos a serem usados (SHEPARD, 1998) e (DROZDENKO et al., 2002), isto é por demais evidente na fase de modelagem, devendo por isso estar disponíveis numa forma que permita que sobre eles sejam aplicados algoritmos de DM (FAYYAD et al., 1996). Esta fase deve ser repetida várias vezes, até que se possa assegurar a qualidade e utilidade dos resultados obtidos.

O pré-processamento dos dados é considerado por muitos autores como a fase crucial para o sucesso do processo de KDD, tal como as estatísticas frequentemente a indicam como a fase mais demorada, chegando a consumir cerca de 80% do tempo total (FAYYAD et al., 1996) e inclui como atividades:

- Tratamento de valores omissos; Avaliação de exceções (*outliers*); Derivação de novos atributos;
- Dispersão de valores; Identificação de atributos duplicados e redundantes;
- Integridade da informação do registro; Identificação de ruído ou poluição dos dados;
- Discretização de atributos qualitativos em quantitativos; Identificação de inconsistências;
- Normalização dos dados; Identificação de valores atribuídos por definição (defaults) e
- Transposição de tabelas.

Essas atividades são em seguida descritas em pormenor:

B.1) Tratamento de Valores Omissos

A existência de valores em branco num determinado atributo suscita o tratamento desse atributo, por meio de técnicas diversas, com origens na estatística e matemática, ou seja, métodos *Bayesianos* e métodos de *Imputação*. A eliminação de registros com valores omissos é uma das soluções indicadas por um dos métodos de imputação, contudo, esta atitude pode introduzir na BD alterações significativas relativamente ao universo em estudo e conduzir a amostras muito pequenas. Na bibliografia foram encontrados outros métodos de imputação mais refinados:

- Substituição pelo valor mais comum do atributo;
- Substituição pelo seu valor médio ou mediana;
- Substituição por um valor resultante da aplicação do método do vizinho mais próximo.

A determinação da técnica a adotar depende do volume de registros com valores omissos e a sua representatividade em termos da BD geral. Uma adoção incorreta de técnicas para tratamento de valores omissos pode introduzir inconsistências na BD e afetar o desempenho dos algoritmos durante a fase de aprendizagem. As razões mais frequentes para a existência de valores omissos nos dados resultam normalmente de fatores processuais ou quando se tratam de dados obtidos a partir de questionários, os problemas podem também advir de recusa de resposta e/ou opções de resposta inadequadas.

B.2) Avaliação de Exceções (*Outliers*)

Verificam-se com alguma frequência nas BD valores anormais para alguns atributos, violando os domínios (conjunto de valores aceitos para esse atributo). O tratamento de exceções resolve-se por recurso a um tratamento manual ou por recurso a valores estatísticos (aplicação de métodos *Bayesianos* ou métodos de *Imputação*). A avaliação de exceções permite um conhecimento genérico sobre os dados, com o qual se desprezam eventos particulares (registro de casos pontuais fora do comum) não afetando outras análises que estejam a ser realizadas.

B.3) Derivação de Novos Atributos

A qualidade dos resultados obtidos na extração de informação em BD depende diretamente dos atributos sobre os quais são aplicados (BARANAUSKAS et al., 2003). Os resultados podem ser imprecisos ou excessivamente complexos, caso não sejam devidamente avaliados e preparados para serem usados.

A derivação de atributos consiste num processo de composição de atributos primitivos, cujo resultado se traduz em novos atributos, possivelmente relevantes para a descrição de um conceito ou objeto. A combinação pode ser resultado de cálculos aplicados sobre os atributos isto é, o atributo *data de nascimento* permite uma informação potencialmente mais relevante se derivar um novo atributo, a *idade*.

B.4) Dispersão de Valores

A dispersão de valores em atributos acontece: (i) porque a natureza assim o determina (em geral, peso, idade, altura); ou, (ii) embora seja um atributo cujo domínio esteja sujeito a uma regra de classes, verifica-se uma dispersão de classes indesejada (nº. filhos, nº. de carros). Muitas dessas situações geram uma tendência na amostra e dificultam o trabalho de alguns algoritmos de aprendizagem automática.

Os classificadores gerados a partir de um conjunto de dados com uma desproporção evidente de classes, apresentam um pior desempenho na classificação da classe minoritária em comparação com os classificadores gerados a partir do mesmo conjunto de dados, mas com uma proporção mais equilibrada de classes.

Como justificativa, encontram-se duas razões: a primeira deriva do fato das regras geradas para a classe minoritária serem baseadas em menos exemplos e conseqüentemente mais ajustadas – o classificador tenderá a aprender limites mais rígidos do conceito. Esse comportamento está relacionado com um problema já bem referenciado na área de aprendizagem: *small-disjuncts*. A segunda razão tem a ver com o fato de dadas às características do domínio existirem mais exemplos de teste da classe minoritária. A classe mais frequentemente prevista é a majoritária existindo uma maior probabilidade de classificar incorretamente exemplos da classe minoritária. Para tornar a distribuição de classes mais equilibrada, existem dois métodos básicos:

- *Under-sampling* – cria uma amostra menor do conjunto de exemplos da classe majoritária;
- *Over-sampling* – consiste em gerar casos a partir dos casos iniciais do conjunto, de forma a aumentar o número de casos da(s) classe(s) minoritária(s).

B.5) Integridade da Informação do Registro

Analisar a integridade dos dados consiste na verificação da coerência da informação ao longo dos registros e envolve, com alguma freqüência, uma análise subjetiva de alguns atributos, tornando esta atividade como uma das mais morosas de todo o pré-processamento. Como exemplo considere-se que quando no atributo *nome* aparece *Teófilo*, no atributo *sexo*, deverá constar, em condições normais, *masculino*.

Um caso especial de verificação dos dados consiste na identificação de casos extremos, os quais não são mais do que combinações de valores raros, mas válidos, que se verificam em simultâneo e estão dentro dos valores dos domínios para um grupo de atributos.

B.6) Identificação de Atributos Duplicados e Redundantes

A redundância ocorre com armazenamento em diversos atributos de informação idêntica, em geral, codificação diferente para conteúdos iguais em tabelas diferentes. A informação torna-se ainda redundante quando o conteúdo de alguns atributos é explicado pela combinação de outros, em geral, havendo os atributos *preço unitário* e *quantidade comprada*, o conteúdo do atributo *total de compra* poderá ser redundante uma vez que se consegue obter pela conjugação dos dois anteriores.

B.7) Identificação de Inconsistências

As inconsistências podem ocorrer quando dados diferentes são representados sob a mesma codificação ou quando o mesmo para os mesmos dados existem codificações distintas, em geral, o atributo *nome_colaborador*, destinado a receber nome de colaborador pode assumir os valores *TCM*, *Mattozo* ou *Teófilo Câmara Mattozo*, sendo que todos esses representam uma mesma pessoa.

B.8) Identificação de Ruído ou Poluição dos Dados

Entende-se aqui como ruído ou poluição dos dados a presença de dados distorcidos, os quais não representam valores verdadeiros. Existem diversas fontes de ruído ou poluição dos dados, uma delas consiste na introdução de valores estranhos (fora do conjunto de valores esperados) em atributos, por vezes com a expectativa de forçar o registro de informação para além do seu contexto. Outra fonte de ruído ou poluição dos dados advém da resistência humana em introduzir os dados corretamente, seja deixando os campos em branco, incompletos ou simplesmente com valores incorretos.

B.9) Identificação de Valores Atribuídos por Definição (*Defaults*)

A maioria dos sistemas gestores de BD permitem a especificação de valores pré-definidos para alguns atributos. Esses valores poderão, mais tarde, vir a gerar padrões significativos e como tal originarem alguns deslocamentos na interpretação dos dados, caso o analista não seja informado da sua existência. Os valores pré-definidos normalmente representam falta de informação em vez de informação relevante. A utilização dos valores pré-definidos pode ser comprometedora na análise de dados em particular quando o resultado final da análise pretende ser a previsão.

C-) Transformação dos Dados

A transformação dos dados consiste no tratamento dos dados com vista à aplicação dos algoritmos de DM, resolvendo assim eventuais limitações que esses possuam. As transformações mais comuns são apresentadas a seguir:

C.1) - Normalização dos Dados

Esse procedimento, à semelhança de outros referenciados, consiste em realizar uma transformação nos dados de modo a acelerar ou melhorar o processo de aprendizagem dos algoritmos e visa resolver um aspecto importante nos dados como é a sua escala de valores, ou seja, no tratamento de dois atributos *altura* e *distância*, a *razão* entre eles é diferente consoante à escala de medida utilizada em cada um deles.

Numa outra perspectiva, os atributos ao estarem em escalas ou domínios diferentes provocam problemas nos métodos de aprendizagem, pois podem eventualmente dar demasiada importância a um atributo com um domínio maior, ou seja, no caso do atributo a_1 pertencer ao conjunto $\{0...10\}$ e o atributo a_2 pertencer ao conjunto $\{2.100...4.010\}$, o algoritmo de aprendizagem utilizado pode atribuir uma importância inadequada ao atributo a_2 em consequência desse apresentar um domínio maior de valores. O escalonamento depende do tipo de dados:

- Entradas – o escalonamento das variáveis de entrada tem efeitos diversos conforme os algoritmos de aprendizagem considerados, de um modo particular, os algoritmos de gradiente descendente (*Back-Propagation*), são bastante sensíveis ao escalonamento;
- Saídas – sempre que se usa mais do que uma saída e se a função de erro é sensível à escala, como acontece no caso da aprendizagem do gradiente descendente, então a diferença de escalas entre as saídas pode afetar a forma como, por exemplo, uma rede neural aprende. No caso de uma saída possuir valores entre 0 e 1, enquanto outra tem valores entre 0 e 1.000.000, o algoritmo irá despende a maior parte do esforço de aprendizagem na segunda saída. Assim, as saídas com a mesma importância devem ser transformadas para a mesma escala de valores.

C.2) - Discretização de Atributos Quantitativos em Qualitativos

Alguns algoritmos possuem como limitação o fato de apenas trabalharem com dados do tipo qualitativo, implicando que na presença de dados do tipo quantitativo seja necessário proceder a transformação dos valores. A transformação de dados realiza-se pela aplicação de correspondências entre grupos de valores quantitativos a valores qualitativos, ou seja, ao atributo *rendimento per capita*, seria possível atribuir classe social para valores como sejam,

para intervalos *até R\$ 2.000,00*, classe baixa; *entre R\$ 2.000,00 e R\$ 7.000,00*, classe média; *mais de R\$ 7.000,00*, classe alta.

C.3) - Discretização de Atributos Qualitativos em Quantitativos

Ao contrário do caso anterior, existem algoritmos que apenas aceitam valores quantitativos na entrada, forçando a transformação dos valores qualitativos iniciais em valores aceitáveis. A transformação desses valores ocorre por meio do estabelecimento de correspondências entre os dois tipos de valores, ou seja, para o atributo *prioridade*, cujos valores poderiam ser: grande, média e baixa poder-se-ia fazer corresponder, respectivamente os valores 1,2 e 3.

C.4) Transposição de Tabelas

A utilização dos dados pelos algoritmos durante a fase de modelagem, ocorre normalmente sobre uma tabela só. Esse fato implica na importação dos dados de várias tabelas para uma única, forçando que todos estejam ao mesmo nível, na mesma tabela e que cada registro possua todos os dados relativos ao objeto em estudo.

O recurso à transposição acontece sempre que seja necessário relacionar o conteúdo de um atributo numa tabela com registros de outra, ou seja, quando se pretende associar num só registro todas as promoções em que um cliente participou, passa-se de uma situação de várias tabelas para uma só tabela.

D-) Interpretação dos Resultados Alcançados

Após a geração dos modelos, é necessário avaliar o desempenho dos mesmos. Caso esta análise não seja efetuada, existe o risco de serem apresentados resultados errados ou sem interesse. Existem vários métodos de amostragem para estimar a capacidade de generalização de um modelo: *Estatística Simples*, *Validação com Divisão da Amostra*, *Validação Cruzada* e *Bootstrapping*.

O método mais popular para a estimação do erro de generalização é a *Validação com Divisão da Amostra*, que se baseia numa divisão dos dados do problema em casos de treino, para a aprendizagem do modelo e casos de validação para estimar o erro de validação. Como

pontos fortes temos a sua simplicidade e rapidez, embora produza uma redução efetiva dos dados disponíveis para treino.

Validação Cruzada é um melhoramento do método de *Validação com Divisão da Amostra*, que permite a utilização de todos os casos disponíveis. Na validação cruzada k -desdobrável, os dados (P) são divididos em k subconjuntos mutuamente exclusivos (P_1, P_2, \dots, P_k) de comprimentos aproximadamente iguais. Os modelos são treinados e testados k vezes. O erro final de generalização é dado pela média dos erros de validação obtidos durante k vezes. Os valores de k podem variar entre 2 e n embora o valor 10 seja o mais popular. A *Validação Cruzada* é notavelmente superior à validação com divisão da amostra para pequenos conjuntos de exemplos de treino. Todavia, é conseguida à custa de um considerável esforço computacional.

Existem várias técnicas de avaliação, que devem ser selecionadas em função do tipo de problema, sendo as mais utilizadas em problemas de classificação, a *Matriz de Confusão*, as *Curvas ROC* e nos casos de regressão medidas como *Mean Absolute Deviation*, *Sum Squared Error*, *Mean Squared Error* e *Root Mean Squared Error*.

D.1) Avaliação de Modelos de Classificação

Quando se trata de problemas de classificação, uma das técnicas mais usadas é a *Matriz de Confusão* (KOHAVI & PROVOST, 1998). A Matriz de Confusão ou de Erros é usada para avaliar o resultado de uma classificação, mapeando os valores previstos por um modelo com os valores desejados.

A matriz de confusão reflete a quantidade de classificações corretas e erradas sobre um conjunto de exemplos T . O número de acertos, para cada classe, é indicado na diagonal principal da matriz $M(C_i, C_j)$, esses acertos podem ser *verdadeiros positivos* (correspondem ao número de exemplos positivos classificados como tal, também designados por TP – *true positive*) e *verdadeiros negativos* (correspondem ao número de exemplos negativos classificados como tal, também designados por TN – *true negative*). Os restantes elementos $M(C_i, C_j)$, para $i \neq j$, representam erros na classificação que podem ser *falsos positivos* (correspondem ao número de exemplos positivos classificados como negativos, também designados por FP – *false positive*) ou *falsos negativos* (correspondem ao número de exemplos negativos classificados como positivos, também designados por FN – *false negative*). A matriz de confusão ideal possui todos os elementos classificadores a nulo, representando a inexistência de erros, sendo possível derivar as seguintes medidas:

- *Precisão (accuracy)* indica uma determinada percentagem de respostas certas ($TP+TF$) em função de todos os casos equacionados ($TP+TF+FP+FN$);
- *Especificidade (erro tipo I)* calcula a percentagem de acertos negativos corretos (TN) relativamente a todos os casos classificados como negativos ($TN+FP$);
- *Sensibilidade (erro tipo II)* reflete a percentagem de acertos positivos corretos (TP) em face de todos os resultados apresentados como positivos ($TP+FN$).

Essas três medidas de desempenho são independentes do custo e das probabilidades das classes.

Outra das medidas de avaliação de classificadores são as curvas ROC. A análise ROC (EGAN, 1975) surgiu no âmbito da “*Teoria da Detecção de Sinal*”, desenvolvida durante a II Guerra Mundial para a análise de imagens de radar. A tarefa consistia em descobrir se um dado sinal era indicador de um barco inimigo, aliado ou simplesmente ruído. A detecção de sinal mede a capacidade do *receiver operator* do radar em fazer essas importantes distinções. A sua utilidade foi reconhecida mais tarde e a sua utilização estendida para a comparação de exames clínicos para diagnósticos médicos. Esta curva permite a avaliação do desempenho de um classificador, sendo apropriado quando existem dois estados possíveis, isto é, duas classes. A curva ROC estabelece a relação entre a taxa de TP ($TP/(TP+FN)$) e a taxa de FP ($FP/(FP+TN)$), variando num determinado *threshold*.

A curva ROC permite visualizar o compromisso entre a sensibilidade (taxa de TP) e a seletividade ($1 -$ taxa de FP) do algoritmo. Na situação ideal, o algoritmo deveria possuir indicadores máximos de sensibilidade e seletividade, ambos iguais a um. A fim de fornecer uma comparação do desempenho dos diferentes classificadores, independente da distribuição de classes ou de custos de erros, existem duas técnicas frequentemente utilizadas: AUC (METZ, 1978) e ROCCH (PROVOST & FAWCETT, 1997):

- A AUC (Area Under Curve) consiste numa métrica de desempenho do classificador obtida por meio do cálculo da área por baixo da curva ROC do classificador, assumindo valores entre 0 e 1 e pode ser interpretada como a probabilidade de um exemplo positivo, escolhido aleatoriamente, ser classificado como tal;
- A análise ROCCH (ROC Convex Hull) permite declarar um subconjunto de classificadores como potencialmente ótimos. Incluídos todos os pontos que constituem as curvas ROC dos diferentes classificadores, sendo formada a convex hull correspondente, é realizada uma análise dos pontos que estão sobre a linha de convex hull. Se um ponto está sobre a convex hull, existe então uma linha tangente a esse ponto

que tem uma taxa de TP superior, sendo o classificador representado por esse ponto considerado como ótimo sob a distribuição assumida correspondente a essa inclinação.

D.2) Avaliação de Modelos de Regressão

Um modelo de regressão é um estimador, que tenta estimar o valor associado a cada um dos exemplos. O principal objetivo de um método de regressão é conceber o “melhor” modelo segundo uma medida da estimativa do erro. Quando se trata de problemas de regressão, o erro (ou resíduo) e é medido por $e = d - d'$ onde d representa o valor desejado e d' o valor ajustados pelo modelo.

Para o conjunto de dados $x_1 \rightarrow d_1, \dots, x_n \rightarrow d_n$, pode ser calculado um erro global:

$$\text{Desvio médio absoluto (MAD)}: \text{MAD} = \frac{\sum_{i=1}^N |e_i|}{N};$$

$$\text{Soma do erro quadrático (SSE)}: \text{SSE} = \sum_{i=1}^N e_i^2;$$

$$\text{Erro médio quadrático (MSE)}: \text{MSE} = \frac{\text{SSE}}{N};$$

$$\text{Raiz do erro médio quadrático (RMSE)}: \text{RMSE} = \sqrt{\text{MSE}}.$$

As medidas MAD e MSE são duas das medidas de avaliação mais usadas no âmbito da regressão. Ambas têm as suas origens na estatística, tendo como intuito definir quão melhor é um estimador para um conjunto de valores. Se ele for uma boa medida de centralidade representa a distribuição dos valores, melhor que qualquer outro valor. A melhor medida de centralidade é aquela que minimiza as medidas MAD ou MSE.

Segundo (TORGO, 1999), essas medidas quando aplicadas à avaliação de modelos de regressão servem a diferentes propósitos. Se na previsão de valores for aceitável cometer alguns erros extremos, desde que seja aproximada na maior parte das vezes do valor real, a medida MAD é a mais adequada. Quando se está perante situações em que é crucial não cometer erros extremos, mesmo que possam existir erros pequenos e frequentes, então a medida MSE deverá ser utilizada. Isto porque, esta medida utiliza o quadrado das distâncias das previsões aos valores reais. Desta forma, grandes distâncias são amplificadas relativamente às pequenas distâncias.

2.2 Mineração de Dados (Data Mining – DM)

Mineração de Dados (DM) é o processo de exploração e análise de grandes quantidades de dados de modo a descobrir padrões e regras que possam ser importantes para a resolução de determinado problema (LINOFF et al., 2000). O DM é visto por alguns autores como o núcleo do processo de KDD, sendo a fase desse que tem recebido mais atenção por parte dos pesquisadores. A escolha das abordagens a serem aplicadas sobre os dados depende diretamente dos objetivos, definidos inicialmente e deverá começar pela expressão das tarefas de alto nível (SHEPARD, 1998) e (DROZDENKO et al., 2002).

Observa-se, entretanto que todo esse processo sugere uma hierarquia, algo que começa em instâncias elementares (embora volumosas) e termina num ponto relativamente concentrado, mas bastante importante – o conhecimento. Esse é um dos conceitos importantes na KDD: encontrar padrões requer que os dados brutos sejam sistematicamente "simplificados" de modo a desvalorizar aquilo que é específico e privilegiar aquilo que é genérico. Numa perspectiva orientada para os processos, as atividades do DM podem-se agrupar em três classes (RAMACHANDRAN, 2001):

- *Descoberta de padrões* – processo de procura de padrões escondidos na BD sem idéias ou hipóteses pré-concebidas acerca daquilo que os padrões poderão ser, em geral, os algoritmos tem a iniciativa de realizar a pesquisa sem qualquer orientação específica do pesquisador, ou seja, tendências e variações ou associações ou afinidades entre elementos;
- *Modelagem preditiva* – esta classe de atividades aplica na prática o resultado da descoberta de padrões e consiste na sua utilização para previsão do futuro. A modelagem preditiva permite ao pesquisador submeter registros com valores desconhecidos em alguns dos seus atributos ao sistema, e esse se baseando unicamente nos padrões descobertos irá prever os valores então desconhecidos para esses campos, em geral, previsão de resultados;
- *Análise exploratória* – Trata-se do processo de aplicar os padrões extraídos com o objetivo de detectar dados anômalos ou pelo menos pouco habituais. Para a descoberta de elementos pouco habituais, primeiro define-se o padrão normal e então, com uma determinada margem de desvio, detectam-se os elementos que se afastam desse padrão. Por contraste com as atividades de descoberta de padrões que ajudam a determinar

“conhecimento normal”, a análise exploratória procura elementos pouco habituais e casos específicos dentro da BD, ou seja, detecção de desvios.

A investigação desenvolvida no âmbito do DM viabilizou quer o surgimento de metodologias de implementação, quer o surgimento de algoritmos de aprendizagem que se distinguem na forma como traduzem a informação descoberta e no processo como é realizada essa descoberta, havendo alguns mais adequados a determinados tipos de problemas e de dados. Como uma etapa do processo de KDD, o DM é sustentado por três pilares fundamentais, dos quais depende o sucesso do projeto (LINOFF et al., 2000) dados; modelos e técnicas; e modelagem.

2.2.1) Metodologias de Mineração de Dados

Atualmente encontram-se disseminadas e bem definidas duas metodologias para o desenvolvimento de DM: metodologia *CRISP-DM* (**CROSS-Industry Standard Process for Data Mining**) e a metodologia *SEMMA* (**Sample, Explore, Modify, Model, Assessment**). Essas metodologias foram desenvolvidas em ambientes diferentes, sendo a primeira por um consórcio composto por organizações de diferentes setores de atividade (indústria, serviços, fornecedores de tecnologia) e a segunda por uma organização fornecedora de soluções de suporte à decisão e Business Intelligence (BI). Esse trabalho descreve sucintamente mais três metodologias que são também bem conhecidas, quais sejam: Metodologia ADRIAANS, a Metodologia KLEMETTINEN e a Metodologia FELDENS.

Atendendo ao fato de, em termos de processos para desenvolvimento de um projeto de Mineração de Dados, a metodologia *CRISP-DM* ser mais completa que a *SEMMA* casos práticos apresentados nesse trabalho foram desenvolvidos segundo uma metodologia própria, porém orientada na concepção *CRISP-DM*. Isto ocorre devido à incorporação das fases de Estudo do Negócio, Estudo dos Dados e Implementação se encontrarem melhor documentada, focalizando todo o processo no estudo do negócio. Isto é, orienta as suas etapas nos objetivos de negócio especificados traduzindo-se numa forma segura e direta de resolução do problema de Mineração de Dados, ao apresentar uma visão mais ampla.

1) Metodologia CRISP-DM

A metodologia *CRISP-DM* foi concebida no final de 1996 e o seu desenvolvimento foi motivado pelo interesse crescente e generalizado, por um lado pelo mercado de DM, e por

outro, pelo consenso de que a indústria necessitava de um processo padronizado (WIRTH, 2000). Os fundamentos desta metodologia para além dos princípios acadêmicos e teóricos baseiam-se na prática, na experiência daqueles que desenvolvem de fato, projetos de DM. O conhecimento prático foi assim incorporado de forma a dar resposta aos requisitos dos pesquisadores, não se centrado unicamente na tecnologia, mas antes na resolução de problemas do negócio (HAN & KAMBER, 2001). A metodologia CRISP-DM é descrita em termos de um processo hierárquico, com um ciclo de vida que se desenvolve em seis fases:

- Estudo de Negócios
- Estudo dos Dados
- Preparação dos Dados
- Modelagem
- Implementação
- Avaliação

As fases não têm uma seqüência fixa, dependendo do resultado e do desempenho das outras fases ou das tarefas particulares de determinada fase (CHAPMAN et al., 2000). A Figura 2.6 apresenta o ciclo de vida da metodologia CRISP-DM, em que as setas indicam a seqüência, ligações e interligação entre as fases.

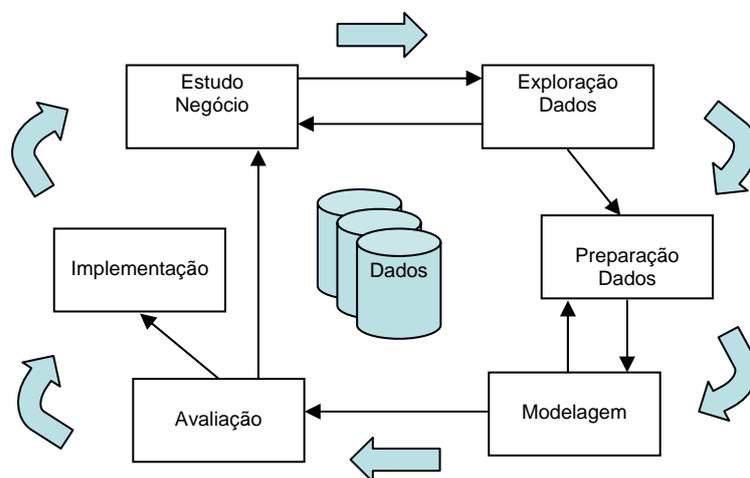


Figura 2.6 – Metodologia CRISP-DM.

Fonte: Adaptada pelo autor a partir de (CHAPMAN et al., 2000).

Na apresentação das fases da metodologia CRISP-DM destacam-se as características mais proeminentes e a documentação produzida. A exposição que se segue tem como base o documento da CRISP-DM (CHAPMAN et al., 2000):

Estudo do Negócio

A abordagem ao *negócio* da organização, centra-se na análise dos objetivos do projeto e nos requisitos (funcionais, técnicos, temporais) segundo a perspectiva organizacional. O conhecimento adquirido nesse estudo é posteriormente utilizado para a definição do problema de DM e na concepção do plano preliminar.

A primeira abordagem do processo de DM consiste em estudar a necessidade da realização do próprio projeto de DM, compreender e enquadrar a perspectiva do problema, os objetivos a atingir e descobrir quais os fatores mais preponderantes que influenciam os resultados, ou seja, prospectar a envolvente do problema a ser resolvido. O estudo do negócio realiza-se pelas seguintes tarefas:

1. *Determinação dos objetivos do negócio* – na fase inicial do projeto é fundamental compreender todos os aspectos que condicionam o negócio como seja, conhecer segundo a perspectiva da organização, os objetivos primários do cliente (fidelização dos clientes atuais prevendo quando esses estão susceptíveis de abandonar);
2. *Avaliação da situação atual* – determinar com exatidão todos os recursos disponíveis para o projeto (recursos humanos, materiais e financeiros). Realizar um levantamento de todos os requisitos, pressupostos e restrições do projeto, o que inclui um programa de realização, compreensibilidade, qualidade dos resultados, segurança, aspectos legais e restrições na disponibilidade dos recursos e tecnológicos. Deverão igualmente ser identificados todos os riscos, ameaças ou eventos que possam comprometer o projeto e respectivos planos de contingência (ações que previnem o risco). Importa referir ainda a importância da elaboração de uma análise de custos e benefícios para o projeto, onde se compare os custos desse com o potencial benefício para o negócio;
3. *Definição dos objetivos de DM* – descrição dos objetivos de DM e os critérios de sucesso do DM (classificação, previsão, segmentação). Como exemplo, tendo por base o histórico das compras efetuadas nos últimos anos, o preço dos produtos e a informação demográfica (idade, rendimentos, cidade, sexo), prever a quantidade que um cliente irá comprar;
4. *Definição do plano para o projeto* – esta tarefa consiste na elaboração de um plano para o projeto que inclua a duração, os recursos, as fases, as subfases, as interações entre os processos, entradas, saídas e dependências. Inclui ainda a elaboração do pressuposto inicial para as ferramentas e técnicas (requisitos ao nível das ferramentas, BD, dos Sistemas Operacionais).

O resultado final do *estudo do negócio* consiste num plano do projeto que inclui a informação acerca do negócio, os seus objetivos e critérios de sucesso, os vários recursos, os requisitos e restrições, os custos e benefícios, os objetivos de DM e os pressupostos das ferramentas e técnicas a utilizar.

Estudo dos Dados

A fase de *estudo dos dados*, inicia-se com obtenção inicial dos dados e prossegue com a sua análise de forma a identificar problemas de qualidade. Para que se possam aplicar as técnicas de DM aos dados, é necessário levar em conta algumas tarefas, quais sejam:

1. *Obtenção inicial dos dados* – consiste na aquisição dos dados e da sua compreensão. Desta tarefa resulta uma lista dos dados adquiridos, a sua localização, os métodos de aquisição, problemas e soluções encontradas.
2. *Descrição dos dados* – uma vez obtidos os dados é necessário descrevê-los, reconhecer o seu formato, o número de registos nas tabelas, identificar os registos e outras características, entretanto descobertas.
3. *Exploração os dados* – o resultado desta tarefa consiste numa listagem inicial de hipóteses e o seu impacto no restante projeto. Para uma melhor exploração utilizam-se, por exemplo, gráficos e histogramas, que indicam características dos dados.
4. *Verificação da qualidade dos dados* – fazer um relatório que inclui problemas de qualidade nos dados e possíveis soluções (normalmente dependem diretamente dos dados e do conhecimento do negócio).

Preparação dos Dados

A fase de *preparação dos dados* envolve todas as atividades associadas à construção do conjunto final de dados, aquele que é usado na ferramenta de modelagem, sofrendo inevitavelmente várias otimizações. Esta fase inclui a seleção de tabelas, registos e atributos, bem como a transformação e limpeza dos dados a usar na ferramenta de modelagem, as subfases são as seguintes:

1. *Seleção de dados* – consiste na escolha dos dados a utilizar na análise. Os critérios para a seleção incluem a relevância dos objetivos de DM e restrições técnicas e de qualidade, como os limites no volume de dados e tipo de dados. No final desta tarefa é ainda realizada uma listagem dos dados incluídos e excluídos e as razões da decisão;

2. *Limpeza de dados* – Complementa a tarefa anterior, existindo várias técnicas que podem ser aplicadas de forma a otimizar a qualidade dos dados, em geral, a normalização dos dados e tratamento dos dados omissos;
3. *Derivação de dados* – realiza-se pela derivação de novos atributos (em geral, determinar o novo atributo *idade*, a partir da *data de nascimento*), criação de novos registros e transformação dos dados (normalização);
4. *Integração de dados* – obtém-se recorrendo a métodos para a criação de novos registros ou valores, cuja informação é uma combinação de múltiplas tabelas ou registros (em geral, junção e agregação de tabelas ou registros);
5. *Formatação de dados* – a última tarefa da preparação dos dados consiste em modificações sintáticas nos dados de modo a que não alterem o seu significado, mas que os tornem utilizáveis pela ferramenta de modelagem.

Modelagem

Esta fase consiste na seleção de várias técnicas de modelagem (árvores de decisão ou regressão linear múltipla) e os seus parâmetros são ajustados de forma a otimizar os resultados. Normalmente, para o mesmo problema de DM existem várias técnicas disponíveis, sendo que algumas têm requisitos específicos para a forma como os dados são apresentados, pelo que pode ser necessário voltar à fase anterior.

Como referido anteriormente, no início do processo, são especificados os problemas e os objetivos do DM, no entanto, apenas nesta fase é que os dados, previamente preparados para a modelagem, são utilizados. A escolha das técnicas deve ser cuidadosa de modo a que satisfazer os objetivos de DM.

Nesta fase são contempladas as seguintes tarefas:

1. *Seleção de técnicas de modelagem* – a seleção da técnica mais apropriada deve ser realizada tendo atenção ao tipo de problema, as ferramentas e os objetivos do DM;
2. *Definição de uma concepção de teste* – importa antes de construir o modelo, definir um procedimento ou um mecanismo para testar o desempenho do próprio modelo;
3. *Construção do modelo* – uma vez selecionada a ferramenta de modelagem, esta é aplicada ao conjunto de dados preparados anteriormente, permitindo a criação de um ou mais modelos. Os vários parâmetros das ferramentas de modelagem devem ser ajustados e os modelos resultantes devem ser convenientemente interpretados e o seu desempenho explicado;

4. *Revisão do modelo* – a interpretação dos modelos deve ser realizada de acordo com o domínio do conhecimento, critérios de sucesso do projeto de DM e com o mecanismo de teste definido. Na avaliação do sucesso de aplicação do modelo deve ser levado em consideração o impacto dos resultados desse no contexto do negócio.

Avaliação

A fase de *avaliação* consiste na validação da utilidade do modelo (ou modelos), na revisão dos passos executados na sua construção e verificação se forem atingidos os objetivos do negócio. Esta fase compreende as seguintes tarefas:

1. *Avaliação dos resultados* – determinação se o modelo atingiu os objetivos do negócio (e de DM) e avaliação do modelo quanto a possíveis lacunas;
2. *Revisão do processo* – análise de todas as fases do processo de modo a realçar eventuais atividades esquecidas e/ou que necessitem de ser repetidas;
3. *Determinação dos próximos passos* – o projeto apenas se considera concluído se todos se todos os passos anteriores foram satisfatórios e os resultados cumpriram os objetivos, devendo então passar para a sua fase de implementação. Caso ocorra o inverso, é necessário então proceder a uma nova iteração das fases anteriores, utilizando novos parâmetros.

Implementação

Uma vez criado o(s) modelo(s) tal não representa o final do projeto. O conhecimento extraído deve ser organizado e apresentado de modo a que o usuário o possa usar. A fase de implementação pode ser tão simples como, por exemplo, gerar um relatório ou pode ser tão complexa como integrar os resultados nos sistemas da organização, dependendo dos requisitos. Em muitos casos é o usuário e não o pesquisador, que executa os passos de implementação sendo, no entanto importante que esse entenda as ações que precisa ser executada de modo a fazer uso dos modelos criados. As tarefas envolvidas nesta fase são:

1. *Planejamento da avaliação dos resultados* – define a estratégia para a implementação dos resultados de DM, incluindo os passos e a forma como executar.
2. *Planejamento da monitorização e manutenção* – consiste na definição de estratégia de monitorização e manutenção e é aconselhável sempre que os resultados do DM (modelos), sejam implementados no domínio do problema como parte da rotina do

quotidiano. Como retorno da monitorização e manutenção é possível verificar se os modelos são usados corretamente.

3. *Produção um relatório final* – é a fase de conclusão do projeto de DM. Elabora-se um relatório final resumindo os pontos mais importantes no projeto, experiência adquirida, explicação dos resultados produzidos e mais importantes.
4. *Revisão do Projeto* – avaliação dos pontos corretos e errados, do que correu bem ou que necessita de ser melhorado. Resumo das experiências mais importantes do projeto, torna-se benéfico para projetos futuros e em situações similares referir as armadilhas, aproximações erradas ou como foram selecionadas as técnicas de DM.

A Aplicação desta metodologia em projetos de DM permite garantir uma maior confiabilidade, menores custos de execução, maior segurança, assim como a sua maior exequibilidade e viabilidade. A metodologia CRISP-DM é extremamente completa e documentada, uma vez que as suas fases estão devidamente organizadas, estruturadas e definidas, permitindo que o projeto possa ser facilmente compreendido ou revisto.

Como resultado da sua aplicação obtém-se um conjunto de documentação sobre todo o processo, em vários relatórios:

- *Estudo do negócio;*
- *Relatório do conjunto inicial de dados;*
- *Relatório da descrição de dados;*
- *Relatório da qualidade dos dados;*
- *Relatório da descrição da amostra de dados;*
- *Relatório da modelagem;*
- *Relatório da avaliação;*
- *Plano de implementação;*
- *Manutenção e relatório final*

2) Metodologia SEMMA

A metodologia SEMMA foi proposta pelo Instituto SAS (SAS Institute Inc. <http://www.sas.com>), que se dedica ao desenvolvimento de soluções para estatística, análise de dados, BI, DM e SAD (SAS, 2007). Esta metodologia surge como resposta à necessidade de definição, padronização e de integração dos processos de DM nos ciclos de produção, para que a solução seja aceite mais facilmente no ambiente do negócio (GROTH, 2000). Mais do que

uma metodologia de DM, é considerada como um auxílio para conduzir um projeto em todas as suas etapas, desde a especificação do problema do negócio até à sua implementação.

O DM é definido pelo Instituto SAS como o “processo de extrair informação valiosa e relações complexas de um grande volume de dados” e foi nesse sentido, que dividiram o processo de DM em 5 etapas (Figura 2.7) – dando origem ao acrônimo SEMMA (SAS, 2007) e (GROTH, 2000).

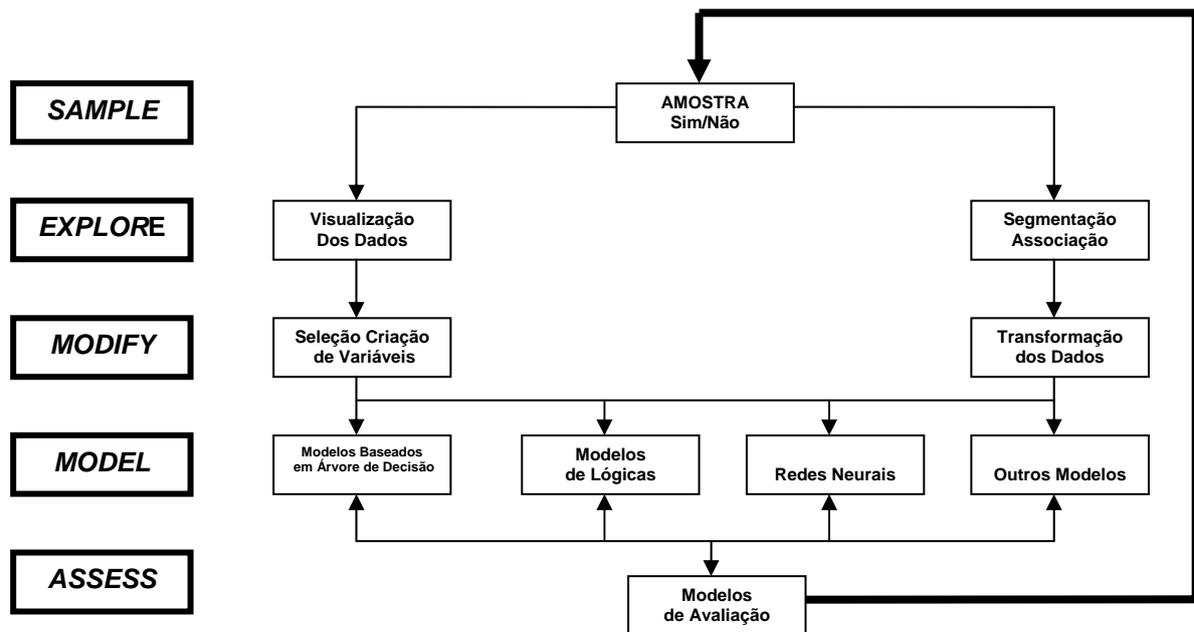


Figura 2.7 – Metodologia SEMMA

Fonte: Adaptada pelo autor a partir de (SAS, 2007)

Numa forma resumida é possível apresentar esta metodologia como um processo com cinco fases, que se inicia com uma amostra (*Sample*) representativa dos dados à qual se aplicam técnicas estatísticas de exploração e de visualização dos dados (*Explore*). Posteriormente são selecionadas e transformadas as variáveis (*Modify*) consideradas mais significativas (as variáveis que sobressaíram na fase anterior), as que são mais relevantes em termos de projeto, e sobre as quais se constroem o modelo (*Model*) (aplicam-se algoritmos no sentido de alcançar os objetivos) e por fim se avalia o modelo (*Assess*). Cada uma das etapas é distinta e corresponde a um ciclo, e as suas tarefas internas podem ser executadas repetidamente sempre que necessário, isto é, pode-se atualizar e ajustar quando surgir nova informação.

A metodologia SEMMA disponibiliza um método de fácil compreensão, possibilitando um desenvolvimento organizado, adequado e conseqüente manutenção dos projetos de DM, isto é, confere uma estrutura para a sua concepção, criação e evolução, de modo a apresentar soluções para os problemas e descobrir os objetivos de DM para o negócio. Esta metodologia

consiste em cinco fases as quais são seguidamente caracterizadas de acordo com a documentação oficial fornecida pelo instituto SAS, disponível no seu endereço oficial na internet.

Sample – Amostragem

A primeira fase da metodologia SEMMA consiste na realização de uma amostragem significativa com a extração de uma quantidade de dados do universo existente. A amostra deve corresponder a um subconjunto de dados que pertencem ao universo onde cada elemento tem as mesmas hipóteses de ser incluído, mas também deve ser pequena de modo a tornar-se rápida e de fácil manipulação.

A realização do processo de amostragem traduz-se numa otimização dos custos, da rentabilidade e do desempenho das etapas seguintes, dado o fato de a manipulação de uma amostra ser mais rápida e fácil do que manipular todo o universo de dados disponíveis.

O desenvolvimento de todo o processo de DM a partir de uma amostra representativa reduz drasticamente o volume e o tempo de processamento necessário para tirar informação crucial para o negócio. Nesse contexto, se o universo de dados tiver um determinado padrão ou tendência bastante determinado, esses estão patentes na amostra, caso contrário, o padrão ou tendência for irrelevante, ao ponto de não ser detectado na amostra, também não é importante para o universo de dados (SAS, 2007).

Explore – Exploração

Uma vez realizado o processo de amostragem, a primeira abordagem realizada sobre os dados consiste em explorá-los visualmente ou numericamente (em geral, gráficos de distribuição e dispersão, histogramas, tabelas de frequência, mapas de associações e segmentação) permitindo em alguns caso detectar as tendências ou agrupamentos inerentes nos dados. A *exploração* ajuda a refinar o processo de descoberta. Se a visualização não revelar claramente as tendências, é possível recorrer ainda ao uso de técnicas mais avançadas de estatísticas, como a distribuição de *Poisson*, *Mínimos Quadrados*, *Qui-Quadrado* e *Regressão Linear*. A etapa de exploração é marcada pela procura de tendências imprevistas e por anomalias de forma a conhecer os dados de uma forma aprofundada e as suas relações.

***Modify* – Modificação**

A fase da *modificação* concentra todas as transformações necessárias com base nos resultados da etapa de exploração. As transformações realizadas podem ser de inclusão de informação (agrupamento de subgrupos significativos de dados), seleção ou introdução de novas variáveis, de forma a obter-se as variáveis mais significativas. O objetivo desta fase consiste em criar, selecionar e transformar as variáveis para o processo de construção do modelo, preparando os dados para a etapa seguinte – a Modelagem.

***Model* – Modelagem**

Uma vez preparados os dados, é possível então prosseguir para a fase de aplicação de algoritmos - *modelagem*. É nesta fase que se definem as técnicas de construção de modelos de DM, onde se incluem as técnicas estatísticas (Regressão Linear, Indução de Probabilidades).

Para a seleção da técnica é necessário levar em consideração que cada modelo tem propriedades e características singulares dependentes dos dados e adequados a situações específicas de DM (as regressão linear múltipla alcançam melhores resultados com dados lineares). A etapa de modelagem tem como objetivo selecionar as técnicas de construção de modelos de forma a prever com confiança os resultados desejados (SAS, 2007).

***Assessment* – Avaliação**

A etapa final da metodologia SEMMA consiste na *avaliação* do modelo de forma a aferir o seu desempenho. Geralmente, a fase de avaliação de um modelo corresponde à aplicação desse a uma amostra de dados selecionada para esse fim (conjunto de teste).

Se o modelo for válido esse deve funcionar tão bem como na amostra que serviu de base à sua construção. A etapa de avaliação tem como objetivo aplicar o modelo à amostra de dados e verificar a seu desempenho, de forma a proceder a ajustes se necessário (SAS, 2007).

3) Metodologia ADRIAANS

Esta abordagem do processo de KDD baseia-se na necessidade das organizações em obterem continuamente novas informações sobre seus dados, por isso não deve ser executado apenas uma vez, mas repetido sempre que novas necessidades de informações aparecerem.

Portanto, nesta abordagem não existe uma etapa específica para entendimento dos dados. É pressuposto que já exista um conhecimento prévio do domínio da base de dados e, conseqüentemente, do objetivo do processo. O processo é composto por seis etapas conforme Figura 2.8 (ADRIAANS et al, 1996).

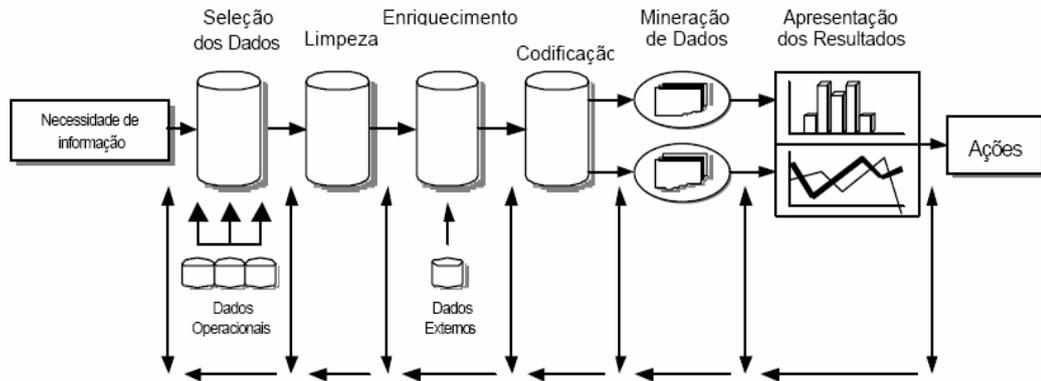


Figura 2.8 - Metodologia ADRIAANS.

Fonte: Fonte: Adaptada pelo autor a partir de (ADRIAANS et al, 1996).

1. *Seleção dos dados:* Nesta etapa ocorre uma análise de todos os dados operacionais da base de dados e são selecionados apenas aqueles que são necessários para alcançar os objetivos do processo. Podem ser feitas novas seleções quando houver outra iteração, ou seja, podem-se incluir dados anteriormente descartados, pois o processo é iterativo, permitindo a retomada de qualquer etapa, independente da etapa em que se encontra;
2. *Limpeza:* Nesta etapa são utilizadas diversas operações de limpeza nos dados, como por exemplo: a remoção de dados duplicados e a correção dos dados. A etapa de limpeza pode ser executada inúmeras vezes, já que é impraticável prever com antecedência todos os problemas de qualidade existente na base;
3. *Enriquecimento:* Algumas informações podem ser incluídas na base de dados para que seja possível atingir os objetivos do processo. Esses dados podem estar disponíveis em outros locais, ou até mesmo podem ser gerados a partir de dados existentes na base de dados e transformados para que seja obtida a informação;
4. *Codificação:* A forma que os dados estão armazenados nas bases de dados pode não ser a representação mais apropriada para a utilização no processo de KDD. Geralmente, os dados têm sua representação apropriada ao contexto da aplicação. Por exemplo, um atributo com valores literais pode não ser adequado a determinados algoritmos mineradores utilizados na etapa de mineração de dados. Para adequá-lo pode ser necessário normalizar esses valores dentro de um determinado intervalo. A codificação é um procedimento criativo, existem diversas maneiras de codificação, assim é difícil

descrevê-las, pois cada caso deve ser analisado individualmente e sua codificação pode variar de acordo com a escolha do algoritmo minerador da próxima etapa;

5. *Mineração de dados*: Esta é a etapa onde os dados são manipulados para que seja extraído o conhecimento. É a etapa que mais exige dos recursos computacionais. O autor afirma que, utilizando inicialmente uma ferramenta de consulta SQL, pode ser possível ter uma visão geral dos dados para então partir para uma análise menos trivial. Nesta primeira tarefa, 80% do conhecimento são extraídos e já podem revelar alguma informação interessante. Entretanto, as informações extraídas por essas consultas podem não ser suficientes, surgindo à necessidade de ser utilizadas técnicas avançadas;
6. *Apresentação dos resultados*: Finalizada a etapa de mineração de dados, resultam informações num formato específico de acordo com a técnica utilizada. Deve-se levar em conta que os dados podem estar codificados ou mesmo que o método utilizado na etapa de mineração gere, como saída, informações em algum formalismo ou representação muito específicos.

Esses resultados devem ser exibidos de forma clara para que sejam de fácil entendimento para quem irá utilizá-los, geralmente pessoas que necessariamente não interpretarão os resultados tão facilmente quanto aquela que conduziu o processo de KDD.

4) Metodologia KLEMETTINEN

(KLEMETTINEN et al, 1997) apresentam uma metodologia que pode ser usada para automatizar aquisição de conhecimento. As fases dessa metodologia são aquelas já definidas pelo autor (FAYYAD et al., 1996), quais sejam: pré-processamento, transformação, descoberta, apresentação e utilização, ver Figura 2.9.

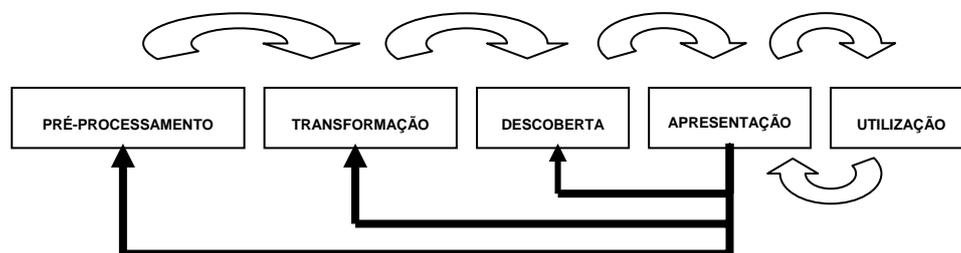


Figura 2.9 - Metodologia KLEMETTINEN.

Fonte: Adaptada pelo autor a partir de (KLEMETTINEN et al., 1997).

No entanto, a maior ênfase é dada nas duas fases centrais dessa metodologia:

- Fase de descoberta de padrões: onde são encontrados todos os padrões potencialmente relevantes para algum critério bastante livre;

- Fase de apresentação: onde são fornecidos métodos flexíveis para iterativa e interativamente criar diferentes visões para os padrões descobertos.

Nas duas primeiras fases do processo, os dados são coletados e preparados de forma adequada para descoberta de padrões. Uma visão geral sobre os dados pode ser produzida nesta fase. Os atributos identificados como irrelevantes são removidos e novos atributos podem ser derivados. Na fase de descoberta de padrões, todos os padrões potencialmente interessantes são gerados do conjunto do *data set*. A apresentação do conhecimento descoberto é uma parte principal dessa metodologia. Nesta fase, os padrões relevantes podem ser localizados de grandes coleções de padrões potencialmente relevantes.

5) Metodologia FELDENS

(FELDENS et al., 1998) propõem uma metodologia integrada, na qual as tecnologias de mineração de dados e *data warehouse*, bem como questões de visualização têm papéis muito importantes no processo. Também supõe uma forte interação entre mineradores de dados e pessoas da organização para questões de modelagem e preparação de dados. As fases definidas para esta metodologia são: pré-processamento, mineração de dados e pós-processamento, conforme Figura 2.10.

A fase de pré-processamento inclui tudo o que é feito antes da mineração de dados, o que significa a análise que é feita na organização a fim de focar o projeto de mineração de dados, a análise dos dados existentes, integração de fontes de dados, transformações de dados, etc.

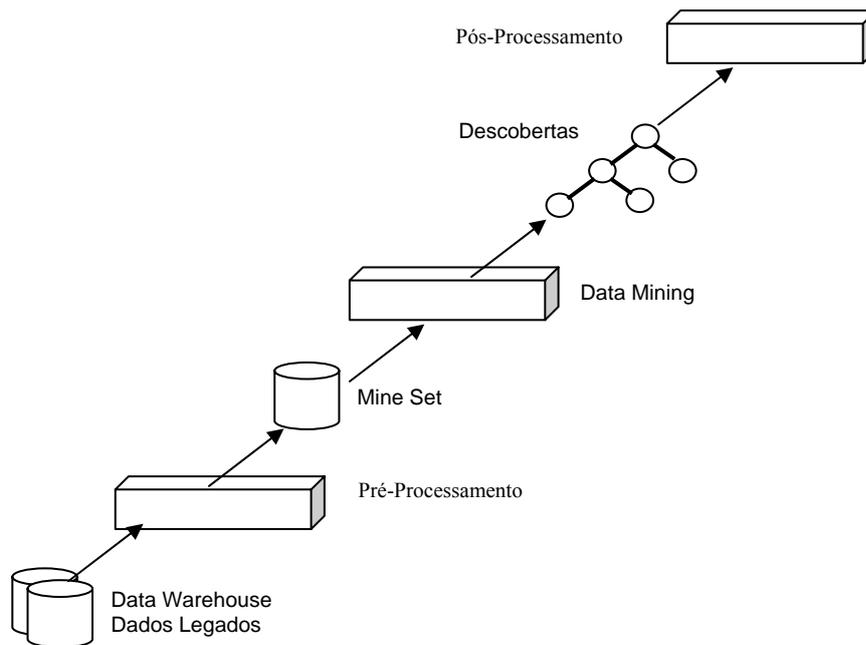


Figura 2.10 - Metodologia FELDENS.

Fonte: Adaptada pelo autor a partir de (FELDENS et al., 1998).

A fase de mineração de dados inclui a aplicação de algoritmos, possivelmente a aplicação repetida. A escolha dos algoritmos pode ser realizada baseando-se na análise que é feita na fase de pré-processamento. A fase de pós-processamento pode ser definida por operações de filtragem, estruturação e classificação. Somente após esta fase, o conhecimento descoberto é apresentado ao usuário. O conhecimento descoberto pode ser filtrado por alguma medida estatística, por exemplo, suporte, confiança ou outro critério definido pelo usuário. Estruturação significa que o conhecimento pode ser organizado de forma hierárquica.

6) Comparação entre as Diferentes Abordagens do Processo de KDD

Em virtude das diversas abordagens presentes na bibliografia, optou-se por realizar um estudo comparativo entre as propostas estudadas. Por meio da análise comparativa, mostrada no Quadro 2.1, foram detectadas algumas similaridades. Porém nenhuma das abordagens adequou-se ao estudo de caso proposto. Conseqüentemente, propõe-se um modelo do processo integrando essas diferentes abordagens. No Capítulo 3, são descritas as etapas do modelo proposto para esse estudo de caso.

As diferentes visões do processo de KDD em etapas, sugerem existir uma trajetória linear das etapas. No entanto, isso não se verifica, uma vez que em cada etapa pode ser identificada a necessidade de retorno para uma das etapas anteriores. Por exemplo, se já numa etapa avançada, onde técnicas já estão sendo aplicadas para descoberta de conhecimento, for

verificada a necessidade de um dado que não havia sido previsto anteriormente, isso pode levar à retomada à fase de seleção.

Assim, o processo ocorre de forma iterativa, pois cada etapa pode possuir interseções com as demais. Da mesma forma, esse processo é interativo, onde a cada etapa as informações geradas são analisadas e enriquecidas com outras informações para obter-se resultados melhores.

Quadro 2.1 - Comparação entre as metodologias FAYYAD, ADRIAANS e CRISP-DM.

Etapas	Metodologia FAYYAD	Metodologia ADRIAANS	Metodologia CRISP-DM
Definição do Objetivo	O objetivo é definido para solucionar o problema que gerou a necessidade de implementação de um processo de descoberta de conhecimento.	O objetivo surge da constante necessidade das organizações em obterem novas informações sobre seus dados	É definido na primeira fase, onde o responsável pelo processo realiza o entendimento do domínio do negócio para então definir o objetivo do processo e critérios de sucesso.
Entendimento dos Dados	Ainda durante a fase de definição do objetivo é realizado o entendimento do domínio e avaliação das possibilidades de sucesso de acordo com os dados disponíveis.	Não existe a etapa de compreensão dos dados. É pressuposto que o objetivo proposto pelo usuário já está vinculado aos dados a serem selecionados.	Descreve os dados disponíveis, avaliando quantidade e problemas de qualidade dos dados. Detecta dados relevantes para formular hipóteses.
Seleção dos Dados	É feito um subconjunto de dados selecionados a partir da(s) base(s) de dados disponíveis. Este subconjunto conterá apenas aqueles dados relevantes para a solução do problema.	Ocorre uma análise dos dados e são selecionados apenas os dados relevantes com a finalidade de reduzir o volume do conjunto de dados.	Tarefa realizada na etapa de preparação dos dados, onde é selecionado um conjunto que pode ser modificado a cada nova iteração.
Limpeza	Realizada na etapa de pré-processamento, porém o autor não define como deve ser conduzida a tarefa de limpeza.	Utiliza algumas operações de limpeza nos dados: de duplicação de dados, correção, etc. Executada inúmeras vezes, já que é impossível prever todos os problemas de qualidade.	Também realizada na etapa de preparação dos dados, todas as tarefas de limpeza realizadas são documentadas.
Transformação	Utiliza algumas técnicas para formatação dos dados, para adequá-los ao algoritmo de mineração. O autor indica que seja primeiramente definida a técnica de mineração e escolhido o algoritmo que é utilizado.	Trata de como incorporar dados externos à base (enriquecimento) e como transformar os dados (codificação) para que possam ser utilizados na etapa de mineração. O autor apresenta métodos de codificação dos dados.	Outra tarefa realizada durante a etapa de preparação dos dados. Onde os dados são formatados para entrada no modelo gerado pelo algoritmo minerador, porém sem alteração do seu significado.
Mineração dos Dados	Considerada a fase mais importante do processo, onde é escolhido o método, definido(s) o(s) algoritmo(s) e realizada a busca pelo conhecimento no banco de dados.	Definição da técnica e dos algoritmos, porém, é garantido que 80% do conhecimento podem ser extraído por consultas SQL e complementado com a utilização de técnicas e algoritmos avançados.	Tarefa de modelagem, onde é definida a técnica, construído o modelo, testado e aplicado no banco de dados.
Resultados	Ocorre a interpretação dos resultados obtidos com o objetivo de melhorar a compreensão do conhecimento. Caso não tenha alcançado os objetivos o processo é retomado.	Os resultados são interpretados com o objetivo de melhorar a compreensão, pois os dados podem estar codificados dependendo do método utilizado na etapa de mineração.	Na fase de avaliação é feita uma validação dos resultados de acordo com as regras reais do domínio da aplicação.

Fonte: Elaborado pelo autor.

2.2.2 Técnicas de Mineração de Dados

Existem cinco técnicas gerais de mineração de dados que englobam todas as outras formas de apresentação e permitem uma visão mais global e apropriada ao assunto. São elas a

classificação, a estimativa, a previsão, a análise de afinidades e a análise de agrupamentos (CARVALHO, 2005).

A-) Classificação

A classificação é uma das mais utilizadas técnicas de mineração de dados, simplesmente porque é uma das mais realizadas tarefas humana no auxílio à compreensão do ambiente em que se vive. O ser humano está sempre classificando o que percebe a sua volta, criando classes de relações humanas diferentes (colegas de trabalho, amigos, familiares, por exemplo...) e dando a cada classe uma forma diferente de tratamento.

A classificação pode ser sintetizada por um processo de discriminação de unidades em classes ou categorias. Assim, classificam-se sabores, amigos, clientes, eventos, entre outros, em categorias, tais como doce / salgado / neutro, bom / mau e legal /ilegal.

Em um processo de mineração de dados, a classificação está especificamente voltada à atribuição de uma das classes pré-definidas pelo analista a novos fatos ou objetos submetidos à classificação. Essa técnica pode ser utilizada tanto para entender dados existentes quanto para prever como novos dados irão se comportar (EURIDITIONHOME, 2004).

Como no mundo físico nada é exatamente igual, por mais semelhante que pareça, para se criar classes é preciso permitir que detalhes sejam desprezados e somente as características principais sejam observadas. A tarefa de classificar geralmente exige a comparação de um objeto ou dado com outros dados ou objetos que supostamente pertençam a classes anteriormente definidas. Para comparar os dados ou objetos utiliza-se uma métrica ou forma de medida de diferenças entre eles, sendo comum em mineração de dados às seguintes tarefas de classificação:

- De clientes em baixo, médio ou alto risco de empréstimo bancário;
- De clientes potencialmente consumidores de um determinado produto a julgar pelo seu perfil;
- De transações financeiras como legais, ilegais ou suspeitas em sistemas de fiscalização do mercado financeiro;
- De ações da bolsa de valores com lucros potenciais baixos, médios e altos, entre outras.

Os algoritmos mais utilizados para esse fim são os de árvores de decisão (PELEGRIN et al., 2005), regressão (HAN & KAMBER, 2001) e redes neurais (HAGAN et al., 1996). São exemplos de tarefas de classificação (GOEBEL & GRUENWALD, 1999): classificar pedidos

de créditos como de baixo, médio e alto risco; esclarecer pedidos de seguros fraudulentos; identificar a forma de tratamento na qual um paciente está mais propício a responder, baseando-se em classes de pacientes que respondem bem a determinado tipo de tratamento médico.

B-) Estimativa (ou Regressão)

A estimativa, ao contrário da classificação, está associada a respostas contínuas. Estimar algum índice é determinar seu valor mais provável diante de dados do passado ou de dados de outros índices semelhantes sobre os quais se tem conhecimento.

Suponha que se deseja determinar o gasto de famílias cariocas com lazer e que para isto se possua índices de gastos de famílias paulistanas com lazer, em função da faixa etária e padrão sócio-cultural. Não se sabe exatamente quanto às famílias potiguares gastam com lazer, mas se pode estimar baseando-se nos dados das famílias paulistanas. Certamente que esta estimativa pode levar a grandes erros, uma vez que Rio Grande do Norte e São Paulo são cidades com geografias diferentes e que oferecem diferentes opções de lazer a seus habitantes.

A estimativa é usada para definir um valor para alguma variável contínua desconhecida como, por exemplo, receita, altura ou saldo de cartão de crédito (HARRISON, 1998). Ela lida com resultados contínuos, enquanto que a classificação lida com resultados discretos. Ela pode ser usada para executar uma tarefa de classificação, convencendo-se que diferentes faixas (intervalos) de valores contínuos correspondem a diferentes classes. “Regressão é aprender uma função que mapeia um item de dado para uma variável de predição real estimada” (FAYYAD et al., 1996)

A arte de estimar é exatamente esta: determinar da melhor forma possível um valor, baseando-se em outros valores de situações semelhantes. Os algoritmos de regressão e as redes neurais são bastante utilizados nesses casos.

C-) Previsão

A previsão, como tarefa típica de DM, está associada à avaliação de um valor futuro de uma variável a partir dos dados históricos do seu comportamento passado. Assim, pode-se prever, por exemplo, se o índice BOVESPA subirá ou descerá no dia seguinte; qual é o valor de determinada ação daqui a um determinado período de tempo; o número de clientes que são

perdidos por uma empresa, em um dado horizonte futuro de tempo; qual é a população de certa cidade daqui a dez anos; entre outras coisas.

A única maneira de avaliar se a previsão foi bem feita é aguardar o acontecimento e verificar o quanto foi acertada ou não a previsão realizada. Sem dúvida, a previsão é uma das tarefas mais difíceis não somente na mineração de dados, mas também no cotidiano das pessoas. Os algoritmos que podem ser utilizados aqui são, dentre outros, as redes neurais, a regressão, e as árvores de decisão.

D-) Análise de Afinidades

A análise de afinidades preocupa-se em reconhecer padrões de ocorrência simultânea de determinados eventos nos dados em análise. Determinar que fatos ocorrem simultaneamente com probabilidade razoável (co-ocorrência) ou que itens de uma massa de dados estão presentes juntos com certa chance (correlação).

O exemplo mais clássico de análise de afinidades é o do carrinho de supermercado, do qual se deseja conhecer quais os produtos que são comumente comprados em conjunto pelos consumidores. Isto possibilita a otimização do layout interno dos supermercados e a realização de vendas dirigidas nas quais os itens são oferecidos já em conjuntos com preços menores. Em termos de algoritmos, a utilização das regras de associação constitui-se no procedimento mais utilizado nesses casos (PELEGRIN et al., 2005).

E-) Análise de Agrupamentos

A análise de agrupamentos visa formar grupos de objetos ou elementos mais homogêneos entre si. Pode ser estabelecido previamente um número de grupos a ser formado, ou então se pode admitir ao algoritmo de agrupamento uma livre associação de unidades, de forma que a quantidade de grupos resultante seja conhecida somente ao final do processo.

Uma clara diferença entre agrupamento e classificação é que na classificação as classes são pré-definidas pelo pesquisador, enquanto que aqui não existe tal requisito. Isto torna esta técnica muito mais complexa do que a classificação. Por exemplo, dadas às classes animal, vegetal e mineral, é relativamente simples classificar a qual dessas classes certo objeto pertence, porém de posse de uma massa de dados sobre o consumo no Brasil, determinar quantas classes ou padrões de comportamento consumista existem é algo bem diferente. A

dificuldade reside no fato de que podem não haver tais classes, ou seja, os dados se distribuem igualmente por todo o espaço possível não determinando nenhuma categoria.

Na análise de agrupamentos, os grupos ou classes são construídos com base na semelhança entre os elementos, cabendo ao analisador das classes resultantes avaliar se essas significam algo útil. Por exemplo, agrupar sintomas pode gerar classes que não representem nenhuma doença explicitamente, uma vez que doenças diferentes podem possuir os mesmos sintomas.

A análise de agrupamentos é normalmente uma técnica preliminar, utilizada quando nada ou pouco se sabe sobre os dados. Segmentar um mercado é uma típica análise de agrupamentos onde consumidores são reunidos em classes representantes dos segmentos desse mercado. Em geral, a técnica de agrupamento é executada por algoritmos estatísticos específicos para esse fim, porém as redes neurais e os algoritmos genéticos (HAN & KAMBER, 2001) são também utilizados nesse sentido.

2.2.3) Escolha da Técnica de Mineração de Dados mais Adequada

A escolha de uma técnica de mineração de dados a ser aplicada não é uma tarefa fácil. Segundo HARRISON (1998), a escolha das técnicas de mineração de dados dependerá da tarefa específica a ser executada e dos dados disponíveis para análise. HARRISON (1998) sugere que a seleção das técnicas de mineração de dados deve ser dividida em dois passos: 1) traduzir o problema de negócio a ser resolvido em séries de tarefas de mineração de dados; 2) compreender a natureza dos dados disponíveis em termos de conteúdo e tipos de campos de dados e estrutura das relações entre os registros. Essa escolha pode ser baseada, também, em critérios para classificação das técnicas. Uma relação desses tipos de critérios é dada por HARRISON (1998).

“Diferentes esquemas de classificação podem ser usados para categorizar métodos de mineração de dados sobre os tipos de bases de dados a serem estudados, os tipos de conhecimento a serem descobertos e os tipos de técnicas a serem utilizadas” (CHEN et al,1996), como pode ser visto a seguir:

- *Com que tipos de bases de dados se deve trabalhar:*

Um sistema de descoberta de conhecimento pode ser classificado de acordo com os tipos de base de dados sobre os quais técnicas de mineração de dados são aplicadas, tais como: base de dados relacionais, bases de dados de transação, orientados a objetos,

dedutivos, espaciais, temporais, de multimídia, heterogêneos, ativos, de herança, base de informação de Internet e bases textuais;

- *Qual o tipo de conhecimento deve ser explorado:*

Vários tipos de conhecimento podem ser descobertos por extração de dados, incluindo regras de associação, regras características, regras de classificação, regras discriminantes, agrupamento, evolução e análise de desvio;

- *Qual tipo de técnica deve ser utilizada:*

A extração de dados pode ser categorizada de acordo com as técnicas de mineração de dados subordinadas. Por exemplo, extração dirigida a dados, extração dirigida a questionamento e extração de dados interativa. Pode ser categorizada, também, de acordo com a abordagem de mineração de dados subordinada, tal como: extração de dados baseada em generalização, baseada em padrões, baseada em teorias estatísticas ou matemáticas, abordagens integradas, etc.

Atualmente, a descoberta de regras de associação parece ser uma das técnicas de mineração de dados mais utilizada, sendo encontrada em diversas pesquisas (CHEN et al, 1996), (HARRISON, 1998) e (MANNILA, 1997).

Segundo (HAIR, J. F. et al., 2005), a técnica de *Regressão Linear Múltipla* se enquadra como o método de análise mais apropriado, pois o trabalho de pesquisa envolve uma única variável dependente que é a produtividade de vendas e um conjunto de variáveis independentes, quais sejam: os indicadores de desempenho de vendas. Esta técnica também fornece um meio de avaliar objetivamente o grau e o caráter da relação entre as variáveis dependente e independentes, pela formação da variável estatística por intermédio das variáveis independentes. Essas, além de sua previsão coletiva da variável dependente, também podem ser consideradas por sua contribuição individual à variável estatística e suas previsões.

A-)Técnica de Regressão Linear Múltipla

A origem do termo “Regressão” deu-se por Francis Galton, quando em um ensaio com pais e filhos ele estudou o relacionamento das alturas dos mesmos. A lei de regressão universal de Galton foi confirmada mais tarde por Karl Pearson, que por meio de um grande ensaio constatou que a altura média dos filhos de pais altos era inferior à altura de seus pais e que a altura média de filhos de pais baixos era superior à altura de seus pais, ou seja, ele concluiu que a altura tanto dos filhos altos como baixos tendem para a média de todos os homens. Na atualidade, a interpretação da regressão é bem diferente. De modo geral pode-se dizer que a

análise de regressão é o estudo de uma variável (a variável dependente) em função de uma ou mais variáveis (as variáveis independentes), com o objetivo de estimar e/ou prever a média populacional ou valor médio da variável dependente, utilizando valores observados por amostragem das variáveis independentes.

Regressão é o termo utilizado para designar uma equação matemática que descreva as relações entre duas ou mais variáveis. Regressão linear é um método para se estimar o valor esperado de uma variável Y (variável dependente), dados os valores de algumas outras variáveis X (variáveis independentes). Assim, dadas duas matrizes de dados, X e Y , a finalidade da regressão é construir um modelo $Y = f(X)$. Tal modelo tenta explicar, ou prever, as variações em Y dada as variações em X . A regressão multivariada leva em consideração as diversas variáveis preditivas simultaneamente, modelando a variável dependente com mais exatidão. Nesse trabalho, a variável dependente são as vendas efetivas e o grupo de variáveis independentes são os indicadores do desempenho de vendas. O modelo de regressão é representado pela equação 2.1.

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i. \quad (2.1)$$

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

As etapas básicas para realizar a análise de regressão são:

- Formular o modelo geral;
- Estimar os parâmetros;
- Estimar coeficientes padronizados da regressão;
- Testes de significância do modelo;
- Determinar as relações entre variáveis;
- Verificar a exatidão das predições;
- Examinar os resíduos;
- Realizar a validação cruzada do modelo.

Um dos aspectos os mais importantes nessa análise é a seleção das variáveis independentes a ser usadas na regressão. As variáveis independentes possíveis que podem

influenciar na produtividade de vendas devem ser listadas a priori, reduzindo o custo da pesquisa (LAROSE, 2006). Entretanto, objetiva-se detectar o subconjunto de variáveis, dentre todas as disponíveis, que seja mais relevante para a formação das vendas.

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

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

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

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

$$\hat{\beta} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}. \quad (2.2)$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}. \quad (2.3)$$

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

$$SSE = \sum (y - \hat{y})^2. \quad (2.4)$$

$$SSR = \sum (\hat{y} - \bar{y})^2. \quad (2.5)$$

$$SST = \sum (y - \bar{y})^2. \quad (2.6)$$

A estatística da regressão pode ser apresentada sucintamente com uso de tabelas da análise de variância (ANOVA). Erros médios (por exemplo, MSE e o MSR) são derivados das equações 2.4, 2.5 e 2.6. Um parâmetro importante é o coeficiente de determinação múltipla, que é definida como:

$$R^2 = \frac{SSR}{SST}. \quad (2.7)$$

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

O uso de análise da regressão pressupõe aderência a um conjunto de suposições:

- As variáveis preditoras devem ser linearmente independentes;
- Os termos do erro devem ser normalmente distribuídos e independentes;
- A variância dos termos de erro deve ser constante;
- Deve ser utilizada uma amostra representativa da população para correta inferência;
- A distribuição da variável dependente deve ter variância aproximadamente constante, isto é, suposição do homoscedasticidade.

As suposições inerentes à análise de regressão múltipla se aplicam às variáveis individuais e a relação como um todo. Nesta técnica, uma vez que a variável estatística (combinação linear de variáveis formada na técnica multivariada determinando-se pesos empíricos aplicados a um conjunto de variáveis usadas coletivamente para prever a variável dependente) e tenha sido determinada, ela atua coletivamente na previsão da variável dependente. Isso necessita de avaliação das suposições não apenas das variáveis individuais, mas também para a variável estatística em si. Desta forma, se deve concentrar no exame da variável estatística e da sua relação com a variável dependente para atender as suposições de regressão múltipla. Essas análises devem ocorrer não apenas nas fases iniciais da regressão, mas também depois que o modelo for ajustado (HAIR, J. F. *et al.*, 2005).

Atender às suposições da análise de regressão é essencial para garantir que os resultados obtidos sejam realmente representativos da amostra e que se tenham os melhores resultados possíveis. Quaisquer violações sérias das suposições precisam ser detectadas e, eventualmente, corrigidas. Para garantir que a pesquisa está atendendo às suposições básicas da análise de regressão deve ser feito o teste das variáveis individuais independentes e dependente, bem como a relação geral após a estimativa do modelo.

Os erros na previsão podem ser resultados de uma ausência real de uma relação entre as variáveis ou elas são causadas por algumas características dos dados não acomodados pelo modelo de regressão. Desta forma, é fundamental se analisar as seguintes características:

Linearidade do fenômeno medido, Variância constante dos termos de erro, Independência dos termos de erro e Normalidade da distribuição dos termos de erro (HAIR, J. F. *et al* , 2005). A principal medida de erro de previsão para a variável estatística é o resíduo que é a diferença entre os valores observados e previstos para a variável dependente, sendo a representação gráfica dos resíduos versus as variáveis independentes ou previstas o um método básico para se identificar as violações de suposições para a relação geral.

2.2.4) Áreas de Aplicação de Técnicas de Mineração de Dados

A seguir, são relacionadas às principais áreas de interesse na utilização de mineração de dados, de acordo com (DRAPER & SMITH, 1981) e (MANNILA, 1997):

- *Marketing e Vendas*: Técnicas de mineração de dados são aplicadas para descobrir preferências do consumidor e padrões de compra, com o objetivo de realizar marketing direto de produtos e ofertas promocionais, de acordo com o perfil do consumidor e encontrar correlações escondidas entre diferentes indicadores de desempenho de vendas;
- *Detecção de Fraudes*: Muitas fraudes óbvias (tais como, a compensação de cheque por pessoas falecidas) podem ser encontradas sem mineração de dados, mas padrões mais sutis de fraude podem ser difíceis de serem detectados, por exemplo, o desenvolvimento de modelos que predizem quem é um bom cliente ou aquele que poderá se tornar inadimplente em seus pagamentos;
- *Medicina*: caracterizar comportamento de paciente para prever visitas, identificar terapias médicas de sucesso para diferentes doenças, buscar por padrões de novas doenças;
- *Instituições Governamentais*: descoberta de padrões para melhorar as coletas de taxas ou descobrir fraudes;
- *Ciência*: técnicas de mineração de dados podem ajudar cientistas em suas pesquisas, por exemplo, encontrar padrões em estruturas moleculares, dados genéticos, mudanças globais de clima, oferecendo conclusões valiosas rapidamente;
- *Controle de Processos e Qualidade*: auxiliar no planejamento estratégico de linhas de produção e buscar por padrões de condições físicas na embalagem e armazenamento de produtos;
- *Base*: detectar padrões de uso de cartão de crédito fraudulento, identificar clientes

“leais”, determinar gastos com cartão de crédito por grupos de clientes, encontrar correlações escondidas entre diferentes indicadores financeiros;

- *Apólice de Seguro*: análise de reivindicações – determinar quais procedimentos médicos são reivindicados juntos, prever quais clientes comprarão novas apólices, identificar padrões de comportamento de clientes perigosos, identificar comportamento fraudulento;
- *Transporte*: determinar as escalas de distribuição entre distribuidores, analisar padrões de carga.

2.3 Medição de Desempenho

Os modelos tradicionais de medição de desempenho têm muitas limitações que os tornam menos aplicáveis frente às condições impostas pelo mercado altamente competitivo. Por isso, muitos autores sugerem um novo conjunto de medidas de desempenho a ser usadas, que seja capaz de fornecer aos gerentes, supervisores e operadores informação relevante em tempo real (GHALAYINI et al., 1996).

Um Sistema de Medição de Desempenho (SMD) deve ser dinâmico e não estático como geralmente ocorre na maioria das empresas. Ele deve ser também sensível às mudanças interna e externa do ambiente da empresa; estar focado na revisão e re-priorização dos objetivos internos quando são significantes; manifestar um alinhamento entre os objetivos internos e as partes críticas da empresa; e também deve ser mantido um programa de melhoria com ganhos significativos para a organização (BITITCI et al., 1997).

2.3.1 Visão Geral sobre Medição de Desempenho

A medição de desempenho, segundo (NEELY, 1998), não é fácil de ser definida. Ela é um conjunto de medidas de desempenho que são usadas para quantificar a eficiência (quão economicamente os recursos são utilizados) e a eficácia (o quanto dos requerimentos dos clientes são realizados) das ações passadas. Porém esta definição ignora o fato que um SMD abrange uma infra-estrutura de suporte. (NEELY, 1998) assim se manifesta sobre uma definição de medição de desempenho que abrange o aspecto da infra-estrutura de suporte: “Um sistema de medição de desempenho possibilita que decisões e ações sejam tomadas com base em informações porque ele quantifica a eficiência e a eficácia de decisões passadas por meio da aquisição (coleta),

compilação (tratamento), categorização (classificação em categorias), análise (busca de padrões nas categorias), interpretação (explicar as implicações em cada categoria e no todo) e disseminação (comunicação das implicações) de informações adequadas”.

Um modelo apresentado por (NEELY et al., 1995) de como a medição de desempenho pode ser esboçada está ilustrada na Figura 2.11.

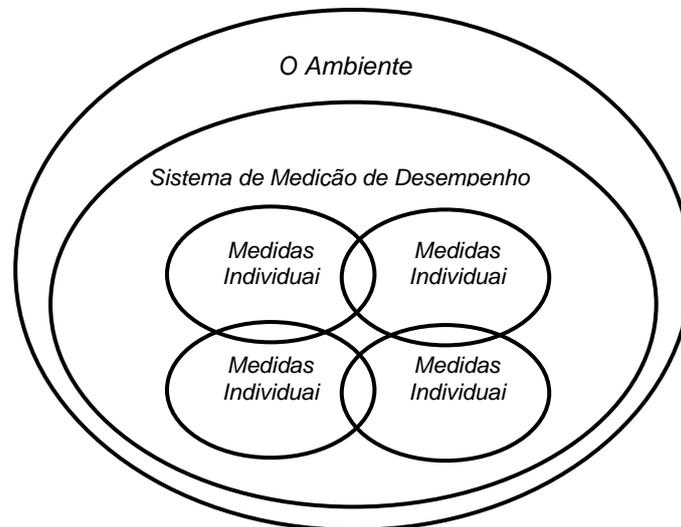


Figura 2.11 - Um modelo para o *design* da medição de desempenho.

Fonte: Adaptada pelo autor a partir de (NEELY et al., 1995).

As medidas de desempenho individuais podem ser caracterizadas por dois parâmetros: um deles é *o quê* deve ser medido, que é a definição das prioridades competitivas (qualidade, flexibilidade, custo, confiabilidade de entrega e rapidez na entrega e de *lead time*); e *o como* deve ser medido, que reflete as categorias de recurso de obtenção dos dados, o tipo de dados, a referência do seu uso e o processo de sua orientação (WHITE, 1996).

A dimensão do SMD já se refere ao agrupamento das medidas individuais e os relacionamentos entre elas. O SMD, nas organizações atuais, é um dos componentes do sistema de gestão (Figura 2.12) e está alinhado com a estratégia dos negócios que reflete o atendimento dos desejos dos *stakeholders* como acionistas, empregados, clientes, fornecedores, entre outros, conforme mostram NEELY et al. (2001). A Figura 2.12 explicita as relações entre os níveis processo de gestão do desempenho, sistema de medição de desempenho e sistema de informação.

O processo de gestão de desempenho alinha as estratégias funcionais e corporativas com o SMD. Esse é influenciado por diversos aspectos na organização: fatores comportamentais e culturais; questões de definição das medidas de desempenho; modo de prover os dados de desempenho para outros sistemas e também a influência da TI. Já o sistema de informação atende às necessidades do SMD. A eficácia do sistema de informação depende de como a informação é usada para gerenciar o desempenho dos negócios.



Figura 2.12 - Posicionamento do sistema de medição de desempenho em uma organizaç3o.

Fonte: Adaptada pelo autor a partir de (BITITCI et al., 1997).

J3 a dimens3o ambiente trata de como as medidas individuais e/ou o SMD s3o utilizados em relaç3o aos ambientes interno (organizaç3o) e externo (concorr3ncia e consumidores) (NEELY et al., 1995). A mediç3o de desempenho possui v3rias raz3es para ser utilizada. Destaca-se dentre as raz3es, os quatro CP's apresentados por (NEELY, 1998), que atuam nas atividades: a) "checar" o posicionamento no mercado; b) "comunicar" a posiç3o para a organizaç3o e os *stakeholders*; c) "confirmar" prioridades para o gerenciamento, tomada de decis3o, controle de custos e investimento; e d) "compelir" o progresso como significado para motivaç3o, comunicaç3o das prioridades da organizaç3o e base para recompensa.

A mediç3o de desempenho pode ser usada nas organizaç3es para servir como uma linguagem a serviç3o da ret3rica (forma de convencimento), induzir uma aç3o robusta (forte consist3ncia) nos membros da organizaç3o e ser a base para o pagamento de recompensas que cria certa ambigüidade (entre os objetivos da organizaç3o e de cada funcion3rio) que preserva a identidade individual de cada um.

A) Import3ncia dos Relacionamentos entre as Medidas de Desempenho

Para que um SMD seja eficiente e eficaz, 3 necess3rio que as informaç3es disseminadas sejam 3teis para a tomada de decis3o e aç3o. Para tanto, certos elementos s3o necess3rios (MARTINS, 1998):

- Exist3ncia de uma rede de relacionamento entre as medidas de desempenho;
- Uso das informaç3es para planejar, controlar e melhorar o desempenho;

- Integração com outras metodologias e ferramentas;
- Integração com sistema de recompensas;
- Indução de atitude nos funcionários;
- Legitimação do discurso da gerência; e
- Uso de medidas de desempenho específicas por alguns setores.

O enfoque desta dissertação abrange alguns desses elementos, mas é dada mais ênfase para a existência de uma rede de relacionamento entre as medidas de desempenho com destaque para o estabelecimento do relacionamento entre as medidas de desempenho. A importância do relacionamento entre as medidas de desempenho é assim destacada por ECCLES & PYBURN (1992): “A questão que deve ser perguntada é por que tem sido tão difícil fazer algo que parece ser tão óbvio - criar um sistema de medição de desempenho abrangente... Um obstáculo significativo para atingir a melhoria na medição é que muito frequentemente os gerentes ignoram um passo crucial. Antes de um sistema de medição de desempenho abrangente ser criado, os gerentes precisam chegar a um acordo sobre o modelo de desempenho dos negócios da empresa — o entendimento deles sobre os relacionamentos entre as ações da gerência e os resultados”.

Como destaca MARTINS (1998), o relacionamento entre as medidas de desempenho não é novo e fica mais importante agora: “Vale a pena observar que essa propriedade não é exclusiva dos novos sistemas de medição de desempenho. Ela sempre existiu mesmo nos sistemas oriundos da Contabilidade de Custos tradicional, tão criticada pela inadequação ao novo padrão de produção emergente. Com o desempenho passando a ser multidimensional, a necessidade de entender e explicitar esse relacionamento ficou mais importante”.

Para que as relações entre a ação gerencial e os resultados sejam compreendidos, é necessário o entendimento do impacto de cada medida de desempenho nas demais medidas de desempenho. Isso é o que permitirá, em boa parte, conforme EVANS et al. (1996), conferir às medidas de desempenho o tão almejado poder de predição e também permitirá, segundo GHALAYINI et al. (1996), "atribuir prioridade aos esforços de maximização do retorno das atividades de melhoria de desempenho."

Existem alguns métodos de construção de modelos de relacionamentos entre as medidas de desempenho de um SMD na literatura (ABREU & MARTINS, 2003). Entretanto, nenhum deles procura construir o relacionamento a partir da identificação de padrões de dados históricos, ou seja, a partir de dados sobre desempenho armazenado. As propostas procuram formalizar o relacionamento esperado assim como no relacionamento de causa-e-efeito expresso no mapa estratégico do *Balanced Scorecard* (KAPLAN & NORTON, 1996).

B) Relacionamento entre as Medidas de Desempenho para a Organização Industrial

Os relacionamentos entre as medidas de desempenho podem auxiliar na manutenção da coerência entre a estratégia de manufatura e a medição de desempenho. Com as mudanças do ambiente industrial nas últimas décadas, aumentaram os riscos de se medir erroneamente o desempenho tanto pelo fato dele, para KUENG et al. (2001), ter passado a ser multidimensional como pelo fato de novas estratégias de manufatura terem surgido para lidar com questões como, aumento das transações comerciais e financeiras em âmbito mundial, redução do ciclo de inovações, entre outros (FERRAZ et al., 1995).

Para os SMD, o claro estabelecimento dos relacionamentos entre as medidas de desempenho contribui para que eles sejam explicitados, fornecendo uma visão ampla de como o desempenho da empresa foi atingido. Ele apresenta as relações entre as medidas de desempenho facilitando o entendimento e a comunicação para os decisores atuarem no controle e nas propostas de melhorias das atividades na manufatura e, portanto, verificando se a estratégia da manufatura está coerente com o que está ocorrendo no sistema de gestão.

A formalização dos relacionamentos entre as medidas de desempenho permite a externalização das relações entre elas facilitando o entendimento dos decisores sobre a estrutura de um SMD. Explicitando-se os relacionamentos entre elas, há um melhor conhecimento sobre como o SMD pode contribuir para o sistema de gestão devido à facilidade de serem analisadas e avaliadas as informações relevantes com maior confiabilidade (MORESI, 2000).

ZARIFIAN (1997) destaca que um dos fatores presentes na nova "física" do desempenho é a qualidade da comunicação interpessoal ao longo de toda a cadeia de atividades, ou seja, a constante troca de informações entre os diferentes saberes profissionais contribui para uma rápida elucidação dos problemas e o progresso a ser atingido na organização.

Com a comunicação interpessoal, há uma maior necessidade das trocas de informações entre as diversas áreas da organização para resolver problemas organizacionais e acelerar o progresso. Os modelos de relacionamentos entre as medidas de desempenho é uma das formas de apresentar uma linguagem comum entre os diferentes decisores, por exemplo, facilitar o entendimento de um decisor com modelo mental voltado para as finanças com um outro decisor com modelo mental voltado para a produção.

2.3.2 Relacionamentos entre as Medidas de Desempenho

Existem na literatura diversas formas de se representar e denominar (classificar) os relacionamentos entre as medidas de desempenho (ABREU & MARTINS, 2003). Essas formas podem ser encontradas na denominação das relações entre as medidas de desempenho, nos métodos que estabelecem relações entre elas e nos modelos de medição de desempenho.

O propósito desse item é apresentar como os relacionamentos entre as medidas de desempenho de um SMD podem ser denominados ou representados em suas diversas formas. Esse item é útil na caracterização dos relacionamentos entre as medidas de desempenho:

- Denominação das relações entre as medidas de desempenho: causa-e-efeito; relação lógica; interdependência; finalidade; *delay*; e relações espúrias;
- Métodos que estabelecem relações entre as medidas de desempenho: modelo causal; desdobramento do modelo causal; pais e filhos; *negative branch*; diagrama de construção de indicadores de desempenho estratégicos; diagrama para a construção de relacionamentos entre as medidas de desempenho; mapa cognitivo; diagrama de causa-e-efeito; diagrama de árvore; e *analytic hierarchy process* (AHP); e
- Relacionamentos entre as medidas de desempenho presentes em modelos de medição de desempenho: SMART (*Strategic, Measurement, Analysis, and Reporting Technique*) também conhecido como *Performance Pyramid*; *Balanced Scorecard* (BSC); *Integrated Dynamic Performance Measurement System* (IDPMS); *Integrated Performance Measurement System* (IPMS) e *Performance Prism*.

A) Relações entre as Medidas de Desempenho Individuais

Na literatura, a mais discutida na literatura dentre as relações entre as medidas de desempenho de um SMD é a chamada relação de *causa-e-efeito* que é, segundo KAPLAN & NORTON (1996), uma das características marcantes do *Balanced Scorecard*.

As relações de causa-e-efeito podem ser definidas, segundo NORREKLIT (2000), quando um evento X precede um evento Y no tempo. A observação do evento X necessariamente, ou altamente provável, implica na observação subsequente de outro evento Y, e os dois eventos podem ser observados no tempo e espaço e logicamente independentes. Segundo NORREKLIT (2000), esses autores justificam a existência do relacionamento entre as medidas de desempenho por conta da casualidade e também justificam a mesma lógica com base na avaliação financeira das ações pelo *Activity-Based Costing* (ABC), que levam a um resultado financeiro. No primeiro caso, a

justificativa do relacionamento é direto e no segundo é algo não claramente direto, pois existem medidas de desempenho de diferentes processos que podem afetar o resultado financeiro de diversas formas e os relacionamentos entre elas não ocorrem somente por casualidade.

Para NORREKLIT (2000), há relacionamentos entre as medidas de desempenho que são baseados em *relações lógicas* que não podem ser verificadas ou determinadas empiricamente. Por exemplo, quando é feita uma pesquisa para se dizer se dois mais dois são quatro, o teste não é feito para saber se isso é verdadeiro, mas para resolver problemas aritméticos.

Na medição de desempenho, por exemplo, um aumento do *market share* com o aumento na lucratividade gerada pela aquisição de produtos pelos clientes é o direcionador para um bom resultado financeiro. Relacionamentos desse tipo são feitos de maneira lógica e não há probabilidade de ocorrer o contrário na relação.

Existe também o relacionamento entre as medidas de desempenho baseado na *interdependência* em que estão presentes correlações positivas e negativas entre essas medidas de desempenho. Por exemplo, uma empresa cujo setor que atua investe fortemente em pesquisa e desenvolvimento (P&D) tem por objetivo bons resultados financeiros investindo em centros para P&D. A empresa usa a medida de desempenho "velocidade de lançamento de novos produtos" e esta direciona o resultado para a medida de desempenho financeira (*Return of Investment*, também chamado de ROI, por exemplo). A interdependência entre as medidas de desempenho se constata quando um fraco desempenho na medida "velocidade de lançamento de novos produtos" gera uma ação gerencial que limita os investimentos nos centros de pesquisa, mas também a mesma ação gerencial pode ocasionar um mau resultado financeiro, pois a empresa atua no setor em que a P&D é fundamental.

NORREKLIT (2000) apresenta também o relacionamento fundamentado na *finalidade* com que uma ação humana ocorre baseada na crença ou desejo de solucionar um problema (finalidade) na organização. Para isso, o executor aloca recursos para resolver o problema e associa medidas de desempenho para fazer o controle da ação e para verificar se o resultado foi atingido. Por exemplo, uma empresa que apresenta preços elevados dos produtos em relação aos concorrentes. Um gerente ou decisor pode decidir fornecer um produto bom ao mercado a um preço baixo para aumentar o *market share* da empresa, levando a um resultado financeiro satisfatório.

Outras formas de relacionamentos entre as medidas de desempenho são encontradas na literatura. As relações do tipo *delav* têm o mesmo princípio das relações de causa-e-efeito; a

diferença é que o tempo em que podem ocorrer os efeitos de uma medida de desempenho é diferente em outras medidas de desempenho (DAVIS & O'DONNELL, 1997).

Por fim, nos relacionamentos das medidas de desempenho, deve se ater ao fato de que podem existir *relações espúrias*. Duas medidas que se relacionam podem estar sendo afetadas indiretamente por uma terceira sem tal fato ser levado em consideração. Isso implica que o relacionamento entre as medidas de desempenho deve estar constantemente sendo monitorado e corrigido quando necessário para verificar se as variáveis espúrias decorrentes de contingências internas e externas não afetam consideravelmente um resultado organizacional (financeiro ou não-financeiro).

Por exemplo, uma empresa de alimentos fabrica produtos com alta velocidade e os entrega sem atraso para os distribuidores. Esse evento ocasionaria um bom resultado financeiro, porém a empresa não percebe pequenas mudanças no comportamento dos consumidores, que podem ser medidas pelo seu grau de satisfação, fazendo com que o produto não seja consumido, afetando conseqüentemente o desempenho financeiro.

B) Métodos para a Construção de Relacionamentos entre as Medidas de Desempenho

Atualmente, nas organizações há a necessidade de se mapear as relações entre as medidas de desempenho, com o intuito de resolver questões como a identificação das medidas de desempenho direcionadoras do resultado, e de aumentar a capacidade preditiva das medidas de desempenho (NEELY, 1999).

Vários métodos podem ser utilizados na construção de relacionamentos entre essas medidas. Aqueles apresentados nesse item são baseados em autores com trabalhos voltados para esta área. LEBAS (1995) apresenta o *modelo causal* para identificar as causas que afetam o desempenho da empresa. O método é apresentado em forma de árvore, onde as raízes são os princípios que formam o desempenho na empresa, o tronco representa os processos de formação das dimensões de desempenho, e as folhas representam as dimensões de desempenho em uma organização. O mesmo autor faz o *desdobramento do modelo causal* para representar a hierarquia dos elementos que compõem as medidas de desempenho na organização da empresa.

Em sua proposta de um método para projetar um SMD com foco nas relações entre indicadores de desempenho, FLAPPER et al. (1996) definem que, em geral, existem dois tipos de relacionamento, que são os indicadores de desempenho usados no contexto de uma mesma função (relações internas), e indicadores de desempenho definidos para funções diferentes (relações

externas). Os autores apresentam uma estrutura de relacionamento entre indicadores do tipo "pais" e "filhos".

Um método apresentado por BOYD & COX III (1997), denominado de *negative branch*, que analisa o impacto das decisões tomadas no desempenho organizacional. O método apresenta os seguintes passos na sua construção:

- A apresentação dos possíveis efeitos positivos que são esperados das ações propostas;
- A apresentação dos possíveis efeitos negativos que são esperados das ações propostas;
- A conexão das soluções propostas com os possíveis efeitos positivos e negativos dos relacionamentos de causa-e-efeito;
- O uso da lógica (se - então) para verificar se as conexões apresentam coerência para os usuários do método.

Esse método desenvolve-se a partir da criação de uma hipótese que gera diversos cenários. Esses dependem das escolhas das conexões entre os efeitos negativos e positivos das ações propostas.

HEREDIA & NATARAJAN (1997) apresentam um *diagrama de construção de indicadores de desempenho estratégicos* que mostra as estratégias, objetivos e medidas de desempenho, onde são utilizadas as perspectivas do *Balanced Scorecard (BSC)*. O uso do diagrama, segundo os autores, auxilia estudos de simulação que podem ser conduzidos para validar as escolhas feitas em etapas anteriores dentro do planejamento de um SMD integrado.

Uma proposta de método para estabelecer relações entre as medidas de desempenho de uma organização é feita por NICKOLS (1999). O método é o *diagrama para a construção de relacionamentos entre as medidas de desempenho* e se baseia em três questões: "Qual é a medida de desempenho-alvo?"; "Como ela é calculada?"; e "Quais são as medidas de desempenho que são variáveis de entrada para a medida de desempenho-alvo?". Respondendo essas três questões, inicia-se a construção de um diagrama de árvore, e o processo de uso das questões para outra medida de desempenho é repetido até serem todas mapeadas no diagrama. O diagrama de árvore mostra as relações entre as medidas de desempenho definidas pelas fórmulas de cálculo de uma medida de desempenho derivada de outras.

O escopo dos diagramas propostos por NICKOLS (1999) é bem definido devido ao uso das relações existentes nas fórmulas consideradas. Esse modelo pode encontrar, em SMD, problemas de utilização nas dimensões de desempenho não-financeiras, pois podem existir medidas de desempenho que não utilizam dados numéricos e também algumas relações entre indicadores que não são necessariamente matemáticas (ABREU & MARTINS, 2003).

O primeiro método apresentado é o *mapa cognitivo*₂ que identifica os efeitos dos fatores no desempenho. Os efeitos poderiam ser classificados em: direto (o fator que afeta está relacionado diretamente a um outro fator); indireto (fatores que afetam um fator que está relacionado indiretamente ao fator em questão); e *self-interactive* (é o efeito do fator nele mesmo). Os sinais positivos ou negativos indicam efeitos positivos ou negativos de um fator recebido de um outro fator de desempenho.

O *diagrama de causa-e-efeito* ajuda a identificar a estrutura hierárquica das medidas de desempenho. Existe um nível O, que é afetado por todas as outras medidas de desempenho, mas ele não afeta todas as outras medidas de desempenho e, conseqüentemente, o próximo nível afeta o nível O, mas não as outras medidas de desempenho.

O *diagrama de árvore* possui o mesmo princípio que o diagrama de causa-e-efeito, mas ele é mais utilizado no intuito de dar uma visão mais clara da estrutura hierárquica das medidas de desempenho.

Por fim, dentre os quatro métodos ilustrados em SUWIGNJO et al. (2000), há o *Analytic Hierarchy Process* (AHP). O processo de quantificação do AHP é baseado, na primeira etapa, na comparação dos fatores de desempenho, em que cada par de fatores é comparado e verificado o seu efeito direto ou indireto em outro fator de desempenho, posteriormente, na segunda etapa, é apurado o quão forte é esse efeito em termos quantitativos.

Na primeira etapa do processo por meio da questão "Comparando-se os fatores A e B, qual deles tem o maior efeito no fator C?", é possível traçar a intensidade do efeito entre os fatores de desempenho. O cálculo para a determinação da prioridade dos fatores é feito a partir da soma dos componentes: efeito direto do fator, efeito *self-interaction*, efeitos indiretos dos outros fatores no fator e efeito do fator nos outros fatores. Os métodos de construção dos relacionamentos entre as medidas de desempenho devem ser usados de acordo:

- Com as necessidades organizacionais;
- Com a adequação do método ao sistema de gestão;
- Com o tipo de ação a ser tomada na organização.

C) Relacionamentos entre as Medidas de Desempenho nos Modelos de Medição de Desempenho

Diversos modelos de medição de desempenho surgiram a partir da década de 80 do século passado, resultantes das mudanças do paradigma fordista/taylorista (MARTINS, 1998). No trabalho desse mesmo autor, foram identificadas 26 novas propostas de modelos de medição de desempenho surgidas nas décadas de 80 e 90 do século passado.

Esse item tem o intuito de apresentar como os principais (mais citados na literatura) modelos de medição de desempenho exploram a questão dos relacionamentos entre as medidas de desempenho. Os modelos analisados são: SMART (*Performance Pyramid*); *Balanced Scorecard* (BSC); *Integrated Dynamic Performance Measurement System* (IDPMS); *Integrated Performance Measurement System* (IPMS); e *Performance Prism*, para as quais são descritos abaixo suas principais características e relações:

C.1.-) SMART *Strategic, Measurement, Analysis and Reporting Technique*

O SMART também conhecido como *Performance Pyramid*, tem o intuito de levar a visão corporativa (objetivos financeiros e de mercado) até os departamentos e os centros de trabalho, onde estão os fluxos de trabalho do dia-a-dia. A Figura 2.13 apresenta a estrutura do modelo de medição de desempenho e também apresenta quais questões são importantes a serem respondidas a partir de cada aspecto do modelo.

Analisando-se a Figura 2.13, verifica-se que o foco é a integração entre as atividades operacionais, de forma que elas administrem as medidas de desempenho operacionais que sustentam a visão da corporação. Os níveis de unidade de negócios (mercado e finanças) e de sistemas de operação do negócio (satisfação do cliente, flexibilidade e produtividade) intermedeiam a ligação entre os departamentos e a alta administração (CROSS & LYNCH, 1990).

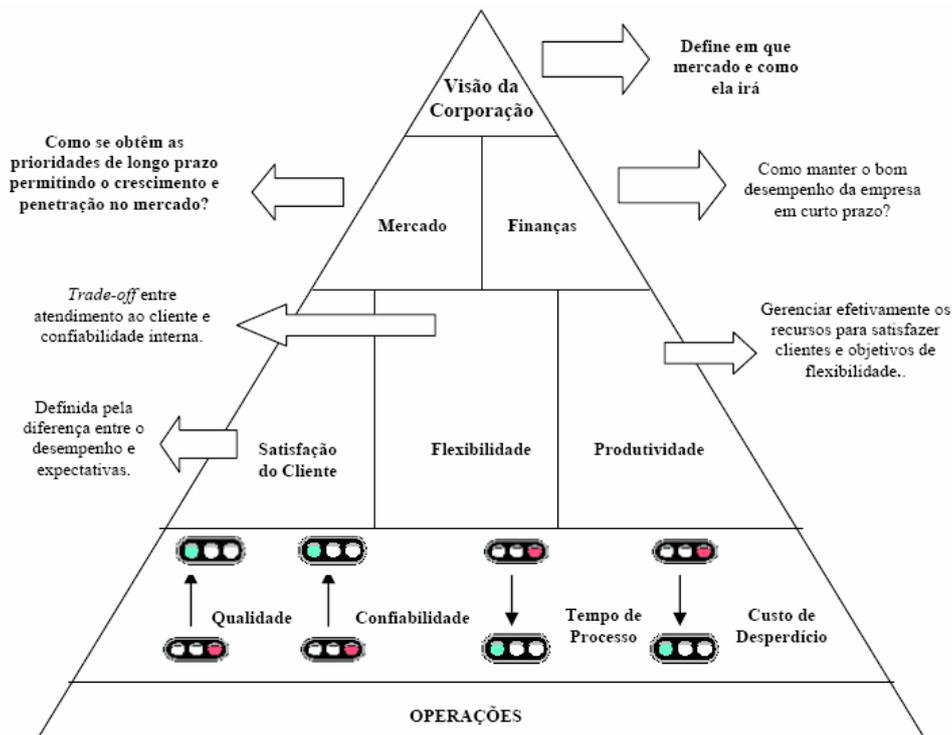


Figura 2.13 - Estrutura do modelo de medição de desempenho *Performance Pyramid*.

Fonte: Adaptada pelo autor a partir de CROSS & LYNCH (1990).

Os relacionamentos entre as medidas de desempenho expressos pelos mesmos autores na *Performance Pyramid* se dão pela transição entre duas linguagens. A linguagem do "dinheiro" (informações provenientes de dados financeiros) para a alta administração, que possui a visão corporativa e que precisa traduzir a linguagem das "coisas" (informações provenientes de dados não-financeiros) da média gerência. Dados de desempenho não-financeiros pertencentes à média gerência são utilizados pela alta administração para a sumarização de medidas de desempenho financeiras como taxas e índices. O inverso ocorre com a média gerência, que deve traduzir a linguagem do "dinheiro" da alta administração para a linguagem das "coisas". Dados de desempenho financeiros são utilizados para a sumarização de medidas de desempenho voltado para produtos específicos, processos, unidades organizacionais, maiores mercados do produto, entre outros.

C.2-) BSC *Balanced Scorecard*

O BSC tem como objetivo prover um conjunto balanceado de medidas de desempenho dentro de quatro perspectivas (financeira, clientes, processos internos de negócio e aprendizagem e crescimento). O BSC prove realimentação sobre os processos de negócio internos e resultados externos para melhorar continuamente o desempenho. As quatro perspectivas sugeridas mostram o

retrato do desempenho operacional bem como os direcionadores - do desempenho futuro (KAPLAN & NORTON, 1996). O *BSC* é uma ferramenta empresarial que traduz a missão e a estratégia da organização em um conjunto compreensível de medidas de desempenho, propiciando a formação de uma estrutura de mensuração estratégia e de um sistema de gestão eficiente (KAPLAN & NORTON, 1996).

Algumas companhias inovadoras estão utilizando o *BSC* para desenvolver suas perspectiva em longo prazo, clarificando e traduzindo visão estratégia, comunicando e integrando objetivos e indicadores, planejamento, estabelecendo metas e alinhando iniciativas, aprimorando o controle e o aprendizado (KAPLAN & NORTON, 1996). O modelo proposto por KAPLAN & NORTON (1996), estabelece um sistema de comunicação a partir de quatro perspectivas. Esta estrutura possibilita a articulação dos resultados pretendidos pela organização avaliando o desempenho atual e as perspectivas futuras. Os indicadores selecionados visam à comunicação das estratégias da organização e contribuem com o alinhamento das iniciativas individuais e coletivas para a realização de metas comuns. A seguir são detalhadas as perspectivas propostas por esses autores.

- *Perspectiva Financeira*

A importância das medidas financeiras está na capacidade de sintetizar as consequências econômicas imediatas de ações consumadas. Indicam se a estratégia da organização sua implementação e execução estão contribuindo para melhoria dos resultados financeiros. (KAPLAN & NORTON, 2004). Os objetivos financeiros podem ser diferentes, dependendo do ciclo de vida da organização, e a perspectiva financeira permite medir e avaliar resultados que o negócio proporciona e necessita para seu crescimento e desenvolvimento, assim como para a satisfação dos seus acionistas (MARTINS, 2002).

- *Perspectiva dos Processos Internos*

É recomendável que as organizações definam completamente cada cadeia de valor dos processos internos. Deve ser iniciado com o processo de inovação, identificação das necessidades atuais e futuras dos clientes e desenvolvimento de novas soluções para essas necessidades, prosseguindo com processos de operações, entrega dos produtos e prestação de serviços aos clientes, e terminando com serviço pós-venda, que complementem valor

proporcionado ao cliente pelos produtos ou serviços de sua organização. (KAPLAN & NORTON, 2001).

No entanto, segundo (MARTINS, 2002), esse desenvolvimento tem que ser sustentado por uma análise e intervenção constante na performance dos processos internos e os processos-chaves do negócio. A melhoria dos processos internos no presente é um indicador chave do sucesso financeiro no futuro.

- *Perspectiva do Cliente*

Clientes são todos aqueles a que uma organização serve, afirma (NICKLOS, 1999). (KAPLAN & NORTON, 2004) complementam que os clientes representam as fontes de receita para atingir os objetivos financeiros da organização. Dessa forma é necessário que as organizações identifiquem os segmentos de clientes e mercados nos quais desejam competir, alinhando medidas essenciais de resultados relacionados aos clientes: satisfação, fidelidade, retenção, captação e lucratividade, com segmentos específicos de clientes e mercado. Mas, para traduzir os processos em sucesso financeiro, as organizações devem e têm, em primeiro lugar, que satisfazer os seus clientes.

- *Perspectiva de Aprendizado e Crescimento*

A perspectiva de aprendizado fornece a infra-estrutura necessária para a consecução de objetivos ambiciosos nas outras perspectivas do *BSC* e que os objetivos são os vetores de resultados. Na opinião de (EPSTEIN & MONZONI, 1998) as organizações podem criar outras “caixas” de indicadores, além das quatro perspectivas propostas pelo *BSC*, devendo contemplar um importante aspecto em relação a fornecedores, que aparentemente foi esquecido pelo *BSC*.

Segundo (KAPLAN & NORTON, 2001), muitos executivos, quando tentam implementar a estratégia do negócio, fornecem aos empregados apenas algumas descrições limitadas sobre o que deveriam fazer e porque aquelas tarefas são importantes. Sem informações claras ou detalhadas é quase impraticável colocar a estratégia em execução.

É preciso utilizar ferramentas para comunicar tanto a estratégia quanto os processo e sistema que ajudarão na sua implementação. A finalidade dos mapas estratégicos é, justamente, fornecer uma representação visual dos objetivos críticos da companhia e os principais relacionamentos entre eles que dirigem o desempenho organizacional. A perspectiva de aprendizado direcionando a sua atenção para as pessoas e para as infra-estruturas de recursos

gestão (*Total Quality Management*, JIT, entre outros), o IDPMS foi desenvolvido (GHALAYINI et al., 1996).

O IDPMS é um modelo baseado nos seguintes fatores: a) integração das equipes de melhoria; b) gestão e medição de desempenho no chão de fábrica; c) integração das áreas gerais de sucesso com medição e indicadores de desempenho associados; e d) integração da medição de desempenho financeira com a medição de desempenho operacional. Um exemplo do IDPMS está na Figura 2.15.

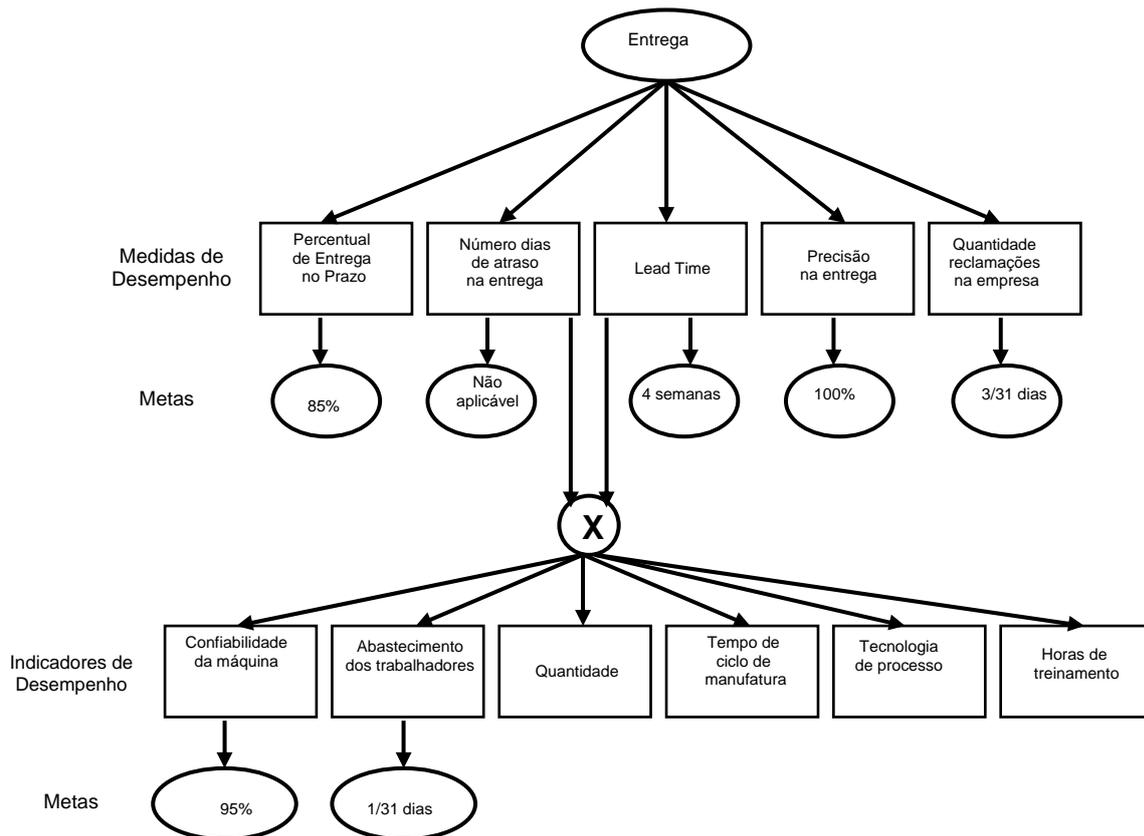


Figura 2.15 - Analisando o desempenho de entrega no IDPMS.

Fonte: Adaptada pelo autor a partir de GHALAYINI et al. (1996).

O IDPMS, segundo GHALAYINI et al. (1996), possui uma estrutura que integra as áreas de sucesso, as medidas de desempenho e suas metas, e os indicadores de desempenho e suas metas. A Figura 2.15 mostra os quatro níveis de integração dos aspectos ligados à medição de desempenho para o fator desempenho da entrega.

C.4-) IPMS *Integrated Performance Measurement System*

O *IPMS* é visto como o meio para desdobrar (traduzir em objetivos) a visão e as estratégias da organização e também permite controlar se os objetivos foram atingidos ou não. Nesse sentido, o SMD integra os esforços na organização. A integração se dá entre a unidade

operacional, a supervisão, a média gerência, um sistema que atua visando à melhoria da organização por meio do *benchmarking* e alta administração (BITITCI et al., 1997).

No IPMS, o SMD está estruturado a partir dos critérios de integração do sistema e das diretrizes. A integração do sistema se dá a partir de subsistemas voltados para cada nível organizacional. Os subsistemas, segundo BITITCI et al. (1997), foram idealizados a partir de um estudo de um grupo de pesquisa da Universidade de Strathclyde e abrangem quatro níveis organizacionais (corporação, unidade de negócio, processos de negócio e atividade).

Cada subsistema é formado por elementos que se relacionam na seguinte ordem: (a) *stakeholders*; (b) medidas de controle; (c) posição no ambiente (se relaciona com o ambiente externo); (d) objetivos de melhoria; e (e) medidas de desempenho internas. O item (e) de um subsistema se relaciona com o item (a) de outro subsistema, ambos condicionados pelas diretrizes dos objetivos de negócio e políticas organizacionais (Figura 2.16).

O relacionamento entre as medidas de desempenho ocorre por meio das interações entre os subsistemas. Os dados de desempenho obtidos por um subsistema são utilizados por outros subsistemas. Por exemplo, na Figura 2.16 o subsistema do nível de unidade de negócios utiliza as metas de desempenho ou os dados históricos de desempenho do subsistema do nível corporativo para estabelecer suas metas, enquanto o subsistema nível corporativo utiliza os dados de desempenho do subsistema do nível unidade de negócios para verificar se as metas estabelecidas pelas medidas de desempenho interna estão sendo atingidas.

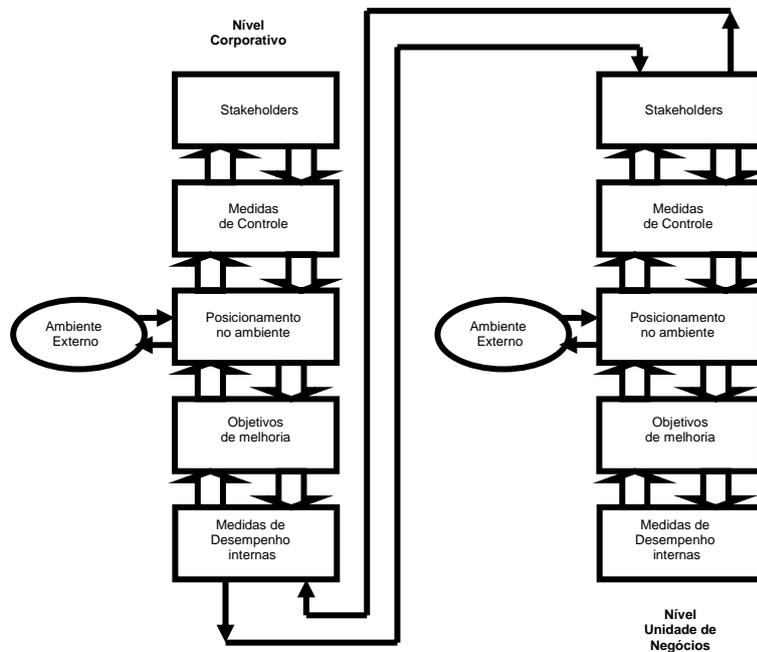


Figura 2.16 - Integração do sistema de medição de desempenho no IPMS.

Fonte: Adaptada pelo autor a partir de BITITCI et al. (1997).

C.5-) Performance Prism

Por fim, o *Performance Prism* é considerado um modelo de medição de desempenho abrangente orientado para os *stakeholders*, que encoraja os executivos a considerar que todos os *stakeholders* de uma organização têm desejos e necessidades a serem atingidas, como também a associar a eles estratégias, processos e capacidades. As escolhas das medidas de desempenho são feitas a partir do desenvolvimento das estratégias advindas das demandas do *stakeholders*. Elas são estabelecidas para cada faceta do modelo ou também para as chamadas facetas do prisma (satisfação dos *stakeholders*, contribuição dos *stakeholders*, estratégias, capacidades e processos) e associadas a um elemento-chave desses.

Os relacionamentos entre as medidas de desempenho ocorrem por meio da definição dos *stakeholders*. Essa definição pode ser feita pelo mapa de sucesso que é construído por meio de diversas sessões de *brainstorming* nas quais foram definidas e discutidas questões críticas para cada faceta do prisma. Definindo-se quais *Stakeholders* podem contribuir e quais aqueles que têm que ser prioritariamente atendidos em suas demandas, é possível estabelecer e verificar o encadeamento das medidas de desempenho para todas as cinco facetas do prisma (NEELY et al., 2001).

Destaca-se que alguns modelos, como o *Balanced Scorecard* e o IDPMS, apresentam de forma mais detalhada como os relacionamentos estão estruturados, enquanto em outros modelos isso é feito de forma obscura, sendo necessário lançar hipóteses de construção de relacionamentos entre as medidas de desempenho ou ter outras ferramentas de apoio (por exemplo, o mapa de sucesso no *Performance Prism*).

2.3.3 Uso da Tecnologia de Informação para Construir Relacionamentos entre as Medidas de Desempenho

Dois aspectos são importantes para os SMD vistos como solucionadores de problemas organizacionais:

- É a questão da busca, que seria desenvolver um *design* do SMD capaz de capturar características relevantes das pessoas, equipes, redes, cadeia de suprimentos e organizações com base em parâmetros pré-definidos (VAKKURI & MEKLIN, 2003);
- É que, mesmo tendo uma busca eficaz, necessariamente não é resolvido o problema de como converter a informação de maneira mais inteligível para a

tomada de decisão. Esta envolve diversos fatores que podem tornar ambígua a decisão racional do uso ou não uso daquela informação encontrada na busca (VAKKURI & MEKLIN, 2003).

Na medição de desempenho, segundo KUENG et al. (2001), as ferramentas de TI podem ajudar de forma substancial as atividades de um SMD, tanto na questão de busca, como na conversão da informação em uma maneira acessível para a tomada de decisão. Vale observar que a TI é um dos apoios e influências para o processo de gestão do desempenho na estruturação do SMD, segundo BITITCI et al. (1997), como mostrado na Figura 2.17.

Ela pode auxiliar um SMD na automatização da atividade de obtenção de informações, o que KUENG et al. (2001) definem como o processo de obtenção de dados relevantes de desempenho provenientes de diferentes recursos, permitindo rapidamente comparar os valores atuais com os dados históricos e as metas e comunicar os resultados para os atores organizacionais incumbidos da tomada de decisão. Segundo esse mesmo autor, existem quatro componentes em um SMD que podem formar um sistema de TI: (i) procedimentos suportados pela TI; (ii) dados; (iii) *software*; e (iv) *hardware*. O quinto componente, que, embora não forme um sistema de TI, é complementar a esse sistema, seriam as pessoas.

ECCLES (1991) considera que a TI possui um importante papel nas cinco atividades fundamentais de auxílio à medição de desempenho, na qual a gestão deve articular uma gramática comum e definir um especial "vocabulário" para o modelo de negócios da companhia. O autor considera a medição de desempenho como uma possibilidade de ser essa gramática. A Figura 2.17 mostra as cinco atividades da medição de desempenho.

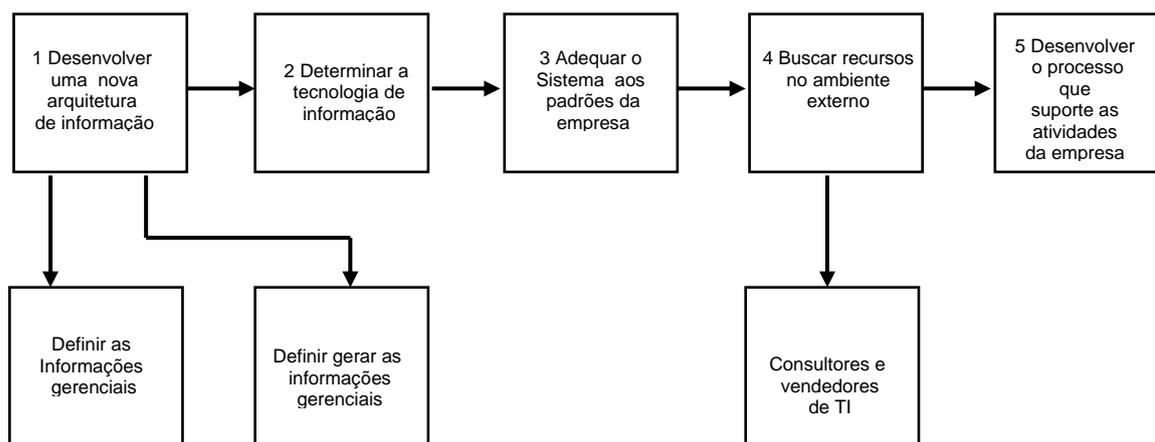


Figura 2.17 - Cinco atividades fundamentais da medição de desempenho.

Fonte: Adaptada pelo autor a partir de ECCLES (1991).

Verifica-se na Figura 2.17 que a TI possui um importante papel nas atividades, principalmente nas atividades 2 e 4, nas quais ela possui um papel de interação com a medição de desempenho tanto na definição da tecnologia de informação apropriada quanto na obtenção de

informações do ambiente externo (clientes, concorrentes, entre outros). Contudo, na realidade, a TI ainda é pouco considerada no auxílio à SMD destacado por KUENG et al. (2001): “De forma geral, os vários modelos de medição de desempenho sugeridos na última década enfatizam muito as dimensões e as medidas de desempenho (indicadores de desempenho) que devem ser considerados. Em contraste, as questões e os aspectos relacionados à implementação que tratam da Tecnologia de Informação como suporte à medição de desempenho (processos de coleta de dados eficientes, armazenamento e gestão dos dados sobre desempenho, e a disseminação dos resultados sobre desempenho) foram negligenciados em alto grau”.

A dificuldade de utilizar os dados históricos sobre desempenho na obtenção de informações reside nos embaraços em manipular, de forma rápida e amigável, uma grande quantidade de dados. Vale ressaltar que uma característica dos novos SMD é trabalhar com uma quantidade maior de medidas de desempenho financeiras e não-financeiras coletadas com maior frequência. Entretanto, o investimento feito na última década pela maioria das organizações em sistemas de informação pode ser aproveitado com a utilização de aplicações de *data warehouse*, OLAP e *Mineração de Dados* para a construção de modelos de relacionamentos a partir de dados históricos de desempenho.

Dentro do panorama apresentado nesse item, esse trabalho propõe um método para o estabelecimento de relacionamentos entre as medidas de desempenho procurando atender a dois aspectos: o aumento da complexidade na análise do desempenho e o uso da TI no suporte ao SMD desenvolvendo uma forma de aplicação das ferramentas da TI por meio da *Mineração de Dados*.

2.4 Vendas Industriais

Atualmente, as organizações industriais vêm enfrentando situações de mercado que fizeram crescer a importância do marketing estratégico no setor industrial da economia, como a crescente competitividade, mercados com baixo índice de crescimento, economia instável e globalizada, e principalmente clientes mais exigentes e sofisticados. Pressões do lado dos custos (aumento da competitividade) e dos preços (clientes mais exigentes) conduzem a organização ao gerenciamento da rentabilidade, e não mais apenas da receita e da participação de mercado. Esse ambiente, como visto anteriormente, reforça a necessidade do desenvolvimento da orientação para mercado e de uma correta segmentação de mercado e seleção dos mercados-alvo em que se deseja atuar. De acordo com WEBSTER (1991)

estrategicamente o foco central do mercado industrial não deve ser em produtos ou mercados, mas sim no relacionamento entre o comprador e o vendedor.

O mercado industrial também é conhecido como mercado empresarial, *business to business* ou ainda mercado organizacional. De acordo com HUTT & SPEH (2002), o volume de dinheiro movimentado pelo mercado industrial supera de forma significativa o volume movimentado pelo mercado consumidor. O mercado industrial pode ser dividido de diversas formas. WEBSTER (1991) sugere a divisão clássica em: construção, equipamento pesado, equipamento de iluminação, componentes, matérias-primas, materiais processados, fornecimento de manutenção e reparo e serviços. KOTLER (2003) descreve que o mercado organizacional congrega todas as organizações que compram bens e serviços a serem utilizados na produção de outros produtos ou serviços que são revendidos, alugados ou fornecidos para outros a um dado lucro.

Os clientes industriais diferem dos consumidores finais em diversos aspectos: são em menor número, o volume de compra é consideravelmente maior, estão geograficamente mais concentrados, a demanda do mercado industrial é derivada do mercado consumidor, as compras envolvem mais compradores que em geral são mais profissionais. WEBSTER (1991) destaca que a diferença entre o consumidor industrial e o final está mais na natureza do cliente do que na natureza do produto, uma vez que em muitos casos o produto comprado é o mesmo - o que difere é a natureza da compra.

Já HUTT & SPEH (2002) apontam que a principal diferença entre o mercado *business to business* e o mercado de consumo é o tipo de cliente e a forma em que o cliente usa o produto. Outra característica importante nos mercados industriais trata da natureza da relação entre comprador e vendedor. Segundo WEBSTER (1991), as relações estão em um continuum de transações puras, passando pela interdependência entre comprador e vendedor e chegando à aliança estratégica. Três aspectos merecem destaque: (i) a empresa consegue migrar de uma dependência das forças do mercado, em transações puras, para uma situação onde virtualmente não há forças de mercado, em alianças estratégicas; (ii) o comprador incorre em custos administrativos adicionais e crescente dependência, obtendo em troca um conjunto de benefícios relacionados à qualidade, confiabilidade e serviços; e (iii) o movimento vai de simplesmente "comprar" um produto para uma decisão de "fazer" o produto ou serviço juntamente com o cliente organizacional. O modelo esquemático de Frederick Webster é apresentado na Ilustração 2.1, a seguir.

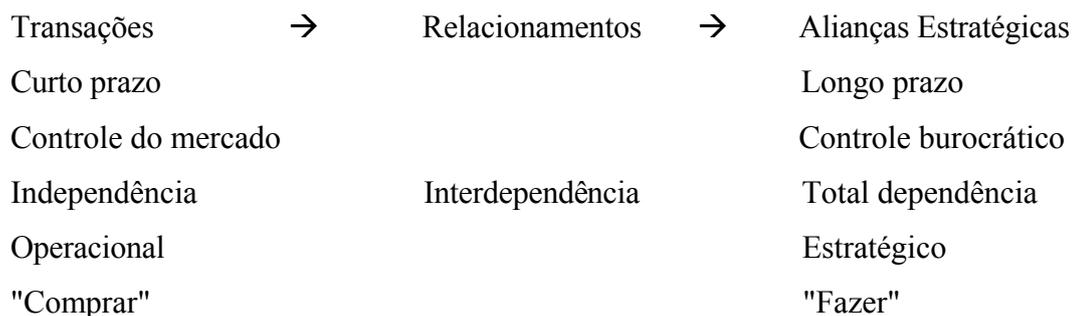


Ilustração 2.1 *The continuum of industrial buying situations.*

Fonte: Adaptada pelo autor a partir de WEBSTER, 1991.

2.4.1 Mercado Empresarial

A atividade de mercado pode ser classificada em mercado de consumo e em mercado empresarial. O mercado de consumo é dirigido às famílias e aos indivíduos, para consumo próprio. O mercado empresarial é direcionado a clientes industrial-empresariais ou institucionais, que utilizariam os bens e serviços demandados no seu processo produtivo, com o objetivo de construir a oferta de seus produtos ao cliente final, o próximo elo da cadeia produtiva (TELLES, 2003).

O mercado empresarial envolve todas as atividades de comercialização de bens e serviços para organizações que os utilizarão, na produção de bens de consumo, bens industriais e serviços ou na sua comercialização (RODRIGUES FILHO & ROMAN, 2004). As principais diferenças entre o mercado empresarial e o mercado de bens de consumo estão na estrutura do mercado, produto, comportamento do comprador, canais de venda, promoção e preço, conforme mostra o Quadro 2.2.

Quadro 2.2 - Diferenças entre o mercado empresarial e o de bens de consumo.

Tipo de Mercado Dimensões	<i>Mercado Empresarial</i>	<i>Mercado de Bens de Consumo</i>
<i>Estrutura de mercado</i>	<ul style="list-style-type: none"> • Geograficamente concentrado • Relativamente poucos compradores 	<ul style="list-style-type: none"> • Geograficamente disperso • Mercado de massa
<i>Produto</i>	<ul style="list-style-type: none"> • Complexidade tecnológica • Produtos <i>Taylor Made</i> (sob medida) • Serviço, expedição e disponibilidade são muito importantes 	<ul style="list-style-type: none"> • Padronizado • Serviços, expedição e disponibilidade são relativamente importantes
<i>Comportamento do comprador</i>	<ul style="list-style-type: none"> • Envolvimento da organização é mais racional • Predomina experiência técnica • Relações estáveis • Reciprocidade 	<ul style="list-style-type: none"> • Envolvimento emocional • Motivos sociais e psicológicos predominam • Menos experiência técnica desejada
<i>Canais de venda</i>	<ul style="list-style-type: none"> • Poucos; predomina a venda direta 	<ul style="list-style-type: none"> • Indireto; múltiplos canais
<i>Promoção</i>	<ul style="list-style-type: none"> • Ênfase na venda pessoal 	<ul style="list-style-type: none"> • Ênfase na propaganda
<i>Preço</i>	<ul style="list-style-type: none"> • Processo complexo de compra • Licitações • Lista de preços nos itens padronizados 	<ul style="list-style-type: none"> • Lista de preços

Fonte: Adaptado pelo autor a partir de (RODRIGUES FILHO & ROMAN, 2004).

Conforme o Quadro 2.2, o Mercado Empresarial apresenta uma estrutura geograficamente concentrada, com relativamente poucos compradores. O produto apresenta grande complexidade tecnológica e os serviços são muito importantes. O relacionamento com o comprador é mais racional e predomina a experiência técnica. Como canal, predomina a venda direta, apesar da Internet esta crescendo bastante para compras repetidas. Na promoção é dada ênfase à venda pessoal e o preço pode variar em função do comprador.

2.4.2 Características do Mercado Empresarial

O mercado empresarial ou organizacional é formado por todas as organizações que compram bens e serviços para produzir outros bens e serviços com o propósito de revendê-los ou alugá-los para obter lucros. Os principais setores que compõem o mercado empresarial são agricultura, exploração florestal e pesca; mineração; manufatura; construção; transporte; comunicação; serviços públicos; financeiro e segurador; distribuição e serviços (KOTLER & KELLER, 2006).

Segundo HUTT e SPEH (2002), o que diferencia o mercado empresarial do mercado de consumo é o tipo de cliente e a forma como o cliente usa o produto. No mercado empresarial os clientes são basicamente organizações (empresas, instituições e governo), sendo os produtos utilizados ao processo produtivo de outros bens e serviços. Para SIQUEIRA (2005), o mercado empresarial é enorme em termos de faturamento e de número de empresas. O segmento mais conhecido é o setor industrial, ou seja, as indústrias de transformação, mas inclui também grande parte da produção agrícola, produtos minerais e florestais.

O mercado industrial evoluiu bastante nos anos 1970 e 1980, com a expansão das multinacionais e transnacionais. O mercado industrial, diante da emergência de mercados mundiais, alargou o enfoque sobre os tradicionais mercados industriais para os mercados mais amplos de negócios, incluindo o governo, alta tecnologia e serviços. A partir da segunda metade dos anos 1980 e década de 1990, os mercados industriais mergulharam em uma fase de intensa competição internacional, o que refletiu nas estratégias das empresas. A literatura especializada passou a falar de mercado *business to business (B2B)*, mercado de negócios ou mercado empresarial.

O mercado empresarial possui características especiais, como apontam HUTT & SPEH (2002):

- Tipo de cliente – organizações (empresas, governo e instituições), sendo uma quantidade relativamente pequena de compradores, se comparada com outros mercados;
- Geograficamente concentrado – a estrutura do mercado B2B costuma ser geograficamente concentrada;
- Compras racionais – o comportamento do comprador demonstra um envolvimento mais racional, onde predomina a experiência técnica, busca por relações estáveis e reciprocidade.

Para SIQUEIRA (2005), as principais diferenças do mercado empresarial são:

- Homogeneidade dos produtos – há tendência a um alto grau de similaridade, normalmente, seguindo especificações de entidades normativas, como Associação Brasileira de Normas Técnicas (ABNT), entre outras; são muito importantes as características e especificações técnicas.
- Situação das compras – as compras empresariais costumam ter alto valor médio e número limitado de compradores.

- Concentração geográfica dos produtores e compradores – concentração regional em muitos dos principais ramos empresariais.

O processo de compra tende a ser mais demorado, Baseado no desempenho e envolvendo várias pessoas na decisão. (RODRIGUES FILHO & ROMAN, 2004) apontam algumas especificidades importantes do mercado empresarial:

- Demanda derivada, muito flutuante e inelástica ao preço – Derivada, pois é gerada por organizações que utilizarão os produtos para atender seus consumidores finais; muito flutuante, ao reagir fortemente às oscilações da demanda final; inelástica ao preço, porque o produto fornecido impacta pouco o preço final do cliente ou em razão do cliente não poder mudar seu processo produtivo.
- Poucos clientes, grandes e concentrados geograficamente – Decisão de compra compartilhada, pelos vários integrantes do Centro de Compras, profissionais tecnicamente orientados, decidindo a compra com critérios objetivos; compra direta do fabricante e relacionamento estreito com o fornecedor.
- Produto definido pelo cliente importante – O cliente sabe o que quer, limitando as manobras do fornecedor; estrategicamente importante para o sistema produtivo do cliente, com um grande número de diferentes aplicações.

2.4.3 O Processo de Compra no Mercado Empresarial

O processo de compra no mercado empresarial é complexo, já que a decisão de compra envolve uma Unidade Tomadora de Decisão (UTD), formada por diversas pessoas da organização, sujeitas a múltiplas influências. A decisão de compra é eminentemente racional, com base em um considerável conjunto de fatores técnicos e econômicos, variando para cada empresa em função do ambiente em que opera e dos custos envolvidos, freqüentemente elevados.

(WEBSTER et al., 1975) denominam a unidade tomadora de decisão centro de compras, composto por "todas as pessoas e grupos que participam do processo decisório de compra e que compartilham algumas metas e riscos provenientes das decisões tomadas". Para os autores, o centro de compras inclui os membros da empresa que exercem um dos papéis:

- Iniciadores – aqueles que solicitam a compra de um produto; podem ser os usuários do produto ou outros membros da empresa;

- Usuários – aqueles que utilizarão o produto ou serviço. Em muitos casos, os usuários iniciam a proposta de compra e ajudam a definir as exigências que devem ser atendidas pelo produto;
- Influenciadores – pessoas que influenciam a decisão de compra, ao fornecer informações sobre a avaliação de alternativas ou estabelecer as especificações de compras.
- Decisores – pessoas que realmente decidem, não importando se têm ou não autoridade formal para fazê-lo.
- Compradores – pessoas com autoridade formal para selecionar o fornecedor e estabelecer os termos da compra. Essa autoridade na prática fica diluída entre os outros membros do processo, ficando os compradores com a responsabilidade de seleção de vendedores e negociação.

Uma pessoa pode assumir todos os papéis em uma situação de compra ou muitas pessoas podem assumir diferentes papéis. Para fazer uma venda no mercado empresarial a empresa precisa conhecer bem as pessoas que compõem a equipe de compras, suas motivações, percepções e poder. As empresas precisam identificar: Quem são os participantes mais importantes do processo decisório? Quais são as decisões influenciadas por eles? Qual a intensidade dessa influência? Quais os critérios de avaliação utilizados por eles? (WEBSTER et al., 1975) definem um modelo seqüencial de atividades 2. em oito estágios percorridos no processo de compra empresarial, conforme mostra o Quadro 2.3.

Quadro 2.3 - Estágios percorridos no processo de compra empresarial.

<i>Estágios da compra</i>	
1	Reconhecimento do problema
2	Definição das características e quantidades necessárias
3	Desenvolvimento das especificações para estabelecer os procedimentos de compra
4	Pesquisa para qualificar potenciais fornecedores
5	Recebimento e análise de propostas
6	Avaliação das propostas e seleção dos fornecedores
7	Seleção e estabelecimento de uma rotina de compra
8	Revisão do desempenho

Fonte: Adaptado pelo autor a partir de (WEBSTER et al., 1975).

O reconhecimento do problema ou de uma necessidade em potencial dá início ao processo de compras. Durante o processo de compras, são definidas as especificações, pesquisados fornecedores potenciais, recebidas e avaliadas propostas, até se tomar uma decisão e

selecionar um fornecedor. Após essa etapa, pode-se decidir manter o fornecedor em uma rotina de compra, sendo seu desempenho revisto periodicamente. SHAPIRO (2001) simplifica o processo de compra em três etapas:

- Negação;
- Reconhecimento do problema e
- Escolha da solução.

Na fase de negação, os membros da equipe de compras simplesmente torcem para que o problema desapareça, preferindo não enfrentá-lo. Essa fase dura por um período indeterminado e algumas empresas deixam para seus concorrentes os clientes que estão nessa fase. Na fase de reconhecimento do problema, o cliente reconhece o problema e deseja que alguém o ajude a esclarecê-lo. É o momento certo para vender um "conceito". Embora possa demorar um tempo até o fechamento do negócio, é a fase na qual é desenvolvida a especificação para a solução do problema e aquele que consegue fazê-lo geralmente ganha a venda. A última fase, que leva algumas semanas, é aquela na qual a equipe de compras escolhe formalmente o fornecedor que lhe oferece a melhor solução.

Segundo HUTT & SPEH (2002), os estágios percorridos no processo de compra variam em função das situações de compra, que podem ser três: compra nova, recompra modificada e recompra direta. Na compra nova, o problema ou necessidade é percebido como sendo diferente das experiências anteriores, sendo um processo mais longo e com a participação de diversos membros da empresa. Nas compras novas, SHAPIRO (2001) ressalta que a possibilidade de surgimento de problemas financeiros e de desempenho em função de uma compra é amplificada pela percepção que o comprador tem do risco a que sua carreira está exposta. Assim, a reafirmação e a redução do risco são muito importantes na venda comercial.

A recompra direta ocorre, quando há uma requisição contínua ou recorrente e a avaliação de novas alternativas é desnecessária. Nesses casos, a equipe de compras estabelece critérios de decisão e trata a compra como uma rotina. A Internet está sendo muito utilizada para dinamizar o processo de compras em caso de aquisições de rotina ou recompra direta. Na recompra modificada, os tomadores de decisão acham que benefícios significativos podem surgir de uma reavaliação de alternativas. A complexidade é menor que uma compra nova, mas envolve a revisão das alternativas de fornecedores e propostas. Na compra empresarial, algumas variáveis são mais relevantes. TELLES (2003) destaca a importância do desempenho, custo, assistência técnica, condições de entrega, condições de pagamento, Bases contratuais e atributos associados, conforme mostra o Quadro 2.4.

Quadro 2.4 - Variáveis utilizadas na análise na compra empresarial.

<i>Variáveis de Análise</i>	<i>Palavras-chave</i>	<i>Argumentação</i>
Desempenho	Eficiência	Produtividade; velocidade; otimização
Custo	Relação custo-benefício; controle	Custos inferiores; retorno sobre o investimento; análise comparada
Assistência técnica	Responsabilidade; acessibilidade	Disponibilidade; compromisso
Condições de entrega	Confiabilidade; compromisso; parceria	Capacitação; histórico de mercado
Condições de pagamento	Viabilidade financeira; relação de parceria	Grau de exposição; fluxo de caixa
Bases contratuais	Credibilidade; relação "ganha-ganha"	Responsabilidade e riscos compartilhados
Atributos associados	Diferenciação	Valor agregado; análise comparada

Fonte: Adaptado pelo autor a partir de TELLES, 2003.

O comprador empresarial busca otimizar ao máximo essas variáveis. Maior desempenho representa maior eficiência do produto e é medido pela velocidade, produtividade e otimização dos recursos. O custo está atrelado à relação custo-benefício e é comparado com outras ofertas, buscando diminuir o custo. A assistência técnica reflete a responsabilidade e acessibilidade do fornecedor, seu compromisso e disponibilidade. As condições de entrega e de pagamento influem diretamente na escolha, em razão dos prazos e recursos disponíveis. As bases contratuais são as estruturas de formalização das condições comerciais, informando responsabilidades e obrigações. Os atributos associados representam valor agregado para o cliente e fonte de diferenciação. A identificação das situações de compra é uma das bases de segmentação em nível macro, que permite à empresa vendedora redefinir seus esforços comerciais em função da situação do comprador.

2.4.4 Segmentação no Mercado Empresarial

A segmentação de mercado pode ser entendida como a subdivisão do mercado global da empresa em parcelas menores, com potencial suficientemente elevado para justificar uma adaptação da política de mercado a cada parcela.

Para TOLEDO & SIQUEIRA (2001), a segmentação no mercado empresarial é instrumento de apoio para definir o escopo tanto da estratégia de negócios como da estratégia de mercado. No primeiro caso, o objetivo é identificar quantos são os segmentos de mercado e em quais deles a empresa poderá competir com vantagem; no segundo, delinea-se o perfil dos segmentos, escolhem-se os segmentos que são alvos da estratégia de marketing (cobertura de

mercado) e busca-se um modo de posicionar-se para competir, de forma única e diferenciada, em cada segmento escolhido. Pela segmentação, segundo (RODRIGUES FILHO & ROMAN, 2004), a empresa visa a atingir alguns objetivos:

- Penetrar a fundo nos segmentos escolhidos e neles aumentar a sua participação;
- Concentrar seus esforços de marketing para economizar recursos financeiros, humanos e técnicos;
- Simplificar os processos de distribuição, venda pessoal, propaganda e promoção (campanhas dirigidas aos segmentos) e controle (menor volume de dados a processar);
- Fortalecer a imagem da empresa e de suas linhas de produto, graças à adaptação, aumentando a lucratividade por linha de produto, etc.

Para HUTT & SPEH (2002), o mercado empresarial pode ser segmentado em várias bases, classificadas de maneira ampla em duas categorias principais: macros segmentação e micros segmentação. A primeira categoria envolve a formação de macros segmentos baseados fundamentalmente nas características da organização compradora, aplicação do produto ou serviço e nas características da situação de compra, como mostra o Quadro 2.5.

Quadro 2.5 - As bases de segmentação em nível macro.

<i>Bases de segmentação em nível macro</i>	<i>Exemplos de classificação</i>
<i>Características da organização compradora</i> Potencial de compra Classificação dos setores industriais Tamanho do comprador Localização geográfica Taxa de utilização do produto Estrutura de compras	Pequeno, médio, grande Classificação da empresa compradora Pequeno, médio, grande, com Base no volume de vendas ou número de empregados País, Estado, cidade, bairro Usa muito, moderadamente, pouco, não usa Centralizada, descentralizada
<i>Aplicação do produto/serviço</i> Mercado final atendido Valor do uso	Varia de acordo com o produto ou serviço Alto, médio, baixo
<i>Características da situação de compras</i> Tipo da situação de compras Estágio do processo de decisão de compra	Compra nova, recompra modificada, recompra direta Estágios iniciais (especificação), estágios finais (decisão) direta finais

Fonte: Adaptado pelo autor a partir de HUTT & SPEH, 2002.

De forma macro as organizações compradoras podem ser segmentadas em função do seu potencial de compra, ramo de atividade, tamanho, localização geográfica, taxa de utilização do

produto e estrutura de compras. A aplicação do produto representa o mercado que é atendido com aquele produto e seu valor para o comprador. As características da situação de compra indicam se a compra é nova ou recompra, o estágio do processo de decisão, se está em fase final ou ainda em negociação.

A segunda categoria baseia-se na divisão dos macros segmentos em segmentos menores ou micros segmentos, similaridades e diferenças entre as unidades tomadoras de decisão. Para isolar os micros segmentos a empresa precisa coletar informações, que podem ser solicitadas à força de vendas ou conduzindo pesquisas para coleta de dados primários. As bases de segmentação micro estão no Quadro 2.6.

Quadro 2.6 - As base de segmentação em nível micro.

<i>Bases de segmentação em nível micro</i>	<i>Exemplos de classificação</i>
<ul style="list-style-type: none"> • Benefício chave • Estratégia de compra • Estrutura da unidade decisória • Importância da compra • Atitude em relação aos vendedores • Inovação organizacional 	<ul style="list-style-type: none"> • Qualidade, pronta entrega, preço • Otimização, satisfeito • Principais participantes da decisão • Muito, médio, pouco importante • Favorável, desfavorável • Inovador, seguidor
<ul style="list-style-type: none"> • Características pessoais • Características demográficas • Estilo de decisão • Risco • Confiança • Responsabilidade de trabalho 	<ul style="list-style-type: none"> • Idade, nível educacional • Normativo, conservador, modelo misto • Corre, evita risco • Alto, médio, baixo • Compra, produção, engenharia

Fonte: Adaptado pelo autor a partir de HUTT & SPEH, 2002.

Para separar os micros segmentos é importante verificar os benefícios chave para a organização compradora, que podem ser: preço, qualidade, pronta entrega, assistência e continuidade do fornecimento. Para CARDOZO (1980), a estratégia de compra pode variar entre satisfeitos e muitos satisfeitos. Os satisfeitos, ao receberem um pedido de compras, entram em contato com os fornecedores conhecidos e colocam o pedido com o primeiro fornecedor que atender aos requisitos do produto; os muitos satisfeitos consideram diversos fornecedores, conhecidos ou não, abrem concorrência e examinam cuidadosamente todas as propostas, antes de escolher um fornecedor. A estrutura da unidade tomadora de decisão também é uma forma de dividir o mercado em segmentos menores, a importância da compra, a atitude em relação aos vendedores, se a organização que faz o pedido é inovadora ou seguidora, além das características pessoais dos

tomadores de decisão: demográficas (idade, nível de educação, personalidade), estilo de decisão, tolerância a risco, confiança e área funcional.

WIND & CARDOZO (1974), citados por HUTT & SPEH (2002), propõem um modelo que percorre seis questões relacionadas aos macro e micros segmentos que devem ser respondidas pela empresa que estiver segmentando:

- Identificar os macros segmentos homogêneos com base nas características da empresa (ex.: tamanho, localização, taxa de uso) e situação de compra (estrutura de compra, tipo);
- Selecionar um conjunto de macros segmentos "aceitáveis" como oportunidades de mercado com Base nos objetivos e recursos corporativos;
- Avaliar cada macro segmento selecionado, se responde ou não ao programa de marketing da empresa. Em caso positivo, selecionar os macros segmentos desejados com base nos custos e benefícios associados àquele segmento.
- Em caso negativo, identificar dentro de cada macro segmento os micros segmentos importantes (com respostas homogêneas), com base nas características da unidade tomadora de decisão do tipo: importância da compra, inovação, características pessoais;
- Selecionar os micros segmentos alvo com base nos custos e benefícios associados à facilidade de atingi-los;
- Identificar o perfil completo de todos os micros segmento com base nas características da organização e da unidade tomadora de decisão.

BONOMA & SHAPIRO (1991) propõem um modelo de segmentação composto por cinco níveis, conforme mostra a Figura 2.18.

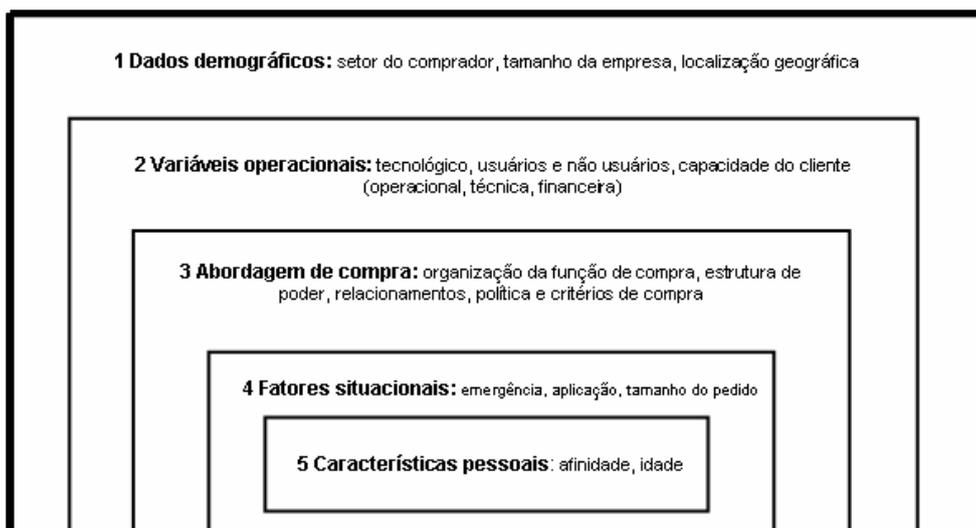


Figura 2.18 - Segmentação por níveis.

Fonte: Adaptada pelo autor a partir de BONOMA & SHAPIRO, 1991.

As bases de segmentação, propostas na Figura 2.18, começam nos níveis externos com características gerais mais facilmente observáveis: 1 Dados demográficos; 2 Variáveis Operacionais; 3 Abordagem de compra, indo até bases específicas no nível intermediário; 4 Fatores Situacionais, até o nível interno e 5 Características pessoais, sendo esses dois mais difíceis de avaliar.

No nível externo, os dados demográficos concentram informações relacionadas às necessidades e padrões de uso do cliente, incluindo dados do setor, tamanho da empresa e localização geográfica. As variáveis operacionais incluem a tecnologia relacionada ao pedido, se a empresa já é usuária do produto ou da marca oferecida, capacidade operacional, técnica e financeira. (RODRIGUES FILHO & ROMAN, 2004) propõem a segmentação dos clientes em função da sua capacidade técnica, sendo definido como tecnicista o cliente que possui grande capacidade técnica, concentrando as discussões nas características do produto ou serviço, e não tecnicista o cliente que não possui grande capacidade técnica, concentrando as discussões na solução do seu problema.

A abordagem de compra, segundo BONOMA & SHAPIRO (1991), é uma das bases de segmentação mais importantes, apesar de negligenciada, e envolve a filosofia de compras da empresa, incluindo as variáveis: organização formal da função compras; estrutura de poder; natureza dos relacionamentos organizacionais existentes; políticas gerais de compras e critérios de compras. Os dados necessários devem ser coletados diretamente com os funcionários da empresa-cliente em perspectiva. (RODRIGUES FILHO & ROMAN, 2004) propõem a segmentação dos clientes em função da sua organização. O cliente organizado possui na sua estrutura de compras todas as informações sobre seu relacionamento com os fornecedores e atua com base nessas

informações. O cliente não organizado carece, na estrutura de compras, de informações sobre o seu relacionamento com os fornecedores e as compras são feitas sem uma base sólida.

Os fatores situacionais estão associados a situações individuais de compra ou pedidos isolados. Entre esses fatores, BONOMA & SHAPIRO (1991) destacam a urgência, a aplicação específica, o tamanho do pedido e o tamanho de uma determinada parte do pedido. Os fatores situacionais podem ter grande influência sobre o nível anterior, a abordagem de compra. Por exemplo, a urgência de uma peça em uma linha de produção pode fazer com que a variável preço tenha menor importância.

Com relação às características pessoais, para BONOMA & SHAPIRO (1991), "as decisões de compras são tomadas por pessoas e não por empresas. (...) entre as variáveis nesse nível, considerado interno, estão a similaridade comprador-vendedor, a motivação do comprador, as percepções individuais e as estratégias de administração de riscos. Compradores avessos a riscos tendem a não mostrar boa receptividade a novos produtos e conceitos, além de evitar fornecedores não testados. Todas essas variáveis representam maneiras significativas para segmentar um mercado".

Os modelos de WIND & CARDOZO (1974) e BONOMA & SHAPIRO (1991) possuem abordagem complexa, com base no comportamento do comprador industrial. TOLEDO & SIQUEIRA (2001), comparando esses dois modelos com um caso prático, concluem que sua adoção não é simples, sendo mais eficaz a segmentação de mercado do que a abordagem de segmentação por tipo de cliente.

2.4.5 Segmentação por Tipo de Cliente no Mercado Empresarial

A segmentação por tipo de cliente no mercado empresarial aparenta ser mais efetiva do que a segmentação de forma macro. A maior parte das empresas faz a separação dos seus clientes no B2B em função do percentual de vendas em valores monetários, montando a Curva ABC de clientes, na qual classifica como "A" os clientes com maiores volumes de vendas, "B" os intermediários e "C" clientes com menor valor de vendas e menor representatividade (COBRA, 2000).

Para aprimorar essa forma simplista de segmentar os clientes, HUGLES (1997) sugere o modelo RFV (Recência, Frequência e Valor), incluindo duas variáveis, além do valor (V) gasto pelo cliente: Recência (R), que reflete a data da última compra, e Frequência (F), que reflete com qual frequência as compras são efetuadas. O resultado é a classificação dos clientes em células, com a combinação das 3 variáveis. Por exemplo, uma empresa pode obter 27 células com seus clientes,

sendo os clientes 333 os que possuem maior recência, frequência e valor e os clientes 111, os menos importantes. Para o autor, a variável que tem maior influência em uma nova compra por parte do cliente é a Recência, pois acredita que a lembrança da marca fará com que o cliente esteja mais disposto a comprar novamente.

PEPPERS & ROGERS (1997) propõem a diferenciação dos clientes em grupos em função do seu valor potencial, separando os clientes de maior valor (CMV), os clientes de maior potencial (CMP) e os clientes que dão prejuízo, denominados *bellow zero* (BZ), conforme abaixo:

- *Clientes de Maior Valor (CMV)* são aqueles clientes com o Valor Real mais alto para a empresa - os que fazem a maior parte dos negócios, geram as mais altas margens, são mais predispostos a cooperar, tendem a ser os mais fiéis e nas quais a empresa tem a mais alta participação (*share of customer*). O objetivo de uma empresa para os CMVs é a retenção, ou seja, conservar esses clientes;
- *Clientes de Maior Potencial (CMP)* são aqueles clientes nos quais o Valor Potencial de gasto com a empresa excede muito o Valor Real, têm o maior potencial de crescimento, o qual pode ser alcançado por meio de vendas cruzadas, manutenção do cliente por um longo período de tempo ou talvez pela alteração de seu comportamento, fazendo-o operar de uma maneira que custe menos para a empresa. O objetivo da empresa é desenvolver seus CMPs;
- *Clientes Bellow Zero (BZ)*, ou abaixo de zero, são aqueles cujo custo operacional é maior que seu valor real e potencial, ou seja, clientes que dão prejuízo à empresa. Exemplos: um BZ pode ser uma empresa que aceita muitos serviços gratuitos, mas não gera receita suficiente para pagar esses serviços. Nesse caso, os autores questionam se a empresa deveria "livrar-se" deles.

SHAPIRO (2001) considera que no mercado empresarial as empresas devem segmentar seus clientes, buscando clientes estratégicos, ou contas estratégicas, clientes que mostram um potencial comercial de longo prazo. O autor considera que a qualidade dos clientes é que faz o futuro da empresa, não a quantidade:

a) Experiência de Compras Anteriores: Clientes Inexperientes ou Experientes

Uma forma interessante de classificação dos clientes é a partir de sua experiência anterior na compra daquele produto ou serviço. SHAPIRO, 2001 classifica os clientes do setor das empresas de plásticos de engenharia em dois grupos: clientes inexperientes e clientes experientes, analisando várias dimensões, conforme o Quadro 2.7.

Quadro 2.7 - Efeito da experiência do cliente no setor de plásticos de engenharia.

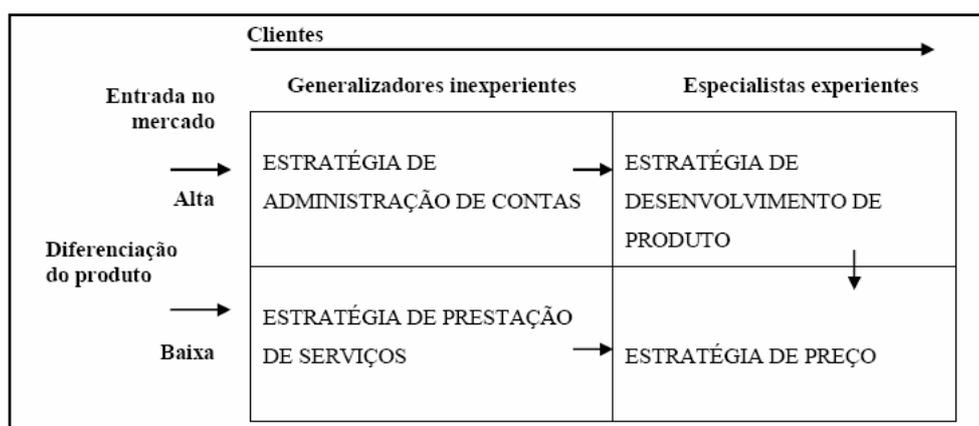
<i>Grupos de Clientes</i>	<i>Clientes Inexperientes</i>	<i>Clientes Experientes</i>
Unidade tomadora de decisão	Engenheiros de aplicações de produto	Compradores
Processo de tomada de decisão	Novas tarefas, dois anos	Recompra rotineira, quatro a cinco compras por ano
Benefícios do produto dominante	Assistência técnica, aplicações, serviço	Desempenho, disponibilidade, preço
Considerações	Posição competitiva vantajosa	Baixo custo
Programas de vendas	Administração da conta realizada por especialistas do setor	Vendas no campo com Base em áreas geográficas
Fatores críticos de sucesso	Administração de conta e tecnologia	Baixo custo dos produtos vendidos, preços baixos ou paridade de preços

Fonte: Adaptado pelo autor a partir de SHAPIRO, 2001.

Como mostra o Quadro 2.7, à medida que os clientes se tornam mais experientes, as decisões de compras deixam de ser tomadas por engenheiros de aplicações do produto e passam a ser tomadas diretamente pelos compradores, diminuindo o tempo da tomada de decisão. Conseqüentemente, os programas de vendas são mais especializados no início e mais pulverizados para clientes experientes. Os fatores críticos de sucesso, no caso de clientes inexperientes, estão associados à administração da conta e tecnologia e, no caso de clientes experientes, ao baixo custo dos produtos vendidos e preços baixos.

O autor sugere que as estratégias de marketing no mercado empresarial sejam moldadas, considerando a experiência do cliente e a diferenciação do produto, conforme o Quadro 2.8.

Quadro 2.8 - Estratégias de marketing empresarial x experiência do cliente.



Fonte: Adaptado pelo autor a partir de SHAPIRO, 2001.

O Quadro 2.8 mostra que, para produtos com alta diferenciação oferecidos para clientes inexperientes (quadrante superior esquerdo), a empresa deve enfatizar a administração de contas por experientes administradores. Segundo o autor, a IBM utilizou essa estratégia na indústria de

computadores por vários anos, incluindo gerentes de alto nível e profissionais em processamento de dados numa mesma unidade tomadora de decisão, assegurando a continuidade do envolvimento de generalizadores inexperientes no processo decisório do cliente.

Para produtos com alta diferenciação oferecidos para clientes experientes (quadrante superior direito), a empresa deve tentar manter o cliente com uma estratégia de constante melhoria no seu produto. Nesse caso, os fornecedores devem enfatizar um determinado benefício - por exemplo, suporte técnico para novas aplicações - que represente valor para o cliente. A força dessa estratégia está em manter a diferenciação do produto aos olhos do comprador, desviando sua atenção da atração causada pelos preços dos concorrentes.

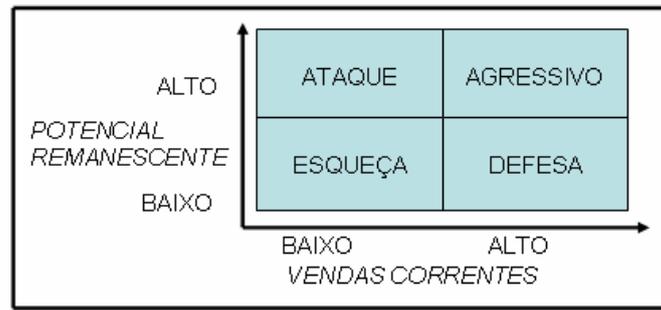
Para produtos com baixa diferenciação e clientes inexperientes (quadrante inferior esquerdo), a estratégia sugerida pelo autor para atrair clientes inexperientes que entram tardiamente no mercado é concentrar-se em prestação de serviços e conscientizar-se de que os clientes percebem os produtos como similares. Como exemplo, a AT&T apresenta-se a esse grupo de clientes com o slogan: "Nossos produtos são normalmente fornecidos junto com consultoria técnica" (SHAPIRO et al., 1994). Essa estratégia está direcionada para um grande número de empresas inexperientes em termos de grandes sistemas de telecomunicações, que desejam suporte de serviços para fazer uma seleção entre os produtos concorrentes.

Para produtos com baixa diferenciação e clientes experientes (quadrante inferior direito), o cliente e as forças da concorrência se somam e acabam conduzindo um determinado mercado a uma condição de extrema sensibilidade a preço. A estratégia sugerida pelo autor, baseada em preços, envolve determinar preços corretos, após garantir qualidade aceitável e disponibilidade do produto.

b) Potencial Remanescente do Cliente: Alto ou Baixo

Para SHAPIRO (2001), a empresa precisa equilibrar os esforços para desenvolver e manter os clientes ativos e ganhar novos clientes. O Quadro 2.9 apresenta como classificar os principais clientes ativos de uma empresa com base nas vendas correntes e no potencial remanescente.

Quadro 2.9 - Segmentação da base de clientes x potencial e vendas correntes.



Fonte: Adaptado pelo autor a partir de SHAPIRO, 2001.

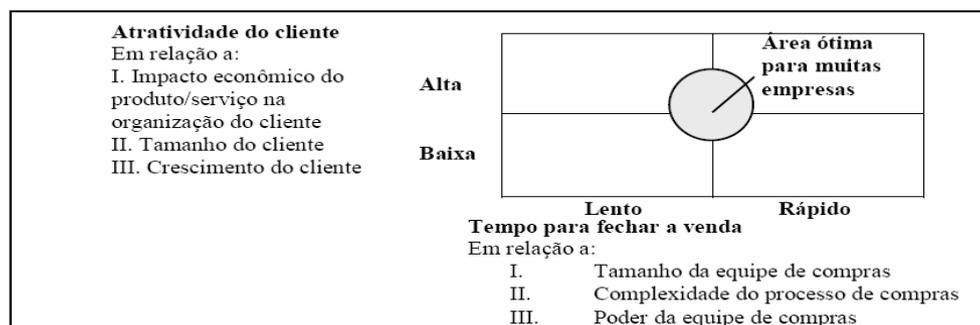
Para o autor, os clientes ativos com altos níveis atuais de compras e elevado potencial remanescente (quadrante superior direito) devem ser a prioridade máxima da empresa e atendidos por profissionais de vendas eficientes em desenvolvimento de clientes, que garantam bom volume de negócios com uma postura mais agressiva. Os clientes ativos do quadrante inferior direito precisam ser atendidos por profissionais que enfatizem o atendimento e garantam os negócios já fechados, com uma postura de defesa. No quadrante superior esquerdo, compras baixas ou inexistentes e alto potencial exigem um profissional de vendas mais voltado para o ataque e clientes com baixo potencial e baixo volume de vendas corrente devem ser praticamente esquecidos.

Segundo SHAPIRO (2001), essa mudança na *cesta* de clientes ativos exige uma transformação da equipe de vendas, que precisa ser segmentada além da tradicional dimensão geográfica ou especializada em produto/serviço, para uma equipe de vendas segmentada em função do potencial do cliente.

c) *Tempo para Fechar a Venda: Lento ou Rápido*

Quando a unidade responsável pelas decisões é grande, com muitas pessoas envolvidas ativamente na decisão de compra, o tempo para fechar a venda geralmente é longo. O tempo para fechar é o menor possível nas empresas com processo simples de tomada de decisão, um pequeno grupo para tomar as decisões, na qual a mesma pessoa controla o problema que o fornecedor está procurando eliminar e o orçamento para resolvê-lo. Segundo SHAPIRO, 2001, é útil combinar as variáveis: atratividade do cliente com tempo para fechar vendas, como mostra o Quadro 2.10.

Quadro 2.10 - Atratividade do cliente x tempo de fechar a venda.



Fonte: Adaptado pelo autor a partir de SHAPIRO, 2001.

Os critérios para determinar a atratividade dos clientes são os mesmos para analisar a atratividade de um segmento, incluindo impacto econômico do produto ou serviço para a organização do cliente, tamanho e taxa de crescimento do cliente. Além desses, o autor inclui o "valor testemunhal", ou seja, a capacidade de recomendar o produto, como importante de ser avaliado em um cliente potencial. Os clientes são classificados como de alta e baixa atratividade (SHAPIRO, 2001). O quadrante superior direito corresponde à situação ideal, que geralmente fica vazia. Muitos clientes grandes e atraentes aparecem no quadrante superior esquerdo, pois seu tamanho reflete complexidade na tomada de decisão. Assim, no caso de uma empresa em início de operação, o quadrante inferior direito, com clientes menos atraentes, mas que fecham uma venda mais rapidamente, pode ser uma boa aposta.

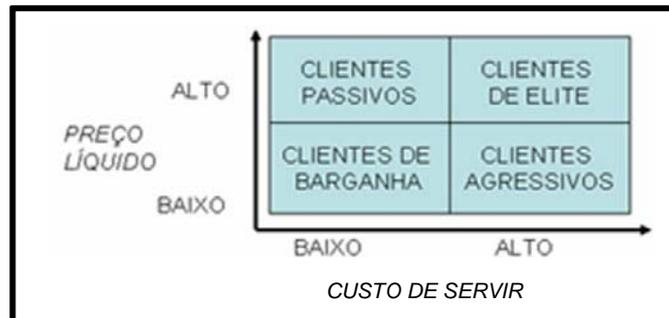
Para SHAPIRO (2001), a capacidade de fechar vendas rapidamente é, em geral, resultado de uma equipe de compras pequena, que tem poder para tomar decisões e se guia por um processo simples e claro. Os formadores de opinião são os grandes "clientes-vitrine", com equipes de compras numerosas, processos complexos e poder de decisão difuso.

d) Lucratividade do Cliente Comparada com o Custo de Atendimento

Os autores consideram a distribuição dos clientes em duas dimensões: preço líquido recebido e custo de servir. Sua premissa é que alguns pedidos possuem custo de fornecimento maior do que outros e que as variações de custo e preço provocam profundas diferenças na lucratividade individual de clientes e pedidos. Lucro, para os autores, é a diferença entre o preço líquido e custo total. O custo total é composto por quatro componentes: custo de pré-venda, de produção, de distribuição e de serviços pós-venda. Muitas empresas trabalham com variações de preço por cliente, o que nem sempre reflete o custo total do produto, havendo alguns clientes que geram prejuízos e grande dispersão da lucratividade. O Quadro 2.11 apresenta a dinâmica de inter-

relacionamento entre vendedor e comprador. O eixo vertical corresponde ao preço líquido, variando entre baixo e alto, e o horizontal, o custo de servir, variando entre baixo e alto.

Quadro 2.11 - Matriz de dispersão dos clientes x preço líquido e custo de servir.



Fonte: Adaptado pelo autor a partir de SHAPIRO, 1987.

No Quadro 2.11, os clientes de elite têm alto custo de servir, porém, estão dispostos a pagar preços altos (quadrante superior direito). No extremo oposto, encontram-se os clientes barganhistas que são sensíveis a preço e relativamente insensíveis a serviços e qualidade, sendo atendidos com baixo custo de servir e preço baixo (quadrante inferior esquerdo).

Os clientes passivos estão dispostos a aceitar preços altos e são menos dispendiosos no custo de servir (quadrante superior esquerdo). Como exemplo, alguns componentes importantes de aviões não podem ser modificados sem que a aeronave seja submetida a testes, permitindo aos fornecedores um preço alto com baixo custo de servir. Os clientes agressivos exigem produtos de mais alta qualidade, melhor serviço e preços baixos (quadrante inferior direito). Compradores agressivos são geralmente poderosos; pelo fato de realizarem compras em grandes volumes, possuem poder de barganha com os fornecedores para negociar preços e obter mais serviços.

A falta de uma contabilidade que calcule os custos por cliente dá respaldo ao mito da correlação entre alto preço e alto custo. Os deslocamentos em direção aos quadrantes dos clientes barganhistas e dos clientes agressivos fazem parte da tendência generalizada que os produtos têm de evoluir da condição de especialidades, com alta margem de lucro, para a condição de mercadorias com pequena margem de lucro. O principal uso da matriz de dispersão dos clientes é o estabelecimento de prioridades para a área Comercial.

Na prática, o erro mais comum na seleção das contas, cometido pelas empresas, é a falta de decisão. A gerência pode achar difícil deixar de lado um negócio que não se encaixe na opção estratégica da empresa, mas deve avaliar aquilo que é ganho (lucros) e que futuras oportunidades poderão ser perdidas (custos), ao tomar a decisão de atender um cliente. Algumas vezes, a melhor resposta é "não" (BONOMA & SHAPIRO, 1991). Outra dificuldade da matriz proposta por SHAPIRO et al (1994) está em estabelecer bons sistemas para o registro de custos, preços e lucros

por cliente. A partir do momento em que essas informações estão disponíveis, a análise não é difícil de ser efetuada.

2.5 Relação entre Produtividade de Vendas e Medição de Desempenho

2.5.1 – Gestão de Desempenho

A gestão do desempenho deve proporcionar uma ligação sistemática entre a estratégia organizacional, os recursos e os processos, de uma forma estruturada, para a obtenção de melhorias contínuas, de modo que todos os envolvidos entendam *onde se está e aonde se quer chegar* para atender os interesses dos *stakeholders*. Para tanto, faz-se necessário definir a visão do estágio desejado para o futuro, e tal estágio se baseia não apenas no potencial de mercado, mas também no conjunto de crenças e valores da organização que define a sua própria missão, e a partir da qual possibilita traçar os caminhos a serem percorridos, guiando-se por ações específicas, de aproveitamento de oportunidades e superação de problemas.

O processo de gestão visa administrar *como* essas ações estão sendo desenvolvidas e *se* caminham em direção aos resultados desejados. É esta a forma pela qual a administração comunica e institucionaliza sua estratégia, define medidas de sucesso, coordena os esforços dos processos-chave em busca de vantagens competitivas para a organização.

De acordo com SINK & TUTTLE (1989) o desempenho de uma organização é função de uma complexa interação entre sete critérios: eficiência, eficácia, qualidade, produtividade, qualidade de vida no trabalho, inovação e lucratividade. Para ATKINSON *et al.*, (1999) o sucesso de uma empresa depende do atendimento dos interesses dos *stakeholders*; ou seja: os clientes valorizam serviço, qualidade e custo; os funcionários, satisfação e segurança; enquanto os sócios e proprietários priorizam o retorno dos investimentos.

O desempenho organizacional depende das suas unidades de negócio; por isso o processo de gestão estratégica tem de levar em conta a necessidade de alinhamento entre as áreas operacionais e as estratégias e políticas da organização como um todo. Essas áreas precisam ser eficazes no que lhes compete fazer, para garantirem o cumprimento da missão da organização. As metas gerenciais são atingidas por meio de atividades operacionais, devendo ser planejadas e executadas de acordo com a direção e as metas definidas para a unidade de

negócios, processo, ou nível individual, o que permite ao gestor alinhar as atividades organizacionais e a alocação de recursos às metas estratégicas.

A chave para atingir a eficácia organizacional é administrar os processos operacionais essenciais, valendo-se de fatores-chave como velocidade, custo, flexibilidade e confiabilidade, os quais atendem as necessidades dos clientes e, portanto, a necessidades estratégicas da organização. CERTO & PETER (1993) sugerem a adoção de esquemas de controle que visem monitorar e avaliar o processo de administração estratégica, a fim de melhorá-lo e assegurar-lhe um funcionamento adequado, fazendo com que as estratégias se desenvolvam da forma como foram planejadas, o que ajuda a alta administração a atingir as metas organizacionais. A definição de um sistema de controle deve espelhar os objetivos da organização e as estratégias competitivas do negócio, de modo a definir medidas (a partir de fatores facilmente quantificáveis) nas quais as relações entre quantidades de um evento atual são comparadas às de um evento anterior, ou aos padrões.

A - Medição de Desempenho

Em busca da excelência, as organizações têm como desafios principais: executar seu trabalho com pontualidade, dentro das especificações, com certa quantidade de recursos, e melhorar continuamente o desempenho dos indivíduos e grupos, da organização e dos sistemas. Por isso, precisará desenvolver instrumentos de medição e avaliação, que alimentem sistematicamente o processo de tomada de decisões, definindo detalhadamente o processo de gestão e avaliando sua eficácia e eficiência, assim como a sua abrangência e consistência, de acordo com SINK & TUTTLE (1989).

A medição do desempenho tem como objetivo prover os gestores com informações adequadas, para que eles possam saber sobre *o que* se fez, *o modo* como se fez, *onde* pode ser melhorado, e, ainda, *quando* o sistema está sob ou fora de controle e, assim, poder motivar e encorajar os elementos da organização a caminharem unidos, em direção aos objetivos estratégicos. Portanto, um sistema de mensuração deve ser moldado de forma a captar e organizar dados, e comunicar resultados de forma clara e rápida, permitindo aos elementos da organização tomar as providências cabíveis, no sentido de melhorar o desempenho global.

No contexto competitivo, mensurar o desempenho é essencial porque, o conhecimento *do que* e *como foi feito*, propicia às pessoas saber como obter melhorias contínuas em condições de

lucratividade. No entanto, o uso eficaz de sistemas de mensuração pode ser dificultado em virtude de alguns paradigmas, segundo SINK & TUTTLE (1989), como o que considera a mensuração uma ameaça que deve ser pautada em precisão, enfoca um único indicador, dando ênfase à produtividade do trabalho, encarando a subjetividade como descuido e os padrões como fatores limitativos do desempenho.

Para esses autores, a quebra desses paradigmas é uma condição essencial para que se desenvolvam medidas e sistemas de mensuração de qualidade. Recomendam que se procure desenvolver medidas essenciais absolutas, partindo de uma visão do sistema de mensuração da organização, definindo operacionalmente os critérios de desempenho, definindo medidas específicas, criando uma lista priorizada e definindo técnicas de aplicação, de acordo com SINK & TUTTLE, (1989).

Para empresas de serviços o uso de medidas não financeiras pode prover uma dimensão adicional valiosa, por fornecer informações que permitem quantificar o posicionamento competitivo do negócio. Por esse motivo tem-se tentado utilizar uma mistura de medidas financeiras e não financeiras, cujos componentes qualidade e flexibilidade, intangíveis possam ser mensuradas juntamente com as medidas tradicionais, como lucratividade e retorno sobre o investimento.

B - Medição e a Gestão da Administração de Vendas

A medição do desempenho da empresa é fundamental para a gestão de vendas. A medição fornece aos gerentes as informações necessárias à tomada de decisões e ao desenvolvimento de ações de melhoria da qualidade e produtividade de vendas. O desenvolvimento de um bom sistema de monitoramento de desempenho é fundamental para o gerenciamento das atividades de administração de vendas. A mensuração de desempenho é uma das mais importantes ferramentas a serem utilizadas para verificar se os objetivos estabelecidos pela empresa estão sendo alcançados, auxiliando ainda na melhor aplicação dos recursos destinados à administração de vendas. A realização desse monitoramento torna-se ainda mais importante no atual cenário em que as atividades relacionadas à qualidade de vendas vêm sendo reconhecidas mundialmente como de elevada importância para geração de valor para o cliente

Os indicadores consistem em expressões quantitativas que representam uma informação gerada, a partir da medição e avaliação de uma estrutura de produção dos processos que a compõem e/ou dos produtos resultantes. A medição e a avaliação referem-se à identificação

dos dados e informações e ao estabelecimento de critérios, especificações ou valores para comparação entre resultados obtidos e padrões ou metas definidas. A medição é a parte inerente da gestão da qualidade, constituindo um sistema de apoio para o planejamento, solução de problemas, tomada de decisões, desenvolvimento de melhorias, controle de processos e motivação dos recursos humanos.

Um processo representa uma tarefa específica ou um conjunto de tarefas realizadas por um indivíduo, uma equipe, um departamento ou por toda a empresa, resultando em um produto ou serviço. Cada processo é um sistema gerencial completo com fornecedores, *inputs* (matérias-primas), *outputs* (bens e serviços) e clientes (internos e externos). A gerência, aqui sendo representada por os diretores, engenheiros, supervisores, encarregados, enfim, todas as pessoas responsáveis pelas decisões no processo, onde cada um desses intervenientes tem duas responsabilidades principais:

- Fazer o trabalho em tempo hábil, dentro das especificações de qualidade, com a quantidade correta de recursos;
- Melhorar, de forma contínua, o desempenho de cada indivíduo, da equipe e da empresa.

Para a execução dessas duas tarefas, são efetuadas intervenções no processo. A análise da possibilidade de uma intervenção é feita por meio de coleta, processamento e avaliação de dados. Esses elementos constituem a medição, estruturada como um subsistema de apoio à tomada de decisões no sistema gerencial.

C – Aspectos Operacionais e de Gerenciamento no Uso de Indicadores de Vendas

Qualquer que seja a ou as perspectivas de mensuração utilizadas, a simples seleção de indicadores e sua mensuração não são suficientes para que um sistema de medição seja eficaz na melhora do desempenho de administração de vendas de uma empresa. O suporte de um sistema de informação, o uso de um sistema dinâmico e a utilização efetiva dos resultados obtidos são fundamentais para o sucesso na medição do desempenho de administração de vendas.

C.1 -) Necessidade do Suporte de um Sistema de Informação

A tecnologia de informação teve um papel fundamental na revolução da medição de desempenho. Tornou a captura, análise e revisão dos dados mais fácil, simplificou o gerenciamento do sistema de medição de desempenho e a manutenção da informação. Embora

sistemas de mensuração possam ser implantados sem o suporte de um sistema de informação, é provável que caso isso ocorra, seu escopo e a habilidade dos gerentes de explorá-lo seja inferior a um que utilize um sistema de informação como suporte.

Um bom sistema de informação precisa dar suporte aos gerentes das áreas, ser capaz de transferir dados dos pontos de coleta para o nível apropriado de gerenciamento, fornecer informação precisa, útil, confiável e fácil de usar e levar em consideração a estratégia, e fatores ambientais relacionados ao negócio assim como a estrutura da organização, seus processos, funções e relacionamentos. Também é necessário que conecte as funções administração de vendas internamente umas com as outras e com outras funções como produção e marketing.

C.2 -) Necessidade de um Sistema Dinâmico de Mensuração de Desempenho

Bons indicadores não permanecem bons para sempre. Para que estratégia e medidas de desempenho estejam alinhadas, e o sistema de medição se mantenha integrado, eficiente e eficaz ao longo do tempo, é necessário que os indicadores empregados sejam revisados sistematicamente e que o sistema seja dinâmico. A habilidade do acompanhamento contínuo é um importante fator de sucesso.

C.3 -) O Uso Efetivo das Medidas e Ações Corretivas para os Desvios

O passo final para a implementação de um sistema de indicadores de desempenho é o monitoramento contínuo do *output*. Uma das chaves da melhoria contínua é o processo constante de monitoramento que identifica as variações e age corretivamente para eliminar as causas das variações e evitar que elas ocorram novamente.

D - Classificação das Medições

As medições podem ser classificadas em tipos diferentes, segundo a finalidade da informação que fornecem e os critérios de avaliação da informação (SINK&TUULE, *apud* PICCHI, 1997):

- Medições para visibilidade: são medições para diagnóstico inicial, realizadas num primeiro momento quando a empresa ainda não controla seus processos. Nesse caso, as medições podem anteceder a realização de intervenções na empresa. Têm por objetivo

identificar pontos fortes e fracos ou disfunções a partir das quais são priorizadas ações de melhoria. Sua finalidade principal é demonstrar o desempenho atual;

- Medições para controle: visam a controlar os processos a partir do momento em que a empresa consegue definir padrões de desempenho para os mesmos. A medição de desempenho passa, então, a ser utilizada na identificação de problemas. Existe um problema sempre que o indicador mostrar um desvio em relação a um padrão estabelecido. Após a detecção do problema, é preciso identificar suas causas e atuar de forma a eliminá-las. Nesse caso, a avaliação é feita comparando-se os resultados com padrões adotados ou convencionais, tais como as médias e os limites de controle superior e inferior;
- Medições para melhorias: quando as empresas decidem intervir no processo, devem ser estabelecidas metas por meio dos seus indicadores (por exemplo, utilizando o *benchmarks* como referência). Nesse caso, a medição permite verificar o impacto das ações de melhoria sobre o desempenho do processo. A avaliação é feita comparando-se o desempenho da variável medida em relação à meta estabelecida.

A utilização das medições também contribui efetivamente para a motivação e envolvimento das pessoas com o desenvolvimento de melhorias, pois permite ao indivíduo um retorno em relação ao desempenho do processo no qual está envolvido e ao seu próprio desempenho.

E - Geração de Indicadores de Desempenho

Para a geração de indicadores em uma empresa, inicialmente devem-se selecionar os processos em relação aos quais se quer desenvolver a medição. Essa seleção deve ser criteriosa, porquanto um número excessivo de medições pode resultar em custos muito elevados e desmotivar as pessoas envolvidas, tendo em vista a carga de trabalho adicional.

Nesse sentido, a definição de um conceito de administração de vendas com os objetivos da empresa e o estabelecimento de estratégias para melhoria da produtividade, é fundamental para a geração de indicadores. Para tanto, a empresa deve assegurar-se de que está medindo os elementos adequados, isto é, aqueles realmente importantes para a melhoria do seu desempenho.

Os processos a serem medidos devem ser aqueles onde se localizam as causas principais de problemas levantados e priorizados pela empresa (por meio de ferramentas da qualidade, por exemplo). Com base nessa priorização, são definidos os processos que irão

sofrer intervenções com base em programas de melhoria da produtividade, surgindo à necessidade de informações que apoiem as decisões a serem tomadas. Os indicadores podem ser classificados em dois tipos principais, apresentados a seguir:

- Indicadores gerenciais: são estabelecidos com o objetivo de acompanhar e impulsionar a implantação das estratégias. A partir do planejamento estratégico da empresa, gera-se um plano de ação, no qual tais estratégias são traduzidas em atividades concretas a serem conduzidas a vários níveis hierárquicos, nos diferentes departamentos e funções dentro da organização. O plano de ação da organização deve incluir ainda a definição de metas, que são os resultados a serem atingidos no futuro. As metas devem ser desdobradas em vários níveis dentro da empresa, bem como os indicadores a elas associados. Dessa forma, os indicadores gerenciais permeiam vários processos dentro da empresa;
- Indicadores operacionais: são estabelecidos em função dos objetivos e tarefas desenvolvidos dentro de cada processo, devendo ser coerentes com os objetivos e estratégias adotadas pela empresa.

Uma vez escolhidos os processos em relação aos quais são desenvolvidas as medições, deve-se fazer uma análise dos mesmos, elaborando-se fluxogramas que possibilitarão identificar produtos ou serviços executados, os clientes e fornecedores internos e externos do processo, bem como determinar funções e responsabilidades e identificar pontos críticos do processo. Os indicadores são estabelecidos sobre os resultados controláveis ou gerenciáveis do processo, isto é, aqueles sobre os quais as pessoas envolvidas no processo têm responsabilidades e podem atuar sobre suas causas, corrigindo desvios e melhorando resultados.

Os resultados de um processo podem ser afetados por várias causas. Pode-se achar necessário verificar algumas dessas causas como meio de garantir um bom resultado, estabelecendo-se, então, indicadores para o seu monitoramento. No TQC (*total quality control*), esses indicadores são denominados itens de controle e itens de verificação.

F - Requisitos dos Indicadores

Considerando-se a situação específica em que for aplicado, o indicador deve atender aos seguintes requisitos (MARTINS, 1998):

- Seletividade: os indicadores devem estar relacionados a aspectos, etapas e resultados essenciais ou críticos do produto, serviço ou processo;

- Simplicidade: devem ser de fácil compreensão e aplicação, principalmente para aquelas pessoas diretamente envolvidas com a coleta, processamento e avaliação dos dados, utilizando relações percentuais simples, médias, medidas de variabilidade e números absolutos;
- Baixo custo: devem ser gerados a custo baixo, de modo que o custo para coleta, processamento e avaliação não deve ser superior ao benefício propiciado pela medida;
- Acessibilidade: os dados para o cálculo do indicador devem ser de fácil acesso;
- Representatividade: os indicadores devem ser escolhidos ou formulados de forma que possam representar satisfatoriamente o processo ou o produto a que se referem;
- Estabilidade: devem perdurar ao longo do tempo, com base em procedimentos rotinizados, incorporados às atividades da empresa ou departamento;
- Rastreabilidade: os dados e informações utilizados devem ser adequadamente documentados, bem como formulários e memórias de cálculo, inclusive o registro do pessoal envolvido;
- Abordagem experimental: é recomendável desenvolver, inicialmente, os indicadores considerados necessários e testá-los. Caso não se mostrem realmente importantes ao longo do tempo, devem ser alterados.

G - Implantação de Medição

A estratégia ideal para aumentar a compreensão e aceitação da medição é envolver no seu desenvolvimento todos aqueles que são afetados ou irão utilizá-la. Para que a medição se torne realmente parte integrante do sistema gerencial, uma série de questões operacionais devem ser levantadas e respondidas para cada uma das três etapas que compõem a medição: a coleta, o processamento e avaliação de dados. Para se obter tais respostas, devem-se considerar quais as informações necessárias e quem são os seus usuários.

1 - Coleta: enfoca a geração dos dados necessários ao fornecimento da informação.

- Onde os dados são obtidos?
- Quem é a pessoa responsável pela coleta?
- De que forma os dados são obtidos?
- Como são armazenados e recuperados?
- Com que frequência os dados são coletados?

2 - Processamento: é por intermédio do processamento que os dados transformam-se em informações.

- Que procedimentos são utilizados para representar as informações?
- Que ferramentas, métodos e programas computacionais são usados para o processamento dos dados?
- Como as informações são armazenadas?
- Qual é o público-alvo das informações?
- Com que frequência são fornecidas as informações?

3 - Avaliação: nela discutem-se as possíveis causas dos resultados obtidos e apontam-se metas.

- Quais os critérios para avaliação?
- Que atitude deve ser tomada em caso de resultados indesejados?
- Como é realimentado o processo?

2.5.2 – Produtividade de Vendas

Uma análise das características do mercado *business-to-business*, em relação ao mercado de vendas direto ao consumidor, mostra que a gestão do relacionamento com o cliente assume uma dimensão muito maior quando aplicada ao mercado B2B. O mercado *business-to-business* é um mercado de tamanho menor em número de clientes e é composto por empresas de grande porte. O valor econômico das negociações é alto. Por isso, o valor de um único cliente ao longo de sua existência pode ser enorme. A perda de um único cliente pode ter um impacto negativo para o faturamento da empresa, o que faz com que maior ênfase seja dada na manutenção dos clientes existentes do que na aquisição de novos clientes. O interesse da empresa que atua nesse segmento é construir um relacionamento baseado na confiança e no interesse mútuos e ainda, entregar produtos e serviços de alto valor ao cliente.

No mercado industrial é quase obrigatória a ênfase da venda pessoal. Mesmo que sejam predominantes racionais as motivações que levam o encarregado de compras ou o departamento especializado a fazer aquisições, o contato pessoal por meio do profissional da venda é determinante. A negociação transforma-se, de fato, numa busca efetuada pelas duas partes com o objetivo de se chegar a uma conclusão satisfatória para ambas. Não se trata de induzir o comprador a adquirir determinado produto ou a comprar mais, mas de descobrir um campo de satisfação para duas necessidades não opostas, mas sim complementares. Dessa forma, as conversações não se desenvolvem apenas sobre o produto ou serviço em si, mas

também sobre os elementos colaterais (condições de pagamento, prazos e condições de entrega) que dão mais agilidade ao mecanismo de pretensões e concessões que constitui a verdadeira essência da negociação. Essa experiência mostra toda sua utilidade sempre que o profissional da venda precisa entrar de novo em contato com as várias pessoas envolvidas no processo de compras de uma empresa, para as quais a arte da negociação é um componente essencial de sua habilidade profissional.

A - O Processo de Vendas

A venda no mercado B2B pode demorar semanas ou meses. É preciso conhecer e respeitar o ritual de aprovação de novos clientes. O vendedor do mercado B2B é um profissional de quem se exige preparo especial, não só no que se refere ao conhecimento do produto ou serviço que oferece, mas também quanto à habilidade de organizar e de padronizar as técnicas de relação interpessoal em função de um processo de compra mais racional do que o habitualmente encontrado no mercado de consumo (B2C).

As empresas tendem a envolver nas negociações de compra de serviços de telecomunicações um número maior de pessoas (decisores e influenciadores), cada uma delas com uma competência própria no processo que levará à aquisição do serviço. Em linhas gerais, a negociação desenvolve-se em três fases:

- *Primeira fase:* É quando se estabelece a relação entre o vendedor e o comprador (esse com frequência, um simples porta-voz de um departamento especializado). Trata-se de uma etapa longa e complexa, finalizada com o conhecimento do produto ou serviço e de sua capacidade de satisfazer a necessidade que origina a compra. Nessa fase, solicitam-se do vendedor informações adicionais e ajuda para resolver eventuais problemas de telecomunicações. Costuma ser um período decisivo para criar estima e cooperação entre vendedores e cliente;
- *Segunda fase:* Nesta fase acontecem o desenvolvimento da negociação e a decisão de comprar ou não o produto ou serviço oferecidos. O vendedor, além de demonstrar a capacidade de o produto ou serviço satisfazer às necessidades do comprador, procura garantir que a entrega se faça dentro das condições, qualidade e do prazo combinados. Um atraso terá influência negativa sobre o processo de telecomunicações do comprador e poderá prejudicar no futuro suas relações com o vendedor e a empresa para a qual ele trabalha;

- *Terceira fase:* Concretizada a compra, o vendedor precisa estar atento a um fenômeno muito freqüente: o surgimento de um conflito entre a empresa vendedora e a organização compradora. A origem desse conflito está na mudança de fornecedor, o que geralmente ocasiona um período de turbulência. Resolver esse conflito significa não só preservar a venda, mas, principalmente, não perder o clima de confiança e perspectiva de uma relação de longo prazo com o cliente.

O processo de vendas é uma seqüência de passos por meio dos quais os vendedores realizam as vendas. Essa seqüência foi, originalmente, proposta por (DUBINSKY & HANSEN, 1981) e pode ser aplicada a diferentes setores em diferentes cenários. Ela foi resultado de um esforço de consolidação sobre as diferentes atividades dos vendedores. Sete passos são, geralmente, reconhecidos como formadores do processo de vendas. DWYER et al. (2000) fazem a seguinte descrição dessas etapas:

- *Prospecção:* Vendedores utilizam diferentes técnicas para identificar clientes em potencial. Um cliente em potencial significa alguém que possui o desejo, necessidade, habilidade e autoridade para comprar;
- *Pré-abordagem:* Informação é coletada a respeito do cliente a ser abordado para que o vendedor se prepare para a visita. A informação é usada para qualificar o cliente em potencial e também para desenvolver a abordagem e a apresentação ao cliente. Esse estágio geralmente termina com o agendamento de uma visita ao cliente;
- *Abordagem:* Esse passo está relacionado aos primeiros minutos de uma visita de venda. O objetivo do vendedor é garantir uma boa impressão inicial, além de despertar a atenção e o interesse do cliente;
- *Apresentação de vendas:* Esta é a principal parte do processo de vendas na qual os vendedores apresentam suas ofertas e os seus benefícios. O objetivo é aumentar o desejo do cliente em relação ao produto;
- *Lidar com objeções e superar resistências:* O vendedor procura lidar com as objeções e superar as resistências para comprar o produto oferecido, por meio das respostas a objeções e enfatizando benefícios em particular para promover a decisão de compra;
- *Fechamento:* Os vendedores iniciam as decisões dos clientes por meio de métodos desenhados para solicitarem pedidos. Da forma mais apropriada e eficaz, os clientes são perguntados a comprar a oferta;
- *Atendimento pós-venda:* Os vendedores continuam a enfatizar a satisfação do consumidor no período após a venda ter sido realizada. As atividades durante esse

tempo incluem redução das preocupações do cliente após a compra, assegurar a entrega dentro do prazo, instalação ou treinamento, prover acompanhamento ou manutenção além de lidar com reclamações e questões. O objetivo é construir boa vontade e aumentar as chances de vendas futuras.

Dentro de cada uma dessas etapas existe uma grande quantidade de técnicas e ferramentas para tornar o processo de vendas mais eficiente e eficaz. Por exemplo, para prospecção de clientes, (WEITZ et al., 2004) sugere que clientes satisfeitos, clientes visitados, contatos e redes de relacionamento, além de atividades de promoções de venda possam ser usadas para gerar novos nomes. Para pré-abordagem tanto (WEITZ et al., 2004) quanto (FUTRELL, 2003) sugerem diversas informações que devem ser buscadas. O uso de técnicas como a realização de perguntas abertas para descobrir necessidades é muito discutida principalmente na literatura de vendas de varejo (PARENTE, 2000).

Para a apresentação de vendas, (FUTRELL, 2003) sugere a importância do vendedor saber relacionar os atributos dos produtos e seus benefícios com as necessidades dos clientes. É importante, também, que a apresentação de vendas seja adaptada, permitindo a exploração das vantagens que a venda pessoal possui em relação a outras formas de atender clientes (WEITZ et al., 2004).

Para lidar com objeções, são colocadas, por exemplo, formulações a respeito de formas diferentes de pagamento, o destaque de atributos do produto, lidar com resistências em relação à empresa e mesmo ao vendedor que devem ser usadas no momento que elas aparecem. Para o fechamento de vendas, diversas técnicas como tomar a frente e solicitar o fechamento da venda, oferecer itens adicionais, considerando que o principal já foi comprado, são sugeridas por (WEITZ et al., 2004) e (PARENTE, 2000). Por fim, é importante destacar que, dependendo do enfoque do autor, essa seqüência proposta de (DUBINSKY & HANSEN, 1981) tem sofrido adaptações, com a adição de etapas ou a consolidação de outras (KOTLER, 2000), (WEITZ et al., 2004), (COBRA, 2000) e (FUTRELL, 2003), no entanto a lógica da seqüência originalmente proposta por ele permanece.

A dificuldade da venda de serviços está na intangibilidade daquilo que se está oferecendo às pessoas, ou, em outras Palavras, na impossibilidade de fazer o cliente ver, tocar ou experimentar aqueles serviços (da mesma forma, é impraticável, na maioria das vezes, para o profissional de vendas demonstrar como eles funcionam). Ao propô-lo a uma empresa industrial, o vendedor pode sugeri-los como complemento de determinado produto principal ou como um bem em si mesmo.

Ninguém imagina a venda de uma máquina – seja um poderoso torno ou uma minúscula calculadora de bolso – dissociada da garantia de um componente serviço de assistência técnica. Em um mundo de produtos categorizados e muito competitivos, o serviço de apoio pós-venda, aliado a envolvente personalidade do vendedor, pode fazer toda a diferença no momento da decisão de compra. No caso, embora se caracterize como complemento do produto principal, o serviço pode assumir grande importância.

Não há problema algum, no entanto, em oferecer os serviços como um bem a parte: é possível, por exemplo, “vender” o trabalho de manutenção da frota de veículos ou dos aparelhos eletrônicos de uma empresa, mesmo que ela tenha adquirido tais produtos de quaisquer outros fornecedores. Encontra-se outro fértil campo de possibilidades em casos de serviços que custariam muito caro à empresa, caso essa se depusesse a executá-los por conta própria. Invariavelmente sai mais barato contratar pequenas firmas especializadas que se encarregam de fornecer, por exemplo, programas de processamento de dados, organização de viagens e convenções e assim por diante. Sobre esse aspecto e levando-se em conta as dificuldades inerentes à venda de qualquer coisa intangível, a negociação se desenvolve com mais agilidade.

Afinal, num certo sentido, é muito mais fácil diferenciar, mesmo que só com uma argumentação, bem elaborada o próprio serviço daquele que a concorrência oferece. A consequência de tudo isso está no fato de que a negociação com bens intangíveis requer do vendedor habilidade especial. Muitas vezes exige-se dele (independentemente de sua competência técnica) a peculiar capacidade de estimular a fantasia do interlocutor, fazendo-o visualizar a aplicação prática e os benefícios decorrentes do serviço em sua empresa.

B - Desempenho Comercial

Os objetivos da mensuração de qualquer atividade, processo ou unidade de negócio são monitorar, controlar e orientar decisões. As medidas de monitoramento procuram acompanhar o histórico do desempenho do sistema de administração de vendas, do tipo de nível de serviço e dos componentes de custos. As medidas de controle procuram acompanhar o desempenho em andamento, e são usadas para refinar o processo de administração de vendas, no sentido de retornar à conformidade, de acordo com padrões de controle. E as medidas de orientação são definidas com o intuito de motivar ações das pessoas para que se chegue aos objetivos perseguidos.

As medidas de desempenho são definidas tendo-se em vista a perspectiva da atividade ou da produtividade. Quando são baseadas na atividade, o foco está na eficiência e na eficácia do esforço do trabalho principal, mas nem sempre no desempenho do processo, como um todo, de satisfação do cliente. Em geral, são expressas em termos de tempo, tais como entrada de pedido, entrega de pedido, seleção de pedido, reclamação por pedido, entrada de pedido por cliente, seleção de pedido por cliente, entrega por cliente, seleção de pedido por produto e entrega por produto.

A mensuração do processo considera a satisfação do cliente em todas as fases de um ciclo de comercialização, ou seja, pré venda, venda e pós venda. Verifica-se o tempo total de desempenho do ciclo, ou a qualidade total do serviço, atributos que medem a eficácia coletiva de todas as atividades requeridas para satisfazer os clientes.

Mecanismos de monitoramento são essenciais para um sistema de avaliação e controle que vise assegurar, à administração, que a operação total está dentro dos parâmetros estabelecidos. O propósito do uso desses mecanismos é sinalizar alguma anormalidade que requeira ações corretivas e prevenção da recorrência de problemas.

A natureza da mensuração requer que diversos níveis de informação sejam desenvolvidos dentro da empresa, e tem como regra geral que, quanto mais alto é o nível de gestão, mais seletivos devem ser os relatórios. Na mensuração administração de vendas quatro níveis de informação podem ser aplicáveis: direção, variação, decisão e política.

- No nível de direção, os fluxos de informações e mensurações estão relacionados com a execução do plano operacional. O foco está nas transações do dia-a-dia e limita-se a regras de decisão predeterminadas, gerando uma base de dados que auxilia a avaliação da eficiência e da eficácia, o monitoramento de tendência e a detecção de exceções;
- As medidas de variação relacionam-se com os desvios dos planos e resultam na identificação de tendências, podendo apresentar indicações de exceções úteis ao nível de direção;
- As medidas de decisão estão relacionadas com modificações no plano operacional, devido à materialização de exceções ou problemas relacionados à variação ou direção. São medidas muito seletivas, que devem ser avaliadas em termos das conseqüências desses problemas na totalidade do sistema;
- As medidas de políticas envolvem mudanças de objetivos e formulação de novas políticas, o que requer a avaliação do desenho do sistema planejado, assim como a do custo total a atingir.

C - Informações Gerenciais de Vendas

O processo de tomar decisões qualificadas em vendas pressupõe que elas sejam tomadas levando em conta informações concretas sobre o assunto em pauta. Largamente utilizada a intuição é uma armadilha perigosa por eivada de paradigmas. Quanto mais distante do dia-a-dia da venda, menos baseadas na realidade e mais apoiadas em convicções pessoais são as decisões.

Muitas das queixas dos gestores comerciais, a respeito do desempenho das suas equipes de venda, é uma coleção de suposições sem fundamento técnico. Avaliar o resultado da venda usando como parâmetro o valor vendido, não dá nenhuma informação sobre as causas que levaram a um desempenho inferior ou superior.

Uma análise de vendas envolve a coleta, classificação, comparação e estudo dos dados de venda da empresa. Isso pode ser feito checando a evolução temporal dos dados ou diversos cruzamentos possíveis e comparações com fontes de dados internas e externas. É importante identificar concentrações de vendas ou mesmo lucros em produtos, vendedores ou territórios, para orientar a política de investimentos. É sempre comum que alguns produtos, vendedores e territórios representem grande parte das vendas (ZOLTNERS et al., 2001). Para mensurar corretamente o resultado do esforço comercial se precisa ter em mão informações que possam dar ao gestor meios de corrigir procedimentos, tomar decisões de aplicação imediata e aferir os resultados.

Quando se fala em sistema de informações gerenciais de vendas, esta-se referido ao conjunto de procedimentos necessários para dar à empresa um conhecimento completo do mercado em que atua e do reflexo das suas decisões comerciais. As palavras chave de um sistema de informações são comparação e tendência. Comparação entre o que se pensa em executar e o que foi efetivamente executado. Tendência é o provável resultado futuro do comportamento das ações empreendidas até aqui.

C.1 - Modelagem de um Sistema de Informações.

O motor da criação de um sistema de informações gerenciais é a curiosidade para esclarecer as causas dos números de vendas. Os sistemas de informação de apoio à decisão têm duas fases distintas bem definidas. No primeiro momento é necessário definir que informações devem ser analisadas e para que analisá-las. No segundo momento se precisa criar as rotinas, ferramentas, rituais e processos de análise.

Todo o esforço da área de vendas está direcionado em levar ao mercado as estratégias e objetivos da empresa. A cesta de produtos ofertados, o preço, as quantidades, prazos de pagamento e prazos de entrega constituem a proposta da empresa para relacionar-se com o mercado escolhido.

Considera-se que a empresa tem claro tanto os segmentos que pretende atender, bem como, tem definido o posicionamento para cada segmento. Quando a empresa não tem claro esses pontos – segmentos e posicionamento, a área de vendas encontra dificuldades sérias em suas atuação. O conceito de que basta contratar vendedores que eles se encarregam do resto é uma forma obsoleta e ineficiente de abordar o mercado.

Algumas empresas atribuem à área de vendas tarefas relativas ao planejamento estratégico e ao marketing. É preciso salientar que toda a habilidade, conhecimento e foco da área de vendas deve ser o de “realizar negócios de curto prazo a partir do que a empresa disponibiliza de produto/serviço, preço e condições de competição com os concorrentes”. A atividade de venda é essencialmente do dia-a-dia e nisto deve estar focada. Decisões sobre produtos e mercados são atribuições de outras áreas. Quando, por qualquer razão, o gerente de vendas é o responsável pela definição das estratégias comerciais faz-se necessário que o mesmo tenha habilitação e conhecimento técnico para conduzir as duas atividades simultaneamente.

Para a correta modelagem de um sistema de informações gerenciais a área de vendas precisa ter claros alguns parâmetros definidos pela empresa, sendo que todas as informações listadas precisam estar disponíveis de forma a propiciar avaliações discriminadas por critérios de mercado ou pela estrutura de vendas, podendo-se ter as seguintes informações para análise:

.Quadro 2.12 – Informações dos clientes.

Classificação ABC por valor	Lista de clientes ordenada por volume financeiro de compra de serviços.	Identificar o grau concentração financeira das vendas.
Classificação ABC por volume	Lista de clientes ordenada por quantidade comprada de serviços.	Identificar o grau de concentração por quantidade das vendas
Classificação ABC por variedade	Lista de clientes ordenada por número de itens diferentes comprados de serviços.	Identificar o grau de concentração por variedade de itens.
Classificação ABC por Margem R\$	Lista de clientes ordenada por margem bruta de contribuição R\$.	Identificar o resultado da venda em cada cliente.
Classificação ABC por Margem %	Lista de clientes ordenada por margem bruta de contribuição %.	Identificar o resultado da venda em cada cliente.
Clientes Ativos	Relação dos clientes que compraram no mês.	Identificar a quantidade de cliente ativos no mês.
Clientes Inativos	Relação dos clientes que não compraram no mês.	Identificar a quantidade de clientes inativos no mês.
Clientes novos no mês	Relação dos clientes novos	Identificar a entrada de novos clientes
Comparativo Ativos x Inativos	Cálculo do % de clientes atendidos no mês.	Verificar aproveitamento da carteira.

Fonte: Elaborado pelo autor.

Quadro 2.13 – Informações dos serviços.

Classificação ABC por valor	Lista de produtos ordenada por volume financeiro de compra de serviços.	Identificar o grau concentração financeira das vendas.
Classificação ABC por volume	Lista de produtos ordenada por quantidade comprada de serviços.	Identificar o grau de concentração por quantidade das vendas
Classificação ABC por Margem R\$	Lista de serviços ordenada por margem bruta de contribuição R\$.	Identificar o resultado da venda em cada produto.
Classificação ABC por Margem %	Lista de serviços ordenada por margem bruta de contribuição %.	Identificar o resultado da venda em cada produto.
Produtos Ativos	Relação dos serviços vendidos no mês.	Identificar a quantidade de produtos ativos no mês.
Produtos Inativos	Relação dos serviços que não venderam no mês.	Identificar a quantidade de produtos inativos no mês.
Comparativo Ativos x Inativos.	Cálculo do % de serviços vendidos no mês.	Verificar aproveitamento do Cesta de produtos.
Efetividade das vendas	Diferença entre o ofertado/proposto e a venda efetiva.	Identificar problemas relacionados à política comercial.

Fonte: Elaborado pelo autor.

Quadro 2.14 – Informações da gestão de vendas.

Relação de Oportunidades Identificadas por Consultor	Montante das oportunidades identificadas no mês.	Identificar a demanda do mercado.
Relação das Propostas feitas no mês por Consultor.	Montante das ofertas feitas no mês.	Identificar o montante de negócios possíveis no mês.
Relação dos Contratos Fechados do mês por Consultor.	Montante da venda real no mês.	Identificar o resultado final das vendas realizadas.
Comparativos entre Oportunidades e Propostas.	Cálculo da efetividade de propostas	Identificar o grau de aproveitamento das oportunidades de negócio.
Comparativos entre Propostas e Contratos Fechados por Consultor.	Cálculo da efetividade da venda	Identificar o grau de aproveitamento das oportunidades de negócio.
Metas de Venda por Consultor	Relação dos objetivos de Venda	Avaliar o grau de atingimento dos objetivos.
Quantidade de Visitas Realizadas	Montante das ofertas feitas no mês.	Identificar o montante de negócios possíveis no mês.
Quantidade de contatos customizados por serviço.	Relação dos contatos customizados feitos	Avaliar a intensidade do esforço

Fonte: Elaborado pelo autor.

Quadro 2.15 – Informações da concorrência.

Relação de Oportunidades Identificadas da Concorrência por Consultor	Montante das oportunidades identificadas da concorrência no mês.	Identificar a demanda do mercado.
Relação das Propostas feitas com base nas Oportunidades da Concorrência no mês por Consultor.	Montante das ofertas feitas no mês.	Identificar o montante de negócios possíveis no mês.
Relação dos Contratos Fechados com base nas Oportunidades da Concorrência do mês por Consultor.	Montante da venda real no mês.	Identificar o resultado final das vendas realizadas.

Fonte: Elaborado pelo autor.

C.2 – Informações Comparativas e de Tendência.

As informações comparativas demonstram o comportamento de uma determinada variável utilizando duas bases diferentes. Por exemplo, na venda média por vendedor, a variável é a venda média e as bases comparativas são os vendedores. O objetivo da informação comparativa é sempre criar um ranking hierarquizado. Quando se utiliza análises comparativas se defronta com a dúvida sobre o número ou percentual “ideal” para aquela variável. Para encontrar esse valor meta temos três fontes:

- Um número ou percentual externo obtido por pesquisa, informação setorial ou de uma empresa “benchmarking”;

- Um número ou percentual meta que se tenha estabelecido;
- O valor médio de toda a empresa para a mesma variável.

Os três valores permitem o estabelecimento de planos de melhoria; O “benchmarking”, o valor meta e o valor médio falam por si mesmos, identificando quais as áreas, funções ou eventos estão com desempenho superior ou inferior daquele valor. Desta forma informações comparativas nos fornecem a possibilidade de discriminar desempenhos e adotar ações corretivas.

As informações de tendência demonstram o comportamento e a tendência de uma determinada variável utilizando à mesma base ao longo do tempo. Por exemplo: venda média nos últimos 12 meses, a variável é a venda média e a base é mensal. As informações de tendência ficam mais bem representadas em forma de gráfico.

As informações de tendência demonstram a trajetória de comportamento da variável durante uma determinada base de tempo, propiciando à empresa atuar sobre esta variável para manter ou alterar sua tendência.

Se a mediação da venda média diária de um determinado produto indica uma tendência de queda se pode promover ações relativas ao preço para corrigir a tendência. As informações listadas apenas indicam um caminho para a construção de um sistema de informações gerenciais de vendas. Reforça-se que só é possível gerenciar, melhorar desempenho e tomar decisões qualificadas a partir do exame das causas dos resultados de vendas.

D - Indicadores de Desempenho da Força de Vendas

Definir indicadores de qualidade e produtividade de vendas é imprescindível para avaliar o desempenho atual e acompanhar o progresso alcançado por empresas do setor envolvidas na implantação de programas de melhoria da qualidade. A existência de indicadores setoriais permite que cada empresa compare seu desempenho com outras empresas do setor e avalie seu nível de competitividade, estabelecendo suas metas para uma melhoria contínua.

O conjunto de indicadores tem um caráter dinâmico. À medida que as empresas implantam a medição dos indicadores propostos, podem ser feitos ajustes e alterações no sistema de indicadores. Essa parceria é estabelecida a partir do comprometimento das empresas do setor com os critérios propostos para a medição dos indicadores. O gerente de vendas pode melhorar de forma significativa o desempenho da sua força de vendas adotando algumas medidas simples de avaliação. A partir do relatório de visitas e de algumas outras informações é possível identificar as diferenças de desempenho entre os componentes da força de vendas.

Na área de gestão de vendas existe um paradigma que é difícil conseguir as informações necessárias porque os vendedores e clientes não as fornecem, ou porque os sistemas da empresa não contemplam os dados necessários ao acompanhamento. Isso pode ser contornado por meio de informações gerenciais conforme abaixo:

- Relatório de Visitas: É o instrumento básico de prestação de contas da atividade do vendedor e do representante, por isso indispensável. Abrir mão desse instrumento é abrir mão do gerenciamento da equipe. Não procede a alegação de que “vendedor não gosta de preencher relatório”, nem a de que exigir esse tipo de documento cria vínculos trabalhistas com o representante. No papel de gerente de vendas é sua a tarefa de manter o registro da atividade da sua equipe de vendas;
- Sistemas Internos: Não há mais justificativa para que os *softwares* utilizados desconheçam as necessidades da área de vendas para o seu bom gerenciamento. As informações necessárias devem ser agregadas aos sistemas de clientes e faturamento para que a área comercial extraia as informações necessárias para avaliar o desempenho da força de vendas.

Para que as informações tenham algum valor é necessário que sejam comparadas entre si, pela sua média ou com um objetivo. Por exemplo: a informação de que o vendedor fez 66 visitas no mês não significa nada se não for comparada com a mesma informação de outro vendedor, a média de visitas de toda a equipe de vendas ou da meta de visitas por mês daquele vendedor ou da equipe de vendas. Exemplos de indicadores que podem ser utilizados para identificar e corrigir falhas no desempenho da força de vendas podem ser:

Quadro 2.16: Sugestão de indicadores de desempenho de vendas.

INDICADOR	FÓRMULA
Novos Clientes	Clientes novos/Total de Clientes
Visitas	Visitas realizadas/Total de clientes da carteira
Perda de Clientes	Clientes que não compraram/ Total de Clientes
Rendimento	Clientes que compraram/Total de Clientes
Proporção da venda	Clientes que não compraram/Clientes que compraram
Clientes sem compra - 60 dias	Clientes sem compra 60 dias/Total de clientes
Clientes sem compra - 90 dias	Clientes sem compra 90 dias/Total de clientes
Valor médio da venda	Valor total dos pedidos/número de pedidos
Visitas	Número de visitas do mês/Dias úteis de trabalho
Produtividade	Número de pedidos do mês/Dias úteis de trabalho
Efetividade	Número de pedidos do mês/Número de visitas realizadas
Volume	Valor dos pedidos no mês/potencial de compra dos clientes que compraram
Participação	Valor dos pedidos do mês/potencial de compra da carteira
Produção diária	Valor dos pedidos do mês/Dias úteis de trabalho
Pedidos rejeitados Número	Número de pedidos rejeitados/Total de pedidos
Pedidos Rejeitados Valor	Valor dos pedidos rejeitados/Valor total dos pedidos.
Média de produtos por pedido	Número total de produtos vendidos/Quantidade de pedidos
Média da quantidade por pedido	Quantidades de todos os pedidos/número de pedidos
Satisfação do Cliente I	Reclamações dos clientes/Visitas a esses clientes
Satisfação do Cliente II	Reclamações dos clientes/Número de clientes ativos
Custo por pedido	Custo total do vendedor/número de pedidos
Custo da venda	Custo total do vendedor/valor dos pedidos
Custo por visita	Custo total do vendedor/número de visitas
Prazo médio de pagamento	Total dos dias de prazo/número de pedidos

Fonte: Elaborado pelo autor.

E – Orientações para a Análise de Resultados

Se o resultado dos indicadores não é satisfatório, análises devem ser feitas, visando garantir a integridade do processo e promover sua melhoria (PDCA) ou mesmo sua estabilidade (SDCA). Ações isoladas e sem planejamento têm efeito paliativo e imediato, mas não garantem a estabilidade do processo a médio e longo prazo, porque o verdadeiro problema não foi identificado e tratado adequadamente, ou seja, o efeito torna-se o único foco dessas ações sem que as causas fossem claramente apresentadas. Planos de ação estruturados têm função de integração dentro da Organização, tamanha a interação entre áreas decorrente de

uma série de interfaces entre processos, o que vai de encontro a crescente necessidade de informações que possibilitem a descentralização das decisões, o atendimento às expectativas dos clientes e a melhoria dos processos produtivos.

Indicadores estão intimamente ligados a processos, pois são frutos de suas necessidades, ou seja, esses existem para evidenciar o dia-a-dia das atividades relacionadas aos processos. Partindo desse princípio, os resultados dos indicadores em si não são os mais importantes aspectos a serem levado em consideração, pois é a fotografia de algo que já aconteceu. O fundamental é a análise que é feita dos resultados e qual a estratégia de melhoria a ser implementada daqui por diante, ou seja, o que é feito em função do resultado alcançado.

Para que o resultado de um indicador seja confiável, devem ser garantidas a rastreabilidade e a acessibilidade das informações que serviram de suporte ao mesmo. O registro e a adequada manutenção e disponibilidade dos dados, resultados e memórias de cálculo são essenciais à pesquisa dos fatores que afetam o indicador. É fundamental, portanto, que as bases de dados não sejam alteradas ou manipuladas, mesmo que para justificar possíveis problemas no processo. Os resultados devem refletir a realidade das atividades e processos de forma a serem úteis à análise e melhoria dos mesmos. As bases de dados são fotografias do que realmente se passou com o processo durante o período de apuração.

Se algo aconteceu fora do esperado, é evidenciado por meio dos resultados dos indicadores associados. Isto é um fato e não pode ser mudado. O conjunto desses registros ao longo do tempo nos dará informações suficientes de análise da evolução histórica de cada indicador, premissa básica à gestão de processos por indicadores, subsidiando a constatação da eficácia das ações que foram implementadas para a solução do problema.

É importante ter sempre em mente que todos os desvios devem ser justificados, analisados e controlados por meio de ações planejadas e monitoradas. Jamais esses desvios devem ser manipulados ou atenuados, pois podem encobrir problemas sérios e recorrentes, desta forma, se estar sempre acobertando situações críticas e crônicas, prejudicando severamente a análise do comportamento processo ao longo do tempo por meio da evolução histórica de indicadores. Associando-se esse registro histórico aos planos de ação aplicados e às respectivas justificativas, temos uma ampla visão do “por que” e “como” determinados desvios aconteceram e a solução utilizada para controlá-los.

Toda tomada de decisão deve levar em conta interpretação das variações inerentes ao processo. Deve-se atentar para a distinção entre variações observadas que indicam tendências ou demonstram aleatoriedade, para evitar-se problemas como culpar pessoas por problemas fora do controle delas, adquirir produtos, serviços ou melhorias desnecessariamente ou ainda

desperdiçar tempo procurando explicações para uma suposta tendência, quando na verdade nada mudou. É de suma importância então o levantamento das causas dessas variações (NEELY et al., 1995).

- Causas comuns – São inerentes ao processo e afetam as suas saídas. São aleatórias e previsíveis. Requerem uma profunda análise do processo e possivelmente demandam mudanças no mesmo. O seu tratamento deve estar associado a diagramas de causa e efeito, planos de ação e gráficos seqüenciais;
- Causas especiais – Não são inerentes ao processo. Surgem em circunstâncias específicas, são imprevisíveis e instáveis. São facilmente identificáveis e estão associadas a fontes pontuais. O seu tratamento deve estar associado a gráficos seqüenciais.

A melhoria contínua de um processo requer a adoção de uma metodologia adequada para a interpretação das variações. O que normalmente ocorre é somente a eliminação das causas especiais, mais facilmente identificáveis. Desta forma o processo fica sujeito às causas comuns não identificadas e, conseqüentemente, não tratadas adequadas.

F - Avaliação e Controle de Resultados de Venda

É difícil avaliar o desempenho de um vendedor. Muito embora suas atividades sejam de grande importância para o alcance dos objetivos da empresa, trabalham longe da matriz da empresa e longe dos olhos do gerente de vendas. Existe uma crença de que o controle sobre os vendedores possa impactar, negativamente, no desempenho de vendas, estudos recentes têm mostrado aspectos positivos do acompanhamento próximo dos homens de venda, levando a um maior desempenho e também satisfação, comprometimento, diminuição de hiatos de expectativa e conflitos de papéis existentes, entre outros aspectos positivos (CRAVENS et al., 2004).

Um erro comum em avaliação de vendas é confundir eficácia com desempenho. Eficácia é o resultado de vendas de um vendedor que incorpora não só o seu esforço, mas também outras variáveis como potencial de território, acontecimentos ocasionais no território, nível de competição e assim por diante. Já desempenho de vendas refere-se aos fatores que estão sob o controle do vendedor. Tanto medidas de desempenho como eficácia podem ser divididas entre medidas de *atividades* e resultados ou ainda a combinação de ambas que são os índices de vendas CHURCHILL et al. (2000), mostrados no quadro 2.17 a seguir.

Quadro 2.17: Medidas complementares para avaliação de desempenho.

Fatores de Avaliação de Resultados		Fatores de Avaliação Atividades		Índices Comumente utilizados para avaliar vendedores			
Pedidos	Número de pedidos	Visitas	Número de visitas	Índices e Despesa	Índice de despesa de vendas: Despesas / vendas		
			Número de visitas planejadas		Custo por visita: Custos totais / número de chamadas		
			Número de visitas não planejadas		Índice de penetração em clientes: Clientes ativos / Total de clientes disponíveis		
	Pedido médio	Tempo e Utilização do Tempo	Dias trabalhados	Índice de Desenvolvimento de Contas e Serviços	Índice de conversão de novos clientes: número de novos clientes / número total de clientes		
			Visitas por dia		Índice de clientes perdidos: Clientes que não compraram / número total de clientes		
			Tempo de vendas <i>versus</i> tempo de não vendas		Índice de vendas por cliente: Total de vendas / número total de clientes		
	Número de pedidos cancelados	Despesas	Total		Valor do pedido médio: Valor das Vendas totais / número total de pedidos		
			Por tipo		Índice de cancelamento de pedidos: Número de pedidos cancelados/ número total de pedidos		
	Clientes	Número de clientes ativos	Atividades não Relacionadas à Venda		Como porcentagem das vendas	Atividade de Visitas e Produtividade	Índice de visitas por dia: Número de visitas / número de dias trabalhados
					Como porcentagem da quota		
Número de novos clientes		Comunicações escritas para clientes potenciais			Índice de visitas por cliente: Número de visitas / Número de clientes		
		Chamadas telefônicas feitas para clientes					
Número de clientes perdidos		Número de propostas formais desenvolvidas			Índice de visitas planejadas: Número de visitas planejadas/ Número total de visitas		
		Colocação de display de vendas					
		Número de encontros com distribuidores					
Número de clientes prospectados		Número de treinamentos desenvolvidos com distribuidores		Índice de pedidos por visita: Número de pedidos / número total de visitas			
		Número de visitas para prestação de serviços					
		Número de Cobranças atrasadas coletadas					

Fonte: Elaborado pelo autor a partir de CHURCHILL et al. (2000).

Além das medidas acima podem ser desenvolvidas medidas subjetivas a respeito do trabalho dos vendedores com o objetivo de enriquecer ainda mais a sua avaliação. Os critérios que podem ser usados é o conhecimento do vendedor em relação à empresa, seus produtos e suas atividades, administração do território (planejamento de visitas, controle de despesas, entre outros), relações com a empresa e consumidores e características pessoais.

Uma discussão importante é o tipo de avaliação que deve ser usada, basicamente considerando os dois tipos citados, as metas de resultados ou metas de atividades. De acordo com (ZOLTNER et al., 2001), se as vendas de uma empresa dependem pouco do esforço de vendas em relação às outras variáveis de *marketing*, o sistema de avaliação deve ser baseado principalmente nas atividades realizadas pelos vendedores. Se, por outro lado, as vendas da empresa dependem fortemente do esforço de vendas, então devem ser baseadas em resultados.

Metodologia da Pesquisa

3.1 Tipo de Pesquisa

Uma pesquisa pode ser definida como um processo formal e sistemático de desenvolvimento do método científico de forma a se descobrir respostas para problemas por meio do uso de procedimentos científicos (GIL, 1994). São vários os autores que ressaltam a importância do planejamento da pesquisa para que se possam obter informações confiáveis e adequadas aos seus propósitos. Segundo SELTZ (1974): "uma vez que o problema de pesquisa tenha sido formulado de maneira suficientemente clara para que se possam especificar os tipos de informações necessárias, o pesquisador precisa criar o seu planejamento de pesquisa que varia de acordo com o objetivo da mesma".

Obviamente, não existe nenhum método que seja melhor ou pior que algum outro. O que se deve é procurar uma melhor adequação entre o método, o objetivo e as condições nas quais uma pesquisa está sendo realizada (YIN, 2005). Segundo BERVIAN & CERVO (2003), "a pesquisa é uma atividade voltada para a solução de problemas teóricos ou práticos com o emprego de processos científicos". Ainda segundo os mesmos autores, a pesquisa científica pode ser classificada em bibliográfica, descritiva e experimental. A pesquisa bibliográfica procura explicar um problema a partir de referências teóricas publicadas. A pesquisa descritiva observa, analisa e correlaciona fatos ou fenômenos (variáveis) sem manipulá-los. E por fim, a pesquisa experimental caracteriza-se por manipular diretamente as variáveis relacionadas com o objeto de estudo.

Esse estudo se encaixa, primordialmente, no perfil da pesquisa descritiva, pois segundo BERVIAN & CERVO (2003), a pesquisa descritiva procura descobrir, com a precisão

possível, a frequência com que um fenômeno ocorre, sua relação e conexão com outros, sua natureza e características, desenvolvendo-se principalmente, nas ciências humanas e sociais. Na pesquisa descritiva, utilizam-se os chamados estudos descritivos, que BERVIAN & CERVO (2003) classificam como estudos e descrições das características, propriedades ou relações existentes na comunidade, grupo ou realidade pesquisada que contém os estudos que visam identificar as representações sociais e o perfil de indivíduos e grupos, como também estudos para identificar estruturas, formas, funções e conteúdos. Para tanto, foram utilizadas ferramentas de pesquisa bibliográfica, estudos exploratórios e elementos de estatística para desenvolvimento e conclusão desta dissertação.

3.2 Problema de Pesquisa

Pelo exposto anteriormente, pode-se observar que um problema existente no campo da medição de desempenho é a falta de métodos de estabelecimento de relacionamentos entre as medidas de desempenho a partir de uma grande massa de dados históricos. Além disso, existe uma tendência de estabelecer o relacionamento esperado de forma que a medição de desempenho possa refletir o futuro desejado. Entretanto, é de grande valia aprender por intermédio daquilo que já foi feito, ou seja, pelas ações passadas refletidas nas medidas de desempenho.

Outro ponto a destacar da revisão bibliográfica feita nesta dissertação é que, com o aumento da complexidade das organizações e a adoção de diferentes estratégias competitivas, como diferenciação, a complexidade da medição de desempenho aumentou, resultando num conjunto de medidas de desempenho muito maior. Dessa forma, fica um tanto quanto difícil manipular dados históricos sobre desempenho para identificação de padrões de relacionamento sem lançar mão de abordagens, técnicas e ferramentas da tecnologia de informação. O uso da tecnologia de informação nas empresas tem gerado uma quantidade enorme de dados na organização, que pode dificultar a sua utilização. Por isso, as abordagens modernas de TI (*data warehouse*, por exemplo) procuram, em um dos seus usos, lidar com a questão da grande quantidade de dados gerados.

As variadas formas de aplicação da tecnologia de informação na medição de desempenho não são, pelo menos na literatura pesquisada, algo amplamente difundido. As aplicações mais comuns se limitam à automatização de atividades de coleta, compilação e categorização dos dados. Muitas aplicações procuram oferecer ao usuário um acesso mais

amigável (*interface*) e fácil (diferentes plataformas de forma remota). Contudo, essas aplicações não causam tanto impacto como as aplicações que permitem análises mais complexas dos dados sobre desempenho. O entendimento de como o desempenho vem sendo atingido e como pode ser atingido permite ampliar muito mais o poder da medição de desempenho para dar suporte e viabilizar decisões com base em informações da eficiência e eficácia de decisões passadas (NEELY, 1998). Como cada vez é maior o número de variáveis de relacionamento entre as medidas de desempenho, a modelagem tem que ser descritiva.

3.3 Detalhamento dos Objetivos da Pesquisa e das Hipóteses Definidas

O objetivo geral de pesquisa da presente dissertação é: *“Existe relação de dependência entre os indicadores de desempenho e a produtividade de vendas? É possível determinar a importância desses indicadores, os tipos de relações eventualmente encontradas ou as inter-relações entre eles?”*. Os objetivos específicos desse estudo são os seguintes:

- Verificar se existe um relacionamento significativa entre as características dos indicadores de desempenho de vendas com a produtividade de vendas;
- Descrever e elucidar, caso exista, uma forma de relacionamento entre os indicadores de desempenho de vendas disponibilizados pelas empresas de telecomunicações e produtividade de vendas.

As hipóteses propostas para esta pesquisa, considerando que as variáveis de desempenho de vendas são as variáveis independentes, e a variável de produtividade de vendas é a variável dependente que se usará, são as seguintes (Figura 3.1):

- **H₀**: Não existe relação entre as características da produtividade de vendas e os indicadores de desempenho de vendas;
- **H₁**: Existe relação entre as características da produtividade de vendas e as variáveis de desempenho de vendas das empresas.

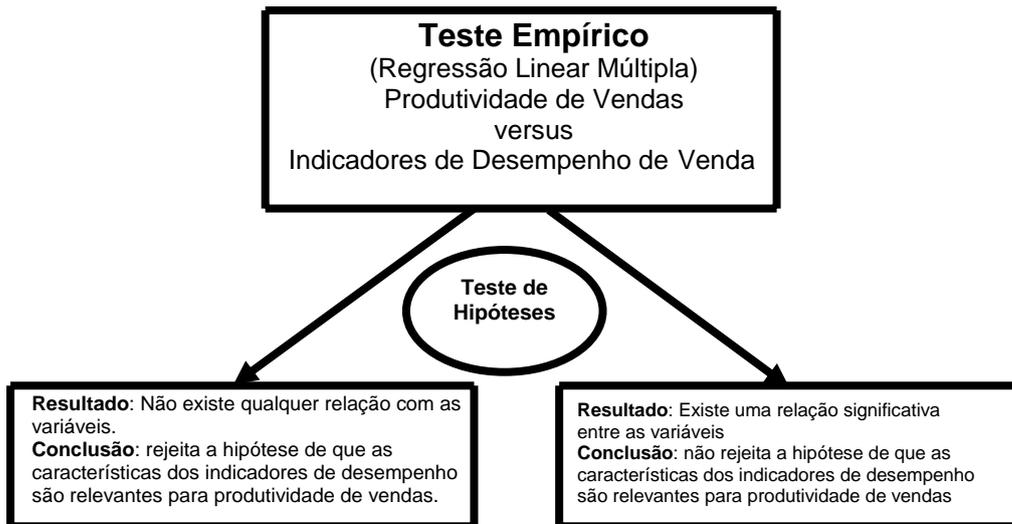


Figura 3.1 - Modelo esquemático do teste de hipóteses.

Fonte: Elaborada pelo autor.

Caso seja confirmado a existência de relacionamento entre as características da produtividade de vendas fixa com as variáveis de desempenho de vendas das empresas, é realizado ajuste no modelo para verificar se a produtividade de vendas fixa é fortemente relacionada com as variáveis especificadas pelos especialistas de vendas em serviços de telecomunicações, ou seja, a Produtividade de Vendas Fixa tem dependência linear com as seguintes variáveis independentes (*CO1* Vendas x Mapeamento da Concorrência, *EO1* Conversão de Contratos Fixa, *FC1* Média de Visitas por Consultor, *GP1* Produtividade de Visitas, *GP2* Eficácia de Visita, *GP5* Produtividade de Oportunidades e *LA2* Volume de Vendas Com Descontos).

3.4 Estabelecimento de Relacionamentos entre as Medidas de Desempenho

Os requisitos necessários para a sua aplicação, é que a empresa faça parte de qualquer atividade industrial ou de serviços, e que possua um sistema de medição de desempenho estruturado e um sistema de informação, com ferramentas de tecnologia de informação modernas, ou seja, uma estrutura de *hardwares*, *softwares*, comunicação e redes que realizem a integração dos computadores. Ela precisa possuir bancos de dados transacionais, onde estão armazenados os dados sobre desempenho, e onde são utilizadas as ferramentas e técnicas de análise.

Para aplicar esse método proposto é necessário criar um modelo conceitual levantando informações da empresa sobre as medidas de desempenho, os relacionamentos pressupostos por tomadores de decisão da organização, quais dados à empresa possui e como

estariam dispostos em um banco de dados. Nesta etapa, é possível definir o desafio a ser realizado na organização para a execução do método e estabelecer um modelo prévio de relacionamentos entre as medidas de desempenho a partir da análise dessas informações. A partir do modelo conceitual, são estabelecidos os relacionamentos entre as medidas de desempenho com base no uso das técnicas e ferramentas de Mineração de Dados. A escolha da(s) técnica(s) e ferramenta(s) é baseada na estrutura dos dados contida no banco de dados histórico e na análise do modelo conceitual obtendo-se um modelo de *data warehouse* propício para a operacionalização de Mineração de Dados. O objetivo da utilização da(s) técnica(s) e ferramenta(s) de Mineração de Dados é verificar o modelo de *data warehouse* mediante a comprovação e aperfeiçoamento das relações entre as medidas de desempenho já existentes e enriquecê-lo com a descoberta de novos padrões de relacionamento entre as medidas de desempenho. Portanto, dentro da necessidade estabelecida pela empresa, o resultado principal do método é um modelo de relacionamentos entre as medidas de desempenho. Para isso, são usados dados que podem estar centralizados em um *data warehouse* ou descentralizados em diversos *data marts*.

3.5 Detalhamento da Metodologia Implementada

A noção de metodologia ao longo desse trabalho pressupõe um conjunto de passos (fases) formado por subconjuntos (atividades) interligados entre si de forma necessária e condicionante. Esta definição permite assumir o conceito de metodologia como o modo particular de organizar ou compor as partes de um conjunto de tal forma que as relações que mantêm entre si sejam as necessárias e suficientes para garantir a integridade e harmonia intrínseca desse conjunto. Em síntese a metodologia é o princípio que organiza a forma (MARCOLLI, 1986).

O desenvolvimento desta pesquisa acontece integrado num contexto de estrutura organizacional, a partir do qual são definidos os objetivos de mercado e então colocado em prática o desenvolvimento das suas atividades (DROZDENKO et al., 2002), (ZWICK et al., 2004). Os resultados obtidos têm aplicação direta em atividades de administração de vendas, sendo, então, avaliados de modo a permitir uma aprendizagem para ações futuras.

A estratégia de desenvolvimento desse projeto obedece a um conjunto de premissas que vão desde a necessidade de fixação dos objetivos para as atividades de mercado (DROZDENKO et al., 2002) e (HUGHES, 1997), a determinação das fontes de dados

(SELLER et al. 1999) e (HUGHES, 1997) internas e externas à organização (COOKE, 2000) e da capacidade tecnológica de processamento de grandes volumes de dados, capazes de suportar atividades de Descoberta de Conhecimento em Bases de Dados (KDD), (LO, 2002) e (ZWICK et al., 2004).

A proposta da metodologia apresentada nesse trabalho tem por objetivo auxiliar a área de Gestão de Negócios em Telecomunicações a utilizar o conhecimento extraído das Bases de Dados (BD) nas suas atividades de gestão, as quais se encontram associadas a objetivos de planejamento mais vastos, que por sua vez correspondem ao reflexo da estratégia organizacional.

O desenvolvimento de um sistema de Descoberta de Conhecimento em Bases de Dados é uma tarefa complexa, principalmente pela característica não determinística desse tipo de sistema. Portanto, é imprescindível o uso de uma metodologia completa e sistemática. “Uma metodologia de engenharia de *software* é um processo para a produção organizada de *software*, com utilização de uma coleção de técnicas predefinidas e convenções notacionais. Uma metodologia costuma ser apresentada como uma série de etapas, com técnicas e notação associadas a cada etapa.” (WESTPHAL et al., 1998).

Os trabalhos que se propõem apresentar uma metodologia para o desenvolvimento de sistemas de descoberta de conhecimento não incluem formalismo na especificação desses sistemas. Normalmente, as metodologias propostas procuram solucionar questões relativas a determinadas etapas do processo de desenvolvimento desses sistemas e não apresentam notação sistêmica para representar as características do sistema como um todo. No capítulo 2, foram relacionados, com o detalhamento necessário, trabalhos que propõem uma metodologia para sistemas de Descoberta de Conhecimento em Bases de Dados.

Em virtude das diversas abordagens presentes na bibliografia, optou-se por realizar um estudo comparativo entre elas, onde foram detectadas algumas similaridades. Porém, nenhuma das abordagens adequou-se ao estudo de caso proposto. Consequentemente, é proposto um modelo do processo integrando essas diferentes abordagens. A metodologia proposta para o desenvolvimento desta pesquisa é recorrente à exploração dos conceitos e características do CRISP-DM, cruzando-o quer com as atividades de vendas, quer com as questões inerentes à integração dos modelos de MD (referido aqui como componente integrante do processo de KDD).

A sistematização desenvolvida foi denominada como **MDAV** (**M**ineração de **D**ados para **A**dministração de **V**endas), tendo sido implementado com a pretensão de ser suficientemente genérico, podendo ser utilizado nas mais diversas e diferentes áreas de

negócios. O estudo de caso analisado situa-se no domínio comercial na área de telecomunicações, pretendendo-se aproveitar a oportunidade surgida de um caso real e uma vez que nesse domínio se pode facilmente evidenciar a sistematização dos passos desenvolvidos e a sua aplicação prática.

As diferentes visões do processo de KDD em etapas sugerem existir uma trajetória seqüencial das etapas. No entanto, isso não se verifica, uma vez que em cada etapa pode ser identificada a necessidade de retorno para uma das etapas anteriores. Por exemplo, se já numa etapa avançada, na qual técnicas já estão sendo aplicadas para descoberta de conhecimento, for verificada a necessidade de um dado que não havia sido previsto anteriormente, isso pode levar à retomada à fase de seleção. Assim, o processo ocorre de forma iterativa, pois cada etapa pode possuir interseções com as demais. Da mesma forma, esse processo é interativo, onde a cada etapa as informações geradas são analisadas e enriquecidas com outras informações para obter-se resultados melhores.

A sistematização desses componentes, integradas numa perspectiva organizacional, permite definir uma estratégia para o desenvolvimento de projetos, a qual está representada esquematicamente pela Figura 3.2. Esta metodologia possui três fases: Estudo do Negócio, Extração de Conhecimento e Avaliação dos Resultados Obtidos. Primeiro, os dados podem ser obtidos a partir de diferentes fontes. Após o seu registro e análise é então criado uma BD de Administração de Vendas, com vista ao suporte de toda a fase seguinte, correspondente à extração de conhecimento. A avaliação dos resultados obtidos é concretizada na fase de aplicação a atividades de Administração de Vendas.

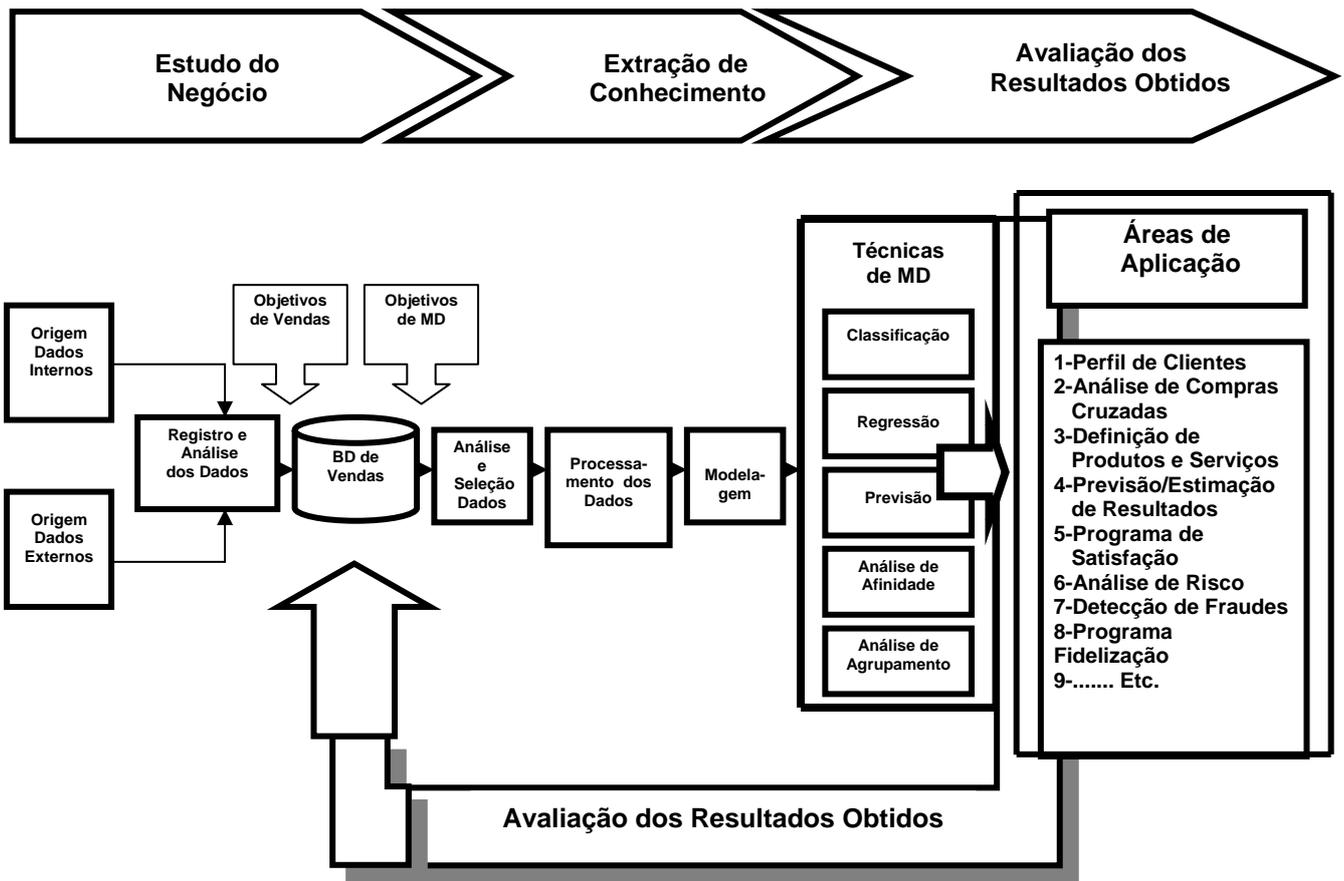


Figura 3.2 – Proposta da metodologia de Mineração de Dados (MDAV).

Fonte: Elaborada pelo autor.

Nesse método, existem três questões fundamentais para o desenvolvimento do trabalho:

- O levantamento de informações sobre a empresa de forma a criar um modelo. Sem esse entendimento, não é possível definir as necessidades dela, os fluxos de informações e dados vitais para o atendimento dos objetivos de negócio;
- A definição do banco de dados. A sua escolha vai depender do tamanho, da quantidade de variáveis presentes e das áreas organizacionais envolvidas. Um *data warehouse* ou um *data mart* conseguem lidar com dados armazenados, complicados e complexos, mas existe também possibilidade, por questão de custo e simplicidade, de aplicar as ferramentas de Mineração de Dados diretamente em bancos de dados relacionais;
- A escolha das técnicas e ferramentas de Mineração de Dados. A escolha da técnica vai depender dos tipos de dados existentes (por exemplo, dados numéricos e não-numéricos) e também de qual tipo de relacionamento a ser pesquisado entre as medidas de desempenho. Por exemplo, uma relação de causa-e-efeito requer uma comprovação a partir de cálculos matemáticos. Já a relação lógica é baseada em

deduções sobre as relações existentes. Vale observar que ambas usam técnicas de Mineração de Dados diferentes. Dessa forma, é possível estabelecer, a partir de uma grande massa de dados históricos sobre o desempenho, o relacionamento entre as medidas de desempenho de um SMD.

O caso da pesquisa aqui descrita desenvolveu-se numa das maiores empresas de telecomunicações do Brasil que comercializa produtos e serviços de telefonia básica, avançada, comunicação de dados, longa distância nacional e internacional e mobilidade. Foram coletados dados de 15 empresas com atuação no Nordeste e Sudeste do Brasil as quais são agentes autorizados e comercializam todos os produtos voltados para a área empresarial. A ferramenta de MD escolhida para o desenvolvimento dos casos de estudo foi o *software* SPSS versão 13.0.

A metodologia proposta (MDAV) se baseia na revisão teórica realizada no capítulo anterior e permitiu a identificação dos requisitos a cumprir no desenvolvimento da pesquisa realizada para verificação dos objetivos propostos anteriormente. Esse capítulo apresenta uma aplicação da metodologia proposta e encontra-se organizado de acordo com as seguintes etapas:

A) Estudo do Negócio

A1) Enquadramento do Projeto com os Objetivos do Negócio:

A2) Obtenção e Exploração dos Dados:

Avaliação dos Dados Internos Disponíveis;

Importação dos Dados e Constituição do BD Inicial;

Filtragem e Limpeza dos Dados;

A3) Constituição da Base de Dados dos Indicadores de Desempenho de Vendas:

Seleção das Informações Elegíveis;

B) Extração de Conhecimentos

B1) Análise e Seleção dos Dados;

B2) Pré-processamento e Transformação dos Dados;

B3) Modelagem;

B4) Avaliação dos Resultados.

C) Avaliação dos Resultados Obtidos

As condições da pesquisa dependeram diretamente dos dados disponibilizados pela empresa promotora havendo, desde o início, por razões de confidencialidade, uma reserva de sigilo quanto à divulgação dos dados, pelo menos no que refere a não divulgação dos parâmetros comerciais. Foram verificadas algumas dificuldades inerentes ao processo de escolha da amostra, que resultaram de dois fatores: em primeiro lugar, da confidencialidade das

informações comerciais, ou seja, a não divulgação de informações sigilosas, no qual houve necessidade de normatizar seus valores e, em segundo lugar, do tamanho dos bancos de dados existentes.

(A) Estudo do Negócio

As atividades realizadas nesse passo foram ser alcançadas com a Mineração de Dados e o que deverá ser feito com os seus resultados, após a obtenção de um conhecimento mais detalhado sobre os dados operacionais existentes. Definir um estudo pode envolver articular uma meta, escolher uma variável dependente ou uma saída que caracterize um aspecto da meta e especificar os campos de dados que são usados no estudo. Bons estudos são unidos para caracterizar aquilo que pode ser descrito com seus dados. Por outro lado, a meta pode ser usada para agrupar tipos similares de dados ou para identificar exceções em um conjunto de dados. A identificação de exceções é geralmente usada na descoberta de fraude ou de dados incorretos.

Para estabelecer os relacionamentos entre as medidas de desempenho usando a Mineração de Dados, é necessário conhecer o SMD e os sistemas de informação da organização onde o método proposto vai ser aplicado. O levantamento das informações abrange três aspectos na definição do modelo conceitual:

- Mapeamento dos possíveis relacionamentos entre as medidas de desempenho;
- Identificações de quais informações são relevantes para os tomadores de decisão e que eles ainda têm dificuldades em obtê-las;
- Levantamentos de como os sistemas de informação da empresa estão estruturados, principalmente na questão do armazenamento dos dados a fim de identificar quais são as melhores possibilidades de aplicação de Mineração de Dados.

Conforme já detalhado no Capítulo 2, o *Balanced ScoreCard* é um modelo de gestão que auxilia a Organização a traduzir a estratégia em objetivos, indicadores, metas e planos de ação, balanceados e alinhados, que direcionam comportamentos e performances. Tal como indicadores registram o dia-a-dia dos processos associados, o BSC o faz com a estratégia da Empresa. Dentro da perspectiva do *Balanced ScoreCard*, a estratégia é tarefa de todos, cada um com a sua parcela do desdobramento de metas corporativas. Isto é alcançado com o alinhamento dos indicadores ao BSC e a sua gestão nos diversos níveis hierárquicos da Empresa, ou seja, a gestão por indicadores alinhada ao BSC permitirá determinar qual a contribuição dada por cada Unidade de Negócio no resultado global da Organização.

Os atributos do BSC nas empresas analisadas, os quais foram disponibilizados por meio do sistema oficial de medição de desempenho da unidade de Negócios Empresariais, foram os seguintes: Aprendizado e crescimento, Capital informacional, Capital organizacional, Concorrência, Excelência operacional, Financeira, Foco do cliente, Gestão de parceiros, Inovação e ação e Mercado. Para o caso desse trabalho, foram identificados, na Unidade de Negócios Empresarial, os seguintes indicadores de desempenho associados com os seguintes atributos do BSC, quais sejam:

- *Concorrência*: Vendas x Mapeamento da Concorrência;
- *Excelência Operacional*: Assertividade de Entrega e Cadastro de Contratos, Retorno de Contratos, Abertura de Ordem de Serviços, Conversão de Contratos e Cancelamento de Vendas;
- *Foco do Cliente*: Média de Visitas por Consultor, Assertividade de Agendamento e Visitação x Agendamento;
- *Gestão de Parceiros*: Produtividade de Visita, Eficácia de Visitas, Cesta de Produtos, Distribuição de Vendas e Produtividade de Oportunidades;
- *Inovação e Ação*: Volume de Vendas Com Descontos, Aprovação de Viabilidade e Assertividade de Análise de Viabilidade.

Esses indicadores estão contemplados em todo fluxo de venda, ou seja:

- *Pré-visita* por meio da oportunidade de vendas, agendamento de visitas e no planejamento da visita;
- *Visita* pelo registro da visita por meio da ficha de visita eletrônica;
- *Pós-visita* na emissão de propostas e contratos e acompanhamento da abertura da ordem de serviços de instalação.

A1-) Enquadramento do Projeto com os Objetivos do Negócio

A existência de uma BD com grandes dimensões evidencia uma das maiores limitações inerentes à sua utilização na maioria das organizações que é a (in)capacidade para extrair informação relevante, para além daquilo que os processos tradicionais permitem. O projeto em estudo promoveu a constituição de uma BD com mais de 200.000 registros permitindo ainda concretizar as práticas enumeradas na administração de vendas de serviços de telecomunicações, com resultados práticos relevantes.

Uma etapa fundamental para todo o processo de KDD é a definição do objetivo, ou seja, descobrir a necessidade de se implementar um sistema para KDD. O objetivo é definido a partir do entendimento do problema que é solucionado. Existem diversos tipos de problemas que podem ser solucionados com esse processo, tais como: otimização de campanhas de marketing, prevenção de fraudes em cartões de crédito, previsões futuras, relacionamento de indicadores de desempenho, entre outros.

Para definição do objetivo, é necessário o entendimento do domínio e dos dados. Para o processo de KDD ser capaz de resolver o problema eficientemente, devem ser descobertos os fatores importantes que influenciam os resultados, ou seja, as variáveis relevantes para o processo. Embora os dados comerciais ofereçam diversas possibilidades de estudo, esse trabalho focaliza-se na análise do relacionamento entre os indicadores de desempenho de vendas e a produtividade de vendas. Especificamente, na tentativa da descoberta de padrões de relações, importância e inter-relações desses indicadores nas atividades comerciais. Esse estudo foi limitado a esse objetivo em virtude das diversas possibilidades de aplicação que surgiram.

Segundo FAYYAD et al. (1996), deve ser verificado se o processo de KDD tem chances de ser bem-sucedido. Para isso, o autor indica a utilização de alguns critérios, que, quando atendidos, determinam uma maior possibilidade de sucesso. Esses critérios são divididos em dois grupos: práticos e técnicos. Os práticos abrangem as condições externas envolvidas ao processo e os técnicos tratam de características relacionadas à base de dados disponível.

Os critérios práticos utilizados nesse estudo de caso abrangem:

- *Impacto Potencial na Aplicação*: Esse critério é medido de acordo com a aplicação, pois as informações extraídas devem ter um valor significativo para a Empresa. O resultado desse estudo pode servir como suporte aos gestores quanto um melhor planejamento comercial;
- *Falta de Alternativa*: Esse critério define quando é recomendada a utilização do processo de KDD. Por exemplo, quando existe uma grande quantidade de dados a serem analisados. Nesse estudo o processo é recomendado, tendo em vista o volume de dados acumulados nos sistemas comerciais;
- *Suporte Organizacional*: Quando há o interesse da Empresa nos resultados do processo. Esse trabalho possui suporte Organizacional das pessoas diretamente ligadas ao dia-a-dia operacional da gestão de resultados comerciais. O interesse

nos resultados é crescente, visto que nenhum estudo desta natureza havia sido realizado nos dados das empresas analisadas;

- *Problema de Ordem Legal*: Esse critério abrange a privacidade dos dados, garantindo a legalidade na utilização dos dados que contém informações privadas. Para esse estudo, foi autorizada a utilização de todos os dados dos sistemas, mantendo em sigilo as informações comerciais.

Os critérios técnicos utilizados nesse estudo de caso abrangem:

- *Disponibilidade dos Dados*: Devem existir dados suficientes para realizar as análises de padrões. Nesse estudo de caso, a quantidade de dados disponíveis mostrou-se suficiente para uma primeira análise de padrões no sistema;
- *Atributos Relevantes*: Mesmo que exista uma grande quantidade de dados esses devem possuir alguma relação com o objetivo do processo, senão os resultados não gerarão informações úteis. Nesse trabalho, existe uma quantidade suficiente de atributos relacionados ao objetivo do processo.
- *Baixo Nível de Ruído*: O nível de ruído é medido quando existe a ocorrência de valores incompletos, dados redundantes, dinâmicos ou volumosos demais. Os dados analisados nesse trabalho apresentaram um nível de ruído bastante reduzido, sendo que esse problema é resolvido durante a fase de limpeza desenvolvida na etapa de preparação dos dados.
- *Conhecimento Prévio*: Baseado no conhecimento do sistema pode-se determinar quais atributos da base de dados são mais importantes para o processo e quais padrões já são conhecidos. Esse critério foi contemplado durante a fase de entendimento, no início do processo, onde foi realizado o entendimento do domínio da aplicação e dos dados.

A2-) Obtenção e Exploração dos Dados

O desenvolvimento desse projeto como o caso em estudo envolve o manuseio de elevados volumes de dados. Decorrente desse fato, pressupõe-se a obtenção de dados, que sejam provenientes de BD internos. Envolvido no processo de obtenção de informação, esta atividade consiste em proceder à obtenção dos dados, os quais normalmente podem ser encontrados em diversos sistemas, em locais desconhecidos do pesquisador, ou mesmo em

suportes que não sejam digitais (em geral, papel). Esta fase de obtenção é morosa e os desafios na sua execução podem ser descritos da seguinte forma (PYLE, 1999):

- *Problemas Legais e Éticos* – podem existir barreiras legais ou éticas que impeçam que os dados sejam disponibilizados para análise. Em geral, no caso das instituições financeiras o acesso aos dados encontra-se regulamentado. É possível que existam também limitações de ordem ética que restrinjam o acesso aos dados como ocorre, por exemplo, nos dados relativos a informações clínicas;
- *Motivos Organizacionais e Estratégicos* – a existência de motivos de natureza organizacional e estratégicos que condicionem o acesso aos dados pois se verifica sempre que os mesmos possam revelar segredos da atividade da Empresa, como em geral, o caso dos dados relativos ao uso fraudulento de cartões de crédito, trata-se de uma informação sobre a qual os emissores de cartões de crédito pretendem o máximo sigilo, bem como de informações comerciais;
- *Razões Políticas* – o acesso aos dados ou à sua titularidade pode estar ligado a pessoas, departamentos ou organizações as quais, pelas mais diversas razões, podem não apoiar as iniciativas de utilização desses dados;
- *Formato dos Dados* – a evolução dos sistemas tem proporcionado ao longo dos anos diferentes formas e formatos para armazenamento e codificação dos dados. Esta heterogeneidade de formatos condiciona o processo de obtenção, na medida em que obriga a processos específicos para cada formato, em geral, os dados que ainda podem estar guardados em sistemas operacionais antigos e descontinuados;
- *Acesso aos Dados* – Para que os dados possam ser utilizados é necessário que os mesmo estejam acessíveis pelo sistema onde se desenvolverá o trabalho de análise. Esse fato obriga a colocação dos dados num sistema ao qual as estações de trabalho possam acessar *on-line*, forçando por vezes a duplicação dos dados;
- *Sistemas Legados* – Como já referido anteriormente, a evolução dos sistemas promoveu, com alguma frequência, a coexistência de legados com sistemas contemporâneos nas organizações. Os sistemas mais antigos, no tempo em que foram desenvolvidos ou porque se perderam com o tempo, não estão dotados de documentação que permita uma extração fácil da informação neles contida. Além disso, é possível a ocorrência de problemas de compatibilidade ao nível dos dados;
- *Aplicações Proprietárias e os Sistemas Gestores de BD* que não possuam equivalência para todos os tipos de dados existentes, levando que a unificação dos mesmos seja mais

uma tarefa que o pesquisador terá que realizar;

- *Granularidade dos Dados* – Representa o nível de detalhe em que os dados são armazenados. Os sistemas tradicionais, para que possam assegurar as suas tarefas, armazenam os dados com o maior detalhe possível (também denominado como registro de transação). No povoamento dos *DataWarehouses* (DW), por motivos de desempenho, ou seja, é comum a totalização ou sumarização de dados, isto é, os dados relativos a vendas de produtos são armazenadas de uma só vez em lugar de se guardar o valor referente a cada transação (data, tipo de serviço, quantidade e preço). Uma vez resumidos e armazenados no DW os dados dificilmente podem voltar a representar detalhe.

A2.1-) Avaliação dos Dados Internos Disponíveis

O sucesso de qualquer projeto de Mineração de Dados depende diretamente dos dados disponíveis, sobre os quais irão ser desenvolvidos. O processo da obtenção de dados, mesmo internos à Empresa, depende de um conjunto de aspectos que condicionam a disponibilidade e a qualidade dos mesmos. Os dados internos disponíveis provinham de dados contidos em diferentes sistemas de informação da Empresa.

O processo de obtenção desses dados envolveu ainda vários recursos em diferentes níveis hierárquicos da Empresa como gestores ou administradores de sistemas de informação, no sentido de se garantir o acesso aos mesmos dados. No momento da constituição das BD são considerados todos os tipos de suportes, havendo para os casos de informação em papel (em geral, formulários, pesquisas ou cupões de desconto) necessidade de se proceder à sua obtenção manual. A natureza das fontes de informação pode ser de três tipos distintos: interna, externa ou de documentos relativos a estudos de mercado e indicadores estatísticos (DROZDENKO et al., 2002):

- Coleta de dados internos, presentes nos diversos sistemas da empresa ou de empresas pertencentes à mesma Empresa; dados internos, mas registrados em suportes não tecnológicos e como tal sujeitos à obtenção manual dos mesmos (formulários, reclamações, sugestões);
- Dados externos, provenientes de empresas do mesmo grupo organizacional ou adquiridas a empresas externas;
- Dados obtidos a partir de indicadores estatísticos ou estudos de vendas realizados.

De acordo com levantamento feito junto às empresas analisadas, a Empresa dispõe de diversos sistemas para registro do seu dia-a-dia operacional. Tais dados são armazenados em bancos de dados distintos. Para esse estudo de caso, foi concebido acesso para análise e utilização dos dados do SA 3, ASE e do SCI.

Os dados foram disponibilizados em formatos distintos: parte em banco de dados Microsoft Access e parte em planilhas Excel. Todos os dados selecionados foram migrados para o banco de dados Access para acesso e manipulação.

- *Sistema de Vendas (SA 3)*: Permite à equipe de vendas do Empresarial o controle e gerenciamento sobre todo o *workflow* de vendas, possibilitando a gestão das oportunidades, gestão de atividades, planejamento de visitas, mapeamento da concorrência, ações comerciais, medição da carteira, parâmetros de desempenho e, conseqüentemente, análise do funil de vendas, além relatórios gerenciais bem estruturados e detalhados;
- *Sistema de Cadastro de Informações (SCI)*: Esse sistema armazena informações das atividades de todos os serviços de telecomunicações comercializados pelos consultores, abrangendo toda a parte contratual, ou seja: visão de contrato e suas condições, identificação dos serviços comercializados, condições de aceite para liberação de solicitação de abertura de ordem de serviço, parâmetros de desempenho, etc.;
- ASE Empresarial: Sistema gerencial que disponibiliza informações relacionadas a ações comerciais, carteira de cliente segmentada, priorização de visitas a serem realizadas, entre outras informações correlatas.

A2.2-) Importação dos Dados e Constituição da BD Inicial

A constituição da BD inicial dos parâmetros de vendas de telecomunicações, consistiu na elaboração de uma lista prévia organizada pelo tipo de informação a que cada registro correspondia, informação do cliente, do serviço comercializado (em nível de detalhes), oportunidades identificadas, mapeamento dos serviços existentes disponibilizados pela concorrência (com registro dos serviços, prazo contratual, data de vencimento, etc.) bem como várias outras informações pertinentes ao negócio. Esta lista foi criada a partir da importação do atributo, *Agente Autorizado*, de BD relativas às vendas realizadas.

Uma vez garantida a disponibilidade dos dados internos, a obtenção de dados a partir de fontes diversas, ainda que apenas a um nível de atributo básico (*Agente Autorizado*) requereu

um processo de unificação, dada a elevada probabilidade de redundância de informação sobre a mesma comercialização, pois se pode encontrar registrado em mais do que uma BD. Nesse caso após a importação das diferentes tabelas, procedia-se a uma ordenação da tabela resultante pelos atributos em análise e manualmente o pesquisador poderia detectar a ocorrência de registros duplicados. Ocorreram poucas situações com duplicação de registros decorrentes da concentração de tabelas de BD distintas proporcionando a coexistência dos mesmos atributos, com o mesmo significado, mas codificados de modo diferente, provocando redundância de informação.

A2.3-) Filtragem e Limpeza dos Dados

Como os dados foram obtidos de diversas fontes, verificaram-se erros, arquivos repetidos ou dados nulos, os quais foram excluídos do conjunto de dados para análise. A filtragem e limpeza dos dados desenvolveram-se pela aplicação das técnicas: Levantamento de incongruências (confusão e redundância de informação para indicar o mesmo objeto), determinação de inconsistências (a importação de dados de sistemas distintos podem justificar a existência de inconsistências como sejam: utilização de métricas diferentes e representações distintas para o mesmo serviço) e violações de domínio (a violação do conjunto de valores admitidos como válidos para um atributo (domínio) permitiu a inserção valores incorretos no BD, os quais alteraram a informação acerca do cadastro realizado).

A3-) Constituição do Banco de Dados dos Indicadores de Desempenho de Vendas

A criação da BD de vendas foi orientada pelos objetivos de mercado, aos quais irão servir como elemento de suporte. O objetivo da utilização das BD de vendas consiste, dentre outros objetivos, na identificação dos serviços mais comercializados, gestão do desempenho de vendas por consultor, etc. A natureza das fontes de informação pode ser de três tipos distintos: interna, externa ou de documentos relativos a estudos de mercado e indicadores estatísticos (DROZDENKO et al., 2002).

- Obtenção de dados internos, presentes nos diversos sistemas da empresa ou de empresas pertencentes à mesma Empresa; dados internos mas registrados em suportes não tecnológicos e como tal sujeitos a recolha manual dos mesmos (em geral, formulários, reclamações, sugestões);
- Dados externos, provenientes de empresas do mesmo grupo organizacional ou

adquiridos a empresas externas;

- Dados recolhidos a partir de indicadores estatísticos ou estudos de mercado realizados.

Os dados foram obtidos de fontes internas da Empresa por meio dos sistemas já referenciados conforme elucidado pela Figura 3.3.

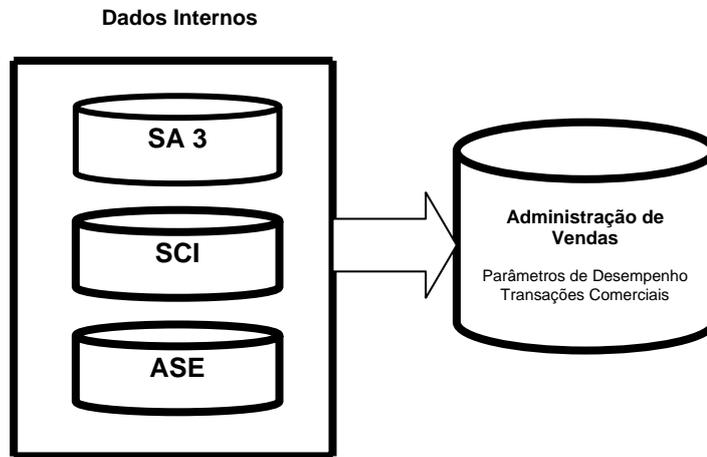


Figura 3.3 – Obtenção de informação de diversos sistemas.

Fonte: Elaborada pelo autor.

Uma vez recolhidos os dados, existe uma etapa intermédia, bastante importante e consumidora de muito do tempo dos pesquisadores, que consiste na seleção dos registros. Face à heterogeneidade das fontes de dados foi necessário proceder a uma uniformização dos mesmos, no sentido de evitar a duplicação de registros, incongruências, inconsistências e violações de domínio.

A definição de uma BD de vendas pode ser dividida em três fases. A primeira, já referida anteriormente, consiste na aquisição dos dados. A segunda refere-se à gestão e manutenção da própria BD, que irá servir de base para a realização da gestão comercial; finalmente, a terceira consiste na atualização das BD com base em rotinas de atualizações ou inserções de dados. As BD de vendas funcionam como armazém central de dados que servirá de base ao desenvolvimento de todas as ações de administração de vendas (SHEPARD, 1998), nomeadamente aquelas que envolvem a extração de conhecimento em BD.

A BD sobre a qual se desenvolveu todo o trabalho de pesquisa resultou da operacionalidade do sistema de administração de vendas. Os dados guardados na BD foram armazenados em tabelas distintas:

- *Parâmetros de Desempenho*: dados relativos às visitas realizadas, agendadas, informações da concorrência, ações comerciais, entre outras;
- *Transações Comerciais*: dados transacionais, relativos a informações sobre a emissão e cadastro de contratos dos serviços de telecomunicações comercializados.

Uma vez obtidos os dados, existe uma etapa intermédia, bastante importante e que é consumidora de muito do tempo, que consiste na seleção dos registros. Face à heterogeneidade das fontes de dados foi necessário proceder a uma uniformização dos mesmos, no sentido de evitar a duplicação de registros, incongruências, inconsistências e violações de domínio.

A.3.1-) Seleção das Informações Elegíveis

Face ao objetivo explícito de se conhecer o relacionamento dos indicadores de desempenho de vendas e em função da dimensão da BD, o projeto centrou-se exclusivamente nas informações que seriam de interesse na aplicação da Mineração de Dados voltada na busca do objetivo proposto. Desta maneira, com as informações relacionadas abaixo é possível reduzir substancialmente as informações necessárias, as quais são relacionadas: Agente Autorizado; Vendas Realizadas com Mapeamento da Concorrência; Vendas Realizadas; Contratos Cadastrados no Prazo; Contratos Cadastrados; Contratos Devolvidos; Ordem de Serviço Aberta; Ordem de Serviço Aberta no Prazo; Ordem de Serviço Aberta Cancelada; Consultores Cadastrados; Dias Úteis; Visitas Realizadas; Visitas Agendadas; Visitas Realizadas e Agendadas; Visitas Realizadas com Oportunidades Identificadas; Contratos Fechados; Consultores com Vendas Realizadas em Toda a Cesta; Nome dos Consultores com Desempenho de Vendas Acima de R\$ 3.000,00; Vendas Realizadas com Mapeamento de Oportunidades; Vendas Realizadas com Descontos; Viabilidades Aprovadas; Solicitações de Viabilidade e Viabilidades Aprovadas no Prazo.

B-) Extração de Conhecimentos

A preparação dos dados é uma etapa de grande importância para todo o processo de KDD que engloba a seleção dos dados, limpeza e transformação. Para o sucesso do processo de KDD é necessário que os dados tenham sido corretamente selecionados, corrigidos e transformados. Embora se tenha amenizado esse problema, a BD inicial apresentou problemas com os dados que tiveram que ser corrigidos nesta etapa. Assim, eliminou-se o processamento desnecessário do algoritmo de mineração de dados. Nesta etapa são estudadas e aplicadas as estratégias para tratamento de dados incorretos, além de alternativas para tratar os registros nulos.

O processo de KDD no âmbito desse projeto desenvolveu-se segundo o processo da metodologia proposta (Figura 24), no qual existem as fases de análise de dados (avaliação da

qualidade); Pré-processamento; Modelagem e finalmente avaliação dos modelos obtidos. O objetivo da KDD no âmbito da sua aplicação em projetos para ser utilizado em Administração de Vendas é transformar dados em resultados práticos permitindo numa fase seguinte atuar com a informação obtida. As fases que compõem a KDD são expostas em seguida.

B.1-) Análise e Seleção dos Dados

A análise dos dados disponíveis torna-se uma fase relevante que irá permitir ao pesquisador a compreensão dos dados, por meio da sua análise e descrição, assim como a seleção dos dados sobre os quais irá desenvolver todo o trabalho de extração de conhecimento. Como já referido anteriormente o conjunto de dados inicial era significativamente maior tanto em número de registros como na quantidade de atributos.

Embora o conjunto de dados utilizado em cada modelo seja relativamente pequeno face à dimensão da BD, foram utilizadas as informações mais relevantes para o desenvolvimento desse trabalho, quais sejam: os indicadores de desempenho de vendas que garantirá mesmo com um número relativamente pequeno de casos, desde que estejam, de alguma forma, garantido que a amostra é representativa do universo a estudar, é possível induzir um padrão genérico de relacionamento das variáveis estudadas.

A preparação de dados envolve as tarefas de seleção e transformação dos dados. Os tipos de dados selecionados podem estar organizados ao longo de múltiplas tabelas. Durante a seleção de dados, o usuário pode necessitar executar junções de tabelas ou eliminar linhas e/ou colunas de tabelas. Os métodos de transformação incluem organizar dados na forma desejada e converter um tipo de dado em outro tipo. Os dados selecionados foram armazenados em um DW, *data mart* ou *data set*. Para facilitar a realização desta fase, deve ser mantido um catálogo sobre as fontes de dados e sobre o que está no DW, *data mart* e *data set* no *meta dados*. Os *meta dados* são dados sobre as estruturas contidas em banco de dados. O *meta dados* facilita o entendimento sobre o conteúdo e a estrutura de um DW, bem como a história das modificações realizadas.

A realização das tarefas desse passo exigiu conhecimento dos dados operacionais e de seus relacionamentos, disponibilidade de tempo do pesquisador e alguns cuidados na escolha de subconjuntos de atributos e de dados. BECHER et al (2000) apresentam uma abordagem para a automatização desse passo e HSU et al (2000) propõem o uso de um sistema semi-automático de limpeza de dados. PYLE (1999) descreve uma metodologia de preparação de

dados. A descrição completa dos atributos constantes na BD cedida pela empresa promotora do projeto encontra-se descrito no item *Derivação de Novas Variáveis*.

Além dos métodos enunciados existiu ainda a sensibilidade do pesquisador, segundo a sua experiência e com conhecimento adquirido ao longo do processo conferiu-lhe uma capacidade de prospectar o tipo e a quantidade de dados que melhor se adapta a cada situação. Embora pese o esforço desenvolvido na filtragem e limpeza de dados, expresso anteriormente, a BD apresentava alguns problemas, principalmente em termos de dados relativos ao processo de entrada de dados, como sejam os dados omissos ou incorretos, os quais tiveram que ser tratados (processo descrito nas secções seguintes). A ocorrência desse tipo de problema justifica-se basicamente pela seguinte razão fundamental: o processo de entrada dos dados das vendas realizadas é manual, proporcionando a inserção de valores incorretamente ou a falha na interpretação dos dados inscritos manualmente pela área de pós venda.

Durante a fase de avaliação dos dados internos disponíveis, foram conhecidos os bancos de dados disponíveis para esse estudo de caso e os dados armazenados. Para o sucesso da aplicação, nesta fase devem ser levados em consideração os critérios técnicos definidos na definição dos objetivos, para que o conjunto de dados resultante esteja apropriado ao restante do processo.

A quantidade de dados é um fator de extrema importância. Para uma solução satisfatória da descoberta de conhecimento é necessária uma amostragem de casos do problema determinado, ou seja, devem existir exemplos que validem o modelo para o efetivo funcionamento do processo. Além do volume de dados, o conjunto de dados selecionado deve conter apenas os dados relevantes ao processo de KDD, ou seja, aqueles que determinam o padrão de comportamento. Nesse trabalho, os dados relevantes são aqueles referentes aos parâmetros dos indicadores de desempenho comercial. Para isso foram utilizados os dados do sistema do SA-3, ASE e SCI.

Para esse estudo, foram disponibilizados os dados do SA-3, ASE e SCI migrados para o formato do banco de dados Microsoft Access, além de outros documentos com dados relevantes ao processo. Contudo, quando se utilizam bases de dados distintas é necessário que os dados sejam migrados para um mesmo tipo de banco de dados ou que os algoritmos de extração estejam preparados para trabalhar com bancos de dados heterogêneos.

Além disso, em virtude do preenchimento dos sistemas ser mensal, os dados se encontravam em arquivos separadas por mês de competência. Foram disponibilizados dados de 2006. Com a finalidade de padronizar o armazenamento e manipulação dos dados foram utilizados todos os dados armazenados no banco de dados Microsoft Access. Após esta etapa,

se seleciona os parâmetros necessários para a confecção dos indicadores de desempenho de vendas conforme já elucidado no item A.3.1 – Seleção das Informações Elegíveis.

Os dados utilizados nesse estudo foram obtidos nos sistemas SA-3, ASE e SCI das empresas que tem como base de elaboração dados dos indicadores de desempenho e produtividade de vendas oficiais da Empresa. Os dados tratados se referem ao período compreendido entre janeiro de 2006 e dezembro de 2006. Todos os dados relativos aos indicadores de desempenho e produtividade de vendas foram obtidos em bases mensais, no mesmo período referenciado.

Por último, ressalte-se que a etapa de coleta de dados teve como objetivo a coleta de dados sobre produtividade de vendas e indicadores de desempenho aplicados ao mercado brasileiro de telecomunicações, mesmo porque, conforme comentado no Capítulo 1 desse trabalho, a principal virtude foi aglutinar esses parâmetros de diversas empresas geograficamente distribuídas pelo Brasil.

B.2-) Pré-processamento e Transformação de Dados

O pré-processamento e a transformação de dados são tidos na literatura como uma fase que envolve uma grande quantidade de conhecimento do domínio e de tempo. Muitos pesquisadores têm citado o fato dos dados obtidos diretamente a partir do BD serem de má qualidade, ou seja, possuírem informações incorretas e imprecisas, além de uma grande quantidade de valores desconhecidos. Embora pese o fato de muitos dos algoritmos utilizados na fase de modelagem terem sido projetados para manipular dados em tais situações, é natural que esses mesmos algoritmos gerem resultados mais precisos caso a maioria dos problemas presentes nos dados tenham sido removidos ou corrigidos.

A Figura 3.4 esquematiza e distingue as atividades relacionadas com o pré-processamento de dados das atividades relacionadas com a transformação de dados.

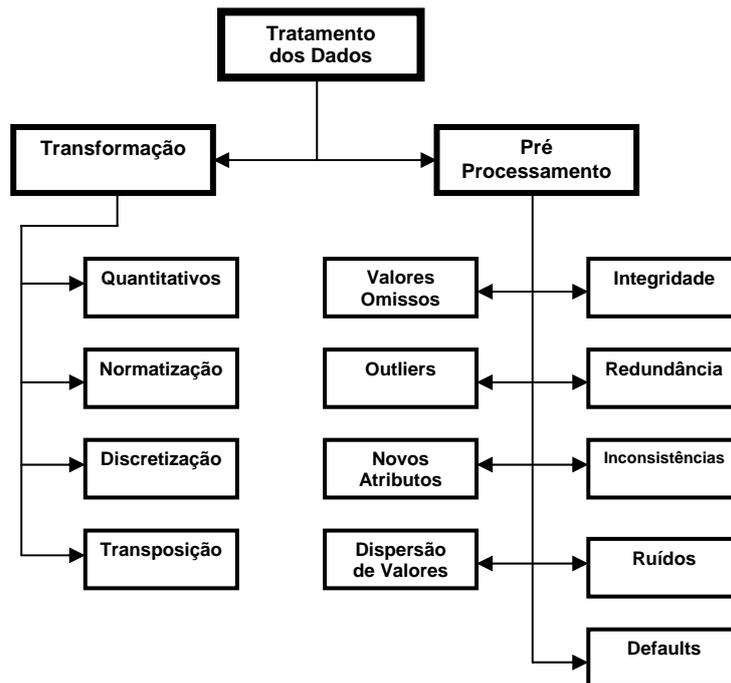


Figura 3.4 – Organização das atividades de pré-processamento e transformação de dados.

Fonte: Elaborada pelo autor.

Verifica-se que cada empresa tem a sua particularidade na aplicação do método, devido aos fatores da indústria e do tipo de modelo de negócio que ela utiliza. Cada empresa possui diferentes dados, que são críticos, pois são armazenados em bancos de dados transacionais e isso afeta a definição do modelo conceitual a ser usado no método proposto. Conhecendo-se a estrutura informacional da empresa, foi detalhada a busca do conhecimento no SMD e no sistema de informação. Os aspectos mais importantes analisados na estrutura do SMD são: classificação, busca de padrões nas categorias e disseminação que envolve análise e disseminação dos dados e informações, pois à medida que se conhecem quais são as medidas de desempenho e como elas se interagem foi possível estabelecer um modelo conceitual com os relacionamentos entre as medidas de desempenho.

O modelo conceitual é criado a partir das necessidades dos tomadores de decisão da empresa, ou seja, de quais informações são importantes para eles analisarem no que tange o estabelecimento dos relacionamentos entre as medidas de desempenho. O modelo conceitual é construído a partir da análise do SMD existente na empresa. A análise do SMD focou nos seguintes aspectos:

- Suporte das medidas de desempenho, prioridades de desempenho da empresa ou processos de negócio;
- O alinhamento das medidas de desempenho com as estratégias da empresa;

- Como as medidas de desempenho estão relacionadas aos processos de negócio;
- As medidas de desempenho que são relevantes e irrelevantes para a questão definida pelos tomadores de decisão;
- Os tipos de relacionamentos entre as medidas de desempenho existentes e seus posicionamentos nas áreas de construção desses relacionamentos.

Em relação ao sistema de informação da empresa, é necessário conhecê-lo, principalmente o que se relaciona à decisão existente na empresa, de modo a verificar quais são os bancos de dados transacionais e como eles são utilizados pelos tomadores de decisão. Esse entendimento auxiliou na utilização das fontes de dados possíveis para o uso das técnicas e ferramentas de Mineração de Dados.

De um modo geral, o pré-processamento dos dados é um processo semi-automático, dependendo da capacidade do pesquisador em identificar os problemas presentes nos dados e de selecionar os métodos mais apropriados para solucionar cada um dos casos. A transformação de dados consiste em trabalhar a representação dos dados a fim de superar quaisquer limitações existentes nos algoritmos utilizados na modelagem. As atividades mais comuns de pré-processamento e transformação de dados foram referenciadas no Capítulo 2.

A fase de pré-processamento e transformação de dados é como já foi referido anteriormente, responsável por grande parte do tempo consumido num projeto de extração de conhecimento em BD (CABENA et al., 1998). Uma vez obtido a BD isenta de todos os registros considerados como inválidos, inicia-se então o trabalho sobre os dados. Esta fase correspondeu ao trabalho de eliminação de ruído, erros e omissões que pudessem de algum modo comprometer o trabalho de modelagem dos dados. Uma vez limpos e pré-processados os dados, foi possível constituir a base de dados de vendas, onde foram incluídas, todas as informações pertinentes à elaboração dos parâmetros de vendas.

Qualquer banco de dados pode conter vários problemas de qualidade nos dados (poluição). Por isso, para o perfeito funcionamento do processo de KDD, é necessário assegurar que os dados utilizados no processo estejam corretos. A etapa de seleção resultou num conjunto dos dados relevantes para a aplicação das técnicas de mineração de dados. Porém, como esses dados são manipulados no dia-a-dia operacional da Instituição, eles podem apresentar problemas que poderão acarretar em atrasos, retornos de etapas e até mesmo a falha do processo de KDD. Por isso, tornou-se necessária a realização desta etapa de limpeza e correção dos dados durante todo o processo. Os procedimentos de pré-processamento desenvolvidos foram concretizando-se nomeadamente nos seguintes aspectos:

Tratamento de Dados Omissos

A existência de valores em branco num determinado atributo suscita o tratamento desse atributo, com duas opções distintas, mas viáveis: eliminação do atributo da BD (no caso de valor omissos da maioria dos registros) ou processamento do atributo, isoladamente ou em função de outros. Relativamente ao último caso, o preenchimento do atributo em registros em que esteja omissos foi considerado nesse trabalho com a seguinte abordagem: preenchimento dos casos omissos com valores equivalentes à média desse atributo em todos os registros de uma determinada condição do serviço. Isto ocorreu poucas vezes na segmentação do serviço comercializado bem como, na entrada dos dados das visitas realizadas. Entretanto, esses fatos não foram relevantes na constituição da BD final dos parâmetros de venda, pois, na derivação de novas variáveis, essas informações não faziam parte da mesma.

Tratamento de Exceções (Outliers)

Verificou-se com pouca frequência a ocorrência de valores anormais para alguns atributos, em alguns registros. Ao contrário do processo de tratamento de dados omissos, os atributos encontravam-se preenchidos, mas com a possibilidade de não corresponderem à realidade. Esta fase procurou recuperar a integridade do registro pelo recurso do tratamento de exceções. As exceções poderiam ser abordadas de dois modos:

- Seleção dos registros com valores excepcionais e conseqüente tratamento manual independente;
- Substituição do valor considerado como anormal pela média desse atributo em todos os outros registros com dados semelhantes em outros atributos (à semelhança do tratamento de dados omissos).

Foi verificado já na fase de consolidação da BD das variáveis a serem estudadas, a existência de cinco *outliers* na variável independente Volume de Vendas Com Descontos, sendo dois da empresa *NJ* e dois da empresa *SG* para além da cerca inferior e um da empresa *SA* para além da cerca superior e três na variável dependente Produtividade de Vendas, sendo todos além da cerca superior com duas participações da empresa *NJ* e um da empresa *NC*, são descritas mais adiante. As outras informações que não faziam parte da constituição dos parâmetros de desempenho de vendas não foi feita uma análise criteriosa, pois eram informações irrelevantes para o objetivo da Mineração de Dados.

Verificação da Coerência da Informação do Registro

A verificação da coerência da informação é de todas as atividades de pré-processamento a mais morosa, dada a dependência direta da avaliação subjetiva do pesquisador em cada situação (registro). Não foi verificada nenhuma situação desse tipo que pudesse comprometer a BD das variáveis a serem estudadas.

Atributos com Valores Dispersos

Esse problema é comum ocorrer nos questionários que tem questões abertas, sem a imposição de qualquer limite máximo para o valor da resposta ou padronização de valores. Esse fato provoca uma dispersão das respostas que acabam por vir a condicionar a sua análise. Como os dados não são oriundos desta situação, não foi verificado esse tipo de problema.

Transposição de Tabelas

A aplicação de muitos algoritmos e ferramentas requer que os dados estejam todos numa única tabela ou, pelo menos em tabelas distintas, mas com a mesma chave primária. Verificou-se na BD em estudo quer a existência de mais do que uma tabela para a mesma entidade (em geral, tabela serviços comercializados e tabela serviços mapeados da concorrência) quer a existência de índices (chaves primárias) heterogêneos impossibilitando a integração e cruzamento de todos os dados numa tabela única. A transposição de tabelas passou então pela homogeneização de atributos (obtida pela integração de todos os dados numa só tabela ou pela utilização de um único atributo como chave primária em todas as tabelas que irão ser utilizadas na modelagem). Após esta operação, não foi verificado nenhum dado inconsistente.

Normalização de Valores

Todos os dados usados nesse trabalho (em bases mensais, no período de janeiro de 2006 a dezembro de 2006) foram disponibilizados em valores originais de cada variável e descritos em detalhes mais a frente. Entretanto os mesmos foram padronizados, cujos valores foram usados na tentativa de se conseguir maiores precisão e significância nos resultados além da sua confidencialidade. Considerando que as empresas analisadas têm tamanhos diferentes e

mercados potencialmente heterogêneos, portanto com volume de vendas bastante diferenciadas, houve necessidade de se padronizar os valores observados. Para tanto, foi adotado o processo nos quais os dados originais foram transformados em novas variáveis com média zero e desvio padrão um (LAROSE., 2006). Isto nos permitirá comparar diretamente o efeito relativo de cada variável independente sobre a variável dependente, facilitando sobremaneira a interpretação dos dados a serem analisados, além de reservar o caráter de confidencialidade dos dados comerciais.

Derivação de Novas Variáveis

Isoladamente ou na sua forma original, nem sempre os dados possuem caráter informativo, havendo por isso necessidade de efetuar cálculos posteriores com vista à exposição da informação associada. Tomando em consideração os requisitos e os dados necessários para a aplicação dos algoritmos na fase de modelagem, a derivação de novas variáveis consistiu na criação de índices de classificação que permita, de algum modo, facilitar a seleção de registros.

Os índices foram segmentados nos cinco atributos do BSC, conforme já descrito no item Estudo de Negócio. As descrições das variáveis que foram utilizadas são as seguintes:

CO₁ Venda x Mapeamento da Concorrência

Medir o percentual de efetividade do mapeamento da concorrência realizado pela empresa na visão por grupo de produtos. Esse indicador mede o sucesso do alinhamento da equipe de vendas com os objetivos estratégicos da empresa controladora buscando liderança no mercado Empresarial por meio de estratégias de ataque à base da concorrência. É a relação entre o Total de vendas com mapeamento da concorrência – Quantidade de clientes com vendas (visão por CNPJ com ordem de serviço ou pedido de venda e que tiveram algum tipo de mapeamento da concorrência. A análise é feita levando em conta a visão de grupo de produtos: mobilidade, voz avançada e comunicação de dados) e o Total de vendas – Quantidade de clientes com vendas (visão por CNPJ com ordem de serviço ou pedido de venda).

EO₁ Assertividade de Entrega e Cadastro de Contratos

Avaliar a eficiência do time de vendas em cadastrar os contratos em prazo inferior ao proposto. Não é considerado para o indicador o tempo médio que os contratos levaram para serem entregues e cadastrados, apenas se o foram dentro do prazo ou não. É a relação

percentual entre o Total de contratos cadastrados no prazo – Quantidade de contratos que foram corretamente cadastrados em até 07 dias corridos, no máximo, após a assinatura do contrato pelo cliente, incluindo o dia do cadastro. Exceção feita a produtos que não necessitam contratos e o Total de contratos cadastrados – Quantidade de contratos no sistema, feitos pelas regionais durante o mês, independentemente o prazo de 07 dias corridos.

EO4 Retorno de Contratos

Avaliar a quantidade de contratos devolvidos que impeçam a abertura da ordem de serviço. É a relação percentual entre o Total de contratos devolvidos pela GGC – Quantidade de contratos que não tiveram ordem de serviço ou pedido de venda aberto devido a pendências internas, ou seja, excluem-se motivos internos da empresa controladora e o Total de contratos cadastrados no sistema – Quantidade de inputs de contratos no sistema, feitos pelas regionais durante o mês, independentemente do prazo de 07 dias corridos.

FC1 Média de Visitas por Consultor

Avaliar desvios da relação diária de visita e número de clientes da carteira por consultor AA. Total de consultores – É a relação percentual entre o Total de consultores cadastrados no sistema de controle e o Total de dias úteis – Total de dias úteis do mês de apuração do indicador. O calendário referencia leva em consideração somente os feriados nacionais e a semana de trabalho de segunda a sexta-feira.

FC2 Assertividade de Agendamento

Avaliar a assertividade das visitas agendadas pelo tele marketing (o agendamento deve acontecer com até 72hs de antecedência). É a relação percentual entre o Total de visitas realizadas de acordo com o agendamento – Quantidade de clientes visitados com espaçamento entre o agendamento e a visita de 72 horas úteis, no mínimo. Para efeito de cálculo do indicador são consideradas visitas e revisitas e o Total de visitas agendadas - Quantidade de clientes visitados com agendamento prévio. Para efeito de cálculo do indicador são consideradas visitas e revisitas.

FC3 Visitação X Agendamento

Medir o grau de participação do processo de agendamento prévio no universo de visitas realizadas pela empresa. É a relação percentual entre o Total de visitas agendadas - Quantidade de clientes visitados com agendamento prévio. Para efeito de cálculo do indicador são

consideradas visitas e revisitas e o Total de visitas – Total de clientes visitados pelos consultores da empresa no mês e apurados no sistema de controle. Para efeito de cálculo do indicador são consideradas visitas e revisitas.

GP1 Produtividade de Visita

Medir a eficiência da equipe de consultores em identificar oportunidades em todas as visitas (visão por cliente). É a relação percentual entre o Total de visitas com oportunidades identificadas – Total de clientes visitados pelos consultores da empresa no mês e apurados no sistema de controle. Para efeito de cálculo do indicador são consideradas visitas e revisitas e se foram identificadas oportunidades, independente da quantidade de oportunidades. A análise é feita levando em conta a visão por cliente e o Total de visitas – Total de clientes visitados pelos consultores da empresa no mês e apurados no sistema de controle. Para efeito de cálculo do indicador são consideradas visitas e revisitas.

GP2 Eficácia de Visitas

Medir o percentual de contratos fechados em relação às visitas realizadas e gerir o funil de vendas do Agente Autorizado. É a relação percentual entre o total de contratos fechados – Total de contratos cadastrados no sistema, dentro dos critérios de vendas e elegíveis ao comissionamento (Ordem de serviço aberta, sem pendência; Móvel – Pedido lançado e aprovado; Ordem de serviço aberta; Longa Distância – cliente fidelizado e chaveado) e o Total de visitas – Total de clientes visitados pelos consultores da empresa no mês e apurados no sistema de controle. Para efeito de cálculo do indicador são consideradas visitas e revisitas.

GP3 Cesta de Produtos

Medir a capacidade da equipe de consultores em prospectar toda a cesta de produtos (Longa Distância, Voz Avançada, Mobilidade e Comunicação de Dados). É a relação percentual entre o Total de consultores com vendas realizadas em toda a cesta – Número de consultores da empresa, que apresentaram pelo menos uma venda (visão por CNPJ com ordem de serviço ou pedido de venda) em cada uma das famílias de produtos do portfólio (Longa Distância, Voz Avançada, Mobilidade e Comunicação de Dados). O foco nesse caso não é a quantidade nem o tipo de produto vendido e sim a ocorrência de vendas em todas as famílias de produtos e o Total de consultores – Quantidade de consultores da empresa cadastrados no sistema de controle, no mês de referência.

GP4 Distribuição de Vendas por Consultores

Medir a participação de consultores verdes e amarelos na equipe de vendas da empresa. É a relação percentual entre o Total de consultores com desempenho de vendas acima de R\$ 3.000,00 – Quantidade de consultores cadastrados no sistema de controle, com classificação verde ou amarela (modelo de atuação), no mês de referência, com base no resultado de vendas do consultor e o Total de consultores – Quantidade de consultores da empresa cadastrados no sistema de controle, no mês de referência.

GP5 Produtividade de Oportunidades

Medir a eficiência da equipe de consultores fecharem negócio nas oportunidades prospectadas. É a relação percentual entre o Total de vendas com mapeamento de oportunidades – Quantidade de clientes com vendas (visão por CNPJ com ordem de serviço ou pedido de venda) e que tiveram algum tipo de mapeamento da oportunidade e o Total de visitas com oportunidades identificadas – Total de clientes visitados pelos consultores da empresa no mês e apurados no sistema de controle. Para efeito de cálculo do indicador são consideradas visitas e revisitas e se foram identificadas oportunidades, independente da quantidade de oportunidades.

IA2 Volume de Vendas Com Descontos

Medir a participação de produtos customizados no volume de vendas (Visão Voz Avançada e Comunicação de Dados). É a relação percentual entre o total de vendas com descontos – Quantidade de clientes com vendas, com aplicação de política de desconto por alçada. Para efeito de cálculo do indicador, somente foram considerados os produtos pertencentes aos grupos Voz Avançada e Comunicação de Dados e o Total de vendas – Quantidade de clientes com vendas (visão por CNPJ com ordem de serviço). Para efeito de cálculo do indicador, somente se considera os produtos pertencentes aos grupos Voz Avançada e Comunicação de Dados.

IA3 Aprovação de Viabilidades

Medir o grau de aprovação de viabilidades em comparação à quantidade total de pedidos. É a relação percentual entre o Total de viabilidades aprovadas – Quantidade de clientes com pedidos de análise de viabilidade aprovados (com data de aprovação cadastrada) no sistema de controle, dentro do mês de referência e o Total de solicitações de viabilidade -

Quantidade de clientes com pedidos de análise de viabilidade cadastrados (qualquer Status) no sistema de controle, dentro do mês de referência.

IA4 Assertividade de Analise de Viabilidade

Avaliar a capacidade da equipe de engenharia comercial em analisar viabilidades dentro do prazo proposto (08 dias corridos entre a data da solicitação e o retorno da aprovação, no máximo). É a relação percentual entre o Total de viabilidades aprovadas no prazo – Quantidade de clientes com pedidos de análise de viabilidade aprovados (com data de aprovação cadastrada) no sistema de controle, dentro do mês de referência e no prazo de 08 dias corridos entre a data da solicitação e a data de aprovação do estudo, inclusive e o Total de viabilidades aprovadas – Quantidade de clientes com pedidos de análise de viabilidade aprovados (com data de aprovação cadastrada) no sistema de controle, dentro do mês de referência.

Em relação às medidas de desempenho disponibilizadas, esta pesquisa desenvolveu-se com a utilização de 17 variáveis descritas anteriormente, as quais fazem parte da sistemática de acompanhamento de desempenho de vendas das empresas analisadas, relativas à composição das variáveis independentes e ainda uma variável dependente, como mostrado no Anexo I. Para aplicar o procedimento de regressão, é selecionada a Produtividade de Vendas (REFX) como a variável dependente (efeito presumido ou resposta a uma mudança nas variáveis independentes) a ser explicada pelas variáveis independentes (causa presumida de qualquer mudança na variável dependente) que representam os indicadores de desempenho das empresas de telecomunicações analisadas. As 17 variáveis a seguir descritas são incluídas como independentes: *CO1* Venda x Mapeamento da Concorrência; *EO1* Assertividade de Entrega e Cadastro de Contratos; *EO4* Retorno de Contratos; *EO5* Abertura de Ordem de Serviço; *EO6* Conversão de Contratos; *EO7* Cancelamento de Vendas; *FC1* Média de Visitas por Consultor; *FC2* Assertividade de Agendamento; *FC3* Visitação X Agendamento; *GP1* Produtividade de Visita; *GP2* Eficácia de Visitas; *GP3* Cesta de Produtos; *GP4* Distribuição de Vendas por Consultores; *GP5* Produtividade de Oportunidades; *IA2* Volume de Vendas Com Descontos; *IA3* Aprovação de Viabilidades; *IA4* Assertividade de Analise de Viabilidade.

B.3-) Modelagem

A fase de modelagem consiste na seleção de técnicas que permitam alcançar os objetivos propostos para a KDD e que devem encontrar-se enquadrados com os objetivos da Empresa e de vendas. Contudo a especificação dos objetivos de MD depende ainda da

especificidade da atividade de venda (definida no âmbito dos objetivos de venda) a que se encontra ligada.

A construção de um modelo é feita por meio de uma técnica de Mineração de Dados, tendo como base os dados transformados e o estudo definido no passo anterior. Um modelo resume grandes quantidades de dados por acumular indicadores. Alguns dos indicadores que vários modelos acumulam são: a) frequências: mostram em qual frequência que certo valor ocorre; b) pesos ou impactos: indicam a influência exercida por algumas entradas na ocorrência de uma saída; c) conjunções: algumas vezes certas entradas têm mais peso juntas do que separadas; d) diferenciação: indica a importância de uma entrada para uma determinada saída do que para outra saída.

Dependendo do tipo de modelo usado para ajustar os dados, existem diferentes formas de entendê-lo. Os indicadores que muitos modelos podem acumular, conforme descritos no passo anterior, podem influenciar no entendimento do modelo, além do tipo de técnica de Mineração de Dados aplicada na construção do modelo.

Especificamente, é utilizado o modelo de regressão linear múltipla que permitirá fazer análise da relação entre as variáveis quantitativas, nomeadamente, entre uma variável dependente (produtividade de vendas) e um conjunto de variáveis independentes (nesse estudo, relacionadas com os indicadores de desempenho de vendas). Com efeito, as variáveis utilizadas no modelo de regressão foram quantitativas com escalas de medida de razão.

Um dos aspectos mais importantes nesta análise é a seleção das variáveis independentes que possam ser utilizadas na regressão, que são aquelas que têm influência na formação da produtividade de vendas. As possíveis variáveis independentes (ou explicativas) que podem influenciar produtividade de vendas devem ser listadas a priori. A definição das variáveis explicativas preliminarmente, economiza tempo e diminui o custo de execução da pesquisa. As variáveis explicativas para a esta análise são aquelas referentes a todas as gerenciadas nos indicadores de desempenho. No entanto, dentre todas as variáveis independentes disponibilizadas, nem todas deverão ser relevantes à formação das vendas.

A-) Diagnóstico do Modelo

Por intermédio da análise dos gráficos de resíduos, prática esta indispensável, é possível a verificação da adequação do modelo, valores discrepantes e também a investigação da atuação das várias variáveis regressoras disponíveis (NETER et al., 1996).

- *Adequação do modelo*: a verificação de que o modelo de regressão ajustado é adequado

aos dados pode ser detectado por meio dos gráficos dos resíduos versus variáveis independentes ou resíduos versus valores preditos. Quando esses gráficos apresentarem algum tipo de padrão sistemático dos pontos, há indicação de que o modelo ajustado não é adequado.

- *Valores discrepantes*: os gráficos dos resíduos versus variáveis independentes ou resíduos versus valores ajustados mostram a presença de pontos discrepantes ou “*outliers*”. Considera-se um ponto discrepante quando é muito maior que o resto em valor absoluto e talvez se estiver afastado de zero por três ou mais desvios padrões (DRAPER & SMITH, 1981 e NETER & WASSERMAN, 1974). Um ponto discrepante pode ser retirado da análise se a causa de sua existência for devido a erro de medida ou de transcrição de valor. Caso contrário deve ser pesquisado e o pesquisador, conhecedor da natureza das variáveis, é quem deve dar a solução.
- *Omissão de variáveis independentes*: a investigação de variáveis independentes importantes não incluídas no modelo pode ser feita pelo gráfico dos resíduos versus cada uma delas. Considera-se que variável deva ser incluída no modelo quando os pontos apresentarem uma relação linear, (NETER et al., 1996).

Nota-se que a análise dos resíduos é de fundamental importância na seleção de um modelo de regressão. Esta análise dos resíduos permite a verificação da validade ou de possíveis violações das suposições do modelo adotado. A validade das suposições é essencial para a obtenção dos estimadores, construção dos intervalos de confiança e testes de hipóteses.

B-) Suposições em Análise de Regressão Múltipla

A regressão linear múltipla é usada para modelar o valor de uma variável dependente escalar baseada em seu relacionamento linear para um ou mais regressores. O modelo da regressão linear supõe que exista um relacionamento linear entre a variável dependente e cada um dos regressores. Para o propósito de testes das hipóteses sobre os valores dos parâmetros, o modelo da regressão linear supõe também o já elucidado no item 2.2.3.A do capítulo 2.

A linearidade da relação entre as variáveis dependente e independentes representa o grau em que a variação da variável dependente está associada à variável independente, expressando o conceito de que o modelo possui as propriedades de aditividade e homogeneidade, ou seja, os modelos lineares prevêm valores que estão sobre uma reta que tem uma taxa constante de variação (coeficiente angular) da variável dependente em relação a uma variação unitária constante na variável independente. O coeficiente de regressão deve ser

constante no domínio de valores da variável independente (MALHOTRA, 2004). O conceito de correlação é baseado em uma relação linear, o que torna uma questão crucial na análise de regressão. A linearidade pode ser examinada por meio de gráficos de resíduos. O método mais adequado para examinar relações bi variadas é o gráfico de dispersão que é um gráfico de pontos baseado em duas variáveis, onde uma variável define o eixo horizontal e a outra o eixo vertical. Uma forte organização de pontos ao longo de uma reta caracteriza uma relação linear ou correlação.

A relação mais comum entre duas variáveis fundamenta-se na dependência funcional de uma variável relativamente à outra. Duas variáveis dizem-se funcionalmente dependentes quando a magnitude de uma delas - variável dependente - é governada pela magnitude da outra - variável independente -, enquanto que a recíproca não é necessariamente verdade (LAROSE, 2006). Dependência não significa, portanto, causalidade. Um gráfico de dispersão é uma ferramenta útil na detecção do tipo de função que está por detrás da estrutura de dependência entre duas variáveis.

A presença de variâncias desiguais (heteroscedasticidade) é uma das violações mais comuns de suposições. Esse diagnóstico também é feito com gráficos de resíduos. Para avaliar o comportamento geral dos resíduos, constrói-se um gráfico de resíduos, contendo no eixo vertical os resíduos padronizados (*standardized residual*) e no eixo horizontal os valores previstos padronizados (*standardized predicted value*). Habitualmente também se constroem diagrama de dispersão para visualizar a relação entre a variável dependente e cada um dos regressores individualmente bem como, constrói-se gráfico de dispersão que apresente os valores observados da variável dependente com os valores preditores (*standardized predicted value*) (HAIR, J. F. et al., 2005).

Como é trabalhado com regressão múltipla, haverá necessidade de se analisar cada variável independente com a dependente, pois no exame dos resíduos são mostrados os efeitos combinados de todas as variáveis independentes. Para tanto, é utilizado os gráficos de regressão parcial que satisfaz a condição acima referenciada. Em gráficos de regressão parcial, um padrão curvilíneo de resíduos indica uma relação não linear entre uma variável independente específica e a variável dependente (HAIR, J. F. et al., 2005). Esse método é muito útil já que se pode dizer quais variáveis independentes específicas violam a suposição de linearidade além de se identificar eventuais observações atípicas.

A independência dos termos de erro pode ser identificada fazendo o gráfico de resíduos em relação a qualquer variável seqüencial possível. Se os resíduos forem independentes, o padrão é aleatório e semelhante ao gráfico nulo de resíduos, pois é assumido em regressão que

cada valor previsto é independente, ou seja, não está relacionado com qualquer outra previsão, não sendo seqüenciados por qualquer variável. A homoscedasticidade é uma suposição relacionada primariamente a relações de dependência entre variáveis, referindo-se à suposição de que a variável dependente exibe níveis iguais de variância ao longo do domínio da variável preditora. É desejável a existência de homoscedasticidade porque a variância da variável dependente sendo explicada na relação de dependência não deve se concentrar apenas em um domínio limitado dos valores independentes (HAIR, J. F. et al., 2005).

O diagnóstico mais simples para verificação de normalidade da distribuição dos termos de erros é um histograma de resíduos, com uma verificação visual para uma distribuição que se aproxima da normal. Um histograma é uma representação gráfica de uma única variável que representa a frequência (valores dos dados) dentro da categoria de dados. O histograma é um gráfico parecido com o de barras, só que o eixo horizontal é escalar contínuo, representando classes de uma variável numérica e não categorias como no caso do gráfico de barras. As barras representam as frequências absolutas registradas no intervalo ou classe definido no referido eixo.

Outra abordagem muito utilizada é o gráfico de probabilidade normal, que compara a distribuição cumulativa de dados reais com a distribuição cumulativa de uma distribuição normal. A distribuição normal forma uma reta diagonal e os dados do gráfico são comparados com a diagonal. Se uma distribuição é normal, a linha que representa os dados segue muito próximo à diagonal. Os gráficos de probabilidades (P-P: *Probability Plots*) visualizam graficamente o ajustamento de uma variável a uma função de distribuição de probabilidades (DANIEL & WOOD, 1971).

Por último, as observações influentes é a categoria mais ampla, incluindo todas as observações que têm um efeito desproporcional sobre os resultados de regressão. Salienta-se que essas observações não são necessariamente inadequadas, no sentido de que devam ser omitidas, pois em muitos casos, elas representam os elementos distintos do conjunto de dados, podendo ter sua ocorrência em função de uma das quatro condições a seguir descritas. Erro em observação ou entrada de dados, uma observação válida, porém explicável por uma situação extraordinária, uma observação excepcional sem explicação convincente ou uma observação comum em suas características individuais, porém excepcional em sua combinação de características (DRAPER & SMITH, 1981 e NETER & WASSERMAN, 1974).

Um gráfico *boxplot* (ver gráfico 4.5) pode ser visto como a representação gráfica de algumas medidas de localização e de dispersão já tratadas. A construção do *boxplot* é baseada na mediana e nos quartis. Como tais medidas são robustas ou resistentes ao impacto de valores

muito elevados ou muito pequenos, resulta que o *boxplot* é igualmente resistente a influência proveniente de valores extremos. A linha central da caixa corresponde à mediana da amostra. Os extremos inferiores e superiores da caixa correspondem aos primeiro e terceiro quartis, respectivamente. A cerca superior termina no valor máximo da amostra e a cerca inferior termina no valor mínimo da amostra. Assim, o *boxplot* espelha de forma simples a estrutura da população subjacente aos dados e nele é possível identificar as seguintes características evidenciadas pelos dados:

- Localização - indicada pela mediana;
- Dispersão - resultante do comprimento da caixa (distância interquartil dF) e do comprimento total entre os extremos dos bigodes (amplitude R);
- Comprimento das caudas da distribuição subjacente - dado pelo comprimento das linhas que definem as caixas.

Pode ser complementada esta análise visual com estatísticas que reflitam a forma de distribuição (assimetria que corresponde graficamente a um arco simples acima da diagonal, chamada de positiva ou abaixo da diagonal, chamada de negativa enquanto curtose se refere a uma elevação ou um achatamento da distribuição, ambas comparadas com a distribuição normal) bem como um teste estatístico de normalidade que é o teste modificado de *Kolmogorov-Smirnov* (JOANES & GILL,1998). O princípio desse teste baseia-se na comparação da curva da frequência cumulativa dos dados, com a função de distribuição teórica em hipótese. Quando as duas curvas se sobrepõem à estatística de teste é calculada por meio da máxima diferença entre ambas. A magnitude da diferença estabelece-se probabilisticamente, segundo a lei de probabilidade dessa estatística, que se encontra tabelada. Se os dados experimentais se afastam significativamente do que é esperado segundo a distribuição em hipótese, então as curvas obtidas devem encontrar-se igualmente afastadas, e por um raciocínio análogo, se o ajustamento ao modelo hipotético é admissível, as curvas têm um delineamento próximo.

B.4-) Avaliação dos Resultados

Uma vez aplicado e desenvolvido o conceito teórico para a indexação dos indicadores de desempenho de vendas, o estudo prosseguiu de acordo com a metodologia de suporte, a MDAV, orientando-se agora para a aplicação dos algoritmos de MD. Tratando-se de um projeto cujo objetivo de negócio é conhecer melhor os relacionamentos entre os indicadores de desempenho de vendas, a transposição desse para a administração de vendas e por

conseqüência para os objetivos da gestão de negócios em telecomunicações, proporcionando o estabelecimento de diferentes objetivos de MD. Assim, o objetivo de MD definido é verificar se, além da existência do relacionamento entre as variáveis anteriormente referenciadas, se confirma a relação definida pelos especialistas de vendas em serviços de telecomunicações.

Nesta etapa, é feito a estimação do modelo de regressão e a avaliação da precisão preditiva geral das variáveis independentes por meio da seleção de um método para especificar o modelo de regressão a ser ajustados, avaliando a significância estatística do modelo geral na previsão da variável dependente e determinando se alguma das observações tem influência indevida nos resultados. Quando se ajusta um modelo de regressão múltipla, pode acontecer que se justifique estatisticamente incluir na equação de regressão todas as variáveis independentes, ou que se incluam apenas algumas dessas variáveis explanatórias. Esta decisão é tomada em função da significância do parâmetro de regressão de cada uma das variáveis, ou pelo acréscimo do coeficiente de determinação, R^2 , provocado pela inclusão dessas variáveis.

Um dos problemas mais difíceis e freqüentes em análise de regressão é a seleção do conjunto de variáveis independentes para serem incluídas no modelo (NETER & WASSERMAN, 1974). O pesquisador deve especificar o conjunto de variáveis independentes a ser empregado para descrever, controlar ou prever a variável dependente. Um problema muito difícil de relacionamento que aparece na seleção de variáveis é quando uma equação de regressão é construída com o objetivo de previsão e envolve muitas variáveis. Talvez, muitas delas, contribuam pouco ou nada para precisão da previsão. A escolha apropriada de algumas delas fornece a melhor previsão, porém o problema é quantas e quais variáveis selecionar.

Tipicamente, o número de variáveis independentes que permanece, após esta seleção inicial, ainda é grande. Posteriormente, muitas dessas variáveis estarão altamente intercorrelacionadas. Portanto, o pesquisador geralmente desejará reduzir o número de variáveis independentes a serem usadas no modelo final. Existem várias razões para isto. Um modelo de regressão com um número grande de variáveis independentes é caro para se utilizar. Dessa forma, modelos de regressão com um número limitado de variáveis independentes são fáceis para se avaliar e estudar. Finalmente, a presença de muitas variáveis independentes altamente intercorrelacionadas, podem adicionar pouco ao poder de previsão do modelo, enquanto retira suas habilidades descritivas e aumenta os erros de previsão. O problema então é como reduzir a lista de variáveis independentes de forma a obter a melhor seleção de variáveis independentes. Esse conjunto precisa ser suficientemente pequeno para que a manutenção dos custos de atualização do modelo seja manuseável e a análise facilitada, e

ainda, deve ser grande o suficiente de forma que seja possível uma descrição, um controle e uma predição adequados.

Muitas vezes quando se dispõem de observações acompanhadas de vários valores correspondentes a vários regressores coloca-se em questão se todos esses são importantes para o modelo. Pode acontecer que um ou mais regressores não contribuam para um melhor ajuste do modelo e como tal não vale a pena considerá-los. Deve ser considerado que a inclusão de regressores faz diminuir o número de graus de liberdade de várias estatísticas e como tal aumenta a incerteza de alguns dos resultados. Chama-se a atenção que, por vezes, a inclusão de muitas variáveis explanatórias no modelo, pouco acréscimo trazem em termos de precisão do modelo, e esse fica muito pouco funcional e muito fictício ou artificial, isto é, muito bem ajustado à amostra particular de valores usados para o cálculo da equação, mas pouco adaptável realidade do problema (WEISBERG, S, 1985).

As técnicas apresentadas a seguir são úteis para confirmar ou afastar hipóteses acerca da inclusão de variáveis explicativas, produzidas a partir do conhecimento do sistema e de seu comportamento. Esta seleção primária de variáveis é, portanto, fundamental para o eventual sucesso da regressão. Existem diversos procedimentos para selecionar um conjunto de variáveis (regressores) tidas como fundamentais ou mais importantes em cada problema.

Os procedimentos de procura para se encontrar o melhor conjunto de variáveis independentes devem ser empregados após o pesquisador ter estabelecido a forma funcional da relação de regressão, ou seja, se as variáveis dadas estão na forma linear, quadrática, etc.; se as variáveis independentes são primeiramente transformadas como, por exemplo, por transformação logarítmica; e se algum termo de interação foi incluído. Nesse ponto, os procedimentos de procura são empregados para reduzir o número de variáveis independentes. Existem muitos procedimentos de seleção, mas nenhum deles pode, comprovadamente, produzir o melhor conjunto de variáveis independentes. Não existe um conjunto ótimo de variáveis independentes, pois o processo de seleção das variáveis possui julgamentos subjetivos. Dentre os procedimentos, pode-se citar como os mais comumente usados: todas as regressões possíveis, *backward*, *forward*, *stepwise* e confirmatória, conforme abaixo:

- *Todas as regressões possíveis*: esse procedimento consiste em ajustar todas as possíveis equações de regressão. Após a obtenção de todas as regressões, devem-se utilizar os critérios para comparação dos modelos ajustados. Alguns critérios que podem ser usados são o R^2 (coeficiente de determinação), MSE (erro médio quadrático) e C_p (estatística de Mallows). Para alguns conjuntos de variáveis, os três critérios podem levar para o mesmo “melhor” conjunto de variáveis independentes.

Esse não é o caso geral, pois diferentes critérios podem sugerir diferentes conjuntos de variáveis independentes. (DANIEL & WOOD, 1971) recomendam, no caso de um grande número de equações alternativas, o critério do erro quadrado total para caracterizar a equação. A principal desvantagem do procedimento de procura de todas as regressões possíveis é a quantidade de esforço computacional necessária. Já que cada variável independente potencial pode ser incluída ou excluída, gerando $(2^p - 1)$ regressões possíveis quando existem p variáveis independentes potenciais (DRAPER & SMITH, 1981);

- *Seleção stepwise (passo a passo)*: é, provavelmente, o mais amplamente usado dos métodos de pesquisa que não requerem a computação de todas as regressões possíveis. Ele foi desenvolvido para economizar esforços computacionais, quando comparado com a abordagem de todas as regressões possíveis, enquanto atinge um conjunto de variáveis independentes razoavelmente bons. Essencialmente, esse método de pesquisa computa uma seqüência de equações de regressão, adicionando ou deletando uma variável independente em cada passo. A rotina de regressão stepwise permite que uma variável independente, trazida para dentro do modelo em um estágio anterior, seja removida subseqüentemente se ela não ajudar na conjunção com variáveis adicionadas nos últimos estágios. Esta rotina conduz a um teste para rastrear alguma variável independente que seja altamente correlacionada com variáveis independentes já incluídas no modelo. A limitação da procura da regressão stepwise é que ela presume a existência de um único conjunto ótimo de variáveis independentes e busca identificá-lo. Como notado anteriormente, não existe freqüentemente um único conjunto ótimo. Outra limitação da rotina de regressão stepwise, é que ela algumas vezes surge com um conjunto de variáveis independentes razoavelmente fracos para predições, quando as variáveis independentes estão altamente correlacionadas (DRAPER & SMITH, 1981);
- *Seleção forward*: esse procedimento de procura é uma versão simplificada da regressão stepwise, omitindo o teste, se uma variável uma vez que tenha entrado no modelo deva ser retirada. Esse procedimento considera, inicialmente, um modelo simples usando a variável de maior coeficiente de correlação com a variável dependente. Uma variável por vez é incorporada até que não haja mais inclusão, e as variáveis selecionadas definem o modelo;
- *Eliminação backward*: esse procedimento de procura é oposto à seleção forward. Ele começa com o modelo contendo todas as variáveis independentes potenciais. O

procedimento de eliminação backward requer mais computações que o método de seleção forward, já que ela começa com o maior modelo possível. Entretanto, ela tem uma vantagem de mostrar ao analista as implicações do modelo com muitas variáveis;

- A abordagem mais simples, mas uma das mais exigentes, para especificar o modelo de regressão é empregar uma perspectiva confirmatória quando o pesquisador especifica completamente o conjunto de variáveis independentes a serem incluídas. Nesse caso, o pesquisador tem total controle sobre a seleção das variáveis. Ainda que a especificação confirmatória seja conceitualmente simples, o pesquisador deve assegurar de que o conjunto de variáveis atinja a previsão máxima, embora mantendo o modelo parcimonioso (HAIR, J. F. et al., 2005).

Se um método confirmatório, de busca seqüencial ou combinatório for escolhido, o critério mais importante é o bom conhecimento do pesquisador sobre o contexto da pesquisa, que permita uma perspectiva objetiva e fundamentada quanto às variáveis a serem incluídas e aos sinais e magnitude esperados de seus coeficientes. Sem esse conhecimento, os resultados da regressão podem ter elevada precisão preditiva sem qualquer relevância gerencial ou teórica. O pesquisador jamais deve ser totalmente orientado por qualquer uma dessas abordagens, mas, em vez disso, deve usá-las após considerar todas cuidadosamente e então aceitar os resultados somente após cuidadosa investigação (HAIR, J. F. et al., 2005).

O impacto da multicolinearidade (a qual se refere à correlação entre três ou mais variáveis independentes que evidenciam quando uma é regressada em relação às outras) é reduzir o poder preditivo de qualquer variável independente na medida em que ela é associada com as outras variáveis independentes (MALHOTRA, 2004). Quando a colinearidade aumenta, a variância única explicada por cada variável independente diminui e o percentual da previsão compartilhada aumenta. Como essa previsão compartilhada pode ser considerada apenas uma vez, a previsão geral aumenta muito mais vagarosamente quando variáveis independentes com multicolinearidade elevada são acrescentadas.

B.4.1-) Seleção de Dados para Teste

A predição é o processo de escolher o melhor resultado possível baseado na análise de dados históricos. O usuário deve analisar a informação descoberta de acordo com sua tarefa de suporte à decisão e objetivos. Portanto, após terem sido realizadas as tarefas de seleção, tratamento e pré-processamento dos dados, se procede agora à aplicação de técnicas de

modelagem avançada, de modo a identificar o modelo preditivo de apoio à identificação do relacionamento dos indicadores de desempenho.

O processo de construção de um modelo de regressão requer repetidas análises sobre o conjunto de dados para checar se o modelo ajusta-se bem os dados. Um modelo ruim pode levar a previsões errôneas. Esse fato pode ocorrer devido à má escolha das variáveis independentes. Uma forma de medir as previsões tendenciosas é deixar alguns dos dados originais fora dos cálculos para determinação do modelo e usá-los para realizar o ajustamento do poder preditivo do modelo (NETER & WASSERMAN, 1974).

O conjunto de dados para a modelagem contempla 180 amostras por variável independente, onde são utilizadas 90 amostras para o subconjunto de testes a fim de refinar o modelo. Com esta operação pretende-se confirmar a generalização do modelo, avaliando-se os resultados obtidos, para que as suas previsões não sejam válidas apenas para o caso específico. É utilizado um conjunto de testes, para avaliar a generalização do modelo e o seu desempenho, conforme o método da divisão da amostra, apresentado no Capítulo 2.

3.6 A Metodologia Proposta MDAV versus CRISP-DM

A metodologia proposta de Mineração de Dados para Administração de Vendas fundamenta o seu desenvolvimento em três componentes principais:

- *Obtenção de Informação*: Sistematiza as atividades de obtenção de dados quanto ao tipo de dados; à sua origem e problemática na sua aquisição;
- *Extração de Conhecimento*: Expõe todos os aspectos relacionados com as atividades de procura e extração de conhecimento em BD, nomeadamente no que respeita às atividades de Mineração de Dados em vendas;
- *Aplicação e Avaliação de Resultados*: Sistematiza e expõe a aplicação dos resultados obtidos com tarefas de DM em atividades de vendas, instanciando para cada um, conjunto de exemplos e técnicas recomendáveis para a sua resolução.

A Figura 3.5 representa um paralelismo entre os componentes e as respectivas atividades da metodologia MDAV com a metodologia de Mineração de Dados CRISP-DM (representada pela redução da Figura 2.6). Em seguida especifica-se o funcionamento desse instrumento bem como se define e caracteriza cada uma das atividades.

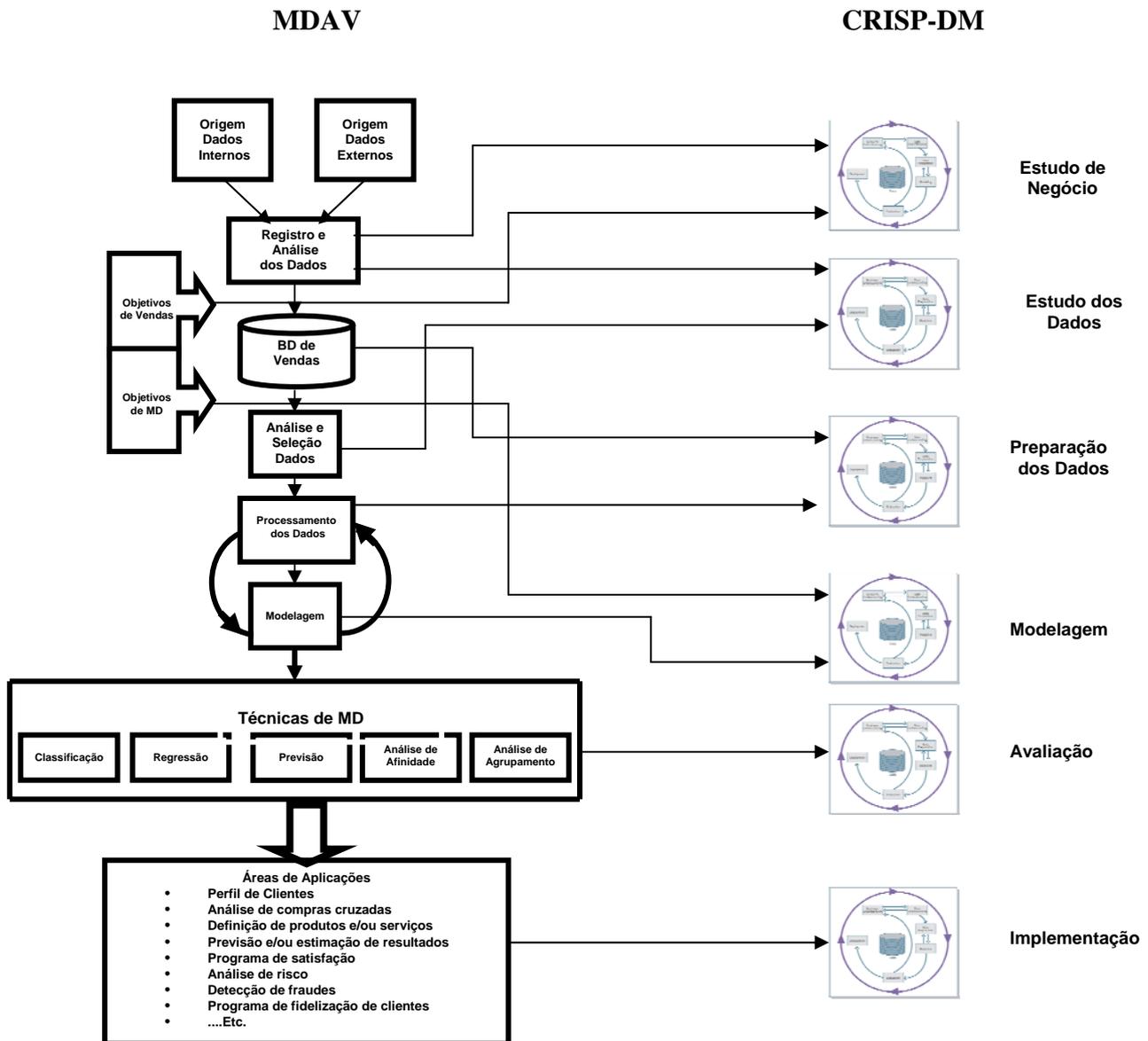


Figura 3.5 – Paralelismo entre a metodologia proposta e CRISP-DM.

Fonte: Elaborada pelo autor.

- O *registro e análise dos dados* incluem toda a atividade de importação de dados e aglomera duas fases da metodologia CRISP-DM (fases de *Estudo do Negócio* e de *Exploração de Dados*). Dado o fato de ambas elaborarem um estudo prévio de modo a melhorar a compreensão do problema real em termos de enquadramento dos objetivos em função do problema de negócio;
- A definição dos *Objetivos de Vendas* encontram um paralelo na fase de *Estudo de Negócio* da metodologia CRISP-DM, pelo fato de ambos os casos considerarem o negócio em que se irá desenvolver o projeto havendo, contudo na metodologia MDAV a particularidade de o mesmo objetivo de venda poder implicar o desenvolvimento de

diferentes atividades de mercado, cada uma com o seu objetivo específico (concorrente com os objetivos gerais de mercado);

- A criação da *BD de Vendas* encontra referencial na metodologia CRISP-DM em termos de *Preparação de Dados* devido à sua natureza seletora de registros. Tal como referido, é no decorrer do processo da criação de BD de vendas que se realiza a seleção de dados que irão servir de base para a realização das ações;
- A fase de definição dos *objetivos de Mineração de Dados* da metodologia MDAV encontra paralelo na metodologia CRISP-DM na fase de *Modelagem*, consistindo na seleção dos algoritmos a serem usados e no efetivo processamento do modelo;
- A tarefa de *Seleção de Dados* tem correspondência dentro da metodologia CRISP-DM em termos da *Exploração de Dados*, pelo fato de incorporar atividade de reconhecimento e explicação dos dados disponíveis;
- O *Pré-Processamento de Dados* define-se em paralelo com a fase da *Preparação de Dados*, sendo em termos da metodologia MDAV mais abrangente. Isto se deve ao fato da sua recursividade ser aplicável em várias atividades de mercado, ou seja, ao existir uma diversidade de atividades de mercado para o mesmo objetivo (mercado), possibilita a coexistência de diferentes objetivos de Mineração de Dados no mesmo projeto;
- A fase da *Modelagem* tem como preocupação a seleção e aplicação de algoritmos de Mineração de Dados que permitam alcançar os objetivos das atividades de vendas. No paralelo entre metodologias, em ambas existe a mesma denominação;
- A colocação de *técnicas de Mineração de Dados* (classificação, regressão, previsão, análise de afinidades e análise de agrupamento) para a definição das atividades de vendas requer a existência de uma avaliação prévia nos resultados obtidos. Por isso, foi estabelecido o paralelo com a metodologia CRISP-DM ao nível da fase de avaliação, uma vez que é nesse momento que os resultados de Mineração de Dados são avaliados em função das necessidades (objetivos) estabelecidas;
- A *Implementação* caracteriza-se pela aplicação dos resultados às atividades de vendas. Os vários modelos (decorrentes dos vários resultados obtidos) devem ser avaliados tendo em consideração os objetivos de mercado, certificando-se de que não existem falhas ou contradições em relação às regras do negócio. Os vários parâmetros das ferramentas de modelagem devem ser ajustados, os modelos resultantes devem ser convenientemente interpretados e o desempenho explicado.

- Face ao exposto, conclui-se que a metodologia MDAV concretiza fases da metodologia CRISP-DM em termos de aplicação de Mineração de Dados em vendas. O desenvolvimento de projetos Administração de Vendas segundo a metodologia MDAV permitirá incorporar nas atividades de mercado, processos já comprovados e demonstrados cientificamente para análise e exploração dos dados.

3.7 Ferramenta SPSS

Para executar uma análise completa, qualquer pesquisador necessita de um conjunto de ferramentas integradas e que lhes facilitem o trabalho. Infelizmente, muitos pesquisadores não dispõem de uma solução flexível que cubra todas as etapas do processo analítico. Uma das formas que os pesquisadores utilizam para dispor de uma solução completa de análise é a de juntar vários *softwares* especializados de diferentes fabricantes. O problema com esta solução é que o pesquisador pode encontrar problemas de incompatibilidade e despender muito tempo em solucionar problemas e não em encontrar resultados. Uma opção adequada é a utilização de uma solução analítica que seja expansível, por meio de módulos adicionais ou de programas de software para funções especializadas. Mas sempre integráveis e compatíveis.

A identificação do modelo é feita utilizando os princípios associados à KDD, procurando-se padrões nos dados dos parâmetros de desempenho de vendas de telecomunicações. O SPSS, por ser uma ferramenta bastante completa, foi a ferramenta de Mineração de Dados escolhida para desenvolver esse trabalho. Oferece um conjunto variado de funcionalidades de Mineração de Dados e grande flexibilidade para criar aplicações personalizadas.

O SPSS é um *software* analítico que fornece aos pesquisadores as funcionalidades de planejamento, acesso aos dados, acesso, preparação, análise dos dados e da disponibilização dos resultados. Devido à integração total com os seus módulos adicionais e com outros produtos especializados de análise SPSS, o tempo do pesquisador é gasto no processo analítico e não em questões menores de incompatibilidade. O pacote estatístico SPSS é um poderoso sistema de análises estatísticas e tratamento dos dados, num ambiente gráfico, em que a utilização mais freqüente, para a maioria das análises a ser efetuada, se resume à seleção das respectivas opções em menus e caixas de diálogo. Dispondo ainda de um editor de comandos, onde o pesquisador poderá recorrer a fim de realizar determinado tipo de análise mais complexa e elaborada. Muitos dos procedimentos estatísticos do SPSS incluem já a elaboração

de gráficos adequados às análises respectivas. São apresentadas a seguir algumas características desta ferramenta (NORUSIS, 2004a) que a torna uma opção bastante atrativa:

- Abrange todo o processo de KDD;
- Fornece uma interface gráfica intuitiva, para programação visual;
- Fornece um variado leque de algoritmos de modelagem como às redes neurais, redes de Kohonen e análises baseadas em árvores de decisão, regras de associação e regressão linear múltipla;
- Permite combinar técnicas e modelos;
- Trabalha com várias plataformas computacionais, operando em sistemas *Windows* e *Unix*;
- Permite o acesso, via *Open Database Connectivity* (ODBC) aos BD relacionais, trabalhando também com arquivos de texto;
- Proporciona a avaliação da qualidade dos dados;
- Permite operações de manipulação dos dados como a filtragem, ordenação e agregação;
- Permite a exploração gráfica dos dados, incluindo as relações entre valores;
- Facilidade de elaboração de estatísticas descritivas sobre os dados.

Resultados da Pesquisa

Nesse capítulo são apresentados os resultados obtidos a partir da análise dos dados relativos a 15 empresas que integram esse estudo, com dados do período de janeiro a dezembro de 2006. Inicialmente, na seção 4.1, é apresentada a metodologia utilizada para estimação e avaliação do modelo de regressão linear das variáveis que relacionam a produtividade de vendas com os indicadores de desempenho das empresas de telecomunicações analisadas. Posteriormente, na seção 4.2, é apresentado o detalhamento dos resultados obtidos na aplicação dos testes realizados nos três métodos do modelo referenciado. Em seguida, na seção 4.3, é realizada a análise comparativa do modelo *Stepwise* em relação ao Combinatório. Na seção 4.4, é feita uma interpretação dos coeficientes do modelo de regressão selecionado e na seção 4.5 é feito uma análise aos pressupostos do modelo. Por último, na seção 4.6, são analisadas as hipóteses levantadas para a verificação da validade do modelo.

4.1-) Metodologia Utilizada para Estimação e Avaliação do Modelo de Regressão

Nesta fase é estimando o modelo de regressão e avaliando a precisão preditiva geral das variáveis independentes por meio da seleção de um método para especificar o modelo de regressão a ser ajustado, avaliando a significância estatística do modelo geral na previsão da variável dependente e determinando se alguma das observações tem influência indevida nos resultados. Quando se ajusta um modelo de regressão múltipla, pode acontecer que se justifique estatisticamente incluir na equação de regressão todas as variáveis independentes, ou que se incluam apenas algumas dessas variáveis explanatórias. Esta decisão é tomada em função da significância do parâmetro de regressão de cada uma das variáveis, ou pelo acréscimo do coeficiente de determinação, R^2 , provocado pela inclusão dessas variáveis.

A fim de comparar diversos modelos de regressão, é feito inicialmente com uma abordagem para especificar o modelo de regressão empregando uma perspectiva confirmatória onde o pesquisador especifica completamente o conjunto de variáveis independentes a serem incluídas, denominado de Especificação Confirmatória. Em seguida é repetido a metodologia, porém com um outro método de busca por meio de uma abordagem combinatória que é um processo de busca generalizada em todas as possíveis combinações das variáveis independentes pelo método Combinatório (*Enter*). Por último, é utilizado o método de busca seqüencial para estimar a equação de regressão com um conjunto de variáveis sendo então acrescentado seletivamente ou eliminado variáveis até que uma medida de critério geral seja alcançada, sendo denominado Seqüencial (*Stepwise*). São não só avaliados os coeficientes ajustados, mas também o impacto potencial das variáveis omitidas para garantir que a significância gerencial seja avaliada juntamente com a significância estatística. Para tanto, a metodologia utilizada nesse trabalho é baseada na seguinte estrutura:

- Um modelo de regressão linear específico (ou seja, contendo as variáveis explicativas especificadas pelos especialistas de negócios da área de telecomunicações) foi ajustado em primeiro lugar e em seguida foram efetuados testes individuais aos coeficientes ajustados;
- Um modelo de regressão linear completo (ou seja, contendo todas as variáveis explicativas disponíveis) foi ajustado em segundo lugar e em seguida foram efetuados testes individuais aos coeficientes ajustados;
- Um novo modelo, *Stepwise* (por conter menos variáveis explicativas que o Combinatório) foi ajustado, tendo-se retirado aquelas variáveis que demonstraram não contribuir para a explicação da variação da variável dependente.

Assumiu-se, conforme já demonstrado anteriormente, que os termos de erro ε_i são independentes e seguem uma distribuição $N(0, \sigma^2)$.

4.2-) Métodos Testados no Modelo de Regressão Linear Múltipla

4.2.1-) Método Especificação Confirmatório

Tendo por base as hipóteses da pesquisa e o conhecimento dos especialistas na área de gestão de negócio de telecomunicações, é apresentado o seguinte modelo

$$REFX_i = \beta_0 + \beta_1 CO1_i + \beta_2 EO1_i + \beta_3 FC1_i + \beta_4 GP1_i + \beta_5 GP2_i + \beta_6 GP5_i + \beta_7 IA2_i + e_i. \quad (4.1)$$

4.2.1.1-) Estimação dos Parâmetros do Modelo

Foi utilizado o programa SPSS para estimar os parâmetros do modelo, por meio do método dos mínimos quadrados.

Tabela 4.1 – Coeficientes de regressão do método confirmatório.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Correlations			Collinearity Statistics		
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF	
1	(Constant)	,000	,044		-,010	,992	-,087	,086						
	Venda x Mapeamento da Concorrência	,293	,050	,293	5,908	,000	,195	,391	,484	,354	,272	,862	1,160	
	Asseritvidade de Entrega e Cadastro de Contratos Fixa	,198	,050	,198	3,984	,000	,100	,296	,394	,247	,184	,857	1,167	
	Média de Visitas por Consultor	,060	,048	,060	1,263	,208	-,034	,155	,177	,081	,058	,930	1,075	
	Produtividade de Visita	,166	,049	,166	3,390	,001	,070	,263	,368	,212	,156	,884	1,131	
	Eficácia de Visitas	,203	,048	,203	4,221	,000	,108	,298	,285	,261	,195	,919	1,088	
	Produtividade de Oportunidades	,251	,049	,251	5,088	,000	,154	,348	,443	,310	,235	,872	1,147	
	Volume de Vendas Customizadas Fixa	,131	,047	,131	2,749	,006	,037	,224	,156	,173	,127	,942	1,062	

a. Dependent Variable: Produtividade de Vendas Fixa

Fonte: Relatório do SPSS.

A Tabela 1 (um *output* do SPSS) apresenta os coeficientes do modelo ajustados, respectivos desvios padrão, resultados dos testes individuais (estatística t e nível de significância), intervalo de confiança, correlações e estatísticas de colinearidade (NORUSIS, 2004a).

Logo, a função de regressão estimada é:

$$REFX_i = 0,293CO1_i + 0,198EO1_i + 0,060FC1_i + 0,166GP1_i + 0,203GP2_i + 0,251GP5_i + 0,131IA2_i. \quad (4.2)$$

4.2.1.2-) Coeficiente de Determinação Múltiplo

O coeficiente de determinação múltiplo (R^2) é determinado pela razão entre a soma dos quadrados da regressão ($SSR = 111,158$) e a soma dos quadrados total ($SST = 230,863$).

Tabela 4.2 – Sumário do modelo da regressão do método confirmatório.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	,694 ^a	,481	,467	,70042	,481	32,368	7	244	,000	1,827

a. Predictors: (Constant), Volume de Vendas Customizadas Fixa, Produtividade de Oportunidades, Média de Visitas por Consulta de Visita, Eficácia de Visitas, Venda x Mapeamento da Concorrência, Assertividade de Entrega e Cadastro de Contratos Fixa

b. Dependent Variable: Produtividade de Vendas Fixa

Fonte: Relatório do SPSS.

Como se pode observar a partir da Tabela 4.2, $R^2 = 0,481$, o que significa que as sete variáveis explicativas constantes na função de regressão são responsáveis por explicar 48,1% da variação da produtividade de vendas (em outras palavras, o modelo explica 48,1% da variação da variável dependente *REFX*).

4.2.1.3-) Análise de Variância

Para efetuar um teste global de significância ao modelo, ou seja, a existência ou não de um relacionamento entre a variável dependente e as variáveis independentes, recorreu-se ao *SPSS* (NORUSIS, 2004b) para construir uma tabela ANOVA (Tabela 4.3), e executar o Teste-*F*, correspondendo às seguintes hipóteses:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0 \text{ (ausência de efeito);}$$

$$H_1 : \beta_1 \neq 0 \vee \beta_2 \neq 0 \vee \beta_3 \neq 0 \vee \beta_4 \neq 0 \vee \beta_5 \neq 0 \vee \beta_6 \neq 0 \vee \beta_7 \neq 0 \text{ (presença de efeito).}$$

A hipótese H_0 significa que a regressão não é significativa, ou seja: que a equação de regressão não explica a variação na variável resposta e que não existe relação linear entre a variável dependente e o conjunto de variáveis independentes utilizadas.

O teste estatístico *F* e o teste de efeito conjunto das variáveis independentes indicam, após análise dos testes de hipótese, que o teste estatístico *F* é igual a 32,368 sendo os graus de liberdade do numerador e do denominador 7 e 244, respectivamente. A análise foi definida ao nível de significância de $\alpha = 5\%$ ($\alpha = 95\%$), sendo que o valor do *F* crítico é $F_c \cong 2,05$. A hipótese que se pretende testar é verificar se as variáveis independentes – *COI*, *EOI*, *FCI*, *GPI*,

GP_2 , GP_5 e IA_2 – exercem conjuntamente efeitos significativos sobre a variável dependente produtividade de vendas, sendo formulada as hipóteses acima referenciadas.

Tabela 4.3 – Variância (ANOVA) do método confirmatório.

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	111,158	7	15,880	32,368	,000 ^a
	Residual	119,705	244	,491		
	Total	230,863	251			

a. Predictors: (Constant), Volume de Vendas Customizadas Fixa, Produtividade de Oportunidades, Média de Visitas por Consultor, Produtividade de Visita, Eficácia de Visitas, Venda x Mapeamento da Concorrência, Assertividade de Entrega e Cadastro de Contratos Fixa

b. Dependent Variable: Produtividade de Vendas Fixa

Fonte: Relatório do SPSS.

Regra de decisão: se $F_{obs} \leq F_c = F_{(95\% ; 7 ; 244)}$ se aceita H_0 . Da tabela da distribuição F , $F_{(95\% ; 7 ; 244)} \cong 2,05$. Logo, como $F_{obs} \geq F_{(95\% ; 7 ; 244)}$, deve-se rejeitar a hipótese nula e se aceita a hipótese H_1 no nível de significância de 5%, ou seja, existe pelo menos um parâmetro diferente de zero. Isto significa que as variáveis têm influência sobre a variável produtividade de vendas, ao nível de 5%, ou seja, devendo ser concluído que a produtividade de vendas está relacionada com as variáveis independentes.

A existência de uma relação de regressão, por si só, não garante que predições úteis podem ser feitas usando esse modelo. Esse teste é apenas uma etapa na verificação de aceitação do modelo.

4.2.1.4-) Testes Individuais dos Parâmetros

Os coeficientes de regressão ajustados são usados para calcular os valores previstos para cada observação e para expressar a variação esperada na variável dependente para cada variação unitária nas variáveis independentes. Nesse trabalho, o objetivo é saber quais indicadores de desempenho (variáveis independentes) têm maior efeito na previsão da produtividade de vendas (variável dependente). Em alguns casos, os coeficientes de regressão não fornecem literalmente essa informação. Para resolver esse problema de explicação é usado um coeficiente de regressão modificado chamado de coeficiente beta os quais são os coeficientes resultantes de dados padronizados que eliminam o problema de lidar com diferentes unidades de medidas, refletindo assim o impacto relativo sobre a variável dependente de uma mudança em um desvio padrão de qualquer variável.

Os testes individuais dos parâmetros β_i são realizados utilizando à estatística t de Student:

$$H_0 : \beta_i = 0 \text{ (ausência de efeito);}$$

$$H_1 : \beta_i \neq 0 \text{ (presença de efeito):}$$

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

em que b_i é o valor ajustado para o parâmetro β_i , e $s(b_i)$ o valor ajustado para o desvio padrão de b_i .

Se $|t^*| \leq t\left(1 - \frac{\alpha}{2}; n - p\right)$, então se deve aceitar a hipótese nula. Caso contrário, deve-se rejeitá-la.

O teste estatístico t do coeficiente β_1 da variável COI – Venda x Mapeamento da Concorrência, $t_1 = 5,908$, para esse parâmetro é $H_0: \beta_1=0$ (ausência de efeito) e $H_1: \beta_1 \neq 0$ (presença de efeito). Os graus de liberdade são 7 e 244, o nível de significância 5%, sendo o valor crítico de $t_c \cong 2,241$. Como $|t_c| \cong 2,241 < |t| = 5,908$, rejeita-se a hipótese nula (H_0) em favor da hipótese alternativa de efeito positivo, com um nível de significância de 5%. Esse resultado indica que a influência da variável Venda x Mapeamento da Concorrência sobre a produtividade de vendas é estatisticamente significativa, ao nível de 5%.

O teste estatístico t do coeficiente β_3 da variável FCI – Média de Visitas por Consultor, $t_3 = 1,263$, para esse parâmetro é $H_0: \beta_3=0$ (ausência de efeito) e $H_1: \beta_3 \neq 0$ (presença de efeito). Os graus de liberdade são 7 e 244, o nível de significância 5%, sendo o valor crítico de $t_c \cong 2,241$. Como $|t_c| \cong 2,241 > |t| = 1,263$, rejeita-se a hipótese (H_1) em favor da hipótese nula, com um nível de significância de 5%. Esse resultado indica que a influência da variável Média de Visitas por Consultor sobre a produtividade de vendas não é estatisticamente significativa porque está dentro da região de aceitação ($sig. \geq 0,05$, ou seja, o teste foi realizado com um nível de significância de 5%).

Efetou-se o cálculo da estatística t para os outros parâmetros a serem testados, recorrendo ao *SPSS* (NORUSIS, 2004a), cujos resultados estão apresentados no quadro dos coeficientes. Como se pode verificar, a partir da Tabela 4.1 e fazendo-se análise similar ao modo como foi realizado acima, foi constatado que apenas a estatística t correspondente ao parâmetro β_3 está dentro da região de aceitação ($sig. \geq 0,05$, ou seja o teste foi realizado com um nível de significância de 5%).

De fato, $t(0,975 ; 244) \cong 2,241$, e $|1,263| < 2,241 < |2,749| < |3,390| < |3,984| < |4,221| < |5,088| < |5,908|$. Deve ser concluído, que ao nível de significância dado (5%), os coeficientes $\beta_1, \beta_2, \beta_4, \beta_5, \beta_6$ e β_7 são significativos. O mesmo não se pode dizer do parâmetro β_3 .

Conclusão: a variável $FC1$ deve ser retirada do modelo, reduzindo a sua complexidade (dimensionalidade). A performance do modelo, dada pelo valor de R^2 , foi pouco afetada (cerca de 1% menor). Gerencialmente, esse indicador de desempenho – Média de Visitas por Consultor – que tem historicamente, nessas empresas, um desempenho satisfatório, deveria ter um forte relacionamento com as vendas. Esse fato (exclusão desta variável) pode caracterizar que as visitas realizadas pelos consultores estão sendo feitas, em sua maioria, sem um planejamento adequado, não estando sendo caracterizados os preceitos definidos pela empresa operadora.

4.2.2-) Método Combinatório (Enter)

Havendo 17 variáveis independentes o modelo de regressão linear caracterizado é do tipo

$$\begin{aligned} REF\text{X}_i = & \beta_0 + \beta_1 CO1_i + \beta_2 EO1_i + \beta_3 EO4_i + \beta_4 EO5_i + \beta_5 EO6_i + \beta_6 EO7_i + \\ & \beta_7 FC1_i + \beta_8 FC2_i + \beta_9 FC3_i + \beta_{10} GP1_i + \beta_{11} GP2_i + \beta_{12} GP3_i + \beta_{13} GP4_i + \\ & \beta_{15} IA2_i + \beta_{16} IA3_i + \beta_{17} IA4 + \varepsilon_i. \end{aligned} \quad (4.3)$$

4.2.2.1-) Estimação dos Parâmetros do Modelo

Foi utilizado o programa *SPSS* para estimar os parâmetros do modelo, por meio do método dos mínimos quadrados.

Tabela 4.4 – Coeficientes de regressão do método combinatório.

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Correlations			Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1												
(Constant)	,000	,044		-,010	,992							
Venda x Mapeamento da Concorrência	,289	,051	,289	5,605	,000	,187	,390	,484	,344	,259	,803	1,246
Assertividade de Entrega e Cadastro de Contratos Fixa	,203	,052	,204	3,945	,000	,102	,305	,394	,250	,182	,801	1,248
Retorno de Contratos Fixa	,010	,052	,010	,188	,851	-,093	,113	-,085	,012	,009	,783	1,278
Abertura de Ordem de Serviço Fixa	-,028	,050	-,028	-,565	,573	-,127	,070	,097	-,037	-,026	,856	1,169
Conversão de Contratos Fixa	-,024	,050	-,024	-,493	,623	-,122	,073	,046	-,032	-,023	,869	1,150
Cancelamento de Vendas Fixa	-,065	,051	-,065	-1,276	,203	-,164	,035	,004	-,083	-,059	,830	1,205
Média de Visitas por Consultor	,030	,051	,030	,587	,558	-,071	,131	,177	,038	,027	,807	1,239
Assertividade de Agendamento	,067	,057	,067	1,189	,236	-,044	,179	,029	,077	,055	,665	1,503
Visitação X Agendamento	-,085	,054	-,085	-1,588	,114	-,191	,021	-,042	-,103	-,073	,737	1,357
Produtividade de Visita	,184	,052	,184	3,536	,000	,081	,287	,368	,225	,163	,787	1,271
Eficácia de Visitas	,161	,058	,161	2,796	,006	,048	,275	,285	,180	,129	,640	1,564
Mix de Produtos	,010	,053	,010	,196	,845	-,095	,116	,155	,013	,009	,749	1,336
Distribuição de Vendas por Consultores	,062	,062	,062	1,006	,315	-,059	,183	,369	,066	,046	,562	1,781
Produtividade de Oportunidades	,233	,051	,234	4,541	,000	,132	,335	,443	,285	,210	,807	1,239
Volume de Vendas Customizadas Fixa	,132	,049	,132	2,679	,008	,035	,230	,156	,173	,124	,873	1,145
Aprovação de Viabilidades Fixa	,026	,050	,026	,513	,608	-,073	,124	,102	,034	,024	,850	1,177
Assertividade de Análise de Viabilidade Fixa	-,058	,053	-,059	-1,106	,270	-,163	,046	-,167	-,072	-,051	,763	1,311

a. Dependent Variable: Produtividade de Vendas Fixa

Fonte: Relatório do SPSS.

A Tabela 4.4 (um *output* do SPSS), apresenta os coeficientes do modelo ajustados, respectivos desvios padrão, resultados dos testes individuais (estatística *t* e nível de significância), intervalo de confiança, correlações e estatísticas de colinearidade (NORUSIS, 2004a).

Logo, a função de regressão estimada é:

$$\begin{aligned}
 REF_{X_i} = & 0,289_1 CO1_i + 0,204 EO1_i + 0,010 EO4_i - 0,028 EO5_i - 0,024 EO6_i - \\
 & 0,065 EO7_i + 0,030 FC1_i + 0,067 FC2_i - 0,085 FC3_i + 0,184 GP1_i + \\
 & 0,161 GP2_i + 0,010 GP3_i + 0,062 GP4_i + 0,234 IA2_i + 0,132 IA2_i + \\
 & 0,026 IA3_i - 0,059 IA4_i.
 \end{aligned}
 \tag{4.4}$$

4.2.2.2-) Coeficiente de Determinação Múltiplo

O coeficiente de determinação múltiplo (R^2) é determinado pela razão entre a soma dos quadrados da regressão ($SSR= 115,610$) e a soma dos quadrados total ($SST = 230,863$).

Tabela 4.5 – Sumário do modelo da regressão do método combinatório.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	,708 ^a	,501	,465	,70181	,501	13,807	17	234	,000	1,910

a. Predictors: (Constant), Assertividade de Análise de Viabilidade Fixa, Volume de Vendas Customizadas Fixa, Produtividade de Oportunidades, Retorno de Contratos Fixa, Aprovação de Viabilidades Fixa, Mix de Produtos, Abertura de Ordem de Serviço Fixa, Conversão de Contratos Fixa, Venda x Mapeamento da Concorrência, Visitação X Agendamento, Cancelamento de Vendas Fixa, Média de Visitas por Consultor, Assertividade de Entrega e Cadastro de Contratos Fixa, Produtividade de Visita, Eficácia de Visitas, Assertividade de Agendamento, Distribuição de Vendas por Consultores

b. Dependent Variable: Produtividade de Vendas Fixa

Fonte: Relatório do SPSS.

Como se pode observar a partir da Tabela 4.5, $R^2 = 0,501$, o que significa que todas as variáveis explicativas constantes na função de regressão são responsáveis por explicar 50,1% da variação da produtividade de vendas (em outras palavras, o modelo explica 50,1% da variação da variável dependente *REFX*).

4.2.2.3-) Análise de Variância

Para efetuar um teste global de significância ao modelo, ou seja, a existência ou não de um relacionamento significativo entre a variável dependente e as variáveis independentes, recorreu-se ao *SPSS* (NORUSIS, 2004b) para construir uma tabela ANOVA (Tabela 4.6), e executar o Teste-*F*, correspondendo às seguintes hipóteses:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = \beta_{15} = \beta_{16} = \beta_{17} = 0 \quad (\text{ausência de efeito});$$

$$H_1 : \beta_1 \neq 0 \vee \beta_2 \neq 0 \vee \beta_3 \neq 0 \vee \beta_4 \neq 0 \vee \beta_5 \neq 0 \vee \beta_6 \neq 0 \vee \beta_7 \neq 0 \vee \beta_8 \neq 0 \vee \beta_9 \neq 0 \vee \beta_{10} \neq 0 \vee \beta_{11} \neq 0 \vee \beta_{12} \neq 0 \vee \beta_{14} \neq 0 \vee \beta_{15} \neq 0 \vee \beta_{16} \neq 0 \vee \beta_{17} \neq 0$$

(presença de efeito).

A hipótese H_0 significa que a regressão não é significativa, ou seja: que a equação de regressão não explica a variação na variável resposta e que não existe relação linear entre a variável dependente e o conjunto de variáveis independentes utilizadas.

O teste estatístico *F* e o teste de efeito conjunto das variáveis independentes, indica, após análise dos testes de hipótese, que o teste estatístico *F* é igual a 13,807 sendo os graus de liberdade do numerador e do denominador 17 e 234, respectivamente. A análise foi definida ao nível de significância de $\alpha = 5\%$ ($\alpha = 95\%$), sendo que o valor do *F* crítico é $F_c \cong 1,70$. A hipótese que se pretende testar é verificar se as variáveis independentes – *CO1*, *EO1*, *EO4*, *EO5*, *EO6*, *EO7*, *FC1*, *FC2*, *FC3*, *GP1*, *GP2*, *GP3*, *GP4*, *GP5*, *IA2*, *IA3* e *IA4* – exercem conjuntamente

efeitos significativos sobre a variável dependente produtividade de vendas, sendo formulada as hipóteses acima referenciadas.

Tabela 4.6 – Variância (ANOVA) do método combinatório.

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	115,610	17	6,801	13,807	,000 ^a
	Residual	115,253	234	,493		
	Total	230,863	251			

a. Predictors: (Constant), Assertividade de Analise de Viabilidade Fixa, Volume de Vendas Customizadas Fixa, Produtividade de Oportunidades, Retorno de Contratos Fixa, Aprovação de Viabilidades Fixa, Mix de Produtos, Abertura de Ordem de Serviço Fixa, Conversão de Contratos Fixa, Venda x Mapeamento da Concorrência, Visitação X Agendamento, Cancelamento de Vendas Fixa, Média de Visitas por Consultor, Assertividade de Entrega e Cadastro de Contratos Fixa, Produtividade de Visita, Eficácia de Visitas, Assertividade de Agendamento, Distribuição de Vendas por Consultores

b. Dependent Variable: Produtividade de Vendas Fixa

Fonte: Relatório do SPSS.

Regra de decisão: se $F_{obs} \leq F_c = F_{(95\% ; 17 ; 234)}$ se aceita H_0 . Da tabela da distribuição F , $F_{(95\% ; 17 ; 234)} \cong 1,70$. Logo, como $F_{obs} \geq F_{(95\% ; 17 ; 234)}$, deve-se rejeitar a hipótese nula e se aceita a hipótese H_1 ao nível de significância de 5%, ou seja, existe pelo menos um parâmetro diferente de zero. Isto significa que as variáveis têm influência significativa sobre a variável produtividade de vendas, ao nível de 5%, ou seja, deve ser concluído que a produtividade de vendas está relacionada com as variáveis independentes.

A existência de uma relação de regressão, por si só, não garante que predições úteis podem ser feitas usando esse modelo. Esse teste é apenas uma etapa na verificação de aceitação do modelo.

4.2.2.4-) Testes Individuais dos Parâmetros

Os testes individuais dos parâmetros β_i são realizados utilizando à estatística t de Student:

$$H_0 : \beta_i = 0 \text{ (ausência de efeito);}$$

$$H_1 : \beta_i \neq 0 \text{ (presença de efeito).}$$

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

em que b_i é o valor ajustado para o parâmetro β_i , e $s(b_i)$ o valor ajustado para o desvio padrão de b_i .

Se $|t^*| \leq t\left(1 - \frac{\alpha}{2}; n - p\right)$, então se deve aceitar a hipótese nula. Caso contrário, se deve

rejeitá-la.

Efetuiu-se o cálculo da estatística t para os 17 parâmetros a serem testados, recorrendo ao SPSS (NORUSIS, 2004a), cujos resultados estão apresentados no quadro dos coeficientes. Como se pode verificar a partir da Tabela 4.4 e fazendo-se análise similar ao modo como foi realizado no item 4.2.1.4, pode ser verificado que as estatísticas t correspondentes aos parâmetros $\beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{12}, \beta_{13}, \beta_{16}$ e β_{17} estão dentro da região de aceitação ($sig. \geq 0,05$, ou seja o teste foi realizado com um nível de significância de 5%).

De fato, $t(0,975 ; 234) \cong 2,241$, e $|0,188| < |0,196| < |0,493| < |0,513| < |0,565| < |0,587| < |1,006| < |1,106| < |1,189| < |1,276| < |1,588| < 2,241 < |2,679| < |2,796| < |3,536| < |3,945| < |4,541| < |5,605|$. Deve ser concluído que, ao nível de significância dado (5%), os coeficientes $\beta_1, \beta_2, \beta_{10}, \beta_{11}, \beta_{14}$ e β_{15} são significativos. O mesmo não se pode dizer dos outros parâmetros.

Conclusão: as variáveis $EO_4, EO_5, EO_6, EO_7, FC_1, FC_2, FC_3, GP_3, GP_4, IA_3$ e IA_4 devem ser retiradas do modelo. Verifica-se que, com a retirada dessas variáveis, os valores a serem obtidos nesse novo modelo são exatamente iguais aos obtidos com o método de busca seqüencial *Stepwise*, descrito em seguida.

4.2.3-) Método Seqüencial (Stepwise)

A partir das conclusões obtidas anteriormente, construiu-se um novo modelo de regressão linear, retirando as variáveis não representativas, por se ter considerado que essas não explicam de forma significativa o comportamento da produtividade de vendas.

O modelo *Stepwise* é então

$$REFX_i = \beta_0 + \beta_1 CO1_i + \beta_2 EO1_i + \beta_3 GP1_i + \beta_4 GP2_i + \beta_5 GP5_i + \beta_6 IA2_i + \varepsilon_i. \quad (4.5)$$

4.2.3.1-) Estimação dos Parâmetros do Modelo

Tabela 4.7 – Coeficientes de regressão do método seqüencial.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	,000	,044		-,010	,992	-,087	,087					
	Venda x Mapeamento da Concorrência	,299	,049	,299	6,039	,000	,201	,396	,484	,360	,279	,869	1,151
	Assertividade de Entrega e Cadastro de Contratos Fixa	,207	,049	,207	4,185	,000	,109	,304	,394	,258	,193	,872	1,146
	Produtividade de Visita	,168	,049	,168	3,435	,001	,072	,265	,368	,214	,159	,885	1,130
	Eficácia de Visitas	,196	,048	,196	4,103	,000	,102	,291	,285	,254	,189	,930	1,076
	Produtividade de Oportunidades	,254	,049	,254	5,147	,000	,157	,351	,443	,312	,238	,874	1,145
	Volume de Vendas Customizadas Fixa	,137	,047	,137	2,891	,004	,044	,230	,156	,182	,133	,952	1,051

a. Dependent Variable: Produtividade de Vendas Fixa

Fonte: Relatório do SPSS.

Da Tabela 4.7, pode ser obtida uma nova função de regressão estimada

$$REFX_i = 0,299CO1_i + 0,207EO1_i + 0,168GP1_i + 0,196GP2_i + 0,254GP5_i + 0,137IA2_i. \quad (4.6)$$

Tabela 4.8 – Sumário do modelo da regressão do método seqüencial.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	,691 ^a	,478	,465	,70128	,478	37,406	6	245	,000	1,801

a. Predictors: (Constant), Volume de Vendas Customizadas Fixa, Produtividade de Oportunidades, Produtividade de Visita, Eficácia de Visitas, Assertividade de Entrega e Cadastro de Contratos Fixa, Venda x Mapeamento da Concorrência

b. Dependent Variable: Produtividade de Vendas Fixa

Fonte: Relatório do SPSS.

4.2.3.2-) Coeficiente de Determinação Múltiplo

A partir da Tabela 4.8, verifica-se que 47,8% da variação na produtividade de vendas observadas, é explicada pelas variáveis exógenas do modelo. Como seria de esperar, o valor de R^2 é inferior no modelo Stepwise, quando comparado com o modelo Combinatório ($R^2 = 0,501$) embora a diferença não seja significativa (como é oportunamente comprovado).

4.2.3.3-) Análise de Variância

A análise da variância permite decompor a variação total observada na variável dependente em variação explicada pela função de regressão (Soma dos Quadrados da Regressão ou SSR) e variação não explicada pela função de regressão (Soma dos Quadrados

dos Erros ou *SSE*). A partir desses dois valores é possível obter a Média da Soma dos Quadrados da Regressão (*MSR*) e a Média da Soma dos Quadrados dos Erros (*MSE*).

Tabela 4.9 – Variância (ANOVA) do método seqüencial.

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	110,375	6	18,396	37,406	,000 ^a
	Residual	120,488	245	,492		
	Total	230,863	251			

a. Predictors: (Constant), Volume de Vendas Customizadas Fixa, Produtividade de Oportunidades, Produtividade de Visita, Eficácia de Visitas, Assertividade de Entrega e Cadastro de Contratos Fixa, Venda x Mapeamento da Concorrência

b. Dependent Variable: Produtividade de Vendas Fixa

Fonte: Relatório do SPSS.

O cálculo do *MSE* e do *MSR* permite efetuar um teste de significância global do modelo, utilizando a estatística *F* (*F-test*).

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0 \text{ (ausência de efeito);}$$

$$H_1 : \beta_1 \neq 0 \vee \beta_2 \neq 0 \vee \beta_3 \neq 0 \vee \beta_4 \neq 0 \vee \beta_5 \neq 0 \vee \beta_6 \neq 0 \text{ (presença de efeito).}$$

A hipótese H_0 significa que a regressão não é significativa, ou seja: que a equação de regressão não explica a variação na variável resposta e que não existe relação linear entre a variável dependente e o conjunto de variáveis independentes utilizadas.

A estatística F^* é calculada a partir da razão entre *MSR* e *MSE*. Quanto o valor de F^* é elevado, significa que uma grande parte da variação dos valores observados para a variável endógena é explicada pela reta de regressão. Caso contrário, quando F^* é reduzido, a reta de regressão apenas explica uma pequena parte da variação da variável endógena.

O critério de aceitação da hipótese nula resulta da comparação entre F^* e o valor da estatística *F* para $p-1$ e $n-p$ graus de liberdade (n é o número de observações e p o número de coeficientes do modelo) (NORUSIS, 2004b). Assim, se $F^* \leq F(1-\alpha; p-1; n-p)$, deve-se aceitar a hipótese nula, caso contrário, rejeita-se.

A partir da Tabela 4.9 pode-se observar que $F^* = 37,406$. Por outro lado, consultada a tabela estatística para a distribuição *F*, verifica-se que $F(95\%;6;245) \cong 2,1$. Logo, como $37,406 > 2,1$, deve ser rejeitada a hipótese nula, ou seja, o modelo é significativo a um nível de significância de 5%.

A existência de uma relação de regressão, por si só, não garante que predições úteis podem ser feitas usando esse modelo. Esse teste é apenas uma etapa na verificação de aceitação do modelo.

4.2.3.4-) Testes Individuais dos Parâmetros

Conforme já analisado, os testes individuais para os parâmetros β_i são realizados utilizando a estatística t de *Student*:

$$H_0 : \beta_i = 0 \text{ (ausência de efeito);}$$

$$H_1 : \beta_i \neq 0 \text{ (presença de efeito).}$$

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

Como se pode observar a partir do quadro de coeficientes já apresentado (Tabela 4.7), no nível de significância de 5%, qualquer dos parâmetros $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ e β_6 é significativo. De fato, determinado o valor da estatística t ($97,5\%;245$) $\cong 2,241$ na tabela da estatística t , verifica-se que $|t_i^*| \geq 2,241$ ($i = 1, \dots, 6$), o que significa que deve ser rejeitada a hipótese nula em todos os testes individuais.

O algoritmo stepwise escolhe $CO_1, EO_1, GP_1, GP_2, GP_5$ e IA_2 como regressores. As vendas são afetadas positivamente por esses indicadores de desempenho, pois para todos eles t foi maior que t ($97,5\%;245$) $\cong 2,241$. Vendas x Mapeamento da Concorrência foi escolhido primeiramente porque é o regressor que é mais altamente correlacionado com as vendas. Os regressores restantes são analisados então para determinar qual, se alguns, é o mais apropriado para a inclusão na etapa seguinte. *Beta In* é o valor do coeficiente estandardizado para o regressor se for incluído em seguida.

Todos os valores significativos são menores que 0.05, assim alguns dos regressores restantes poderiam ser adequados se incluídos no modelo. Para escolher a melhor variável a adicionar ao modelo, olha-se na correlação parcial, que é a correlação linear entre o regressor proposto e a variável dependente após ter sido removido o efeito do modelo atual. Assim, Produtividade de Oportunidade foi escolhido em seguida porque tem a correlação parcial mais elevada.

4.2.3.4-) Variabilidade da Produtividade de Vendas em Função das Correlações

O coeficiente de correlação semiparcial é a correlação entre a variável dependente produtividade de vendas e um regressor parcializado do(s) restante(s) regressor(es). O quadrado desta quantidade dá-nos a proporção da variabilidade da produtividade de vendas explicada exclusivamente pelo regressor em causa, enquanto que o coeficiente de correlação

parcial é a correlação entre duas variáveis quando ambas foram parcializadas de terceiras variáveis, sendo que o quadrado desta quantidade dá-nos a proporção da variabilidade da produtividade de vendas não associada a um regressor que está associada ao outro regressor (LAROSE, 2006). Em outras palavras, responde à questão: “quanto da variância da produtividade de vendas que não é estimada pelas outras variáveis independentes na equação é estimada por esta variável”.

Pode ser dimensionada essa variabilidade (Tabela 4.10) usando os dados de correlações semiparcial e parcial contidos na Tabela 4.7.

Tabela 4.10 – Produtividade de vendas em função das correlações do método seqüencial.

Indicador de Desempenho	Correlações					
	Semiparcial			Parcial		
	sr	sr ²	%	pr	pr ²	%
CO1 Venda x Mapeamento da Concorrência	0,279	0,0778	7,8	0,360	0,130	13,0
GP5 Produtividade de Oportunidades	0,238	0,0566	5,7	0,312	0,097	9,7
GP1 Produtividade de Visita	0,159	0,0253	2,5	0,214	0,046	4,6
EO1 Assertividade de Entrega e Cadastro de Contratos	0,193	0,0372	3,7	0,258	0,067	6,7
GP2 Eficácia de Visitas	0,189	0,0357	3,5	0,254	0,065	6,5
IA2 Volume de Vendas Com Descontos	0,133	0,0177	1,8	0,182	0,033	3,3

Fonte: Elaborado pelo autor a partir dos dados do SPSS.

Conclusão: o modelo *Stepwise* é gerencialmente adequado.

4.3-) Teste F Generalizado do Modelo *Stepwise* em Relação ao Modelo Combinatório

Ao construir o modelo *Stepwise* a partir do modelo Combinatório, retirando onze variáveis independentes, verifica-se que a variação total da variável endógena do modelo que é explicada pela função de regressão diminui. Ou seja, a soma do quadrado dos erros no modelo *Stepwise* é sempre superior à soma do quadrado dos erros no modelo Combinatório, ou $SSE_s > SSE_c$.

Após o teste t sugerir as variáveis independentes a serem usadas na equação, é importante examinar se a variável dependente pode ser explicada pelas variáveis sugeridas tão adequadamente quanto por todas as variáveis. Para isto, testam-se as hipóteses:

$$H_0 : \beta_{q+1} = \beta_{q+2} = \dots = \beta_{q+p} = 0, q < p;$$

$$H_1 : \beta_k \neq 0, \text{ para algum } k=q+1, \dots, p.$$

em que q representa os coeficientes não usados na equação.

O teste F generalizado pretende determinar se o acréscimo de SSE resultante da redução do modelo é ou não significativo. Se não for significativo, então de fato as variáveis excluídas não contribuíam significativamente para diminuir a variação não explicada pelo modelo. Para a realização do teste F generalizado, foi utilizada a seguinte estatística:

$$F = \frac{SSE_s - SSE_c}{GLE_s - GLE_c} \bigg/ \frac{SSE_c}{GLE_c}. \quad (4.7)$$

Em que GLE_s e GLE_c representam os graus de liberdade associados ao erro no modelo Stepwise e no modelo Combinatório, respectivamente. Esse valor é depois comparado com o valor de $F(1-\alpha; GLE_s-GLE_c; GLE_c)$, e se for inferior, então deve ser aceito a hipótese de que não existem diferenças significativas entre SSE_s e SSE_c .

A partir das Tabelas 4.6 e 4.9 (ANOVA), obtidas por intermédio do SPSS, e já apresentadas, pode-se calcular:

$$F = \frac{120,488 - 115,253}{245 - 234} \bigg/ \frac{115,253}{234} = 0,966. \quad (4.8)$$

Se $F \leq F(p-q; n-p-1)$, o teste não rejeita H_0 ; caso contrário o teste rejeita H_0 em favor de H_1 . A aceitação de H_0 indica que a variação da variável dependente é tão adequadamente explicada como o conjunto de todas as variáveis independentes (CHATTERJEE & PRICE, 1977). Da tabela da estatística F pode ser verificado que $F_{(95\%; 11; 245)} \cong 1,85$. Como $0,966 < 1,85$, então não existem diferenças significativas entre os dois modelos, ou seja, as variáveis retiradas do modelo completo não contribuíam significativamente para a redução da variabilidade não explicada (medida pelo SSE).

Conclusão: o modelo *Stepwise* é o mais apropriado para ser utilizado pela gerência na explicação da Produtividade de Vendas com os Indicadores de Desempenho, estando o mesmo com características gerenciais adequadas para a finalidade que se propôs.

4.4-) Interpretação dos Coeficientes do Modelo Seqüencial

Os valores dos coeficientes das variáveis podem ser interpretados da seguinte forma:

a-) O valor do coeficiente ajustado para a variável COI (0,299) significa que mantendo todas as outras variáveis constantes, para cada venda realizada com mapeamento da

concorrência em relação ao total de vendas, a produtividade aumenta, em média, aproximadamente R\$ 299,00 sempre que o volume de negócios aumentar R\$ 1.000,00 normatizados. Por outro lado, 7,8% da variabilidade das vendas deve-se exclusivamente a CO_1 quando as vendas realizadas estão mapeadas da concorrência e que as não explicadas pelas outras variáveis independentes, CO_1 explica 13%;

b-) O valor do coeficiente ajustado para a variável GP_5 (0,254) significa que mantendo todas as outras variáveis constantes, para cada venda realizada com mapeamento de oportunidades em relação ao total de visitas com oportunidades identificadas, à produtividade aumenta, em média, aproximadamente R\$ 254,00 sempre que o volume de negócios aumentar R\$ 1.000,00 normatizados. Por outro lado, 5,7% da variabilidade das vendas deve-se exclusivamente a GP_5 quando a equipe de consultores fecham negócios nas oportunidades prospectadas e que as não explicadas pelas outras variáveis independentes, GP_5 explica 9,7%;

c-) O valor do coeficiente ajustado para a variável GP_1 (0,168) significa que mantendo todas as outras variáveis constantes, para cada visita realizada com oportunidades identificadas em relação ao total de visitas, a produtividade aumenta, em média, aproximadamente R\$ 168,00 sempre que o volume de negócios aumentar R\$ 1.000,00 normatizados. Por outro lado, 2,5% da variabilidade das vendas deve-se exclusivamente a GP_1 quando a equipe de consultores identificam oportunidades nas visitas realizadas e que as não explicadas pelas outras variáveis independentes, GP_1 explica 4,6%;

d-) O valor do coeficiente ajustado para a variável EO_1 (0,207) significa que mantendo todas as outras variáveis constantes, para cada contrato cadastrado no prazo em relação ao total de contratos cadastrados, a produtividade aumenta, em média, aproximadamente R\$ 207,00 sempre que o volume de negócios aumentar R\$ 1.000,00 normatizados. Por outro lado, 3,7% da variabilidade das vendas deve-se exclusivamente a EO_1 quando os contratos dos serviços comercializados são entregues e cadastrados no prazo e que as não explicadas pelas outras variáveis independentes, EO_1 explica 6,7%;

e-) O valor do coeficiente ajustado para a variável GP_2 (0,196) significa que mantendo todas as outras variáveis constantes, para cada contrato fechado em relação ao total de visitas realizadas, a produtividade aumenta, em média, aproximadamente R\$ 196,00 sempre que o volume de negócios aumentar R\$ 1.000,00 normatizados. Por outro lado, 3,5% da variabilidade das vendas deve-se exclusivamente a GP_2 quando os contratos dos serviços comercializados são coerentes com as visitas realizadas e que as não explicadas pelas outras variáveis independentes, GP_2 explica 6,5%;

f-) Finalmente, o valor do coeficiente ajustado para a variável IA_2 (0,137) significa que mantendo todas as outras variáveis constantes, para cada venda realizadas com customização em relação ao total de vendas, a produtividade aumenta, em média, aproximadamente R\$ 137,00 sempre que o volume de negócios aumentar R\$ 1.000,00 normatizados. Por outro lado, 1,8% da variabilidade das vendas deve-se exclusivamente a IA_2 quando os serviços comercializados são menores que 30% de desconto e que as não explicadas pelas outras variáveis independentes, IA_2 explica 3,3%.

4.5-) Análise dos Pressupostos do Modelo Seqüencial

O método de regressão linear múltipla, apesar de não ser muito complexo de implementar, está estruturado num conjunto de hipóteses fundamentais que nem sempre se verificam na prática, quais sejam: a ausência de multicolinearidade, a homoscedasticidade, a normalidade da distribuição dos termos de erros e a ausência de autocorrelação nos erros. Nesta secção, analisa-se a natureza e as conseqüências práticas da verificação ou não desses pressupostos.

4.5.1-) Multicolinearidade

Comparando o modelo Combinatório com o *Stepwise*, se verifica que a inclusão ou exclusão de variáveis não altera significativamente os coeficientes ajustados b_i para a função de regressão. Por si só, esta observação é um sinal da não existência de multicolinearidade, uma vez que quando as variáveis independentes do modelo estão correlacionadas, os coeficientes ajustados variam consideravelmente como conseqüência da introdução ou exclusão de variáveis no modelo.

Por outro lado, fazendo-se análise da matriz de correlação entre as variáveis do modelo, pode ser verificado que, entre as variáveis independentes, não existem valores superiores a 0,5, conforme Tabela 4.11, obtida a partir do *SPSS* (NORUSIS, 2004a), apresentada em seguida.

Tabela 4.11 – Correlações de Pearson do método seqüencial.

		Venda x Mapeamento da Concorrência	Assertividade de Entrega e Cadastro de Contratos Fixa	Produtividade de Visita	Produtividade de Oportunidades	Eficácia de Visitas	Volume de Vendas Customizadas Fixa	Produtividade de Vendas Fixa
Venda x Mapeamento da Concorrência	Pearson Correlation Sig. (2-tailed) N	1 252	,228** ,000 252	,223** ,000 252	,259** ,000 252	,097 ,125 252	,114 ,070 252	,484** ,000 252
Assertividade de Entrega e Cadastro de Contratos Fixa	Pearson Correlation Sig. (2-tailed) N	,228** ,000 252	1 252	,246** ,000 252	,256** ,000 252	,069 ,278 252	-,006 ,921 252	,394** ,000 252
Produtividade de Visita	Pearson Correlation Sig. (2-tailed) N	,223** ,000 252	,246** ,000 252	1 252	,157* ,013 252	,151* ,017 252	,089 ,161 252	,368** ,000 252
Produtividade de Oportunidades	Pearson Correlation Sig. (2-tailed) N	,259** ,000 252	,256** ,000 252	,157* ,013 252	1 252	,160* ,011 252	,001 ,993 252	,443** ,000 252
Eficácia de Visitas	Pearson Correlation Sig. (2-tailed) N	,097 ,125 252	,069 ,278 252	,151* ,017 252	,160* ,011 252	1 252	-,148* ,019 252	,285** ,000 252
Volume de Vendas Customizadas Fixa	Pearson Correlation Sig. (2-tailed) N	,114 ,070 252	-,006 ,921 252	,089 ,161 252	,001 ,993 252	-,148* ,019 252	1 252	,156* ,013 252
Produtividade de Vendas Fixa	Pearson Correlation Sig. (2-tailed) N	,484** ,000 252	,394** ,000 252	,368** ,000 252	,443** ,000 252	,285** ,000 252	,156* ,013 252	1 252

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Fonte: Relatório do SPSS.

Ou seja, não existindo nenhum valor $r_{X_i X_j} \geq 0,5$, pode ser concluído pela inexistência de multicolinearidade (NORUSIS, 2004b).

O SPSS fornece ainda outro indicador, mais formal, da existência de multicolinearidade: o Fator de Inflação da Variância (FIV). O FIV é uma medida do grau em que a correlação entre as variáveis independentes faz inflacionar a variância associada à distribuição dos coeficientes ajustados da regressão.

$$FIV_i = (1 - R_i^2)^{-1} \quad i = 1, 2, \dots, p - 1. \quad (4.9)$$

Em que R_i^2 corresponde ao coeficiente de determinação múltipla de um modelo de regressão linear e a variável X_i é dependente das restantes variáveis X_j . Uma maneira adequada para verificar a colinearidade de duas ou mais variáveis é por meio do valor de Tolerância ou do Fator de Inflação de Variância (*variance inflation factor VIF*) que é o seu inverso. Essas medidas dizem o grau em que cada variável independente é explicada pelas demais variáveis independentes. Pequenos valores de Tolerância ou elevados de FIV denotam colinearidade elevada, sendo usual a utilização de valores de referência para a Tolerância na ordem de 0,10, o que corresponde a um valor de FIV acima de dez. A Tabela 4.12, mostra que os valores do FIV

são bastante próximos de um, pelo que se pode concluir pela não existência de Multicolinearidade.

Tabela 4.12 – Estatísticas de colinearidade do método seqüencial.

Model		Collinearity Statistics	
		Tolerance	VIF
1	Venda x Mapeamento da Concorrência	,869	1,151
	Assertividade de Entrega e Cadastro de Contratos Fixa	,872	1,146
	Produtividade de Visita	,885	1,130
	Eficácia de Visitas	,930	1,076
	Produtividade de Oportunidades	,874	1,145
	Volume de Vendas Customizadas Fixa	,952	1,051

a. Dependent Variable: Produtividade de Vendas Fixa

Fonte: Relatório do SPSS.

Uma outra maneira que ajuda na determinação da ocorrência de problemas com Colinariedade (ou multicolineariedade) é por meio dos autovalores (*eigenvalues*). Os autovalores (*eigenvalues*) fornecem uma indicação de quantas dimensões distintas existem entre as variáveis independentes. Quando diversos *eigenvalues* são perto de zero, as variáveis são altamente intercorrelacionadas e as pequenas mudanças nos valores dos dados podem conduzir a grandes mudanças nas estimativas dos coeficientes (NORUSIS, 2004b).

Os Índices de Condições são as raízes quadradas das relações do autovalor (*eigenvalue*) maior para cada autovalor (*eigenvalue*) sucessivo. Um índice de condições superior a 15 indica um problema possível e um índice superior a 30 sugere um problema sério com colinariedade. As proporções da variância são as proporções da variância da estimativa esclarecida por cada componente principal associado com cada um dos *eigenvalues*. Colinariedade é um problema quando um componente associado com um índice de condições elevado contribui substancialmente à variância de dois ou mais variáveis.

Tabela 4.13 – Diagnóstico de Colinearidade do Método Seqüencial.

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions						
				(Constant)	Venda x Mapeamento da Concorrência	Assertividade de Entrega e Cadastro de Contratos Fixa	Produtividade de Visita	Eficácia de Visitas	Produtividade de Oportunidades	Volume de Vendas Customizadas Fixa
1	1	1,763	1,000	,00	,12	,11	,11	,04	,12	,00
	2	1,157	1,235	,00	,03	,00	,01	,28	,01	,48
	3	1,000	1,328	1,00	,00	,00	,00	,00	,00	,00
	4	,878	1,418	,00	,00	,19	,24	,33	,14	,12
	5	,832	1,456	,00	,13	,26	,31	,07	,22	,07
	6	,716	1,570	,00	,71	,02	,00	,03	,28	,19
	7	,655	1,641	,00	,00	,41	,34	,25	,24	,13

a. Dependent Variable: Produtividade de Vendas Fixa

Fonte: Relatório do SPSS.

O diagnóstico de colinearidade, por meio da Tabela 4.13, confirma que não existe nenhum problema com a multicolineariedade. Diversos *eigenvalues* são muito maiores que zero, indicando que os regressores não são intercorrelacionadas e que apenas na ocorrência de grandes mudanças nos valores dos dados podem conduzir a pequenas mudanças nas estimativas dos coeficientes. Os índices de condições são computados como as raízes quadradas das relações do maior *eigenvalue* para cada *eigenvalue* sucessivo.

4.5.2-) Heteroscedasticidade

A presença de variâncias desiguais (heteroscedasticidade) é uma das violações mais comuns de suposições. Esse diagnóstico pode ser feito com gráficos de resíduos. Para avaliar o comportamento geral dos resíduos, constrói-se um gráfico de resíduos, contendo no eixo vertical os resíduos padronizados (*standardized residual*) e no eixo horizontal os valores previstos padronizados (*standardized predicted value*). Também devem ser construído um diagrama de dispersão para visualizar a relação entre a variável dependente (Produtividade de Vendas) e cada um dos regressores individualmente bem como, constrói-se gráfico de dispersão que apresente os valores observados da variável dependente com os valores preditores (*standardized predicted value*).

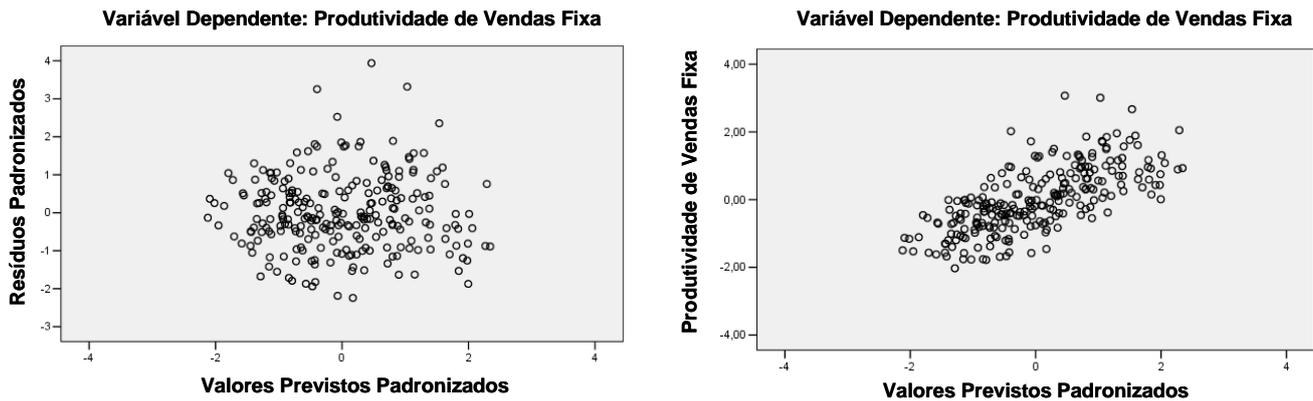


Gráfico 4.1 – Resíduos padronizados & resíduos REFX do método seqüencial.

Fonte: Relatório do SPSS.

O objetivo do Gráfico 4.1 é detectar a existência de uma relação sistemática entre os valores ajustados da variável dependente e os erros. Pela configuração da dispersão de valores, se pode concluir pela existência de homocedasticidade, dado não haver um padrão de variação dos termos de erro ajustados pelo modelo relativamente aos valores ajustados da variável dependente.

Observando o Gráfico 4.2, também se pode concluir pela não existência de uma relação sistemática entre os valores ajustados do termo de erro, e as variáveis independentes do modelo. Mais uma vez, essas observações corroboram a hipótese da não existência de heteroscedasticidade, ou seja, esses gráficos de ajustamento dos resíduos não apresentam nenhum tipo de tendência que faça suspeitar da validade dos pressupostos do modelo de regressão, pois os resíduos distribuem-se segundo uma faixa horizontal em torno do zero, sem denotar qualquer padrão de distribuição.

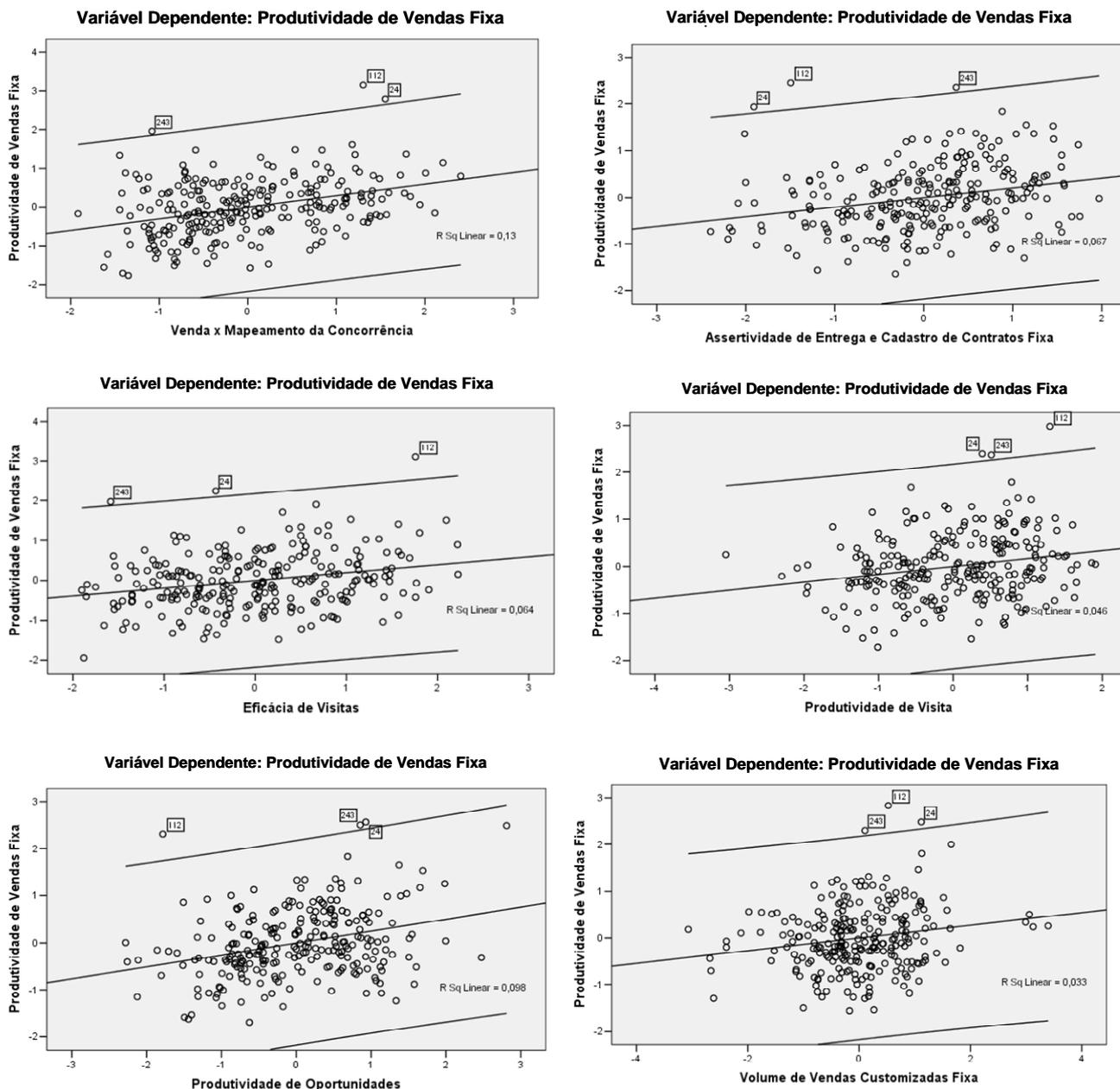


Gráfico 4.2 – Regressões parciais das variáveis independentes do método sequencial.

Fonte: Relatório do SPSS.

A independência dos termos de erro pode ser identificada fazendo o gráfico de resíduos em relação a qualquer variável sequencial possível. Se os resíduos forem independentes, o padrão é aleatório e semelhante ao gráfico nulo de resíduos, pois é assumido em regressão que cada valor previsto é independente, ou seja, não está relacionado com qualquer outra previsão, não sendo seqüenciados por qualquer variável. A homoscedasticidade é uma suposição relacionada primariamente a relações de dependência entre variáveis, referindo-se à suposição de que a variável dependente exibe níveis iguais de variância ao longo do domínio da variável preditora. É desejável a existência de homoscedasticidade porque a variação da variável

dependente sendo explicada na relação de dependência não deve se concentrar apenas em um domínio limitado dos valores independentes.

Nos gráficos das regressões parciais das variáveis independentes mostrados anteriormente, pode ser verificado que nenhum padrão foi encontrado, bem como não existe nenhum tipo de tendência que nos faça suspeitar da validade dos pressupostos do modelo de regressão. Como o padrão verificado parece ser aleatório e semelhante ao gráfico nulo de resíduos, se concluí que os resíduos são independentes, confirmando a suposição da independência dos termos de erro, bem como aponta para a homoscedasticidade no conjunto das variáveis independentes.

4.5.3-) Normalidade

O diagnóstico mais usual para verificação de normalidade da distribuição dos termos de erros é um histograma de resíduos, com uma verificação visual para uma distribuição que se aproxima da normal. Um histograma é uma representação gráfica de uma única variável que representa a frequência (valores dos dados) dentro da categoria de dados. O histograma traduz a distribuição de frequências, sendo possível analisar a simetria e o achatamento da amostra. O histograma é um gráfico que procura refletir a estrutura (forma) da população de onde foi retirada a amostra. O histograma é uma representação constituída por uma sucessão de retângulos (barras) adjacentes em que cada um tem por base um intervalo de classe e por área a frequência relativa (ou absoluta) dessa classe. Desta forma a área total do gráfico deve ser unitária (ou igual à dimensão da amostra).

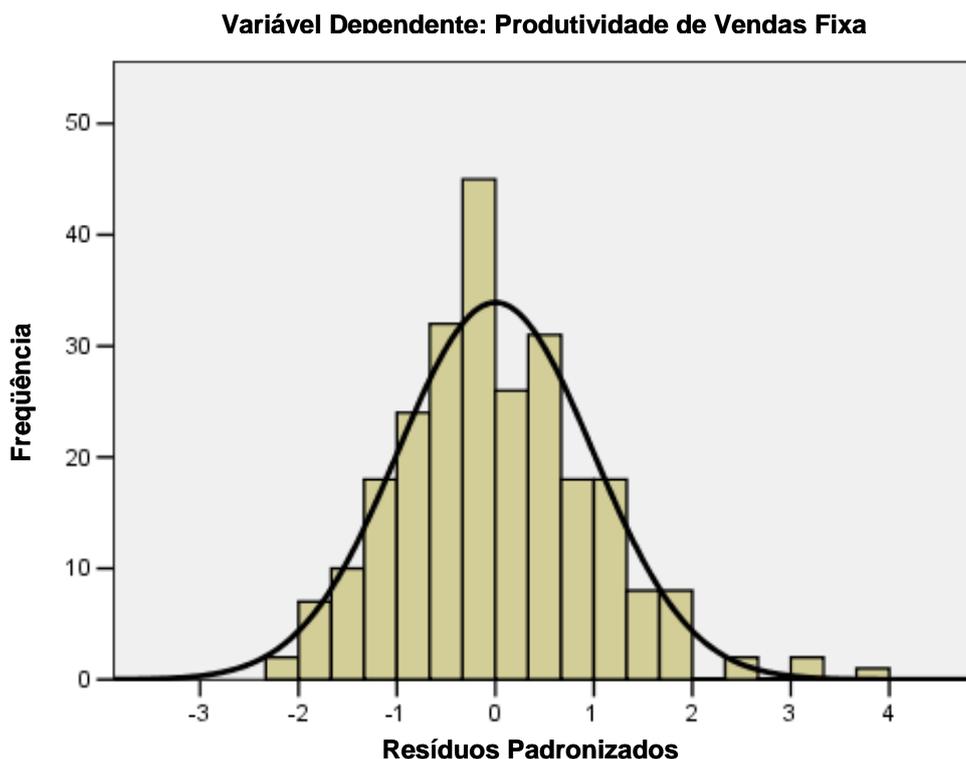


Gráfico 4.3 – Histograma de resíduos da variável dependente do método seqüencial.

Fonte: Relatório do SPSS.

Uma outra abordagem muito utilizada é o gráfico de probabilidade normal, que compara a distribuição cumulativa de dados reais com a distribuição cumulativa de uma distribuição normal. A distribuição normal forma uma reta diagonal e os dados do gráfico são comparados com a diagonal. Se uma distribuição é normal, a linha que representa os dados segue muito próximo à diagonal. Os gráficos de probabilidades (P-P: *Probability Plots*) visualizam graficamente o ajustamento de uma variável a uma função de distribuição de probabilidades. Esse tipo de gráfico representa no eixo horizontal as frequências relativas acumuladas observadas na amostra (*observed cummulative probability*) e, no eixo vertical, a função de distribuição de probabilidades esperada (*expected cummulative probability*). A diagonal do gráfico representa um ajustamento perfeito da amostra à função de distribuição de probabilidades. Quanto mais os pontos se afastam da diagonal, ou se distribuem segundo um determinado padrão, menor é o ajustamento da amostra à distribuição teórica.

No histograma com o ajustamento à distribuição normal que teve o aspecto mostrado no Gráfico 4.3, pode ser verificado que existe uma distribuição consistente dos resíduos em relação à distribuição teórica, ou seja, segue aproximadamente a forma da curva normal. Existe apenas uma elevação um pouco maior, nomeadamente na zona central da distribuição, sendo aceitável perto da curva normal, porém com a existência de pontos residuais positivos relativamente grandes (que é tratado mais adiante por meio do diagnóstico de *Casewise*).

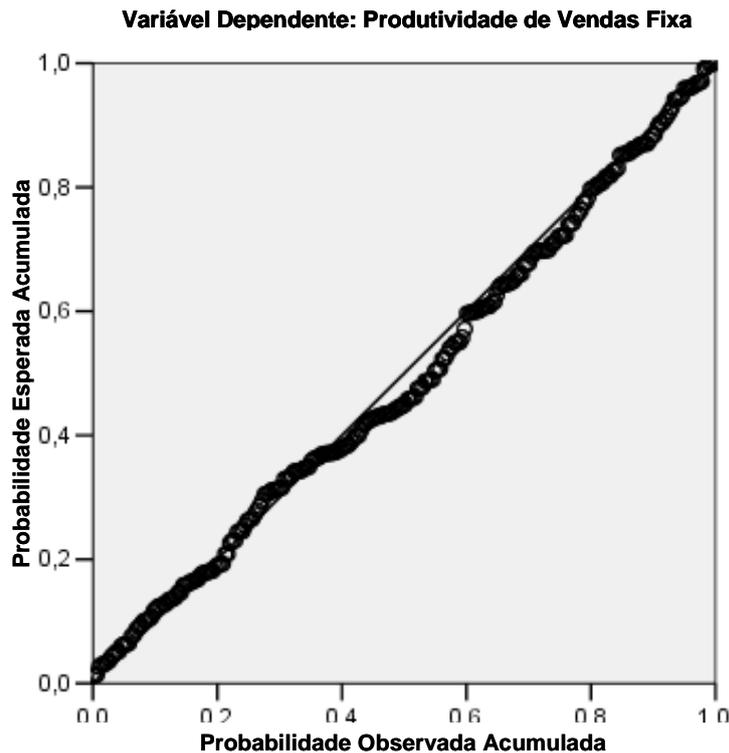


Gráfico 4.4– Probabilidade normal *PP-plot* dos resíduos do método seqüencial.

Fonte: Relatório do SPSS.

Conforme pode ser observado por meio do Gráfico 4.4, nota-se uma tendência de distribuição uniforme, ou seja, os pontos encontram-se bastante próximos de uma reta, sem desvios substanciais ou sistemáticos de modo que não há razão para duvidar da normalidade dos erros, sendo os resíduos considerados representativos de uma distribuição normal. Nem o histograma nem o gráfico P-P indicam que a suposição da Normalidade foi violada.

4.5.4-) Autocorrelação dos Erros

A Autocorrelação ocorre quando os termos de erro não são independentes, ou seja, $Cov(\varepsilon_i, \varepsilon_j) \neq 0$. A autocorrelação entre os resíduos pode ser detectada pelo método gráfico ou por meio do teste *d* de Durbin-Watson. Na análise gráfica por meio do gráfico de resíduos (e_i) versus valores unitários ajustados pelo modelo de regressão (\hat{y}), observou-se que os resíduos estavam distribuídos aleatoriamente em torno da reta não apresentando nenhum padrão definido. Desta forma, houve a necessidade de se aplicar o teste *d* de Durbin-Watson para se detectar a existência de autocorrelação positiva ou negativa. O teste de autocorrelação Durbin-Watson é obtido a partir dos resíduos do modelo no qual são estabelecidos os valores críticos do limite inferior (d_l) e do limite superior (d_u) que é verificado por meio de tabelas em função

de p (regressores), n (número de observações) e α (nível de significância), sendo que: para $d < 2$ temos: se $d < d_i$ rejeita-se H_o , ou seja, que os resíduos aleatórios não são autocorrelacionados e se aceita a autocorrelação, se $d_i < d < d_u$ então o teste é inconcluso e se $d > d_u$ se aceita H_o ; para $d > 2$ temos: se $d < 4 - d_u$ se aceita H_o e se aceita a autocorrelação, se $(4 - d_u) < d < (4 - d_i)$ então o teste é inconcluso e se $d > (4 - d_i)$ rejeita-se H_o .

A existência de autocorrelação foi testada recorrendo ao teste de Durbin-Watson ($p=6$, $n=252$ e $\alpha=5\%$). No modelo gerado a estatística d (Tabela 4.8) estimada foi $d = 1.801 < 2$ e $d_u = 1.757$ enquadrando-se no intervalo de valores associados à aceitação da hipótese nula (então se aceita H_o). Por conseqüência, é de se admitir que a eficiência dos estimadores obtidos pelos métodos dos mínimos quadrados, dada à inexistência de autocorrelação entre eles, ou seja, não há evidência de relação de dependência dos valores dos resíduos aleatórios não sendo, portanto, os mesmos autocorrelacionados.

4.5.5-) Análise dos Pontos Influentes

É comum em trabalho experimental que se confronte com situações em que, ao obter ou analisar um conjunto de dados, se depare com um ou mais valores que aparentemente diferem razoavelmente dos outros. Esses valores produzem por vezes, dependendo da amplitude do seu afastamento, conclusões errôneas e a distorção de parâmetros. As suas causas podem ser variadas, como sejam erros humanos ou instrumentais, ou outro tipo de situações anômalas. Da sua identificação depende muitas vezes a validade das conclusões que são obtidas. A média e o desvio padrão dependem da remoção ou não desses valores, e uma vez que a discussão sobre a acurácia e precisão dos dados depende desses parâmetros, torna-se evidente o cuidado a ter quanto à decisão relativamente à sua eliminação ou não, devendo sempre ser fundamentada a opção tomada. A eliminação desses valores pode igualmente incorrer num erro, como, por exemplo, sobreestimar a precisão dos dados, ou aceitar um modelo que não é válido, sendo esse pressuposto a base da classificação do *outlier* (HAIR, J. F. et al., 2005).

Uma grande vantagem do *boxplot* é a de permitir visualizar informações importantes sobre a forma dos dados. Para além de todas as características já mencionadas, o *boxplot* revela-se excepcionalmente útil na comparação de várias amostras. A representação de *boxplots* paralelos (vários *boxplots* representados num só gráfico) pode facilitar a comparação de amostras ou de coleções de dados. Esta representação permite formular conjecturas acerca das semelhanças ou diferenças com respeito às medidas de localização e de dispersão já mencionadas.

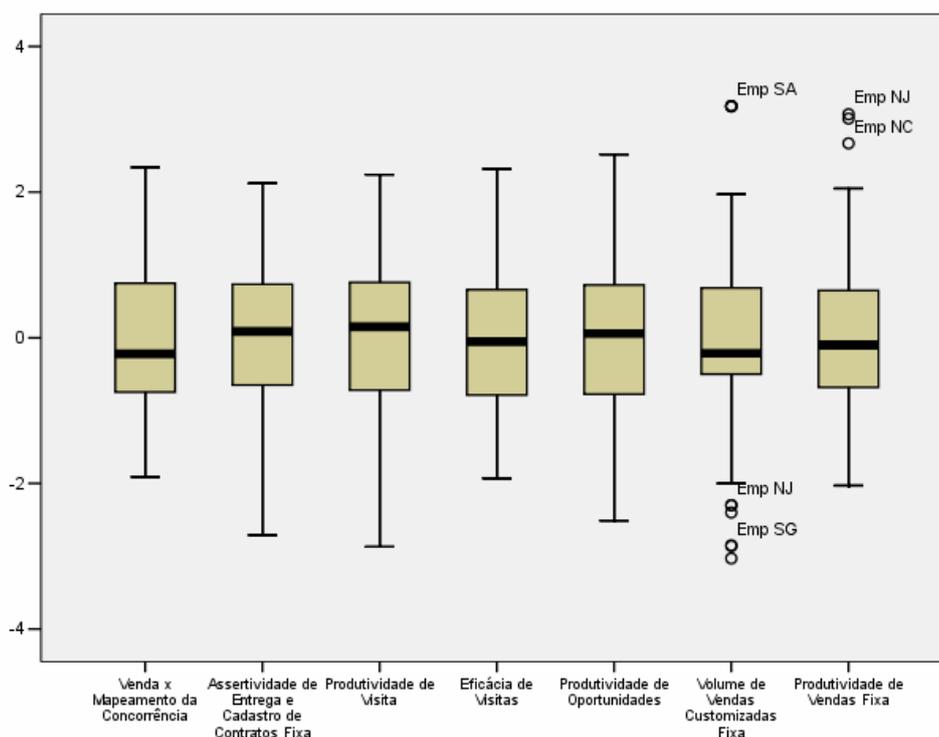


Gráfico 4.5 – *Boxplot* do método seqüencial.

Fonte: Relatório do SPSS.

Eventuais assimetrias podem ser facilmente visualizadas por meio do deslocamento da mediana para um dos lados da caixa e por meio de uma diferença significativa no comprimento da caixa. Também a presença de *outliers* apenas para um dos lados da caixa pode ser indicativa de assimetria. Assim, no Gráfico 4.5 pode ser observado pequenas assimetrias tanto negativas quanto positivas (concentração de observações elevadas) já que a mediana está mais próxima do terceiro quartil do que do primeiro ou do primeiro quartil do que do terceiro, a cerca superior ou inferior é mais curto que o inferior ou superior, porém dentro dos parâmetros aceitáveis e existem cinco *outliers* na variável independente Volume de Vendas Com Descontos, sendo dois da empresa *NJ* e dois da empresa *SG* para além da cerca inferior e um da empresa *SA* para além da cerca superior e três na variável dependente Produtividade de Vendas, sendo todos além da cerca superior com duas participações da empresa *NJ* e um da empresa *NC*, além de cinco outros na variável independente Volume de Vendas Com Descontos tendo sido causado por guerra de preço provocado pela concorrência.

A Tabela 4.14 indica as estatísticas para os casos individuais que são identificados por seu número do caso. As estatísticas são indicadas para casos com residuais padronizados mais do que um número especificado de desvios padrão longe da média. Os resíduos padronizados

são resíduos ordinários divididos pelo desvio padrão da amostra dos resíduos. Os resíduos padronizados têm a média 0 e um desvio padrão 1.

Tabela 4.14 – Diagnóstico de Casewise da regressão do método seqüencial.

Case Number	Std. Residual	Produtividade de Vendas Fixa	Predicted Value	Residual
24	3,318	3,01	,6831	2,32691
112	3,938	3,07	,3081	2,76187
243	3,254	2,02	-,2622	2,28216

a. Dependent Variable: Produtividade de Vendas Fixa

Fonte: Relatório do SPSS.

Os valores observados da variável dependente são indicados, onde se pode também ver o valor predito e o valor residual para cada caso. O valor predito é o valor para o caso predito pelo modelo. O residual é a diferença entre o valor observado da variável dependente e o valor predito. Os casos observados acima ocorreram na empresa *NB* no mês de Dez/06 (caso 24), na empresa *NJ* no mês de Abr/06 (caso 112) e na empresa *SDT* no mês de Mar/06 (caso 243), sendo valores válidos, porém fora da curva.

4.6 -) Verificação dos Testes de Hipóteses

O estudo de viabilidade de uso de análise de regressão linear múltipla situa-se no domínio comercial do setor corporativo de telecomunicações, utilizando-se medição de desempenho para a identificação de relações entre parâmetros de vendas e de indicadores de desempenho. Dentro desse contexto, a questão de pesquisa da presente dissertação foi: “*Existe relação de dependência entre os indicadores de desempenho e a produtividade de vendas? É possível determinar a importância desses indicadores, os tipos de relações eventualmente encontradas ou as inter-relações entre eles?*”.

Para responder essas questões foi utilizado o processo de Mineração de Dados por meio da técnica de *Regressão Linear Múltipla* envolvendo uma única variável dependente que foi a produtividade de vendas e um conjunto de variáveis independentes, quais sejam: os indicadores de desempenho. Para tanto, foram utilizados o Método de Especificação Confirmatória, o Método Combinatório (no SPSS é denominado de *Enter*) e o Método Seqüencial (no SPSS é denominado de *Stepwise*) por meio da estimação dos parâmetros, do coeficiente de determinação múltiplo, da análise de variância e dos testes individuais dos parâmetros.

Tabela 4.15 – Comparativo dos métodos testados.

	Métodos Testados		
	Especificação Combinatória (Enter)	Especificação Confirmatória	Especificação Sequencial (Stepwise)
Parâmetros do Modelo	REFXi = 0,289 CO1i + 0,204 EO1i + 0,010 EO4i - 0,028 EO5i - 0,024 EO6i - 0,065 EO7i + 0,030 FC1i + 0,067 FC2i - 0,085 FC3i + 0,184 GP1i + 0,161 GP2i + 0,010 GP3i + 0,062 GP4i + 0,234 GP5i + 0,132 IA2i + 0,026 IA3i - 0,059 IA4i	REFXi = 0,293 CO1i + 0,198 EO1i + 0,060 FC1i + 0,166 GP1i + 0,203 GP2i + 0,251 GP5i + 0,131 IA2i	REFXi = 0,299 CO1i + 0,207 EO1i + 0,168 GP1i + 0,196 GP2i + 0,254 GP5i + 0,137 IA2i
Coefficiente de Determinação Múltiplo	0,501	0,481	0,478
Teste F	13,807 >> 1,70	32,368 >> 2,05	37,406 >> 2,10
Teste Estatístico t	0,188 < 0,196 < 0,493 < 0,513 < 0,565 < 0,587 < 1,006 < 1,106 < 1,189 < 1,276 < 1,588 < 2,241 < 2,679 < 2,796 < 3,536 < 3,945 < 4,541 < 5,605	1,263 < 2,241 < 2,749 < 3,390 < 3,984 < 4,221 < 5,088 < 5,908	t ⁱ * ≥ 2,241 (i = 1, ..., 6)
Variáveis Excluídas	Nenhuma	FC1	EO4, EO5, EO6, EO7, FC1, FC2, FC3, GP3, GP4, IA3 e IA4
Quantidade de Variáveis na Equação	17	7	6
Validação da Hipótese H ₀	Não	Não	Não
Validação da Hipótese H ₁	Sim	Sim	Sim

Fonte: Elaborada pelo autor a partir dos dados do SPSS.

O teste de hipótese é uma regra usada para decidir se uma hipótese estatística deve ser rejeitada ou não. O objetivo do teste de hipótese é decidir se uma hipótese sobre determinada característica da população é ou não apoiada pela evidência obtida de dados amostrais. Os testes de hipóteses necessários em uma análise de regressão foram os seguintes: teste de hipótese para a significância do modelo, teste de hipótese para o parâmetro β_i e o teste de hipótese para um subconjunto de parâmetros (NETER & WASSERMAN, 1974). Conforme pode ser verificado por meio da Tabela 4.15, as duas hipóteses levantadas inicialmente foram testadas utilizando-se ferramentas estatísticas adequadas a esse tipo de pesquisa. As hipóteses identificadas, cujo detalhamento encontra-se no capítulo 3 e que foram agora analisadas, são as seguintes:

1-) H₀: Não existe relação entre as características da produtividade de vendas e os indicadores de desempenho de vendas. Por intermédio do Método Combinatório (Tabela 4.15) pode ser verificado que as informações disponibilizadas pelo coeficiente de determinação múltiplo, Teste F e o teste estatístico t comprovam a não validação desta hipótese tendo em vista que os coeficientes

$\beta_1, \beta_2, \beta_{10}, \beta_{11}, \beta_{14}$ e β_{15} obtiveram valores de significância $sig < 0,05$ sendo, portanto significantes, caracterizando o relacionamento dos indicadores de desempenho de vendas associados a esses coeficientes com a produtividade de vendas;

2-) **H1:** Existe relação entre as características da produtividade de vendas e as variáveis de desempenho de vendas das empresas. Por meio do Método Confirmatório (Tabela 4.15) pode ser verificado que as informações disponibilizadas pelo coeficiente de determinação múltiplo, Teste F e o teste estatístico t comprovam a não validação desta hipótese tendo em vista que o coeficiente β_3 obteve um valor de significância $sig > 0,05$ não sendo, portanto estatisticamente significativa. Esta exclusão da variável $FC1$, a qual está associada ao coeficiente β_3 , caracteriza a não confirmação das variáveis, especificadas na equação de regressão linear múltipla, definidas pelos especialistas de vendas de telecomunicações. Por meio do Método Seqüencial (Tabela 4.15) pode ser verificado que as informações disponibilizadas pelo coeficiente de determinação múltiplo, Teste F e o teste estatístico t comprovam também a validação desta hipótese tendo em vista que os coeficientes $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ e β_6 obtiveram valores de significância $sig < 0,05$ sendo, portanto significativos, caracterizando o relacionamento dos indicadores de desempenho de vendas associados a esses coeficientes com a produtividade de vendas. Porém esta relação é diferente da especificada pelos especialistas de vendas em serviços de telecomunicações.

Conclusões e Recomendações

5.1 Conclusões

Esse trabalho consistiu em utilizar técnicas de mineração de dados (MD) no tratamento de um banco de dados histórico proveniente da medição de desempenho organizacional na área de vendas corporativas. A proposta foi a de estabelecer os relacionamentos entre as medidas de desempenho já existentes e descobrir novos relacionamentos entre elas.

Teve como motivação o fato do sistema de gestão das empresas e, conseqüentemente, o sistema de medição de desempenho, precisar refletir as mudanças ocorridas no ambiente industrial nas últimas décadas. Um desses principais reflexos foi a necessidade do sistema de medição de desempenho se tornar multidimensional e não mais somente voltado para a avaliação dos aspectos financeiros e de produtividade. Por isso, torna-se mais complexo o entendimento de como os relacionamentos entre as medidas de desempenho afetam as decisões organizacionais.

Assim, foi delineado um método baseado no emprego de regressão linear múltipla para analisar o efeito que os indicadores de desempenho de vendas poderiam ter sobre o desempenho da produtividade de vendas. Para tanto se utilizou equações de regressão cujos parâmetros de desempenho pudessem ser favoráveis à formação de uma equação de produtividade de vendas. Mediante análises estatísticas e comerciais criteriosas, as equações foram definidas, sendo os seus respectivos coeficientes de determinação ajustados. Foram realizados testes de hipóteses dos principais parâmetros, visando à validação ou não dos modelos de regressão e a análise da qualidade de seus ajustes.

A pesquisa permitiu verificar que existe relação entre as características da produtividade de vendas e as variáveis de desempenho. Constatou-se que a relação entre a produtividade de vendas e os indicadores de desempenho acontece diferente do modelo atualmente utilizado. A variável independente *Média de Visitas por Consultor*, por exemplo, foi excluída do modelo gerado por não ter sido considerada estatisticamente significativa para a composição da variável dependente Produtividade de Vendas.

Os resultados mostram que, dos dezessete indicadores de desempenho de vendas utilizados pelo modelo atual em execução, seis desses indicadores foram significativos na explicação da produtividade de venda baseada no tratamento dos dados históricos, isto é, no modelo gerado pela pesquisa. Os coeficientes de todas as variáveis foram positivos, indicando que o melhor desempenho de um dos fatores significativos conduz a um maior volume de negócio realizado. O modelo proposto revelou bom ajuste e com boa capacidade de explicar o comportamento do volume de vendas realizado.

A variável independente *Vendas versus Mapeamento da Concorrência* foi, em relação a produtividade de vendas, que apresentou os resultados mais importantes da pesquisa, para o conjunto das amostras utilizadas. Essa variável mostrou uma correlação positiva significativa ao nível de 5% (valor da probabilidade associada à estatística *t*). Os resultados foram consistentes em todos os casos em que existe significância, apresentando uma forte evidência de que a empresa deste setor que atua com foco no mapeamento da concorrência e na produtividade de oportunidades poderá ter melhor desempenho comercial.

Os resultados obtidos permitem sugerir que as empresas devem adotar um planejamento de vendas com ênfase nos seguintes principais indicadores de desempenho: Vendas versus Mapeamento da Concorrência, Produtividade de Oportunidades, Conversão de Contratos, Eficácia de Visitas, Produtividade de Visitas e Volume de Vendas com Descontos. Também, os parâmetros individuais do modelo, validados pela hipótese referenciada, caracterizam que o mapeamento da concorrência deve ser mais eficaz do que realizar vendas com preços com aplicação da política de descontos por alçada em relação aos valores nominais (o que caracteriza o indicador de desempenho *Volume de Vendas com Descontos*).

Em resumo pode-se concluir que:

- O modelo usado foi adequadamente especificado, isto é, tem a forma funcional correta;
- Não houve omissão de variáveis relevantes nem inclusão de variáveis irrelevantes;
- Teve uma estrutura de parâmetros tão simples quanto possível;

- A Produtividade de Vendas pode ter dependência com os Indicadores de Desempenho de Vendas e ser representada pela equação de regressão linear múltipla: $REFX = 0,299 COI + 0,254 GP_5 + 0,168 GP_1 + 0,207 EO_1 + 0,196 GP_2 + 0,137 IA_2$;
- Revelou boa qualidade de ajustamento e significado (o coeficiente de determinação foi relativamente elevado para a tarefa de explicação e os coeficientes ajustados tiveram os sinais corretos e uma magnitude realista);
- O Mapeamento da Concorrência sinalizou ser bem mais eficaz do que realizar Vendas com Descontos;
- A exclusão da variável independente Média de Visitas por Consultor pode caracterizar que as visitas realizadas pelos consultores estão sendo feitas, em sua maioria, sem um planejamento adequado, não estando sendo caracterizados os preceitos definidos pela empresa operadora.
- A sistematização desenvolvida **MDAV - Mineração de Dados para Administração de Vendas**, teve aplicação prática podendo ser viabilizada a sua utilização em diversas áreas de gestão de negócios;
- O estudo de caso realizado explorou a validade do modelo, principalmente para as atividades relacionadas ao processo de gestão de melhoria na produtividade de vendas.

5.2 Sugestões e Recomendações

Novos estudos podem ser realizados, com o objetivo de melhorar a análise sobre a relação entre os indicadores de desempenho e o desempenho comercial das empresas. Por exemplo, sugere-se utilizar testes estatísticos alternativos, por exemplo, um modelo de análise conjunta. Ou, ainda, incluir outras empresas ou utilizar um horizonte temporal diferente do utilizado nesta pesquisa. Recomenda-se ainda a aplicação experimental de outros algoritmos para mineração dos dados selecionados para avaliação de desempenho e compatibilidade com o objetivo da aplicação, dando continuidade à pesquisa. Além disso, seria importante que os próximos estudos sejam focados em:

- Análise de agrupamentos para verificação de empresas e/ou grupos de serviços na cesta de produtos, nos segmentos fixo e móvel, disponibilizados para os Clientes;
- Previsão de vendas no setor corporativo de telecomunicações, utilizando indicadores de desempenho nos segmentos de mobilidade e fixo, por meio de outras técnicas de Mineração de Dados;

- Análise do relacionamento, no segmento de mobilidade, entre os indicadores de desempenho com as vendas realizadas;
- Determinação do tamanho da força de vendas adequado ao desafio de vendas proposto.

Referências

- ABREU, A. L. T.; MARTINS, R. A. *Construção de relacionamento entre as medidas de desempenho: uma análise de literatura*. In: Encontro Nacional de Engenharia de Produção, 13, 2003. Anais ... Ouro Preto – MG: ABEPRO/USP, 2003.
- ADRIAANS, Pieter; Zanting, Dolf; “*Data Mining*”, Addison-Wesley, England, 1996.
- ANUÁRIO TELECOM. Disponível em: <http://www.anuariotelecom.com.br>. Acessado em Fevereiro de 2007.
- ARNETT, Dennis B; Menon, Anil, Wilcox, James B.; “*Using Competitive Intelligence: Antecedents and Consequences*”, *Competitive Intelligence Review*, Vol. 11(3), 2000.
- ATKINSON, A. *The strategy-focused organization*. *CMA Management*, Hamilton, Vol. 75, No. 2, Abr. 2001. ISSN: 12075183.
- BARANAUSKAS J A.; Monard M C.; “*Combining Symbolic Classifiers from Multiple Inducers*”; *Knowledge Based Systems*; Vol. 16, Elsevier Science, 2003.
- BECHER, J.D.; Berkhin, P.; Freeman, E. *Automating Exploratory Data Analysis for Efficient Data Mining*. In: KDD 2000, Boston MA USA, 2000.
- BERVIAN, Amado L., CERVO, Pedro A. *Metodologia Científica*. 5. ed. São Paulo: Prentice Hall, 2003.
- BITITCI, U. S; Carrie, A. S.; McDevitt, L. *Integrated performance measurement systems: a development guide*. *International Journal of Operations & Production Management*, V17, No 5. 1997.
- BONOMA, Thomas; Shapiro, Benson P. *Sucesso e marketing industrial: a obtenção de lucros através da racionalização do mercado*. São Paulo: Harbra, 1991.
- BOYD, L. H.; COX III, J. F. *A cause-and-effect approach to analyzing performance measures*. *Production and Inventory Management Journal*, Vol. 38, No.3, 3rd quarter 1997.
- BRADLEY, P., U. Fayyad, & O. Mangasarian. *Data mining: Overview and optimization opportunities*. Technical Report MSR-TR-98-04, Microsoft Research Report, Redmond, WA.1998.
- BROWN, S.; “*Data quality: Relatively critical and critically relative*”; *DM Review*; 2002.
- CABENA, P.; Hadjinian, P.; Stadler, R.; Verhees, J.; Zanasi, A. “*Discovering Data Mining – From Concept to Implementation*”, Prentice-Hall, 1998.
- CARDOZO, Richard N. *Situational Segmentation of Industrial Markets*. *European Journal of Marketing*. Bradford, Vol. 14, No. 5/6, 1980.

- GARDNER, S. R. (1998). *Building the data warehouse*. Communications of the ACM 41(9)
- CARVALHO, Luís Alfredo Vidal de. *Data Mining: A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração*. Rio de Janeiro, editora Ciência Moderna, 2005.
- CERTO, S. C. e PETER, J. P. *Administração estratégica: planejamento e implantação da estratégia*. São Paulo: Makron Books, 1993.
- CHAPMAN, Pete; Clinton, Julian; Kerber, Randy; Khabaza, Thomas; Reinartz, Thomas; Shearer, Colin; Wirth, Rüdiger; “*CRISP-DM 1.0 – Step-by-Step data mining guide*”; CRISP-DM Consortium, 2000
- CHATTERJEE, S.; PRICE, B. *Regression Analysis by Example*. USA: John Wiley & Sons, 1977.
- CHEN, Ming-Syan; Han, Jiawei; YU, Philip S. *Data Mining: An Overview from Databases Perspective*. Disponível: <http://citeseer.nj.nec.com/5126.html> (acessado em 25/11/2006), 1996.
- CHURCHILL, A. G.; Ford, N. M.; Walker, O. C.; Jonhston, M. W.; Tanner, J. F. *Sales Force Management*. Boston: Irwin McGraw-Hill, 6ª edição. 2000.
- COBRA, M. *Administração de Vendas*. São Paulo: Atlas, 1994.
- CODY, W. F., J. T. Kreulen, et al. (2002). "The integration of business intelligence and knowledge management". *IBM Systems Journal* 41(4).
- COOKE, Simon; “*Database Marketing: strategy or tactical tool?*” *Marketing Intelligence & Planning*, Vol 12, No 6, 2000.
- CRAVENS, D. W.; Marshal, G. W.; Lassk, F. G.; Low, G. S. *The control factor. Marketing Management*. Jan/Fev, 2004.
- CROSS, K. F.; LYNCH, R. L. *Managing the corporate warriors*. *Quality Progress*, Vol. 23, No. 4, apr. 1990.
- DANIEL, C.; WOOD, T. E. *Fitting Equations to Data*. New York: John Wiley & Sons, Inc, 1971.
- DAVIS, A.; O’DONNELL, L. *Modeling complex problems: system dynamics and performance measurement*. *Measurement Accounting*, Maio 1997.
- DRAPER, N. R. & SMITH, H. *Applied Regression Analysis*. New York: John Wiley & Sons, Inc, 1981.
- DROZDENKO, Ronald; Drak Perry D, “*Optimal Database Marketing*”, SAGE Publications, Thousand Oaks, USA, 2002
- DUBINSKY, A. J.; HANSEN, R. W. *The salesforce management audit*. *California Management Review*, V.XXIV, No 2, Winter, 1981.

DWYER, Hill. J.; Martin, W. *An empirical investigation of critical success factors in the personal selling process for homogeneous goods*. Journal of Personal Selling & Sales Management, Vol. XX, No. 3, Summer, 2000.

ECCLES, R. G. *The Performance Measurement Manifesto*. Harvard Business Review. jan./fev., 1991.

ECCLES R. G.; PYBURN P. J. *Creating a comprehensive system to measure performance*. Management Accounting, Vol. 74, No. 4, 1992.

EGAN, J. (1975), *Signal detection theory and ROC analysis*, Academic Press, New York.

EPSTEIN, M.; MANZONI, J.F. *Implementing corporate strategy: from Tableaux de Board to Balanced Scorecards*. European Management Journal, Vol. 16, No.2, abril de 1998.

EVANS, H.; Sahworth, G.; Gooch, J.; Davies, R. *Who needs performance management?* Management Accounting – CMA, Vol.74, No. 11, 1996.

EURIDITIONHOME. *Data Mining Tutorials*, Resources. Disponível em: <http://datamining.eruditionhome.com>. Acessado em Janeiro de 2007.

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

FEELDERS A.J.; “*Statistical Concepts*”; In Berthold, M; Hand D.J. (Eds); *Intelligent Data Analysis: An Introduction* 2ªed; Springer-Verlag; 2002.

FELDENS, Miguel Artur; Moraes, R.L.; Pavan, A.; Castilho, J.M.V. *Towards a methodology for the discovery of useful knowledge combining data mining, data warehousing and visualization*. In: XXIV CLEI (Conferência Latino-Americana de Informática). Quito, Equador. Disponível: <http://jacui.inf.ufrgs.br/~FELDENS/clei98.html> (acessado em 22/10/2006), 1998.

FERRAZ, J. C.; Kupfer, D.; Haguenaer, L. *Competitividade, padrões de concorrência e fatores determinantes*. In: *Made in Brazil: desafios competitivos para a indústria*. Rio de Janeiro. Campus. 1995

FLAPPER, S. D. P.; Fortuin, L.; Stoop, P. P. M. *Towards consistent performance management systems*. International Journal of Operations & Production Management. Vol. 16, No. 7, 1996

FUTRELL, C. *Vendas: Fundamentos e Novas Práticas de Gestão*. São Paulo: Saraiva, 2003.

GARDNER, S. R.. *Building the data warehouse*. Communications of the ACM 41(9), 1998.

GARVINL, D. A., P. R. Nayak, A. N. Maira, & J. L. Bragar. *Aprender a Aprender*. HSM Management. 1998.

GHALAYINI, A. M. & Noble, J. S. *The changing basis of performance measurement*. International Journal of Operations & Production Management. Vol16, No 8, dez/1996.

- GIL, Antonio C. *Métodos e técnicas de pesquisa social*. 4 ed., São Paulo: Atlas, 1994.
- GOEBEL, M.; GRUENWALD, L. *A survey of data mining and knowledge discovery software tools*. Acm Sigkdd, San Diego, Vol.1, No.1, 1999.
- GROTH, Robert; "Data Mining: Building Competitive Advantage", Prentice Hall PTR, USA, 2000.
- HAIR, J.; Tatham, Ronald L; Andersin, Rolph E.; Black, Willian. *Análise Multivariadas de Dados*. 5º Edição. Bookman, 2005.
- HAGAN, M. T., Demuth, H. B., Beale, M.; "Neural network design", PWS Publishing Company, 1996
- HAN, Jiawei; KAMBER, Micheline; "Data mining – Concepts and Techniques", Academic Press Morgan Kaufmann Publishers, San Francisco, USA, 2001.
- HARRISON, T. H., *Intranet Data Warehouse*, São Paulo, Berkeley Brasil.1998
- HEREDIA, J. A.; NATARAJAN, R. *A methodology framework for developing an integrated performance measurement system*. In Ribeira, J.; Prats, M. J. (eds.) *Managing service operations – lessons from the service and the manufacturing sectors*. Barcelona, Euroma/IESE – University of Navarra, 1997
- HUGLES, Artur M.; *Data base marketing estratégico*. São Paulo: Makron, 1997.
- HSU, W.; Lee, M.L.; Liu, B.; Ling, T.W. *Exploration Mining in Diabetic Patients Databases: Findings and Conclusions*. In: KDD 2000, Boston MA USA, 2000.
- HUTT, Michael; SPEH, Thomas. *B2B: Gestão de Marketing em Mercados Industriais e Organizacionais*. Porto Alegre: Bookman, 2002.
- INMON, W. H., "What is a Data Warehouse?" Center for the Application of Information Technology (CAIT), <http://www.cait.wustl.edu/cait>, vol.1, 2003. (acessado em Abril 2007).
- I. SARAFIS, A. M. S. Zalzala, P.W. T.. *A genetic rule-based data clustering toolkit*. In Congress on Evolutionary Computation (CEC), Honolulu, USA. 2002.
- JOANES, D. N. & GILL, C. A. (1998) *Comparing measures of sample skewness and kurtosis*. Journal of the Royal Statistical Society (Series D): The Statistician 47 (1).
- KAPLAN, R. S.; NORTON, D. P. *Organização orientada para a estratégia: como as empresas que adotam o balanced scorecard prosperam no novo ambiente de negócios*, Rio de Janeiro: Campus, 2001.
- KAPLAN, Robert S.; NORTON, David P. *A estratégia em ação. Balanced Scorecard*. 10.ed. Rio de Janeiro: Campus, 2004;
- KAPLAN, Robert S.; NORTON, David P. *Linking the Balanced Scorecard to Strategy*. California Management Review, Vol.39, 1, Fall 1996

- KIMBALL, R.. *Data Warehouse Toolkit*. São Paulo: Makron Books. 1997.
- KLAUBER Poe, V., P., & S. BROBST (1998). *Building a Data Warehouse for Decision Support*. New Jersey: Prentice-Hall.
- KLEMETTINEN, M.; Mannila, H.; Toivonen, H. *A data mining methodology and its application to semi-automatic knowledge acquisition*. In: DEXA Workshop, 1997.
- KOCK Jr., N. F., R. J. McQueen, & M. Baker. *Learning and process improvement in knowledge organizations: A critical analysis of four contemporary myths*. The Learning Organization, 1996.
- KOHAVI, R., PROVOST, F. (1998), *Glossary of Terms, Machine Learning*.
- KOTLER, Philip. *Administração de Marketing*. São Paulo: Prentice Hall do Brasil. 2000.
- KOTLER, Philip. *Princípios de Marketing*. São Paulo: Prentice Hall do Brasil 9 ed. 2003.
- KOTLER, Philip; KELLER, Kevin L. *Administração de Marketing*. São Paulo: Pearson Prentice Hall, 12 ed. 2006.
- KUENG, P.; Meier, A.; Wettstein, T. *Performance measurement systems must be engineered*. Communications of AIS, Vol. 7, article 3, 2001
- LEBAS, M. J.. *Performance measurement and performance management*. International Journal of Production Economics, Amsterdam, Vol.41, No.1/3, Oct. 1995.
- LAROSE, Daniel T. 2006. *Data Mining Methods and Models*, a John Wiley & Sons, inc;
- LINOFF, Gordon.; Berry, Michael.; “*Mastering Data Mining: The Art and Science of Customer Relationships Management*”, John Wiley & Sons Inc, USA, 2000.
- LO, Victor S.; ” *The True Lift Model - A Novel Data Mining Approach to Response Modeling in Database Marketing*”; SIGKDD Explorations; vol 2 n2; 2002.
- YIN, Robert K. *Estudo de caso: planejamento e métodos*. Porto Alegre: Bookman, 3 ed., 2005.
- PARENTE, J. *Varejo no Brasil: gestão e estratégia*. São Paulo: Atlas, 2000.
- MALHOTRA, K. N., *Pesquisa de Marketing*, Bookman, 2004
- MANNILA, H.. *Data mining: Machine learning, statistic and databases*. In Proceedings of the 8th International Conference on Scientific and Statistical Database Management, 1997.
- MARTINS, R. A. *Sistemas de medição de desempenho: um modelo para estruturação do uso*. Tese (Doutorado em Engenharia de Produção) - Escola Politécnica da Universidade de São Paulo, Universidade de São Paulo, São Paulo, 1998.
- MARCOLLI, A.; “*Teoria del Campo*”, G. C. Sansoni Editore, Firenze, 1986.
- MARTINS, R. A. (2002). *The use of performance measurement information as a driver in designing a performance measurement system*. In: NEELY, A.; Walters, A.; Austin, R. performance measurement and management: research and action. Proceedings of the Third

Performance Measurement and Management Conference, 2002, Performance Measurement Association, Boston, MA (USA), July 17-19, 2002.

METZ, C.E. (1978), *Basic Principles of ROC analysis*, volume Semin Nucl Med 8.

MORESI, E. A. D. *Delineando o valor do sistema de informação de uma organização*. CI. INF. Brasília, Vol. 29, No. 1, jan/abr. 2000.

NEELY, A. *The performance measurement revolution: why now and next?* Int. Journal of Operations and Production Management, Vol. 19, No. 2, 1999.

NEELY, A. *Measuring business performance*. London, The Economist, 1998.

NEELY, A. D.; Gregory, M. J.; Platts, K. W. *Performance measurement system design – a literature review and research agenda*. International Journal of Operations & Production Management, Vol. 15, No. 4, 1995.

NEELY, A.; Adams, C.; Crowe, P. *The performance prism in practice*. Measuring Business Excellence, Vol. 5, No. 2, 2001.

NETER, J.; WASSERMAN, W. *Applied Linear Statistical Models*. Richard D. Irwin, Inc, Illinois, 1974.

NETER, J., Wasserman, W., Kutner, M. H.; Nachtshelm, C. J. *Applied Linear Regression Models*. 3^a ed., Times Mirror Hiher Group, Inc., Boston, 1996.

NICKOLS, F. *Measurement-based analysis: hooking what you do to the bottom line (on line)*. 1999. Disponível em <http://hone.att.net/~NICKOLS/mba.htm>. Acessado em Junho de 2006.

NORREKLIT, H.. *The balance on the balanced scorecard - a critical analysis of some of its assumptions*. Management Accounting Research, London, Vol.11, No.1, Mar. 2000.

NORUSIS, M. 2004. *SPSS 13.0 Guide to Data Analysis*. Upper Saddle-River, N.J.: Prentice Hall, Inc..

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

PELEGRIN Diana C., Diego Paz Casagrande, Merisandra Côrtes de Mattos, Priscyla Waleska Targino de Azevedo Simões, Rafael Charnovscki, Jane Bettiol. *As Tarefas de Associação e de Classificação na Shell de Data Mining Orion*. Disponível em: <http://www.dcc.unesc.net/sulcomp/artigos/sessaoOral/22103.pdf>. Acessado em Fevereiro de 2007.

PEPPERS, Don; ROGERS, Martha. *Empresa 1:1 instrumentos para competir na era da interatividade*. Rio de Janeiro: Campus, 1997.

PINTO, F, Santos, M.F, Cortez, P, Quintela, H, “Data Preprocessing for Database Marketing”, Data Gadgets 2004, Málaga Spain, 2004.

PRIOR, Vernon; “*The Language of Competitive Intelligence: Part One*”, *Competitive Intelligence Review*, vol. 9(2).

PROVOST, F., FAWCETT, T. (1997), *Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions*, in *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, AAAI Press.

PYLE, Dorian; “*Data Preparation for Data Mining*”, Morgan Kaufmann Publishers Inc, San Francisco, CA, 1999.

RAMACHANDRAN, Pushpa; “*White Paper - Mining for Gold*”, Wipro Technologies, Santa Clara CA, USA, 2001.

RODRIGUES FILHO, Lino N.; ROMAN Amigo, Ricardo J. *Determinantes da competitividade no marketing industrial*. *Revista de Administração da USP – RAUSP*. São Paulo, Vol. 35, No. 1, jan/mar. 2000.

SAS, <http://www.sas.com> (url) (acessado em 17-3-2007).

SELLER, Marianne; Gray, Paul; “*A Survey of Database Marketing*”, California, March 1999.

SELLETIZ, W. C. *Métodos de pesquisa nas relações sociais*. São Paulo: EPU, 1974.

SHAPIRO, Benson P. *Velocidade para fechar venda*. HSM Management. São Paulo. Ano 5, No. 28, 2001.

SHAPIRO, Benson P.; Rangan, V. Kasturi; Moriarty, Rowland, T.; Ross, Elliot. *Manage Customers for Profits (Not Just Sales)*. *Harvard Business Review*. Boston, V. 65, n. 5, 1987.

SHEPARD, David; “*The New Direct Marketing: How to Implement A Profit-Driven Database Marketing Strategy*”, ed David SHEPARD Associates, McGraw-Hill, 3ª ed; 1998.

SIQUEIRA, Antonio C. B. *Marketing Empresarial, Industrial e de Serviços*. São Paulo: Saraiva, 2005.

SINK, Scott; TUTTLE Thomas C. *Planning and measurement in your organization of the future*. Norcross, Georgia: Industrial of industrial engineers, 1989.

S. MITRA, S. K.; PAL, P. M.. *Data mining in soft computing framework: A survey*. 2002. *IEEE Transactions on Neural Networks* 13(1).

SUWIGNJO, P.; Bititci, U. S.; Carrie, A.S. *Quantitative models for performance measurement system*. *International Journal of Production Economics*, Vol. 64, 2000.

TELLES, Renato. *Marketing Empresarial*. São Paulo: Saraiva, 2003.

TOLEDO, Geraldo L.; SIQUEIRA, Antonio B. *Bases de segmentação de mercado no marketing industrial*. *Revista de Administração da USP – RAUSP*, São Paulo, Vol. 36, No.4, 2001.

TORGO, L., *Inductive Learning of Tree-based Regression Models*, PhD thesis, Universidade do Porto, Portugal, 1999.

- ZARIFIAN, P. *Organização e sistema de gestão: à procura de uma nova coerência*. Gestão & Produção, Vol. 4, 1997.
- ZOLTNERS, A. A.; Sinha, P.; ZOLTNERS, G. A. *The Complete Guide to Accelerating Sales Force Performance*. New York: Amacom, 2001.
- ZWICK, Detlev, Nikhilesh Dholakia “*Whose Identity Is It Anyway? Consumer Representation in the Age of Database Marketing*”; Journal of Macromarketing, Vol. 24, No. 1, 2004.
- WEBSTER F. E. *Industrial marketing strategy*. USA: John Wiley & Sons, 3 ed, 1991.
- WEBSTER, Jr., Frederick, E., Wind, Yoram. *O Comportamento do Comprador Industrial*. São Paulo: Atlas, 1975.
- WEISBERG, S. 1985. *Applied Linear Regression*. New York: John Wiley & Sons.
- WEITZ, B. A.; Castleberry, S. B.; Tanner, J. F.. *Selling – Building Partnerships*. New York: McGraw-Hill, 2004.
- WESTPHAL, C.; Blaxton, T.; “*Data Mining Solutions: Methods and Tools for Solving Real-World Problems*”, John Wiley & Sons, New York, 1998.
- W. FRAWLEY, G. PIATETSKY-SHAPIRO, C. M. *Knowledge discovery in databases: An overview*. AI Magazine. 1992.
- WIND, Yoram; CARDOZO, Richard N. *Industrial market segmentation*. Industrial Marketing Management. New York, Vol. 3, Iss. 2 Mar. 1974.
- WIRTH, Ruediger; “*CRISP-DM Position Statement*”, 6th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, USA, 2000.
- WHITE, G. P. *A survey and taxonomy of strategy-related performance measures for manufacturing*. International Journal of Operations & production Management. Vol. 16, No.3, 1996.
- WITTEN, I. H. & FRANK, E.. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann Publishers. 1999.
- ZHOU, Z. H.. *Three perspectives of data mining*. Artificial Intelligence journal. 2003.
- VAKKURI, J.; MEKLIN, P. *The impact of culture on the use o performance measurement information in the university setting*. Management Decision. 41/8, 2003.

ANEXO

Descrição das Variáveis Utilizadas na Pesquisa

Nº	BALANCED SCORECARD VERSUS NOME DO INDICADOR	FÓRMULA DE CÁLCULO	FONTE DE COLETA	MEDIÇÃO	
				META	FREQUENCIA
				UNIDADE	SENTIDO
CO Concorrência					
CO.1	Venda x Mapeamento da concorrência	Total de vendas com mapeamento da concorrência ----- Total de Vendas	SCI e SA3	20,0% %	Mensal U
EO Excelência Operacional					
EO.1	Assertividade de entrega e cadastro de contratos	Total de contratos cadastrados no prazo ----- Total de contratos cadastrados no SCI	SCI	90,0% %	Mensal U
EO.4	Retorno de contratos	Total de contratos devolvidos pela GGC ----- Total de contratos cadastrados no SCI	SCI	7,0% %	Mensal D
EO.5	Abertura de OS	Total de OS abertas dentro do prazo ----- Total de OS abertas no mês	SCI	90,0% %	Mensal U
EO.6	Conversão de contratos	Total de OS abertas no mês ----- Total de contratos cadastrados no SCI	SCI	90,0% %	Mensal U
EO.7	Cancelamento de vendas	Total de OS canceladas no mês ----- Total de OS abertas no mês	SCI	3,0% %	Mensal D
FC Foco do Cliente					
FC.1	Média de visitas por consultor	Total de visitas ----- Total de consultores x Total de dias úteis	SA3	3,0 visitas / dia	Mensal U
FC.2	Assertividade de agendamento	Total de visitas realizadas de acordo com o agendamento ----- Total de visitas agendadas	SA3 e ASE	70,0% %	Mensal U
FC.3	Visitação X Agendamento	Total de visitas agendadas ----- Total de visitas	SA3 e ASE	80,0% %	Mensal U
GP Gestão de Parceiros					
GP.1	Produtividade de Visita	Total de visitas com oportunidades identificadas ----- Total de visitas	SA3	90,0% %	Mensal U
GP.2	Eficácia de visitas	Total de contratos fechados ----- Total de visitas	SCI e SA3	15,0% %	Mensal U
GP.3	Cesta de produtos	Total de consultores com vendas em todo a Cesta ----- Total de consultores	SCI e SA3	30,0% %	Mensal U
GP.4	Distribuição de vendas (consultores Verdes + Amarelos)	Total de consultores verdes + amarelos ----- Total de consultores	SCI e SA3	60,0% %	Semanal U
GP.5	Produtividade de Oportunidades	Total de vendas com mapeamento de oportunidades ----- Total de visitas com oportunidades identificadas	SCI e SA3	50,0% %	Mensal U
IA Inovação e Ação					
IA.2	Volume de vendas Com Descontos (VA + Dados)	Total de vendas com descontos ----- Total de vendas	SCI	30,0% %	Mensal D
IA.3	Aprovação de viabilidades	Total de viabilidade aprovadas ----- Total de solicitações de viabilidade	SA3	70,0% %	Mensal U
IA.4	Assertividade de análise de viabilidade	Total de solicitações de viabilidade atendidas no prazo ----- Total de solicitações de viabilidade atendidas	SCI	95,0% %	Mensal U

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)