

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

USO DO DATA MINING NO ESTABELECIMENTO DE
RELACIONAMENTOS ENTRE AS MEDIDAS DE DESEMPENHO

FLAVIO AUGUSTO CUSTODIO

DISSERTAÇÃO DE MESTRADO

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

**USO DO DATA MINING NO ESTABELECIMENTO DE
RELACIONAMENTOS ENTRE AS MEDIDAS DE DESEMPENHO**

Flavio Augusto Custodio

**Dissertação de Mestrado apresentada ao
Programa de Pós-Graduação em
Engenharia de Produção da
Universidade Federal de São Carlos,
como parte dos requisitos para a
obtenção do título de Mestre em
Engenharia de Produção.**

Orientador: Prof. Dr. Roberto Antonio Martins

Agência Financiadora: CNPq

SÃO CARLOS

2004

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

C987ud

Custodio, Flavio Augusto.

Uso do data mining no estabelecimento de relacionamentos entre medidas de desempenho / Flavio Augusto Custodio. -- São Carlos : UFSCar, 2005.
120 p.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2004.

1. Processo decisório. 2. Medição de desempenho. 3. Data mining (mineração de dados). 4. Medidas de desempenho. 5. Sistemas de medição de desempenho. I. Título.

CDD: 658.5036 (20^a)



UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Exatas e de Tecnologia
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO
Rod. Washington Luís, Km. 235 - Caixa Postal 676
CEP. 13565-905 - São Carlos - SP - Brasil
Fones: (016) 260-8238 – (ramal 232)
Fax: (016) 260-8238 (r. 232)
Email : ppg-ep@power.ufscar.br

FOLHA DE APROVAÇÃO

Aluno: FLAVIO AUGUSTO CUSTODIO

DISSERTAÇÃO DE MESTRADO DEFENDIDA E APROVADA EM 30/09/04
PELA COMISSÃO JULGADORA:

PROF. DR. ROBERTO ANTONIO MARTINS
(Orientador - PPGEP/UFSCar)

PROF. DR. NÉOCLES ALVES PEREIRA
(PPGEP/UFSCar)

PROF. DR. ANTONIO CARLOS DOS SANTOS
(DC/UFSCar)

PROF. DR. EDSON WALMIR CAZARINI
(EESC/USP)

Presidente da Coordenação de Pós-Graduação
Prof. Dr. Dário Henrique Alliprandini

AGRADECIMENTOS

Ao Prof. Dr. Roberto Antonio Martins pela orientação e compreensão durante o desenvolvimento deste trabalho.

Ao Prof. Dr. Antonio Carlos dos Santos pelo seu inestimável auxílio para este trabalho.

Ao CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico, pelo apoio financeiro.

A toda minha família, meu pai, minha mãe e irmãos por terem me incentivado durante todo o trabalho.

Aos meus colegas de pós-graduação do Departamento de Engenharia de Produção, em especial ao Luciano Silva, Daniela, Mirela, Marcela, André, Tatiane, Ricardo Mergulhão, Gian, Lesley, Eduard e Zuin, pela convivência e troca de experiências, nos auxiliando nos momentos de angústias e dúvidas.

Aos professores e funcionários do Departamento de Engenharia de Produção.

À todos aqueles que de alguma forma deixaram uma parcela de contribuição ao desenvolvimento deste trabalho.

ABSTRACT

This work aims to propose a method to analyze the relationships between performance measures in a Performance Measurement System using historical performance data stored in a datawarehouse or operational data store. There is a problem in the performance measurement area that it doesn't have methods to create relationships models between performance measures. The present methods that we have in academic researches don't help to build the relationships concerning historical performance data. Therefore, there is a trend to build the relationship between performance measures to reflect the desirable future, but it is also true that we have to learn about the past actions. Nowadays, with the increasing complexity in the organizations environment it is very difficult to handle historical data about performance to identify relationship patterns without using concepts, techniques and tools of the Information Technology (IT) field. The variables contained in the performance measurement models are increasing continually so it is important to understand the complex net of relationships between performance measures in an organization. The stakeholders in the organization see the relationships between performance measures as trivial, but this doesn't help because the relationships are partial and subjective and the stakeholders that articulate the variables in most of the cases are accountable by the performance. It's expected that decision makers participate and share their models of relationships between performance measures and that it be the most comprehensive as possible. This work is important because it proposes to use the data mining philosophy to help building a method to understand relationship between performance measures with performance historical data. Hence, it will be possible to define and communicate the relationships between performance measures to the users of the organization and increase the use of performance measurement models. The proposed method presents a process to build and find relationships between performance measures data using data mining techniques. The IDEF0 procedure was used to present our approach.

Key-Words: Performance Measurement Systems; Relationships between Performance Measures; Data Warehouse; Data Mining

RESUMO

O objetivo deste trabalho é propor um método para o estabelecimento dos relacionamentos entre as medidas de desempenho de um sistema de medição de desempenho a partir de dados históricos sobre desempenho armazenados em um banco de dados, utilizando a abordagem *data mining*. Um problema no campo da medição de desempenho é a falta de métodos de criação de modelos de relacionamentos entre as medidas de desempenho. Os existentes, encontrados na literatura, não tratam de como construir o relacionamento a partir de dados históricos de desempenho. Além disso, existe uma tendência de estabelecer o relacionamento esperado de forma que a medição de desempenho reflita o futuro desejado. Entretanto, é de grande valia aprender por intermédio daquilo que já foi feito, ou seja, pelas ações passadas. Com o aumento da complexidade das organizações, fica um tanto quanto difícil manipular dados históricos sobre desempenho para a identificação de padrões de relacionamento sem lançar mão de conceitos, técnicas e ferramentas da tecnologia de informação. Em face de o número de variáveis envolvidas ser cada vez maior, é importante a busca do entendimento da complexa teia de relacionamento existente entre as medidas de desempenho numa organização. Este relacionamento é visto pelas pessoas nas organizações como algo corriqueiro. Entretanto, o que pode ser improdutivo é que esses relacionamentos são parciais e pessoais, visando a articular as variáveis por cujo desempenho as pessoas, na maioria dos casos, tinham responsabilidade. O ideal é que a maioria dos tomadores de decisão compartilhem do mesmo modelo de relacionamento entre as medidas de desempenho e que ele fosse a mais abrangente possível. Portanto, a relevância deste trabalho é procurar desenvolver uma forma de aplicação da abordagem *data mining* a fim de auxiliar na construção de um método para o estabelecimento dos relacionamentos entre as medidas de desempenho com base em dados de desempenho históricos. Assim, será possível formalizar e disseminar o relacionamento entre as medidas de desempenho para uma gama maior de pessoas numa organização, podendo melhorar o uso da medição de desempenho. O método proposto procura abranger todo o processo de construção do relacionamento com aplicação de *data mining* e não somente a aplicação de uma ou outra técnica específica dele. A apresentação da proposta é feita utilizando-se a prática IDEF0.

Palavras-Chave: Sistema de Medição de Desempenho; Relacionamentos entre as Medidas de Desempenho; *Data Warehouse*; *Data Mining*

SUMÁRIO

1	INTRODUÇÃO.....	1
1.1	Estrutura do Trabalho.....	2
2	MEDIÇÃO DE DESEMPENHO.....	4
2.1	A Medição de Desempenho no Contexto Industrial.....	4
2.2	Visão Geral sobre Medição de Desempenho.....	12
2.2.1	A importância dos relacionamentos entre as medidas de desempenho.....	15
2.2.2	O relacionamento entre as medidas de desempenho para a organização industrial.....	16
2.3	Relacionamentos entre as Medidas de Desempenho.....	18
2.3.1	As relações entre medidas de desempenho individuais.....	19
2.3.2	Métodos para a construção de relacionamentos entre medidas de desempenho	22
2.3.3	Os relacionamentos entre as medidas de desempenho nos modelos de medição de desempenho.....	31
2.4	Uso da Tecnologia de Informação para Construir Relacionamentos entre as Medidas de Desempenho.....	38
2.5	Considerações e Síntese do Capítulo.....	41
3	RELAÇÃO ENTRE <i>DATA MINING</i> E MEDIÇÃO DE DESEMPENHO.....	44
3.1	Contextualização do <i>Data Mining</i> em um Ambiente Organizacional.....	44
3.1.1	Sistema de informação e tecnologia de informação.....	45
3.1.2	Sistemas de apoio à decisão.....	51

3.1.3	<i>Data warehouse</i>	55
3.1.4	OLAP.....	60
3.2	<i>Data Mining</i>	62
3.2.1	Visão geral.....	63
3.2.2	Projeto de <i>data mining</i>	66
3.2.3	Técnicas de <i>data mining</i>	72
3.2.4	Ferramentas de <i>data mining</i>	76
3.3	O <i>Data Mining</i> na Medição de Desempenho.....	82
3.4	Considerações e Síntese do Capítulo.....	84
4	PROPOSIÇÃO DE UM MÉTODO PARA ESTABELECIMENTO DE RELACIONAMENTOS ENTRE AS MEDIDAS DE DESEMPENHO.....	85
4.1	Proposta de Pesquisa.....	86
4.2	Visão Geral do Método para Estabelecimento de Relacionamentos entre as Medidas de Desempenho.....	88
4.3	Detalhamento do Método Proposto.....	89
4.3.1	Levantamento de informações sobre a empresa.....	92
4.3.2	Operacionalização do <i>data mining</i>	103
4.3.3	Uso dos resultados obtidos.....	110
5	CONCLUSÕES.....	112
	REFERÊNCIAS BIBLIOGRÁFICAS.....	115

ANEXO A 123

APÊNDICE B 126

LISTA DE QUADROS

QUADRO 2.1 – Tipos de Redes de produção.....	8
QUADRO 2.2 – Vantagens e desvantagens dos métodos para a construção de relacionamentos entre as medidas de desempenho.....	30
QUADRO 2.3 – Exemplo de medidas de desempenho para cada faceta do prisma.....	37
QUADRO 2.4 – Características e críticas dos relacionamentos entre as medidas de desempenho nos modelos de medição de desempenho....	38
QUADRO 2.5 – Componentes do Sistema de Medição de Desempenho.....	40
QUADRO 3.1 – Características do processamento dos sistemas de informação.....	51
QUADRO 3.2 – Diferenças entre banco de dados operacionais e <i>data warehouse</i>	59
QUADRO 3.3 – Possíveis operações analíticas em um projeto de <i>data mining</i>	71
QUADRO 3.4 – Principais técnicas e ferramentas utilizadas nas operações de <i>data mining</i>	72
QUADRO 3.5 – Vantagens e desvantagens das ferramentas de <i>data mining</i>	81
QUADRO 3.6 – As técnicas de <i>data mining</i> e as aplicações na medição de desempenho.....	82
QUADRO 4.1 – Estratégias aplicadas em sistemas de produção de diferentes setores.	96
QUADRO 4.2 – Informações críticas para o sistema de informação gerencial em quatro setores industriais.....	97

LISTA DE FIGURAS

FIGURA 2.1 – Modelo de organização baseada na indústria eletrônica.....	6
FIGURA 2.2 – Um modelo para o <i>design</i> da medição de desempenho.....	13
FIGURA 2.3 – Posicionamento do sistema de medição de desempenho em uma organização.....	14
FIGURA 2.4 – Relacionamento de causa e efeito.....	19
FIGURA 2.5 – Diagrama de influência apresentando as relações <i>delay</i>	21
FIGURA 2.6 – Modelo causal.....	22
FIGURA 2.7 – Desdobramento do modelo causal.....	23
FIGURA 2.8 – Relacionamentos entre indicadores de desempenho baseado em “pais” e “filhos”	23
FIGURA 2.9 – Dois cenários para a medida de desempenho “eficiência” a partir do uso do <i>Negative Branch</i>	24
FIGURA 2.10 – Diagrama de indicadores de desempenho estratégicos.....	25
FIGURA 2.11 – Mapa cognitivo.....	27
FIGURA 2.12 – Diagrama de Causa-e-Efeito.....	27
FIGURA 2.13 – Diagrama de árvore.....	28
FIGURA 2.14 – Questionário de comparação dos fatores de desempenho.....	28
FIGURA 2.15 – Estrutura do modelo de medição de desempenho <i>Performance Pyramid</i>	32
FIGURA 2.16 – Mapa estratégico no <i>Balanced Scorecard</i>	33
FIGURA 2.17 – Analisando o desempenho entrega no IDPMS.....	34
FIGURA 2.18 – Integração do Sistema de Medição de Desempenho no IPMS.....	36
FIGURA 2.19 – Cinco atividades fundamentais da medição de desempenho.....	40

FIGURA 3.1 – Funções de um sistema de informação.....	47
FIGURA 3.2 – Integração do sistema de informação com a estratégia de negócio.....	49
FIGURA 3.3 – Inter-relacionamentos entre os sistemas de informação.....	52
FIGURA 3.4 – Arquitetura de um sistema de apoio à decisão.....	54
FIGURA 3.5 – Tipos de <i>data warehouse</i> e <i>data marts</i> em uma organização.....	57
FIGURA 3.6 – A relação entre <i>data warehouse</i> e <i>data mining</i>	65
FIGURA 3.7 – Estágios do processo de <i>data mining</i>	68
FIGURA 3.8 – Exemplo de uso da técnica de classificação.....	73
FIGURA 3.9 – Exemplo de separação dos dados.....	74
FIGURA 3.10 – Exemplo de uso da técnica de descoberta de associações.....	75
FIGURA 3.11 – Estrutura de uma rede neural.....	76
FIGURA 3.12 - Arquitetura da rede neural.....	77
FIGURA 3.13 – Exemplo da ferramenta árvore de decisão.....	78
FIGURA 3.14 – Representação gráfica de funções lineares e não lineares.....	79
FIGURA 3.15 – Arquitetura do ILP.....	80
FIGURA 4.1 - Modelo de pesquisa proposto pelos autores.....	87
FIGURA 4.2 – Geração do modelo conceitual.....	88
FIGURA 4.3 – Verificação do modelo de pesquisa.....	88
FIGURA 4.4 – Visão geral do método para o estabelecimento de relacionamentos entre as medidas de desempenho.....	90
FIGURA 4.5 – Detalhamento do método para estabelecimento de relacionamentos entre as medidas de desempenho.....	91
FIGURA 4.6 – Adquirir conhecimento sobre a empresa, as medidas de desempenho e sistema de informação.....	93

FIGURA 4.7 – Geração dos dados a partir de um modelo de negócios.....	94
FIGURA 4.8 – A escolha do tipo de banco de dados para a aplicação das técnicas e ferramentas de <i>data mining</i>	100
FIGURA 4.9 – Formação do modelo de <i>data warehouse</i> final.....	102
FIGURA 4.10 – Operacionalizar a <i>data mining</i>	103
FIGURA 4.11 – Escolha das técnicas e ferramentas de <i>data mining</i>	105
FIGURA 4.12- Fases para o estabelecimento das técnicas e ferramentas de <i>data mining</i> na construção de relacionamentos entre as medidas de desempenho.....	106
FIGURA 4.13 – Construção dos resultados.....	110
FIGURA A1 – Elementos Básicos Presentes no IDEF0.....	123
FIGURA A2 – Decomposição da estrutura com base no conceito diagramas pais e filhos.....	124
FIGURA A3 – Descrição de alguns elementos presentes na metodologia IDEF0.....	125

LISTA DE TABELAS

TABELA 2.1 – Matriz de comparação dos fatores de desempenho.....	29
--	----

LISTA DE SIGLAS

ABC:	<i>Activity-Based Costing</i>
AHP:	<i>Analytic Hierarchy Process</i>
BD:	Banco de Dados
BSC:	<i>Balanced Scorecard</i>
BLS:	<i>Bureau Labor Statistics</i>
CEP:	Controle Estatístico de Processo
DEA:	<i>Data Envelopment Analysis</i>
DMU:	<i>Decision Making Unit</i>
ERP:	<i>Enterprise Resource Planning</i>
HOLAP:	<i>Hybrid OLAP</i>
ILP:	<i>Inductive Logic Programming</i>
IDPMS:	<i>Integrated Dynamic Performance Measurement System</i>
IPMS:	<i>Integrated Performance Measurement System e Performance Prism</i>
IDEF:	<i>Integration Definition for Function Modeling</i>
IT:	<i>Information Technology</i>
JIT:	<i>Just-in-Time</i>
KDD:	<i>Knowledge Discovery Databases</i>
MOLAP:	<i>Multidimensional OLAP</i>
OLAM:	<i>On-Line Analytical Mining</i>
OLAP:	<i>On-Line Analytical Processing</i>
P&D:	Pesquisa e Desenvolvimento
QME:	Quadrado Médio do Erro
ROI:	<i>Return Of Investment</i>

ROLAP:	<i>Relational OLAP</i>
SAD:	Sistemas de Apoio à Decisão
SMD:	Sistemas de Medição de Desempenho
SMART:	<i>Strategic, Measurement, Analysis, and Reporting Technique</i>
SQL:	<i>Structured Query Language</i>
TI:	Tecnologia de Informação
TQM:	<i>Total Quality Management</i>
WCM:	<i>World Class Manufacturing</i>

1. INTRODUÇÃO

Certamente a medição de desempenho é um tópico atual no campo da Engenharia de Produção e áreas afins. Alguns autores argumentam que a medição de desempenho vive uma revolução desde o início da década passada (ECCLES, 1991; SINK, 1991; NEELY, 1998; NEELY, 1999). A principal razão para tanto é provavelmente a inadequação dos sistemas de medição de desempenho (SMD) tradicionais desenvolvidos com base na Contabilidade Gerencial do início do século passado, coerente com o paradigma da produção em massa (MARTINS, 1998).

Na eterna busca da competitividade, muitas empresas investiram uma grande quantidade de recursos em novas tecnologias, sistemas de informação e programas de mudanças, porém elas não alteraram significativamente a forma de avaliarem o desempenho (MARTINS, 2002). A medição de desempenho tradicional foi construída durante a década de 20 do século XX quando o ambiente interno e externo das organizações industriais era bastante diferente das atuais (JOHNSON & KAPLAN, 1993).

Desde a descoberta e tomada de consciência da inadequação da medição de desempenho tradicional, o trabalho de pesquisadores e praticantes sobre o assunto, refletido na literatura especializada, está na sua maior parte concentrada na proposição de novos modelos de medição de desempenho, enquanto outras dimensões como o uso da informação proveniente das medidas de desempenho vem tendo pouca ênfase (MARTINS, 1998; KUENG et al., 2001). A informação pode ser utilizada no suporte para a tomada de decisão e ação, segundo Hopwood¹ (1974 *apud* BLENKINSOP & BURNS, 1992), contribuindo para um adequado comportamento de gerentes e colaboradores na organização.

Um outro aspecto não muito considerado na literatura sobre medição de desempenho é a aplicação de tecnologia de informação (TI) nas suas mais variadas facetas e não somente em atividades mais tradicionais, tais como: compilar e armazenar (KUENG et al., 2001). A aplicação de TI pode suportar os procedimentos de análise e interpretação dos dados sobre desempenho, ajudando a esclarecer relacionamentos entre as medidas de desempenho (OLVE et al., 1999). Isso pode conduzir o uso da medição

¹ Hopwood, A. G. Accounting and Human Behavior. London, U.K.: Haymarket, 1974.

de desempenho para além do controle cibernético já constatado por alguns estudos como sendo muito comum nas organizações (MARTINS, 1998; KUENG et al., 2001).

Dentre as inúmeras possibilidades de aplicação de TI na medição de desempenho, destacam-se as abordagens *data warehouse* e *data mining* que procuram auxiliar o usuário da informação sobre desempenho a transformar uma extensa coleção de dados, muito comum à grande maioria das empresas atualmente, em informação que guiará a tomada de decisão e subsequente ação com vistas a manter e/ou aumentar a vantagem competitiva de uma organização.

Neste contexto, o objetivo desta dissertação é propor um método para o estabelecimento do relacionamento entre as medidas de desempenho de um SMD a partir de dados históricos armazenados em bancos de dados modelados como um *data warehouse*. Para o estabelecimento dos relacionamentos, será utilizada a abordagem *data mining*.

A proposição do método para o estabelecimento de relacionamentos entre as medidas de desempenho será realizada a partir da revisão da literatura, visando criar um guia estruturado tendo como ponto de partida os dados históricos de desempenho e como chegada o estabelecimento dos relacionamentos entre as medidas de desempenho pelos tomadores de decisão.

Espera-se que o método seja útil para que futuros pesquisadores possam replicar em outras organizações ou buscar estudar uma questão de pesquisa específica envolvendo as relações entre medição de desempenho e TI.

1.1 Estrutura do Trabalho

O trabalho está dividido nas seguintes partes:

- revisão da literatura;
- apresentação do método para o estabelecimento de relacionamentos entre as medidas de desempenho a partir do uso do *data mining*; e
- apresentação das considerações finais e propostas futuras do trabalho.

O capítulo dois apresenta a importância dos relacionamentos entre as medidas de desempenho de um SMD. Para tanto, são apresentados o papel da medição de desempenho no ambiente industrial, uma visão geral sobre medição de desempenho, a importância e a contribuição dos relacionamentos entre as medidas de desempenho

para as organizações e as diversas denominações e representações dos relacionamentos entre as medidas de desempenho. Por fim, é mostrada a problemática da pesquisa enfocando a questão do aumento da complexidade das dimensões de desempenho e a negligência ou mau uso da TI na medição de desempenho. Considerações sobre o capítulo são apresentadas.

O capítulo três apresenta o *data mining*. Inicialmente é mostrada uma visão geral sobre o seu papel na geração do conhecimento na empresa, no sistema de informação, na TI e também nas abordagens modernas de TI (*data warehouse* e o *On-Line Analytical Processing*, conhecido também como OLAP), contribuem para o *data mining*. São apresentadas as operações analíticas, as técnicas, as ferramentas, as principais características de um projeto de *data mining*, qual a contribuição da abordagem para a medição de desempenho e ,mais especificamente, para os relacionamentos entre as medidas de desempenho. Considerações sobre o capítulo são apresentadas.

No capítulo quatro, é apresentado o método para o estabelecimento do relacionamento entre as medidas de desempenho revelando a estratégia de pesquisa para a criação do método, a proposta do método o estabelecimento de relacionamentos entre as medidas de desempenho. O detalhamento deste método a partir do uso da representação IDEF0.

Por fim, no capítulo cinco, são apresentadas as conclusões do trabalho com considerações finais sobre o método, a contribuição e limitações dele e as propostas futuras deste trabalho.

2. MEDIÇÃO DE DESEMPENHO

A fim de que se possa sistematizar e descrever o método para o estabelecimento de relacionamentos entre as medidas de desempenho de um SMD, com o uso do *data mining*, é necessário ter conhecimento da problemática desses relacionamentos dentro do contexto da medição de desempenho.

Neste sentido, o objetivo deste capítulo é apresentar a relevância desses relacionamentos. Os aspectos apresentados são:

- a importância dos relacionamentos entre as medidas de desempenho para a organização;
- diferentes representações e denominações dos relacionamentos entre as medidas de desempenho; e
- o uso da tecnologia de informação nos relacionamentos entre as medidas de desempenho.

Antes de apresentar cada aspecto que contribui para o entendimento da sua importância, é apresentada uma visão geral sobre a relevância da medição de desempenho nos dias atuais e como os relacionamentos entre as medidas de desempenho de um SMD se inserem nesse contexto.

2.1 A Medição de Desempenho no Contexto Industrial

Nas últimas décadas, a maioria das organizações industriais tem passado por diversas mudanças tanto no ambiente interno quanto no externo. As mudanças sócio-econômicas que afetaram o ambiente industrial são apoiadas em três pilares, segundo Hoffman & Kaplinsky (1988): a transição da maquinofatura para sistemofatura; a difusão da microeletrônica; e a mudança no relacionamento entre as firmas.

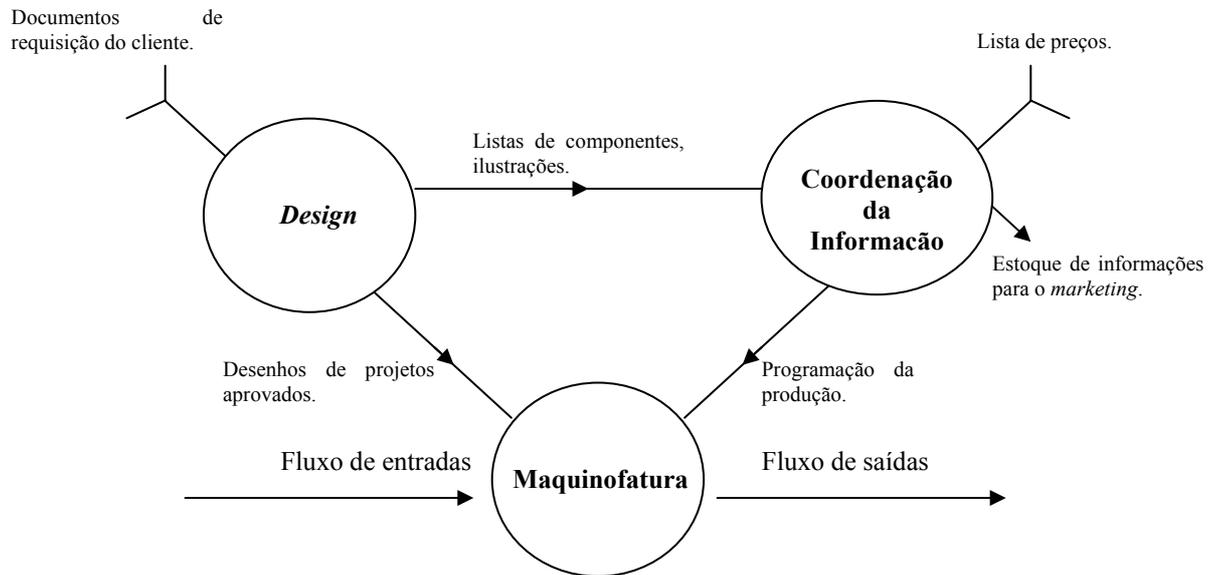
No primeiro pilar de transição entre os dois modelos de produção, o fator principal de mudança se deve principalmente à exaustão do modelo fordista por causa da crise do modelo clássico de produção baseado no fordismo e taylorismo (MARTINS, 1998). Veltz & Zarifian (1994) apontam três elementos do princípio da eficiência do modelo clássico que entraram em crise:

- a. a mão-de-obra direta como fator preponderante da eficiência dos sistemas de manufatura;
- b. o problema da produtividade se resume numa fórmula sobre o objetivo dominante, no caso custos; e
- c. a temporalidade linear do regime de exploração regular dos sistemas de manufatura.

A transição da era da maquinofatura para a sistemofatura ocorreu devido às dificuldades econômicas do modelo fordista de manter o processo de acumulação das economias nos níveis nacional e global, à insatisfação na realização do trabalho pelas pessoas e a inflexibilidade do maquinário. A sistemofatura surgiu como substituta baseada no novo processo de trabalho que enfoca as indústrias eletrônica e automotiva japonesa que se fundamentam nos baixos níveis de estoque, enfoque produto–inovação, trabalhadores com multi-habilidades e, conseqüentemente, com multitarefas, controle da qualidade, ênfase maior no fluxo da informação na organização e flexibilidade no trabalho (HOFFMAN & KAPLINSKY, 1988).

O segundo pilar é a difusão da microeletrônica que gerou a presença de um enfoque sistêmico com as novas tecnologias (tecnologia digital). Isso pode ser notado nas três esferas de produção de uma firma moderna que são o *design*, a maquinofatura e a coordenação da informação. A coordenação da informação monitora e atua sobre as outras duas esferas que também se inter-relacionam. Outro fator importante da microeletrônica é a questão da automação na substituição do trabalho, que pode atuar tanto nas atividades das esferas de produção, no relacionamento das atividades nas esferas e também na relação entre as esferas de produção (HOFFMAN & KAPLINSKY, 1988).

A Figura 2.1 mostra um exemplo de modelo de organização que gerencia as entradas e saídas baseadas na indústria eletrônica. Três esferas compõem o modelo organizacional: a maquinofatura; *design*; e coordenação da informação.



Fonte: Hoffman & Kaplinsky (1988, p.57).

FIGURA 2.1 – Modelo de organização baseada na indústria eletrônica.

De acordo com a Figura 2.1, a maquinofatura recebe o apoio do *design* nos desenhos dos projetos aprovados para executar a produção e a coordenação da informação na programação da produção dos equipamentos e mão-de-obra. O *design* atua de acordo com a necessidade dos clientes. A coordenação da informação atua de acordo com a lista e ilustrações de componentes a partir do *design*, da lista de preços e também fornece informações importantes para o *marketing* a partir das listas obtidas.

O terceiro pilar é a mudança do relacionamento entre as firmas e está fortemente ligado aos princípios da indústria japonesa, principalmente na relação entre os fornecedores e as montadoras (típico da indústria automobilística). O primeiro princípio é a proximidade entre fornecedores, baseado nas novas tecnologias organizacionais como o *Just-in-Time*² e zero defeito, que pregam o forte relacionamento entre montadoras e fornecedores; o segundo princípio é a natureza do progresso tecnológico, que, com o avanço da produção sistêmica, provoca uma interdependência entre diferentes subsistemas na geração do produto final; e o terceiro princípio é a

² O objetivo do sistema JIT é reduzir os estoques, de modo que os problemas fiquem visíveis e possam ser eliminados por meio de esforços concentrados e priorizados, possibilitando a melhoria contínua do processo produtivo. O JIT procura transformar o processo de produção, aproximando-o do fluxo contínuo, principalmente por meio do aprimoramento do projeto dos produtos e da modificação do *layout*, utilizando o conceito de células de manufaturas e linhas (Correa et al., 1993).

crescente tendência para o conteúdo tecnológico, o que requer grandes investimentos em desenvolvimento do produto (HOFFMAN & KAPLINSKY, 1988).

Desde dos fins da década de 80 do século XX, surgiram novos fenômenos que afetam o ambiente industrial como, segundo Ferraz et al. (1995), a globalização (maior quantidade de transações comerciais e financeiras), especialização (crescente agrupamento dos produtos em uma mesma base tecnológica), concentração e descentralização geográfica. Diante desses fenômenos surgiram os chamados *players* (grandes empresas que atuam em diversos setores), que passaram a atuar em diversos países por meio de aquisições ou *joint-ventures* (associação entre duas empresas para participar de determinado empreendimento) com as empresas locais (WONGTSCHOWSKI, 1999).

Esta atuação dos *players* gerou dois fenômenos: a formação de cadeia de empresas alinhadas ao *player*; ou a associação de empresas locais ou de um mesmo setor como forma de sobreviverem e se protegerem contra as investidas dos *players*. O Quadro 2.1 mostra exemplos de redes de empresas que estão de acordo com o fenômeno de cadeias e que representam novas formas de relacionamento entre elas.

Essas mudanças no ambiente externo e interno das organizações, principalmente nas empresas de manufatura, criaram novas demandas e um novo padrão tecnológico. Zarifian (1997, p.77) alerta que:

Elas (as mudanças organizacionais) só poderão chegar a resultados positivos se, paralelamente, os sistemas de gestão forem submetidos a uma reforma igualmente profunda. Caso contrário, ou seja, se os sistemas de avaliação e controle do desempenho econômico dessas empresas não se modificarem, as mudanças organizacionais serão, na melhor das hipóteses, bloqueadas e, na pior, produzirão (já estão produzindo) uma série de efeitos negativos.

Um dos sistemas de apoio à gestão da empresa que necessitaram de uma reforma profunda é o SMD. Eles se originaram nos fins do século XIX e se tornaram inadequados ao novo ambiente operacional e competitivo com o qual a maioria das empresas opera atualmente. Isso se deve ao desenvolvimento de novas técnicas de gestão da produção e o aumento do uso da tecnologia e automação pelas empresas (MENEZES, 2002).

QUADRO 2.1 – Tipos de Redes de produção

Tipos de Redes	Definição	Exemplo
Rede estratégica	guiada estrategicamente por uma grande empresa, a rede é relativamente estável e orientada para junção de firmas para obter vantagens estratégicas. Os membros da rede são ligados fortemente à firma central, mas conservam sua autonomia e competitividade. Esta forma de rede é mais adequada para mercados estáveis.	uma empresa do setor de eletrônicos orientada estrategicamente para o desenvolvimento de produto atua como firma central, e as empresas de manufatura, como fornecedores dependentes da estratégia desta.
Empresa virtual	firmas independentes que trabalham juntas para explorar uma oportunidade de negócio unindo suas manufaturas. A cooperação é temporária e dura apenas na vigência do projeto. As firmas têm específicas competências que podem ser combinadas sinergicamente.	associação de empresas do setor de auto-peças com montadora de automóveis para a criação de um novo modelo de carro.
Redes regionais	firmas altamente especializadas situadas espacialmente próximas em clima cooperativo uma com as outras. Os relacionamentos individuais não são estáveis e cada uma tem latentes relacionamentos com um grande número de potenciais parceiros.	empresas do setor têxtil que atuam em diferentes atividades (fiação, confecção, <i>design</i>); que estão concentradas e se relacionam cooperativamente em uma determinada região.
Redes operacionais	as transações são relativamente padronizadas e há preocupação em adicionar valores simples às atividades ao invés de aumentar a complexidade do processo. Um sistema de informação inter-organizacional forma a base para o uso de mecanismo de mercado a fim de coordenar as trocas.	um grande hipermercado mantém uma rede estável com diversos fornecedores. A coordenação da informação visa obter mercadorias para o hipermercado nas melhores condições custo/ benefício.

Fonte: Adaptado de Pfohl & Buse (2000).

Fazendo um histórico da evolução dos SMD, segundo Ghalayini & Noble (1996), a primeira fase da medição de desempenho, que teve início por volta de 1880, era voltada para as medidas de desempenho financeiras como lucro, retorno sobre capital investido e produtividade dos recursos.

No início da década de 80 do século XX, o meio acadêmico e empresarial começou a perceber fortes limitações das medidas financeiras de cunho contábil. Segundo Ittner & Lacker (1998), as limitações dessas medidas são:

- 1) são medidas de desempenho históricas e voltadas apenas para a análise do passado;
- 2) falta-lhes capacidade preditiva em explorar o desempenho futuro;
- 3) estão voltadas para o curto prazo podendo gerar comportamento individual incorreto;
- 4) provêm poucas informações para soluções de problemas novos e recorrentes;
- 5) não capturam em tempo suficiente as mudanças nos negócios;

- 6) são medidas de desempenho muito agregadas e sumarizadas dificultando uma ação gerencial;
- 7) refletem a estrutura funcional da organização e não os processos inter-funcionais que vêm ganhando importância atualmente; e
- 8) têm dificuldade na quantificação dos aspectos intangíveis na organização como, capital intelectual.

Como respostas a essas mudanças ambientais, tanto internas quanto externas, a partir da década de 80 surgiram os novos SMD, que buscaram inserir novas prioridades competitivas (não voltadas somente para custo), tecnologias e novas filosofias de administração da produção (GHALAYINI & NOBLE, 1996).

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., 1997).

Dada a relevância desse novo conjunto de medidas de desempenho, um modelo de medição de desempenho eficaz (atendimento dos objetivos) e eficiente (uso dos recursos disponíveis) deve considerar os seguintes fatores, que mudaram a realidade das organizações em sua maioria (NEELY, 1999):

1. mudança na natureza do trabalho: Na década de 50 e 60 do século XX o trabalho direto (produção) representava mais de 50% do custo do produto, a partir da década de 80, o trabalho direto está entre 5 e 10%. Estes dados mostram a necessidade de mudança do sistema tradicional de medidas financeiras;
2. aumento da competição: O crescimento da competição em nível global, como, por exemplo, a abertura dos mercados, a desregulamentação de diversos setores (telefonia, energia elétrica, dentre outros), geraram impactos no sistema de medição de desempenho que são:
 - busca das empresas em se diferenciar dos competidores em termos de qualidade de serviço, flexibilidade, inovação e rapidez na resposta;

- conciliação do SMD com o planejamento estratégico da organização; e
 - redução dos níveis hierárquicos dentro das empresas, com o intuito de facilitar os meios de comunicação na organização.
3. iniciativas de melhoria específica: surgimento de técnicas ligadas à medição de desempenho, como o método Taguchi, ciclo de DEMING (qualidade), custos da qualidade, Controle Estatístico de Processo (CEP), entre outros. Tanto o *Total Quality Management* (TQM) como o *World Class Manufacturing* (WCM) consideram o SMD de vital importância;
 4. prêmios da qualidade nacionais e internacionais: Os prêmios de reconhecimento incentivam as empresas a se manterem competitivas. Elas fornecem um completo diagnóstico da organização em vários enfoques como Política, Organização, Informação, Padronização, Recursos Humanos, Garantia da Qualidade, Manutenção, Melhoria, Efeitos e Planos Futuros, o que equivale dizer que são uma forma de auto-avaliação da medição do desempenho do negócio;
 5. mudanças nas regras organizacionais: Há uma ênfase na maior participação de profissionais na área contábil e do pessoal de recursos humanos no desenvolvimento de medidas de desempenho. As medidas de desempenho passaram a integrar o sistema de gestão de desempenho, sendo relacionadas com recompensas e bônus;
 6. mudanças das demandas externas: aumento da pressão de diversos *stakeholders* (consumidores/clientes, da sociedade, dos competidores, entre outros); destaques para a desregulamentação de vários setores públicos em que a medição de desempenho se tornou um fator de cobrança da sociedade, segundo o qual o consumidor dá maior valor ao produto do que ao preço. Outra questão é a mudança da relação cliente/fornecedor, muito presente na indústria automobilística, que se caracteriza pela intensa pressão das montadoras sobre os fornecedores e também pelo intuito de seleciona-los; e

7. o poder da tecnologia de informação: O papel da TI é importante hoje em dia para a captura e análise de dados; verifica-se rápido crescimento da demanda do sistema de informação para o gerenciamento do processo de negócios e também para executivos.

Os novos SMD apresentam diversas características destacadas por Neely et al. (1995) e Martins (1998). Convém citar que nem todas elas estão presentes nos sistemas encontrados nas empresas, pois fatores como tamanho da empresa, grau tecnológico envolvido e modelo de negócios influenciam na presença ou não dessas características; contudo o predomínio delas auxilia na caracterização do SMD na empresa. As características seriam:

- ser congruente com a estratégia competitiva da organização;
- ter medidas de desempenho financeiras/ não financeiras e internas/ externas;
- direcionar e dar suporte a melhoria contínua;
- identificar tendência e progressos;
- facilitar o entendimento das relações de causa-e-efeito;
- ser facilmente inteligível para os funcionários;
- abranger todo o processo de negócio, desde o fornecedor até o cliente;
- disponibilizar informações sobre o desempenho em tempo real para toda a empresa;
- ser dinâmico;
- influenciar a atitude dos funcionários: os SMD atuam como base para o pagamento de recompensas na organização;
- avaliar o grupo e não somente o indivíduo;
- permitir o *benchmarking* competitivo;
- ser composto de medidas de desempenho de eficiência e eficácia;
- estar ligado ao processo de negócios;
- ter um número pequeno de medidas de desempenho;
- ser parte do aprendizado organizacional e individual;
- ser composto de medidas de desempenho do processo integrado e do resultado;
- estar integrado com outros sistemas de gestão; e

- mostrar o desempenho do passado, do presente e do futuro.

Portanto, pelas características apresentadas, um sistema de medição de desempenho 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).

2.2 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.

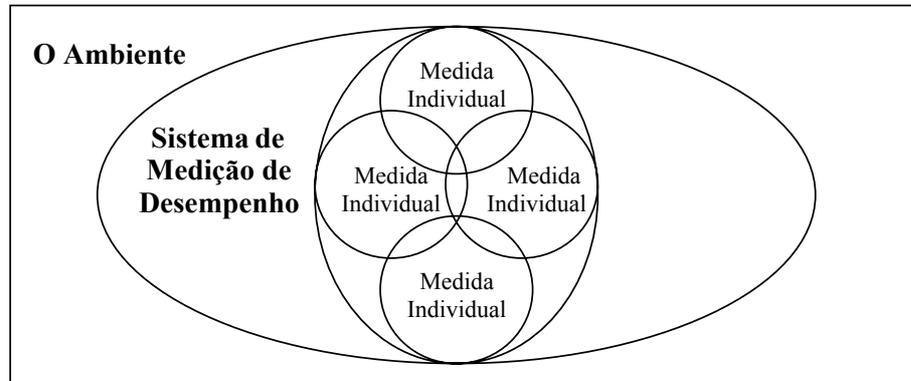
Neely (1998, p.5) assim se manifesta sobre uma definição de medição de desempenho que abrange o aspecto da infra-estrutura de suporte:

Um sistema de medição de desempenho possibilita que decisões e ações sejam tomadas com base em informações³ 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.

A medição de desempenho abrange um conjunto de pessoas, processos, métodos e ferramentas, que geram, analisam, expõem, avaliam e revisam dados e informações sobre as múltiplas dimensões de desempenho nos níveis individual, grupal, operacional e geral da organização, em seus diversos elementos constituintes (MACEDO-SOARES & RATTON, 1999).

³ Convém destacar a diferença entre dados e informação. Dados são ruídos ou imagens coletadas, por meio de processos organizacionais, nos ambientes interno e externo que não foram processados, correlacionados, integrados, avaliados ou interpretados de qualquer forma. Já na informação os dados passam por algum tipo de processamento (formatação, tradução, fusão, impressão, entre outros) para serem exibidos em uma forma inteligível às pessoas que a utilizam (MORESI, 2000).

Um modelo apresentado por Neely et al. (1995) de como a medição de desempenho pode ser esboçada está ilustrada na Figura 2.2.



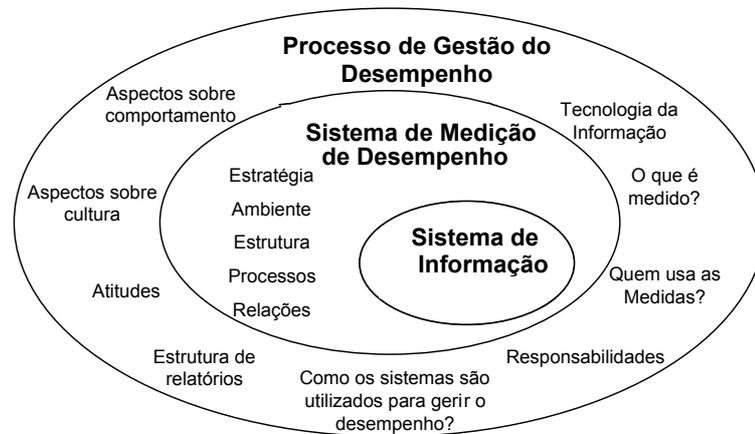
Fonte: Neely et al. (1995, p.81).

FIGURA 2.2 – Um modelo para o *design* da medição de desempenho.

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 2.3) 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 2.3 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. Este 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.



Fonte: Adaptado de Bititci et al. (1997, p.525).

FIGURA 2.3 – Posicionamento do sistema de medição de desempenho em uma organização

Já a dimensão ambiente trata de como as medidas individuais e/ou o SMD são utilizados em relação aos ambientes interno (organização) e externo (concorrência e consumidores) (NEELY et al., 1995).

A medição de desempenho possui várias razões para ser utilizada. Destaca-se dentre as razões, os 4CP's apresentados por Neely (1998), que atuam nas atividades: a) “checar” o posicionamento no mercado; b) “comunicar” a posição para a organização e os *stakeholders*; c) “confirmar” prioridades para o gerenciamento, tomada de decisão, controle de custos e investimento; e d) “compelir” o progresso como significado para motivação, comunicação das prioridades da organização e base para recompensa.

A medição de desempenho pode ser usada nas organizações para servir como uma linguagem a serviço da retórica (forma de convencimento), induzir uma ação robusta (forte consistência) nos membros da organização e ser a base para o pagamento de recompensas que cria uma certa ambigüidade (entre os objetivos da organização e de cada funcionário) que preserva a identidade individual de cada um (ECCLES⁴ et al., 1992 *apud* MARTINS, 1998).

⁴ ECCLES, R. G.; NOHRIA, N.; BERKLEY, J. D. Beyond the hype – rediscovering the essence of management. Boston, Harvard Business School Press, 1992.

2.2.1 A importância dos relacionamentos entre as medidas de desempenho

Para que um SMD seja eficiente e eficaz, é necessário que as informações disseminadas sejam úteis para a tomada de decisão e ação. Para tanto, certos elementos são necessários (MARTINS, 1998):

- existência de uma rede de relacionamento entre as medidas de desempenho;
- uso das informações 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 será 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. Em um estudo empírico, foi observado que cada entrevistado da medição de desempenho tinha o seu próprio modelo mental, mas um modelo formal de relacionamento inexistia (MARTINS, 1998).

A importância do relacionamento entre as medidas de desempenho é assim destacada por Eccles & Pyburn (1992, p.42):

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 freqüentemente 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, p.115), 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 Brown (1996), 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. (1997, p.218), “atribuir prioridade aos esforços de maximização do retorno das atividades de melhoria de desempenho.”

Geralmente os modelos mentais dos usuários da medição de desempenho são incompletos e podem gerar conflitos na hora da tomada de decisão resultando em um desempenho fraco, pois não há consenso de como é o desempenho desejado ou de como este pode ser atingido. Por outro lado, a formalização de uma rede de relacionamento pode reduzir o efeito dos modelos mentais individuais. Assim, o suporte dado por um SMD à tomada de decisão e a condução da ação podem ser melhorados (MARTINS, 1998).

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 armazenados. As propostas procuram formalizar o relacionamento esperado assim como no relacionamento de causa-e-efeito expresso no mapa estratégico do *Balanced Scorecard* (KAPLAN & NORTON, 1996a).

2.2.2 O 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 manufatura⁵ 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

⁵ A estratégia pode atuar para diversas finalidades, destaca-se nesta dissertação, o uso da estratégia para a manufatura. Entende-se por manufatura um sistema que engloba as atividades de produção e atividades inter-relacionadas como desenvolvimento de produto/ serviço, engenharia/ suporte técnico, *marketing*, compras, contabilidade e finanças e recursos humanos (SLACK et al., 1999).

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 externalização⁶ das relações entre elas facilitando o entendimento dos decisores sobre a estrutura de um SMD e 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.

⁶ Externalização significa a transformação do conhecimento tácito em explícito (FERRARI & TOLEDO, 2001).

2.3 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.

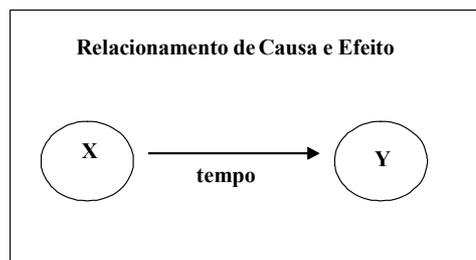
O propósito deste item é apresentar como os relacionamentos entre as medidas de desempenho de um SMD podem ser denominados ou representados em suas diversas formas. Este item é útil na caracterização dos relacionamentos entre as medidas de desempenho.

- 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*.

2.3.1 As 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 (1996a), 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 (Figura 2.4).



Fonte: Elaborado pelo autor.

FIGURA 2.4 – Relacionamento de Causa-e-Efeito

No trabalho da mesma autora, é apresentada outra forma de relações entre as medidas de desempenho que podem estar presentes no modelo *Balanced Scorecard*, pois, na sua opinião, elas são descritas de forma ambígua por Kaplan & Norton (1996b).

Segundo Norreklit (2000), esses autores justificam a existência do relacionamento entre as medidas de desempenho por conta da casualidade (o mesmo princípio da Figura 2.4) e também justificam a mesma lógica com base na avaliação financeira das ações pelo *Activity-Based Costing*⁷ (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.

⁷ O modelo ABC vê o desempenho como a busca da minimização dos custos dos processos internos de negócios que geram valor para os clientes (LEBAS, 1995).

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.

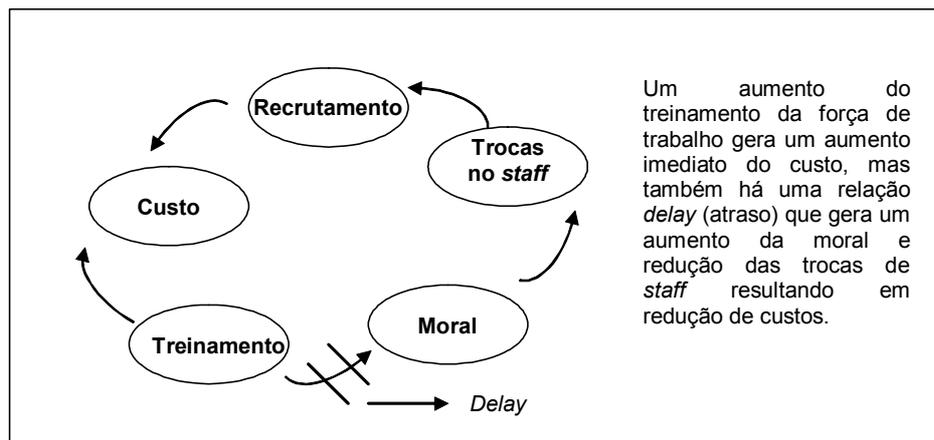
Existe também o relacionamento entre as medidas de desempenho baseado na interdependência em que estão presentes correlações positivas e negativas entre essas medidas de desempenho. Por exemplo, uma empresa cujo setor que atua investe fortemente em pesquisa e desenvolvimento (P&D) tem por objetivo bons 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, baseado 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 *delay* 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). A Figura 2.5 apresenta um exemplo.



Fonte: Davis & O'Donnel (1997, p.19).

FIGURA 2.5 – Diagrama de influência que apresenta as relações *delay*.

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 (BRYMAN, 1989). 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).

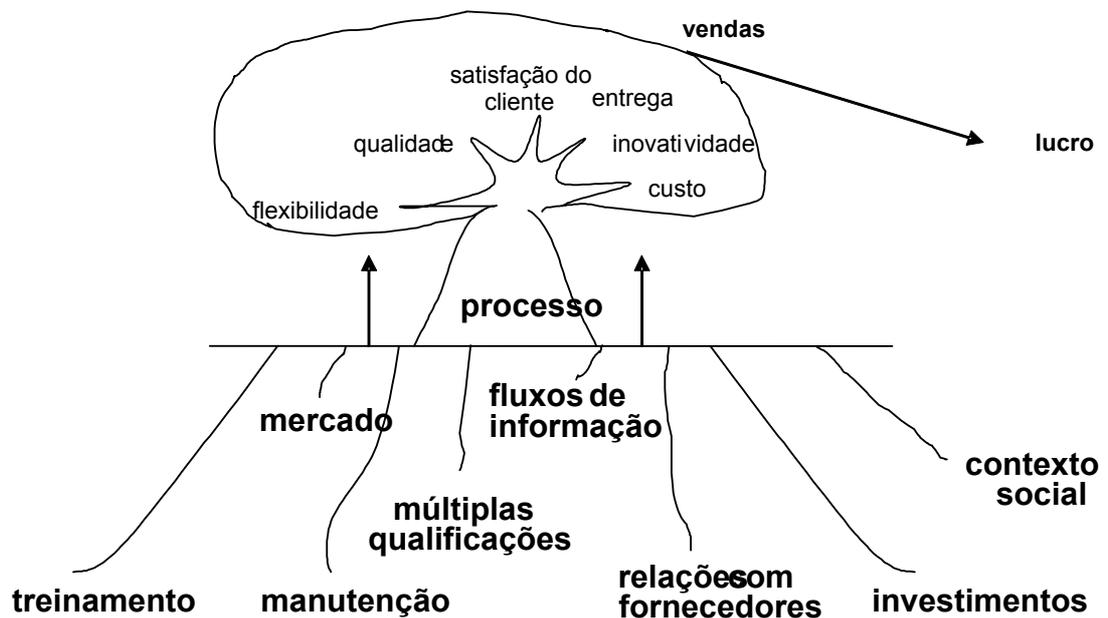
Por exemplo, uma empresa de alimentos fabrica produtos com alta velocidade e os entrega sem atraso para os distribuidores. Esse evento ocasionaria um bom resultado financeiro, porém a empresa não percebe pequenas mudanças no comportamento dos consumidores, que podem ser medidas pelo seu grau de satisfação,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

2.3.2 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 neste 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, como mostrado na Figura 2.6, 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.

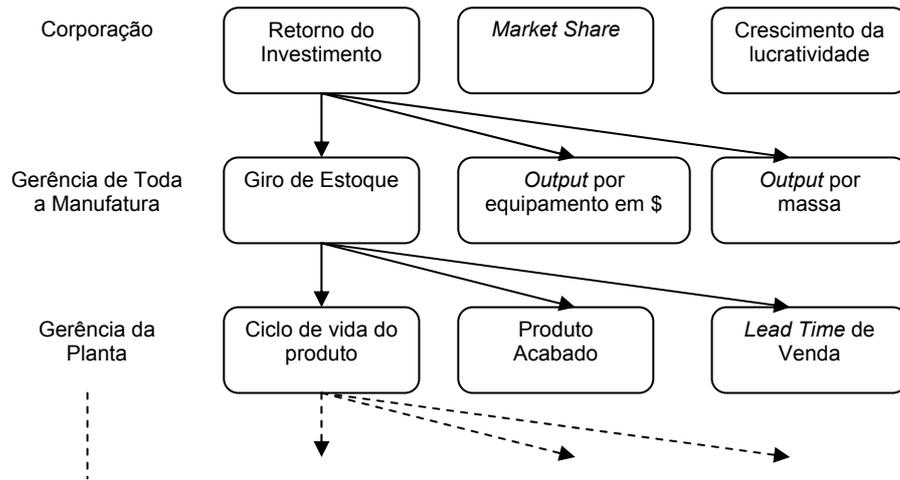


Fonte: Lebas (1995, p.28)

FIGURA 2.6 – Modelo causal

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 (Figura 2.7). Nessa figura, é mostrado o desdobramento a partir de uma medida de desempenho que é o retorno de investimento.

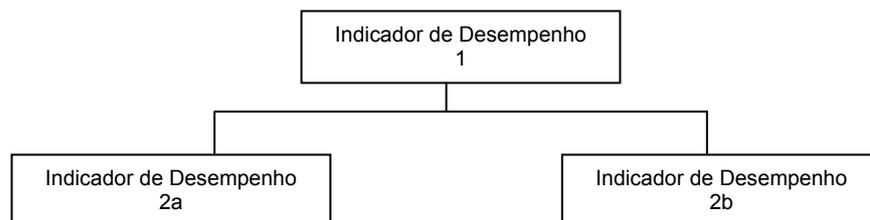
Descrevendo o exemplo da Figura 2.7, o retorno do investimento de uma organização só é satisfatório se houver um rápido giro de estoque, ou seja, uma rápida venda dos produtos aliada uma alta produtividade (em termos financeiros) dos equipamentos e também uma alta quantidade (utilizando-se um indicador de desempenho físico) de saída dos produtos vendidos.



Fonte: Lebas (1995), p.28.

FIGURA 2.7 – Desdobramento no modelo causal

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” conforme é ilustrado na Figura 2.8.

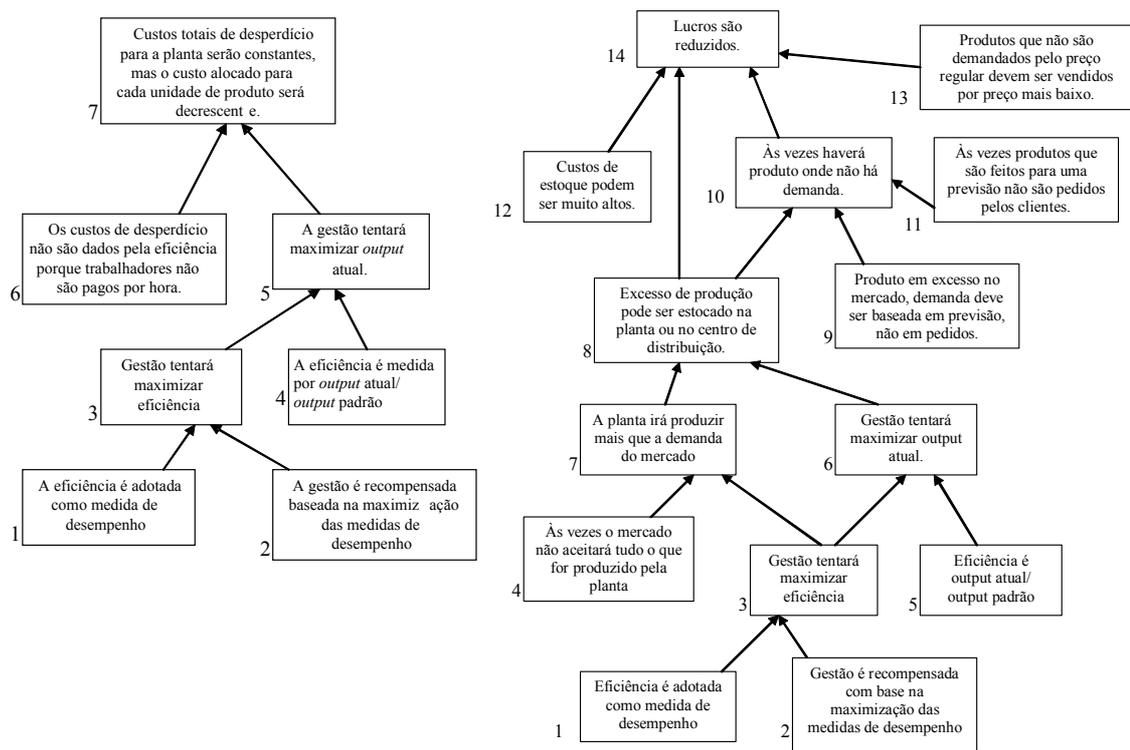


Fonte: Flapper et al. (1996, p.33).

FIGURA 2.8 – Relacionamentos entre indicadores de desempenho baseado em “pais” e “filhos”.

Nota-se, na Figura 2.8, que o indicador de desempenho 1 (pai) pode relacionar com mais de um indicador de desempenho (2a e 2b). O diagrama “pais” e “filhos” é útil nos relacionamentos verticais entre os indicadores de desempenho.

Um método apresentado por Boyd & Cox III (1997), denominado de *negative branch*, que analisa o impacto das decisões tomadas no desempenho organizacional, ilustrado na Figura 2.9.



Fonte: Boyd & Cox III. (1997, p. 29 e 30).

FIGURA 2.9 – Dois cenários para a medida de desempenho “eficiência” a partir do uso do *Negative Branch*

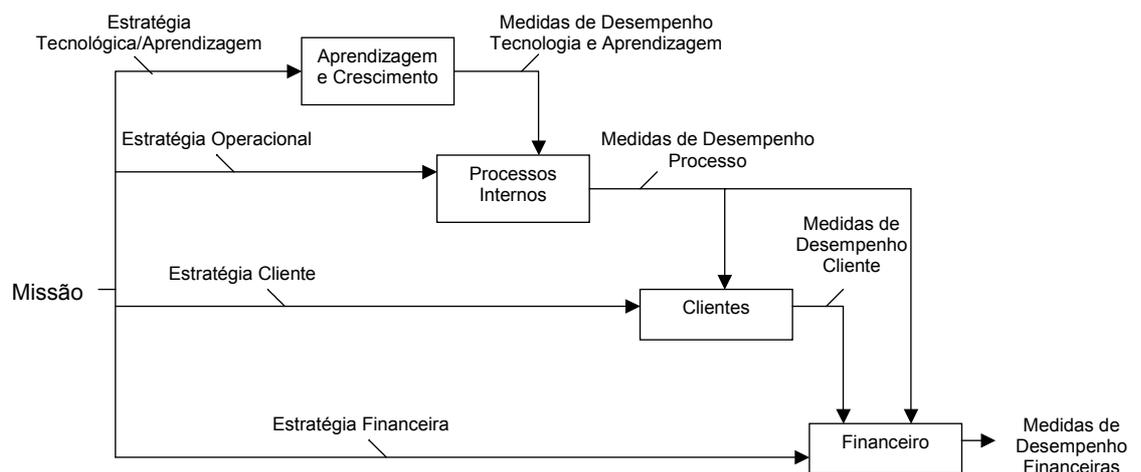
O método apresenta os seguintes passos na sua construção:

- 1 a apresentação dos possíveis efeitos positivos que são esperados das ações propostas;
- 2 a apresentação dos possíveis efeitos negativos que são esperados das ações propostas;
- 3 a conexão das soluções propostas com os possíveis efeitos positivos e negativos dos relacionamentos de causa-e-efeito; e

- 4 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. Estes dependem das escolhas das conexões entre os efeitos negativos e positivos das ações propostas. Na Figura 2.9, a hipótese proposta é se a eficiência pode ser adotada como medida de desempenho. São utilizados nesta figura dois cenários de uma empresa de processamento de papéis, tendo como ponto de partida a adoção da medida de desempenho “eficiência”. O primeiro cenário trabalha com a questão dos custos decrescentes, e o segundo, com a questão dos efeitos negativos quando a produção se torna maior que a demanda. Pode-se observar que os dois cenários trabalham com relações de causa-e-efeito.

Heredia & Natarajan (1997) apresentam um diagrama de construção de indicadores de desempenho estratégicos que mostra as estratégias, objetivos e medidas de desempenho (Figura 2.10, onde são utilizadas as perspectivas do *Balanced Scorecard*). 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.



Fonte: Heredia & Natarajan. (1997, p. 64).

FIGURA 2.10 – Diagrama de Indicadores de Desempenho Estratégicos

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?”.

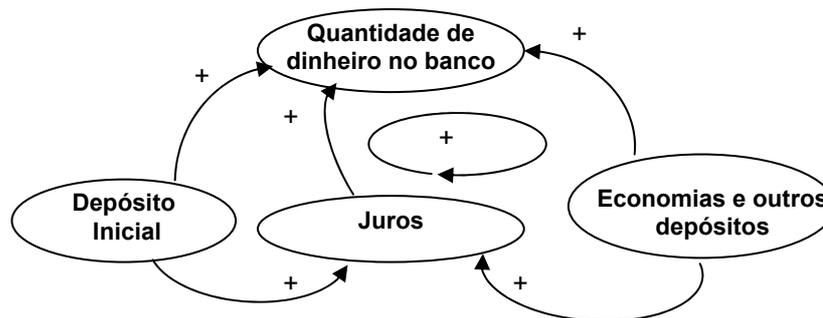
Respondendo essas três questões, inicia-se a construção de um diagrama de árvore, e o processo de uso das questões para outra medida de desempenho é repetido até serem todas mapeadas no diagrama. O diagrama de árvore mostra as relações entre as medidas de desempenho definidas pelas fórmulas de cálculo 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. Este 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).

Para explicar e ilustrar os próximos quatro métodos (mapa cognitivo, diagrama de causa e efeito, diagrama de árvore e *Analytic Hierarchy Process*, também chamado de AHP), é apresentado um exemplo em Suwignjo et al. (2000) no qual a quantidade de dinheiro depositado é influenciada pelos fatores depósito inicial, juros e economias e outros depósitos.

O primeiro método apresentado é o mapa cognitivo, que identifica os efeitos dos fatores no desempenho, como mostrado na figura 2.11. 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.

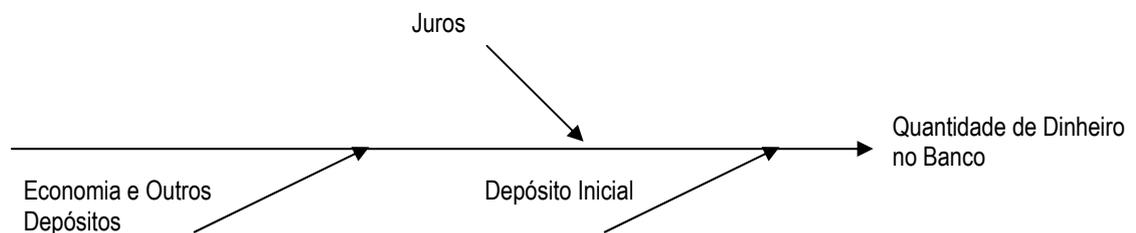
A Figura 2.11 mostra que o fator “quantidade de dinheiro no banco” é aumentado a partir da relação direta com o fator “juros”, já os fatores “depósito inicial” e “economias e outros depósitos” se relacionam, tanto direta quanto indiretamente, a partir do fator “juros” com o fator “quantidade de dinheiro no banco”. Existe também o efeito *self interactive* no fator “juros” em que a quantidade de dinheiro recebida no mês decorrente do fator “juros” irá aumentar o mesmo fator no mês seguinte.



Fonte: Suwignjo et al. (2000, p. 233).

FIGURA 2.11 – Mapa Cognitivo.

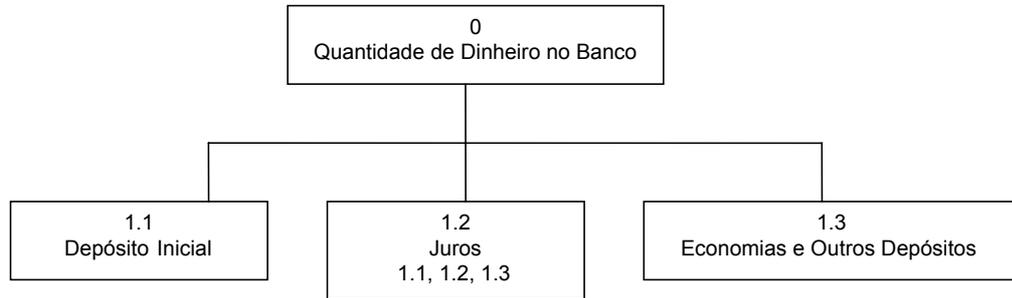
O diagrama de causa-e-efeito ajuda a identificar a estrutura hierárquica das medidas de desempenho. Existe um nível 0, que é afetado por todas as outras medidas de desempenho, mas ele não afeta todas as outras medidas de desempenho e, conseqüentemente, o próximo nível afeta o nível 0, mas não as outras medidas de desempenho. No caso da Figura 2.12, o nível 0 é o fator “dinheiro guardado no banco”, que é afetado por todos os outros fatores.



Fonte: Suwignjo et al. (2000, p. 233).

FIGURA 2.12 – Diagrama de Causa-e-Efeito

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. Na Figura 2.13, o fator “quantidade de dinheiro no banco” é o primeiro nível hierárquico, enquanto os outros fatores pertencem ao segundo nível hierárquico. Nota-se que o fator “juros” também afeta indiretamente a “quantidade de dinheiro no banco” por meio dos outros fatores de desempenho.



Fonte: Suwignjo et al. (2000, p. 233).

FIGURA 2.13 – Diagrama de Árvore

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 (Figura 2.14).

Nível: 0										
Fator: Quantidade de dinheiro no banco										
Sub –Fatores: Depósito inicial, Juros e Economia e Outros Depósitos										
Coluna	Absoluto	Muito Forte	Forte	Fraco	Igual	Fraco	Forte	Muito Forte	Absoluto	Coluna
1.1 Depósito inicial			V							1.2 Juros
1.1 Depósito inicial							V			1.3 Economias e outros depósitos
1.2 Juros								V		1.3 Economias e outros depósitos

Fonte: Suwignjo et al. (2000, p.235).

FIGURA 2.14 – Questionário de comparação dos fatores de desempenho

Na Figura 2.14, é feita a comparação de fatores verificando-se qual é o maior efeito no fator “quantidade de dinheiro no banco”. Por exemplo, comparando os fatores “juros” e “depósito inicial”, verifica-se que o segundo fator causa um efeito forte no fator quantidade de dinheiro no banco.

A segunda etapa do processo de quantificação, seguindo o exemplo mostrado em Suwignjo et al. (2000), é feita a partir da resposta à questão “Quão forte é

esse efeito?” Com isso são estabelecidas relações entre os fatores de desempenho em bases quantitativas (Tabela 2.1).

TABELA 2.1 – Matriz de comparação dos fatores de desempenho

Quantidade de Dinheiro no banco	Depósito inicial	Juros	Economias e outros depósitos	Prioridade
Depósito inicial	1	5	1/5	0,212
Juros	1/5	1	1/8	0,062
Economias e outros depósitos	5	8	1	0,726

Fonte: Suwignjo et al. (2000, p.235).

A Tabela 2.1 é o resultado das comparações feitas a partir da primeira etapa do processo (exemplo da Figura 2.13). 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. Por exemplo, a prioridade do fator “depósito” é feito pela soma do seu efeito direto e pelo seu efeito nos outros fatores. Já os outros efeitos são nulos. No exemplo utilizado por Suwignjo (2000), foi utilizado um *software* para a realização do cálculo.

Várias análises podem ser feitas a partir do método, como, por exemplo, o fator depósito inicial tem um efeito cinco vezes mais forte em relação ao fator juros. Outra análise possível da Tabela 2.1 é que o fator depósito inicial tem o segundo maior efeito no fator quantidade de dinheiro no banco.

Neste item, foram apresentados diversos métodos que podem ser usados no diagnóstico ou estabelecimento de relacionamentos entre as medidas de desempenho. Cada método, conforme mostrado no Quadro 2.2, tem suas vantagens e desvantagens na forma de elas serem aplicadas na organização.

QUADRO 2.2 – Vantagens e desvantagens dos métodos para a construção de relacionamentos entre as medidas de desempenho

Método	Vantagens	Desvantagens
Modelo Causal	permite uma visão encadeada de como o desempenho pode ser medido e de como ele pode afetar a gestão da empresa.	o modelo causal é generalista, não detalhando a forma como esse modelo pode ser utilizado. Por exemplo, como a manutenção afeta a dimensão qualidade na empresa?
Desdobramento do Modelo Causal	permite verificar como as medidas de desempenho afetam os outros níveis organizacionais da empresa.	o método não abrange as relações horizontais entre as medidas de desempenho.
“Pais” e “Filhos”	é eficaz na abrangência da relações verticais entre as medidas de desempenho.	não abrange relações horizontais entre as medidas de desempenho e também não situa a medida de desempenho em um dado nível organizacional.
<i>Negative Branch</i>	permite o desenvolvimento de cenários sobre os possíveis efeitos de uma ou mais medidas de desempenho.	a aplicação do método requer uma forte liderança e a atividade tem que estar bem organizada, pois membros da organização com linguagens diferentes estarão participando do processo.
Diagrama de Construção de Indicadores de Desempenho Estratégicos	auxilia o alinhamento das medidas de desempenho com os objetivos estratégicos da organização	o método não abrange as medidas de desempenho que não são estratégicas e também não diz o que acontecerá com as medidas de desempenho que não fazem mais parte da estratégia.
Diagrama Para a Construção de Relacionamentos entre as Medidas de Desempenho	permite o detalhamento das medidas de desempenho facilitando o entendimento sobre qual é o uso desta medida.	o método não é aplicado para relações entre as medidas de desempenho que não sejam matemáticas.
Mapa Cognitivo	abrange diversas formas de relacionamento entre as medidas de desempenho (vertical, horizontal, financeira, não financeira, entre outros).	não permite identificar a intensidade das relações entre as medidas de desempenho.
Diagrama de Causa e Efeito	posicionamento dos efeitos das medidas de desempenho em escala temporal.	não abrange relações entre medidas de desempenho que ocorrem não temporalmente como, por exemplo, as relações lógicas
Diagrama de Árvore	permite o posicionamento hierárquico das medidas de desempenho frente a um dado nível organizacional	não abrange relações horizontais entre as medidas de desempenho
<i>Analytic Hierarchy Process</i>	quantifica o quão forte é a relação entre as medidas de desempenho	o processo de quantificação envolve etapas que requerem escolhas subjetivas dos participantes do processo, e pode distorcer o resultado.

Fonte: Elaborado pelo autor.

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; e

- com o tipo de ação a ser tomada na organização.

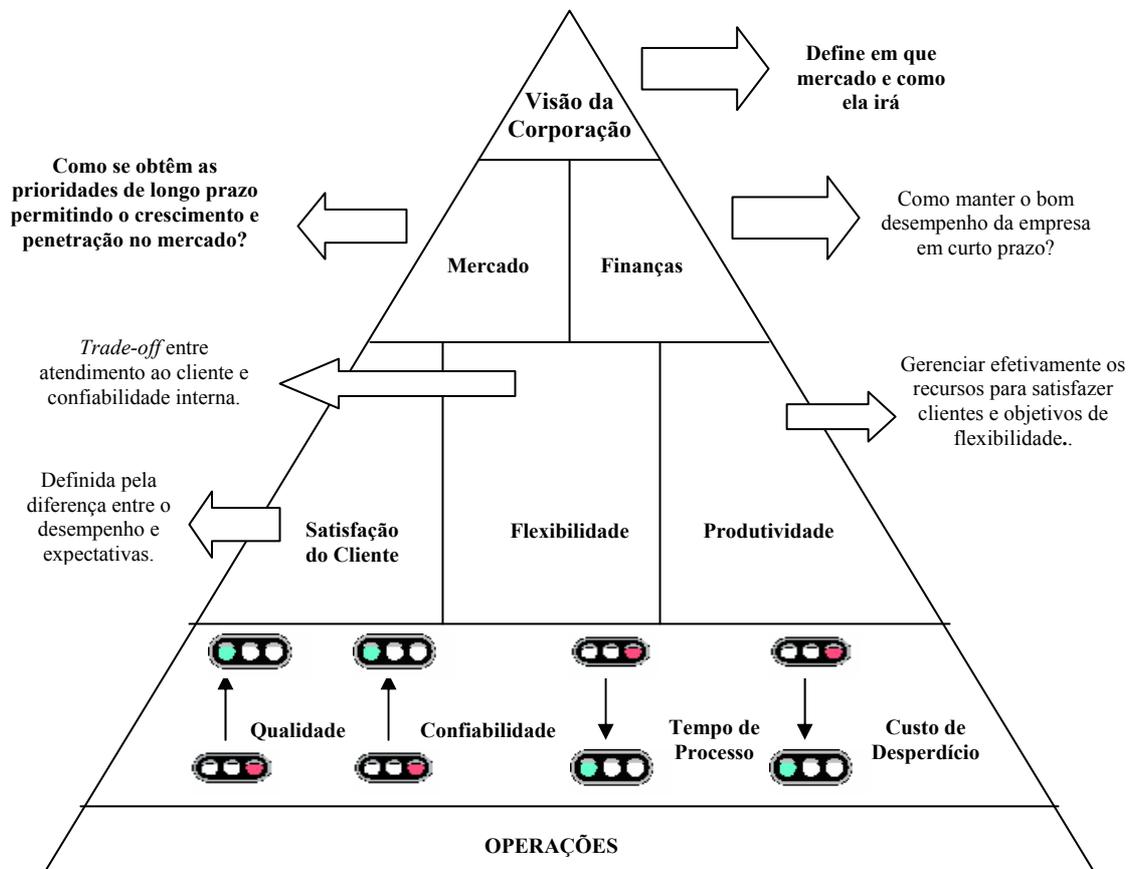
2.3.3 Os relacionamentos entre as medidas de desempenho nos modelos de medição de desempenho

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.

Este item tem o intuito de explorar 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*.

O *Strategic, Measurement, Analysis, and Reporting Technique* (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 2.15 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 2.15, 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).



Fonte: Adaptado de Cross & Linch (1990).

FIGURA 2.15 – Estrutura do modelo de medição de desempenho *Performance Pyramid*

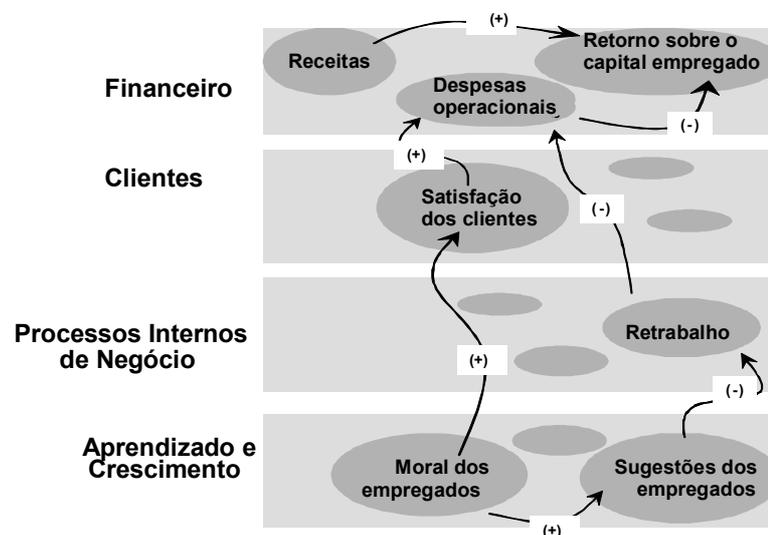
Os relacionamentos entre as medidas de desempenho expressos pelos mesmos autores na *Performance Pyramid* se dão pela transição entre duas linguagens. A linguagem do “dinheiro” (informações provenientes de dados financeiros) para a alta administração, que possui a visão corporativa e que precisa traduzir 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.

O *Balanced Scorecard* 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 provê *feedbacks* 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, 1996a).

Sobre o relacionamento entre as medidas de desempenho, Kaplan & Norton (1996b) defendem a identificação e explicitação das relações, denominadas de causa-e-efeito, entre os objetivos e as medições das perspectivas do BSC para serem gerenciadas e validadas. Nota-se, na Figura 2.16, que a perspectiva financeira representa os resultados da organização, enquanto as demais são diretoras do resultado.



Fonte: Kaplan & Norton (1996c, p.83).

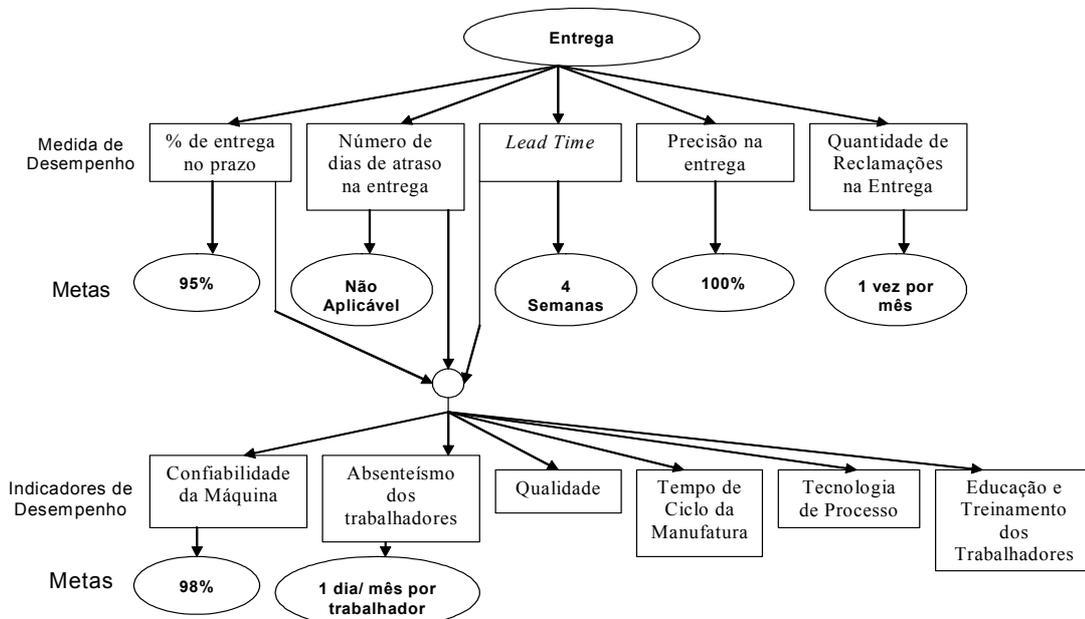
FIGURA 2.16 – Mapa estratégico no *Balanced Scorecard*

Para esse mapa estratégico, como visto na Figura 2.16, “estratégia é um conjunto de hipóteses sobre causa-e-efeito ... a cadeia de causa-e-efeito deve permear todas as quatro perspectivas” (KAPLAN & NORTON, 1996a, p.65).

O *Integrated Dynamic Performance Measurement System (IDPMS)* foi baseado na idéia de que a medição de desempenho deve ser flexível, relevante para cada nível gerencial e capaz de mudar quando as necessidades organizacionais mudam. Os

SMD antigos focam muito o controle e o monitoramento, e pouco destaque dão as atividades de melhoria. Há, também, a falta de uma associação entre melhoria e medição de desempenho. Para superar essas deficiências e também acompanhar os surgimento de novas ferramentas de gestão (*Total Quality Management*, JIT, entre outros), o IDPMS foi desenvolvido (GHALAYINI et al., 1997).

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 2.17.



Fonte: Ghalayini et al. (1997, p.219).

FIGURA 2.17 – Analisando o desempenho entrega no IDPMS

O IDPMS, segundo Ghalayini et al. (1997), possui uma estrutura que integra as áreas de sucesso, as medidas de desempenho e suas metas, e os indicadores de desempenho⁸ e suas metas. A Figura 2.17 mostra os quatro níveis de integração dos aspectos ligados à medição de desempenho para o fator desempenho da entrega.

O Integrated Performance Measurement System (IPMS) é visto como o meio para desdobrar (traduzir em objetivos) a visão e as estratégias da organização e

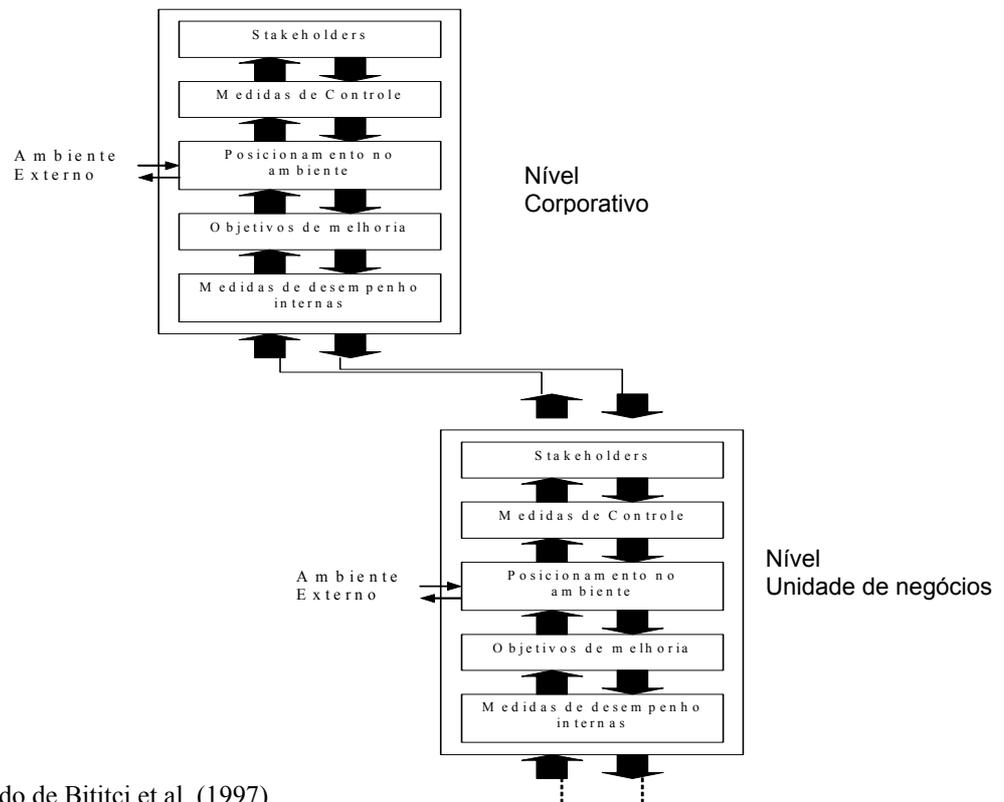
⁸ Para os autores, os indicadores de desempenho e suas metas estão voltados para o chão-de-fábrica e representam o desdobramento das medidas de desempenho e suas metas como pode ser visto na Figura 2.16. Nesta dissertação, medição de desempenho e indicadores de desempenho são considerados sinônimos.

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 a 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 2.18).

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 2.18, 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 internas estão sendo atingidas.



Fonte Adaptado de Bititci et al. (1997).

FIGURA 2.18 – Integração do Sistema de Medição de Desempenho no IPMS

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 destes. Um exemplo de medidas de desempenho é apresentado no Quadro 2.3 em uma organização não-governamental voltada para o atendimento de jovens ingressantes no mercado de trabalho.

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.

QUADRO 2.3 – Exemplo de medidas de desempenho para cada faceta do prisma

Facetas do Prisma	Satisfação dos Stakeholders	Estratégias	Processos	Capacidades	Contribuição dos Stakeholders
Elementos-Chave	Jovens Trabalhadores	Grupos de Melhoria	Eventos organizados	Pessoas	Fundadores
Medidas de Desempenho	Satisfação dos jovens trabalhadores + <i>survey</i> de tendências das necessidades	Número de novos produtos e serviços oferecidos	Número de participantes por evento	“Investimento em Pessoas” para cada nível	Nível e % de fluxo de caixa para cada tipo de rendimento

Fonte: Neely et al. (2001, p.11).

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. Por exemplo, no Quadro 2.3, analisando-se a faceta “processos”, verifica-se que, com o aumento do número de participantes por eventos voltados para os jovens trabalhadores haverá um impacto positivo na medida “satisfação dos jovens trabalhadores” e no nível de fluxo de caixa dos fundadores. Essa lógica pode ser estendida para as facetas “estratégia” e “capacidade”.

Dos principais modelos de medição de desempenho apresentados, existem certas características que são únicas para os modelos no que tange aos relacionamentos entre as medidas de desempenho, como mostrado no Quadro 2.4. Também nesse quadro são apresentadas críticas aos modelos mostrando-se a ausência de algumas características presentes nos relacionamentos entre as medidas de desempenho.

QUADRO 2.4 – Características e críticas dos relacionamentos entre as medidas de desempenho nos modelos de medição de desempenho

Modelos de Medição de Desempenho	Características dos Relacionamentos entre as Medidas de Desempenho	Críticas ao Modelo no que tange o Relacionamento entre as Medidas de Desempenho
<i>Performance Pyramid</i>	Abrange o relacionamento entre as medidas de desempenho financeiras e não-financeiras, especialmente na transição entre elas.	O modelo não explica de modo detalhado como é feita a transição entre as linguagens do “dinheiro” com a linguagem das “coisas”.
<i>Balanced Scorecard</i>	Integra os relacionamentos entre as medidas de desempenho financeiras e não-financeiras por meio de relações de causa e efeito.	Não levam em consideração outras formas de relacionamento como relações lógicas e interdependências.
IDPMS	Apresenta, de forma detalhada, as relações entre as medidas de desempenho da alta administração até o chão de fábrica	Embora o modelo de SMD abranja um componente dinâmico na empresa que é a melhoria, é difícil de ser aplicado em ambientes dinâmicos nos quais o modelo de negócios da empresa é regularmente questionado.
IPMS	Apresenta uma visão abrangente de como as medidas de desempenho se interagem em diferentes níveis organizacionais por meio do conceito de subsistemas	Não explica como é feita a interação entre as medidas de desempenho entre os subsistemas e também que tipo de relacionamentos entre as medidas de desempenho podem ser abrangidos nesse modelo
<i>Performance Prism</i>	Os relacionamentos entre as medidas de desempenho são estabelecidos a partir das demandas dos <i>stakeholders</i> .	O modelo não apresenta um procedimento de como são estabelecidos os relacionamentos entre as medidas de desempenho. São somente abertas possibilidades para a definição dos relacionamentos entre as medidas de desempenho, como o uso de sessões de <i>brainstorming</i> para a construção dos relacionamentos.

Fonte: Elaborado pelo autor.

Dentre as críticas gerais apresentadas no Quadro 2.4, 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*).

2.4 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. O primeiro aspecto 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 predefinidos (VAKKURI & MEKLIN, 2003).

O segundo aspecto 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 2.3.

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.

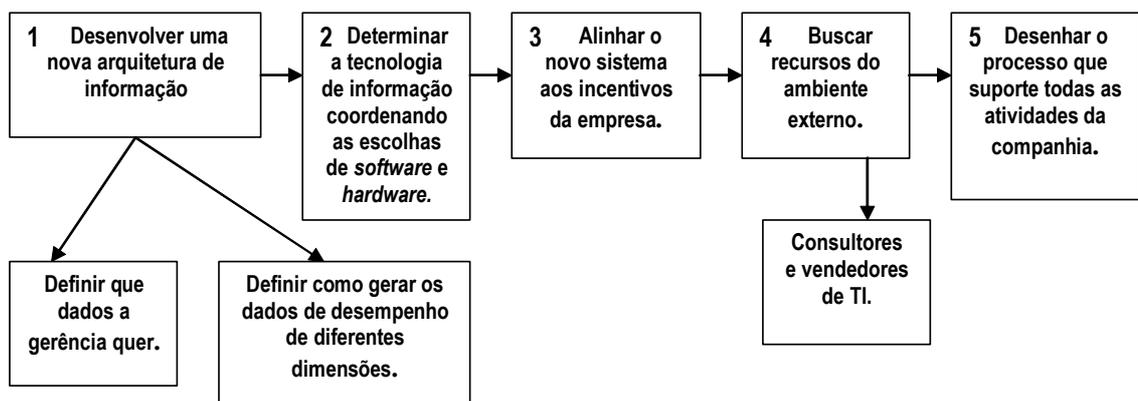
Segundo esse mesmo autor, existem quatro componentes em um SMD que podem formar um sistema de TI: (i) procedimentos suportados pela TI; (ii) dados; (iii) *software*; e (iv) *hardware*. O quinto componente, que, embora não forme um sistema de TI, é complementar a este sistema, seriam as pessoas. O Quadro 2.5 mostra as principais atividades de cada um dos cinco componentes.

QUADRO 2.5 – Componentes do Sistema de Medição de Desempenho

Pessoas	Procedimentos	Dados	Software	Hardware
<ul style="list-style-type: none"> - dono do processo; - contadores para medidas individuais; - pessoas que atuam e mantêm o SMD; - fornecedores de dados; e - usuários internos e externos do SMD; - <i>stakeholders</i> do SMD internos e externos. 	<ul style="list-style-type: none"> - regras e procedimentos para definição dos indicadores de desempenho; - regras para a gestão dos dados e informações; - regras para a comunicação dos dados e informações; e - regras para o uso dos resultados do SMD. 	<ul style="list-style-type: none"> - dados relevantes de desempenho; - relevância dos indicadores de desempenho; - informações obtidas de cálculos que geram resultados; - gestão dos metadados⁹ que descrevem os indicadores de desempenho. 	<ul style="list-style-type: none"> - <i>software</i> para extração, transformação e carregamento dos dados; - <i>software</i> para gestão da base dados ou <i>data warehouse</i>; - <i>software</i> para a análise dos dados; e - <i>software</i> de visualização e comunicação. 	<ul style="list-style-type: none"> - computadores pessoais ou outras unidades de visualização; - servidores; - infra-estrutura de comunicação; e - sistema de estocagem dos dados e informações.

Fonte: Kueng et al. (2001, p.8).

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 2.19 mostra as cinco atividades da medição de desempenho.



Fonte: Adaptado de Eccles (1991).

FIGURA 2.19 – Cinco atividades fundamentais da medição de desempenho.

Verifica-se na Figura 2.19 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

⁹ Ver item 3.1.3 para uma explicação do significado de metadados.

apropriada quanto na obtenção de informações do ambiente externo (clientes, concorrentes, entre outros).

Contudo, na realidade, a TI ainda é pouco considerada no auxílio a SMD destacado por Kueng et al. (2001, p.5):

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 *data mining* para a construção de modelos de relacionamentos a partir de dados históricos de desempenho.

Dentro do panorama apresentado neste item, este 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 do *data mining*.

2.5 Considerações e Síntese do Capítulo

Nas últimas décadas, o ambiente industrial passou por diversas mudanças tanto interna como externamente. As empresas buscaram novas formas de competição a fim de aumentarem sua competitividade superando o modelo tradicional centrado no fordismo e taylorismo. Houve uma mudança no ambiente interno, na qual a coordenação da informação ganha importância, e uma maior proximidade com os

stakeholders (principalmente com fornecedores e clientes) é verificada com o advento de novas filosofias de gestão (TQM, JIT, entre outros).

Todas essas mudanças geraram um impacto no sistema de gestão da empresa e, conseqüentemente, nos SMD. Houve neste o surgimento de novas características como a integração entre medidas de desempenho financeiras e não-financeiras, a busca do suporte do SMD para a melhoria e não somente para controle e monitoramento, e a integração com outros sistemas de gestão. Além disso, com os avanços da TI, é possível atualmente gerar informações sobre desempenho em tempo real para os tomadores de decisão.

Um aspecto que ganha ênfase no SMD, dada a questão das múltiplas dimensões na avaliação do desempenho, são os relacionamentos entre as medidas de desempenho, os quais o estabelecimento do relacionamento entre as medidas de desempenho auxiliam no diálogo entre os membros da organização, com possíveis modelos mentais diferentes.

Neste capítulo, foram apresentadas três formas de abordar a questão desses relacionamentos: por meio de denominações das relações entre as medidas de desempenho; de métodos que estabelecem relações entre essas medidas de desempenho; e de relacionamentos entre as medidas de desempenho presentes em modelos de medição de desempenho.

Verificou-se que, na denominação entre as relações entre as medidas de desempenho, as predominantes na literatura são as relações de causa-e-efeito; as outras relações são pouco abordadas dada a forte influência do *Balanced Scorecard*.

Já os métodos que estabelecem essas relações, embora existam certa variedade deles, estão centrados em três aspectos:

- 1) o efeito dos relacionamentos entre as medidas de desempenho no sistema de gestão na empresa;
- 2) as relações entre as medidas de desempenho em níveis organizacionais diferentes; e
- 3) busca do estabelecimento de relações entre as medidas de desempenho baseada em relações matemáticas.

Quanto aos modelos de medição de desempenho, a questão dos relacionamentos entre as medidas de desempenho é mais aprofundada em alguns deles

(BSC e IDPMS) e pouco nos outros. No geral, os modelos abordam relacionamentos entre as medidas de desempenho de níveis organizacionais diferentes, principalmente abordando a questão da relação entre medidas de desempenho financeiras e não-financeiras. É pouco considerada a questão dos relacionamentos horizontais entre as medidas de desempenho de um dado nível organizacional.

Por fim, destaca-se o aspecto do uso da TI no estabelecimento de relacionamentos entre as medidas de desempenho em termo de uso das abordagens, métodos e ferramentas na análise e disseminação das informações para a organização. Ela pode auxiliar o estabelecimento desses relacionamentos com o uso de abordagens modernas como o *data warehouse*, OLAP e *data mining*. Isto será visto no Capítulo 3.

3. A RELAÇÃO ENTRE *DATA MINING* E MEDIÇÃO DE DESEMPENHO

A tecnologia de informação pode auxiliar no estabelecimento de relacionamentos entre as medidas de desempenho em um sistema de medição de desempenho. O auxílio pode vir na coleta de dados históricos de desempenho a partir de um *data warehouse* e na análise desses dados, como, *drill down*, utilizando técnicas e ferramentas analíticas eficazes como a abordagem *data mining*.

O propósito do capítulo é mostrar como o *data mining* pode ser utilizado no estabelecimento de relacionamentos entre as medidas de desempenho de uma SMD.

Para tanto, é apresentado primeiramente o contexto em que o *data mining* atua na geração de conhecimento nas empresas. Destacam-se o papel do sistema de informação nas empresas e o uso da tecnologia de informação no apoio à tomada de decisão. Também é mostrada a importância do *data warehouse* e do OLAP no uso do *data mining*.

Tendo apresentado o contexto, busca-se detalhar a abordagem *data mining*. É mostrada uma visão geral da abordagem, operações analíticas, técnicas utilizadas e possíveis ferramentas que podem ser aplicadas em organizações.

Por fim, é discutida a relação entre *data mining* e medição de desempenho indicando-se as suas possíveis aplicações, principalmente no estabelecimento de relacionamentos entre as medidas de desempenho. Considerações sobre o capítulo são apresentadas.

3.1 Contextualização do *Data Mining* em um Ambiente Organizacional

Segundo Costa et al. (2000), a busca do conhecimento sempre foi uma das maiores atividades exercidas pelo homem. Nas últimas décadas, o mundo dos negócios armazenou uma grande quantidade de dados. Hoje, as empresas reagem mais rapidamente às mudanças no mercado precisando comparar e melhorar o seu desempenho continuamente para incorporar a melhor prática. Por exemplo, uma prática interessante seria a ferramenta de apoio denominada *Data Envelopment Analysis* (DEA), que relaciona um conjunto de dados contendo múltiplas entradas e saídas. O conjunto de dados possui diversos *Decision Making Unit* (DMU) que relaciona somente um *input* com um *output*. O DEA se propõe, por meio de modelagem matemática, a

detectar a DMU mais eficiente e também a usar esta DMU como comparação (*benchmarking*) para as outras DMU's (COOPER et al., 2004).

Por exemplo, um DMU que relaciona investimento em P&D/faturamento com ROI para cada empresa no setor de farmacêuticas. Busca-se nesse setor encontrar a DMU ótima para usá-la como meta de desempenho para outras empresas (COOPER et al., 2004).

Nesse contexto, surgiu nas últimas décadas o *data mining* como uma dos novos campos da Ciência da Computação. Com a geração de um volume cada vez maior de dados, é essencial tentar aproveitar o máximo possível o investimento em TI. Uma forma de utilizar esses repositórios de dados é tentar descobrir se há informações úteis escondidas neles. Um banco de dados de transações comerciais pode, por exemplo, conter diversos registros indicando os produtos que são comprados em conjunto. Quando isso é descoberto, podem-se estabelecer estratégias para otimizar os resultados financeiros da empresa. Essa já é uma vantagem suficientemente importante para justificar todo o processo. Contudo, embora essa idéia básica seja facilmente compreensível, há sempre uma dúvida sobre como um sistema é capaz de obter esse tipo de relação (NAVEGA, 2002).

Inicialmente, para o entendimento do contexto do *data mining*, é importante a compreensão do sistema de informação nas organizações destacando-se a sua importância atual e suas principais características. Dentre dessas características destaca-se a TI como uma ferramenta que vem ganhando importância com as mudanças no ambiente industrial (LAUDON & LAUDON, 2002).

Um dos tipos de sistema de informação que vêm ganhando importância com os avanços da TI é o sistema de apoio à decisão (SAD) que é útil para o processo decisório nas organizações. Destaca-se o surgimento de novas abordagens de TI dentro do SAD como o *On Line Analytical Processing* (OLAP), o *data warehouse* e o *data mining*.

3.1.1 Sistema de informação e tecnologia de informação

O ambiente industrial passou por diversas mudanças nas últimas décadas. Elas afetaram o ambiente interno das organizações, com destaque, nesta dissertação,

para os sistemas de informação. Segundo Laudon & Laudon (2002), os principais fatores que afetaram os sistemas de informação nas empresas foram:

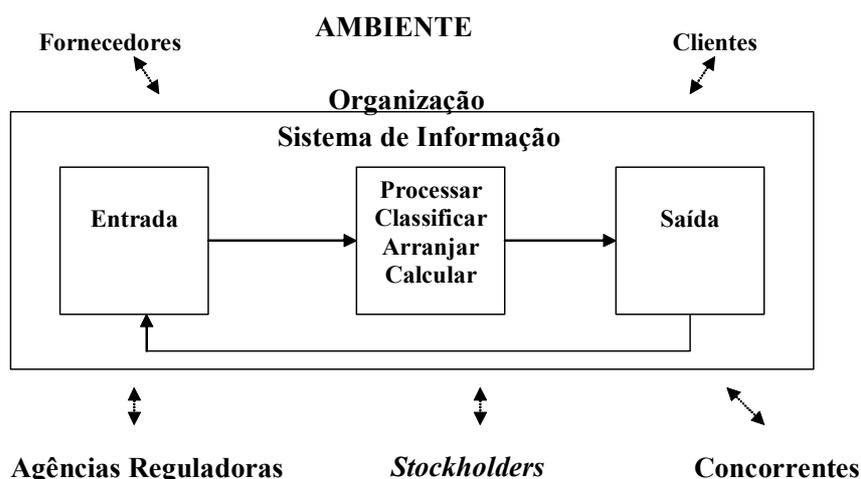
- a emergência e o fortalecimento da economia global: As firmas atualmente têm que competir em mercados globalizados e ter um sistema de distribuição, controle e gestão coordenado globalmente entre as firmas;
- a transformação socioeconômica industrial: Houve um crescimento da importância do conhecimento e dos serviços nas empresas tornando a base do surgimento de novos serviços e produtos;
- a transformação dos negócios das empresas: Houve mudança de um sistema de produção em massa para um sistema descentralizado, menos hierarquizado, com arranjo flexível, orientado para o cliente a fim de obter coordenação junto com os empregados e formação de grupos de trabalho; e
- a emergência das empresas digitais¹⁰: Vem ocorrendo desde meados da década de 90 do século passado com o intensivo uso da TI nos negócios da empresa.

Um sistema de informação pode ser definido de várias formas em uma organização. Laudon & Laudon (2002) o definem como uma interligação de componentes que trabalham juntos para coletar, processar, estocar, disseminar informações para suporte à tomada de decisão, à coordenação, ao controle, análise e à visualização. Um sistema de informação baseado em computadores é definido, segundo Laudon & Laudon (2002), como o uso confiável de *hardware* e *software* para processar e disseminar informações em uma organização.

Um sistema de informação pode ajudar o *staff* de uma organização a visualizar problemas complexos, analisar problemas e criar novos produtos. Ele produz as informações que as organizações precisam para tomar decisões, controlar operações, analisar problemas e criar novos produtos e serviços. A Figura 3.1 mostra as suas funções na visão de Laudon & Laudon (2002).

¹⁰ Uma empresa digital pode ser definida como uma organização em que todos os processos de negócios e os relacionamentos com clientes, fornecedores, empregados são realizados digitalmente, sem nenhum contato pessoal (LAUDON & LAUDON, 2002).

De acordo com essa figura, a entrada captura os dados coletados do ambiente interno e/ou externo da empresa para o processo do sistema de informação. O processamento realiza a conversão, manipulação e análise dos dados para torná-los mais compreensíveis para os usuários. A saída distribui a informação processada para os usuários que a utilizam nas atividades organizacionais.



Fonte: Laudon & Laudon (2002, p.9).

FIGURA 3.1 – Funções de um sistema de informação

Os sistemas de informação são um conjunto de normas e procedimentos que visam captar informações na organização, apresentando a cada nível (operacional, tático e estratégico), o que lhe cabe e tendo como por um dos objetivos subsidiar a tomada de decisões (OLIVEIRA, 1993). Segundo Prates (1994), os sistemas de informação são formados pela combinação estruturada de vários elementos, a saber:

- a informação, os recursos humanos (pessoas que coletam, armazenam, recuperam, processam, disseminam e utilizam as informações);
- as tecnologias de informação (o *hardware* e o *software* usados no suporte aos sistemas de informação); e
- as práticas de trabalho (métodos utilizados pelas pessoas no desempenho de suas atividades), organizadas de tal modo que permitam o melhor atendimento dos objetivos da organização.

Um sistema de informação é multidimensional. Ele deve ser entendido pelos aspectos organizacionais, de gestão e tecnológicos (LAUDON & LAUDON, 2002).

O planejamento e o desenvolvimento do sistema de informação de uma organização, segundo Moresi (2000), devem resultar de uma análise que proporcione obter uma estimativa prévia sobre as seguintes dimensões:

- estratégica: em que se analisa o impacto do sistema em relação ao ambiente externo da organização, incluindo-se aspectos relacionados à competitividade, posicionamento no ambiente, visão do cliente, diferencial de mercado, entre outros;
- organizacional: onde são analisados aspectos relativos às necessidades de informação da organização para a sua boa estruturação, coordenação operacional e atendimento às demandas externas;
- econômica: em que se incluem os aspectos relacionados com a sua importância econômica com especial atenção à redução de custos, melhorias de produtividade e eficiência, ganhos financeiros e outros aspectos que possam ser mensuráveis; e
- capacitação da organização: refere-se ao cuidado quanto à capacitação em face do conjunto de tecnologias disponíveis, de forma que a organização esteja preparada para acompanhar as possibilidades de utilização e acomodar a evolução tecnológica.

O sistema de informação em sua estrutura interna contém diversos sistemas que têm funções diferentes em uma organização. Na Figura 3.2 é apresentado um *portfolio* dos quatro sistemas.

As aplicações¹¹ dos quatro sistemas estão relacionadas com as estratégias dos negócios na organização. A Figura 3.2 procura mostrar que os quatro sistemas estão inter-relacionados. A proposta do *portfolio* é analisar as aplicações atuais e as potenciais nos quatro sistemas e verificar o impacto dessas aplicações nos negócios atuais e futuros (WARD & PEPPARD, 2002).

¹¹ Aplicações referem-se ao uso da tecnologia de informação nas atividades e/ou processos de negócios (Ward & Peppard, 2002)



FIGURA 3.2 – Integração do sistema de informação com a estratégia de negócio

Fonte: Adaptado de Ward & Peppard, 2002.

Segundo Laudon & Laudon (2002), a tecnologia utilizada no uso do sistema de informação é composto por *hardware*, *software*, banco de dados, comunicação e redes que realizam a integração dos computadores. Todos esses componentes podem ser chamados de infra-estrutura para a tecnologia de informação.

Para os mesmos autores, a TI é definida como o conjunto de tecnologias citadas no parágrafo anterior com o objetivo de fornecer um *portfolio* de recursos na transformação de dados em informações relevantes para a organização.

Nas organizações, o uso da TI promove a coordenação interdepartamental, cristalizada pela relação entre cliente e fornecedor interno, na qual as diversas etapas do processo produtivo precisam estar integradas de modo a estimular a cooperação interna, aumentar a capacidade de resposta a imprevistos e dar flexibilidade às operações da empresa (VALLE, 1996).

Para esse mesmo autor, para que a incorporação dos desenvolvimentos da TI seja bem-sucedida, é necessário que haja reestruturação (ou redefinição) das formas de organização dos sistemas produtivos e do modo de gerenciá-los. Se antes a competição estava baseada em custos e quantidades, com empresas extremamente hierarquizadas, com sistemas administrativos rígidos, vários níveis de supervisão e comunicação horizontal incipiente: hoje o formato organizacional precisa ser modificado para tornar-se compatível com o novo ambiente competitivo, no qual as empresas líderes em seus segmentos de mercado são aquelas que têm por estratégia concorrencial a qualidade e a diferenciação dos seus produtos e serviços. Além disso,

existem outras oportunidades a serem exploradas pelas empresas a partir da adoção desse tipo de tecnologia. São elas:

- redução dos níveis hierárquicos devido ao aumento das informações disponíveis ao nível operacional, não havendo necessidade de a decisão ser tomada em nível gerencial;
- crescente delegação de responsabilidades valorizando o uso de equipes de trabalho *ad hoc* e também formação de equipes multifuncionais com real poder de decisão na organização; e
- descentralização dos processos decisórios e de controle diante do crescente uso da comunicação de colaboradores a distância;

Para Moresi (2000), a convergência das tecnologias de informação e dos sistemas de informação tem afetado o processo de trabalho nas organizações. Antes de implantar qualquer sistema é importante analisar a situação da empresa para que se possam determinar satisfatoriamente os requisitos do projeto, ou seja, compatibilizar o sistema com o modelo de negócios da empresa. Existem diversos desafios nos quais os sistemas de informação e conseqüentemente a TI podem ser explorados:

- problemas bem definidos, bem estruturados e com objetivos claros;
- problemas bem estruturados, com objetivos claros e incertezas quanto às necessidades dos usuários;
- problemas que não estão estruturados e cujos objetivos não são claros;
- grande interação do usuário com o sistema e/ou a aceitação do usuário é importante; e
- questões complexas, em que existem diversas variáveis inter-relacionadas, e necessidade de uma abordagem contingencial para o desenvolvimento de sistemas de informação.

Houve uma evolução muito grande nas últimas décadas na área de informática e nas tecnologias da informação. Todavia, essa evolução permitiu um grande avanço apenas no nível operacional da empresa. Porém, especificamente para os seus níveis táticos e estratégicos, somente agora nos últimos anos é que a TI começou a trazer benefícios, disponibilizando ferramentas computacionais adequadas de apoio ao gerenciamento dos negócios. Surgiram, assim, os sistemas de apoio à decisão, bem

como ferramentas que apoiariam essa categoria de sistema de informação, a exemplo de *data warehouses*, *data mining* e diversas ferramentas e técnicas para análise de dados (FREITAS JÚNIOR et al., 2000).

3.1.2 Sistemas de apoio à decisão

Dado o atual ambiente industrial competitivo no qual diversas aplicações de TI são voltadas para diferentes níveis organizacionais, um sistema de informação único e centralizado não é o mais adequado para as organizações (LAUDON & LAUDON, 2002; WARD & PEPPARD, 2002). Ele não pode prover toda a informação necessária para todos os membros da organização devido aos diferentes interesses, especialidades e níveis organizacionais.

Em face desse contexto, Laudon & Laudon (2002) apresentam seis tipos de sistemas de informação com suas características de processamento, de acordo com o Quadro 3.1.

QUADRO 3.1 – Características do processamento dos sistemas de informação

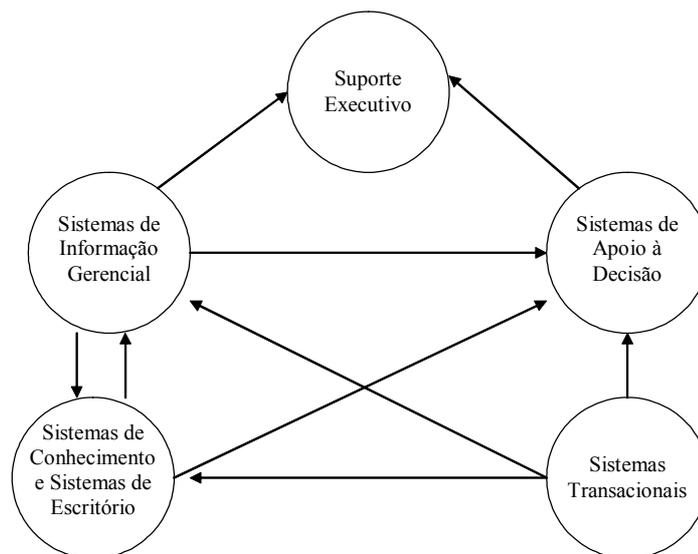
Tipo de Sistema	Entrada	Processamento	Saída	Usuários
Suporte Executivo (<i>Executive Support Systems</i>)	Dados agregados internos e externos	Gráficos, simulações, interação	Projeções, respostas para questões lógicas	Gerentes seniores
Sistemas de Informação Gerencial (<i>Management Information Systems</i>)	Dados transacionais sumarizados; alta complexidade dos dados, uso de modelos simples	Relatórios rotineiros, modelos simples, análise de baixa complexidade	Sumários e relatórios excepcionais	Média gerência
Sistemas de Apoio à Decisão (<i>Decision-Support Systems</i>)	Dados de baixa complexidade para análise utilizando-se ferramentas e modelos analíticos	Interativo, simulação e análise	Relatórios especiais, decisão sobre a análise e respostas para questões lógicas.	Gerentes de <i>staff</i> , especialistas
Sistemas de Conhecimento (<i>Knowledge Work Systems</i>)	Especificações de <i>design</i> , bases de conhecimento	Simulações, modelagem.	Modelos, gráficos	Especialistas, <i>staff</i> técnico
Sistemas de Escritório (<i>Office Systems</i>)	Documentos, tabelas de horário	Gestão dos documentos, programações, comunicações	Documentos, tabelas de programação e correio	Trabalhadores administrativos
Sistemas Transacionais (<i>Transaction Processing Systems</i>)	Transações, eventos	Classificação, listagem, fusão, atualização	Relatórios detalhados, listas e sumários	Pessoas ligadas a operações e supervisores

Fonte: Laudon & Laudon (2002, p.41).

Pela análise do Quadro 3.1, verifica-se que o Suporte Executivo e Sistema de Informação Gerencial (com o apoio dos Sistemas de Escritório) atendem respectivamente aos níveis estratégico e tático da empresa.

Já os outros sistemas atuam na geração de conhecimento para a empresa. O Sistema Transacional movimenta os dados por diferentes áreas da empresa e opera como suporte aos Sistemas de Conhecimento e Apoio à Decisão, que atuam na análise dos dados gerando informações relevantes para a empresa.

Os seis tipos de sistemas de informação estão inter-relacionados entre si como ilustra a Figura 3.3. Por exemplo, a produção de informação e/ou dados dos sistemas transacionais atendem a demanda por dados e/ou informação do sistema de informação gerencial, sistemas de conhecimento e sistemas de apoio à decisão.



Fonte: Laudon & Laudon (2002, p.46).

FIGURA 3.3 – Inter-relacionamentos entre os sistemas de informação.

A Figura 3.3 mostra ainda que é possível fazer com que as informações fluam facilmente entre diferentes áreas da organização atendendo a diversos *stakeholders*. Porém, a integração de diferentes sistemas pode ser dispendiosa, complexa e requerer enorme gasto de tempo.

Muitas organizações estão implementando *enterprise systems*, também conhecido como *Enterprise Resource Planning* (ERP). No ERP é possível integrar as diferentes fontes de dados e informações que estão fragmentadas em diferentes sistemas na empresa facilitando o acesso de todos.

Um sistema ERP pode também coletar dados de diferentes processos de negócios e armazená-los em um único e simples repositório de dados, para serem usados em outras áreas da empresa. Esse repositório pode atuar como um *data warehouse* permitindo o uso analítico dos dados.

Dentre os seis sistemas apresentados, destacam-se neste trabalho os sistemas de apoio à decisão (SAD) que são baseados em computador e que auxiliam o processo decisório utilizando-se modelos e ferramentas analíticas para resolver problemas não-estruturados (problemas não-rotineiros) em uma organização (FREITAS JÚNIOR et al., 2000; LAUDON & LAUDON, 2002).

O SAD, segundo Ward & Peppard (2002), conjuntamente com o Sistema de Suporte Executivo e Sistema de Informação Gerencial atua nas aplicações críticas para o sucesso dos negócios futuros. Eles criam um suporte para que as organizações conduzam os seus negócios e também auxiliam na manutenção da vantagem competitiva nas empresas.

O SAD ajuda os gerentes a tomarem decisões sobre problemas não-rotineiros, ou seja, problemas que mudam rapidamente e não são facilmente especificados. Ele utiliza informações e/ou dados dos sistemas transacionais e sistemas de informação gerencial, como também busca informações provenientes do ambiente externo (LAUDON & LAUDON, 2002).

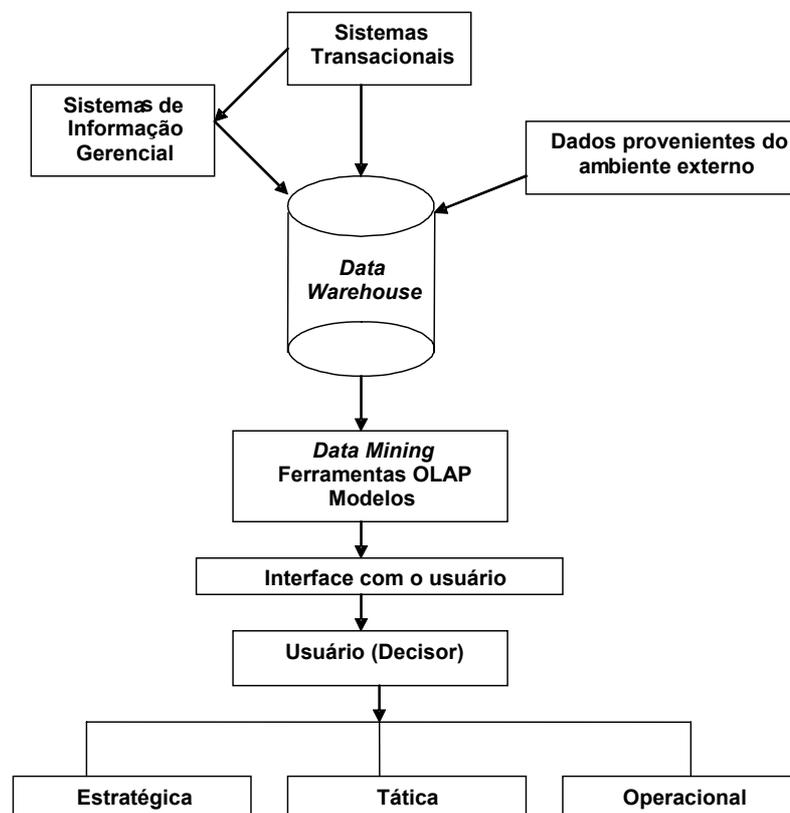
No processo decisório, o volume de informações e dados colocados à disposição do decisor deve estar na medida certa. Se ele for excessivo, os dados e informações pertinentes à solução do problema serão mascarados por aqueles considerados espúrios. Para resolver esse problema, é necessário escalonar a informação em uma hierarquia capaz de diferenciar as necessidades nas diversas situações, o que reforça a importância de reconhecer que a informação possui valor (MORESI, 2000).

De acordo com Bispo (1998), o SAD contribui para o processo decisório desde que contenha as seguintes características:

- ser voltado para problemas menos estruturados e menos especificados com os quais os gerentes se deparam;
- combinar o uso de modelos ou técnicas analíticas a funções tradicionais de acesso e recuperação de informações;

- concentrar-se especificamente nos recursos que facilitem o uso para pessoal não especializado em computação; e
- enfatizar a flexibilidade e a adaptabilidade de acomodar mudanças no ambiente e na abordagem do processo decisório.

O SAD é composto por dois elementos que são o banco de dados e um sistema de modelos (conjunto de métodos, técnicas e ferramentas utilizadas na transformação dos dados em informações relevantes, como o *data mining*) conforme mostrado na Figura 3.4. Ele interage principalmente com os usuários responsáveis pela tomada de decisão na empresa (FREIRE, 2000; LAUDON & LAUDON, 2002).



Fonte: Adaptado de Laudon & Laudon (2002) e Freire (2000).

FIGURA 3.4 – Arquitetura de um sistema de apoio à decisão.

Para Freire (2000), o banco de dados é um conjunto de recursos que manipula e controla tudo o que se refere a dados dentro de um SAD. Ele deve ser independente para ser capaz de lidar com informações estruturadas, semi estruturadas e não-estruturadas.

O sistema de modelos contém métodos, ferramentas e técnicas computacionais usadas para a análise de dados. As análises podem ser simulações, cálculos, resolução de problemas matemáticos, entre outros (FREIRE, 2000).

Os usuários do SAD podem ser considerados os tomadores de decisão na organização. Eles são responsáveis pelo estudo do ambiente em que atuam e pela determinação de possíveis linhas de ação. Para Moresi (2000), a tomada de decisão é um processo complexo de investigação, escolha e análise das informações em cujo processo o SAD pode auxiliar bastante.

Dados os avanços da TI na década de 90 do século passado, surgiram novas ferramentas de SAD tanto para o sistema de banco de dados como para o de modelos, como é esquematizado na Figura 3.4.

Dentre as novas abordagens do SAD que surgiram na década de 90 do século passado, destacam-se o *data warehouse*, o OLAP e o *data mining*, que são mostrados nos próximos itens.

3.1.3 Data warehouse

Data warehouse é uma abordagem baseada nas necessidades de informação da organização, que busca atender os critérios de investimento e auxiliar na obtenção de vantagens competitivas (JONES, 1998). Um *data warehouse* permite processamento analítico e informacional, e as informações contidas nele devem atender às necessidades da gerência no processo de tomada de decisão.

Ele teve a sua evolução histórica, segundo Haley & Watson (1998), no início da década de 70 do século passado quando houve o desenvolvimento da primeira arquitetura de informação que tinha um estoque de dados separado e customizado que servia como suporte à tomada de decisão.

Em meados da década de 80 do século passado, várias corporações de setores como telecomunicações, bancos, entre outras, iniciaram a construção de *data warehouses*, antes do termo se tornar conhecido no final dessa década por meio de Bill Inmon. Com a diversificação do mercado nessa época em micro-segmentos, houve a

necessidade de introdução de tecnologias especializadas que foram fornecidas pela IBM e Teradata (KELLY¹², 1997 *apud* HALEY & WATSON, 1998).

No início dos anos 90 do século passado, as aplicações de *data warehouse*, dados os avanços da TI, se tornaram disponíveis fazendo com que elas se tornassem um desenvolvimento interessante para o mundo da informática (HALEY & WATSON, 1998).

Os dados de um *data warehouse* são modelados a partir do conceito de multidimensionalidade e são representados por meio do cubo de dados, que é a melhor forma de compará-los, pois permite que diversas variáveis sejam trabalhadas conjuntamente. Por exemplo, um cubo de dados “vendas” a qual cada aresta do cubo de dados representa uma dimensão (tempo, região, produto). A representação do cubo é feita geralmente pela tabela-fato que contém as dimensões e para cada uma delas existe a tabela-dimensão com os seus atributos, e elas se interagem em uma banco de dados relacional (FELIPE, 2000).

Para que o *data warehouse* seja bem aplicado para o atendimento das necessidades de gestão da empresa, vários temas são importantes para o seu uso eficiente:

- escalabilidade: capacidade para lidar com a variação do tamanho de um banco de dados, da complexidade para lidar com um grande fluxo de entrada e saída de dados e da facilidade para interagir com funções provenientes do analisador de informações, como o OLAP (TANLER, 1997);
- segurança: Deve-se garantir um controle dos usuários e esquemas de autorização de acessos para algumas partes do conteúdo do *data warehouse* (TANLER, 1997);
- gestão de metadados: Metadados são dados explicando dados. Fazendo uma analogia, uma lista telefônica seria um metadados onde se obtêm dados para fazer chamadas telefônicas. Para gerenciar os metadados, é preciso haver integração de três tipos de metadados. O primeiro é o metadados da banco de dados operacional, que define a sua estrutura; o segundo é o *warehouse-level*, que define a maneira de

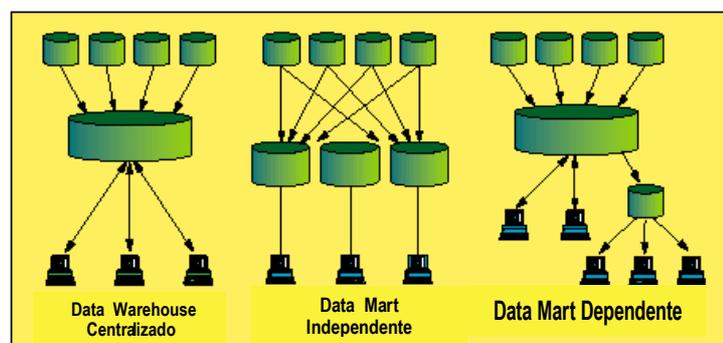
¹² Kelly, S. Data Warehousing in Action. Wiley, New York, 1997.

converter dados em informações e o terceiro é o *business-level*, que mapeia os metadados do *warehouse* para o uso nos negócios (JACOB & SEN, 1998);

- monitoramento e desempenho: É necessária uma medição de desempenho global para todos os escopos do *data warehouse* e das aplicações interligadas com o analisador de informações para poder garantir um alto desempenho em todas as etapas de carregamento, manutenção, auditoria e análise (JACOB & SEN, 1998); e
- *data archive* ou *data mart*: Uma questão importante em um *data warehouse* é como centralizar os dados (*data archive*) ou descentralizar (*data mart*). O primeiro tem a vantagem de manter a integridade e a eficiência para a gestão do banco de dados. Já o segundo tem a vantagem de atender as necessidades de segmentos específicos ou grupos de usuários da organização (TANLER, 1997).

Destacando-se o *data mart* como um aspecto importante do *data warehouse*, a empresa pode criar *data marts* para cada segmento e facilitar a interação dos usuários. Ele está sendo muito discutido pelas empresas por dificultar o trabalho com grandes arquiteturas, o atendimento das necessidades específicas de todos os segmentos e também por ser um processo longo e custoso de implementação (BASHEIN & MARKUS, 2000).

O *data mart* são dados “limpos” já em um formato analítico voltado para uma área ou projeto da organização (SAS, 2000). O *data warehouse* e *data mart* podem ser estruturados, segundo Gardner (1998), em três formas mostradas na Figura 3.5.



Fonte: Gardner (1998, p.54).

FIGURA 3.5 – Tipos de *data warehouse* e *data marts* em uma organização.

De acordo com a Figura 3.5, o *data warehouse* centralizado representa o agrupamento de dados e/ou informações de todos os bancos de dados transacionais e de informação gerencial. Já no *data mart* independente, os *data marts* obtêm os seus dados diretamente dos bancos de dados transacionais e de informação gerencial, o que não ocorre com o *data mart* dependente, que obtêm os seus dados e/ou informações a partir do *data warehouse*.

Bashein & Markus (2000), baseados em estudos de caso próprios, constataram quatro padrões distintos para o uso do *data warehouse*:

- relatório gerencial e suporte à decisão: O *data warehouse* tem o intuito de facilitar a produção e distribuição de relatórios gerenciais rotineiros para a organização, reduzindo custos, erros e tempo de obtenção dos relatórios;
- sistemas e integração organizacional: Servem como substituto ou transição para integração de dados, sistemas transacionais ou processos de negócios;
- novos dados de produtos: São realizados na área de vendas e utilizados para a divulgação de novos produtos ou serviços através da obtenção de informações contidas no *data warehouse*; e
- *data mining*: Utiliza uma variedade de técnicas indutivas, como inteligência artificial, para encontrar padrões de informações nos grandes bancos de dados. Geralmente as aplicações de *data mining* focam o ambiente interno da empresa.

Convém destacar a diferença entre banco de dados operacionais ou transacionais e o *data warehouse*. Segundo Quispe (2003), os bancos de dados operacionais armazenam as informações necessárias para as operações momentâneas do domínio. São utilizados por todos os usuários para registrar e executar operações pré-definidas, por isso seus dados podem sofrer constantes mudanças conforme as necessidades atuais do domínio. Para não ocorrer redundância, os dados históricos não ficam armazenadas por muito tempo. Este tipo de bancos de dados não exige grande capacidade de armazenamento. O Quadro 3.2 apresenta as principais diferenças entre os bancos de dados operacionais e o *data warehouse*.

Por intermédio do Quadro 3.2, é possível verificar que o *data warehouse* é utilizado para o armazenamento de dados históricos com uma estrutura propícia para o uso de ferramentas de suporte à tomada de decisão, pelas suas características de permitir consultas estruturadas a banco de dados, de conter dados formatados para análise (em uma estrutura de cubo de dados), dados detalhados e resumidos e com integridade periódica dos dados.

QUADRO 3.2 – Diferenças entre banco de dados operacionais e *data warehouse*

Características	Bancos de dados	Data Warehouse
Objetivo	Operações eventuais do domínio	Analisar o domínio
Uso	Operativo	Informativo
Tipo de processamento	<i>Online Transactional Processing (OLTP)</i>	OLAP
Unidade de trabalho	Inclusão, alteração, exclusão.	Consulta
Número de usuários	Milhares	Centenas
Tipo de usuário	Operadores	Gerenciadores
Interação do usuário	Somente predefinida	Predefinida e <i>ad-hoc</i>
Condições dos dados	Dados operacionais	Dados Analíticos
Volume	Megabytes – gigabytes	Gigabytes – terabytes
Histórico	Mensal, Bimestral.	Anual
Granularidade	Detalhados	Detalhados e resumidos
Redundância	Não ocorre	Ocorre
Estrutura	Estática	Variável
Manutenção desejada	Às vezes	Freqüentemente
Acesso a registros	Dezenas	Milhares
Atualização	Contínua (tempo real)	Periódica (em lote)
Integridade	Eventualmente	A cada atualização
Número de índices	Poucos/simples	Muitos/complexos
Intenção dos índices	Localizar um registro	Aperfeiçoar consultas

Fonte: Adaptado de Kinball (1996); Singh (2001); e Quispe (2003).

Como a TI vem crescendo em importância para as organizações, é desejável a integração de diversas ferramentas dela, como o *data warehouse* integrado com o *data mining* para utilização de informações analíticas (FEELDERS et al., 2000).

Essa integração é importante no processo de *data mining*, principalmente na fase de pré-processamento dos dados, que corresponde às atividades que visam gerar uma representação conveniente para os algoritmos de garimpagem a partir do banco de dados. Nem sempre a forma utilizada para armazenar dados é favorável para a execução do processo de *data mining*. Por isso, as organizações que dispõem de *data warehouses* têm maior facilidade de aplicar o *data mining*, pois as informações estão armazenadas de maneira mais favorável e em alguns casos sintetizadas (RODRIGUES FILHO & SHIMIZU, 2002). Portanto, o uso efetivo do *data mining* acontece quando é combinado com o *data warehouse*, quando os dados já estão limpos e consistentes (BISPO, 1998).

3.1.4 OLAP

O OLAP pode ser definido como uma aplicação de diversas funções de análise dos dados para o tratamento e geração de informações úteis em um *data warehouse* (TANLER, 1997). O uso do OLAP no *data warehouse* se deve à necessidade de dados históricos sumarizados e consolidados para facilitar o uso analítico (FELIPE, ,2000).

Tanler (1997) destaca as seguintes funções analíticas no OLAP:

- query and reporting: O OLAP permite que os usuários formulem questões sem necessitar da interação com a linguagem de programação, por exemplo, *Structured Query Language* (SQL);
- análise multidimensional: Possui uma robusta estrutura de dados computacionais e capacidade de navegação de dados. A análise multidimensional permite que os usuários penetrem no *data warehouse* em uma dimensão para iniciar a análise e, então, dado a algum interesse, pode-se transitar rapidamente para outras dimensões a fim de aumentar a capacidade de análise da informação;
- análise estatística: Tem o intuito de reduzir uma grande quantidade de dados em simples relacionamentos, que são frequentemente expressos em fórmulas matemáticas; e

- data mining: A mais complexa das funções do OLAP busca usar sofisticados padrões de reconhecimento e aprendizado de algoritmos para identificar relacionamentos entre os dados.

Segundo Felipe (2000), o OLAP tem como meta fornecer ao usuário, que geralmente não é especialista em linguagens de programação, uma interface intuitiva e poderosa para tarefa comuns de análise de negócio.

Esta interface intuitiva e poderosa do usuário advém da manipulação do OLAP no cubo de dados com diversas operações, sendo as mais se destacadas são:

- *roll-up*: que aumenta o nível de agregação de dados;
- *drill-down*: que fornece uma visão detalhada dos dados, oposto ao *roll-up*;
- *pivoting*: rotação que se faz entre os eixos do cubo, a fim de mudar o ângulo de visão de dados; e
- *slice and dice*: que consiste na possibilidade de mudar a ordem das dimensões, alterando, assim, a orientação pela qual os dados são visualizados.

As abordagens de *data warehouse* e OLAP são complementares: a primeira armazena e mantém os dados, e a segunda converte-os em informações estratégicas. Há diversas formas de se manipular dos dados em um *data warehouse*. Felipe (2000) destaca:

- servidor ROLAP (Relational OLAP): ferramenta responsável pela manipulação de dados do *data warehouse* implementando em banco de dados relacionais. A vantagem está na eficiência para o armazenamento de grandes conjuntos de dados, devido ao fato de se poder armazenar dados esparsos, de forma mais compacta, em tabelas;
- servidor MOLAP (Multidimensional OLAP): ferramenta responsável pela manipulação de dados armazenados em banco de dados multidimensionais. Esses servidores executam as operações OLAP diretamente sobre as matrizes; e
- servidor HOLAP (Hybrid OLAP): adoção de uma forma de armazenamento em dois níveis, podendo atuar uma parte como

ROLAP e outra como MOLAP a fim de manipular um conjunto de dados densos e esparsos.

Existem outros tipos de servidores, com menos destaque, como, o OLAP pode ser explorado nos terminais por clientes na própria empresa deles ou também ser utilizado via *web* (FORSMAN, 2004).

Existem ferramentas em que se aplica o *data mining* a partir do OLAP. Destaca-se, segundo FELIPE (2000), o OLAM (*On-Line Analytical Mining*) que propicia a aplicação do OLAP em um ambiente com múltiplas variáveis e cujas respostas têm que ser dadas rapidamente. Ele executa múltiplas tarefas de *data mining* como análise de associações, classificação, agrupamento de banco de dados, entre outros.

3.2 Data Mining

Segundo Lee & Siau (2001), em meados da década passada havia somente 50 pesquisadores interessados na conferência de *data mining* e descoberta do conhecimento. Em 2001, apenas um grupo¹³ de mensagens eletrônicas (*newsletter*) voltado para o *data mining* e para a descoberta do conhecimento continha cerca de 4000 leitores.

Diversas empresas usam os computadores para capturar detalhes das transações de negócios como informações de cartões de créditos, bancos, vendas, manufatura, telecomunicações e uma miríade de outras transações. O *data mining*, com suas técnicas e ferramentas, pode ser aplicado na mineração de textos e dados que são úteis para os tomadores de decisão, usando *queries* para a extração de informações relevantes e criando processos com uma série de novos *queries* com vistas a gerar algoritmos que podem ser preditivos, descritivos ou intervencionistas.

Dada a importância do *data mining*, a apresentação deste item será feita a partir da descrição de diversos aspectos de acordo com a qual essa abordagem pode ser utilizada. Esse detalhamento será útil no entendimento de como o *data mining* pode atuar no método para o estabelecimento de relacionamentos entre as medidas de desempenho.

¹³ Esse grupo de mensagens eletrônicas foi feito por Gregory Piatetsky-Shapiro e está no endereço eletrônico <http://www.kdnuggets.com>. Em 20 de Agosto de 2004, esse grupo continha mais de 11 mil assinantes.

3.2.1 Visão geral

O *data mining* explora bancos de dados históricos para obter informações relevantes sobre eles, com o propósito de auxiliar a tomada de decisão. Ele busca obter padrões e relações existentes nesses dados para a geração de informações úteis (CABENA et al., 1997; TANLER, 1997; LINDEN, 2001; LAUDON & LAUDON, 2002).

A evolução histórica do *data mining*, segundo CABENA et al. (1997), iniciou-se na década de 60 do século passado com a criação, por Frank Rosenblatt, do “Perceptron”, que foi um precursor das modernas redes neurais (redes inspiradas na estrutura do cérebro, com o objetivo de apresentar características similares ao comportamento humano) e que solucionava problemas simples. Em 1969, Minsky e Papert desenvolveram um modelo de rede neural mais complexo que o “Perceptron”, denominado “engenharia do conhecimento”, que buscava soluções por meio de observações alimentadas pela própria máquina.

No início dos anos 70 do século passado, com base no modelo de “engenharia do conhecimento”, começaram a ser desenvolvidos alguns sistemas especialistas, sendo os mais conhecidos o MYCIN (diagnósticos médicos) e XCON VAX (configuração de computadores) (CABENA et al., 1997). Os sistemas especialistas são programas inteligentes de computador que usam conhecimentos e procedimentos de inferências, para resolver problemas que são bastante complexos e/ou complicados, de forma a requererem para sua solução muita perícia humana (METAXIOTIS & PSARRAS, 2003).

Os sistemas especialistas são aplicados em diversas áreas, sendo os de maior destaque: *marketing*, área bancária e financeira, previsões e tendências de negócios (METAXIOTIS & PSARRAS, 2003). Os sistemas especialistas estão englobados dentro das áreas de *data mining*. Destaca-se o seu uso em modelos preditivos, pois requer o uso de programas que inferem do conjunto de dados predições de valores desejados pelos decisores.

Em meados da década de 80, houve um aumento considerável de pesquisas e pessoas interessadas no meio acadêmico em *machine learning*, gerando o desenvolvimento de novas técnicas para soluções de problemas altamente complexos. Aliado a esse avanço do *machine learning*, houve também um grande aumento da

disponibilidade de grandes bancos de dados comerciais em que modelos de *machine learning* passaram a ser aplicados com maior facilidade, destacando-se suas aplicações na área de *marketing* (CABENA et al., 1997).

No fim dos anos 80, surge o novo termo que buscava englobar conceitualmente todos os outros conhecidos na área (inteligência artificial, *machine learning*, entre outros) denominados *Knowledge Discovery Databases* (KDD), que tem o objetivo de encontrar padrões e similaridades em base de dados. Dentro deste contexto, uma das etapas do processo do KDD foi chamada de *data mining* em um congresso realizado em Montreal em 1995 (CABENA et al., 1997).

Para o mesmo autor, em fins da década de 90 do século passado, com o crescimento da popularidade e da utilização cada vez maior do termo *data mining* por parte dos vendedores, ele passou a ser utilizado como um substituto do termo KDD como um processo de extração de conhecimento de uma base de dados.

O *data mining* está sedimentado em quatro pilares: Inteligência Artificial, Estatística, KDD e banco de dados (FAYYAD et al., 1996; COSTA et al., 2000). Embora haja visões diferentes sobre a relação entre KDD e *data mining*, nesta dissertação é considerado o KDD como um dos pilares importantes para o uso do *data mining*.

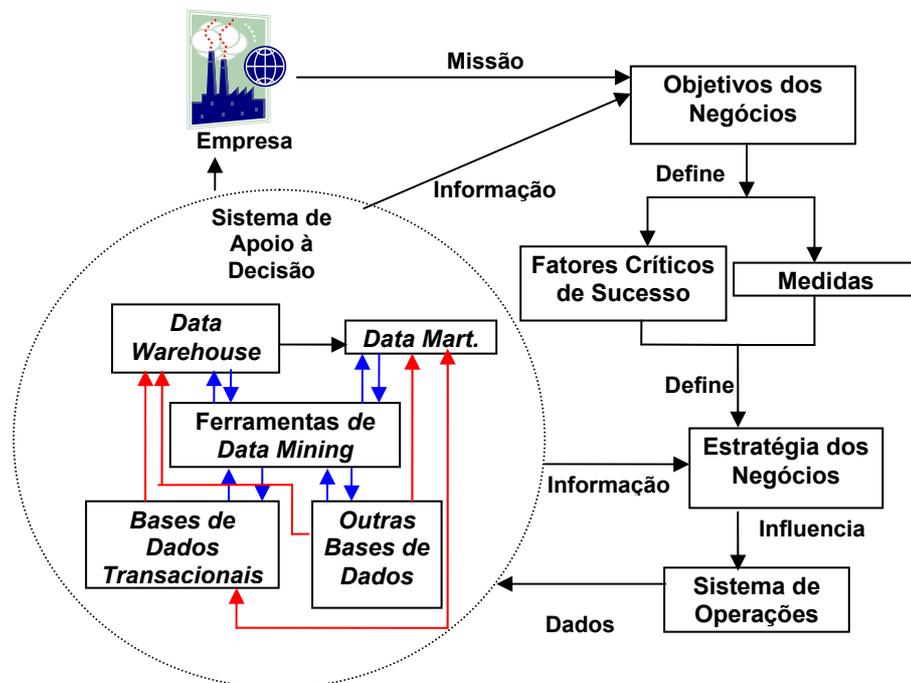
A contribuição da Inteligência Artificial para o *data mining* se reflete na busca de padrões compreensíveis que podem ser interpretados como um conhecimento útil na tomada de decisão (FAYYAD et al., 1996). Ela envolve muitos termos complexos e de utilização ampla, sendo necessário simplificar os conceitos empregados sobre o assunto para melhor compreensão de como ela é utilizada pelo *data mining* (COSTA et al., 2000).

Já a Estatística contribui para o *data mining* provendo uma linguagem que quantifica a incerteza dos resultados quando há a tentativa de inferir (entender) padrões gerais para uma particular amostra de toda uma população (FAYYAD et al., 1996).

O KDD é importante para o *data mining*, pois suporta a utilização e integração de vários métodos de descoberta de conhecimento, a maioria dos quais é baseada em pesquisa e pode ser usada em modo interativo e/ou automatizado (COSTA et al., 2000). Ele apóia o *data mining* na análise de dados e descoberta de algoritmos

com limitações de eficiência computacional produzindo uma série de padrões ou modelos a partir de um conjunto de dados (FAYYAD et al., 1996).

Por fim, o banco de dados permite um eficiente ganho no acesso aos dados por meio de um conjunto de técnicas que permitem o agrupamento e a ordenação das operações que acessam os dados, e otimiza *queries* que constituem a base dos algoritmos que exploram grandes bancos de dados (FAYYAD et al., 1996). Uma das representações dos bancos de dados é o *data warehouse*, que, conjuntamente com outros bancos de dados, dentro do contexto organizacional pode auxiliar o *data mining* da seguinte forma, como ilustra a Figura 3.6.



Fonte: Elaborado pelo autor.

FIGURA 3.6 – A relação entre *data warehouse* e *data mining*.

De acordo com a Figura 3.6, a empresa tem uma missão que estrutura os objetivos de seus negócios da empresa, que define quais fatores críticos (custo, qualidade, entrega, velocidade e flexibilidade) e quais ações a serem tomadas são relevantes para ela. Com isso, a empresa tem definido o seu modelo de negócios. A sua estratégia de negócios é definida a partir desse modelo de negócios e é aplicada a todos os processos de negócios relevantes para a empresa que compõem o seu sistema de operações. Esses processos geram dados para os sistemas de informação, cujo foco, nesta dissertação, são os SAD, que retornam com informações relevantes para que a

estratégia de negócios possa ser testada e redefinida caso não atenda o modelo de negócios da empresa.

Dentro do SAD, como mostrado na Figura 3.6, existem diversos bancos de dados que armazenam os dados dos processos de negócio. Todos esses bancos, entre eles o *data warehouse* e o *data mart*, interagem (um dos tipos de interação dos banco de dados) com as ferramentas de *data mining* desenvolvidas a partir de técnicas para atender às necessidades dos usuários. O SAD gera informações relevantes para a estratégia de negócios da empresa e até mesmo para repensar os objetivos dos negócios da empresa como componente do modelo de negócios.

O *data warehouse*, como componente do sistema de apoio à decisão da empresa, é o ponto de partida do processo de conversão dos dados em informação útil. Dentro do SAD, o *data warehouse* é o elemento fornecedor de dados e/ou informações para o sistema de modelos, cujo foco desta dissertação é o *data mining*.

3.2.2 Projeto de data mining

Um projeto de *data mining* pode ser realizado a partir de três possíveis abordagens, segundo Thuraisingham (1999):

1. *top-down*: São sugeridas hipóteses, que se não validadas são submetidas a uma revisão e é feito o lançamento de uma nova hipótese;
2. *bottom-up*: Os dados são analisados e extraídos padrões. Duas vertentes surgem desse método: é o *bottom-up* supervisionado¹⁴, em que os tomadores de decisão têm uma idéia do que estão procurando e o *bottom-up* não-supervisionado¹⁵ em que os tomadores de decisão não têm idéia do que estão procurando;e
3. método híbrido: O que combina os dois primeiros métodos, cuja análise dos dados e extração de padrões podem gerar hipóteses para serem validadas em um projeto de *data mining*.

¹⁴ Exemplo de *bottom-up* supervisionado: Um analista de uma fábrica de componentes plásticos inicia a busca por padrões utilizando as variáveis produtos e clientes em uma banco de dados a partir da questão: Que tipos de peças o maior cliente compra em termos de quantidade?

¹⁵ Exemplo de *bottom-up* não-supervisionado: Um analista de uma fábrica de componentes plásticos inicia a busca por padrões utilizando as variáveis produtos e clientes em uma banco de dados de forma aleatória, sem ter questão definida.

Dois aspectos devem ser levados em conta no projeto de *data mining*: o primeiro deles é aquele que para ser denominado *data mining*, deve ter a busca centrada em métodos automatizados para extrair padrões ou modelos a partir de dados (FAYYAD et al., 1996).

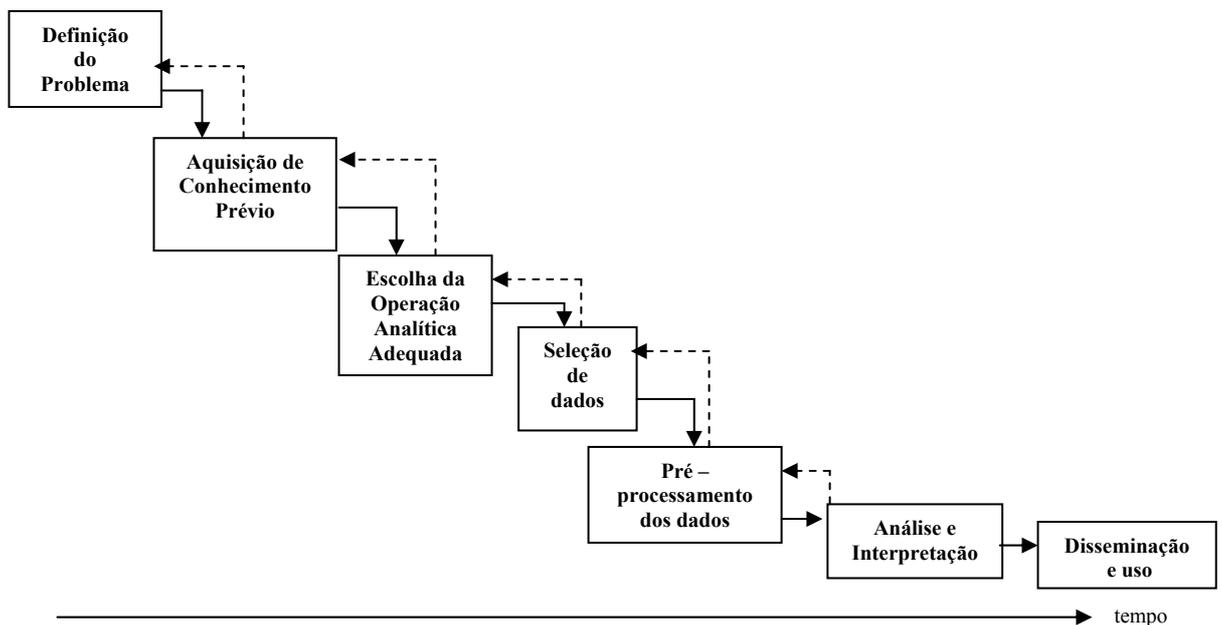
Segundo aspecto é aquele que, embora os métodos automatizados atuais sejam capazes de descobrir padrões "válidos e novos", ainda não apresenta uma solução eficaz para determinar padrões ótimos, ou seja, com riscos quase nulos de erros ou falhas. Por essa razão, o *data mining* ainda requer uma interação muito forte com analistas humanos, que são, em última instância, os principais responsáveis pela determinação do valor dos padrões encontrados. Além disso, a condução da exploração de dados é também tarefa fundamentalmente confiada a analistas humanos, um aspecto que não pode ser desprezado em nenhum projeto que queira ser bem sucedido (NAVEGA, 2002).

Há três razões para desenvolver um projeto de *data mining* (RODRIGUES FILHO & SHIMIZU, 2002):

- a) visualização dos dados: As organizações precisam dar significado a uma quantidade cada vez maior de informações armazenadas em seus banco de dados. O objetivo é qualificar e organizar os dados a serem trabalhados e encontrar novas formas de visualizá-los de modo transparente para os usuários;
- b) descoberta de novos conhecimentos: O objetivo é explicitar relacionamentos ocultos, padrões e correlações entre os diferentes dados existentes nos bancos de dados nas organizações ; e
- c) maior precisão dos dados: É necessário obter dados cada vez mais consistentes para processamento e análises futuros, evitando-se que a organização trabalhe com dados errados, incompletos ou contraditórios.

Nesta dissertação, as três razões para o projeto de *data mining* se encaixam no estabelecimento de relacionamentos entre as medidas de desempenho. Porém, a maior ênfase nesta dissertação será a razão descoberta de novos conhecimentos, pois ela trabalha com a análise de relacionamentos e padrões entre a entidade a ser minerada e as outras.

Um projeto de *data mining* apresenta diversas etapas de desenvolvimento, como ilustra a Figura 3.7. Convém destacar algo pouco mencionado em trabalhos acadêmicos e também não utilizado por Feelders (2000), que é a necessidade da escolha das operações analíticas de *data mining* como direcionador para análise dos resultados. Essas operações são baseadas na definição do problema e na aquisição de conhecimento prévio.



Fonte: Adaptado de Feelders et al. (2000).

FIGURA 3.7 - Estágios do processo de *data mining*

A definição do problema, segundo Feelders et al. (2000) consiste na identificação de questões para o uso do *data mining*. Um exemplo de uma típica questão a ser identificada seria: “Que grupo de aplicadores tenderá a ter problemas em recuperar créditos financeiros?”. É também importante determinar como os resultados são usados. Eles podem ser usados de forma descritiva (descrever um segmento ou grupo), preditiva (descoberta de relações que são usadas como instrumento preditivo) e intervencionista (resultados que podem gerar intervenção ao sistema a ser modelado).

O segundo estágio é a identificação do conhecimento prévio para uso no projeto. Nesse estágio, é importante verificar a razão da criação dos bancos de dados na organização. Geralmente, elas não são criadas para o propósito de análise, mas para suporte aos processos de negócios vitais. Isso representa uma armadilha para o projeto, pois não há um *design* estatístico formulado para ele. Por isso, é importante a

identificação dos possíveis vieses e da exclusão ou limitação dos efeitos que podem limitar a identificação dos padrões (FEELDERS et al., 2000).

O terceiro estágio é a escolha da operação analítica adequada para o projeto. As operações de *data mining* servem para quaisquer aplicações de negócios e podem ser mutuamente exclusivas, ou seja, várias operações podem ser realizadas conjuntamente para as operações de negócios (CABENA et al., 1997). Esse estágio mostrado na Figura 3.7 não é proposto por Feelders et al. (2000).

O quarto estágio é a seleção de dados, que busca decidir quais dados são relevantes para responder a questão da pesquisa no banco de dados. A seleção pode ser feita como “*open mind*”, ou seja, sem restringir os dados a partir da questão de pesquisa ou restringir a análise para uma hipótese pré-especificada (FEELDERS et al., 2000).

O pré-processamento de dados é o quinto estágio que possui dois aspectos: a derivação de novos atributos, que possibilita ao sistema, no qual se aplica a abordagem de *data mining*; e adicionar novos atributos derivados de outros existentes em uma tabela para o uso do *data mining*. Por exemplo, uma empresa possui uma relação entre valores monetários de vendas e áreas geográficas no mundo todo para 3 anos. Se quer adicionar o atributo segmentos de negócios da empresa e relacioná-las com as outras duas variáveis para poder entender quais são os segmentos de negócios mais importantes financeiramente para cada área geográfica.

O segundo aspecto é a agregação, em que a entidade a ser minerada se relaciona com diversas entidades do banco de dados. Um algoritmo de *data mining* agrega todas as entidades que se relacionam com a entidade a ser minerada em uma única tabela. Às vezes, as escolhas têm que ser cuidadosas, pois podem se perder atributos talvez essenciais para uma entidade vista como um objeto (FEELDERS et al., 2000). Por exemplo, a Secretária de Transportes da cidade de São Paulo quer detectar as principais causas de acidentes no trânsito. Para isso, ela tem que associar entidades a fim de identificar os principais problemas como veículos, pedestres, animais, objetos fixos, entre outros, até os decisores verificarem que o mapeamento dos acidentes é satisfatório.

O sexto estágio é a análise e interpretação. Nessa fase, os três tipos de *expertise* são requeridos: a) o especialista na aplicação de domínio é requerido para interpretar os resultados e indicar como devem ser explorados; b) o especialista em

dados é indicado para explicar padrões estranhos que podem ser dados poluídos ou outras causas como conversão de dados; e c) o especialista em *data mining* que procura interpretar tecnicamente os resultados, traduzindo para uma linguagem comum (FEELDERS et al., 2000).

O último estágio é a disseminação e uso dos resultados no projeto, que podem ser usados como um sistema especialista do qual é gerado um conjunto de conhecimentos para tomar decisão ou sistema de apoio à decisão (FEELDERS et al., 2000). Nesta dissertação, o uso do resultado é para o apoio à decisão.

Segundo Feelders (2000), para um projeto de *data mining* ser plenamente realizável, é necessário que as seguintes premissas sejam colocadas:

- um banco de dados de desempenho, se possível na forma de um *data warehouse* ou *data mart*, pois facilita a obtenção de dados para a análise;
- um conjunto de especialistas nas operações, ferramentas e técnicas de *data mining*, na interpretação dos dados e na interpretação dos resultados; e
- um membro da organização profundo conhecedor das atividades da empresa e do problema que ela enfrenta para auxiliar a equipe de *expertise*.

Várias operações analíticas podem ser usadas no projeto de *data mining*. Pode ser apenas uma operação ou até várias delas (CABENA et al., 1997; THURASINGHAM, 1999). O Quadro 3.3 apresenta as principais operações analíticas em um projeto de *data mining*.

As operações analíticas podem ser comparadas com as três razões para se desenvolver um projeto de *data mining*, apresentado por Rodrigues Filho & Shimizu (2002). Na primeira razão, que é a visualização de dados, podem ser aplicadas as operações analíticas “detecção de desvios” com as ferramentas e técnicas de visualização que facilitam a análise dos usuários e a operação analítica agrupamentos em bancos de dados que facilita a visualização dos dados a partir da formação de um *cluster* de dados que permite a detecção de uma característica para análise pelo decisor.

QUADRO 3.3 – Possíveis operações analíticas em um projeto de *data mining*

Operações	Definição	Técnicas Utilizadas
Modelos Preditivos	Predizer comportamentos ou resultados utilizando-se dados históricos por meio de ferramentas de encadeamento lógico como redes neurais, árvore de decisão, métodos de regressão, entre outros	Classificação e Predição de valor
Agrupamentos em bancos de dados	Agrupa os dados em termos das características desejadas pelos decisores	Agrupamento via redes neurais e separação demográfica
Análise de Associações	Buscar padrões de relacionamentos entre o fator desejado pelos decisores e outros fatores pertencentes ao banco de dados construindo regras e condições para se obter o relacionamento	Descoberta de associação, descoberta de padrões sequenciais e descoberta de seqüências similares no tempo.
Deteccção de desvios	Busca desvios ou falhas em populações de dados	Estatística e visualização gráfica

Fonte: Adaptado de Cabena et al., (1997).

Já a descoberta de novos conhecimentos pode ser aplicada às operações analíticas, modelos preditivos e análise de associações, pois estas permitem um ganho de conhecimento para os usuários tanto na aquisição de padrões e relacionamentos dos fatores a serem analisados como também na capacidade de predizer comportamentos futuros.

Já a terceira razão, que é a maior precisão dos dados, se encaixa na operação analítica “deteccção de desvios”, em que são aplicadas técnicas e ferramentas estatísticas com o intuito de eliminar ou limitar os vieses e dados estranhos no uso do *data mining* permitindo uma maior consistência nos dados.

Nesta dissertação serão apresentadas as técnicas e ferramentas que compõem as operações analíticas conforme se mostra no Quadro 3.4.

QUADRO 3.4 – Principais técnicas e ferramentas utilizadas nas operações de *data mining*

Operações	Principais Ferramentas
Modelagem Preditiva	<p><u>Classificação</u>: OLAP¹⁶ (Linden, 2001), árvore de decisão (CABENA et al., 1997; LINDEN, 2001), Lógica Fuzzy (WATANABE, 2003) ou rede neural (CABENA et al., 1997; LINDEN, 2001)</p> <p><u>Predição de valor</u>: regressão linear, não linear e múltipla (CABENA et al., 1997; LINDEN, 2001)</p>
Agrupamento de base de dados	<p><u>Via rede neural</u>: método Kohonen (CABENA et al., 1997)</p> <p><u>Separação demográfica</u>: <i>Condorset</i> (CABENA et al., 1997)</p>
Análise de Associações	Técnicas de associações: <i>Market Basket Analysis</i> (THURASINGHAM, 1999), <i>Inductive Logic Programming</i> (THURASINGHAM, 1999)
Detecção de desvios	<p><u>Estatística</u>: Sumarização estatística dos dados (CABENA et al., 1997; LINDEN, 2001)</p> <p><u>Visualização</u>: Uso de softwares estatísticos (CABENA et al., 1997), planilhas eletrônicas (CABENA et al., 1997) ou OLAP (LINDEN, 2001)</p>

Fonte: Adaptado de Cabena et al., 1997; Thuraisingham, 1999; Linden, 2001 e Watanabe, 2003.

Convém destacar as diferenças entre técnicas e ferramentas de *data mining* nesta dissertação. Técnicas seriam os procedimentos para a aplicação das operações analíticas de *data mining*. Já as ferramentas representam o uso das técnicas por meio de *softwares* comerciais e livres, algoritmos, entre outros. Serão detalhadas nos próximos itens cada técnica e ferramenta de *data mining* mostrada no Quadro 3.4.

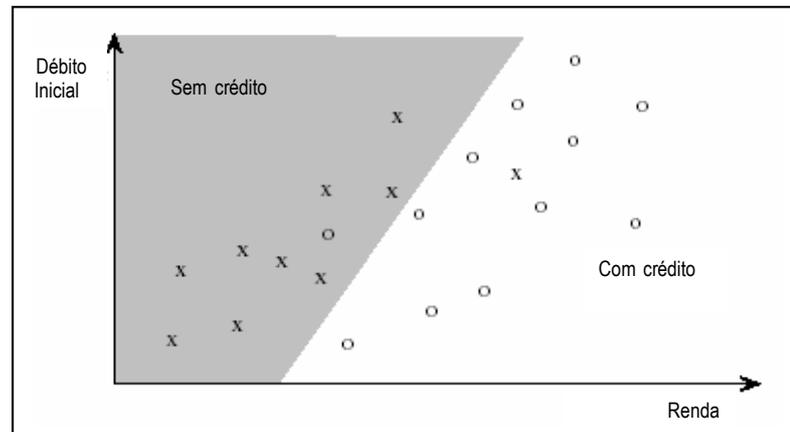
3.2.3 Técnicas de *data mining*

No Quadro 3.3 foram apresentadas as principais operações analíticas de *data mining* relacionadas com as técnicas da mesma abordagem. Agora serão apresentadas as principais técnicas de *data mining*. Dentro dos modelos preditivos, segundo Cabena et al. (1997), as principais técnicas utilizadas são a classificação e a predição de valor.

A classificação consiste em determinar uma função ou critério que classifica um registro de dados em uma dentre as diversas classes predefinidas (FELIPE, 2000). Por exemplo, na Figura 3.8, apresenta uma agência bancária em que é utilizada uma equação linear relacionando *debit* (débito inicial) e *income* (renda) para separar em *no loan* (sem crédito) e *loan* (com crédito) duas classes representadas pelos

¹⁶ Neste item, embora o OLAP seja uma abordagem de TI, é considerado em um dos seus aspectos como uma ferramenta analítica para o uso do *data mining*.

X (pessoas que têm débito no banco) e O (pessoas que têm crédito no banco). Pode-se verificar que a função linear se aproxima muito da realidade dos clientes do banco.



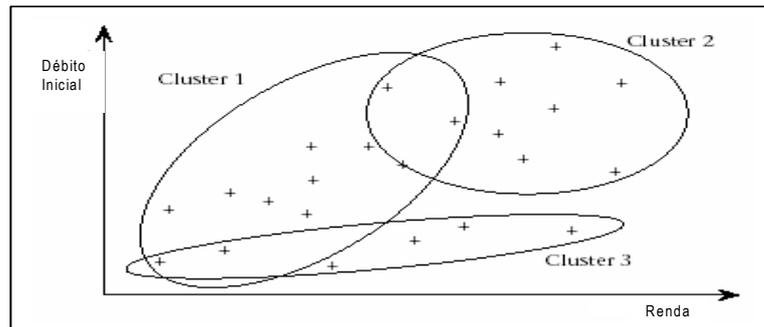
Fonte: Fayyad et al. (1996, p. 44).

FIGURA 3.8 - Exemplo de uso da técnica de classificação

Já a técnica predição de valor busca informar antecipadamente valores usando uma série de valores (dados) já existentes (CABENA et al., 1997). Busca-se nessa técnica encontrar uma função que mapeie um conjunto de dados, colocando-os como valores reais em uma variável de predição (FAYYAD et al., 1996). Um exemplo de uso dessa técnica é a determinação da probabilidade de um paciente sobreviver a uma cirurgia utilizando dados históricos de sucessos e fracassos na cirurgia de outros pacientes.

Na operação analítica de agrupamento em bancos de dados, destacam-se duas técnicas que são o agrupamento via redes neurais e separação demográfica (CABENA et al., 1997). As duas técnicas buscam identificar um conjunto de categorias finitas ou *clusters* para descrever os dados (ver na Figura 3.9 um exemplo de separação de dados em uma agência bancária baseada nos critérios de débitos e investimentos).

Nos agrupamentos via redes neurais, os *inputs* são agrupados e apresentados. Por se tratar de um aprendizado supervisionado, *outputs* são definidos. A partir disso inicia-se o processo de conexões entre *inputs* e *outputs* buscando-se realizar a ligação mais adequada entre eles estabelecendo-se pesos para as conexões (CABENA et al., 1997).



Fonte: Fayyad et al. (1996, p.45).

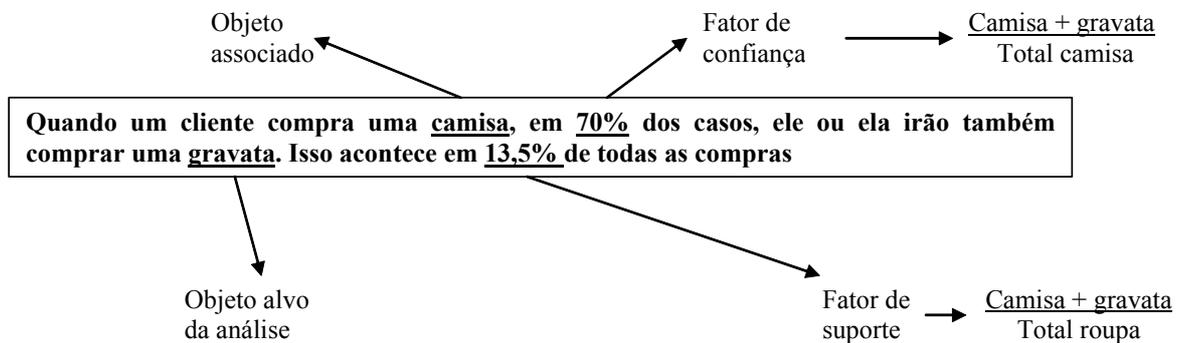
FIGURA 3.9 - Exemplo de separação dos dados

Segundo o mesmo autor, a separação demográfica constrói segmentos comparando cada registro de dados com todos os segmentos criados e depois adiciona-os a um deles ou cria um novo caso não encontre compatibilidade. Por exemplo, uma agência de pesquisas realiza uma atualização anual do banco de dados demográfico contendo inúmeras informações sobre as características da população de uma determinada região. Na variável “cor do cidadão” existem os segmentos “brancos”, “negros”, “mulatos”, “amarelos” e outros tipos. Nessa nova atualização a base de dados recebe uma quantidade de dados considerável para a cor “pardos”, o que faz com que a ferramenta de gerenciamento do banco de dados crie um novo segmento “pardos” para a variável cor dos cidadãos.

Na operação analítica de análises de associações existem, segundo Cabena et al. (1997), três técnicas: descoberta de associações, descoberta de padrões seqüenciais e descoberta de seqüências similares no tempo.

A descoberta de associações tem o propósito de encontrar registros de dados que implicam na presença de outros registros da mesma transação (BISPO, 1998). Um exemplo de uso dessa técnica está na Figura 3.10.

Na descoberta de padrões seqüenciais há a detecção de padrões entre registros de dados relacionados com um conjunto de outros registros em transações diferentes de um banco de dados em um dado período de tempo (CABENA et al., 1997). Por exemplo, dados de clientes que realizaram compras em um supermercado são comparados em períodos subseqüentes e procura-se descobrir associações entre os produtos comprados por esses clientes nesses períodos seqüenciais.



Fonte: Cabena et al. (1997, p.81).

FIGURA 3.10 - Exemplo de uso da técnica de descoberta de associações

Por fim, dentro da operação análise de associações, há a descoberta de seqüências similares no tempo que é definida pela busca de todas as ocorrências, similares ocorrências ou seqüências similares para uma seqüência-alvo com dados em série temporal (CABENA et al., 1997). Por exemplo, em uma empresa de autopeças são monitoradas as vendas de transmissões leves e transmissões pesadas em um período de um ano, quando então busca-se associar as vendas desses dois produtos para encontrar padrões de sazonalidade entre eles.

Na operação analítica “detecção de desvios”, são apresentadas as últimas técnicas que são a estatística e a visualização. Elas focalizam a descoberta das mais significantes mudanças nos dados previamente medidos ou que utilizam valores normativos (FAYYAD et al., 1996).

As técnicas de análise estatística detectam a presença de dados estranhos ou inválidos e os possíveis vieses nos dados que podem inviabilizar a mineração de dados (CABENA et al., 1997). Por exemplo, uma empresa tem um banco de dados contábil que contém informações sobre salários dos funcionários digitados pelo pessoal administrativo. Realizando-se uma análise estatística descritiva (com o uso de valores máximos e mínimos e contagem), verifica-se que diversos dados foram digitados incorretamente podendo enviesar uma posterior análise financeira.

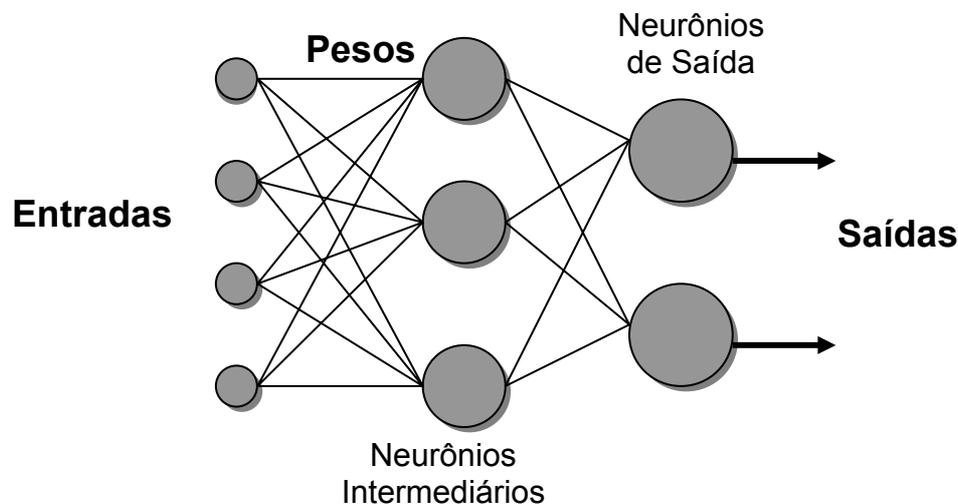
Já a técnica de visualização possibilita ao usuário obter uma compreensão melhor e mais intuitiva dos dados e os resultados apresentados (FELIPE, 2000). Ela auxilia as outras técnicas de *data mining*, mostradas nas Figuras 3.8 e 3.9, em

que foi utilizada a visualização gráfica bidimensional para facilitar o entendimento das técnicas de classificação e, na operação analítica, o agrupamento em bancos de dados.

3.2.4 Ferramentas de *data mining*

Nessas técnicas de *data mining* apresentadas, há um conjunto de ferramentas que auxiliam a sua utilização conforme visto no Quadro 3.4.

Dentre as principais ferramentas de *data mining*, tem-se a rede neural que simula o pensamento cognitivo humano modelando o inerente paralelismo dos circuitos neurais encontrados no cérebro usando modelos matemáticos como função dos circuitos. A ferramenta é mesclada de um nível de nós de *input* ou entradas (variáveis independentes), um ou mais níveis de nós intermediários (ou escondidos) e um nível de nós de *output* ou saídas (variáveis dependentes), como mostra a Figura 3.11. A rede neural só trabalha com dados numéricos (SPRANGLER et al., 1999).

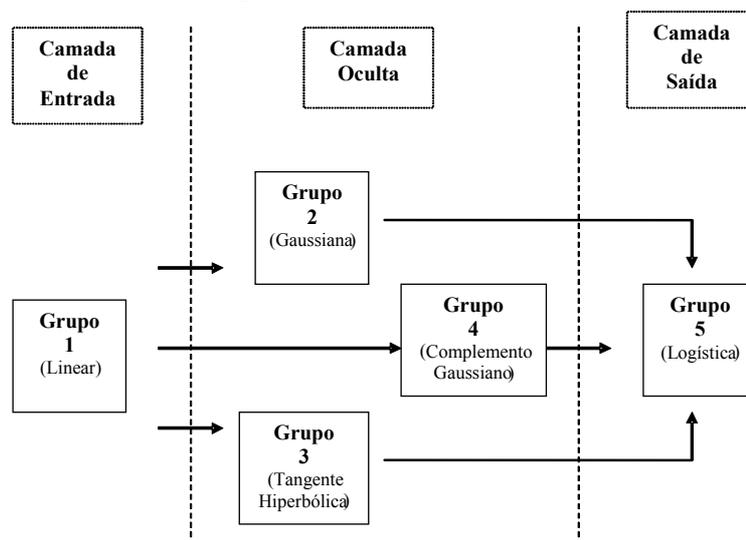


Fonte: Tafner, 1998.

FIGURA 3.11 – Estrutura de uma rede neural

Segundo o mesmo autor, a rede neural utiliza otimização não-linear com o algoritmo denominado retropropagação, em que os valores de *input* são propagados pela rede derivando uma série de valores de *output* e comparando com os valores de *output* previamente definidos para calcular a diferença entre os dois números. Se houver diferenças significantes, o algoritmo retorna ao nível de *input* ajustando os pesos entre várias conexões com base nos gradientes deles, para reduzir a soma total dos erros até um nível aceitável.

Um exemplo de utilização das redes neurais artificiais foi apresentado por Salle et al. (2001) na predição dos parâmetros de desempenho das aves em recría, pertencentes a uma integração avícola localizada no Sul do Brasil. Foram utilizados os registros de 11 lotes de reprodutoras pesadas (galinhas) para a análise por redes neurais artificiais, no período compreendido entre 09/11/1997 a 10/01/1999. Os dados utilizados corresponderam a 273 linhas de registros semanais. Os modelos de redes neurais artificiais foram comparados e selecionados como melhores, utilizando-se os critérios de Coeficiente de Determinação Múltipla (R^2), Quadrado Médio do Erro (QME), bem como a análise de gráficos, plotando a predição da rede *versus* a predição menos o real. Um aspecto interessante no trabalho desses autores foi a arquitetura montada para a rede neural como está mostrada na Figura 3.12.



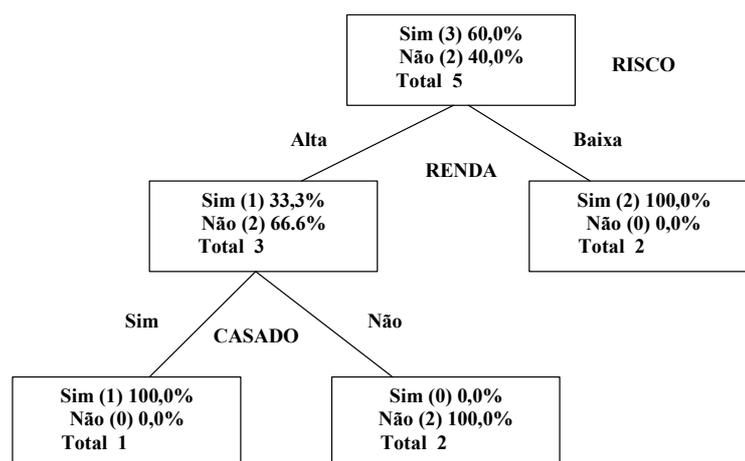
Fonte: Salle et al. (2001, p.261).

FIGURA 3.12 - Arquitetura da rede neural

A arquitetura utilizada foi de retropropagação com redes recorrentes supervisionadas e três grupos de neurônios ocultos e diferentes funções de ativação. A camada de entrada (grupo de neurônios número 1) utilizou uma função de escala linear $[-1,1]$. O primeiro grupo de neurônios da camada oculta, grupo 2, utilizou função gaussiana. O segundo grupo oculto, grupo 3, utilizou função tangente hiperbólica, e o terceiro grupo oculto, grupo 4, utilizou a função de complemento da Gaussiana. A camada de saída (grupo 5) utilizou uma função de ativação logística (logística sigmóide). As ligações entre neurônios foram ajustadas para taxa de aprendizagem e momento, em 0,1, e os pesos iniciais, entre +0,3 e -0,3.

A *lógica Fuzzy ou nebulosa* foi estruturada em 1965 pelo Dr. Lofti A. Zadeh da Universidade da Califórnia e surgiu da necessidade do tratamento de problemas onde existe uma região cujas classificações tornam-se imprecisas (VANDERLEI FILHO et al., 2002). Ela não trata as variáveis como tendo apenas um estado, mas sim n estados, cada qual com um certo grau de associação, ou seja, uma casa não é grande e sim 0,8 grande, 0,2 média e 0,0 pequena. Isso gera a criação de conjuntos nos quais os dados podem ser inseridos. Essas características lhe permitem um melhor tratamento de problemas onde existam fronteiras imprecisas. A lógica nebulosa tem sido empregada em diversas áreas de modo isolado ou em associação com outras técnicas, por exemplo, árvores de decisão (WATANABE, 2003).

A ferramenta *árvore de decisão* pode ser utilizada tanto em modelos preditivos quanto descritivos (FELIPE, 2000). A árvore de decisão torna explícita a representação de classificações como uma série de cortes binários que induzem o conhecimento de uma forma estruturada, facilitando o entendimento e validade. A árvore de decisão pode ser transformada em um equivalente conjunto de regras, ou seja, em vez da representação em árvore, pode ser feito um algoritmo apresentando as regras para cada série de corte binário (SPRANGLER et al., 1999). A título de exemplo, na Figura 3.13 são estabelecidas regras para clientes que receberão créditos em uma agência bancária.



Fonte: Felipe (2000, p. 30).

FIGURA 3.13 - Exemplo da ferramenta árvore de decisão

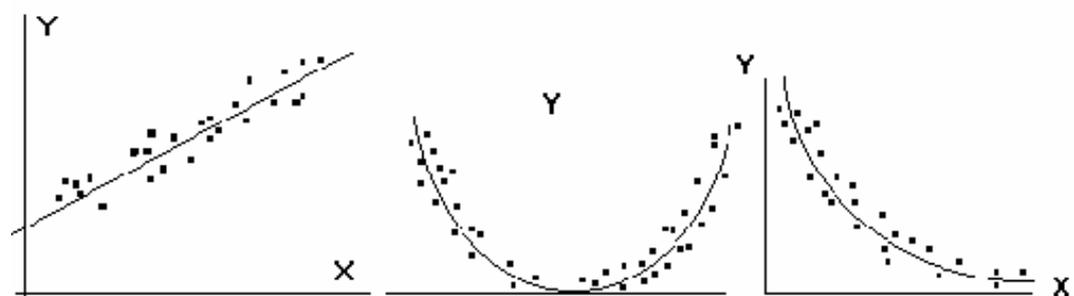
De acordo com a Figura 3.13, são escolhidos cinco clientes para verificar se o perfil deles está de acordo com o que a agência bancária deseja-clientes de renda alta e casados. Verifica-se na figura que apenas um cliente se encaixa nesse perfil.

Há também as ferramentas de regressão, que buscam prever valores a partir de um conjunto de dados. Para isso são construídas funções matemáticas a partir de variáveis definidas pelo usuário (FAYYAD et al., 1996). Existem diversas formas de regressão, a mais simples delas é a regressão linear, que prediz o valor a partir de uma função linear.

Também existem ferramentas de regressão mais complexas, segundo Felipe (2000) como:

- regressão múltipla, em que há a presença de mais de uma variável dependente;
- regressão não-linear segundo a qual o comportamento de um conjunto de dados não pode ser satisfatoriamente explicado a partir de uma reta; e
- métodos de classificação que consistem em um conjunto de técnicas para predição utilizando-se combinações lineares e não-lineares como variáveis de entrada.

Para se estabelecer as funções matemáticas, a representação gráfica auxilia na visualização de tipo de regressão é mais adequado para um conjunto de dados, como é mostrado na Figura 3.14.



Fonte: Polito, 1997.

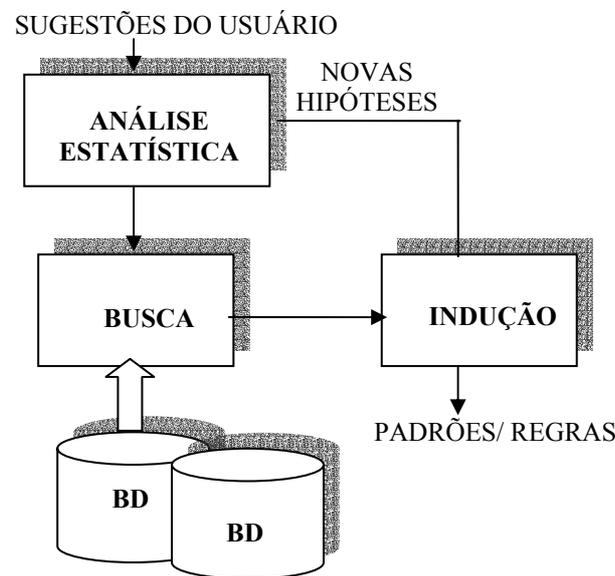
FIGURA 3.14 - Representação gráfica de funções lineares e não lineares.

Por exemplo, na Figura 3.14 o primeiro gráfico mostra que a regressão linear é a mais adequada, visto que o conjunto de dados pode aparentemente ser geometricamente descrito, de forma satisfatória, como uma reta.

Uma das ferramentas utilizadas na operação analítica análise de associações é o *market basket analysis*, que localiza e agrupa itens (ou produtos) a partir de regras (THURAISINGHAM, 1999). Ela é bastante utilizada em *marketing*, pois realiza análise combinatória dos produtos juntamente com a localização geográfica do cliente, disponibilizando investimentos em *marketing* específicos para determinados produtos e regiões (SILVA, 2000).

Essa ferramenta pode utilizar sistemas especialistas, segundo Metaxioxis & Parras (2003), para realizar a busca e estabelecer critérios que limitam essa tarefa. O grande desafio dessa ferramenta é determinar os critérios de busca das associações e eliminar os itens desnecessários (THURAISINGHAM, 1999).

Outra ferramenta utilizada na análise de associações é o *Inductive Logic Programming (ILP)*, que infere regras para um conjunto de dados. O ILP é uma lista "não-encomendada" apresentada aos usuários por meio da detecção de tendências dentro de grupos de dados, ou de "regras" sobre o dado, com pouca interação do usuário (SILVA, 2000), como mostrado na Figura 3.15.



Fonte: Polito, 1997.

FIGURA 3.15 – Arquitetura do ILP.

Por exemplo, em um banco de dados demográfico, o sistema de indução pode, primeiramente, examinar como as idades são distribuídas e pode perceber uma interessante variação para aquelas pessoas cujas profissões são listadas como atletas profissionais. Esta hipótese é, então, considerada relevante, e o sistema criará uma regra

tal como (Se Profissão = Atleta; Então Idade < 30). Esta regra pode ter uma "margem de precisão" de 70% ligada a ela, ou seja, de cem de todos os atletas, setenta estão abaixo de 30 anos. Entretanto, este padrão pode não sustentar as idades dos bancários ou professores no mesmo banco de dados (POLITO, 1997).

Por fim, a ferramenta sumarização estatística dos dados realiza uma descrição compacta de um subconjunto ou conjunto de dados (FAYYAD et al., 1996). Vários *softwares* (ver lista de *softwares* no Apêndice B) podem realizar uma série de operações estatísticas no conjunto de dados, por exemplo, média, moda, mediana, desvio padrão, variância, entre outros. Os resultados são apresentados em uma tabela e/ou gráfico para a realização da análise pelo usuário. O Quadro 3.5 apresenta um resumo das principais ferramentas apresentadas.

QUADRO 3.5 – Vantagens e desvantagens das ferramentas de *data mining*

FERRAMENTAS	VANTAGENS	DESVANTAGENS
Rede Neural	- Alto grau de otimização entre os resultados reais e os preditos	- A análise só pode ser feita em dados numéricos - Não se tem acesso ao modo de funcionamento do processamento dos dados
Lógica Fuzzy	- Permite uma melhor segmentação de sistema, pois não se baseia somente em duas opções (por exemplo, verdadeiro e falso). - Pode ser aplicada em dados não numéricos.	- É necessário um profundo conhecimento do domínio da aplicação para poder modelá-lo e descrever os comportamentos próximos da realidade em que está sendo feita a aplicação. Esse profundo conhecimento já não é tão necessário em ferramentas como a árvore de decisão.
Árvore de Decisão	- Fácil visualização da estrutura hierárquica das informações	- Necessidade de dados com pouco número de variáveis. - A árvore de decisão só toma decisão binariamente, ou seja, com apenas duas opções, não abrangendo as outras.
Regressão	- Útil na predição de valores a partir de um grande conjunto de dados - Fácil entendimento do comportamento das variáveis a partir de análise gráfica	- Limitado a dados numéricos - Depende do comportamento dos dados - Presença de grandes variações nos dados pode inviabilizar o uso da ferramenta.
<i>Market Basket Analysis</i>	- Fácil entendimento dos resultados, pois as associações entre os dados são baseadas em regras lógicas.	- Limitada apenas à área de <i>Marketing</i> .
<i>Inductive Logic Programming</i>	- Fácil entendimento dos resultados, pois se baseiam em regras lógicas.	- Pouca interação do usuário, que pode tornar difícil o entendimento de como foi o processamento dos dados na ferramenta
Sumarização Estatística dos Dados	- Fácil visualização e entendimento do processo de obtenção das informações e dos próprios resultados	- Limitada a dados numéricos.

Fonte: Elaborado pelo autor.

Existem inúmeras ferramentas para aplicação das técnicas de *data mining*. Cada uma delas apresentada nesta dissertação tem a sua vantagem e desvantagem como pode ser visto no Quadro 3.5.

3.3 O *Data Mining* na Medição de Desempenho

As operações analíticas e técnicas de *data mining* podem auxiliar na análise de dados oriundos da medição de desempenho. No Quadro 3.6 estão relacionadas as operações analíticas e as técnicas de *data mining* com as possíveis aplicações na medição de desempenho.

QUADRO 3.6 - As técnicas de *data mining* e as aplicações na medição de desempenho

Operações Analíticas	Técnicas	Aplicações na medição de desempenho
Modelagem Preditiva	Classificação	Separação das medidas de resultado (<i>lagging</i>) e direcionadoras do resultado (<i>leading</i>) utilizando algum critério como tempo de espera da resposta das medidas de desempenho
	Predição de valor	Predizer uma medida de desempenho a partir de outras medidas relacionadas utilizando algum método de regressão
Agrupamento de base de dados	Agrupamento via redes neurais	Rearranjo de dados numéricos de desempenho de um <i>data mart</i> visando otimizar os relacionamentos entre as medidas de entrada, processo e saída.
	Separação Demográfica	Agregação ou desagregação de dados de desempenho em classes com base em um critério visando atender alguma necessidade estratégica da companhia
Análise de Associações	Descoberta de associação	Determinação de relacionamentos com base em regras entre as medidas de desempenho em um mesmo evento
	Descoberta de padrões seqüenciais	Determinação de relacionamentos com base em regras entre as medidas de desempenhos em eventos seqüenciais, ou seja, eventos ocorridos em tempos diferentes
	Descoberta de seqüências similares no tempo	Análise de valores de variáveis de desempenho em um período visando à descoberta de relacionamentos entre as variáveis
Detecção de desvios	Estatística	Útil na estimação da probabilidade de um relacionamento de causa-e-efeito ocorrer entre as medidas de desempenho
	Visualização	Apresentação dos resultados em formato gráfico, matriz ou tabela facilitando a análise de um maior número de variáveis de desempenho

Fonte: Elaborado pelo autor.

Numa análise do Quadro 3.6, pode-se notar que a modelagem preditiva, que contém as técnicas de classificação e predição de valor, se aplica na medição de desempenho para o estabelecimento da posição de cada medida de desempenho frente a um critério classificatório estabelecido pelo decisor, e na predição da finalidade e do impacto gerado por cada medida de desempenho.

O agrupamento de base de dados, que pode ser feita via rede neural ou demográfica, visa organizar os dados de desempenho proveniente de uma base de dados. A organização dos dados pretende facilitar e otimizar a análise perante o problema estabelecido no projeto de *data mining*. Geralmente, essa operação de *data mining* é acompanhada de outras operações como modelagem preditiva e análise de associações.

A análise de associações, como as associações simples, sequenciais ou similares no tempo, visa estabelecer relacionamentos entre as medidas de desempenho com o estabelecimento de uma regra que dita essa relação. Os relacionamentos entre as medidas de desempenho podem ocorrer no mesmo ou em diferentes períodos, ou até mesmo quando elas atendem a produtos ou processos diferentes.

A detecção de desvios, com as técnicas estatísticas e de visualização, auxilia na determinação de probabilidades de um relacionamento entre as medidas de desempenho, como também localiza pequenas flutuações nos dados de desempenho que podem indicar algum desvio ou viéses nos dados. Essa operação pode servir de apoio a outras, como modelagem preditiva e análise de associações.

Pode-se verificar que o uso da abordagem de *data mining* é útil para a medição de desempenho. Dentre as três razões, apresentadas por Rodrigues Filho & Shimizu (2002), para se realizar um projeto de *data mining* como contribuição para a medição de desempenho, estão: a) visualização de dados como definidor de um posicionamento das medidas de desempenho frente a um critério classificatório estabelecido pelo decisor e melhor agrupamento dos dados de desempenho seguindo um critério estratégico; b) descoberta de novos conhecimentos no estabelecimento dos padrões de relacionamentos entre as medidas de desempenho; e c) dados mais consistentes para detectar possíveis desvios na população que contém os dados de desempenho.

3.4 Considerações e Síntese do Capítulo

Conforme visto neste capítulo, o *data mining* é uma abordagem útil na exploração de bancos de dados históricos para a obtenção de informações relevantes para a tomada de decisão.

Ele é uma abordagem atual que possui uma variedade de recursos tanto em nível de *software* como de *hardware* para gerar informações relevantes a partir de grandes bancos de dados. O *data mining* pode apoiar diversos sistemas de informações da empresa, principalmente os Sistemas de Apoio à Decisão (SAD), Suporte Executivo e Sistemas de Conhecimento, e é o foco nesta dissertação para o SAD.

Existem diversas abordagens, formas e condições de se aplicar o *data mining*. No geral, ele é estruturado a partir de uma necessidade organizacional da empresa.

Essa necessidade gera recursos para a aplicação da abordagem, sendo o primeiro deles num banco de dados histórico, que pode ser ou não um *data warehouse*. O *data warehouse* facilita a aplicação das ferramentas de *data mining* devido ao fato de os dados estarem organizados como cubo de dados, facilitando a manipulação desses.

Neste trabalho, o *data mining* foi apresentado em operações analíticas que contêm as principais áreas de atuação da abordagem. Cada operação analítica contém diversas técnicas, que são procedimentos para a aplicação das principais ferramentas de *data mining*.

Na medição de desempenho, o *data mining* pode auxiliar de diversas formas, com destaque para o estabelecimento de relacionamentos entre as medidas de desempenho com base em regras, agrupamentos dos dados de desempenho em *clusters* ou em algum sistema classificatório definido pelos decisores e na redução da probabilidade de os dados de desempenho possuírem vieses.

4. PROPOSIÇÃO DE UM MÉTODO PARA ESTABELECIMENTO DE RELACIONAMENTOS ENTRE AS MEDIDAS DE DESEMPENHO

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 mais o número de variáveis

envolvidas cresce, torna-se importante a busca do entendimento da complexa teia de relacionamentos existente entre as medidas de desempenho numa organização.

Outro fator importante, já demonstrado em pesquisa empírica por Martins (1998), é que o relacionamento entre as medidas de desempenho é algo corriqueiro para os entrevistados. Mesmo que a pesquisa não permita generalizações, o relacionamento foi encontrado em maior e menor grau. Entretanto, o que pode ser improdutivo é que esses relacionamentos são parciais e pessoais, visando articular as variáveis por cujo desempenho os entrevistados, na maioria dos casos, tinham responsabilidade. O ideal seria que a maioria dos tomadores de decisão compartilhasse o mesmo modelo de relacionamento entre as medidas de desempenho e que ele fosse mais abrangente possível. Isso poderia reduzir os conflitos e erros na busca dos objetivos estabelecidos. Nesse sentido, nesta dissertação é proposto um método, a partir da aplicação da abordagem de *data mining*, para o estabelecimento de relacionamentos entre as medidas de desempenho em banco de dados históricos de desempenho.

Antes de o método proposto ser apresentado, é interessante demonstrar o método científico que pautou o desenvolvimento da proposta.

4.1 Proposta de Pesquisa

Nesta dissertação, a pesquisa consistiu em uma investigação da literatura para a proposição de um método que possa estabelecer relacionamentos entre as medidas de desempenho a partir do uso do *data mining*.

Método pode ser definido como uma série de regras para tentar resolver um problema (MAZZOTTI & GEWANDSZNAJDER, 1995). É um caminho, processo ou meio para se chegar a um determinado objetivo (FERREIRA, 2004).

O método utilizado nesta dissertação se baseia na lógica dedutiva, que procura demonstrar ou justificar, partindo de leis ou teorias aplicadas a uma situação específica na busca da solução, explicando o conteúdo da pesquisa (LAKATOS & MARCONI, 1995). Para isso, são utilizados teorias, conceitos e hipóteses já comprovados da literatura para a proposição do método.

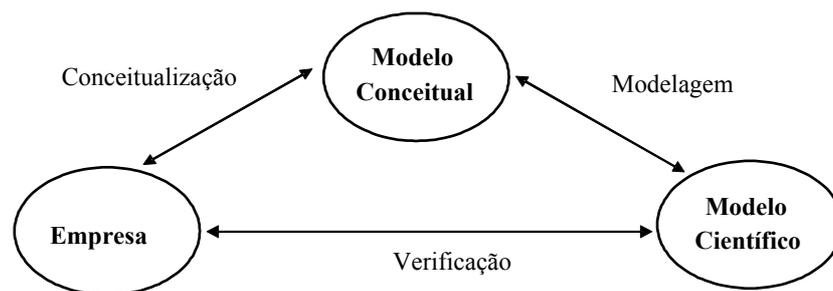
Esse método tem como princípio a realização de uma modelagem de empresa, visando permitir a operacionalização do *data mining* no estabelecimento de

relacionamento entre as medidas de desempenho. Para isso, a modelagem tem que ser descritiva.

A modelagem é descritiva, quando, segundo Fawcett & Downs (1986), ela tem o intuito de descrever ou classificar dimensões específicas ou características de indivíduos, grupos, situações ou eventos comumente encontrados em observações discretas.

A modelagem proposta no método se baseia em Bertrand & Fransoo (2002). Para esses autores, ela busca primariamente criar um modelo que descreve adequadamente o processo que está ocorrendo no ambiente em que vai ser aplicado o método.

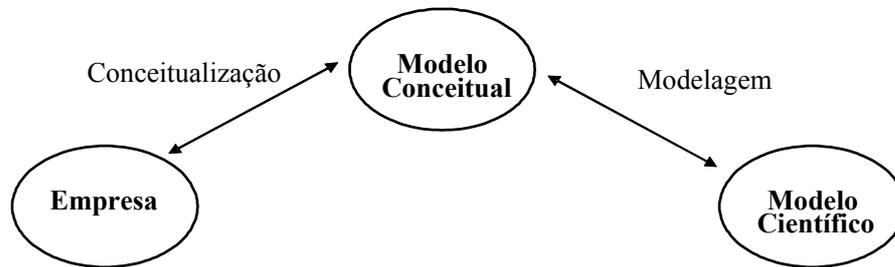
Para entendê-lo melhor, é utilizado o modelo proposto por Mitroff et al. (1974), apresentado na Figura 4.1.



Fonte: Bertrand & Fransoo (2002, p. 252)

FIGURA 4.1 – Modelo de pesquisa proposto pelos autores

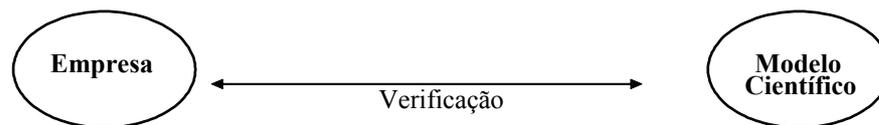
O modelo de pesquisa inicialmente gera o modelo conceitual, que busca levantar informações do ambiente em que vai ser aplicado. Este modelo conceitual pode ser obtido de diversas formas, por exemplo, entrevistas não-estruturadas (sem roteiro de perguntas), consulta a documentos, entre outros. Com o modelo conceitual criado, é feita a sua adequação (modelagem) ao ambiente em que vai ser aplicado permitindo a obtenção do modelo científico pronto para ser feita a verificação (ver Figura 4.2).



Fonte: Bertrand & Fransoo (2002, p. 252)

FIGURA 4.2 – Geração do modelo conceitual.

Posteriormente, a utilização do modelo de pesquisa consiste na fase empírica do projeto, que é a verificação do modelo científico. É geralmente realizada uma pesquisa quantitativa visando a comprovação do modelo (Figura 4.3).



Fonte: Bertrand & Fransoo (2002, p. 252)

FIGURA 4.3 – Verificação do modelo de pesquisa.

Em princípio, o projeto de pesquisa estabelecido seguiria os passos acima. Entretanto, devido à impossibilidade de encontrar uma organização que permitisse a aplicação do método e ao tempo limitado para o desenvolvimento desta dissertação, o escopo foi reduzido à apresentação do método para estabelecimento de relacionamentos entre as medidas de desempenho, o que é feito a seguir.

4.2 Visão Geral do Método para Estabelecimento de Relacionamentos entre as Medidas de Desempenho

O método proposto para ser implementado em uma empresa necessita de requisitos e da meta aonde se pretende chegar com o estabelecimento dos 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 serão 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 a empresa possui e como estariam dispostos em um banco de dados. Nesta etapa, será 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.

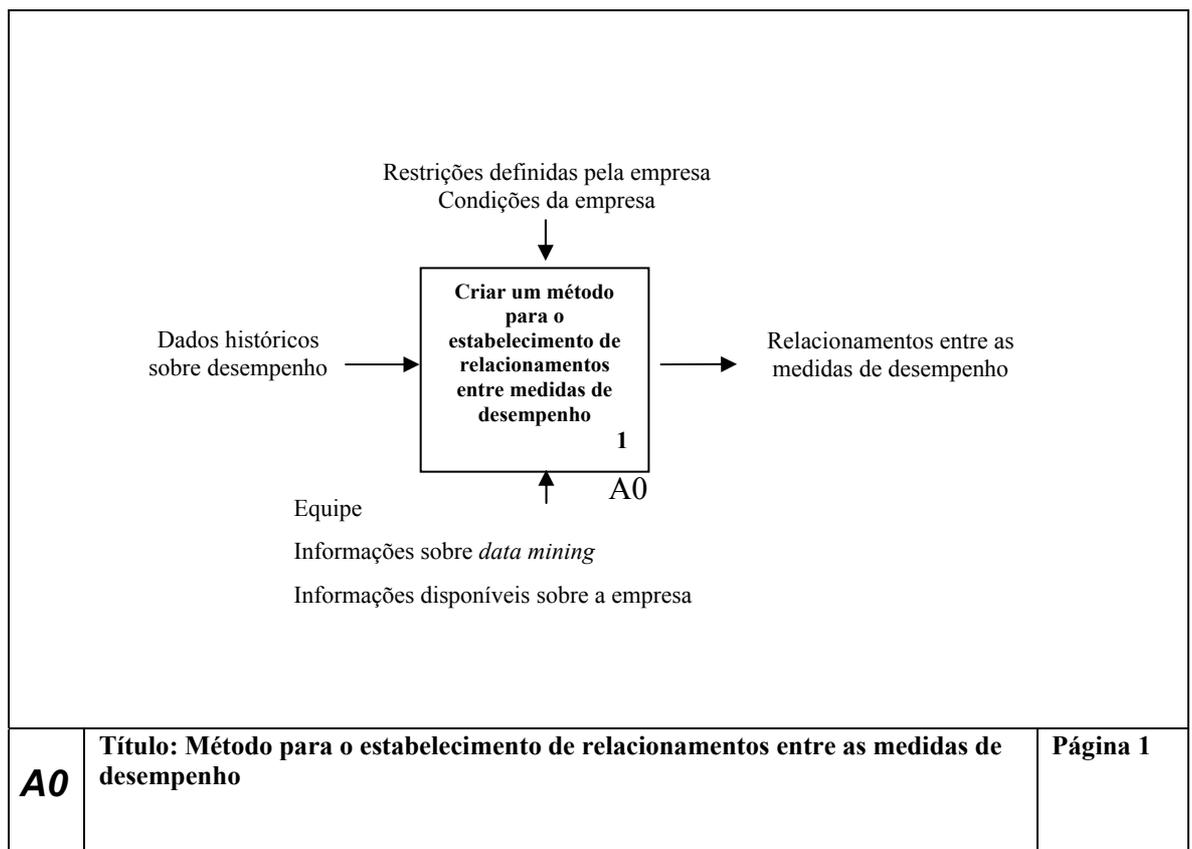
E, a partir do modelo conceitual, são estabelecidos os relacionamentos entre as medidas de desempenho com base no uso das técnicas e ferramentas de *data mining*. A escolha da(s) técnica(s) e ferramenta(s) será baseada na estrutura dos dados contida no banco de dados histórico e na análise do modelo conceitual obtendo-se um modelo de *data warehouse* propício para a operacionalização do *data mining*. O objetivo da utilização da(s) técnica(s) e ferramenta(s) de *data mining* 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*. O usuário (tomador de decisão) consulta diretamente o *data warehouse* a fim de extrair os dados para o uso do *data mining* e também pesquisa utilizando o *data mining* a fim de obter informações relevantes para a tomada de decisão.

4.3 Detalhamento do Método Proposto

No item anterior, foi apresentada uma visão geral do método proposto. Agora ele será detalhado e, para tanto, o *Integration Definition for Function Modeling* (IDEF0) é utilizado. No Anexo A, é possível encontrar um detalhamento do IDEF0.

A Figura 4.4 ilustra como, a partir de dados históricos de desempenho, são estabelecidos os relacionamentos entre as medidas de desempenho. Para isso, são usados recursos como uma equipe, informações obtidas na literatura e na empresa e os bancos de dados. Também são estabelecidos mecanismos que delimitarão o método, que são as restrições e condições da empresa.



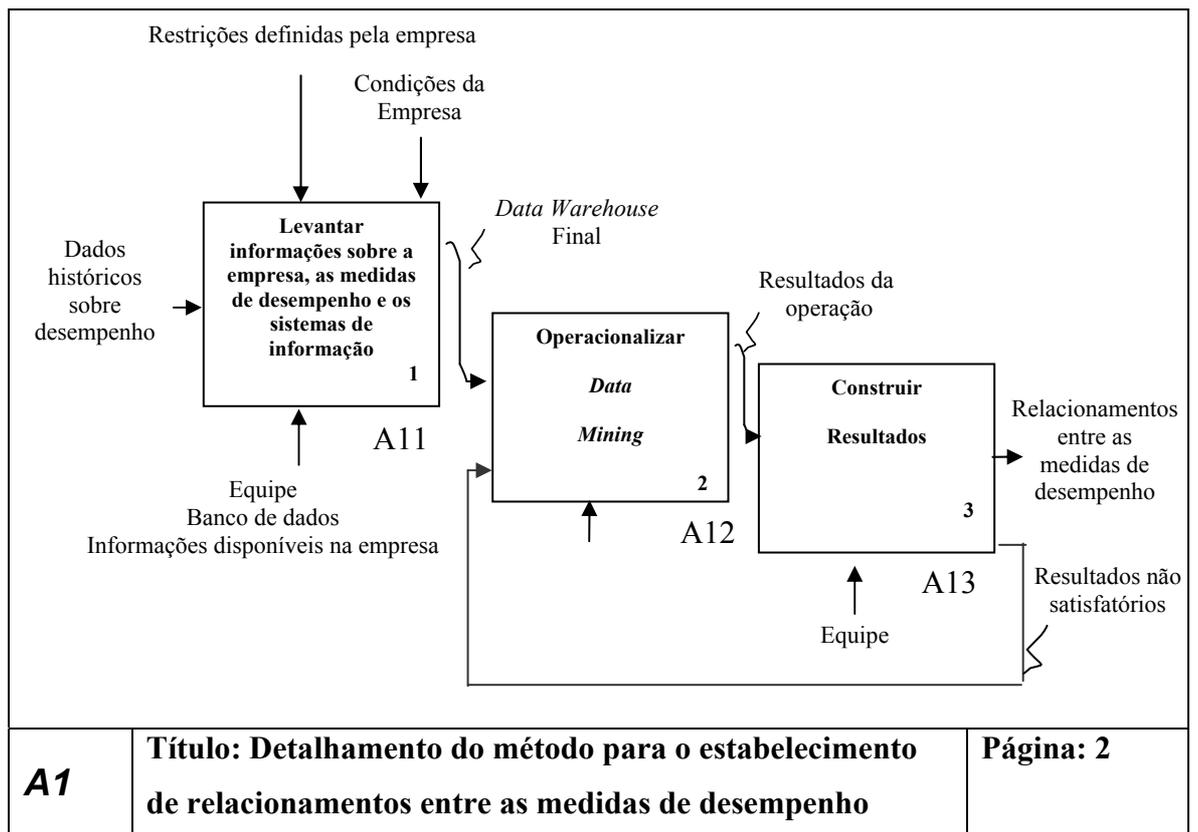
Fonte: Elaborado pelo autor.

FIGURA 4.4 – Visão geral do método para o estabelecimento de relacionamentos entre as medidas de desempenho.

Na Figura 4.5, é apresentado um diagrama filho a partir da Figura 4.4, ilustrando as três fases propostas para o estabelecimento dos relacionamentos entre as seguintes medidas de desempenho:

1. levantar as informações necessárias para o estabelecimento de relacionamentos entre as medidas de desempenho a partir de dados históricos sobre desempenho e da estrutura organizacional criada para manter estes dados;

2. operacionalizar o *data mining* em um modelo criado a partir das informações levantadas; e
3. usar os resultados obtidos pela equipe para a realização da análise e também para a disseminação dos resultados.



Fonte: Elaborado pelo autor.

FIGURA 4.5 – Detalhamento do método para o estabelecimento de relacionamentos entre as medidas de desempenho.

A forma mais recomendada para a aplicação do método proposto é por meio de uma equipe que tenha conhecimento sobre todos os seus aspectos que são: as informações da empresa, a operacionalização do *data mining* e a construção dos resultados. Essa equipe pode ser formada por um ou mais especialistas em tecnologia de informação voltados para a interpretação dos dados, para a aplicação das técnicas e ferramentas de *data mining*, e interpretação dos resultados, pelo menos por um engenheiro de produção que consiga traduzir a linguagem do ambiente produtivo para facilitar o trabalho dos especialistas em TI e também um membro da equipe que tenha conhecimentos sobre a prática de medição de desempenho da empresa.

Neste método, existem três questões fundamentais para o desenvolvimento do trabalho. A primeira delas é 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.

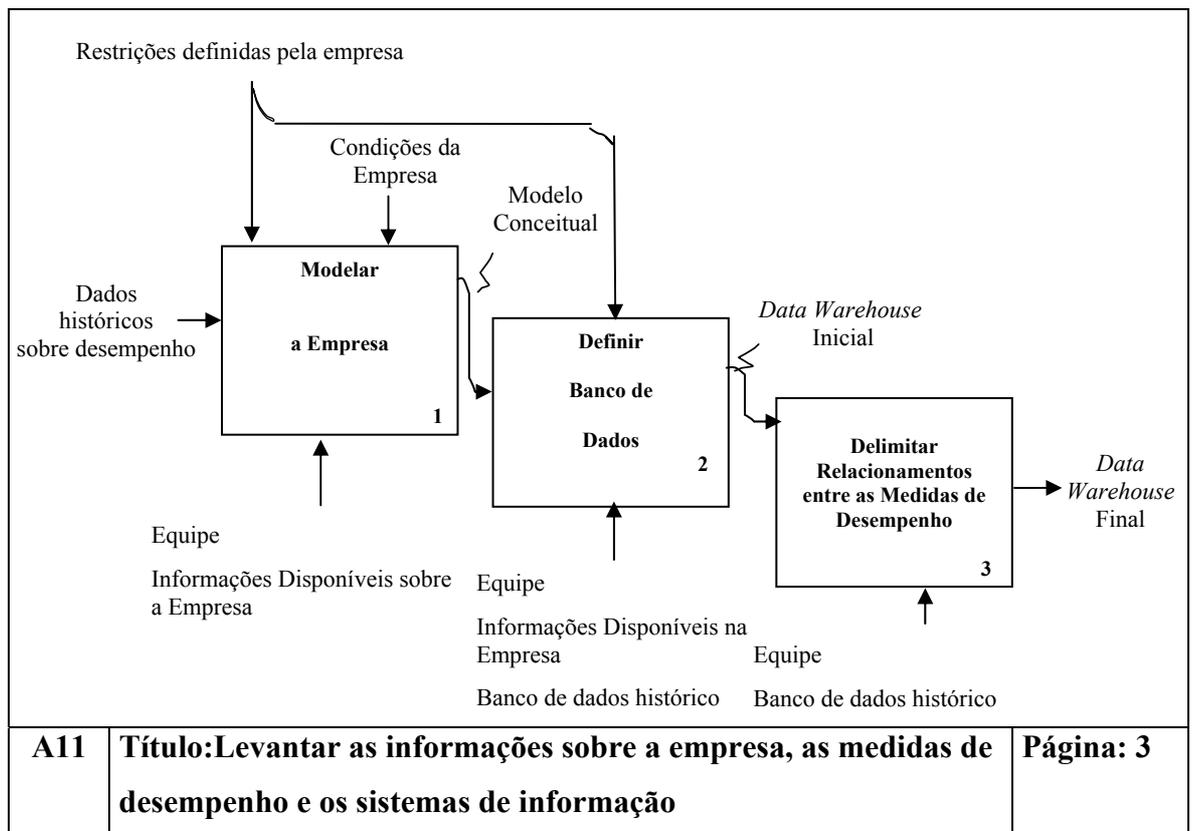
A segunda questão é a definição do banco de dados. A sua escolha vai depender do tamanho, da quantidade de variáveis presentes e das áreas organizacionais envolvidas. Um *data warehouse* ou um *data mart* conseguem lidar com dados armazenados, complicados e complexos, mas existe também possibilidade, por questão de custo e simplicidade, de aplicar as ferramentas de *data mining* diretamente em bancos de dados relacionais.

Por fim, a terceira questão é a escolha das técnicas e ferramentas de *data mining*. 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 investigado 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 *data mining* diferentes. Dessa forma, será possível estabelecer, a partir de uma grande massa de dados históricos sobre o desempenho, o relacionamento entre as medidas de desempenho de um SMD.

Essas três questões serão mais detalhadas no decorrer do item 4.3.

4.3.1 Levantamento de informações sobre a empresa

Para estabelecer os relacionamentos entre as medidas de desempenho usando o *data mining*, é necessário conhecer o SMD e os sistemas de informação da organização onde o método proposto vai ser aplicado. A Figura 4.6 detalha a forma como isso pode ser feito. A equipe deverá decidir a melhor forma de levantamento dessas informações, pois depende das contingências de cada situação, sendo quase impossível prevê-las. Portanto o modelo conceitual precisa ser capaz de descrever os relacionamentos entre as medidas de desempenho e também quais os tipos de dados e banco de dados que existem na empresa.



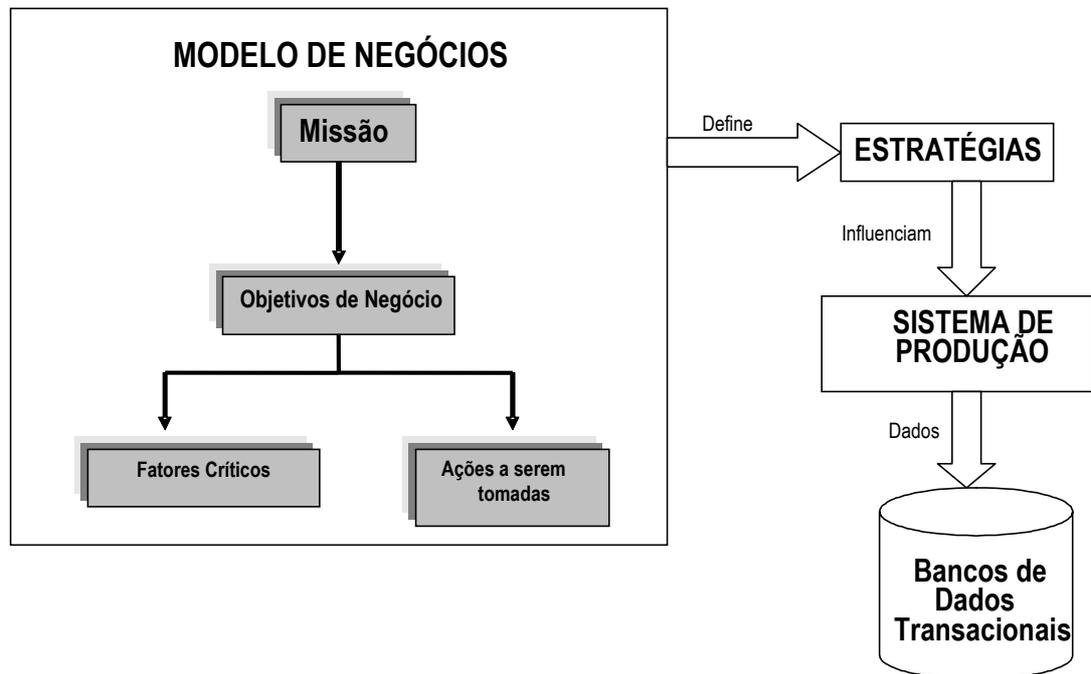
Fonte: Elaborado pelo autor.

FIGURA 4.6 – Adquirir conhecimento sobre a empresa, as medidas de desempenho e sistema de informação.

O levantamento das informações abrange três aspectos na definição do modelo conceitual:

- a) mapeamento dos possíveis relacionamentos entre as medidas de desempenho;
- b) identificação de quais informações são relevantes para os tomadores de decisão e que eles ainda têm dificuldades em obtê-las; e
- c) levantamento 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 do *data mining*.

Na modelagem da empresa, a equipe se deparará com questões acerca do modelo de negócios e da estratégia de negócios da empresa, e de como isso afeta a sua estrutura informacional, conforme mostra a Figura 4.7.



Fonte: Elaborado pelo autor.

FIGURA 4.7 – Geração dos dados a partir de um modelo de negócios

Vale observar que um modelo de negócios é estruturado a partir da missão da empresa. Ela define os escopos do mercado, da indústria, da tecnologia envolvida, da estrutura hierárquica e da área geográfica envolvida. Eles condicionam os objetivos de negócio da empresa e estes, por sua vez, definem quais são os fatores críticos de sucesso (por exemplo, custo, qualidade, entrega, velocidade e flexibilidade) relevantes e quais as ações a serem tomadas para o alcance desses objetivos.

Para exemplificar como isto condiciona o método proposto, dois exemplos de modelos de negócios são apresentados a seguir:

1. uma empresa que participa de uma rede de distribuição de componentes eletrônicos, e cuja a sua missão é satisfazer os clientes (outras empresas) entregando produtos no prazo certo e com qualidade. O objetivo de negócio da empresa é fechar contratos de negócios com diversos clientes que exigem entrega no prazo, e produtos seguindo normas de qualidade. Para isso, a empresa se foca nos fatores críticos (qualidade e entrega) e determina ações visando atender às normas da qualidade e investindo em logística; e

2. uma empresa do setor têxtil que atua verticalmente desde a obtenção da resina para a fibra até a distribuição do vestuário para as lojas de marca. A sua missão é oferecer o melhor preço para as lojas a fim de estar à frente dos concorrentes. O objetivo é ter o controle de todas as atividades que diretamente contribuem para a realização do produto. Para isso, a empresa se foca no fator crítico custo e visa reduzi-los nos produtos em processo e na matéria-prima.

Em ambos os exemplos, é possível verificar que o tomador de decisão terá que decidir em paradigmas diferentes. No primeiro caso, a decisão procura focar necessidades definidas pelo cliente, enquanto no outro as decisões sempre envolvem uma redução de custos e comparação com os outros concorrentes.

A partir do modelo de negócios, é definida qual estratégia é eficaz e eficiente no atendimento dos objetivos. Diversas estratégias podem ser estabelecidas, que influenciam o sistema de produção da empresa, determinando a melhor configuração na sua consecução.

Em Custodio et al. (2002), são apresentadas estratégias focadas no sistema de produção para quatro setores industriais como mostra o Quadro 4.1. São apresentados dois tipos de estratégia: a) aquelas que atuam na quantidade e variação de produtos acabados; e b) aquelas que atuam na forma como o cliente realiza o pedido. As estratégias estabelecidas pela empresa atendem ao seu modelo de negócios.

Verifica-se que os tipos de dados obtidos nos bancos de dados transacionais nos setores do Quadro 4.1 variam conforme o tipo de setor.

De acordo o Quadro 4.1, cada sistema de produção é classificado conforme a estratégia utilizada. O sistema de produção pode ser focado no processo (que se caracteriza por ser flexível tanto no arranjo de pessoas e máquinas para buscar atender uma grande variedade de produtos especificados pelos clientes) ou no produto (que se caracteriza por uma produção contínua ou quase contínua com alto volume de produção e baixa variedade de produtos) e também na produção para estoque (que se caracteriza pela formação de estoque de produto acabado sem haver um pedido confirmado pelo cliente) ou na encomenda (o produto acabado só é produzido quando há um pedido feito pelo cliente).

QUADRO 4.1 – Estratégias aplicadas em sistemas de produção de diferentes setores.

	Moldagens de Plásticos por Injeção	Bobinas e Chapas de Aços	Semicondutores	Meias feitas por pedidos especiais
Estratégia para variações de volume e grau de variedade de produtos	A moldagem por injeção neste tipo de sistema produtivo se caracteriza por uma produção intermitente, com grande variedade de produtos. O processo é flexível, pois a injetora permite a troca de moldes para o produto e cada produto tem um fácil ajuste de pressão e temperatura na máquina, tornando o processo crítico para a obtenção de um produto de qualidade.	Trata-se de uma produção contínua de alto volume de produção com equipamentos, materiais e pessoas dedicados somente para a obtenção do aço.	Caracteriza-se por ser um processo <i>job shop</i> onde existem variedades de produtos, dos quais cada um passa por seqüenciamento de operações diferentes; por isso há necessidade de um processo flexível em que o fácil rearranjo de máquinas, equipamentos e o conhecimento das pessoas do processo são determinantes.	Por se tratar de produção de baixa diversificação, e não existem diferenças na seqüência de operações dos produtos, permite-se uma especialização das máquinas para a obtenção do produto semi-acabado ou acabado, característico do setor têxtil.
Estratégia para atender os pedidos feitos pelos clientes	A previsibilidade da demanda é alta na relação com os clientes (geralmente firmas), embora haja uma grande variedade deles. A maturidade do mercado dos produtos gerados pelos clientes garante que irá haver necessidades de novos produtos acabados.	A produção é feita para estoque devido ao fato de o processo ser estático e qualquer mudança na tecnologia de processo gera altos custos para a empresa, outro fator menos determinante é a alta demanda presente na necessidade de diversas cadeias produtivas.	Trata-se de um setor de alta tecnologia com rápidos ciclos de inovação, e que permite uma imprevisibilidade de demanda e também do grau de especificidade exigida pelos clientes.	Por atender pedidos específicos feitos pelo cliente e pelo baixo volume de produção, as empresas deste setor usam uma estratégia de diferenciação na operação final para o produto semi-acabado garantindo uma variedade de produtos.
Classificação	Sistema focado no processo e produção para estoque	Sistema focado no produto e produção para estoque	Sistema focado no processo e produção por encomenda	Sistema focado no produto e produção por encomenda

Fonte: Custodio et al. (2002, p. 5)

Custodio et al. (2002) apresenta, ainda, as informações críticas do sistema de informação gerencial para os quatro setores analisados, conforme o Quadro 4.2. O sistema de informação gerencial recebe os dados do sistema de produção e transforma em informações relevantes para os tomadores de decisão. Então, para cada setor há uma variação de quais seriam os dados relevantes presentes nos bancos de dados transacionais, conforme mostra o Quadro 4.2.

QUADRO 4.2 – Informações críticas para o sistema de informação gerencial em quatro setores industriais.

	Moldagem de plásticos por injeção.	Bobinas e chapas de aço.	Semicondutores.	Meias feitas por pedidos especiais.
Informações críticas para o sistema de informação gerencial	<p>- Incerteza acerca de quando será feito o pedido do cliente, só se sabe quanto de cada produto será feito por um período, mas não se tem idéia quando será feito o pedido de um produto para a empresa.</p> <p>- Informações sobre os equipamentos para saber se a empresa tem condições de produzir um novo produto requisitado pelo cliente.</p> <p>- Informações sobre a matéria-prima é importante (principalmente a cor), pois pode afetar o processo.</p> <p>- <i>Benchmarking</i> com os concorrentes.</p> <p>- Planejamento de recursos de materiais devido à variedade de produtos fabricados.</p> <p>- Informações regulares sobre a situação dos equipamentos para evitar uma manutenção corretiva.</p> <p>- Rastreabilidade é um fator crítico para saber a localização e quantidade de matérias-primas e estoque em processo e acabado.</p>	<p>- É importante a origem e composição das matérias-primas, principalmente do minério, pois afeta as reações presentes no alto forno.</p> <p>- Com o aumento da variedade de produtos, as informações referentes à etapa produtiva de lingotamento e laminação são críticas para o produto.</p>	<p>- Informações sobre os concorrentes são vitais, pois se trata de uma indústria de alta tecnologia e rápidos ciclos de inovação. Então, informações constantes são necessárias sobre novos <i>softwares</i> especializados e ferramentas de <i>hardware</i>.</p> <p>- Informações sobre pessoas com alta capacidade intelectual e criativa para o <i>design</i> dos semicondutores.</p> <p>- Como a variedade de requisitos formadores do semicondutor é alta, a informação sobre o planejamento de processo a fim de definir a melhor combinação para a fabricação do produto é importante.</p> <p>- Decisão sobre fragmentar ou não a cadeia de valor do semicondutor é vital atualmente, pois os processos em cada etapa são complexos e caros por etapa. A tendência nessa indústria é se especializar em uma etapa na cadeia com pessoal capacitado e bem treinado.</p> <p>- Informações sobre os outros participantes da cadeia de semicondutores é importante, pois esse tipo de indústria está formalizado em redes industriais, necessitando das informações de outras firmas para a sua tomada de decisão local.</p>	<p>- Informações sobre os clientes referentes ao prazo de entrega de pedidos especiais. Geralmente os clientes pedem os produtos em cima da hora, gerando dificuldades para a programação da produção;</p> <p>- Dificuldade em atender aos requisitos especificados pelo cliente nas meias por encomenda, devido as informações virem em cima da hora ou pela falta de bom senso do cliente na especificação do pedido;</p> <p>- <i>Benchmarking</i> com os concorrentes para verificar o desempenho deles.</p>

Fonte: Custodio et al. (2002, p. 7)

Por exemplo, nos quatro setores os dados críticos que estão nos bancos de dados transacionais seriam:

- moldagem de plásticos para injeção: dados sobre o processo produtivo, clientes, concorrentes e rastreabilidade do produto;

- bobinas e chapas de aço: dados sobre especificações das matérias-primas e sobre processos produtivos críticos de lingotamento e laminação;
- semicondutores: dados sobre concorrentes e mercado, recrutamento de pessoal capacitado, *design* do processo, e
- meias feitas por pedidos especiais: dados sobre clientes e concorrentes.

Verifica-se pelo que foi apresentado na Figura 4.4, que cada empresa tem a sua particularidade na aplicação do método devido aos fatores da indústria e ao tipo de modelo de negócios que ela utiliza. Cada empresa possui diferentes dados, que são críticos, 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, será detalhada a busca do conhecimento no SMD e no sistema de informação. A equipe deverá concentrar-se na estrutura do SMD, que, de acordo com que foi mostrado no item 2.2 por Neely (1998), seria coleta, tratamento dos dados, classificação em categorias, busca de padrões nas categorias, esclarecimento das implicações de cada categoria e no todo, e disseminação de dados e informações adequadas.

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 conhecendo-se quais são as medidas de desempenho e como elas se interagem é 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 deve focar 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; e
- os tipos de relacionamentos entre as medidas de desempenho existentes e seus posicionamentos nas áreas de construção desses relacionamentos (ver mais detalhes no item 3.2).

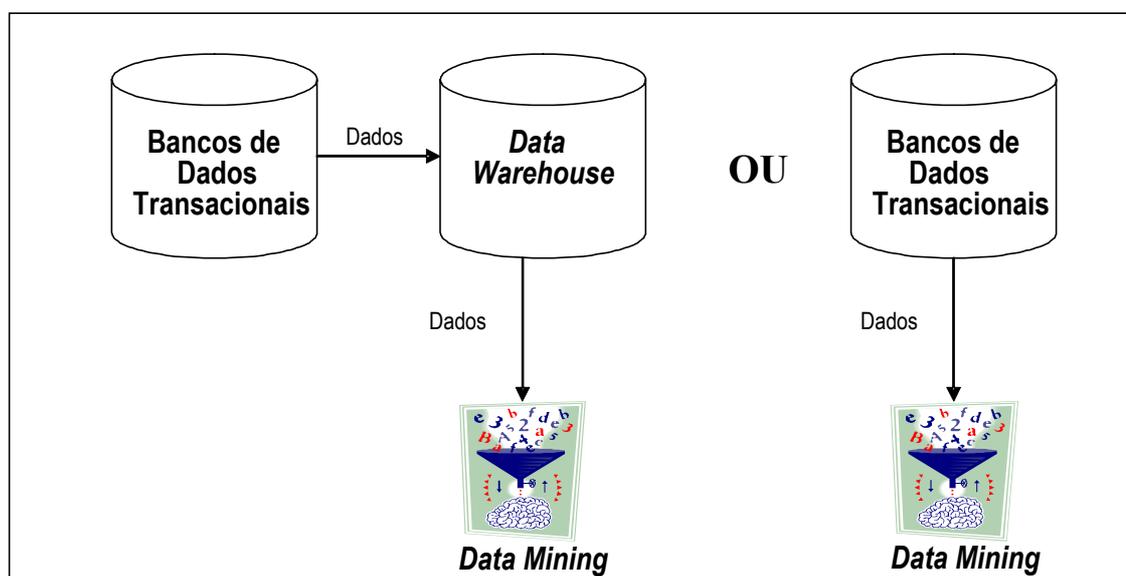
Já acerca do sistema de informação da empresa, é necessário conhecê-lo, principalmente o que se relaciona à decisão existente na empresa, de modo a verificar quais são os bancos de dados transacionais e como eles são utilizados pelos tomadores de decisão. Esse entendimento auxilia na utilização das fontes de dados possíveis para o uso das técnicas e ferramentas de *data mining*.

No entendimento do SAD, os aspectos interessantes a serem analisados para a criação do modelo conceitual são como está estruturado o SAD em termos de pessoas, *hardware* e *softwares* e também como o processo decisório acontece na empresa.

Um aspecto fundamental no apoio à decisão é a definição do banco de dados. Ela estabelece um modelo inicial de *data warehouse*, que pode ser um banco de dados centralizado ou um *data mart*. A utilização de um banco de dados levará em conta o tamanho dele, a quantidade de variáveis presentes para a aplicação do *data mining* e os esquemas de segurança presente no acesso aos dados.

Tais fatores podem influenciar no método, na delimitação dos relacionamentos entre as medidas de desempenho, já que uma maior complicação e complexidade na estrutura dos dados disponível requerem um esforço computacional maior em termos de *hardware*. Dependendo da escolha, isso pode tornar inviável financeiramente o uso da estrutura de determinado banco de dados histórico.

Uma questão importante dentro da definição do banco de dados em uma empresa é a relação entre bancos de dados transacionais e *data warehouse*, conforme ilustra a Figura 4.8. As duas opções são viáveis para o uso das técnicas e ferramentas de *data mining*. Embora neste trabalho seja considerada a opção do uso do *data warehouse*. É importante explicitar as duas opções e mostrar por que neste método é importante o seu uso.



Fonte: Elaborado pelo autor.

FIGURA 4.8 – A escolha do tipo de banco de dados para a aplicação das técnicas e ferramentas de *data mining*

Para mostrar isso, duas questões importantes a serem analisadas são:

1. É necessário uso da *data warehouse*?; e
2. Quais são as diferenças do uso de bancos de dados transacionais e de *data warehouse*?

A primeira questão é sobre a necessidade ou não do uso da *data warehouse* na aplicação do método, já que essa atividade demanda custo e esforço organizacional para realizá-la.

O *data warehouse* armazena dados analíticos, destinados às necessidades dos usuários de diferentes níveis organizacionais no processo de tomada de decisão. Isso pode envolver consultas complexas que necessitem acessar um grande número de registros. Por isso, é importante a existência de muitos índices criados para acessar as informações da maneira mais rápida possível. O *data warehouse* armazena dados históricos e, por isso, precisa ter uma grande capacidade de processamento e armazenamento de dados que se encontram de duas maneiras, detalhados e resumidos (QUISPE, 2003).

O *data warehouse* adiciona valor para os tomadores de decisão, provendo dados sobre questões que envolvem decisões nos níveis tático e estratégico da empresa.

Tais questões precisam estar presentes nas possibilidades de resposta do modelo de *data warehouse*, que permite o uso de ferramentas analíticas, dentre elas o *data mining*, para um determinado conjunto de dados. (SINGH, 2001).

Portanto, o uso do *data warehouse* é recomendado no método proposto, pois os relacionamentos entre as medidas de desempenho estão presentes em diferentes níveis organizacionais na empresa, não ficando restrito ao nível operacional. Por isso, necessita da análise dos tomadores de decisão na resolução de questões complexas e que adicionam valor à empresa.

A segunda questão trata das diferenças do uso entre banco de dados transacionais e *data warehouse*. Um *data warehouse* difere de um banco de dados transacional nos seguintes aspectos:

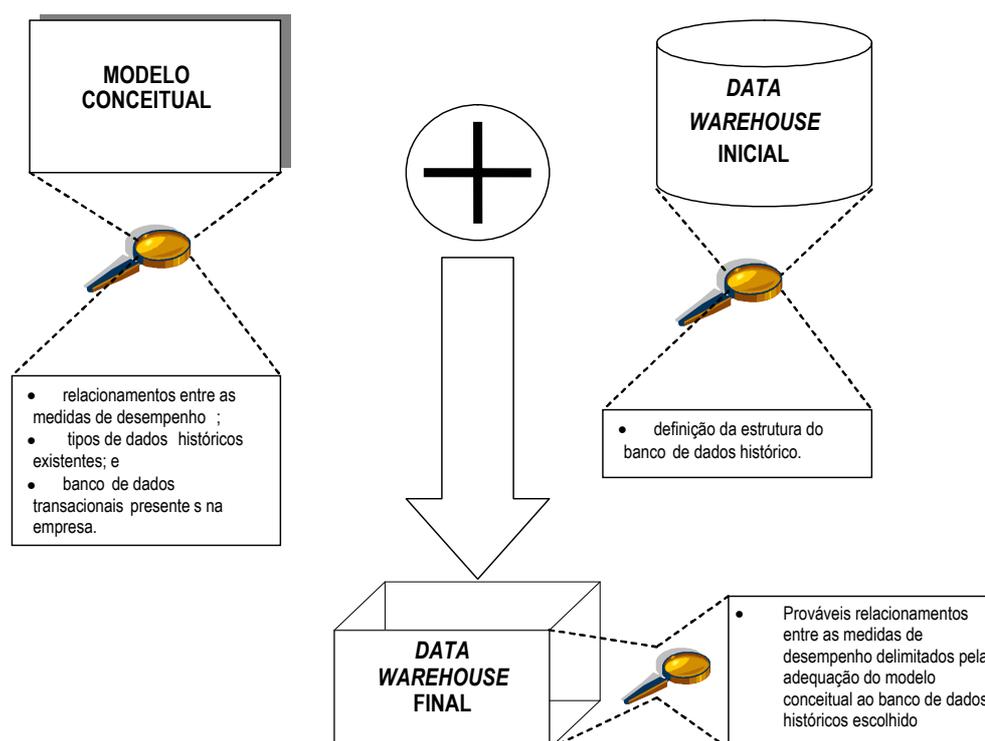
- É acessível aos usuários que possuem conhecimento limitado de sistemas informatizados ou de estruturas de dados devido a uma eficiente gestão dos metadados;
- Oferece suporte à decisão devido à possibilidade de consultas complexas envolvendo problemas não estruturados, o que é necessário numa análise gerencial;
- Disponibiliza dados, dependendo da cultura organizacional da empresa, para todos os usuários que queiram acessar, não ficando restrita a um grupo ou departamento;
- Está associado a um período de tempo definido, por exemplo, o do calendário fiscal; e
- É utilizado para consultas de leitura no expediente, diferentemente dos bancos de dados transacionais, que estão sempre ativos para atualizações.

Nesta proposta, o uso do *data warehouse* em relação aos bancos de dados transacionais se justifica pelo fato de as medidas de desempenho estarem associadas a diferentes aplicações de TI, as quais estão associadas bancos de dados transacionais. O *data warehouse* é modelado no intuito de agregar e uniformizar as diferentes fontes de dados, conseqüentemente facilitando a análise dos relacionamentos entre as medidas de desempenho.

Por fim, de posse das informações da empresa por meio do modelo conceitual e definido o modelo de *data warehouse* inicial, serão delimitados pela equipe quais os tipos de relacionamentos entre as medidas de desempenho serão investigados com a aplicação de *data mining*, respondendo a duas questões:

1. qual(is) critério(s) competitivo(s) ou área(s) de negócio será abrangido na utilização das medidas de desempenho?; e
2. qual(is) tipo(s) de medidas de desempenho conterão o banco de dados de desempenho?

A Figura 4.9 ilustra a formação do modelo de *data warehouse* final, que é formado a partir da análise do modelo conceitual. Tal análise é feita a partir do critério de que se tem um banco de dados histórico definido e da necessidade de informação a ser atendida para os tomadores de decisão.



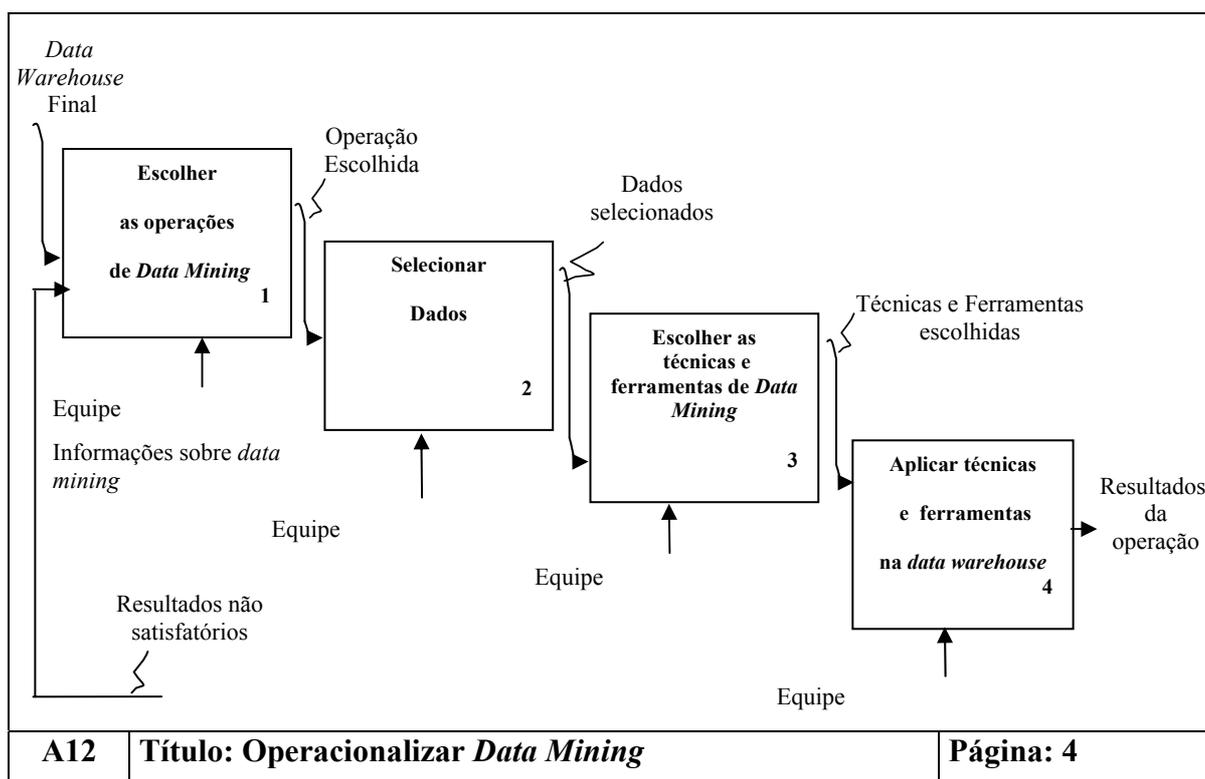
Fonte: Elaborado pelo autor.

FIGURA 4.9 – Formação do modelo de *data warehouse* final

Com as respostas dessas questões, tem-se um modelo de *data warehouse* final que será utilizado na realização das escolhas necessárias para a operacionalização do *data mining*.

4.3.2 Operacionalização do *data mining*

Com o levantamento de informações sobre a empresa em que será aplicado o método, inicia-se o uso do *data mining*. A Figura 4.10 apresenta como será operacionalizado e aplicado o *data mining* no estabelecimento de relacionamentos entre as medidas de desempenho.



Fonte: Elaborado pelo autor.

FIGURA 4.10 – Operacionalizar o *data mining*.

A operacionalização se inicia com a escolha das operações analíticas de *data mining* passíveis de uso de acordo com o tipo de relacionamento que se deseja investigar. A operação analítica a ser utilizada tem também que estar adequada com o modelo de *data warehouse* final, definido na etapa anterior.

A escolha da operação analítica de *data mining* é baseada em três parâmetros presentes no modelo de *data warehouse* final:

- necessidade de informação dos tomadores de decisão;
- características dos relacionamentos entre as medidas de desempenho definidos no modelo conceitual; e
- tipo de banco de dados histórico escolhido.

A operação analítica, conforme mostrado no item 3.2.2, delimita qual a área de atuação do *data mining* no método proposto, podendo ser para a criação de modelos preditivos, agrupamento de base de dados, análise de associações e detecções de desvios.

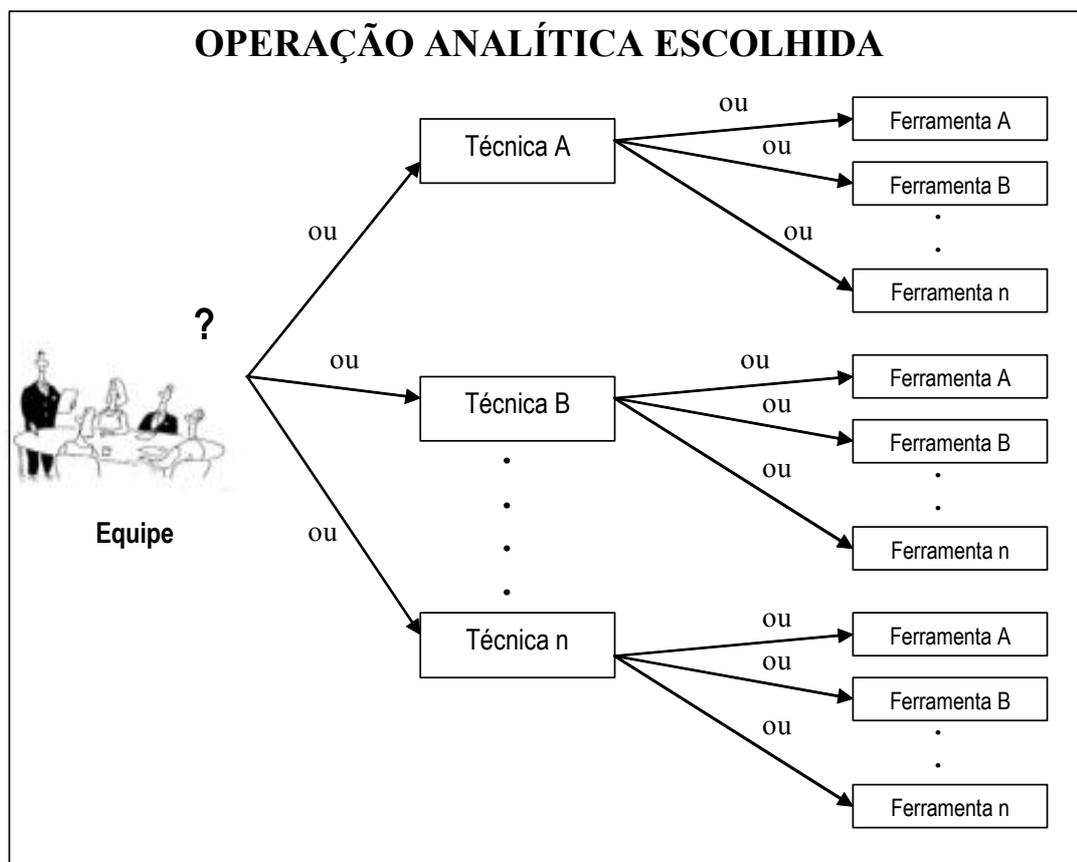
Dependendo da empresa e das necessidades dos tomadores de decisão, poderá haver também duas operações analíticas escolhidas, uma auxiliando a outra. Por exemplo, a operação analítica agrupamento em bancos de dados pode auxiliar na formação de vários *clusters* permitindo que posteriormente a operação de análise de associações atue em um dos *clusters* definidos para a busca de padrões de relacionamento entre as medidas de desempenho.

Posteriormente, o foco migra para o recorte dos dados necessário para a aplicação do *data mining*, no qual os dados serão selecionados e preparados para facilitar o uso da abordagem. Vale ressaltar a importância de verificar se o conjunto de dados não contém dados distorcidos, que provavelmente são registros mal efetuados, ou eventualmente observações anômalas, que, incluídas nas análises, vão distorcer a validade das conclusões (HAND, 1998).

O recorte dos dados tem como critério a delimitação dos relacionamentos entre as medidas de desempenho. Os dados selecionados também permitirão o uso de filtros para que os novos dados provenientes das medidas de desempenho sejam utilizados para a operacionalização do *data mining*.

A partir do recorte dos dados e da escolha da(s) operação(ões) analítica(s), são definidas as técnicas e ferramentas de *data mining*. A Figura 4.11 apresenta a forma como a equipe terá que escolher as técnicas e ferramentas de *data mining*.

Na literatura existem diversas técnicas de *data mining* (maiores detalhes podem ser encontrados no item 3.2.3). As técnicas de *data mining* são selecionadas a partir da operação analítica escolhida. Podem existir várias para cada operação analítica e cabe à equipe que aplicará o método definir qual a melhor e que esteja adequada ao modelo de *data warehouse* final definido na etapa anterior.



Fonte: Elaborado pelo autor.

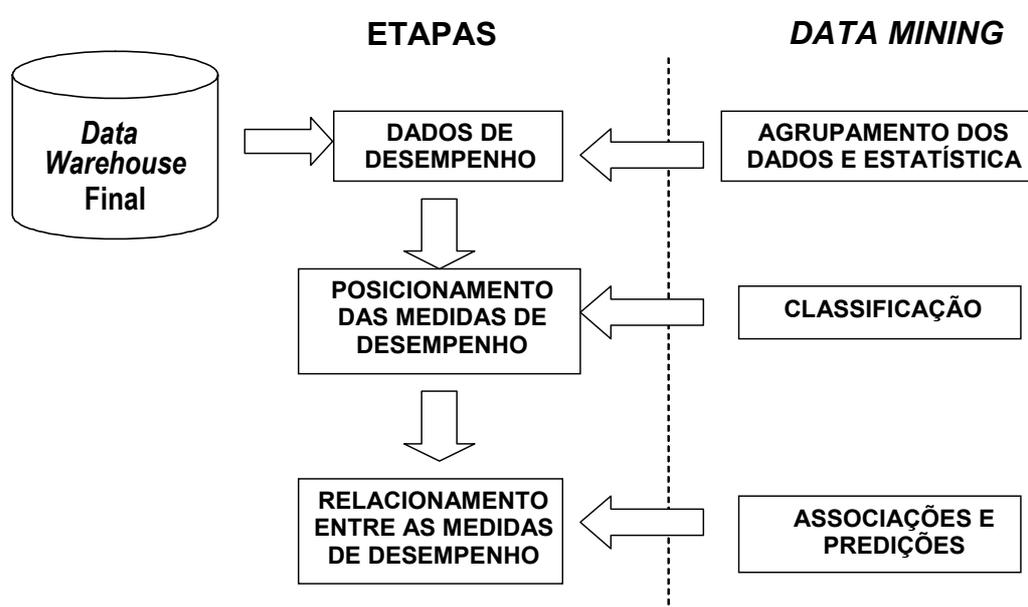
FIGURA 4.11 – Exemplo de escolha das técnicas e ferramentas de *data mining*.

Por exemplo, caso a operação analítica escolhida seja de modelos preditivos, vão existir técnicas a serem escolhidas como classificação e predição de valor. A técnica é escolhida de acordo com os relacionamentos entre as medidas de desempenho delimitados no modelo conceitual, ou seja, a necessidade dos tomadores de decisão no estabelecimento dos relacionamentos entre as medidas de desempenho definirá a técnica.

A escolha das técnicas de *data mining* está presente em três etapas do processo de operacionalização do *data mining* que são:

1. organização dos dados de desempenho no atendimento da necessidade dos tomadores de decisão;
2. verificação de como as medidas de desempenho estão alinhadas e distribuídas na empresa; e
3. definição das melhores formas de criar e analisar associações e predições dos relacionamentos entre as medidas de desempenho.

A partir dessas etapas, a Figura 4.12 mostra um roteiro de uso das técnicas de *data mining* no estabelecimento dos relacionamentos entre as medidas de desempenho. Esse roteiro foi estabelecido a partir da literatura pesquisada, buscando-se focar na questão de quais seriam as melhores técnicas de *data mining* a serem utilizadas no estabelecimento de relacionamentos entre as medidas de desempenho. Com tal roteiro, a equipe pode trabalhar na escolha das ferramentas adequadas para o uso das técnicas.



Fonte: Elaborado pelo autor.

FIGURA 4.12 - Fases para o estabelecimento das técnicas e ferramentas de *data mining* na construção de relacionamentos entre as medidas de desempenho.

Conforme a Figura 4.12, os dados de desempenho obtidos na *data warehouse* final podem passar pelas operações analíticas de *data mining* agrupamento em bancos de dados e detecção de desvios. Às vezes, os dados não estão organizados de acordo com a necessidade dos tomadores de decisão e não permitem uma análise direta dos relacionamentos entre as medidas de desempenho.

Para a operação analítica de agrupamento em bancos de dados, são utilizadas as técnicas de *clusterização* (para maiores informações sobre formação de *clusters* no item 3.2.3). Por meio delas, é possível identificar um conjunto de categorias que permite descrever os dados provenientes da medição de desempenho. A formação de *clusters* permite que sejam identificadas tendências nos dados.

Já para a detecção de desvios, podem ser utilizadas as técnicas de análise estatística, como sumarização e visualização gráfica, a fim de eliminar vieses e dados inválidos que podem prejudicar as atividades posteriores realizadas pelo *data mining*. Dados numéricos sobre o desempenho podem conter valores muito altos ou extremamente pequenos que fogem da variação normal da maioria dos dados apresentados. Essa operação analítica pode eliminar essas distorções.

Para ambos os casos, o uso dessas operações analíticas é opcional para a equipe que aplicará o método, pois se trata de uma preparação dos dados para o estabelecimento dos relacionamentos entre as medidas de desempenho. Às vezes, a operacionalização da *data mining* não necessita dessa preparação em face do tamanho e da quantidade de variáveis presentes no trabalho

A segunda etapa é o posicionamento das medidas de desempenho frente a um sistema classificatório estabelecido pelos tomadores de decisão. Os membros da organização criam um sistema classificatório podendo ser baseados em diferentes critérios como áreas organizacionais, demanda dos *stakeholders*, critérios competitivos, entre outros.

O estabelecimento de critérios classificatórios permite o uso da operação analítica de modelagem preditiva para direcionar as medidas de desempenho. O posicionamento dos dados provenientes da medição de desempenho é útil para estabelecer os relacionamentos entre as medidas de desempenho, pois esse posicionamento se aproxima do que realmente acontece no uso delas no sistema de gestão da empresa.

Esta etapa também é opcional, pois às vezes a equipe pode encontrar os dados em um sistema já satisfatório para eles. O cubo de dados presente no modelo de *data warehouse* final divide-os em dimensões que já permitem aos tomadores de decisão verificarem como as medidas de desempenho estão distribuídas nas áreas organizacionais e/ou alinhadas com os processos de negócios e estratégias.

Por fim, a terceira etapa, que é o foco central do método, consiste no uso da operação analítica descoberta de associações e/ou modelagem preditiva para o entendimento dos relacionamentos entre as medidas de desempenho.

O estabelecimento dos relacionamentos entre as medidas de desempenho consiste em três aspectos que são:

- o estabelecimento de relacionamentos entre as medidas de desempenho não percebidos no modelo conceitual;
- a formalização dos relacionamentos entre as medidas de desempenhos já existentes; e
- a comprovação dos relacionamentos detectados no modelo conceitual.

Para a operação analítica de descoberta de associações, existem diversas técnicas de associações em regras lógicas definidas pelos próprios decisores na análise de eventos organizacionais no mesmo ou em diferentes períodos de tempo.

Essas associações permitem a comprovação dos relacionamentos entre as medidas de desempenho já que existem tipos de relacionamentos, como os de causa-e-efeito, que precisam ser comprovados a partir de cálculos matemáticos, e essas técnicas permitem tal comprovação. As técnicas de associações também auxiliam na descoberta de novos relacionamentos entre as medidas de desempenho já que, por exemplo, nas relações de causa-e-efeito há efeitos que levam uma grande quantidade de tempo para serem detectados e que, às vezes, não são percebidos pelos tomadores de decisão.

Já a operação analítica de modelos preditivos pode ser aplicada nesta etapa a partir das técnicas de predição de valor. Ela pode ser utilizada pelos tomadores de decisão como uma forma de formalizar os relacionamentos entre as medidas de desempenho. Por exemplo, em uma organização os tomadores de decisão percebem que a medida de desempenho “giro de estoque” se relaciona por causa-e-efeito com a medida de desempenho “produtividade”, mas eles não sabem a intensidade do relacionamento. As técnicas de predição de valor podem criar uma equação matemática a partir dessas duas medidas de desempenho podendo estabelecer uma relação matemática entre eles, como também determinar qual o nível ótimo de giro de estoque para certos níveis de produtividade.

Para cada técnica escolhida para o uso das operações analíticas será necessária uma ferramenta de *data mining*. Como mostrado na Figura 4.11, existem diversas ferramentas para cada técnica. A equipe deve estabelecer critérios de escolha para a obtenção da melhor. Os critérios a serem aplicados seriam:

1. adequação das ferramentas de *data mining* à estrutura física (*hardware*) do sistema de apoio à decisão da empresa em que vai ser aplicado o método;

2. adequação das ferramentas de *data mining* ao banco de dados histórico; e
3. análise de custo presente na aquisição e manutenção da ferramenta.

O primeiro critério se refere à estrutura física da empresa em termos de *hardware*. É importante verificar no modelo de *data warehouse* final, os equipamentos da empresa em que será aplicado o método proposto, pois existem ferramentas que exigem grande capacidade computacional em relação a outras. Por exemplo, bancos de dados centralizados exigem maiores recursos em termos de armazenamento do que os *data marts*.

O segundo critério se refere à questão do *software* (a lista de alguns pode ser encontrada no Apêndice B) ou algoritmos usados pelas ferramentas de *data mining*. É necessário analisar se o *software* ou algoritmo é compatível com o banco de dados histórico da empresa em termos de linguagem de programação e dos fabricantes envolvidos. Pode ocorrer que o fabricante do banco de dados históricos não tenha suporte para uma aplicação de *data mining* útil para a empresa. Uma opção para tal dificuldade é a criação de algoritmos compatíveis com a linguagem do banco de dados histórico, o que vai requerer um esforço a mais na utilização no método proposto.

Já o terceiro critério se refere aos custos envolvidos na obtenção das ferramentas. Dois tipos de custos são importantes: o custo de aquisição; e o de manutenção. A escolha de criação de algoritmos em vez da aquisição de *software* é uma opção mais interessante quando a empresa possui um considerável capital intelectual. Senão, o uso de *software* é a opção mais interessante. Dentre as opções de *softwares*, existem os comerciais e públicos. Este último possui custo de aquisição muito baixo, mas geralmente há poucas opções no mercado que atendam às necessidades da empresa.

Tendo estruturado o uso do *data mining*, a aplicação da técnica escolhida é realizada a partir do *software* ou algoritmo escolhido de uma determinada ferramenta de *data mining*. A aplicação deve consistir em um teste-piloto baseado em aprendizado supervisionado realizado pela equipe da qual os resultados são previamente conhecidos. Tendo sucesso nessa atividade, será realizada a aplicação em si, que pode ser feita de forma supervisionada ou não.

4.3.3 Uso dos resultados obtidos

Na última etapa serão apresentados os resultados transacionais obtidos da aplicação de *data mining*, conforme mostra a Figura 4.13. Os resultados serão utilizados para a análise como também para a divulgação dos relacionamentos estabelecidos ou descobertos entre as medidas de desempenho.

A análise dos resultados consiste em verificar se eles são satisfatórios ou não. Resultados satisfatórios são as informações obtidas pelo uso do *data mining* que são úteis para o estabelecimento e verificação dos relacionamentos entre as medidas de desempenho.

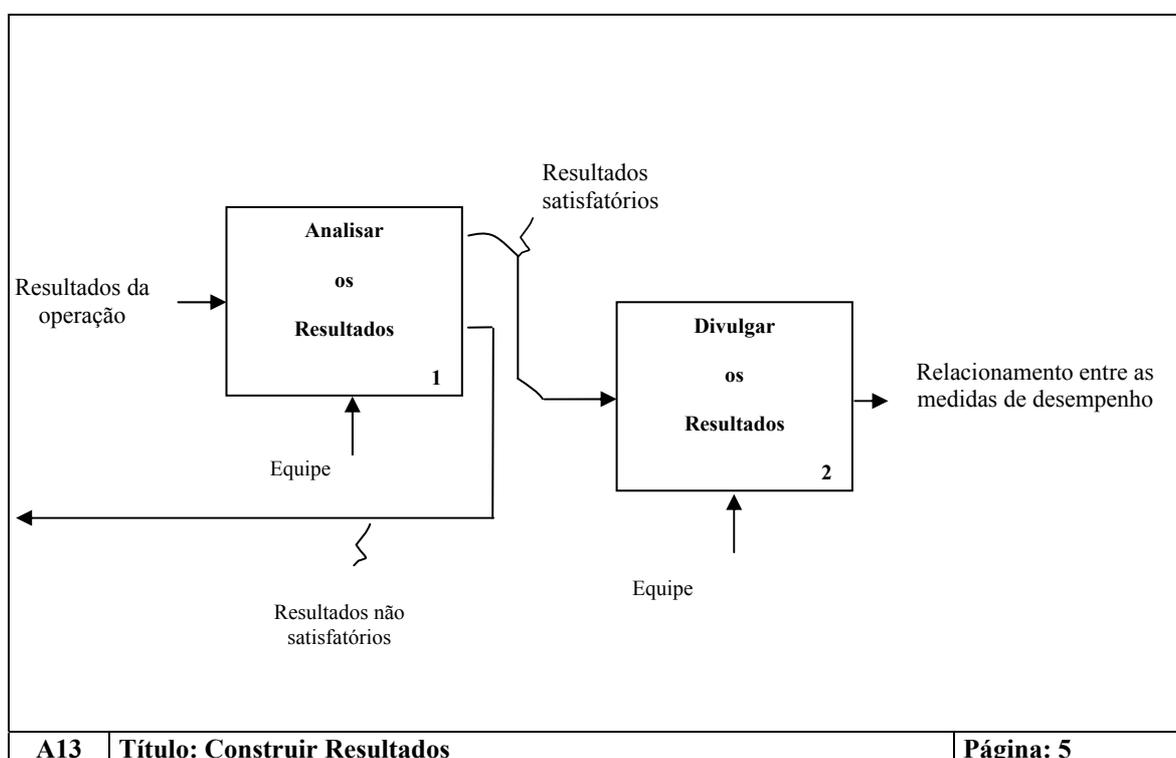


FIGURA 4.13 – Construção dos resultados.

Os resultados satisfatórios têm que atender três critérios para satisfazer os tomadores de decisão:

- formalização dos relacionamentos entre as medidas de desempenho;
- comprovação dos relacionamentos entre as medidas de desempenho existentes; e
- descoberta de novos relacionamentos entre as medidas de desempenho.

Caso não haja um resultado satisfatório, o processo retornará à etapa de operacionalização de *data mining*, como mostra a Figura 4.10, para a realização de uma nova análise. Esse processo iterativo continuará até que a equipe esteja satisfeita com os resultados apresentados e, depois disso, os resultados serão disseminados para a organização, visando o entendimento pelos tomadores de decisão dos relacionamentos entre as medidas de desempenho obtidas pelo método proposto.

A divulgação dos resultados da aplicação do método na empresa pode ser feita internamente para atender aos tomadores de decisão, como também externamente para atender alguns *stakeholders* como os acionistas. Isso dependerá de quais questões são importantes para os tomadores de decisão no entendimento dos relacionamentos entre as medidas de desempenho.

5. CONCLUSÕES

Este trabalho teve como objetivo apresentar um método para o estabelecimento dos relacionamentos entre as medidas de desempenho a partir do uso do *data mining*. Ele consiste em utilizar o processo de *data mining* num banco de dados históricos provenientes da medição de desempenho a fim de poder estabelecer os relacionamentos entre as medidas de desempenho já existentes e descobrir novos relacionamentos entre elas.

O desenvolvimento deste trabalho teve como origem o fato de o sistema de gestão da empresa e, conseqüentemente o SMD, ter sofrido modificações em face de mudanças no ambiente industrial nas últimas décadas. Uma das modificações do SMD foi o fato de a medição de desempenho ter se tornado multidimensional e não mais somente estar voltada para a medição de desempenho nos aspectos financeiros e de produtividade. Por isso, se torna mais complicado e complexo o entendimento de como os relacionamentos entre as medidas de desempenho afetam as decisões organizacionais.

Outro fator relevante para este trabalho é o fato de a TI ter avançado bastante nas últimas décadas em suas aplicações para os processos de negócio da empresa, mas na medição de desempenho pouco foi utilizado nas etapas de análise e disseminação das informações sobre o desempenho.

Dentro dessas etapas nos SMD, duas abordagens de TI se destacam: o *data warehouse* e o *data mining*. Ambas permitem que o método para o estabelecimento dos relacionamentos entre as medidas de desempenho seja aplicado com base na capacidade de armazenamento, análise e disseminação de dados históricos provenientes da medição de desempenho.

O *data warehouse* tem o papel de repositório de dados de desempenho no método proposto. Esse banco de dados histórico é modelado tendo como critérios o tamanho do banco de dados, o número de variáveis presentes na análise e o esquema de segurança para os usuários. O *data warehouse* é eficaz para relacionamentos entre as medidas de desempenho que envolvem diferentes áreas e departamentos da organização, e problemas complexos e complicados, geralmente não estruturados, que adicionam valor a empresa.

Já o *data mining* é uma abordagem que propicia para o método proposto um conjunto de técnicas e ferramentas divididas em quatro operações analíticas, que são os modelos preditivos, agrupamentos em bancos de dados, análise de associações e detecção de desvios.

Dentre as operações analíticas, destaca-se a análise de associações por apresentarem um conjunto de técnicas que relacionam as medidas de desempenho em um mesmo evento ou em eventos temporais diferentes. Quaisquer necessidades dos tomadores de decisão, no método apresentado, envolvem esse tipo de operação analítica.

A operacionalização do *data mining* no método é um processo que depende do modelo de negócios da empresa e de como as estratégias derivadas do modelo de negócio afetam os processos da empresa. O uso do *data mining* envolve escolhas das operações analíticas, das técnicas e das ferramentas de *data mining* e busca atender às necessidades dos tomadores de decisão em termos do estabelecimento dos relacionamentos entre as medidas de desempenho.

Portanto, a proposta desta dissertação procura englobar a questão dos relacionamentos entre as medidas de desempenho com o uso das abordagens de TI como o *data warehouse* e *data mining*, a partir da apresentação de um método de estabelecimento dos relacionamentos entre as medidas de desempenho.

A principal contribuição deste trabalho é proporcionar um método para que o relacionamento entre as medidas de desempenho possa ser estabelecido a partir de uma grande quantidade de dados históricos. O método poderá tanto servir de base para a construção de hipóteses e condução de novas pesquisas como para ajudar as empresas na adoção e utilização da sistemática construída do modelo validado de relacionamentos das medidas de desempenho, já que um dos aspectos importantes presentes no método é a criação de um modelo conceitual.

Para a medição de desempenho, o principal ganho deste trabalho é que o método, ao ser aplicado, pode facilitar o uso da comunicação entre diferentes áreas da empresa pelo fato de permitir o entendimento de como as medidas de desempenho aplicadas em suas áreas podem se refletir em toda a organização.

Dada a dificuldade de se encontrar uma organização que permitisse a aplicação do método proposto, devido a desencontros de interesses em relação à

organização convidada para o projeto, algumas etapas não puderam ser detalhadas totalmente. Outras não o foram também porque dependem única e exclusivamente das contingências a serem encontradas em cada aplicação. Contudo, a proposta apresentada permite encaminhar minimamente o uso do *data mining* como forma de estabelecer relacionamentos entre as medidas de desempenho.

Como propostas futuras de continuidade deste trabalho, tem-se principalmente que o método seja aplicado em uma organização como forma de validar preliminarmente o método proposto. Outras possibilidades de desenvolvimento de pesquisa são:

- a. estudar os efeitos da aplicação do método proposto na melhoria da utilização da informação sobre desempenho nas atividades de controle e melhoria de processos de negócio;
- b. discutir sobre quais os principais efeitos que a tecnologia de informação pode ter em um sistema de medição de desempenho;
- c. aplicar um sistema especialista voltado para a predição de relacionamentos entre as medidas de desempenho; e
- d. investigar sobre a utilização do método proposto diretamente em bancos de dados transacionais.

ANEXO A

INTEGRATION DEFINITION FOR FUNCTION MODELING (IDEF0)

A prática IDEF0 usa uma notação para a representação gráfica que mostra o processo e suas atividades componentes por meio do fluxo de informações existentes entre funções. Estas funções são mapeadas, possibilitando uma visão gradativamente detalhada do processo. Esse detalhamento é feito para cada função ou atividade, por meio de sucessivas explosões das funções.

A representação usual do IDEF0 é mostrado na Figura A1.

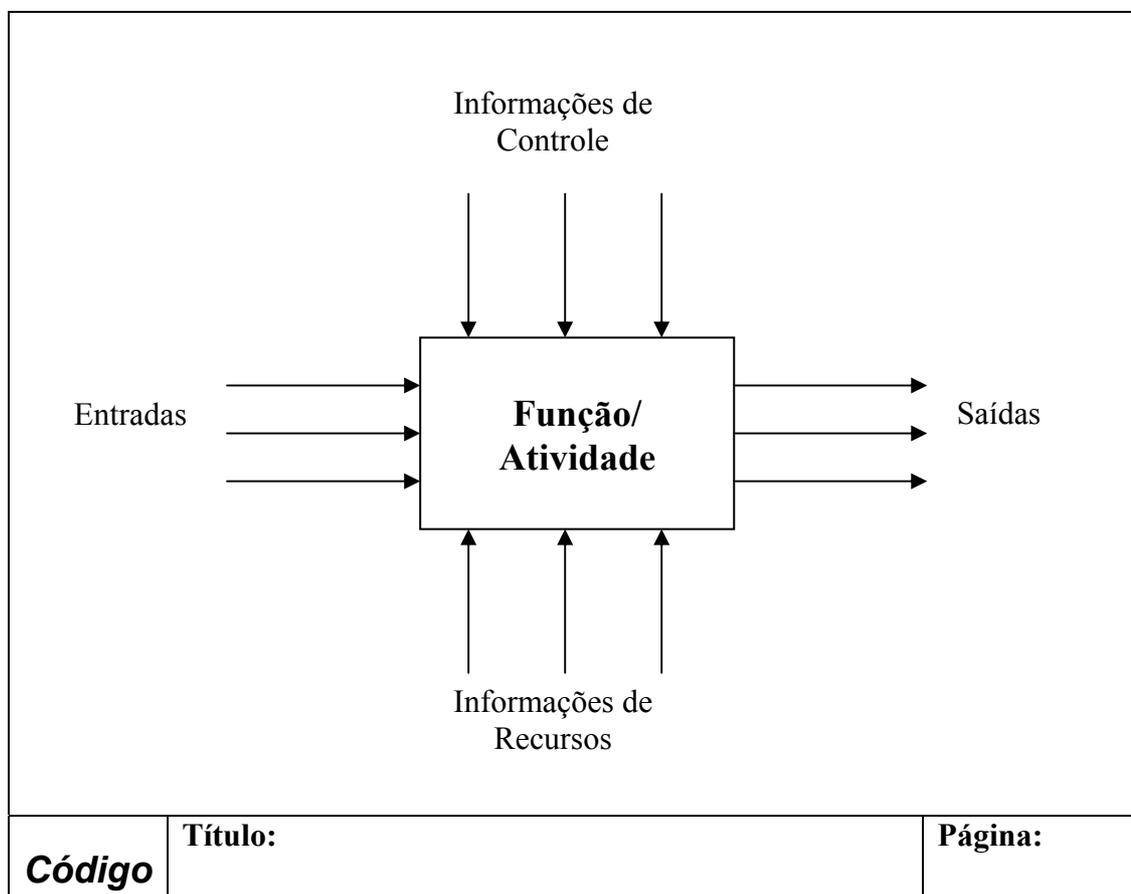


FIGURA A1 – Elementos Básicos Presentes no IDEF0.

Fonte: Draft Federal Information (1993), p.11.

O detalhamento das funções atividades é feito no conceito de diagrama pais e filhos, em que uma função/ atividade é explodida em um novo diagrama contendo relação com o diagrama pai. Uma representação de diagrama pais e filhos para o IDEF0 pode ser encontrada na Figura A2.

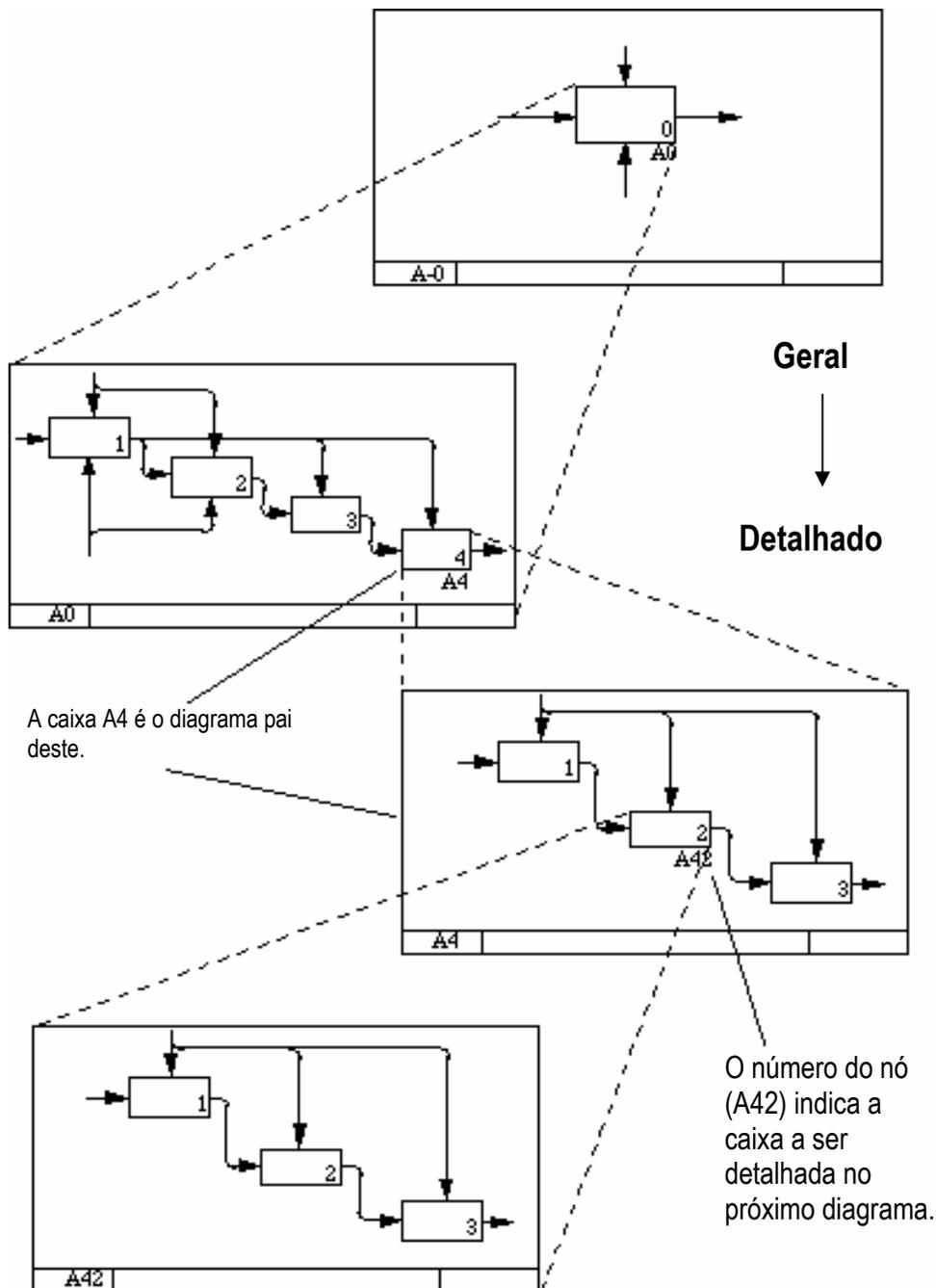


FIGURA A2 – Decomposição da estrutura com base no conceito diagramas pais e filhos.

Fonte: Draft Federal Information (1993), p.16.

O grau de decomposição da estrutura é definido pelo usuário da metodologia que determina qual o nível de detalhamento satisfatório para o seu processo de trabalho.

Convém destacar alguns elementos apresentados no item 4.3, que estão definidos na Figura A3.

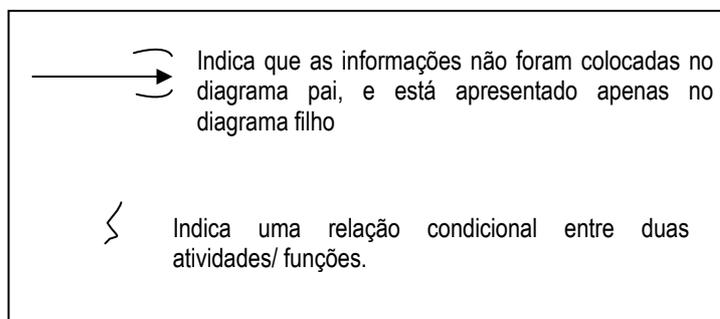


FIGURA A3 – Descrição de alguns elementos presentes na metodologia IDEF0

Fonte: Draft Federal Information (1993).

O IDEF0 pode ser aplicado em diversas áreas de negócios. Ele é prático no sentido de descrever, de modo detalhado, as decisões, ações e atividades de um sistema.

APÊNDICE A

LISTA DE *SOFTWARES*

Esta lista foi baseada nas informações do site <http://www.mestradoinfo.ucb.br/aluno/esilva/datamining/index.htm#resumo> e atualizada na primeira semana de junho de 2004. Também foram inseridas as informações dos *softwares* públicos Weka e Winindams.

Baseada em Casos	
Software	Empresa – Endereço Eletrônico
ALICE d'ISoft	ISoft - http://www.alice-soft.com/html/prod_alice.htm
CASE Advisor & Case Advisor Webserver	Sententia Software Inc.: - http://www.cs.sfu.ca/~isa/isaresearch.html#systems
HelpDesk-3	TreeTools Informatica Ltda: - http://www.treetools.com.br/tti/
Induce-It	Inductive Solutions, Inc. - http://www.inductive.com/softcase.htm
Intellix Designer 5.0	Intellix Denmark ApS: - http://www.intellix.com/products/designer/designer.html
Kaidara	Kaidara Advisor - http://www.kaidara.com/products/advisor.htm
Spotlight	Case Bank Support Systems Inc.: - http://www.casebank.com/products/spotlight.asp
Geração de Conhecimento	
Software	Empresa – Endereço Eletrônico
ALICE d'ISoft	ISoft - http://www.alice-soft.com/html/prod_alice.htm
AnswerTree	SPSS - http://www.spss.com/answertree/
Ascential Enterprise Integration Suite	Ascential Software - http://asi.ascential.com/products/eis.html
BusinessMiner	Business Objects, Inc. - http://www.businessobjects.com/
Clementine	SPSS - http://www.spss.com/clementine/
Cognos PowerPlay®	Cognos Corporation - http://www.cognos.com/products/businessintelligence/analysis/
Contiguous Connection Model (CCM)	Applied Technical Systems - http://www.apptechsys.com/ProductsWhatisCCM.asp
DataDetective	Sentient Information Systems - http://www.sentient.nl/
DataEngine	MIT GmbH - http://www.dataengine.de/english/sp/index.htm
DataMite	Logic Programming Associates - http://www.lpa.co.uk/ind_pro.htm
DataX	Zaptron - http://www.zaptron.com/intro.htm#datax
DB2 Intelligent Miner	IBM – http://www-306.ibm.com/software/data/iminer/
Decision Force	Pro-Action - http://www.proaction.gr/datamining.htm
Enterprise Miner	SAS – http://www.sas.com/technologies/analytics/datamining/miner/index.html
IManageData	BioComp Systems - http://www.biocompsystems.com/products/imanagedata/index.htm
Insight 5	Datanautics - http://www.datanautics.com/products/insight/
K.wiz	ThinkAnalytics- http://www.thinkanalytics.com/productServices/kwiz/index.htm
KnowledgeMiner 5.0	Script Software - http://www.knowledgeminer.net/
LexiQuest	SPSS - http://www.spss.com/lexiquest/lexiquest_mine.htm
Magnify	Magnify, Inc. - http://www.magnify.com/technology.shtml#

Mathsoft Enterprise Framework	MathSoft™ - http://www.mathsoft.com/solutions/framework/0,,0,00.html
Mining Manager®	Angoss Software Corp. - http://www.angoss.com/Products/indexp.html
Nuggets	Data Mining Technologies, Inc. - http://www.data-mine.com/bin/site/templates/default.asp?norelay_hier_id_155A25150=2&area_2=content%2Dproducts&NC=182X
Oracle Database 10g Enterprise	Oracle - http://otn.oracle.com/products/bi/odm/index.html
Pilot® BusinessAnalyzer	Pilot Software, Inc. - http://www.pilotsoftware.com/pba/
See5 e C5.0	RuleQuest Research - http://www.rulequest.com/see5-info.html
SuperQuery	Azmy Thinkware Inc. - http://www.azmy.com/
Teradata Warehouse Miner	NCR - http://www.teradata.com/t/go.aspx?id=94277
VisiRex 2.1	CorMac Technologies - http://cormactech.com/visirex/index.html
Weka 3	The University of Waikato - http://www.cs.waikato.ac.nz/ml/weka/
Winimdams	Unesco - www.unesco.org/webworld/idams/
WizWhy	WizSoft, Inc. - http://www.wizsoft.com/default.asp?win=7&winsub=7
XpertRule® Miner	Attar Software - http://www.attar.com/pages/info_xm.htm
Zoom 'n View	SkyGate International A/S - http://www.skygate.dk/productdesc.php?id=14
Lógica Fuzzy	
Software	Empresa – Endereço Eletrônico
Fide™	Zaptron - http://www.aptronix.com/fide/
FuzzyDecisionDesk	Fuzzy Logik Systeme GmbH - http://www.fuzzy.de/
FuzzyQuery	Sonalysts - http://www.sonalysts.com/fuzzy.htm
FuzzyTECH for Business	Inform Software - http://www.fuzzytech.com/
Redes Neurais	
Software	Empresa – Endereço Eletrônico
Braincel	Promised Land Technologies - http://promland.com/
BrainMaker	California Scientific Software - http://www.calsci.com/
Minotaur	Neural Technologies - http://www.neuralt.com/products/minotaur.htm
Neuralyst	Cheshire Engineering Corporation - http://www.cheshireeng.com/Neuralyst/index.html
NeuroShell	Ward Systems Group, Inc - http://www.wardsystems.com/products.asp
Visualização dos Dados	
Software	Empresa – Endereço Eletrônico
Advisor Solutions Inc	ADVIZOR Solutions Inc. - http://www.advizorsolutions.com/products/software_products.asp
Cognos Visualizer	Cognos Corporation - http://www.cognos.com/products/businessintelligence/visualization_dashboard/index.html?sblid=mnutop_products
Cygron DataScope	Mindmaker - http://www.cygron.com/
Data Desk 6 & Data Desk/XL	Data Description - http://www.datadesk.com/
FYI Visual™	FYI Corporation - http://www.fyicorporation.com/products.php
Heatmaps	NeoVision Hypersystems - http://www.neovision.com/
iBase	i2 Inc - http://www.i2inc.com/Products/iBase/default.asp
IDL	Research Systems, Inc. - http://www.rsinc.com/idl/index.asp
Inxight VizServer™	Inxight Software - http://www.inxight.com/products/vizserver/
Mathsoft Enterprise Framework	MathSoft™ - http://www.mathsoft.com/solutions/framework/0,,0,00.html

MM-Data Mining	<u>Maxus Systems International</u> - http://www.maxusystems.com/datamining.html
Modeling	<u>Alterian</u> - http://www.alterian.com/Standard_1.asp?category=3&id=15499
Net Charts Report Suite 4.5	<u>Visual Mining Inc</u> - http://www.visualmining.com/products/index.html
NetMap	<u>NetMap Analytics</u> - http://www.altaanalytics.com/
OpenViz	Advanced Visual Systems - http://www.avis.com/software/soft_b/openviz/index.html
PV-WAVE	Visual Numerics, Inc. - http://www.vni.com/products/wave/index.html#sharingw.vni.com/
SAS/SPECTRAVIEW	SAS - http://www.sas.com/products/spectraview/index.html
Spotfire DecisionSite	<u>Spotfire</u> - http://www.spotfire.com/products/decision.asp
VantagePoint	VI/Visualize, Inc. - http://www.visualizetech.com/products
Visualizer Workstation	Computer Science Innovations, Inc. - http://picard.csihq.com/CSI/technologies.shtml
VisualMine	Artificial Intelligence Software SpA - http://www.visualmine.com/
NeuroSolutions	NeuroDimension, Inc. - http://www.nd.com/

REFERÊNCIAS BIBLIOGRÁFICAS

- ABREU, A.L.T.; MARTINS, R.A. Construção de relacionamentos entre as medidas de desempenho: uma análise da literatura. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 13, 2003. **Anais ...** Ouro Preto - MG: ABEPRO/USP, 8p., 2003.
- BASHEIN, B.J.; MARKUS, M.L. Data warehouses: more than just mining. **Executive Report-Financial Executive Resources Foundation**, v.7, n.2, p.1-8, jan. 2000.
- BERTRAND, J.W.M; FRANSOO, J.C. Modelling and Simulation – Operations Management Research Methodologies Using Quantitative Modeling. **International Journal of Operations & Production Management**, v.22, n.2, p.241–264, 2002.
- BISPO, C.A.F. **Uma análise da nova geração de sistemas de apoio à decisão**. Dissertação (Mestrado). Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 1998.
- BITITCI, U. S.; CARRIE, A. S.; McDEVITT, L. Integrated performance measurement systems: a development guide. **International Journal of Operations & Production Management**, v.17, n.5, p.522-534, 1997.
- BLENKINSOP, S.A.; BURNS, N. Performance measurement revisited. **International Journal of Operations & Production Management**, v.12, n.10, p.16-25, 1992.
- BOYD, L. H.; COX, III, J. F. A cause-and-effect approach to analyzing performance measures. **Production and Inventory Management Journal**, v.38, n.3, p.25-32, 3rd quarter 1997.
- BRYMAN, A. **Research methods and organization studies**. London: Uniwin Hyman, 1989.
- BROWN, M.G. **Keeping score – using the right metrics to drive world-class performance**. New York, Quality Resources, 1996.
- CABENA, P.; HADJINIAN, P.; STADLER, R.; ZANASI, A. **Discovering Data Mining: From Concept to Implementation**. New Jersey: Prentice Hall PTR, 198p., 1997.
- COOPER, W.W.; SEIFORD, L.M.; ZHU, J. Data envelopment analysis: History, Models and Interpretations, In: ____ **Handbook on Data Envelopment Analysis**, eds W.W. Cooper, L.M. Seiford and J. Zhu, Chapter 1, p. 1-39, Kluwer Academic Publishers, Boston, 2004.
- CORREA, H.L.; GIANESI, N.; IRINEU, G. **Just in Time, MRP II e OPT: um enfoque estratégico**. São Paulo: Atlas, 1993.

- COSTA, L.C.; KLOSTER, N.J.A; GRAEML,K.S.; MAZIA, R.F. Data mining ... ou procurar uma pepita de ouro em uma poça de lama. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 10., 2000, São Paulo. **Anais ...** São Paulo: ABEPRO/USP, 8p., 2000.
- CROSS, K. F.; LYNCH, R. L. Managing the corporate warriors. **Quality Progress**, v.23, n.4, p.54-59, apr. 1990.
- CUSTODIO, F. A.; SANTOS, A. C.; MARTINS, R. A. Configuração da Modelagem de Data Warehouse Sob a Ótica dos Tipos de Sistema de Produção In: IX Simpósio de Engenharia de Produção - SIMPEP, 9., 2002, Bauru. **Anais do IX SIMPEP**. 2002.
- DAVIS, A.; O'DONNELL, J. Modeling complex problems: system dynamics and performance measurement. **Management Accounting**, p. 18-20, maio 1997.
- DRAFT FEDERAL INFORMATION. Integration Definition for Function Modeling (IDEF0). **Processing Standards Publication 183**. December, 1993.
- ECCLES, R.G.; PYBURN, P.J. Creating a comprehensive system to measure performance. **Management Accounting**, v.74, n.4, p.41-44, 1992.
- ECCLES, R.G. The performance measurement manifesto. **Harvard Business Review**, v.69, n.1, p.131-137, 1991.
- EVANS, H.; ASHWORTH, G.; GOOCH, J.; DAVIES, R. Who needs performance management? **Management Accounting – CMA**, v.74, n.11, p.20-25, 1996.
- FAYYAD, U.; MADIGAN, D.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases, **AI Magazine**, v.17, n.3, p. 37-54, 1996.
- FAWCET, J.; DOWNS.F. The relationship of theory and research. Norwalk, CT: **Appleton Century Crofts**, 1986.
- FEELDERS, A.; DANIELS, H.; HOLSHEIMER, M. Methodological and practical aspects of data mining. **Information & Management**, v. 37, p.271-281, 2000.
- FELIPE, J. C. **Extração de conhecimento sobre metadados de aplicações multimídia em um ambiente multidimensional orientado a objetos**. Dissertação (Mestrado). Programa de Pós-Graduação em Ciência da Computação, Universidade Federal de São Carlos, São Carlos, 2000.
- FERRARI, F.M.; TOLEDO, J.C. Proposição e aplicação de um modelo para análise da gestão do conhecimento no processo de desenvolvimento de produto. In: CONGRESSO BRASILEIRO DE GESTÃO DE DESENVOLVIMENTO DE PRODUTO, 3.,2001. **Anais...** Florianópolis, SC, 8p., 25-27 Setembro de 2001.

- FERRAZ, J.C.; KUPFER, D.; HAGUENAUER, L. Competitividade, Padrões de concorrência e fatores determinantes. In: ___ **Made In Brazil: desafios competitivos para a indústria**. Rio de Janeiro: Campus, cap.1, p. 1-53, 1995.
- FERREIRA, A.B.H. **Novo Dicionário Aurélio - Século XXI**. Disponível em: <http://www.uol.com.br/aurelio/>. Acesso em: 13 de Maio de 2004.
- FLAPPER, S. D. P.; FORTUIN, L.; STOOP, P. P. M. Towards consistent performance management systems. **International Journal of Operations & Production Management**, v.16, n.7, p.27-37, 1996.
- FORSMAN, S. **OLAP Council White Paper**. Disponível em: <http://www.bi3.net>. Acesso em: 08 de abril de 2004.
- FREIRE, J. E. **Uma abordagem sobre os colaboradores na atual sociedade da informação**. Dissertação (Mestrado). Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2000.
- FREITAS JÚNIOR, O.G.; COSTA, E.B.; COSTA, M.P.S. Um Sistema de Informação Gerencial para uma Companhia de Bebidas baseado na Tecnologia Data Warehouse. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 10., 2000, São Paulo. **Anais ...** São Paulo: ABEPRO/USP, 8p., 2000.
- GARDNER, S.R. Building the Data Warehouse. **Communications of the ACM**. v. 41, n. 9, p. 52-60, Setembro 1998.
- GHALAYINI, A.M.; NOBLE, J.S.; CROWE, T.J. An integrated dynamic performance measurement system for improving manufacturing competitiveness. **International Journal of Production Economics**, v.48, n.3, p.207-225, 1997.
- GHALAYINI, A. M.; NOBLE, J. S. The changing basis of performance measurement. **International Journal of Operations & Production Management**, v.16, n.8, p.63-80, 1996.
- HALEY, B. J.; WATSON; H. J. Managerial Considerations. **Communications of the ACM**, Estados Unidos, v. 41, n.9, p. 32-37, set. 1998.
- HAND, D.J. Data mining: statistics and more? **The American Statistician**, vol.52, n.2, p.112-118, 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, p.59-64, 1997.

- HOFFMAN, K.; KAPLINSKY, R. Driving force: the global restructuring of technology, labour, and investment in the automobile and components industries. In: _____. **The point of transition – from manufacture to systemofacture**. London: Westview, p.31-71, 1988.
- JACOB, V.S.; SEN, A. Industrial–strength data warehousing. **Communications of the ACM**, v.41, n.9, p. 29-31, set. 1998.
- JOHNSON, H.T.; KAPLAN, R. S. **Contabilidade gerencial - a restauração da relevância da contabilidade nas empresas**. Rio de Janeiro : Campus, 239p., 1993.
- JONES, K. An introduction to data warehousing: what are the implications for the network? **International Journal of Network Management**, v.8, p. 42-56, 1998.
- ITTNER, C. D.; LARCKER, D. F. Innovations in performance measurement: trends and research implications, **Journal of Management Accounting Research**, 10, p.205-239, 1998.
- KAPLAN, R. S.; NORTON, D. P. Linking the balanced scorecard to strategy. **California Management Review**, v. 39, n. 1, pp. 53- 79, fall 1996a.
- _____. **The balanced scorecard – translating strategy into action**. Boston, Harvard Business School Press, 1996b.
- _____. Using the balanced scorecard as a strategic management system. **Harvard Business Review**, v.74, n.1, p.75-85, jan./feb. 1996c.
- KIMBALL, R. **The Data Warehouse Toolkit**, John Wiley & Sons Inc., New York, 1996.
- KUENG, P.; MEIER, A.; WETTSTEIN, T. Performance measurement systems must be engineered. **Communications of AIS**, v.7, article 3, 27 p., 2001.
- LAKATOS, E.M.; MARCONI, M.A. **Fundamentos em metodologia científica**. 3 ed. São Paulo, Atlas, 1995.
- LAUDON, K. C; LAUDON, J. P. **Management Information Systems: Managing the Digital Firm**, Prentice Hall, 7th edition, Upper Saddle River, New Jersey, 2002.
- LEBAS, M. J. Performance measurement and performance management. **International Journal of Production Economics**, v.41, n.23-25, pp.23-35, 1995.
- LEE, S.J.; SIAU, K. A review of data mining techniques. **Industrial Management & Data Systems**, v.101, n.1, pp. 41-46, 2001.
- LINDEN, A. *Advanced Mining for Information and Content*. In: GARTNER SYMPOSIUM ITXPO 2001, 2001, Orlando, Florida. **Paper....** Orlando, Florida, pp.1-18, 2001.

- MACEDO- SOARES, T. D. L. V. A.; RATTON, C. A. Medição de desempenho e estratégias orientadas para o cliente: resultados de uma pesquisa de empresas líderes no Brasil. **Revista de Administração de Empresas**, v. 39, n. 4, Out./ Dez. 1999.
- MARTINS, R.A. **Sistema de medição de desempenho: um modelo para estruturação do uso**. Tese (Doutorado).Universidade de São Paulo, Escola Politécnica, São Paulo, 1998.
- _____. 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, p.371-378, 2002.
- MAZZOTTI, A.J.A.; GEWANDSZNALADAR, F. **O método nas ciências naturais e sociais**. São Paulo, Pioneira, 1998.
- MENEZES, M.T. **Efeitos na medição de desempenho após a implementação do ERP: Estudo de Casos**. Dissertação (Mestrado).Universidade Federal de São Carlos, Departamento de Engenharia de Produção, São Carlos, 2002.
- METAXIOTIS, K.; PSARRAS, J. Expert Systems in business; applications and future directions for the operations researcher. **Industrial Management & Data Systems**. vol.103, n. 5, pp. 361-368, 2003.
- MITROFF, I.I., BETZ, F., PONDY, L.R. AND SAGASTI, F. On managing science in the systems age: two schemas for the study of science as a whole systems phenomenon. **Interfaces**, Vol. 4, No. 3, pp. 46-58. 1974.
- MORESI, E.A.D. Delineando o valor do sistema de informação de uma organização. **Ci. Inf., Brasília**, v. 29, n. 1, p. 14-24, jan./abr. 2000.
- NAVEGA, S. Princípios Essenciais do Data Mining. In: Anais do Infoimagem, 2002, São Paulo. **Anais**: Cenadem, 9p., 2002
- NEELY, A.; ADAMS, C.; CROWE, P. The performance prism in practice. **Measuring Business Excellence**, v. 5, n. 2, p. 6- 12, 2001.
- NEELY, A.; GREGORY, M.; PLATTS, K. – Performance Measurement System Design – A Literature Review and Research Agenda **International Journal of Operations & Production Management**, v. 15, N.4, p. 80–116, 1995.
- NEELY, A. **Measuring business performance**. London: The Economist, 1998.

- _____. The performance measurement revolution: Why now and What Next? **International Journal of Operations & Production Management** – v. 19, n.2, p.205-228, 1999.
- NICKOLS, F. **Measurement-based analysis: hooking what you do to the bottom line** [online]. 1999. Disponível em: <http://home.att.net/~nickols/mba.htm>. Acesso em: 15 de Maio de 2003.
- NORREKLT, H. The balance on the balanced scorecard - a critical analysis of some of its assumptions. **Management Accounting Research**, v. 11, p. 65- 88, 2000.
- OLIVEIRA, D.R., **Sistemas de Informações Gerenciais: estratégicas, táticas, operacionais**. São Paulo: Atlas, 1993.
- OLVE, N.G.; ROY, J.; WETTER, M. **Performance drivers – a practical guide to using the Balanced Scorecard**. New York: John Wiley & Sons, 347p, 1999.
- PFOHL H. C.; BUSE H. P., Inter-organizational logistics systems in flexible production networks, **International Journal of Physical Distribution & Logistics Management**, v. 30, n. 5, p. 388 – 408, 2000.
- POLITO, M. **Data Mining**. 1997. Universidade Estadual do Rio de Janeiro. Monografia. Disponível em: <http://hps.infolink.com.br/mpolito/mining/mining.htm>. Acesso em: 20 de Julho de 2002.
- PRATES, M. Conceituação de Sistemas de Informação do Ponto de Vista do Gerenciamento, **Revista do Instituto de Informática**, PUCCAMP, Março/Setembro, 1994.
- QUISPE, N.R.P. **Técnicas e ferramentas para a extração inteligente e automática de conhecimento em banco de dados**. Dissertação (Mestrado). Universidade de Campinas, Faculdade de Engenharia Elétrica e de Computação, Campinas, 2003.
- RODRIGUES FILHO, J.A.F.; SHIMIZU,T. Data mining: Conceitos Básicos e Aplicações. **Boletim Técnico da Escola Politécnica da USP – Departamento de Engenharia de Produção**, São Paulo, BT/PRO/104,11p., 2002.
- SALLE, C.T.P.;GUAHYBA, A.S.; WALD, V.B.; SILVA, A.B.; SALLE, F.O.; FALLAVENA, L.C.B. **Uso de Redes Neurais Artificiais para Estimar Parâmetros de Produção de Galinhas Reprodutoras Pesadas em Recria**. Revista Brasileira de Ciência Avícola. v.3, n.3, p.257-264, Dez 2001.
- SAS. The Quality Data Warehouse. **SAS e-Intelligence**. Estados Unidos. 12p. 2000.
- SILVA, E.M. **Avaliação do estado da arte e produtos data mining**. 2000. Disponível em: <http://www.mestradoinfo.ucb.br/aluno/esilva/datamining/index.htm#resumo>. Acesso em: 25 de Setembro de 2001.

- SINGH, H.S. **Data warehouse: conceitos, tecnologias, implementação e gerenciamento.** São Paulo: Makron Books, 382p., 2001.
- SINK, D.S. The role of measurement in achieving world class quality and productivity management. **Industrial Engineering**, v.21, n.6, p.23-28, 1991.
- SLACK, N.; HERLAND, C.; CHAMBERS, S. **Administração da Produção.** São Paulo: Atlas, 1999.
- SPRANGLER, W.E.; JERROLD, H.M.; VARGAS, L.G. Choosing Data-Mining Methods for Multiple Classification: Representational and Performance Measurement Implications for Decision Support . **Journal of Management Information Systems**, v.16, n.1, p.37-62, 1999.
- SUWIGNJO, P.; BITITCI, U.S; CARRIE, A.S. Quantitative models for performance measurement system. **International Journal of Production Economics**, v.64, p. 231-241, 2000.
- TAFNER, M.A. **O Que São as Redes Neurais Artificiais.** 1998. Universidade Estadual de Campinas. Revista "Cérebro & Mente". Disponível em: <http://www.epub.org.br/cm/n05/tecnologia/rna.htm>. Acesso em: 11 de Abril de 2004.
- TANLER, R. **The intranet data warehouse.** New York: John Wiley & Sons, Inc., 331p., 1997.
- THURASINGHAM, B. **Data Mining: Technologies, Techniques, Tools, and Trends.** CRC Press LLC, 274p., 1999.
- VAKKURI, J.; MEKLIN, P. The impact of culture on the use o performance measurement information in the university setting. **Management Decision**. 41/8, p.751-759, 2003.
- VALLE, B.M. Tecnologia da informação no contexto organizacional. **Ci. Inf.**, Brasília, v. 25, n. 1, 1996.
- VANDERLEI FILHO, D.; VALENÇA, M.J.S; LUDERMIR, T.B.; SILVA, G.P.F. Uma proposta *fuzzy* na avaliação de desempenho de bibliotecas universitárias brasileiras. **SNBU 2002.** Rio de Janeiro, 15 p., 2002.
- VELTZ, P.; ZARIFIAN, P. De la productivité des ressources à la productivité par l'organisation. **Revue Française de Gestion**, n.97, p.59-66, Jan./Fev. 1994.
- WARD, J.; PEPPARD, J. **Strategic Planning for Information Systems.** Wiley Series in Information Systems, 640 p., 2002.

WATANABE, E.M. **O método de análise hierárquica aplicado ao desenvolvimento do produto**. Dissertação (Mestrado).Universidade de Campinas, Faculdade de Engenharia Mecânica, Campinas, 2003.

WHITE, G.P. A survey and taxonomy of strategy-related performance measures for manufacturing. **International Journal of Operations & Production Management**, v.16, n.3, p.24-61, 1996

WONGTSCHOWSKI, P. **Indústria Química: riscos e oportunidades**. São Paulo: E. Blücher, 1999.

ZARIFIAN, P. Organização e sistema de gestão: à procura de uma nova coerência. **Gestão & Produção**, v.4, n.1, p.76-87, 1997.