UNIVERSIDADE FUMEC
FACULDADE DE CIÊNCIAS EMPRESARIAIS - FACE

ROGER PAUER ROCHA VIANA

DATA MINING:
Auxiliando na tomada de decisões estratégicas nas empresas

BELO HORIZONTE
2013

ROGER PAUER ROCHA VIANA

DATA MINING:

Auxiliando na tomada de decisões estratégicas nas empresas

Projeto de pesquisa realizado na Universidade FUMEC, no curso de Ciência da Computação, apresentado à disciplina Trabalho de Conclusão de Curso.

Orientadores:
Professor Leonardo Arruda Ribas
Professor Osvaldo Manoel Corrêa
Convidado:
Hudson Ludgero

BELO HORIZONTE
2013

ROGER PAUER ROCHA VIANA

DATA MINING:
Auxiliando na tomada de decisões estratégicas nas empresas

Trabalho de Conclusão de Curso realizado na Universidade FUMEC, no curso de Ciências da Computação, apresentado à disciplina Trabalho de Conclusão de Curso.

Professor Osvaldo Manoel Corrêa (TCC)
Universidade Fumec

Professor Leonardo Arruda Ribas (Orientador)
Universidade Fumec

BELO HORIZONTE
2013

Primeiramente a Deus por me dar forças nessa caminhada difícil, aos meus pais pela vida e por tudo que fizeram e ainda fazem por mim. Aos meus orientadores que me apoiaram em todos os momentos, especialmente o Leonardo Ribas que foi muito solidário e participativo.

Roger Pauer Rocha Viana.

RESUMO

Atualmente cada vez mais empresas investem em sistemas para auxiliar na dinâmica do seu negócio. Esses sistemas alimentam bases de dados com transações que são a realidade cotidiana do negócio daquela empresa. Esses dados são sempre compartilhados por especialistas em sistemas, disponibilizando-os para fácil acesso visando a rápida recuperação por quem necessite de alguma informação sobre os mesmos.

O que ocorre na maioria das vezes é que ao se ver um grande volume de dados as pessoas não conseguem interpretá-los de forma correta, pois isso requer um nível de capacidade técnica e analítica muito grande até mesmo para especialistas envolvidos.

Dessa forma, se faz necessário o uso de técnicas e ferramentas que possam extrair dessas bases de dados informações e conhecimento para que usuários possam utilizá-las visando o benefício empresarial, buscando oportunidades, riscos e também realizar planejamentos de médio e longo prazo. E é nesse contexto que as técnicas de Data mining se aplicam.

Este trabalho visa demonstrar o Business Intelligence conceitualmente e tecnicamente, bem como o processo de descoberta de informações em base de dados com enfoque na exposição ampla do data mining, falando sobre suas principais fases e algorítmos

Após o conhecimento das técnicas, pode se destacar os benefícios obtidos pelas empresas que se utilizam das ferramentas de inteligência de negócios (Business Intelligence), sobre tudo o data mining, no auxílio da tomada de decisão.

Palavras chave: Inteligência de negócio, Data mining, Business Intelligence, Banco de dados, KDD.

ABSTRACT

Currently more and more companies invest in systems to assist in the dynamics of your business. These systems feed databases with transactions that are the daily reality of the business of that company. These data are always shared by experts in systems, making them available for easy access for the rapid recovery for those who need some information about them.

What happens most often is that when we see a large amount of data people can not interpret them correctly, because it requires a level of technical and analytical ate very large even for experts involved.

Thus, it is necessary to use techniques and tools that can extract information such databases and knowledge so that users can use them in order to benefit business, seeking opportunities, risks, and also conduct planning for medium and long term. It is in this context that the data mining techniques are applied.

This paper seeks to describe the overall Business Intelligence demonstrating it conceptually and technically, as well as perform the approach of the aspects of the process of knowledge discovery in databases, data mining exhibiting widely and their algorithms.

After knowing the techniques can highlight the benefits obtained by companies that use the business intelligence tools, especially data mining, as an aid in decision making.

Key words: Business Intelligence, Data mining, Data Base, KDD.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS

B.I	*Business Intelligence*
D.W	*DataWarehouse* (Armazém de dados)
DOLAP	*Desktop On-Line Analytical Processing* (Processamentoanalítico*desktoponline*)
EIS	*Enterprise Information System* (Sistemas de informação empresarial)
ETL	*Extracttransformandload* (Extração Transformação Carga)
HOLAP	*Hybrid On-Line AnalyticalProcessing* (Processamento analítico híbrido *online*)
IBM	*International Business Machines* (é uma empresa estadunidense voltada para a área de informática)
KDD	*Knowledge Discovery in Databases* (Descoberta de conhecimento em base de dados)
K-NN	*K-NearestNeighbors* (K-Vizinhos mais Próximos)
KPI	*Key Performance Indicador* (Indicadores Chave de Desempenho)
MOLAP	*Multidimensional On-Line AnalyticalProcessing* (Processamento analítico multidimensional *online*)
OLAP	*On-line AnalyticalProcessing* (Processamento analítico *online*)
OLTP	*On-line TransactionProcessing* (Processamento de transações *online*)
ROLAP	*Relational On-Line AnalyticalProcessing* (Processamento analítico relacional *online*)
SAC	Serviço de Atendimento ao Consumidor
S.I	Sistemas de informação

SUMÁRIO

INTRODUÇÃO

Na atualidade em todos os segmentos comerciais o mercado esta cada vez mais competitivo. São milhares de empresas ofertando produtos e serviços semelhantes buscando mais clientes para se consolidar, crescer ou sair de uma crise. Dados e mais dados dessas empresas estão armazenados em suas bases de dados que são de importância histórica e revelam o dia a dia do negócio.

Mas e se todo esse montante de dados pudesse se transformar em informação que respondesse questões como: Qual o próximo passo a tomar? Qual produto deve ser intensificado a produção? Qual estratégia utilizar em determinada região? Qual produto deve ser retirado de produção?

Visando extrair essas informações surgiu o processo de B.I sendo que uma das suas principais técnicas é o data mining essa a qual possibilita análise de grande volume de dados através de sub técnicas e ferramentas.

Através do data mining grandes empresas mundo a fora tem conseguido se destacar e sair na frente dos concorrentes prevendo tendências, moldando seus produtos ou serviços conforme perfil de consumo dos mesmos e acima de tudo maximizando seus lucros.

Com as técnicas de data mining podemos proporcionar uma inteligência competitiva a nível empresarial que é um diferencial frente aos concorrentes podendo levar a empresa a atingir suas metas mais rapidamente.

São inúmeras as empresas nacionais e internacionais que se utilizam dessa técnica nos dias atuais, dentre elas podemos citar grandes corporações como: Telefônica, Sprint, Itaú, Golden Cross, dentre outras.

Assim sendo, este trabalho aborda o Business Intelligence de uma forma geral descrevendo mais detalhadamente as técnicas de data mining, buscando analisar os benefícios obtidos pelas mesmas no processo de tomada de decisão empresarial e inteligência competitiva.

Este trabalho está dividido em três capítulos, no primeiro capítulo foi trabalhado o Business Intelligence como um todo, descrevendo seu conceito, histórico, técnicas de armazenamento de dados e modelagem multidimensional.

O segundo capítulo apresenta a parte técnica do data mining, descrevendo seus conceitos, histórico, algorítmos e também o processo de descoberta de conhecimento em base de dados.

O terceiro capítulo busca demonstrar os benefícios da utilização do data mining no processo de tomada de decisão, além dos seus diferencias gerados para proporcionar uma maior inteligência competitiva nas empresas que o utilizam. Ao final deste capítulo são citados exemplos de sucesso de empresas que deixaram de perder clientes ou aumentaram seus lucros utilizando-se de data mining.

Pretende-se com este trabalho promover não só a atualização de conhecimento sobre o tema, mas destacar os benefícios gerados para uma gestão empresarial mais eficiente, baseado em informações obtidas por data mining.

CAPÍTULO I – DESCREVENDO O BUSINESS INTELLIGENCE.

1.1. Histórico do Business Intelligence (B.I.)

No atual ambiente computacional das empresas vemos uma grande massa de dados sendo gerada todos os dias, essa massa trás as informações cotidianas do negócio e suas regras específicas, mas essas mesmas informações se trabalhadas da forma correta podem nos trazer dados novos levando a ter uma certa percepção das pessoas responsáveis por gerir as empresas.

Para entendermos melhor esse cenário devemos destacar a revolução do conhecimento e da informação que se iniciou na virada do século XX e que evolui gradativamente.

A Tabela abaixo demonstra os detalhes da evolução dos Sistemas de Informação (S.I.) ao longo dos anos:

Período	Característica dos S.I.	Papel dos S.I.nos s
1950 a 1960	Processamento de Dados (ênfase Mudanças Técnicas)	Sistemas de Processamento Eletrônico de Dados - Processamento de transações, manutenção de registros e aplicações contábeis tradicionais.
1960 a 1970	Relatórios Administrativos (ênfase Controle Gerencial)	Sistemas de informação gerencial-Relatórios administrativos de informações pré-estipuladas para apoio a tomada dedecisão.
1970 a 1980	Apoio a Decisão (ênfase Controle Gerencial)	Sistemas de Apoio a Decisão – Apoio interativo e ad hoc ao processo de tomada de decisão gerencial.
1980 a 1990	Apoio Estratégico ao Usuário Final (ênfase Atividades Institucionais Essenciais)	Sistemas de computação do usuário final ‑ Apoio direto a computação para a produtividade do usuário final e colaboração de grupos de trabalho. Sistemas de informação executiva (EIS) ‑ Informações críticas para a alta administração. Sistemas especialistas ‑ Conselho especializado baseado no conhecimentopara os usuários finais.

		Sistemas de informação estratégica-Produtos e serviços estratégicos paravantagem competitiva.
A partir de 1990	Empresa e Conexão em Rede Global (ênfase Atividades Institucionais Essenciais)	Sistemas de informação interconectados‑ Para o usuário final, a empresa e acomputação, comunicações ecolaboração Interorganizacional, incluindooperações e administração globais na Internet, intranets, extranets e outras redes empresariais e mundiais.

Tabela 1: Evolução dos Sistemas de Informação.
Fonte: Adaptado de LAUDON e LAUDON; O' BRIEN (2001; 2001 apud SILVA JUNIOR, 2006).

A história do Business Intelligence, da maneira conhecida por nós atualmente, é iniciada na década de 70 quando os primeiros produtos de B.I. foram disponibilizados para os analistas de negócios.

Barbieri, (2001, p. 2), nos relata Seymour Pappert, um do grandes professores do MIT (Instituto de tecnologia do Massachussets), que na década referida já dizia que os dados e seus correlatos seriam responsáveis por uma revolução na sociedade, comparável até mesmo com a imprensa inventada por Gutemberg.

O maior problema dos primeiros produtos de B.I. era a necessidade de uma intensa e exaustiva programação, não disponibilizando a informação em tempo hábil e nem de uma forma muito flexível, também se exigia um alto custo de implantação. Serain (2007)

Após o surgimento dos sistemas gerenciadores de banco de dados relacionais, micro computadores e interfaces gráficas, vieram então os produtos realmente direcionados aos analistas de negócio, possibilitando uma maior rapidez e flexibilidade de análise sobre as informações.

Apos o entendimento do breve histórico do B.I. pode se adentrar melhor no assunto conhecendo seus conceitos.

1.2. Conceitos de Business Intelligence

A competitividade do mercado atual deixa o cliente com diversas opções de produtos e serviços similares. É cada vez mais importante que as empresas comecem a levantar informações sobre os dados dos seus sistemas transacionais, buscando encontrar respostas para melhorar um produto, criar ofertas etc. Desta forma uma empresa que conseguir se beneficiar de informações que antecipem a visão do cliente poderá conquistar uma preferência no seu segmento de negocio, buscando com isso uma consolidação, expansão ou afastar uma possível crise.

Conhecer a produção, custo de determinado produto ou serviço, o volume de vendas, etc., são exemplos simples de controle que às vezes muitas empresas não conseguem mensurar o que as faz perder dinheiro e tempo.

Nesse cenário que o B.I.se encaixa sendo um ramo computacional que visa extrair todas as informações das bases de dados transacionais e transformá-las em informação para que os profissionais dos setores gerenciais possam retirar vantagem e obter a inteligência de negócio. Barbieri conceitua B.I. como:

> O conceito de BI de forma mais ampla pode ser entendido com à utilização de variadas fontes de informações para se definir estratégias de competitividade dos s da empresa. O Universo hoje padece de um mal clássico. Possui uma montanha de dados, mas enfrenta grande dificuldade na extração de informações a partir dela.
> (BARBIERI, 2001, p. 34)

B.I. pode se constituir de uma vasta categoria de técnicas e ferramentas para extração, armazenamento e transformação de dados. Estas tecnologias acabam produzindo um ambiente de conhecimento onde há produção sistemática de informação é ágil e consistente.

Para um melhor entendimento dos processos de B.I, não podemos prosseguir no assunto sem antes termos uma idéia sobre banco de dados relacional. Bancos de dados são ferramentas que armazenam conjuntos de registros dispostos em estrutura regular, dificultando assim o tratamento dessa informação.

Nos próximos tópicos será abordado de forma mais ampla alguns conceitos como data warehouse, data mart, OLAP.

1.3. Data Warehouse

Barbieri define DataWarehouse da seguinte forma:

> Data warehouse, cuja tradução literal é armazém de dados, pode ser definido como um banco de dados destinado a sistemas de apoio a decisão e cujos dados foram armazenados em estruturas logicas dimensionais, possibilitando o seu processamento analítico por ferramentas especiais.
> (BARBIERI, 2001, p. 51)

O D.W.tem por características:

a) Baseia-se em assuntos: o D.W. é organizado em torno de assuntos macro de uma organização, tais como clientes, vendas e produtos e não em função de processos ou operações cotidianas. O D.W. foca em modelar os dados para o processo de tomada de decisão;

b) Integrado: é construído integrando diversos tipos de bases de dados, que em muitas vezes são heterogêneas, de forma a tornar as informações consistentes;

c) Varia conforme o tempo: armazena as informações numa perspectiva histórica.

Para termos um melhor entendimento sobre D.W.podemos fazer uma comparação entre o mesmo e os bancos de dados transacionais, que armazenam as informações cotidianas da empresa, esses são utilizados por todos os funcionários para registrar os dados atendendo a regras de negócio, por isso seus dados podem sofrer constantes mudanças.

Por não ocorrer redundância nos dados e as informações históricas serem geralmente armazenadas em dispositivos de backup e apagadas, este tipo de banco de dados reduz a capacidade de armazenamento se tratando de dados históricos. Já em D.W. são gerados dados analíticos, destinados às necessidades da gerência no processo de tomada de decisões. Isto pode envolver consultas complexas que necessitam acessar um grande número de registros.

Um D.W.armazena informações históricas de muitos anos e por isso deve ter uma grande capacidade de processamento e armazenamento dos dados que se encontram de uma forma mais sintética.

Carlos Barbieri (2001, p. 51) nos diz que: “A idéia de D.W. é armazenar os dados em vários graus de relacionamento e sumarização, de forma a facilitar e agilizar os processos de tomada de decisão por diferentes níveis gerenciais”.

Após o conhecimento de D.W faz se necessário falar sobre os seus subconjuntos, os data marts.

1.4. Data Mart

Data marts são subconjuntos de dados de um Data warehouse. Geralmente são dados referentes a um assunto ou área mais especifico como departamento de vendas, departamento de Estoque, etc.; ou então em diferentes níveis de sumarização como, por exemplo: Vendas trimestrais, Vendas mensais, Vendas semestrais.

Barbieri, (2001, p. 50), define Data Mart como: “O termo Data Mart (mercado de dados) significa, deposito de dados que atende a certas áreas especificas da empresa e voltados (também) para o processo decisório gerencial.”

Seus dados são provenientes do D.W, desnormalizados e passam por um processo de indexação para suportando assim intensa pesquisa.

Numa visão comparativa dos dados, onde devemos considerar os requisitos escopo, integração, tempo, agregação, análise e dados voláteis, percebemos que a diferenciação está no requisito de escopo, pois enquanto o DW é pensado para atender a empresa como um todo, o data mart é criado para atender um subconjunto da empresa.

Atender um subconjunto da empresa pode ser a reunião de dados de outros setores, já que, poucas vezes um único setor contém ou gera toda informação que a empresa necessita.

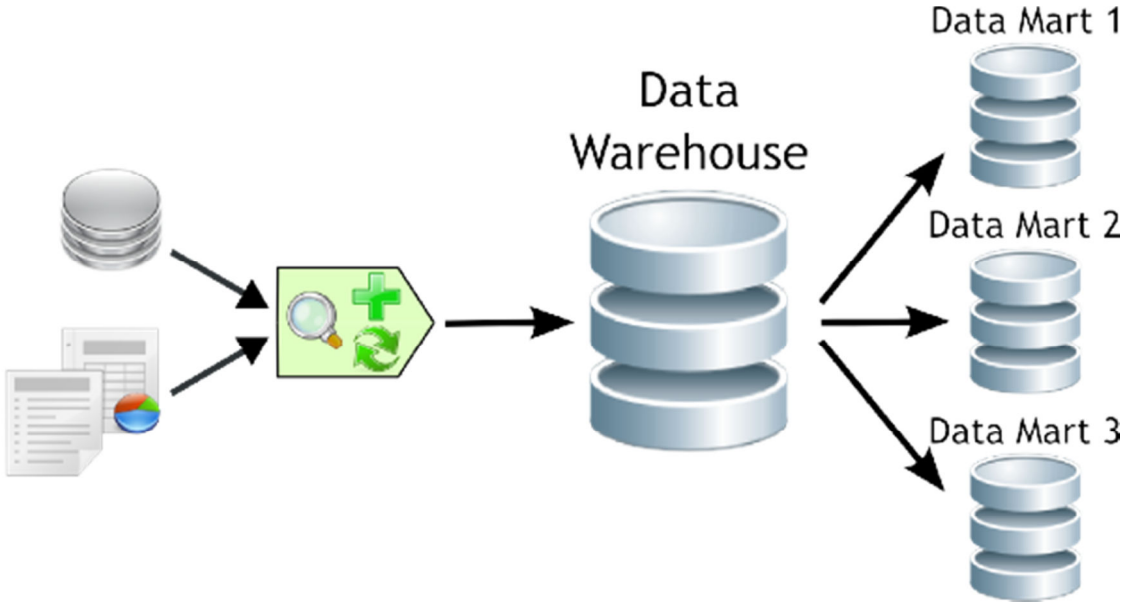

Figura 1: Esquema de um Data Mart
Fonte: Disponível em <http://www.dataprix.net/files/uploads/250image/HEFESTO%20v2_0/data%20mart%20-%20top%20down.png>. Acesso em: 21 abr. 2013.

Antes de se implementar um DW é necessário a realização seu projeto em termos de definição e modelagem dos dados, nesse quesito atualmente dois modelos são os mais dominantes, são eles o Star Schema e SnowFlake.

Segundo Moreira, (MOREIRA, 2006), o modelo star schema é como um modelo em formato de estrela, onde todas as tabelas relacionam-se diretamente com a tabela de fatos.

Sendo assim as tabelas dimensionais devem conter todas as descrições que são necessárias para definir uma classe como Produto, Tempo ou Loja nela mesma, ou seja, as tabelas de dimensões não são normalizadas no modelo estrela, então campos como categoria, departamento, marca contém suas descrições repetidas em cada registro. (MOREIRA, 2006)

Desta forma as tabelas dimensão tem seu tamanho aumentado pela repetição das descrições de forma textual em todos os registros.

Nas tabelas dimensão temos as principais características de um evento e nas tabelas fato, os fatos ocorridos, geralmente com as métricas e as chaves para as características correspondentes das tabelas dimensionais.

Na figura abaixo vemos a representação do modelo star schema:

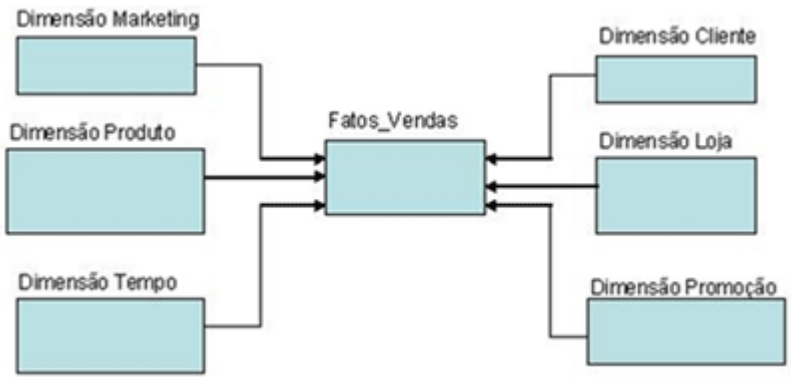

Figura 2: Modelo Star Schema
Fonte. Disponível em: <http://conteudo.imasters.com.br/3836/03.gif>. Acesso em: 24 abr. 2013.

Para Moreira (MOREIRA, 2006), no modelo Snowflake tem se o relacionamento das tabelas dimensão com as tabelas fatos, porem algumas das dimensões relacionam-se apenas entre si, isto acontece com intuito de normalização dessas tabelas dimensionais, buscando diminuir o espaço ocupado por estas tabelas.

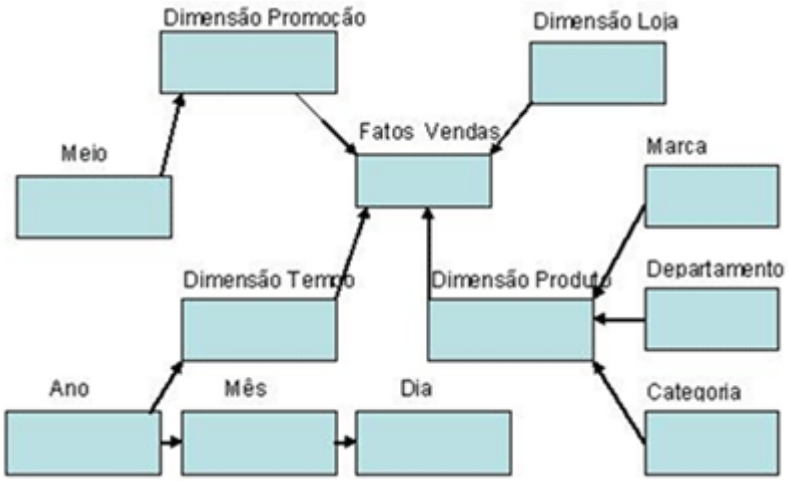

Figura 3: Modelo SnowFlake
Fonte. Disponível em: <http://conteudo.imasters.com.br/3836/04.gif>. Acesso em: 24 abr. 2013.

Após os assuntos descritos faz se necessário falar sobre outro tema muito importante em B.I que é OLAP.

1.5. OLAP

1.5.1. Origem

De acordo com Cynthia Aurora Anzanello, (2005, p.5), bases de análise Multidimensional para OLAP não são uma nova tecnologia. A IBM desenvolveu e implementou a primeira linguagem com análise multidimensional, no fim da década de 1960, esta linguagem foi chamada de APL.

Esta ferramenta foi definida matematicamente, baseando-se em símbolos gregos, foi utilizada por usuários finais e grande consumidora de recursos, sendo muito usada entre as décadas de 1980 e 1990 em aplicações de apoio a negócios. Acompanhando a evolução dos sistemas, na década de 1990, uma nova classe de ferramentas foi introduzida no mercado, essas ferramentas foram nomeadas de OLAP.

Ferramentas de OLAP tem a maioria dos princípios introduzidos pela linguagem APL, mas, com maior integração na utilização dos dados fontes.

1.5.2. OLAP x OLTP

Antes de adentramos mais no mundo OLAP é necessário o entendimento das suas diferenças perante um ambiente OLTP

De acordo com Henrique (2012), o OLTP (On-line Transaction Processing) faz a captura das transações de um sistema e armazena no banco de dados. Bases desse tipo são utilizadas em sistemas que registram pequenas transações realizadas em tempo real e que ocorrem frequentemente sendo de forma rápida.

Por salvar apenas um curto histórico dos dados, não é recomendado seu uso como base de dados adequada para ajudar na tomada de decisões.

Kimball (2002), *apud* Henrique (2012), diz que os sistemas OLTP tem sua modelagem relacional que visam eliminar ao máximo a redundância, de forma que uma transação que gere alterações no estado do banco de dados, atue o mais precisamente possível.

Com isso os dados normalizados estão distribuídos em diversas tabelas, o que traz uma considerável complexidade à criação de uma consulta por um usuário final. Sendo assim, esta prática não parece ser a ideal para o projeto de D.W, onde estruturas mais simples, com menor nível de normalização devem ser buscadas.

O OLAP (On-line Analytical Processing) é destinado à tomada de decisões, oferecendo uma visualização dos dados orientada à análise, além de uma navegação mais flexível e rápida. O OLAP recebe dados do OLTP para que se possam ser feitas as análises, contém dados atuais e históricos. (Henrique, 2012)

Através de pesquisa e um estilo de navegação simplificado, usuários finais podem rapidamente analisar inúmeros panoramas, gerar relatórios, identificar tendências e fatos relevantes sem se preocupar com tamanho, complexidade, e fonte dos dados.

Henrique, (2012), nos diz que o setor gerencial de uma empresa utiliza-se do OLAP para as tomadas de decisões, e assim é feito o planejamento estratégico.

A tabela 2 abaixo demostra as principais características entre OLAP e OLTP:

Características	OLTP	OLAP
Operação Típica	Atualização	Análise
Telas	Imutável	Definida pelo Usuário
Nível de Dados	Atomizado	Altamente Sumarizado
Idade dos Dados	Presente	Histórico, Atual e Projetado
Recuperação	Poucos registros	Muitos registros
Orientação	Registro	Arrays
Modelagem	Processo	Assunto

Tabela 2: Características dos sistemas OLAP e OLTP
Fonte. Disponível em: <http://social.technet.microsoft.com/wiki/contents/articles/6934.oltp-x-olap-pt-br.aspx>. Acesso em: 27 abr. 2013.

1.5.3. Multidimensionalidade

De acordo com Anzanello, (2005, p. 6), a visão multidimensional é composta por consultas que proporcionam dados sobre medidas de desempenho, decompostas por uma ou mais dimensões dessas medidas. Podendo também serem selecionadas pelas dimensões ou pelo valor da medida. As visões multidimensionais disponibilizam as técnicas básicas para cálculo e análise necessários pelas aplicações de B.I.

Para se obter a visão multidimensional, Anzanello (2005, p. 6), nos diz que é necessário compreender outras características:

a) Cubo: Estrutura que guarda os dados em formato multidimensional, tornando sua análise mais fácil.
b) Dimensão: Unidade de análise que reúne dados relacionados. As dimensões vêm a se transformar em cabeçalho de colunas e linhas, como exemplo períodos temporais, linhas de produto, regiões de venda.
c) Hierarquia: Formada por todos os níveis de uma dimensão, pode ou não ser balanceada. Na hierarquia balanceada os níveis mais baixos são correspondentes entre si, no entanto isto não acontece nas hierarquias não balanceadas no qual a equivalência hierárquica não existe. Podemos exemplificar com uma dimensão geografia, onde o nível país não contém um subnível Estado para um determinado elemento e contém para outro.
d) Membro: Pode ser definido como subconjunto em uma dimensão. Cada nível hierárquico tem elementos adequados aquele nível. Anzanello (2005, p. 6)

1.5.4. Arquiteturas

Anzanello (2005), nos descreve os tipos de OLAP mais utilizados, sendo apresentados a seguir:

a) MOLAP (Multidimensional On-Line Analytical Processing) o armazenamento de dados é feito de forma multidimensional, é implementado de conforme a ferramenta OLAP utilizada, sendo regularmente implementado em bancos de dados relacionais, no entanto não na terceira forma normal. Além disso, o acesso aos dados acontece de forma direta no banco de dados do servidor multidimensional.

b) ROLAP (Relational On-Line Analytical Processing) os dados são armazenados no modelo relacional como também suas consultas são processadas pelo gerenciador do banco relacional.
c) DOLAP (Desktop On-Line Analytical Processing) é uma alteração existente para disponibilizar a portabilidade dos dados. A vantagem oferecida é arquitetura e a redução do tráfico na rede.
d) A arquitetura mais atual é a HOLAP (Hybrid On-Line Analytical Processing), na qual ocorre uma mistura entre ROLAP e MOLAP. A vantagem é que com a combinação de tecnologias pode-se obter o que há de melhor em ambas, o alto desempenho do MOLAP e a escalabilidade do ROLAP. Anzanello (2005)

Tendo se visto o OLAP pode ser abordado um outro processo também muito importante em B.I. que é o ETL

1.6. ETL

Para Ribeiro (2011), ETL vindo do inglês Extract Transform Load, (Extração, Transformação e Carga) é o processo que tem como objetivo a realização de toda a parte de extração de dados de fontes diversas, transformação para atender às necessidades de negócios e carga dos dados em um D.W.

Os projetos de D.W consolidam dados de diferentes fontes, a maioria dessas fontes tendem a ser bancos de dados relacionais ou arquivos de texto, mas podem existir outros tipos de fontes também como planilhas Excel, etc; um sistema ETL precisa ter a capacidade de comunicação com todo o tipo de fonte de dados.

Lima (2010), nos diz que é uma fase extremamente crítica de um D.W, envolvendo a movimentação dos dados de origem das diversas fontes existentes.

Como já falado anteriormente as etapas do processo são extração, transformação e carga dos dados.

A extração, conforme IBL (2003), busca a captação de dados de fontes diversas, no qual cada sistema pode utilizar diferentes formatos de dados, um formato dos mais comuns, conforme já dito, são os arquivos texto.

De acordo com Lima (2010), o processo de transformação contém também o processo de limpeza dos dados. Na limpeza são removidas as inconsistências obtidas entre diversificadas fontes de dados participantes do processo de ETL. Na transformação é feita a padronização dos dados oriundos de vários sistemas com formatos diferentes.

IBL (2003), nos diz que o estágio de transformação dos dados é onde devemos aplicar regras ou funções nos dados extraídos para que não venham a ocorrer problemas em sua carga nas bases de dados de destino.

O processo de transformação ainda pode conter regras como:

a) Seleção de algumas, ou nenhuma, colunas para carregar.
b) Padronização de valores codificados como, por exemplo, se o sistema fonte tem a definição de 1 e 2 para sexo masculino e feminino respectivamente, mas o D.W opta pelo armazenamento de M e F para masculino e feminino.
c) Dados derivados ou calculados.
d) Unificação ou junção de dados de fontes heterogêneas.
e) Sumarização ou agregação dos dados.
f) Geração de chaves substitutas (surrogate keys).
g) Operações de pivot, transformação de linhas em colunas e vice-versa.
h) Quebra de uma ou mais colunas em varias outras colunas.

Como o volume de dados pode ser muito grande, segundo Lima (2010), há muitos casos que não temos condições de processar as extrações e transformações em uma janela de tempo no qual o D.W. não está sendo utilizado, fazendo-se necessário o uso das chamada staging áreas, para que possamos executar os processos com sucesso.

A Staging Area é uma parte do D.W responsável por receber o ETL das informações dos sistemas transacionais legados, para posterior geração dos Data Marts de destino. Tem como principais características possuir uma estrutura similar as fontes de dados de origem (visando um ETL mais rápido), ser fora do acesso dos usuários para consulta, dentre outras.

A fase de carga faz o carregamento dos dados para o D.W.e dependendo das necessidades da organização esse processo tende a variar. Em Alguns D.W's pode haver a substituição dos dados existentes semanalmente por dados atualizados, enquanto outros adicionam os dados a um tempo pré-determinado. IBL (2003).

Conforme visto, foi descrito neste capítulo o business intelligence de uma forma geral, fazendo a sua conceituação, histórico e demonstrando suas principais metodologias para uma correta armazenagem e consulta aos dados de forma rápida e eficiente.

Todo esse conteúdo é de grande importância se tratando de inteligência de negócios, ficando pendente a exposição de uma técnica muito relevante e que se bem aplicada permite ganhos fantásticos paras as empresas: O data mining. Este será retratado no próximo capítulo sobre um aspecto mais técnico.

CAPÍTULO II – DATA MINING SOBRE O ASPECTO TÉCNICO

Este capítulo visa abordar o data mining de uma forma mais técnica, mas antes é necessário falar sobre o processo de descoberta de informação em base de dados (KDD), no qual o data mining é uma de suas etapas.

2.1 KDD

2.1.1 Definição e histórico

KDD (Knowledge Discovery in Databases) cuja tradução é descoberta de conhecimento em base de dados, é o procedimento de extração de informações de base de dados, que cria relações de interesse para serem analisadas pelos especialistas, bem como o auxilio da validação de conhecimento extraído.

O termo KDD possui varias etapas relacionadas, sendo elas: seleção, pré-processamento, transformação, data-mining e interpretação enquanto que data mining é usado apenas para a fase de descoberta do processo de KDD. Goldschmidt e Passos, (2005, p.2), mencionam que o data mining dentro de um processo de KDD é apenas uma etapa.

O termo KDD surgiu no final da década de 1980, mais precisamente em 1989, sendo um novo ramo da computação, visando com a extração de conhecimento, uma maneira automatizada de explorar as crescentes bases de dados e reconhecer os padrões existentes através da modelagem de fenômenos do mundo real. (Goldschmidt e Passos, 2005, p. 3)

2.1.2 Processo

O processo de KDD é dinâmico, apesar de ter uma definição parecida a de data mining, deve ser composto de várias etapas em sequência, podendo haver retorno a etapas anteriores, isto é, as descobertas realizadas (ou a falta delas).

Ocasionalmente, este processo leva a novas hipóteses e descobrimentos. Neste caso, o usuário tem a escolha de se decidir pela retomada dos processos de mineração, ou uma nova escolha de atributos, por exemplo, para comprovar as hipóteses que apareceram ao longo do processo.

O processo de KDD de descoberta de dados é composto por varias etapas operacionais, a figura 4 nos apresenta essas etapas.

Segundo Goldschmidt e Passos, (2005, p.2), para um melhor entendimento do processo é necessário primeiro uma apresentação dos principais elementos das aplicações de KDD que são:

a) O problema onde o processo de KDD vai ser aplicado. Esse problema pode ter como característica 3 elementos: conjunto de dados envolvido no problema , o especialista com domínio da aplicação e objetivos da aplicação.
b) Os recursos disponíveis para resolução dos problemas descritos, entre eles pode se ressaltar: o especialista em KDD, as ferramentas de KDD e a plataforma computacional disponível.
c) Resultados conseguidos com a aplicação dos recursos no problema. Abrange os modelos de conhecimento encontrados ao longo da aplicação de KDD e o histórico das ações feitas.

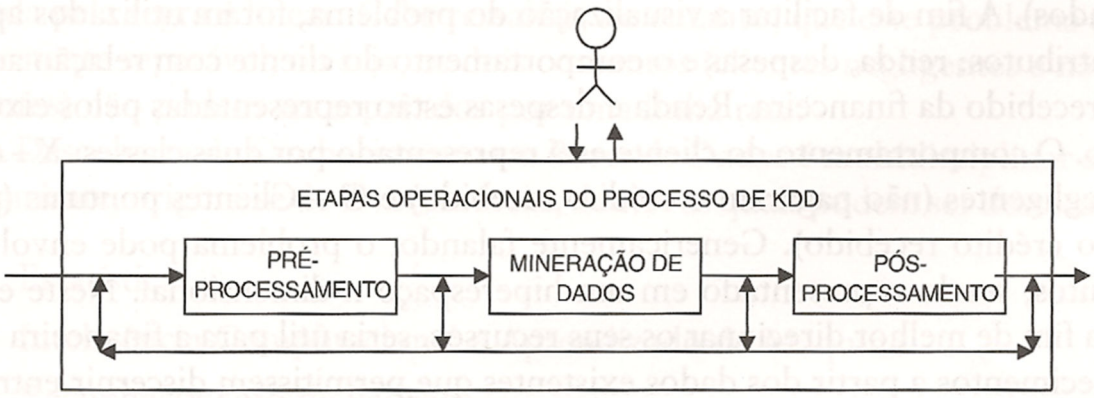

Figura 4: Etapas operacionais do processo de KDD.
Fonte: Goldschmidt e Passos, 2005, p. 3.

No tópico seguinte serão detalhadas as principais fases do processo de KDD.

2.1.3 Fases principais do processo de KDD

Para Prass (2012), as principais fases do processo de KDD são: seleção, pré-processamento e limpeza, transformação, data mining, interpretação e avaliação.

A figura 6 a seguir nos demonstra as fases descritas dentro do processo.

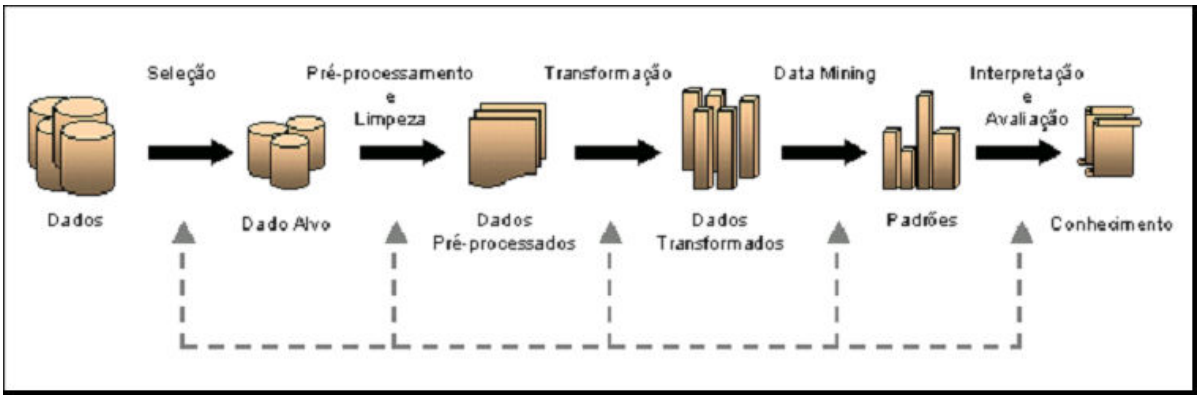

Figura 5: Principais fases do processo de KDD.
Fonte: Prass, Fernando Sarturi, 2012. Disponível em:< http://fp2.com.br/blog/wp-content/uploads/2012/08/kdd.png>. Acesso em: 30 Abr. 2013

Vamos então descrever detalhadamente as fases demonstradas na figura 5:

a) Seleção: Esta fase, no qual selecionamos os dados, é a primeira no processo de descobrimento de informação e possui um impacto significativo sobre a qualidade do resultado final no processo de KDD, sendo que nesta fase escolhemos o conjunto de dados que irão conter todas as possíveis variáveis (podendo ser denominadas de características ou atributos) e também os registros (podendo ser denominados de casos ou observações) a serem analisados. A escolha dos dados geralmente fica a critério de um especialista do domínio, alguém que entende do assunto tratado.

   De acordo com Prass (2012), o processo de seleção é bem complexo, onde os dados podem vir de diversas fontes (D.W, planilhas, sistemas legados) e podem possuir formatos diversos.

b) Pré-processamento e Limpeza: O Pré-processamento e limpeza dos dados é uma parte fundamental em um processo de KDD, pois a qualidade dos dados vai ser determinante na eficiência dos algorítmos de data mining.

   Segundo Goldschmidt e Passos (2005), nesta etapa deverão ser realizadas atividades que eliminem os dados redundantes e inconsistentes, realizem a recuperação dos dados incompletos, e ainda avaliem dados possivelmente discrepantes ao conjunto.

   A participação de um especialista do domínio é essencial, pois na maioria dos casos somente alguém que entende do assunto é capacitado a dizer se um dado é discrepante ao conjunto ou é simplesmente um erro de digitação.

   Nesta fase também utilizamos métodos de redução ou transformação para diminuir o número de variáveis envolvidas no processo de KDD, objetivando com isto melhorar a performance do algorítmo de análise.

Prass nos diz que:

> Identificar de dados inapropriados dentro do conjunto selecionado é problemático, e isto dificulta a automatização desta fase. Definir um dado como "ruim" dentro do conjunto depende da estrutura do mesmo e também de que aplicação é dada a ele. (PRASS, 2012)

c) Transformação dos Dados: Transformar os dados é a fase do KDD anterior a fase de Data Mining. Após ser realizada a seleção, limpeza e pré-processados, os dados tem a necessidade de serem armazenados e formatados corretamente para que os algorítmos possam ser utilizados.

   Prass (2012), nos diz que em grandes corporações é comum encontrar computadores executando diferentes sistemas gerenciadores de Bancos de Dados (SGDB), onde estes dados dispersos devem ser agrupados em um repositório único. Também, nesta fase, há possibilidade de obtenção de dados faltantes através do processo de transformar ou combinar outros dados, assim esses dados obtidos são chamados de dados derivados.

   Um simples exemplo de um dado que pode ser calculado a partir de outro dado é a idade de um indivíduo, podendo ser encontrada a partir de sua data de nascimento.

d) Data Mining: Goldschmidt e Passos (2005), nos descrevem essa etapa dentro do processo de KDD como sendo a principal, onde ocorre uma busca efetiva por conhecimentos novos e úteis a partir dos dados utilizando-se de algorítmos, que são fundamentados em técnicas que buscam, segundos determinados paradigmas, produzir modelos de conhecimento através da exploração dos dados.

   O objetivo da etapa de data mining, como já dito anteriormente, é fornecer informações ás corporações que as possibilitem montar melhores estratégias de marketing, vendas, suporte, melhorando assim os seus negócios.

   Essa fase é o assunto principal desse capítulo, após esse tópico continuaremos a detalhá-la melhor.

e) Interpretação e Avaliação: o data mining trás com ele uma série de idéias e técnicas para uma variedade de campos. Estatísticos, pesquisadores de Inteligência Artificial e administradores de bancos de dados utilizam se de técnicas diferentes para interpretar e avaliar os resultados obtidos com o data mining para chegar a um fim: a informação.

Após conceituado e conhecido o processo de KDD, pode-se então adentrar na abordagem do data mining, começando pela sua definição logo a adiante.

2.2. Definindo o Data Mining

Data mining, ou mineração de dados trata-se do processo de análise de dados utilizando se de técnicas para exploração, de forma a descobrir novos padrões e relações interessantes podendo representar informações de grande relevância. Devido ao grande montante de dados esses padrões dificilmente seriam descobertos com métodos mais tradicionais como consultas a base de dados ou relatórios.

Os padrões podem ser definidos como sendo uma afirmação sobre uma distribuição de probabilidade, podendo ser expressos na forma de regras, sejam elas por fórmulas e funções, entre outras.

> Os conceitos de garimpagem de dados (Data Mining) estão relacionados com a nova tendência (para aplicações comerciais) de se buscar correlações escondidas em altos volumes de dados, nem sempre evidentes, principalmente no tratamento cotidiano dos sistemas de informações.
> (BARBIERI, 2001, p. 178)

O interesse existente por este tipo de informação se dá principalmente ao fato de que as instituições estão coletando e armazenando cada vez mais dados e como consequência do baixo valor de meios de armazenamento e computadores e também do aumento da capacidade de ambos.

Com a maior utilização de D.W, tende a aumentar a quantidade de informações disponíveis. Conforme já mencionado anteriormente, métodos tradicionais de análise de dados, não são apropriados para grandes volumes de dados, pois podem criar relatórios informativos sobre os dados, mas não conseguem analisar o conteúdo destes relatórios a fim de obter conhecimentos importantes.

Para Barbieri (2001, p. 178), o Data Mining é uma forma de se capitalizar em cima de informações, na tentativa de descobrir padrões de comportamento de clientes ou estilos de ações fraudulentas em cartões de credito, seguradoras etc.

A técnica de mining buscar algo a mais que somente interpretação dos dados existentes, almejando principalmente a realização de previsões com possíveis fatos e correlações não explicitadas em um D.W. ou D.M.

No fundo, com as técnicas de Data Mining visamos identificar atributos e indicadores capazes de melhor definir uma situação especifica.

Barbieri (2001, p. 179), nos cita o exemplo de uma empresa de seguros no qual as ferramentas de OLAP nos responderiam perguntas do tipo: “Qual o valor médio de pagamentos de seguros de vida para não fumantes, na região sul do estado, em agosto de

determinada data?". O uso das ferramentas de Mining para o exemplo acima nos trariam melhores atributos de clientes, capazes de ajudarem como previsores de possíveis acidentes de automóvel. A figura 6 nos demonstra uma visão dos exemplos de atributos tratados por técnicas de data mining em um ambiente diversificado de B.I.

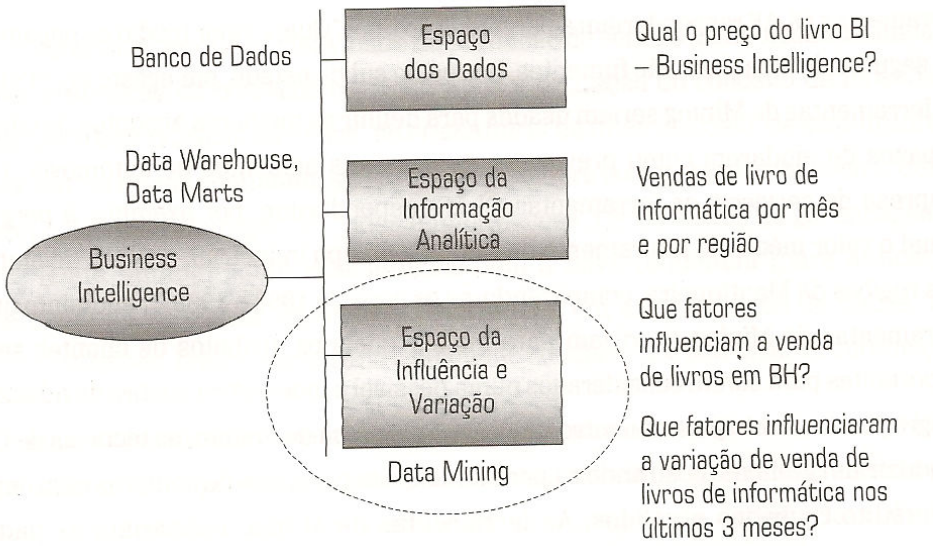

Figura 6: Visão geral dos diversos espaços de conhecimento
Fonte: Barbieri, 2001, p. 180

Feita a definição, o passo seguinte é falar sobre as principais metodologias para aplicação de data mining.

2.3 Métodos de Data Mining

2.3.1 Redes Neurais

De acordo com Goldschmidt e Passos (2005, p.175), redes neurais artificias são modelos matemáticos que se baseiam nos princípios de funcionamento dos neurônios biológicos e na estrutura do cérebro humano. Modelos esses que tem a capacidade de adquirir, armazenar e utilizar conhecimento experimental e visão uma simulação computacional da habilidade dos seres humanos como generalização, aprendizado, associação e abstração.

Segundo Goldschmidt e Passos (2005, p.176), suas principais características são:

a) Busca paralela: Nas redes neurais o conteúdo fica distribuído pela estrutura das redes, desta forma a busca pela informação ocorre de uma forma paralela e não sequencial.

b) Aprendizado por experiência: As redes neurais buscam aprender padrões sobre os dados explorados utilizando-se de um processo de repetidas apresentação dos dados a rede, procurando assim abstrair modelos de conhecimento.
c) Generalização: Redes neurais tem a capacidade de generalizar seu conhecimento com base em exemplos anteriores permitindo a mesma lidar com ruídos e distorções nos dados.
d) Abstração: Capacidade em perceber quais são características relevantes em um conjunto de dados de entrada.
e) Robustez e degradação gradual: Com essa característica a perda de um conjunto de neurônios artificias não causa necessariamente um mal funcionamento desta rede, pois a informação fica distribuída em toda a rede.

Numa rede neural artificial os neurônios artificiais são arranjados em camadas conectas. A figura 7 abaixo nos demonstra a estrutura de uma rede neural simples. Os círculos tendem a representar os neurônios e as linhas representam as conexões.

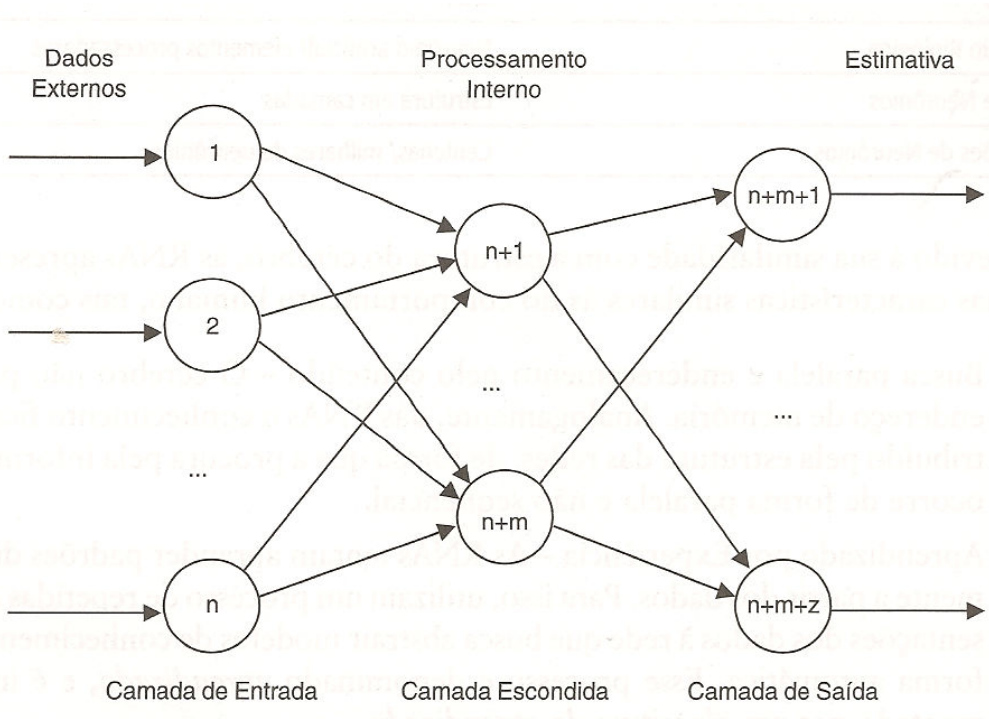

Figura 7: Arquitetura de uma rede neural artificial.
Fonte: Goldschmidt e Passos, 2005, p. 176.

A camada que recebe os dados é denominada camada de entrada e a que exibe o resultado camada de saída. Nas camadas internas ocorre o processamento da rede, uma rede neural pode ter várias camadas internas dependendo da complexidade do problema.

Para Goldschmidt e Passos (2005, p. 85), a topologia da rede neural varia em função do problema e da representação adotada para os dados, no geral aplicações de data mining a camada de entrada recebe os dados pré- processados de uma base de dados. A rede processa esses dados produzindo uma saída variando conforme a aplicação.

A seguir será descrito alguns algorítmos de aprendizado mais utilizados em redes neurais.

a) Back-Propagation: Goldschmidt e Passos (2005, p. 85), nos descrevem que esse é um algorítmo de aprendizado supervisionado, suja aplicação é adequada a tarefas dentro do data mining como classificação, regressão ou previsão de series temporais. Seu objetivo principal é minimizar a função de erro entre a saída gerada pela rede neural e a saída real esperada. Utilizando o método do gradiente descendente.
b) Kohonen: Geralmente ele é baseado em uma forma de competição entre os elementos processadores, suas principais aplicações são as tarefas de clusterização (agrupa dados em conjuntos semelhantes) e detecção de regularidades (o sistema deve extrair características relevantes nos padrões de entrada dos dados).

2.3.2 Algorítmos Genéricos

Goldschmidt e Passos nos definem algorítmos genéricos da seguinte forma:

> Algorítmos genéricos são modelos computacionais de busca e otimização de soluções em problemas complexos, inspirados nos princípios evolutivos de Charles Darwin e também na reprodução genética. Resumidamente, algorítmos genéricos são técnicas que procuram obter boas soluções para problemas complexos por meio da evolução de populações de soluções codificadas em cromossomas artificiais. (GOLDSCHMIDT e PASSOS, 2005, p. 195)

Para Muniz (2008), algorítmos genéticos possuem uma solução potencial para um problema específico numa estrutura parecida a de um cromossomo humano, fazendo uso de operadores de seleção e cross-over a essas estruturas mantendo informações críticas referentes à solução do problema.

Para compreender de uma melhor forma como trabalha os algorítmos genéticos, vamos verificar um modelo matemático, a maximização da função f(x) = x2 irá ajudar a compreender todo o seu processo.

Vamos maximizar f(x) = x2 no intervalo de 0 a 31. Iniciamos a população de cromossomos com 4 escolhidos aleatoriamente.

x1 = 13, x2 = 24, x3 = 8, x4 = 19

Realizando o calculo da função de adaptação (f(x) = x2) para cada termo teremos:

f(x1) = 169, f(x2) = 576, f(x3) = 64, f(x4) = 361

Podemos ver que a melhor solução nesta geração é x2. Muniz (2008)

A adaptação geral vem a sero somatório de todas as adaptações de cada cromossomo, 1170. Percentualmente temos x1 tem participação de 14%, x2 de 49%, x3 de 6% e x4 de 31%. Vamos sortear 4 números aleatórios entre 0 e 100 para verificamos em que ponto da reta entre 0 e 100 esses números encontram-se e então realizar a cópia dos cromossomos.

O cromossomo x1 será copiado uma vez, o cromossomo x2 vai ser reproduzido duas vezes, o cromossomo x3 não deve reproduzido, pois nenhum sorteio aleatório caiu dentro da faixa de 6% entre 64% e 69% e o cromossomo x4 será reproduzido também uma vez. Muniz (2008)

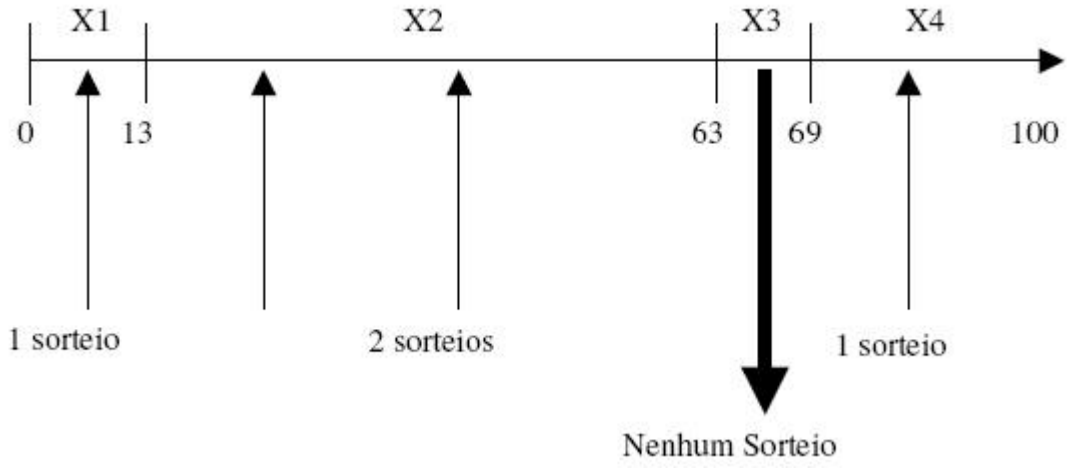

Figura 8: Algorítmo Genérico
Fonte: Muniz, Vander Emiro, 2008. Disponível em:<http://www.devmedia.com.br/imagens/10-07-2007pic03.jpg>. Acesso em: 28 Abr. 2013

A nova geração após a reprodução será de: x1 = 13, x2 = 24, x3 = 24 e x4 = 19.

De acordo com Muniz (2008), pode se notar que x2 = x3 nesta nova geração e que o x3 da geração anterior não se reproduziu, pelo motivo da pouca adaptação, desta forma não há nenhum representante seu nesta nova geração. A nova geração mostra a mescla das soluções bem-sucedidas da geração anterior que se uniram e se reproduziram.

Há possibilidade continuar o processo de evolução, mas ele pode ser interrompido se o valor for considerado suficiente ou até atingir o valor máximo da função f(x) no intervalo de 0 a 31.

2.3.3 Algorítmos baseados em Instâncias

De acordo com Goldschmidt e Passos (2005, p. 98), a expressão de método baseado em instância, indica que o método leva em consideração as instâncias ou os registros existentes na base de dados. Um dos principais métodos que se baseiam em instâncias é denominado de K-NN.

Esse método é frequentemente utilizado em aplicações envolvendo a tarefa de classificação pois trata-se de um método de fácil entendimento e implementação.

No seu processamento o algorítmo K-NN considera os seguintes passos:

a) Cálculo da distância do novo registro a cada um dos registros na base de referência.
b) Identificação dos k registros na base de referência que demonstraram menor intervalo em relação ao novo registro.
c) Verificação da classe mais frequente entre os k registros identificados no passo anterior.
d) Comparação da classe apurada com a classe real, computando erro ou acerto do algorítmo. Este passo deve apenas ser utilizado quando as classes dos novos registros são conhecidas e se quer avaliar o desempenho do método K-NN na base de dados em questão. Caso contrario não deve ser utilizado. Goldschmidt e Passos (2005, p. 99)

   Considerando o exemplo em um contexto de analise de credito avaliando-se a possibilidade de concessão ou não do credito a clientes. A base de dados de referencia encontra-se na figura 9 abaixo. O conjunto está dividido em duas classes: os negligentes representados por "*", representados por "●" estão os não negligentes.

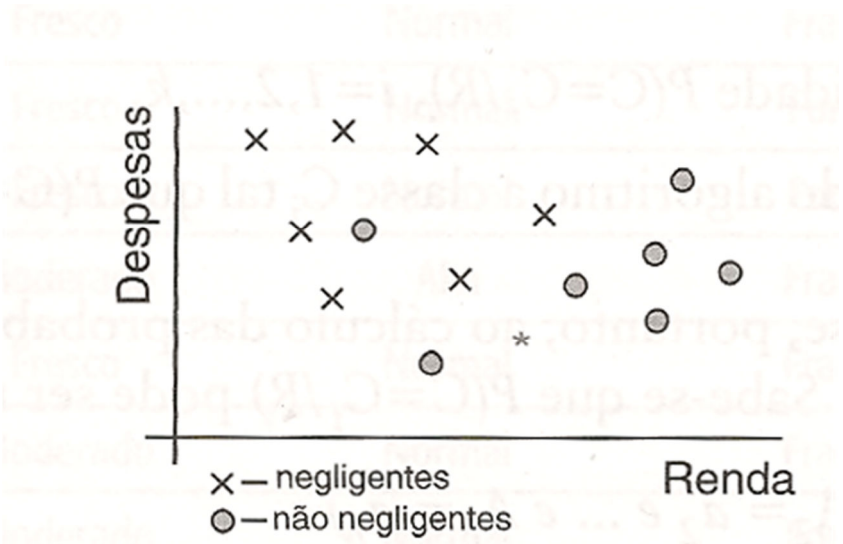

Figura 9: Conjunto dos clientes que receberam crédito.
Fonte: Goldschmidt e Passos, 2005, p. 99.

Apresentando-se um novo registro, representado por "*", faz se o cálculo da distância entre este novo registro e todos os registros já existentes na base de dados. Considerando que o numero de k de vizinhos mais próximos seja 3, apenas os 3 registros com menor distância ao novo serão considerados.

Assim avaliando os resultados na figura 10 abaixo, observamos que a classe com maior ocorrência dentro da área delimitada pelo algorítmo K-NN foi cliente entre não negligentes. Pela aplicação do algorítmo K-NN no exemplo apresentado, o credito seria concedido ao cliente solicitante. Goldschmidt e Passos (2005, p. 99)

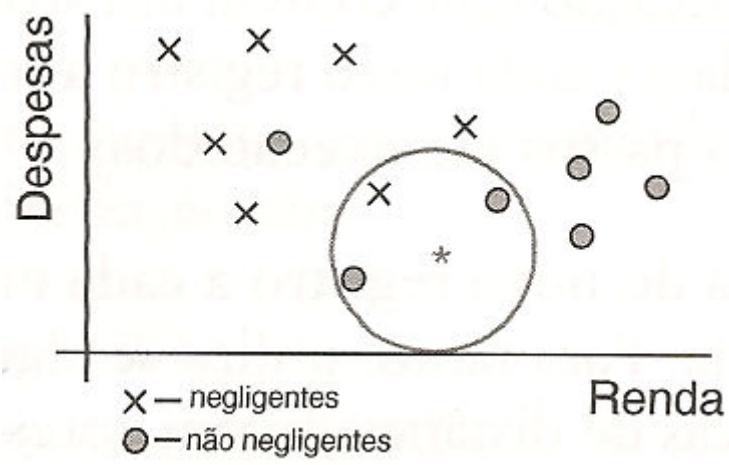

Figura 10: Resultado do K-NN.
Fonte: Goldschmidt e Passos, 2005, p. 100.

2.3.4 Métodos Estatísticos

Segundo Goldschmidt e Passos (2005, p. 100), vários são os algorítmos de data mining que se utilizam de princípios estatísticos, dentre eles podemos citar:

a) Classificador Bayeasiano;
b) K-Means;
c) K-Modes;
d) K-Prototypes;
e) K-Medoids;

Descrever cada um tornaria esse capítulo muito extenso, vamos descrever apenas o classificador Bayeasiano.

2.3.4.1 Classificador Bayeasiano

Pichiliani (2006), nos diz que este algorítmo tem essa nomenclatura porque é baseado na teoria da probabilidade de Bayes. Seu objetivo é o calculo da probabilidade, de que um novo dado faça parte de alguma classe estabelecida previamente.

Ainda segundo Pichiliani (2006), essa ação preventiva pode ser nomeada como classificação estatística, porque é baseada completamente em probabilidades. Esta classificação também pode ser denominada de simples ou ingênua, a mesma leva em consideração que o efeito do valor de um atributo sobre uma determinada classe é independente dos valores dos demais atributos.

Uma característica deste tipo de algorítmo é que ele necessita de um conjunto de dados já classificado previamente, ou seja, ele é voltado para tarefas preditivas. Com base neste conjunto de dados prévio, chamado também de conjunto de treinamento, o algorítmo recebe como entrada uma nova amostra desconhecida, e retorna como saída a classe mais provável para esta amostra com base em cálculos probabilísticos.

De acordo com Pichiliani (2006), seu funcionamento pode ser explicado da seguinte forma:

Inicialmente, cada classe do conjunto de treinamento tem sua probabilidade calculada. O cálculo é feito dividindo-se o número de dados de determinada classe pelo número total de dados do conjunto de treinamento.

Feito isso, calcula-se a probabilidade da inserção de um novo dado para cada classe que existe. Na sequência, é feita a multiplicação do valor obtido pela probabilidade da

classe calculada inicialmente na etapa de treinamento. Com as probabilidades para cada classe calculada, verifica-se qual é a classe que possui maior probabilidade de conter o novo dado.

2.3.5 Métodos Específicos

Goldschmidt e Passos (2005, p. 105), nos dizem são algorítmos desenvolvidos especificamente para implementar alguma tarefa de data mining.Vamos dar ênfase nesse tópico ao algorítmo Apriori, que é um algorítmo de descoberta de associações.

Apriori é um algorítmo tradicional no aprendizado de regras associativas, sendo utilizado com bases de dados que possuem transações.

Sendo comum em data mining associativo, dado conjuntos de itens, o algorítmo tenta achar subconjuntos semelhantes que se encontrem acima do nível de confiança que o usuário definiu.

Para Pichiliani (2008), o algorítmo Apriori utiliza uma aproximação botton-up, no qual subconjuntos são estendidos um item por vez, método também chamado de geração de candidatos, e grupos de candidatos são submetidos a teste a partir de bases de dados. O algorítmo finaliza quando extensões válidas não são mais achadas.

Apriori utiliza busca em largura e uma estrutura de árvore hash para contar com eficácia conjuntos de itens candidatos, criando conjuntos de itens candidatos de tamanho k baseando se em conjuntos de itens de tamanho k-1.

Goldschmidt e Passos (2005, p. 105), nos dizem que: "... Um k-itemset somente pode ser frequente se todos os seus (k-1) itemsets forem frequentes".

O algorítmo exclui os candidatos que não possuem um sub-padrão frequente,o conjunto candidato contém todos os conjuntos de itens frequentes de tamanho k.

Em seguida ele faz uma busca na base de dados relacional determinando conjuntos de itens regulares entre os candidatos. É utilizada a árvore hash para guardar conjuntos candidatos, se utilizando da mesma para apontar quais são os itens de maior frequência.

A árvore hash tem conjuntos de itens nas folhas e tabelas hash nos nós internos.

2.3.6 Métodos baseados em indução de árvores de decisão

Segundo Muniz (2008), árvores de decisão são representações gráficas no qual os nós significam amostras e as folhas significam categorias.

Uma árvore de decisão aponta uma classe numérica para uma entrada padrão filtrando a amostra por testes feitos na árvore. Cada teste tem mutualmente resultados exaustivos e exclusivos. Muniz (2008)

Quando uma amostragem de uma população está sendo estudada com o objetivo de realizar alguma dedução indutiva, árvores de decisão são os modelos mais usados. Muniz (2008)

Abaixo tem se o exemplo de uma árvore de decisão, para um sistema de aprovação escolar na figura 11 abaixo.

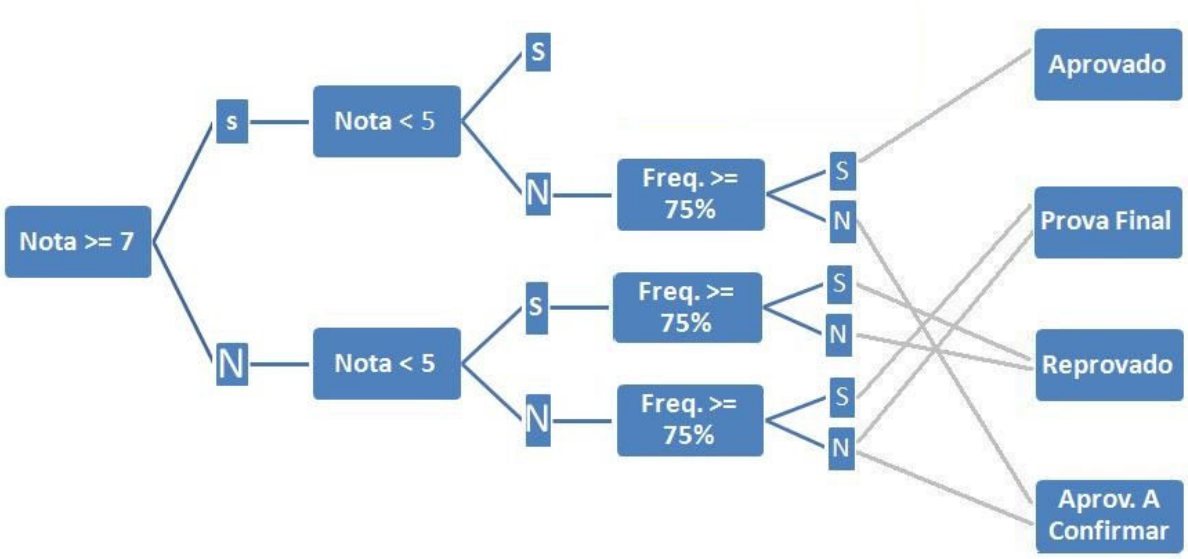

Figura 11: Árvore de decisão.
Fonte: Aula de eng. de software -14/11.
Disponível em:<http://turmadeasg.files.wordpress.com/2008/11/arvore-de-decisao1.jpg >. Acesso em: 28 Abr. 2013

De acordo Pichiliani (2006), para se implementar um algorítmo baseado em árvore de decisões devemos incialmente gerar o nó raiz da árvore, nessa fase cada classe do conjunto de treinamento possui a sua probabilidade calculada. Devemos agora encontrar os nos da árvore que ainda podem ser divididos para geração de novos nós, se não houver mais nenhum nó que possa ser dividido o algorítmo finaliza.

Para cada nó do conjunto de nós que podem ser divididos se deve realizar a escolha de um atributo que melhor qualifica os dados,nesta escolha deve-se excluir todos os atributos que não foram utilizados ainda no caminho que começa deste o nó raiz até o nó que será dividido. Além de considerar os atributos que já foram utilizados, também devemos fazer uma analise quantitativa de nós folha que o atributo gera e a quantidade de nós não folhas optando pelo atributo que mais gere nós folha e que menos gere nós divisíveis.

Após a escolha do atributo, é criado e desenhado o nó e as suas ramificações de acordo com todos os valores possíveis para o atributo. A criação das ramificações gera novos nós que devem verificados em seguida.

2.3.7 Métodos baseados em Lógica Nebulosa

Para Goldschmidt e Passos (2005, p. 183), a Lógica Nebulosa objetiva modelar o modo aproximado de raciocínio humano, buscando criar sistemas computacionais capazes de tomar decisões racionais em um ambiente incerto e impreciso. A Lógica Nebulosa disponibiliza um mecanismo de manipulação das informações imprecisas, como os conceitos de pequeno, alto, muito, pouco, bom, ruim, quente, frio, etc, fornecendo assim uma resposta próxima a uma questão baseada em conhecimentos não exatos, incompletos ou parcialmente confiáveis.

Diversos métodos de data mining foram adaptados de forma a incorporar a flexibilidade proporcionada pela Lógica Nebulosa, sendo um deles o algorítmo de Wang-Mendel, concebido para aplicação na tarefa de previsão de series temporais.

Segundo Goldschmidt e Passos (2005, p. 113), "O método Wang-Mendel consiste em abstrair regras nebulosas a partir de conjuntos de dados históricos, utilizamos esses dados para definir os antecedentes de os consequentes das regras nebulosas".

Temos que considerar então X(k), K=1,2,... uma série temporal, em que X(k)$[U^-;U^+]$ em *m* conjuntos nebulosos de comprimento igual, devendo *m* ser um valor impar.

Pode ser visto na figura 12 abaixo a ilustração de uma serie temporal sendo dividida em 7 conjuntos nebulosos.

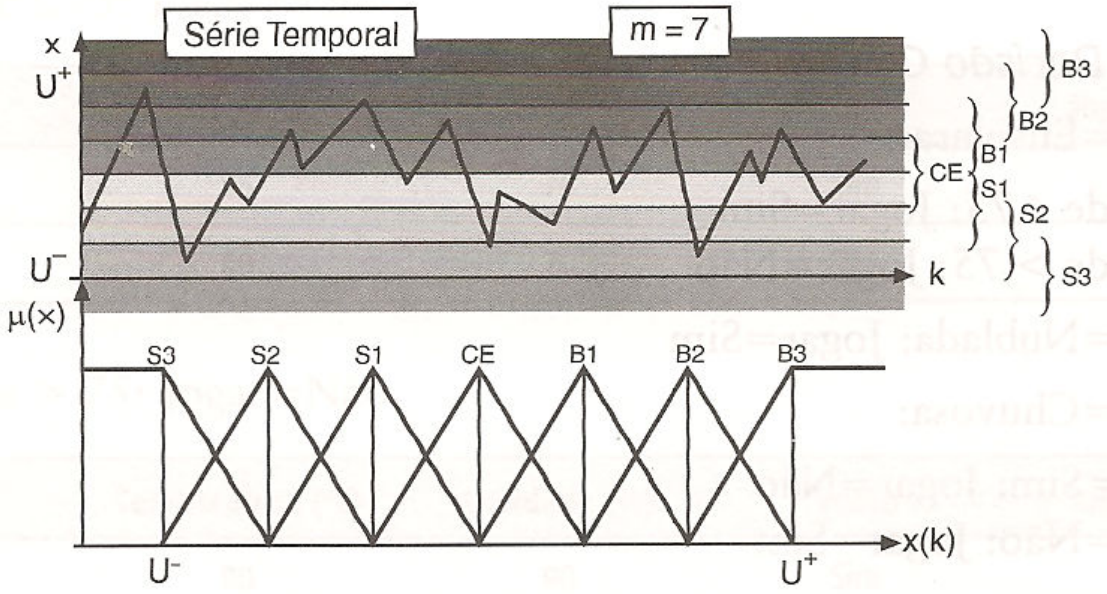

Figura 12: Divisão da serie temporal em conjuntos nebulosos.
Fonte: Goldschmidt e Passos, 2005, p. 114.

Foram demonstrados vários algorítmos utilizados em técnicas de data mining, bem como suas principais características e definições.

Todo esse conteúdo abordado foi necessário para um melhor entendimento do data mining sobretudo num aspecto técnico. Assim pode se concluir que o data mining é uma importante etapa no processo de descoberta de conhecimento em base de dados, nos trazendo informações preciosas que se bem usadas podem trazer um grande auxilio na tomada de decisão estratégica nas empresas. Assunto esse que será discutido no capítulo seguinte.

CAPÍTULO III – DATA MINING NO AUXILIO NA TOMADA DE DECISÃO ESTRATÉGICA NAS EMPRESAS

3.1 O processo de tomada de decisão nas organizações

Segundo Porto (2008), o atual paradigma que as organizações estão inseridas está mais dinâmico, exercendo grande influência nas mesmas. Mediante a isso, é necessário que os gestores tenham a percepção do que os ambientes interno e externo da organização tem a indicar quanto a oportunidade e ameaças, visando fazer escolhas com base na realidade organizacional.

Porto (2008), ainda nos diz que fatores como globalização, avanço tecnológico, desenvolvimento das telecomunicações e o menor tempo de processamento das informações tornam o ambiente de uma organização mais complexo, fazendo com que os gestores tenham sempre que reavaliar o processo de tomada de decisão, exigindo assim uma visão sistemática e cautela.

Para Raskin (2009), o aperfeiçoamento no processo de tomada de decisão deve ser um constante pensamento das organizações, a busca de novas informações para a avaliação de novas possibilidades ajuda os gestores nesse processo.

O grande montante de dados gerados diariamente nos sistemas transacionais das organizações é um desafio para os gestores na árdua tarefa de se converter bits em informação útil. Fazer uso da informação que está contida implicitamente em todo o volume de dados dos sistemas legados de uma corporação é de extrema utilidade, para isso devemos nos utilizar de soluções de data mining. Raposo (2010)

Raposo (2010), ainda nos afirma que a utilização do Data mining nos possibilita uma analise dos dados na busca de padrões de que gerem valor para organização, sendo o cada vez mais utilizado por revelar estruturas de conhecimento que auxiliam na tomada de decisão.

Com o uso do data mining surge uma gama de oportunidades, gerando aprendizado e dados adicionais que podem gerar influência na criação de estratégias organizacionais, garantindo vantagem competitiva e possibilitando melhoras nos produtos e serviços.As técnicas de data mining podem vir a ser um pilar nas empresas modernas. Raposo (2010)

A figura 13 logo abaixo demostra que a informação primaria, é a base da tomada de decisão.

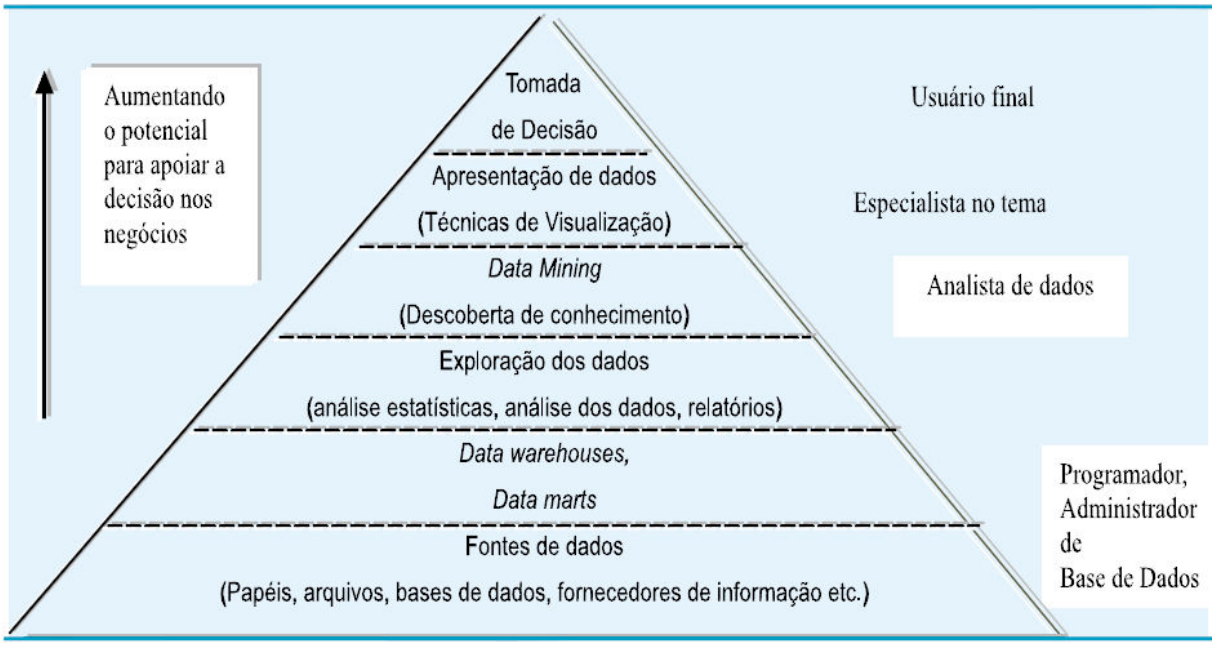

Figura 13: A importância da informação na tomada de decisão.
Fonte: Cabena*et ai*, 1997 & Tyson, 1998citado por Quoniam , Tarapanoff, Júnior, Alvares, 2001, p. 20

Desta forma, segundo Porto (2008), uma decisão de qualidade e acertada se baseia no adequado uso da informação em um processo decisório, levando a escolha de opções que gerem resultados benéficos a organização.

3.2 Inteligência Competitiva e a utilizaçãodo data mining

De acordo com Hilsdorf (2010), Inteligência competitiva é uma maneira proativa de se obter e articular informações que gerem valor para a organização através de análises de tendências e cenários, sendo que essas informações podem ser referentes ao comportamento da concorrência, clientes e do mercado em geral.

A inteligência de negócios objetiva a melhoria e ampliação das condições de competitividade. Seu uso reorienta modelos de negócios, metas, planejamentos e tem um ganho em termos evolutivos do negócio da organização frente aos seus concorrentes, isso ocorre através da antecipação das tendências de mercado com informações obtidas de fontes mistas. Hilsdorf (2010)

Em uma era no qual o conhecimento e inovação são fatores base para competitividade, a inteligência competitiva permite ações proativas ou invés de ações reativas, que são muito comuns no mundo dos negócios.

Um exemplo, conforme Hilsdorf (2010), de como o mercado em geral é reativo se trata da maneira da obtenção de informações pelas empresas de clientes e concorrentes através da mídia, clientes em geral, dentre outros. As organizações que tem essa postura só tomam conhecimentos dos fatos de relevância quando os mesmos já ocorreram ou estão ocorrendo. Isso é um cenário que a inteligência de negócios busca reverter através de um modelo preditivo de administração, procurando oportunizar condições ofertadas pelas tendências observadas antevendo os movimentos das empresas concorrentes.

Seguindo essa linha de raciocínio Montini (2009), nos diz que o data mining atualmente é bastante utilizado para a detecção de comportamento de clientes, perfis de utilização de cartão de crédito, perfis de telespectadores, perfis de pacientes que tem doenças específicas, dente outros, o que auxilia na elaboração de campanhas de marketing.

Hilsdorf (2010), nos diz que as 500 maiores empresas norte americanas possuem setor ou profissionais exclusivamente empenhados no monitoramento das ações de empresas concorrentes, coletando dados sobre seus erros e acertos, e observando minuciosamente seus passos e a estratégia das mesmas para não serem surpreendidos.

Para Hilsdorf (2010) as vantagens que a inteligência competitiva trás consigo são:

a) Diminuição das surpresas em relação às ações das organizações concorrentes.
b) Apontamento de oportunidades e ameaças.
c) Formulação de planejamento baseado em conhecimento obtido atrás de informações.
d) Aprendizado através do acompanhamento de ações assertivas e errônias dos concorrentes.
e) Entendimento do impacto de ações estratégicas sobre o mercado
f) Revisão, realinhamento das estratégias.
g) Verificação e melhora da sustentabilidade do nosso negócio.

Calegari (2012), citado por Oliveira (2012), nos diz que com as informações implícitas obtidas pelas técnicas de data mining,a empresa vai encontrar conhecimentos que não são óbvios nem triviais, sendo um passo a frentes e tratando de estratégia analítica.

As técnicas de data mining geram informações que possibilitam a realização de uma comunicação mais adequada conforme o publico, tendo um maior entendimento dos clientes, definindo também rentabilidade de produtos e planejamento de produção. Montini (2009)

Desta forma segundo Hilsdorf (2010), as empresas que fazem o bom uso da inteligência competitiva tem um aprendizado mais rápido e maior eficiência nas mudanças

perante seus concorrentes, sendo vistas com bons olhos pelos clientes e como inovadoras perante o mercado.

De acordo com Calegari (2012), citado por Oliveira (2012), as empresas se utilizam de data mining pela preocupação com a competitividade, sendo que uma rápida avaliação das informações é um fator de diferenciação.

A IDC (2012), citado por Oliveira (2012), tem uma projeção de 20% de crescimento até 2015 do seu setor estratégico, que engloba as técnicas de data mining.

Lucro e redução de custos, além do melhor entendimento das necessidades do cliente, são fatores da aplicação do data mining nas organizações. Alessandra (2012), *apud* Oliveira (2012).

Varejo, telecom, transporte e internet encontraram no data mining uma forma de surpreender o cliente e também achar consumidores propensos a inadimplência,bem como auxilio na definição de estratégias de cobranças . Esse setores também são pioneiros na utilização do data mining, pois lidam com grandes bases de dados estruturadas e não e estruturadas, sendo alta a competitividade nesses ramos. Oliveira (2012)

Montini (2012), descreve varias aplicabilidades do data mining, que auxiliam a inteligência competitivaem diversos setores:

a) Varejo e-commerce: faz uso da técnica para realizar a um cruzamento da cesta de compras com produtos do perfil do cliente, estimulando assim as vendas.
b) Setor bancário: realiza análises sobre possíveis fraudadores de cartão de credito, risco de pagamentos, perfis de investidores e os indica aplicações.
c) Varejo supermercadista: cria estratégias promocionais através do perfil de dados obtido pelo cruzamento das vendas, buscando assim a aumento das mesmas.
d) Pequenas e médias empresas: determinar perfil de clientes no seu segmento, mix de venda de produtos etc.
e) Agricultura: previsão e planejamento de vendas para o mercado interno e externo bem como gestão de estoque.

Conforme nos diz a Oracle (2012), *apud* Oliveira (2012), tem ocorrido uma maior procura nos últimos tempos pela sua ferramenta de data mining, denominada Oracle Data Mining. Isso ocorreu pela exigência de respostas mais rápidas do mercado a dúvidas e também pelo amadurecimento das organizações perante a análise de dados.

Conforme visto, vários são benefícios da utilização do data mining pelas empresas, se faz necessário ainda ver os ganhos da sua utilização, obtendo indicadores genéricos empresariais como *churn, cross-selling,* dentre outros.

3.3 Benefício da utilização do data mining com indicadores genéricos.

Segundo a StatSoft, a utilização de ferramentas de data mining nos trás vantagens em obtenção de indicadores de KPI comumente analisados em ambiente empresarial.
Segue abaixo o detalhamento desses benefícios nos indicadores:

a) *Churn Analysis*: o entendimento do cliente se tornou um fator crucial no atual mercado competitivo e globalizado. Não é uma tarefa trivial definir o comportamento de um cliente. Podemos levantar questão simples como: Como prever a migração de clientes para empresas concorrentes?

   A StatSoft, nos define *Churn Analysis* como sendo o estudo da previsão ao cliente que esteja em eminente ameaça de deixar de consumir determinado produto ou serviço em detrimento do consumo de outro produto ou serviço similar de outra empresa.

   Utilizamos das técnicas de data mining árvore de classificação, redes neurais dentre outras para obtenção de informações de *churn*

b) *Market Basket Analysis*: o objetivo do *Market Basket Analysis,* conforme nos diz a StatSoft, é uma cesta de produtos mais rentável. Questões importantes são tratadas por esse indicador como associação de venda de produtos, disposição dos produtos de maneira que estimula um maior consumo, fidelização de clientes, etc.

c) *ClusteringAnalysis* (Identificação de grupos e perfis de clientes): para StatSoft, são técnicas para agrupamento de clientes em grupos de características ou perfis similares em uma população analisada, descobrindo assim possibilidades de novas oportunidades para produtos e serviços, ou como melhor direcioná-los para cada conjunto.

   Árvore de decisão, K-NN dentre outros são exemplos de técnicas de data mining utilizadas para obtenção desse indicador. Sendo assim um grande ganho com as informações obtidas do *Clustering Analysis* é o direcionamento mais preciso de ações de marketing.

d) *Cross-selling-Up-selling*: permite-nos analisar preventivamente custos, uso, dentre outros comportamento dos clientes com o benefício de maximização de vendas fortalecimento do relacionamento com o mesmo. Essa informação é obtida através do próprio montante de dados dos sistemas utilizados pelas organizações. StatSoft

e) Conhecimento para concessão de crédito: De acordo com StatSoft, a concessão de crédito envolve perigo eminente ao processo de empréstimo financeiro. É feita uma análise através das técnicas de data mining utilizando métodos estatísticos, para prever

esse risco com base em cruzamento de dados históricos do cliente solicitante. Gerando assim estimativas consideráveis aceitáveis para decisão do empréstimo ou não.

f) *Risk Management*: Para StatSoft é a gerência de riscos de forma previsiva, através de abordagens efetivas de custos reduzindo ameaças a empresa.

g) Text Mining: se trata do agrupamento e classificação de informações de texto através de técnicas de data mining muito utilizado em dados referentes à SAC.

Geralmente é feita uma análise em dados não estruturados os transformando em dados estruturados. StatSoft

Empresas de telefonia, por exemplo, podem se utilizar desses indicadores para analisar uma base de dados sobre reclamações, sugestões e elogios conseguindo verificar o nível de satisfação melhorando assim o serviço. StatSoft

A utilização correta das informações contidas nos indicadores apresentados nesse tópico bem como técnicas de B.I em si pode fazer toda diferença para o sucesso ou ampliação de uma empresa, então devemos conhecer alguns desses casos de sucesso.

3.4 Casos de sucesso no uso de B.I e técnicas de data mining.

De acordo com Gurovitz (1997), uma gigante do varejo norte americano descobriu em sua massa de dados que a venda de fraldas descartáveis estava relacionada a venda de cerveja. O perfil de consumir deste cenário foi de homens que saiam a noite para comprar fraldas e também levavam algumas unidade de cerveja. Foi realizada uma nova disposição dos produtos os colocando próximos maximizando a venda de ambos.

Gurovitz (1997), nos diz que o banco Itaú, que é pioneiro no uso de D.W. no Brasil, tinha um percentual bem baixo (apenas 2%) de respostas em cima de envios, superiores a milhões, de malas diretas para seus correntistas. Foi realizada uma mudança nesse processo emitindo cartas somente a clientes cuja a análise dos dados filtram os que tem uma maior chance de resposta.

O retorno das malas enviadas aumentou em 30%, o que além da redução de custo de correspondência (cerca de 75%), aumentou a efetividade do serviço.

A empresa de Telefonia norte americana Sprint, conseguiu prever com 61% de segurança a troca operadora de telefonia dos consumidores em um período de 2 meses.

Baseado nessas informações elaborou estratégias de marketing conseguindo evitar a perda de 120.000 de seus clientes, o que representam 35 milhões de dólares em faturamento.Gurovitz (1997)

A Telefônica, companhia do ramo de telecomunicações, manteve sua receita anual de 150 milhões de dólares ao detectar através, de técnicas de B.I, que mais da metade das ligações de manutenção eram originadas de empresas rivais. Assim, fez reparos imediatos mantendo milhares de clientes insatisfeitos. Gurovitz (1997)

De acordo com Gurovitz (1997), o governo do estado de Massachusetts, no Estados Unidos, processava informações financeiras através da impressão das telas dos terminais de grande porte. Após a utilização de D.W conseguiu otimizar o processo reduzindo tempo e custo. Somente com papel a economia foi de 250.000 dólares.

A Golden Gross conseguiu verificar que os usuários que mais cancelavam os seus planos de saúde eram os que menos utilizavam, realizando ações de marketing direcionadas nesse publico alvo. Gurovitz (1997)

Ao findar este capítulo é possível dimensionar a grande diferença que informações conseguidas num processo de data mining podem proporcionar para quem as utiliza corretamente. Seus benefícios mudam rumos em uma empresa, podendo alterar inclusive seu planejamento estratégico, lucros e posição de mercado.

Todos os conteúdos expostos levam a crer que o data mining é uma peça chave para as empresas que querem se destacar em um mercado cada vez mais competitivo.

CONCLUSÃO

No mercado competitivo atual tem se visto cada vez menos empresas oferecerem serviços ou produtos inovadores. Muitas vezes a concorrência é tão árdua entre as empresas de determinado setor que faz o cliente ver esse mercado de uma forma muito homogênea, levando o mesmo a não optar por esta ou aquela empresa.

Desta forma, é difícil uma expansão ou consolidação empresarial sem a inteligência adequada. Muitos profissionais de gerência não se atentam que ter o conhecimento do próprio negócio, e principalmente antever o cliente é uma forma muito eficaz de se conseguir uma inteligência que faça a empresa se destacar, estando sempre um passo a frente das demais.

Esse conhecimento está contido no grande montante de dados gerados diariamente pelos sistemas transacionais legados, que escondem informações vitais principalmente sobre o comportamento dos clientes.

Sendo assim, surge no ramo computacional o B.I, que conforme descrito no primeiro capítulo, é uma ciência que tem técnicas e metodologias de correta armazenagem e consulta aos dados de forma mais rápida e eficiente, visando sobretudo a inteligência empresarial.

Sua técnica mais relevante é o data mining que proporciona o encontro das informações ocultas através de algorítmos que baseiam em tendências, associações, simulações artificias da inteligência humana, para uma análise dos dados.

Assim, é possível obter benefícios como criação de perfis de clientes, podendo se direcionar com mais eficácia as campanhas de marketing, fazer o reposicionamento de produtos ou o aperfeiçoamento de processos deficientes garantindo a melhoria em ações estratégicas e maximizando lucros.

Em uma sociedade que se baseia cada vez mais em informação o uso de técnicas de data mining vem a ser um fator primordial na busca de inteligência competitiva em ambientes empresariais.

Com base no que foi exposto nesse documento, principalmente pela ênfase dos exemplos citados, pode-se dizer que o data mining é uma técnica fundamental para o processo de tomada de decisão e crescimento das empresas.

REFERÊNCIAS

ANZANELLO, Cynthia Aurora. **OLAP conceitos e utilização**. Instituto de Informática, UFRGS. Disponível em: <http://www.inf.ufrgs.br/~clesio/cmp151/cmp15120021/artigo_cynthia.pdf>. Acesso em 02 set. 2011.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. **Data Mining**: Um guia pratico. São Paulo: Editora Campus, 2005.

BARBIERI, Carlos. **BI**: Modelagem e tecnologia. Rio de Janeiro: Axcel Books do Brasil, 2001.

GUROVITZ, Helio. **O que cerveja tem a ver com fraldas?** Disponível em: <http://exame.abril.com.br/revista-exame/edicoes/0633/noticias/o-que-cerveja-tem-a-ver-com-fraldas-m0053931>. Acesso em 26 Maio 2013

HENRIQUE, Ozimar. **OLTP x OLAP**. Disponível em: <http://social.technet.microsoft.com/wiki/contents/articles/6934.oltp-x-olap-pt-br.aspx>. Acesso em 24 abr. 2013.

HILSDORF, Carlos. **O que é inteligência competitiva?** Disponível em: <http://www.administradores.com.br/artigos/administracao-e-negocios/o-que-e-inteligencia-competitiva/44824/>. Acesso em 29 Maio 2013

IBL - Informatica Brasileira LTDA. **Conceito – Extração, Transformação e Carga.** Disponível em: <http://www.infobras.com.br/portugues/produtos_conceito_etl.asp>. Acesso em 30 Abr. 2013

INMON, W.H; TERDEMAN, R.H; IMHOFF, Claudia. **Data Warehousing**: Como transformar informações em oportunidades de s. São Paulo: Editora Berkeley, 2001.

LANA, Rogério Adilson. **Inteligência competitiva**: Fator- Chave para o sucesso das organizações no novo milênio.
Disponível em: <http://www.abraic.org.br/v2/artigos_detalhe.asp?c=793>. Acesso em 02 set. 2011.

LAUDON,K.C.; LAUDON, J.P.**Sistemas de informação gerenciais:**Administrando a empresa digital. 5. ed. São Paulo: Pearson Prentice Hall, 2004.

LIMA, Carlos Alberto Lorenzi. **ETL – Extração, Transformação e Carga de dados**. Disponível em: <http://litolima.com/2010/01/13/etl-extracao-transformacao-e-carga-de-dados/>. Acesso em 30 Abr. 2013

MONTINI, Alessandra. **O poder do data mining para o avanço dos negócios.** Disponível em: <http://www.administradores.com.br/noticias/administracao-e-negocios/o-poder-do-data-mining-para-o-avanco-dos-negocios/26135/>. Acesso em 30 Maio 2013

MONTINI, Alessandra. **Varejo e bancos são os setores que mais utilizam data mining.** Disponível em: <http://www.metaanalise.com.br/inteligenciademercado/index.php?option=com_content&view=article&id=6425:varejo-e-bancos-sao-os-setores-que-mais-utilizam-o-data-mining-&catid=8:carreira&Itemid=358>. Acesso em 30 Maio 2013
STATSOFT – StatSoft South América LTDA. **Soluções empresariais Avançadas.** Disponível em: <http://www.statsoft.com.br/pt/conteudo.php?con=0000000016>. Acesso em 30 Maio 2013

MOREIRA,Eduardo**. Modelo Dimensional para Data Warehouse** Disponível em: < http://imasters.com.br/artigo/3836/gerencia-de-ti/modelo-dimensional-para-data-warehouse/>. Acesso em 24 abr. 2013.

MUNIZ, Vander Emiro. **Data Mining:** conceitos e casos de uso na área da saúde. Disponível em: < http://www.devmedia.com.br/data-mining-conceitos-e-casos-de-uso-na-area-da-saude/5945 >. Acesso em 28 Abr. 2013

O' BRIEN, J. A. Sistemas de Informação e as decisões gerenciais na era daInternet. São Paulo: Saraiva, 2001.

OLIVEIRA, Déborah. **Data mining ganha espaço na estratégia empresarial.** Disponível em: <http://computerworld.uol.com.br/tecnologia/2012/03/16/data-mining-ganha-espaco-na-estrategia-empresarial/>. Acesso em 30 Maio 2013

PICHILIANI, Mauro. **Data Mining na Prática:** Classificação Bayesiana. Disponível em: <http://imasters.com.br/artigo/4926/sql-server/data-mining-na-pratica-classificacao-bayesiana/>. Acesso em 30 Abr. 2013

PICHILIANI, Mauro. **Data Mining na Prática:** Regras de Associação. Disponível em: <http://imasters.com.br/artigo/7753/sql-server/data-mining-na-pratica-regras-de-associacao/>. Acesso em 30 Abr. 2013

PICHILIANI, Mauro. **Data Mining na Prática:** Árvores de decisão. Disponível em: <http://imasters.com.br/artigo/5130/sql-server/data-mining-na-pratica-arvores-de-decisao/>. Acesso em 30 Abr. 2013

PORTO, Maria Alice. **Tomada de decisão nas organizações.**Disponível em: <http://www.artigos.com/artigos/sociais/administracao/tomadas-de-decisao-nas-organizacoes-3412/artigo/#.UbC4xPm1EmN>. Acesso em 28 Maio 2013

PRASS, Fernando Sarturi.**Uma visão geral sobre as fases doKnowledge Discovery in Databases (KDD).** Disponível em:<http://fp2.com.br/blog/index.php/2012/um-visao-geral-sobre-fases-kdd/>. Acesso em 24 Abr. 2013

RAPOSO, Marcel Antunes. **A importância do data mining na tomada de decisões.** Disponível em: <http://dbrain.com.br/2010/06/a-importancia-do-data-mining-na-tomada-de-decisoes/>. Acesso em 30 Maio 2013

RASKIN, SaraFichman. **Tomada de decisão e aprendizado organizacional**. Disponível em: <http://www.batebyte.pr.gov.br/modules/conteudo/conteudo.php?conteudo=1121>. Acesso em 28 Maio 2013

RIBEIRO, Viviane. **O que é ETL?** Disponível em: <http://vivianeribeiro1.wordpress.com/2011/06/28/o-que-e-etl-2/>. Acesso em 30 Abr. 2013

SERAIN, João Sidemar. **Por que business intelligence.** Disponível em: <http://imasters.com.br/artigo/5415/gerencia-de-ti/por-que-business-intelligence/>. Acesso em 28 abr. 2013

SILVA JUNIOR, Ovídio F. P. **Modelo de informações estratégicas aplicadas asistemas de inteligência organizacional na gestão pública de pesquisaagropecuária**: o caso da EPAGRI. 2006, 233 f. Tese (Doutorado em Engenharia deProdução) ‑ Universidade de Santa Catarina, Florianópolis, 2006.