

Universidade Federal de Minas Gerais  
Instituto de Ciências Exatas  
Departamento de Ciências da Computação

EMERSON JOSÉ ROCHA REIS

**BENEFÍCIOS DO USO DE TÉCNICAS DE MINERAÇÃO DE DADOS PARA A  
INTELIGÊNCIA COMPETITIVA DAS ORGANIZAÇÕES**

Belo Horizonte  
2011

Universidade Federal de Minas Gerais  
Instituto de Ciências Exatas  
Departamento de Ciências da Computação  
Especialização em Informática: Ênfase: Análise de Sistemas

**BENEFÍCIOS DO USO DE TÉCNICAS DE MINERAÇÃO DE  
DADOS PARA A INTELIGÊNCIA COMPETITIVA DAS  
ORGANIZAÇÕES**

por

EMERSON JOSÉ ROCHA REIS

Monografia de Final de Curso

Prof. Marcos André Gonçalves  
Orientador

Belo Horizonte  
2011

EMERSON JOSÉ ROCHA REIS

**BENEFÍCIOS DO USO DE TÉCNICAS DE MINERAÇÃO DE DADOS PARA A  
INTELIGÊNCIA COMPETITIVA DAS ORGANIZAÇÕES**

Monografia apresentada ao Curso de Especialização em Informática do Departamento de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Especialista em Informática.

Área de concentração: Análise de Sistemas

Orientador: Prof. Marcos André Gonçalves

Belo Horizonte  
2011

Aos meus pais, aos meus avós e às minhas irmãs, por serem o alicerce de minha vida, e também à minha namorada Nadielle, por sempre me incentivar e ajudar a enfrentar todas as dificuldades.

## **AGRADECIMENTOS**

Primeiramente agradeço a Deus, pela saúde e inteligência, sem as quais eu não teria chegado até aqui.

Agradeço aos meus pais, meus avós e minhas irmãs, pelo apoio incondicional, pela educação, pelo respeito, carinho e amor que me foram transmitidos em todos os gestos, me ajudando a transpor todos os obstáculos que encontrei em meu caminho.

Agradeço a minha namorada Nadielle pelo amor, carinho, paciência e compreensão nos momentos difíceis, e também pelos momentos alegres de paz e cumplicidade.

Agradeço aos colegas de curso, que apesar do curto período de convivência, foi possível trocar boas experiências.

Agradeço ao Professor Marcos André Gonçalves, pela orientação neste trabalho.

Enfim, agradeço a todos que, direta ou indiretamente contribuíram para a conclusão de mais esta etapa.

*“O único caminho para desvendar os limites do possível é aventurar-se além dele,  
através do impossível.”  
(Arthur C. Clarke)*

## RESUMO

No mercado competitivo da atualidade, as organizações buscam ser mais inteligentes a partir da análise das informações dos seus ambientes interno e externo, visando obter vantagem sobre seus concorrentes. A Inteligência Competitiva tem o propósito de realizar essa monitoração informacional permitindo às empresas agirem de forma antecipada e reduzir os riscos de operação. Os Sistemas de Apoio à Decisão, dos quais faz parte a Mineração de Dados têm grande potencial para dar suporte a essas atividades, facilitando no processamento, obtenção e disseminação de informações gerenciais, uma vez que a Mineração de Dados tem o intuito de descobrir conhecimento não implícito em bancos de dados. Este trabalho tem por objetivo apontar os benefícios do uso de técnicas de Mineração de Dados para a Inteligência Competitiva das organizações, verificando ainda o valor agregado dos resultados obtidos e os custos de tal processo. Buscou-se reunir um conteúdo que possibilitasse verificar se é compensador investir em tal tecnologia para apoiar as atividades de Inteligência Competitiva. Para tanto, foi feita, primeiramente, uma pesquisa bibliográfica qualitativa visando embasar o conhecimento de ambos os temas. Em seguida realizou-se uma busca por trabalhos acadêmicos que relatassem a aplicação de técnicas de Mineração de Dados em empresas reais para, a partir do material coletado, efetuar uma análise crítica com o intuito de responder aos questionamentos do trabalho. Foram analisadas questões como porte da empresa, metodologia utilizada no processo, e natureza das informações geradas. Ao final, apesar das dificuldades encontradas, foi possível apontar os benefícios investigados, e adicionalmente apresentar recomendações para a realização de um processo de Descoberta de Conhecimento de sucesso.

**Palavras-chave:** Inteligência Competitiva, Gestão do Conhecimento, Mineração de Dados, Descoberta de Conhecimento em Banco de Dados.

## **ABSTRACT**

In today's competitive marketplace, organizations seek to be more intelligent based in analysis of information from its internal and external environments, in order to gain an advantage over their competitors. The Competitive Intelligence has the purpose of conducting this monitoring informational allowing companies to act in advance and reduce the risks of operation. The Decision Support Systems, which include the Data Mining, has great potential to support these activities, facilitating the processing, retrieval and dissemination of management information, since the Data Mining aims to discover knowledge not implicit in databases. This paper aims to point out the benefits of using Data Mining techniques for Competitive Intelligence of organizations, checking the value of the results obtained and the costs of such a process. It sought to bring together content that enables to check whether it is worthwhile investing in such technology to support the activities of Competitive Intelligence. Thus, it was made first, a qualitative literature search for based the knowledge of both subjects. Then a search was made for academic papers to report the application of data mining techniques in real companies, for make a critical analysis of the collected material, in order to answer the questions of labor. It cover such issues as company size, the methodology used in the process and nature of the information generated. In the end, despite the difficulties, it was possible to point out the benefits investigated further and make recommendations for the implementation of a successful process of knowledge discovery.

**Keywords:** Competitive Intelligence, Knowledge Management, Data Mining, Knowledge Discovery in Databases.



## LISTA DE FIGURAS

FIG. 1	O Processo de Inteligência Competitiva nas Organizações.....	18
FIG. 2	O Processo KDD.....	31

## **LISTA DE QUADROS**

QUADRO 1:	Comparativo entre os trabalhos analisados.....	59
-----------	--	----

## LISTA DE SIGLAS

BD	Banco de Dados
BI	<i>Business Intelligence</i>
DW	<i>Data Warehouse</i>
GC	Gestão do Conhecimento
IC	Inteligência Competitiva
KDD	<i>Knowledge Discovery in Database</i>
MD	Mineração de Dados
OLAP	<i>On-Line Analytical Processing</i>
SAD	Sistema de Apoio à Decisão
SGBD	Sistema Gerenciador de Banco de Dados
TI	Tecnologia da Informação
VC	Vantagem Competitiva
UML	<i>Unified Modeling Language</i>

## SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>13</b>
1.1 Motivação .....	14
1.2 Objetivos .....	15
1.2.1 Objetivo Geral .....	16
1.2.2 Objetivos Específicos .....	16
1.3 Metodologia .....	16
1.4 Estrutura do Trabalho .....	17
<b>2. INTELIGÊNCIA COMPETITIVA NAS ORGANIZAÇÕES .....</b>	<b>19</b>
2.1 Gestão do Conhecimento .....	19
2.2 Inteligência Competitiva .....	20
2.3 Vantagem Competitiva .....	23
2.4 Inteligência Competitiva e a Tecnologia da Informação .....	24
<b>3. MINERAÇÃO DE DADOS E O PROCESSO DE DESCOBERTA DE CONHECIMENTO .....</b>	<b>27</b>
3.1 O Processo de Descoberta de Conhecimento em Bancos de Dados .....	27
3.2 As etapas do Processo KDD .....	28
3.2.1 Pré-Processamento .....	28
3.2.2 Mineração de Dados .....	30
3.2.3 Pós-Processamento .....	33
3.3 A Mineração de dados e a Inteligência Competitiva .....	34
<b>4. CASOS DE APLICAÇÃO DA MINERAÇÃO DE DADOS E OS BENEFÍCIOS PARA A INTELIGÊNCIA COMPETITIVA .....</b>	<b>36</b>
4.1 Critérios da pesquisa .....	36
4.1.1 Seleção dos casos de estudo .....	36
4.1.2 Critérios para análise do material .....	37
4.2 Caso de Estudo 1: Site de Comércio Eletrônico .....	38
4.2.1 O trabalho .....	38
4.2.2 A empresa .....	39
4.2.3 O Processo de Mineração de Dados .....	39
4.2.3.1 Metodologia .....	40
4.2.3.2 Tarefas, Técnicas e Algoritmos .....	42

4.2.3.3 <i>Hardware e Software</i> .....	43
4.2.3.4 Custos .....	43
4.3.3.5 Resultados.....	44
4.3 Caso de Estudo 2: Loja de Materiais para Construção .....	45
4.3.1 O trabalho .....	46
4.3.2 A empresa .....	46
4.3.3 O Processo de Mineração de Dados .....	47
4.3.3.1 Metodologia .....	47
4.3.3.2 Tarefas, Técnicas e Algoritmos .....	50
4.3.3.3 <i>Hardware e Software</i> .....	50
4.3.3.4 Custos .....	51
4.3.3.5 Resultados.....	51
4.4 Caso de Estudo 3: Empresa de abastecimento de água .....	54
4.4.1 O trabalho .....	54
4.4.2 A empresa .....	55
4.4.3 O Processo de Mineração de Dados .....	56
4.4.3.1 Metodologia .....	56
4.4.3.2 Tarefas, Técnicas e Algoritmos .....	58
4.4.3.3 <i>Hardware e Software</i> .....	58
4.4.3.4 Custos .....	59
4.4.3.5 Resultados.....	59
4.5. Análise dos Casos de Estudo .....	61
4.6 Benefícios da Mineração de Dados para a Inteligência Competitiva .....	65
<b>5. CONCLUSÃO .....</b>	<b>72</b>
5.1 Considerações Finais.....	72
5.1 Trabalhos Futuros .....	73
<b>REFERÊNCIAS.....</b>	<b>75</b>

## 1. INTRODUÇÃO

A informação é crucial para que uma organização sobreviva, seja para conquistar novos clientes, seja para se sobressair sobre os concorrentes, ou ainda para melhorar seus processos. Informação passou a ser um recurso da empresa, tal qual os recursos humanos, materiais e financeiros, e juntamente com o conhecimento científico e tecnológico tornou-se o centro das atenções das empresas em novos projetos (VALENTIM *et. al.*, 2003 *apud* CARVALHO, 2001).

Assim, é de suma importância a correta manipulação e gerência dessa informação, e de mecanismos que consigam usá-la de maneira inteligente de forma a aumentar a competitividade da empresa. Surge assim a Inteligência Competitiva (IC), que visa o uso inteligente das informações internas e externas à organização em benefício da empresa para se sobressair no seu ambiente.

A Inteligência Competitiva é, então, o processo que tem o propósito de investigar o ambiente interno e externo da organização, coletando informações de modo a reduzir os riscos, agir antecipadamente para proteger a organização, e estabelecer estratégias de ação a curto, médio e longo prazo (Adaptado de VALENTIM *et. al.*, 2003 e ABRAIC, 2011).

Uma empresa capaz de gerir suas informações transformando-as em conhecimento de valor agregado possui um diferencial muito importante, que a torna mais competitiva diante dos concorrentes (AZARIAS P.; MATOS S. N.; CARVALHO H. G.).

Sendo assim, uma empresa sem IC não consegue sobreviver em um mercado competitivo e volátil como o de hoje.

Porém, para um melhor gerenciamento das informações de uma organização e suporte à sua IC, são necessários sistemas de informação, pois são ferramentas capazes de organizar o fluxo de informação, alertando para novas oportunidades e ameaças, ajudando no planejamento estratégico (AZARIAS P.; MATOS S. N.; CARVALHO H. G. *apud* GOMES e BRAGA, 2004).

Sistemas de Apoio à Decisão (SAD) são muito importantes no suporte ao trabalho gerencial com as informações de uma empresa, dentre os quais estão as técnicas de Mineração de Dados (MD), que se destacam não só por facilitar o

trabalho com grandes bases de dados, mas também por possibilitar a descoberta de conhecimento não implícito a partir destas bases. Isso é de grande valia para a IC da organização.

Segundo Tan P. N., Steinbach M., e Kumar V. (2009) Mineração de Dados é “o processo de descoberta automática de informações úteis em grandes depósitos de dados” com o intuito de “descobrir padrões úteis e recentes que poderiam, de outra forma, permanecer ignorados”.

Gonçalves (2001) diz que a mineração de dados é a intersecção de técnicas de inteligência artificial, aprendizado de máquina e tecnologias de banco de dados, de modo a ser utilizada na predição de valores de saída a partir de informações estatísticas obtida em uma grande base de dados.

Dessa forma, a IC e as Tecnologias da Informação (TI) estão intimamente ligadas e sua interação possibilita a geração e difusão de conhecimento inovador no contexto social e econômico da organização, que possibilita à organização obter vantagem diante dos concorrentes (VALENTIM *et. al.*, 2003).

Baseando-se nas informações apresentadas a respeito da mineração de dados, esta é tida como de grande valia para a geração de conhecimento em grandes bases de dados, representando um importante instrumento para a Gestão do Conhecimento (GC) e a IC.

Este trabalho pretende abordar as principais vantagens do uso de técnicas de MD no que diz respeito aos benefícios trazidos para a IC de uma organização, baseando-se em relatos de aplicações práticas destas técnicas em empresas reais.

Desta forma, o tema desse trabalho é interdisciplinar focado na área de Banco de Dados (BD), mais especificamente na Mineração de Dados, e também faz relação à área da Administração conhecida como Inteligência Competitiva.

## **1.1 Motivação**

Um dos mais importantes insumos de uma organização é a informação, e mais importante ainda é saber lidar com ela, de forma a aproveitá-la em benefício da empresa, para que esta sobreviva no mercado competitivo. Diversas ferramentas de apoio à decisão são indicadas para maximizar o aproveitamento das informações de

uma empresa, gerando conhecimento, de modo a beneficiá-la e obter vantagens frente aos concorrentes.

A Mineração de Dados é uma destas ferramentas, da qual muito se fala, ressaltando suas vantagens, campos de aplicação, técnicas, algoritmos utilizados, etc. No entanto, no que diz respeito ao valor agregado para a Inteligência Competitiva da organização, as informações disponibilizadas são mais escassas e as vantagens reais que justificam o investimento, são pouco claros.

A alta administração de uma organização não opta por utilizar técnicas de descoberta de conhecimento sobre suas bases de dados somente por se ouvir dizer que são boas e que trazem vantagens à empresa. Isso porque tal processo necessita de ferramentas e mão-de-obra que, provavelmente demandam alto custo, além, é claro, das dificuldades que podem ser encontradas no desenvolvimento de um projeto do tipo. Decisões como esta necessitam de uma análise mais detalhada no que tange aos benefícios reais trazidos à empresa pela utilização da Mineração de Dados para competitividade da empresa frente ao mercado. Ao se investir em uma melhoria de processo, espera-se um retorno satisfatório.

Surge então uma oportunidade de pesquisa, cujo material produzido seja capaz de reunir casos reais de aplicação da técnica de Mineração de Dados, analisando seus benefícios para a inteligência da organização. Tal material traria informações capazes de orientar quanto à decisão de se utilizar ou não técnicas de MD na base de dados de uma empresa, tendo como exemplo casos reais de sua aplicação.

O presente trabalho objetiva tal análise tentando responder perguntas como: qual o valor agregado da informação gerada por ferramentas de mineração de dados para a inteligência competitiva da empresa? Em relação aos custos de tal processo, seria vantajoso aplicá-lo?

## **1.2 Objetivos**

Neste tópico são apresentados o objetivo geral e os objetivos específicos do presente trabalho, que nortearão todo o seu desenvolvimento.



### **1.2.1 Objetivo Geral**

O objetivo do presente trabalho é a análise dos benefícios gerados para o processo de Inteligência Competitiva pela utilização de técnicas de Mineração de Dados para a descoberta de conhecimento na base de dados de uma organização. Tal análise será feita através de pesquisa qualitativa de modo a fazer um levantamento de estudos de casos aplicados em empresas reais, verificando os resultados obtidos, de modo a sintetizar as vantagens proporcionadas para a inteligência da organização.

### **1.2.2 Objetivos Específicos**

- Entender o processo de Gestão de Conhecimento e Inteligência Competitiva nas organizações.
- Conhecer o Processo de Descoberta de Conhecimento em Bancos de Dados, as técnicas de Mineração de Dados e sua aplicabilidade.
- Pesquisar e estudar casos reais de aplicação de técnicas de Mineração de Dados contidos na literatura.
- Fazer uma análise dos casos reais de utilização da Mineração de Dados identificando os benefícios gerados no que tange à Inteligência Competitiva da organização.

### **1.3 Metodologia**

A metodologia de desenvolvimento deste trabalho será embasada em pesquisa bibliográfica qualitativa. Para tanto, inicialmente será pesquisado e abordado sobre Inteligência Competitiva e Gestão do Conhecimento nas organizações, também conceituando o que vem a ser Vantagem Competitiva, bem como a relação desses conceitos com a Tecnologia da Informação.

Em seguida o foco da pesquisa será nas técnicas de Mineração de Dados e o Processo de Descoberta de Conhecimento em Bancos de Dados e sua aplicabilidade.

Por fim, a pesquisa será mais aprofundada no que diz respeito aos casos de aplicação das técnicas e Mineração de Dados em bases de dados de empresas reais, de forma a possibilitar uma análise crítica do material coletado que demonstre os reais benefícios deste processo para a Inteligência Competitiva de uma organização. Questões como natureza das informações geradas e sua utilidade, custos do processo (*hardware*<sup>1</sup>, *software*<sup>2</sup>, mão-de-obra), técnicas utilizadas, vantagens e desvantagens, serão o foco deste estudo.

#### 1.4 Estrutura do Trabalho

Este trabalho está estruturado da seguinte forma:

- Capítulo 1: O presente capítulo trata da introdução ao tema do trabalho, apresentando a motivação para o desenvolvimento da pesquisa, seus objetivos principais e específicos, bem como a metodologia adotada para o desenvolvimento do mesmo.
- Capítulo 2: Faz uma revisão bibliográfica abordando a Inteligência Competitiva nas organizações. Para tanto são conceituados a Gestão do Conhecimento, Inteligência Competitiva e Vantagem Competitiva. Ao final também é abordado a relação entre Inteligência Competitiva e as Tecnologias da Informação.
- Capítulo 3: Neste capítulo é abordada, através de revisão bibliográfica, a Mineração e Dados enfatizando o processo do qual faz parte, o Processo de Descoberta de Conhecimento em Bancos de Dados, bem como suas etapas: Pré-processamento, a própria Mineração de Dados, e o Pós-processamento.
- Capítulo 4: Apresenta, por meio de pesquisa qualitativa, uma descrição dos trabalhos de aplicação de Mineração de Dados em empresas reais selecionados para estudo, bem como é feita uma análise crítica dos mesmos, confrontando as informações e conhecimento obtidos com as vantagens proporcionadas à Inteligência Competitiva das referidas empresas.

---

<sup>1</sup> *Hardware: parte física do computador, compreendendo equipamentos, circuitos e componentes que permitem o processamento computacional.*

<sup>2</sup> *Software: parte lógica do computador, sendo os programas que permite ao usuário realizar uma determinada tarefa.*

- Capítulo 5: Apresenta a conclusão do trabalho, ressaltando os resultados obtidos, considerações finais, e propostas de trabalhos futuros.
- Capítulo 6: Lista as referências bibliográficas utilizadas para a elaboração do trabalho.

## 2. INTELIGÊNCIA COMPETITIVA NAS ORGANIZAÇÕES

Este capítulo faz uma revisão bibliográfica sobre a Inteligência Competitiva, abordando sua relação com a Gestão do Conhecimento (GC) nas organizações, e ressaltando também como os dois conceitos possibilitam a obtenção de Vantagem Competitiva (VC) no ambiente em que as mesmas estão inseridas. Além disso, o uso de Tecnologias da Informação é evidenciado para possibilitar a implementação desses processos.

### 2.1 Gestão do Conhecimento

Para entender melhor o que vem a ser Inteligência Competitiva, é preciso também conceituar Gestão do Conhecimento, visto que estes conceitos têm um íntimo relacionamento.

Conforme Valentim *et. al.* (2003), “a gestão do conhecimento é uma das bases que amparam o processo de inteligência competitiva nas organizações”, que quando baseada na coletividade entre os indivíduos componentes da organização, possibilita a obtenção de vantagem sobre a concorrência e “potencializa a exploração de novas idéias para fomentar a inovação.” Terra, sendo mais abrangente diz que a Gestão do Conhecimento e a Inteligência Competitiva se completam.

Valentim diz ainda que a Gestão do Conhecimento está relacionada ao capital intelectual e aos fluxos informais da organização, enquanto que os fluxos formais são abordados pela Gestão da Informação. A Inteligência Competitiva trabalha com as duas visões, focando na estratégica da empresa.

Carvalho e Terra falam que um Sistema de Gestão do Conhecimento tem uma perspectiva interna, visa o controle das informações e conhecimentos gerados dentro da própria empresa, enquanto que um Sistema de Inteligência Competitiva tem uma perspectiva externa, olha para o mundo do lado de fora, buscando entender os seus desafios competitivos. Carvalho completa ainda que “um Sistema de Gestão do Conhecimento facilita o fluxo e a criação de conhecimento, permitindo, assim, que a inteligência também consiga ser gerada mais facilmente”.

Dessa forma, Gestão do Conhecimento é “a capacidade que as mesmas (organizações) possuem, através de seus processos, de criar e utilizar o conhecimento” (AZARIAS P.; MATOS S. N.; CARVALHO H. G. e TERRA), sendo considerada “a arte de gerar valor a partir de bens intangíveis da organização” (CARVALHO *apud* SVEIBY, 1998, e DRUCKER, 1993).

Reforçando essa definição, Antoniazzi E. Ap. e Stefano S. R. (2007) definem que a função da gestão do conhecimento é “organizar as principais políticas, processos e ferramentas gerenciais e tecnologias” com o intuito de melhor assimilar os processos de geração de conhecimento “para gerar resultados para a organização.”

Assim, a Gestão do Conhecimento “é utilizada com o objetivo de fornecer a informação para as pessoas certas no momento certo, auxiliando-as na tomada de decisão, de maneira a aumentar o desempenho da organização” (AZARIAS P.; MATOS S. N.; CARVALHO H. G. *apud* ROCHA, 2005). Então, é muito importante o gerenciamento das informações da mesma forma que os outros recursos da organização para que estas sejam utilizadas com mais eficácia (VALENTIM *et. al.*, 2003).

A Gestão do Conhecimento tem por finalidade, portanto, o controle do fluxo de informações internas da empresa visando entender os processos de geração de conhecimento para permitir a tomada de decisão, representando também um insumo para o processo de Inteligência Competitiva. É perceptível que isso só é possível com o auxílio de Tecnologia da Informação, como os Sistemas de Apoio à Decisão.

## **2.2 Inteligência Competitiva**

O valor das informações e seu gerenciamento através da Gestão do Conhecimento é uma preocupação cada vez maior das empresas, que buscam estruturar o conhecimento que possuem, juntamente com a coleta de informações ambientais externas, por meio de práticas como a Inteligência Competitiva, visando um melhor posicionamento da organização no seu ambiente competitivo (AZARIAS P.; MATOS S. N.; CARVALHO H. G.).

Segundo Valentim *et. al.* (2003):

*“A inteligência competitiva é o processo que investiga o ambiente onde a empresa está inserida, com o propósito de descobrir oportunidades e reduzir os riscos, bem como diagnostica o ambiente interno organizacional, visando o estabelecimento de estratégias de ação a curto, médio e longo prazo.”*

Valentim propõe através da FIG. 1, um esquema do processo de Inteligência Competitiva que condiz com sua definição, facilitando a compreensão.



FIG. 1: O Processo de Inteligência Competitiva nas Organizações

Fonte: (VALENTIM)

A ABRAIC (2011) completa essa definição dizendo que a Inteligência Competitiva é um processo que melhora a tomada de decisão estratégica e operacional, minimizando os riscos, agindo antecipadamente, para proteger a organização. A mesma visão é compartilhada por diversos outros autores como Azarias P.; Matos S. N.; Carvalho H. G. (*apud* GOMES e BRAGA, 2004), e Antoniazzi E. Ap. e Stefano S. R. (2007), que citam ainda que a Inteligência Competitiva possibilita a geração de recomendações que consideram eventos futuros, e não somente relatórios, que simplesmente justificam decisões passadas.

Segundo Antoniazzi E. Ap. e Stefano S. R. (2007) é necessário também estar atento à concorrência e às novas tecnologias, de maneira que a empresa aja antecipadamente diante de ameaças e oportunidades, permitindo melhorar sua posição competitiva.

A Inteligência Competitiva tem tido muita importância ultimamente, visto que as empresas atualmente estão inseridas em um ambiente de incerteza dos mercados, tornando difícil a tomada de decisão, que quando feita de forma errada podem acarretar grandes prejuízos. Um Sistema de Inteligência Competitiva pode auxiliar as empresas a lidar com esse ambiente tão adverso, permitindo tomar decisões mais seguras e em tempo real (ABRAIC, 2011).

De acordo com Valentim, a inteligência competitiva necessita do mapeamento e coleta de dados, informações e conhecimento produzidos internamente e externamente à organização de forma ágil, o que permite dar maior segurança às estratégias estabelecidas pela organização. Dessa forma o processo de Inteligência Competitiva é, então, de “fundamental importância para que as organizações sintam-se capazes de atuarem no mundo globalizado”.

Assim, o processo de Inteligência Competitiva possui duas etapas principais Valentim *et. al.* (2003):

- Prospecção Informacional: é a técnica que identifica dados, informação e conhecimento relevantes para a organização;
- Monitoramento Informacional: é o mecanismo de observação e acompanhamento constante dos dados prospectados.

Através da prospecção informacional é possível levantar uma fonte de dados essenciais ao negócio da empresa, que alimentam os sistemas de Inteligência Competitiva, possibilitando a obtenção de conhecimento valioso à tomada de decisão (Valentim *et. al.*, 2003).

Outros autores, como Carvalho (*apud* JOHNSON, 1998) também definem a inteligência competitiva como um produto, que pode ser de três tipos:

- Inteligência Estratégica, quando dá apoio às decisões estratégicas, como alianças e parcerias, aquisições ou fusões;
- Inteligência Tática, quando dá apoio às atividades do dia-a-dia da empresa, como busca de vantagens de mercado, novos fornecedores, novos contratos, etc.

- Contra Inteligência (ou Inteligência Defensiva), quando visa defender o bem intelectual da organização, o conhecimento que a torna competitiva.

Terra propõem algumas etapas para nortear o processo de implementação da Gestão de Conhecimento juntamente com a Inteligência Competitiva. Tais processos, segundo ele, são essenciais para o sucesso desse processo:

- Alinhamento com os objetivos estratégicos da empresa;
- Forte mudança cultural;
- Implementação de novos processos informais e formais;
- Uso de tecnologias de informação e comunicação;
- Governança bem articulada e estruturada;
- Envolvimento de Pessoas Experientes;
- Medição de resultados.

Algumas metodologias e ferramentas também podem e devem ser utilizadas para a implementação de um Sistemas de Inteligência Competitiva, como: i) Forças de Porter; ii) Fatores Críticos de Sucesso; iii) Análise SWOT; iv) *Balanced Scorecard*; e v) Benchmarking (ANTONIAZZI E. Ap. e STEFANO S. R, 2007).

Portanto, entende-se por Inteligência Competitiva o processo organizacional implementado com o objetivo de monitorar os fluxos de dados, informações e conhecimentos obtidos no ambiente interno e externo da mesma, de maneira a gerar conhecimento útil para seu posicionamento competitivo e antecipado diante de ameaças e oportunidades externas, o que pode vir a representar vantagem frente a seus concorrentes.

### **2.3 Vantagem Competitiva**

Quando se fala que a Inteligência competitiva pode trazer vantagem a uma organização, então logo surge o termo Vantagem Competitiva, também importante ser conceituado.

Conforme BARNEY (2007) “uma empresa possui vantagem competitiva quando é capaz de gerar maior valor econômico do que empresas rivais”, e explica esse valor econômico como sendo a diferença entre os benefícios proporcionados ao cliente pelo produto ou serviço e o custo de sua produção, podendo essa vantagem ser temporária ou sustentável.



Ainda, segundo BARNEY (2007), a vantagem competitiva pode ser medida através do desempenho contábil, e também econômico de uma empresa. A primeira medida está relacionada à análise das informações contidas nos demonstrativos e balanços da empresa, e a segunda relativa aos ganhos acima do seu custo de capital, isto é, a empresa ser capaz de satisfazer os interesses econômicos de seus credores e acionistas.

PORTER (1992, pag. 31 e 39) fala que a Vantagem Competitiva é tida nas diversas atividades desempenhadas dentro de uma empresa relacionadas ao projeto, produção, *marketing*<sup>3</sup>, e à entrega e suporte de seu produto. Afirma também que o desenvolvimento e mudanças de tecnologias, e o modo como são empregadas, são muito importantes para a obtenção de vantagem competitiva. Tais tecnologias podem alterar o custo relativo de produção de produtos e serviços e ainda possibilitar a sua diferenciação, sendo que tal processo, em algumas empresas, é vital.

Dessa forma, pode-se concluir que a Gestão do Conhecimento se bem desenvolvida, ampara um processo de Inteligência Competitiva em uma empresa, capaz de gerar, entre outros benefícios, Vantagem Competitiva à mesma para poder sobreviver e crescer no mercado imprevisível e globalizado de hoje.

## **2.4 Inteligência Competitiva e a Tecnologia da Informação**

A origem da Inteligência Competitiva vem dos métodos que órgãos governamentais tinham para identificar e avaliar informações relacionadas à Defesa Nacional. A partir daí foram adaptadas para serem usados no meio empresarial sendo incorporadas técnicas da Ciência da Informação, Tecnologia da Informação, e pela Administração (ABRAIC, 2011).

A Tecnologia da Informação é, então, importante aliado da Inteligência Competitiva, pois sendo um processo que lida com grande volume de informações, seu processamento manual é difícil, necessitando de ferramentas que auxiliem nessa tarefa, ferramentas essas capazes de impulsionar a estratégia da empresa

---

<sup>3</sup> Marketing: Termo da Administração; significa um conjunto de atividades que visam maximizar as vendas de produtos e serviços ao consumidor, usando estratégias diversas como estudo de mercado, promoções, publicidades e assistência pós-venda.

em vantagem competitiva sustentável (AZARIAS P.; MATOS S. N.; CARVALHO H. G.).

Azarias P.; Matos S. N.; Carvalho H. G. confirmam essa visão dizendo que a organização que é capaz de utilizar tecnologia para transformar dados em informações e conhecimento de valor para a tomada de decisão é considerada uma organização inteligente. Valentim *et. al.* (2003 *apud* TEIXEIRA FILHO, 2001) completa afirmando que a empresa que conseguir melhor perceber o potencial de utilização de tecnologias emergentes alcançará maior vantagem competitiva em seu setor de atuação.

Levando em conta a divisão da Inteligência Competitiva em dois processos: prospecção e monitoramento organizacional, as Tecnologias da Informação são muito úteis e realizam esses processos de forma ágil e eficaz (VALENTIM *et. al.*, 2003).

Dessa forma, a inteligência competitiva e as tecnologias de informação estão intimamente ligadas e sua interação possibilita a geração e difusão de conhecimento inovador no contexto social e econômico da organização (VALENTIM *et al*, 2003). O uso de tecnologias deve permitir ainda, que essas inovações sejam obtidas antes dos concorrentes para que possam resultar em alguns casos, em vantagens insuperáveis (MACEDO D. C.; REZENDE, M. A. *apud* TOLEDO, 2007).

No entanto, apesar das tecnologias disponíveis, executivos tem tido receio em tomar decisões erradas devido ao grande volume de informações (VALENTIM *et. al.* 2003).

O grande receio é a incerteza dos benefícios advindos do investimento em tecnologia, pois são resultados subjetivos e de difícil verificação, necessitando de um amplo gerenciamento dos sistemas de informações de modo a alcançar resultados mais efetivos (MONTGOMERY C. A.; PORPTER M. E., 1998, pag. 87).

O desafio está na correta utilização de tecnologias através de metodologias e técnicas, e profissionais capacitados para que os benefícios possíveis na teoria sejam obtidos na prática. Dziekaniak (2010) também diz que apesar do crescente uso de tecnologia da informação não garante o sucesso, pois o conhecimento gerando por elas deve ser usado de forma competente pelos gestores.

Portanto, é evidente a necessidade de sistemas de informação, como os sistemas de apoio à decisão e *Business Intelligence* (BI), para a utilização inteligente

das informações de uma organização. Porém mais importante que utilizar esse tipo de tecnologia é sua correta utilização, de modo a dar o devido suporte ao processo de Inteligência Competitiva. A organização deve ser capaz de aprender com o conhecimento gerado por ferramentas como a Mineração de Dados, para justificar seu investimento.

### **3. MINERAÇÃO DE DADOS E O PROCESSO DE DESCOBERTA DE CONHECIMENTO**

Este capítulo aborda, a partir de uma revisão bibliográfica rápida, as principais características da Mineração de Dados, que faz parte de um processo maior denominado Descoberta de Conhecimento em Banco de Dados. Por fim, é feito um cruzamento entre os conceitos de Mineração de Dados e Inteligência Competitiva, no que diz respeito ao potencial de utilização de tais técnicas para amparar o processo de tomada de decisão nas organizações.

#### **3.1 O Processo de Descoberta de Conhecimento em Bancos de Dados**

Os grandes avanços da tecnologia, principalmente na coleta e armazenagem de dados, fizeram com que as empresas acumulassem uma enorme quantidade de dados. Pelo fato da magnitude dessas bases, tem sido muito desafiadora a extração de informações úteis, uma vez que as técnicas tradicionais de análise de dados não são capazes de lidar com grandes conjuntos de dados (TAN P. N.; STEINBACH M.; KUMAR V., 2009).

Áreas como Ciência, Medicina e Engenharia, que acumulam dados muito rapidamente, logo focaram no desenvolvimento de ferramentas eficientes e escaláveis capazes de lidar com diferentes tipos de dados, em grandes quantidades. Surgiram então as técnicas de Mineração de Dados, que utilizam métodos tradicionais combinados com algoritmos e técnicas sofisticados, como estatística e inteligência artificial, com o objetivo de tornar possível processar grandes volumes de dados. Isso permite o uso dos dados processados por diversas aplicações de inteligência de negócios (TAN P. N.; STEINBACH M.; KUMAR V., 2009).

“Sistemas de conhecimento têm se tornado cada vez mais utilizados e necessários frente ao aumento constante de dados e informações, bem como do grande volume gerado e gerido pelas empresas e organizações” (DIZIEKANIAK, 2010). Dessa forma métodos e técnicas foram criados para serem capazes de trabalhar com grande quantidade de dados extraindo conhecimento útil às

organizações para a tomada de decisão, dentre as quais a Mineração de Dados está inclusa.

Para entender o que é Mineração de Dados e como funciona, é preciso também compreender o processo do qual faz parte, o Processo de Descoberta de Conhecimento em Banco de Dados.

O Processo de Descoberta de Conhecimento em Banco de Dados, conhecido como KDD (do inglês *Knowledge Discovery in Databases*), “é o processo geral de conversão de dados brutos em informações úteis” (TAN P. N.; STEINBACH M.; KUMAR V., 2009).

Conforme Tan P. N., Steinbach M., Kumar V. (2009) e Azarias P.; Matos S. N.; Carvalho H. G. (*apud* GOLDSCHMIDT e PASSOS, 2005), o processo KDD é dividido em três etapas: Pré-processamento, Mineração de Dados e Pós-processamento.

A Mineração de Dados é a etapa central do processo KDD, onde efetivamente ferramentas, técnicas e algoritmos agem sobre as bases de dados vasculhando por padrões entre os atributos de forma a descobrir conhecimento.

As etapas do processo KDD serão abordadas de forma mais abrangente nos próximos tópicos.

### **3.2 As etapas do Processo KDD**

Este tópico descreve as três etapas que constituem o Processo de Descoberta de Conhecimento em Banco de Dados – KDD, enfatizando sua etapa principal, a Mineração de Dados.

#### **3.2.1 Pré-Processamento**

Esta etapa é responsável pela coleta, organização e tratamento dos dados, com o propósito de prepará-los para a etapa seguinte: a mineração de dados. Pelo fato da diversidade de formas como os dados podem ser captados e armazenados, esta etapa é considerada a mais trabalhosa do processo KDD (TAN P. N.; STEINBACH M.; KUMAR V., 2009).

É também nesta etapa que se define o contexto do problema ou o que se espera conseguir com a aplicação da mineração de dados, definindo uma estratégia de ataque.

Barbosa (2007) diz que se nesta etapa os dados forem captados a partir de *Data Warehouses* (DW), o trabalho é facilitado, pois de certa forma já foi feita uma limpeza nos mesmos e estão armazenados de maneira eficiente. Segundo Barbosa (2007 *apud* INMON, 2005), *Data Warehouse* é um conjunto de dados orientados por assunto criado a partir de diversos bancos de dados operacionais, que pode ser chamado de *Data Mart* caso apóie somente uma área específica de uma empresa, contendo informações históricas, com o intuito de apoiar decisões gerenciais,

Desta forma, o pré-processamento possui quatro etapas (CORRÊA, 2007):

- Seleção dos dados: Tem por finalidade identificar os registros e atributos relevantes ao processo KDD em face do problema a ser resolvido, captando-os a partir das diversas fontes e colocando-os em uma mesma base.
- Limpeza dos dados: Objetiva trabalhar, sobre a base reunida, os dados faltantes, ruidosos ou inconsistentes, tratando-os manualmente, ou se for o caso, eliminando-os.
- Codificação dos dados: Consiste em adequar os formatos dos dados reunidos às ferramentas de mineração de dados que serão utilizadas na etapa seguinte.
- Enriquecimento dos dados: Nesta última sub-etapa o objetivo é melhorar as informações da base reunida, através da agregação de novos atributos obtidos a partir dos já existentes, como, por exemplo, adicionando totalizadores de variáveis numéricas e a criação de faixas de valores para atributos contínuos.

“O sucesso ou fracasso de um processo de descoberta de conhecimento pode, muitas vezes, estar associado ao tempo e aos critérios utilizados nesta etapa de preparação” (CORRÊA, 2007). Portanto, é preciso muito empenho na realização do pré-processamento dos dados.

### 3.2.2 Mineração de Dados

Na etapa de Mineração de Dados ou *Data Mining*, no inglês, é feita uma busca de conhecimento sobre os dados preparados na etapa anterior, através de técnicas e algoritmos específicos. Primeiro se escolhe a tarefa de mineração a ser realizada, em seguida a técnica, que por consequência implica na escolha de um algoritmo para a realização dos trabalhos.

Mas o que vem a ser realmente a Mineração de Dados?

Existem muitas definições para o termo Mineração de Dados na literatura. Dependendo da abordagem, estas definições podem focar mais nas tecnologias usadas, área de pesquisa Estatística ou Inteligência Artificial, ou mesmo no próprio processo de descoberta de conhecimento (CORRÊA, 2007).

Algumas destas definições abordam a Mineração de Dados como sendo o próprio processo de descoberta de conhecimento (CORRÊA, 2007). O processo KDD e a Mineração de Dados foram considerados termos sinônimos por diversos autores até 1995 (BARBOSA, 2007). Outras definições mais atuais já consideram a Mineração de Dados como uma etapa do processo KDD (CORRÊA, 2007).

Conforme Tan P. N., Steinbach M., Kumar V. (2009):

*“A mineração de dados é um processo de descoberta automática de informações úteis em grandes depósitos de dados, com o intuito de descobrir padrões úteis e recentes que poderiam de outra forma permanecer ignorados.”*

Gonçalves (2001) e Valentim *et. al.* (2003) também compartilham dessa visão, sendo que o primeiro ainda acrescenta que a mineração de dados é uma intersecção entre inteligência artificial, aprendizado de máquina e tecnologias de banco de dados, sendo capazes, muitas vezes de prever resultados futuros a partir de uma gama de dados de entrada.

O principal objetivo da Mineração de Dados é, então, a descoberta de conhecimento, usando uma metodologia com o intuito de extrair informações preditivas das bases de dados (BARBOSA, 2007).

O método de descoberta de conhecimento deve ser flexível, eficiente e genérico, e a Mineração de Dados em si, deve ser compreensível por usuários humanos, trazer resultado corretos tanto quanto possível, e ser surpreendente no

que tange a descoberta de conhecimento novo, interessante e útil (BARBOSA, 2007 *apud* FREITAS, 2000).

Para isso, a Mineração de Dados conta com diversas técnicas e algoritmos, uma vez que esse processo está intimamente relacionado à Estatística, Inteligência Artificial e Aprendizado de Máquina. Alguns exemplos podem ser: árvores de decisão, redes neurais, sistemas baseados em regras e programas estatísticos, podendo ser usados isoladamente ou combinados (BARBOSA, 2007).

Tan P. N., Steinbach M., Kumar V. (2009) falam que essas técnicas podem desempenhar quatro tarefas principais, detalhadas a seguir:

- Modelagem de previsão: Consiste em prever o valor de uma variável alvo baseando-se em uma função das variáveis explicativas, isto é, construir um modelo que obtenha um valor para a variável a ser prevista com menor índice de erro possível, a partir da análise de variáveis históricas disponíveis. Essa modelagem possui dois tipos: classificação, quando a variável alvo é binária, e regressão, quando a variável alvo possui valor contínuo. A modelagem de previsão pode ser usada, por exemplo, para identificar quais clientes responderão a uma campanha de *marketing*, ou determinar perturbações no ecossistema da Terra, ou ainda julgar a probabilidade de um paciente ter uma doença.
- Análise de associação: Objetiva identificar padrões em que as características das variáveis analisadas estão altamente associadas, de modo a extrair os padrões mais interessantes de forma eficiente. Essa análise pode indicar que quando ocorre um determinado evento, tantos outros eventos ocorrem também, sendo usada, por exemplo, na identificação de genes que funcionam de forma associada ou na determinação de quais páginas da internet serão acessadas juntas por um usuário.
- Análise de grupo: Busca identificar grupos com características semelhantes, intimamente relacionados entre si. Isso através da análise de diversos atributos, traçando e definindo agrupamento de modo que os elementos pertencentes a eles sejam mais parecidos entre si do que com elementos dos outros grupos. A análise de grupo pode ser usada no agrupamento de clientes e para compactar dados.



- Detecção de anomalias (ou fatores estranhos): Identifica características que sejam significativamente diferentes do restante dos dados, porém evitando rotular dados normais como anômalos. Essa tarefa tem sucesso quando possui uma alta taxa de detecção de elementos anômalos e baixa taxa de alarme falso. Pode ser usada na detecção de fraudes, invasões de redes, doenças e perturbações no meio ambiente.

A escolha das técnicas de mineração de dados a serem usadas é muito relativa, dependendo da tarefa específica a ser executada e dos tipos e qualidade dos dados disponíveis (BARBOSA, 2007 e DIAS *apud* HARRISON, 1998). Essa escolha pode ser norteada seguindo duas etapas, conforme Dias (*apud* BERRY e LINOFF, 1997) sugerem: i) Traduzir o problema de negócio a ser resolvido em séries de tarefas de mineração de dados; ii) Compreender a natureza dos dados disponíveis em termos de conteúdo e tipos de campos de dados e estrutura das relações entre os registros.

A Mineração de Dados, tida como uma pesquisa sobre uma base dados, pode ser feita de três maneiras (BARBOSA, 2007):

- automatizada pelo sistema, percorrendo o banco de dados;
- direcionada, com um analista responsável pela geração de hipóteses a serem trabalhadas;
- aprendizagem supervisionada, quando um ‘professor ensina’ o sistema sobre a validade ou não de uma premissa.

No entanto, a Mineração de Dados não é uma tarefa simples. É preciso tanto recursos de *hardware* e *software*, como também mão-de-obra com grande capacitação, e, em decorrência disso, grandes investimentos financeiros, para que os resultados desse processo sejam satisfatórios.

Tan P. N., Steinbach M., Kumar V. (2009) dizem que são necessários sistemas de banco de dados robustos e técnicas de computação distribuída<sup>4</sup> e paralela de alto desempenho para dar suporte às ferramentas de Mineração de Dados.

---

<sup>4</sup> Computação distribuída: Processamento paralelo realizado por dois ou mais computadores conectados por uma rede com o objetivo de realizar uma tarefa comum, que seria impossível ou demandaria muito tempo para conclusão se feita por uma única máquina.

Dziekaniak (2010) também afirma que apesar de existirem processos que podem auxiliar na condução do processo de Mineração de Dados, sua automatização não é simples, exigindo o trabalho de profissionais com bastante conhecimento.

A Mineração de Dados é então, uma técnica de extração de conhecimento não implícito em grandes bases de dados, considerada de difícil implantação devido a exigências de grandes investimentos com equipamentos e profissionais capacitados.

Assim, a Mineração de Dados é a etapa central do processo KDD, onde diversos métodos e técnicas de tratamento de dados são utilizados para a descoberta de conhecimento relevante à organização para a tomada de decisão, que é preciso, no entanto, ser refinado posteriormente na etapa de pós-processamento.

### **3.2.3 Pós-Processamento**

Após a mineração de dados é necessário fazer o pós-processamento, que objetiva tratar o conhecimento obtido a partir da aplicação da mineração de dados, de forma a avaliar sua utilidade e assegurando que resultados relevantes sejam incorporados aos sistemas de apoio à decisão (TAN P. N.; STEINBACH M.; KUMAR V. 2009):

“O novo conhecimento só é encontrado após a interpretação dos relatórios resultantes do processo de mineração de dados” (BARBOSA, 2007).

No entanto, mesmo após a etapa de pós-processamento pode ocorrer que os resultados não foram satisfatórios, sendo, então, necessário retornar à etapa de mineração de dados, ou às etapas iniciais, para rever parâmetros ou até mesmo as técnicas que foram utilizadas, de modo a executar o processo novamente para atingir os objetivos definidos (CORRÊA, 2007).

O processo KDD é, desta forma, iterativo, pois em qualquer que seja a etapa em que se encontra o andamento do mesmo, pode ser necessário voltar às etapas anteriores, até porque o quanto antes erros ou desvios são corrigidos melhores resultados são alcançados (BARBOSA, 2007).

Assim após a abordagem de todas as etapas do processo KDD, este pode ser representado resumidamente, conforme o diagrama de atividades da UML (*Unified Modeling Language*) abaixo (FIG. 2).

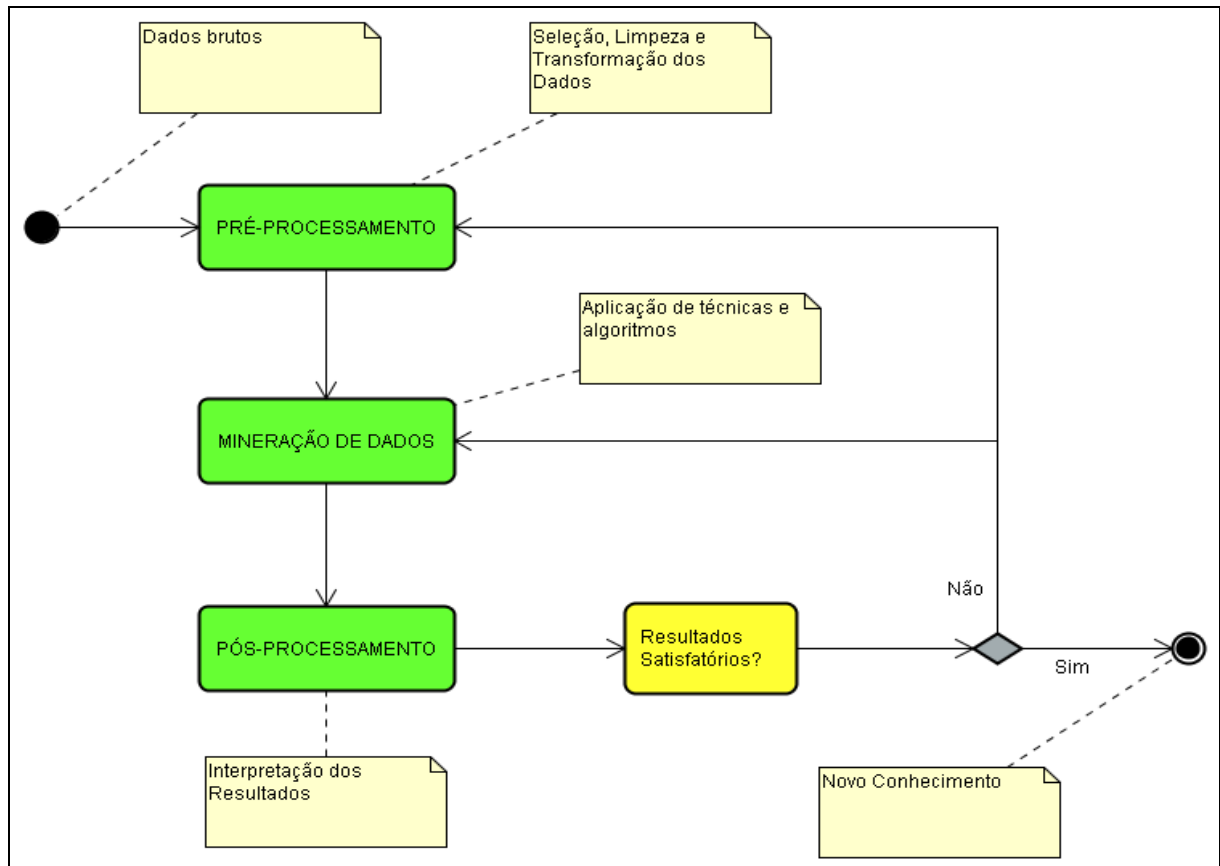


FIG. 2: O processo KDD.

Fonte: Próprio autor.

### 3.3 A Mineração de dados e a Inteligência Competitiva

Considerando que a Mineração de Dados é um processo de descoberta de conhecimento não implícito, então, no âmbito da Inteligência Competitiva, as informações obtidas a partir de tal processo são muito valiosas.

Macedo D. C. e Rezende M. A. dizem que a utilização de Mineração de Dados auxilia os tomadores de decisão no processo de Inteligência Competitiva, identificando informações relevantes de forma ágil, além de tornar a tomada de decisão mais segura e confiável, representando um diferencial competitivo para a empresa. E ainda acrescenta que organizações com um sistema de inteligência

competitiva que utiliza uma ferramenta de mineração de dados possuem vantagens sobre os concorrentes no processo de tomada de decisão, pois correm menos riscos de tomar decisões incorretas e infundadas. Além disso, esse processo torna-se mais ágil, objetivo e confiável, fundamentado sempre em dados reais.

Segundo Antoniazzi E. Ap. e Stefano S. R. (2007), a Inteligência Competitiva sendo um processo amplo, conta com muitas técnicas e ferramentas para sua realização, dentre as quais fazem partes as ferramentas de *Business Intelligence*, e conseqüentemente a Mineração de Dados.

De fato, a Mineração de Dados é tida como benéfica às organizações, mas, apesar disso, muitos desconhecem seu potencial no suporte à tomada de decisão. A pouca utilização dessa poderosa ferramenta deve-se provavelmente à dificuldade de escolha da técnica mais adequada ao contexto de uso, onde deve-se considerar vários parâmetros, como: características gerais da ferramenta, conexão a bancos de dados, critérios de desempenho computacional, critérios de funcionalidade, critérios de usabilidade, etc. (DIAS).

## **4. CASOS DE APLICAÇÃO DA MINERAÇÃO DE DADOS E OS BENEFÍCIOS PARA A INTELIGÊNCIA COMPETITIVA**

Este capítulo apresenta os benefícios do uso de técnicas de Mineração de Dados para a Inteligência Competitiva das Organizações. Para isso foi feita uma pesquisa qualitativa em busca de casos de aplicação de tais técnicas em empresas reais, de modo a analisar os resultados e o teor das informações e conhecimento gerados, avaliando suas vantagens e aplicabilidade no que diz respeito à inteligência da organização.

### **4.1 Critérios da pesquisa**

Este tópico descreve os critérios estabelecidos tanto para a pesquisa em busca de casos de aplicação da Mineração de Dados, quanto para a análise do material levantado.

#### **4.1.1 Seleção dos casos de estudo**

A pesquisa em busca de casos reais de aplicação de Mineração de Dados seguiu alguns critérios, tanto para nortear as buscas, tanto visando a análise que seria feita em seguida. Chegou-se à conclusão de que para ser feita uma análise abrangente, seria necessário reunir estudos de casos aplicados em diferentes contextos. Os critérios para a pesquisa foram os seguintes:

- Quanto ao trabalho e ao processo:
  - Trabalhos acadêmicos de universidades de destaque, de nível de mestrado e doutorado;
  - Aplicação de diferentes técnicas de mineração de dados;
  - Utilização de diferentes algoritmos e *softwares*;
- Quanto às empresas foco do estudo:
  - Empresas de diferentes portes;
  - Empresas com diferentes atividades econômicas;

- Diferentes contextos organizacionais;

Esses critérios guiaram todo o processo de pesquisa de forma a reunir um bom material de estudo.

#### **4.1.2 Critérios para análise do material**

Após a realização de uma extensa pesquisa, foram selecionados e analisados três trabalhos acadêmicos de grandes faculdades do Brasil, aplicados em empresas de portes e ramos econômicos diferentes, onde foram utilizados técnicas e *softwares* variados, em contextos diversos.

Para a análise dos trabalhos, também foram adotados uma série de etapas de maneira a sintetizar as informações dos mesmos, focando nos resultados obtidos, de modo a confrontá-los com a inteligência competitiva adquirida pelas empresas foco do referido estudo. Portanto, seguiram-se os seguintes passos para a análise e síntese do material:

- Descrição do trabalho: Foco no autor, nível do trabalho (mestrado ou doutorado), universidade onde foi desenvolvido, de modo a expor a sua credibilidade, além de abordar a motivação que levou ao seu desenvolvimento, seus objetivos e expectativas.
- Empresa foco do estudo de caso do trabalho: Aborda questões como porte da empresa, atividade econômica, local de atuação, ambiente de mercado em que está inserida, e a situação das bases de dados da empresa, ressaltando suas necessidades que se esperam ser satisfeitas com a realização do trabalho.
- Processo de Mineração de Dados: Esta etapa visa a descrição do trabalho de Mineração de Dados realizado, expondo a metodologia de aplicação, as tarefas realizadas, técnicas, algoritmos, *software* e *hardware* utilizados, bem como os custos do processo.
- Os resultados: Por fim, são mostrados os resultados alcançados pelo trabalho, onde o autor avalia as vantagens e desvantagens do processo, além das dificuldades encontrados, apresentando a conclusão a que se chegou,

isto é, se o trabalho atendeu ou não às expectativas e objetivos estabelecidos.

Dessa forma, tendo um roteiro pré-estabelecido, a análise foi feita de forma melhor organizada em busca dos objetivos do presente trabalho. Os casos de estudo selecionados são descritos e analisados nos próximos tópicos.

## **4.2 Caso de Estudo 1: Site de Comércio Eletrônico**

Este tópico descreve os principais pontos do primeiro trabalho selecionado, que trata da aplicação da Mineração de Dados na análise do perfil de clientes ativos e não ativos e do padrão de compras feitas por estes em um site de comércio eletrônico.

### **4.2.1 O trabalho**

Este primeiro caso de estudo é uma dissertação de mestrado desenvolvida por Leandro da Silva Carvalho em 2009, intitulada “Mineração de Dados Aplicada ao Entendimento do Comportamento do Consumidor para dar Suporte ao Processo de Tomada de Decisões”, na ocasião sendo requisito para obtenção do título de Mestre em Engenharia Civil pela Universidade Federal do Rio de Janeiro (CARVALHO, 2009).

O autor propôs a aplicação de técnicas de Mineração de Dados sobre as bases de dados de um *software* de comércio eletrônico que apóia as vendas de um restaurante, uma vez que foi verificada a presença de muitos clientes cadastrados no mesmo, mas que nunca fizeram nenhum pedido. O comportamento foi julgado estranho, pois é possível navegar pelo site sem se cadastrar, e a razão de ser feito o cadastro é para fazer o pedido.

Diante do problema Carvalho (2009) então identificou a oportunidade e necessidade de se trabalhar sobre estas bases de dados, identificando o que diferenciava um cliente ativo de não ativo, de modo a determinar o que o faz cadastrar-se e não comprar. Ele propôs ainda a identificação das relações de compra dos clientes, isto é, estabelecer um perfil de compra.

O objetivo era construir uma base de conhecimento para a geração de vantagem competitiva, a partir da extração de padrões sobre as bases de dados da empresa, bem como tentar prever o comportamento de novos clientes. Com isso seria possível interagir com seus clientes baseando-se nas suas próprias características pessoais, visando retê-los, além de atrair novos consumidores.

#### **4.2.2 A empresa**

A empresa foco do estudo é um tradicional restaurante italiano, chamado La Mole, situado na cidade do Rio de Janeiro, que possui vários pontos de atendimento em diferentes bairros, além de atender também a cidade de Niterói. O restaurante, na época, possuía 15 unidades e cerca de 1.000 colaboradores.

Em 2004, o La Mole decidiu adotar a solução NETCOMMERCE (*software* desenvolvido pelo próprio autor do trabalho, em 2000) para expandir seus negócios, disponibilizando mais um canal de atendimento aos clientes e oferecendo entregas a domicílio em boa parte da capital carioca.

A empresa, desde então, vem acumulando uma valiosa base de dados, porém não realiza nenhum trabalho de análise sobre os mesmos. Dessa forma, há a necessidade do tratamento dessas informações com o intuito de gerar conhecimento para ser utilizado na tomada de decisão.

#### **4.2.3 O Processo de Mineração de Dados**

Este tópico descreve a metodologia utilizada pelo autor para a descoberta de conhecimento na base de dados do site de comércio eletrônico do restaurante La Mole, bem como, as tarefas de mineração, técnicas, algoritmos, *softwares* e *hardware* utilizados. Ao final também são apresentados os custos do processo e seus resultados, que mostram, do ponto de vista do autor, as vantagens, desvantagens e dificuldades encontradas na realização da mineração de dados.



#### 4.2.3.1 Metodologia

No processo de descoberta de conhecimento o autor adotou algumas etapas para se conseguir sucesso. A base de dados operacional do sistema NETCOMMERCE possuía informações sobre três anos de transações. Durante o pré-processamento, foi criado um armazém de dados (*Data Warehouse*) visando reunir informações de qualidade para as etapas seguintes. No tópico 3.2.1 do presente trabalho foi ressaltada a importância desta metodologia de organização de banco de dados para o processo de mineração.

Dessa forma, primeiramente foram selecionados da base de dados do sistema somente os dados necessários para a realização do processo de descoberta de conhecimento. Em seguida foi feita a limpeza dos mesmos eliminando algumas linhas e colunas desnecessárias, dados duplicados, além da criação de colunas e tabelas de apoio à análise. Os clientes ativos foram separados dos não-ativos em tabelas diferentes, pois o objetivo era descobrir os clientes não-ativos que poderiam se tornar ativos. Por fim, foram criadas as dimensões do *Data Warehouse* e os dados selecionados integrados.

Após a criação do armazém de dados o autor efetuou uma análise estatística do conjunto de dados reunido para se conhecer melhor as informações a serem trabalhadas. Para os clientes foram criados gráficos analisando a faixa etária, tempo de relacionamento com a empresa, sexo, região de residência, loja pelas quais já foram atendidos, bem como a separação entre clientes ativos e não-ativos, revelando que existem na base mais clientes não-ativos do que ativos.

Para os pedidos foram criados gráficos para as lojas que os atenderam, dia da semana em que foi realizado, se era feriado ou não, horário, região de entrega e valor. Conforme o autor, esta simples análise já revelou grandes descobertas nos dados, sendo que os mesmos não possuíam grandes desigualdades, prevendo dificuldades em sua segmentação.

Após a contextualização com os dados partiu-se para o processo de mineração de dados em si. Primeiro foi realizada a segmentação dos dados, isto é, agrupamento dos clientes com base em características intrínsecas do conjunto de dados que permitem uma separação entre os mesmos.

Foi necessário realizar o processo várias vezes com diferentes parâmetros para que os resultados fossem satisfatórios. Os clientes ativos e não-ativos foram analisados como um todo para verificar se houve algum período em que ocorreu o cadastramento sem motivo de vários usuários. A segmentação dos clientes foi feita então, considerando o local de residência dos clientes por regiões.

Depois da segmentação foi realizada a análise de associação dos pedidos feitos pelos clientes, definindo padrões de compra, objetivando sugerir produtos aos clientes. As regras de associação eram comprovadas por consultas efetuadas ao banco de dados verificando se as mesmas não eram geradas por poucos clientes que eram muito ativos e compravam sempre os mesmos produtos.

A base de dados do site possuía 183 produtos diferentes distribuídos em 14 categorias. A análise de associação foi realizada primeiro com todos os produtos, depois feita por categorias. Por último, a associação foi conduzida levando em conta as relações entre os pedidos e os perfis de clientes.

Nesse processo foram necessários também vários modelos até se chegar a resultados relevantes devido à homogeneidade da base de dados, sendo necessário às vezes, a criação de subcategorias de produtos para melhorar a qualidade das informações geradas. O autor levantou ainda a importância de um analista de negócios para analisar o contexto das regras geradas de forma a guiar o processo para a obtenção de melhores resultados.

Por fim, após a análise de regras de associação, tentou-se construir um modelo preditivo de classificação para predizer se, com base no seu perfil, um novo cliente tenderia a ser ativo ou não-ativo, através da estipulação de probabilidades de cada evento ocorrer. Também seria interessante essa predição para o valor do pedido de um novo cliente.

O objetivo desta etapa era verificar se a empresa estava perdendo clientes para os concorrentes e descobrir a razão de clientes se cadastrarem e não realizarem pedidos.

Vários modelos com diferentes parâmetros foram criados, cujos com maiores probabilidades de acerto chegaram a 60%, o que constatou que o perfil de clientes ativos e não ativos eram muito parecidos, não sendo possível identificar características intrínsecas que os diferenciassem de forma eficiente, devido à homogeneidade dos dados.

Os resultados obtidos no processo de mineração de dados feito pelo autor, que compreende o pós-processamento serão apresentados no tópico 4.2.3.5.

#### 4.2.3.2 Tarefas, Técnicas e Algoritmos

Carvalho (2009) utilizou no processo descrito acima três tarefas de mineração de dados para a descoberta de conhecimento: análise de grupo, também chamada de clusterização ou segmentação; análise de associação; e modelagem de previsão. Essas tarefas são descritas no tópico 3.2.2 do presente trabalho.

Na segmentação dos clientes o autor utilizou o algoritmo *K-means* de segmentação de dados, caracterizado por ele como sendo de baixo custo computacional e com alta escalabilidade.

Na geração de regras de associação dos pedidos e entre pedidos e clientes, foi utilizado o algoritmo *Apriori*.

Segundo o autor o algoritmo Apriori consiste em duas etapas. Primeiro encontra-se na base de dados os subconjuntos de itens freqüentes em relação ao total de transações de acordo com um suporte mínimo (valor de freqüência) estabelecido previamente para a seleção desses subconjuntos.

Em seguida o subconjunto selecionado é avaliado conforme um valor de confiança mínima pré-estabelecido, isto é, a probabilidade de um item ocorrer em conjunto com outro, em busca daquelas associações que são interessantes e realmente prováveis de ocorrer.

Por último, na modelagem de previsão que visava a classificação de novos clientes foi utilizado um algoritmo de árvore de decisão contido na ferramenta utilizada para mineração de dados

A árvore de decisão consiste na disposição de características de forma hierárquica, formando literalmente uma árvore, de modo que a complexidade de um problema seja ramificada em problemas menores, porém de menor complexidade. Inicialmente é definido o nó raiz, que é ramificado em sub-nós e estes em mais sub-nós, até chegar nos nós terminais da árvore, que representam uma determinada classe. Novos registros são então classificados navegando-se pela árvore, conforme suas características, até chegar nó terminal que mais lhe convier, determinando sua classe.

#### 4.2.3.3 Hardware e Software

O trabalho de Carvalho (2009) não especifica os recursos de *hardware* que foram utilizados para a realização do processo de descoberta de conhecimento.

Quanto ao *software*, primeiramente é apresentado as características do NETCOMMERCE, sistema que dá suporte às atividades de comércio eletrônico do restaurante La Mole.

O NETCOMMERCE foi desenvolvido pelo autor em 2000, e começou a ser utilizado pela La Mole em 2004. As páginas do *software* foram desenvolvidas em linguagem de programação ASP em conjunto com componentes desenvolvidos em *Visual Basic 6*, funcionando em um servidor de aplicação com sistema operacional *Windows Server 2003*, por meio do *software Internet Information Service 6*. O sistema gerenciador de banco de dados (SGBD) é o *Microsoft SQL Server 2005*.

Durante as análise estatísticas dos dados reunidos no *Data Warehouse*, utilizou-se a planilha eletrônica *Microsoft Excel 2003* como ferramenta principal e o *MatLab 7.0* para suporte na construção de gráficos mais elaborados.

Para o processo de mineração de dados em si, o *software* utilizado foi o SGBD *Microsoft SQL Server 2008*, versão de avaliação, que contém todos os algoritmos abordados no trabalho do autor.

Carvalho (2009) avalia a ferramenta de mineração de dados como sendo muito proveitosa nas atividades desempenhadas. Segundo ele possui muito material de ajuda, é de fácil instalação e utilização, além de se integrar facilmente ao *Microsoft Excel*, razão que facilita o uso das informações mineradas por administradores.

#### 4.2.3.4 Custos

O trabalho de Carvalho (2009) não especifica os custos do processo de descoberta de conhecimento conduzido por ele.

#### 4.3.3.5 Resultados

Com base no estudo do trabalho feito por Carvalho (2009), é possível enumerar diversos conhecimentos descobertos, sendo eles indicados pelo próprio autor para utilização em futuras campanhas de *marketing*. Devido às numerosas informações obtidas, os principais pontos serão sintetizados nos próximos parágrafos.

No processo de segmentação de clientes pode-se chegar a informações que indicaram as regiões e perfis com maior número de clientes não-ativos, locais onde campanhas de *marketing* devem atuar mais intensamente com o intuito de torná-los ativos. Os clientes que tendem a ser não-ativos são:

- Residentes nos bairros Vila Isabel, Andaraí e Grajaú;
- Residentes no bairro São Conrado e atendidos pela loja do bairro Barra;
- Residentes no bairro Botafogo com alta faixa etária;
- Residentes nos bairros Centro e Santa Rosa, do sexo masculino;
- Residentes no bairro Gávea, do sexo feminino e com baixa faixa etária.

Segundo o autor, foi muito difícil interpretar os ‘clusters’ (agrupamentos) gerados pela ferramenta, mas apesar dos dados serem muito homogêneos foi possível extrair alguns padrões interessantes. Porém ele ressalta que os resultados seriam mais interessantes se os dados fossem melhor subdivididos, implicando em melhor compreensão e menos esforço computacional.

Já na análise de regras de associação, os clientes tendem a fazerem pedidos conforme as regras abaixo:

- ‘Medalhão à piamontesa’ geralmente é pedido com ‘couvert família’;
- Quem pede ‘carne’ ou ‘pizzas’ tende a pedir também ‘refrigerante pet’ enquanto que quem pede ‘massas’ tende a pedir ‘refrigerante lata’;
- Geralmente ‘cervejas’ são pedidas em consequência do pedido de ‘entradas e saladas’, ‘massas’ ou ‘carnes’;
- Clientes do sexo feminino não tendem a comprar ‘lasagna à bolognesa’ (regra negativa);
- Clientes residentes no bairro Centro, do sexo masculino, tendem a comprar ‘bebidas’;

- Clientes residentes no bairro Copacabana, do sexo feminino, tendem a comprar ‘carnes’;
- Nas segundas-feiras os clientes tendem a comprar ‘aves’;
- As ‘massas’ são mais vendidas nas terças-feiras;
- Nas quintas-feiras é muito vendido ‘massas’ com ‘bebidas’;

Segundo o autor, houve muita dificuldade em conseguir resultados expressivos com essa técnica, sendo necessária a criação de vários modelos até se chegar ao mais adequado.

A última tarefa de mineração de dados realizada pelo autor, a classificação dos clientes para a previsão de novos clientes, como ativos ou não-ativos, foi apontada como sendo a mais difícil, devido a base de dados ser muito homogênea e sendo difícil identificar características que diferenciasssem de forma específica um tipo de cliente do outro.

No entanto, segundo o autor, prever o tipo de cliente nesse caso, não é crítico para o negócio como, por exemplo, em concessões de crédito. Dessa forma, os modelos classificadores gerados que tenham mais de 50% de probabilidade de acerto são válidos para orientar os administradores no trato com os clientes.

De maneira geral, Carvalho (2009) conclui que descobriu conhecimentos bastante relevantes, principalmente para campanhas de *marketing* e, por isso obteve êxito em seu trabalho.

Ele considera que o processo foi muito trabalhoso, e que a qualidade dos dados é primordial para obter bons resultados. Deve-se ter em mente um problema para guiar o processo, e um analista de negócios é de grande importância na interpretação das informações geradas, para avaliar aquelas com real proveito.

#### **4.3 Caso de Estudo 2: Loja de Materiais para Construção**

Este tópico descreve o segundo trabalho selecionado para estudo, que trata da aplicação de técnicas de mineração de dados às bases de dados de uma loja de materiais de construção.

#### 4.3.1 O trabalho

O segundo trabalho analisado, com o título “*Data Mining* no Varejo: estudo de caso para loja de materiais de construção”, é de autoria de André Gustavo Schaeffer, sendo desenvolvido em 2003, como requisito para a obtenção do título de mestre em Ciências da Computação pela Universidade Federal do Rio Grande do Sul (SCHAEFFER, 2003).

Nesse trabalho o autor aplica o processo de Mineração de Dados sobre a base de dados do sistema que dá suporte às atividades operacionais de uma empresa que trabalha com venda de materiais de construção.

Sabia-se previamente que empresas de construção civil são bons clientes, mas e quanto aos clientes pessoas físicas, quais os melhores perfis de clientes?

O objetivo, dessa forma, era promover o cruzamento de informações cadastrais dos clientes com informações das vendas realizadas, para gerar informação útil para o negócio da empresa, que podem embasar ações para gerar lucro.

#### 4.3.2 A empresa

A empresa onde Schaeffer (2003) aplicou o processo de mineração de dados é a MK Móveis e Materiais de Construção, situada na cidade de Erechim, no Rio Grande do Sul, loja de materiais de construção, que também compreende uma fábrica de móveis.

A empresa em questão possui 18 anos de existência, atendendo o Rio Grande do Sul, Santa Catarina, Paraná e Minas Gerais. Os materiais de construção são vendidos para clientes finais, foco do trabalho de Schaeffer (2003), e os móveis são vendidos para revendedores.

A loja de materiais de construção conta com 6 funcionários compreendendo os dois proprietários, e possui uma aplicação que mantém dados sobre seus clientes, produtos, vendas e estoque, sendo que nunca foi feita qualquer análise sobre a mesma.

### 4.3.3 O Processo de Mineração de Dados

Este tópico descreve o processo de descoberta de conhecimento conduzido pelo autor no trabalho em análise, apontando sua metodologia, tarefas, técnicas e algoritmos de mineração utilizados, custos e resultados obtidos.

#### 4.3.3.1 Metodologia

A metodologia utilizada por Schaeffer (2003) chama-se CRISP, que, segundo ele, foi criada em 1996 pelos pesquisadores das empresas *Daimler Chrysler*, *SPSS* e *NCR*. As etapas sugeridas pela metodologia para condução do processo de mineração de dados e descoberta de conhecimento são as seguintes:

- Conhecimento do negócio: compreende determinar os objetivos do negócio, avaliar a situação das bases de dados, definir os objetivos da mineração de dados e o plano de projeto;
- Conhecimento dos dados: consiste em coletar os dados iniciais, descrevendo e explorando os mesmos, e verificando sua qualidade;
- Preparação dos dados: etapa que seleciona os dados a serem minerados, limpando-os, construindo novos, integrando-os e formatando-os;
- Modelagem: compreende na seleção das técnicas de modelagem, e construir o modelo de mineração e, em seguida avaliá-lo;
- Avaliação: etapa que avalia os resultados obtidos, revendo o processo e determinando os passos seguintes;
- Aplicação dos resultados: etapa de responsabilidade dos tomadores de decisão, que planejam a aplicação dos resultados, bem como sua monitoração e manutenção, e revendo todo o projeto.

Dessa forma, primeiro o autor procurou conhecer todo o negócio e realizou entrevistas com os proprietários da empresa. Com isso foram definidos os objetivos do processo de mineração de dados:

- a) Conhecer o perfil do cliente que compra na loja;
- b) Conhecer o perfil do cliente associado com as compras que o mesmo faz na loja;



- c) Criar uma lista dos produtos mais rentáveis e verificar quais os clientes com mais tendência a realizar compras desses produtos;
- d) Através de uma análise de cesta de mercado, conhecer quais produtos estão associados em transações de venda;
- e) Tentar prever se o cliente que está comprando pela primeira vez na loja tende a não voltar mais;
- f) Conhecer o perfil dos clientes que compram pela primeira vez na loja.

Constatou-se também a necessidade da alteração do sistema para captação de dados essenciais ao processo de mineração de dados no que diz respeito ao cadastro de clientes e de vendas. Sendo feito estas alterações, a aplicação que dá suporte aos negócios da empresa capturou dados durante 8 meses: de maio a dezembro de 2002, resultando em 40.100 itens vendidos em 6.422 transações a serem mineradas.

Assim, os dados foram coletados da base operacional, transformada em arquivo texto e em seguida importada, primeiro para o banco de dados onde foram feitos melhoramentos nos dados. Esses melhoramentos consistiam na estruturação das tabelas para adequar ao processo de mineração de dados, exclusão de registros nulos, duplicados, ou fora do padrão, além da transformação dos dados, com a criação de atributos e totalização de valores. Alterações foram feitas também na classificação de produtos da loja, que estavam mal descritos e classificados. Segundo o autor esse etapa foi muito trabalhosa.

Inicialmente o banco de dados de trabalho de mineração de dados possuía apenas a tabela de clientes e a tabela de vendas, que foram depois integradas em uma só cruzando seus registros. Esta tabela foi transformada em arquivo texto, conforme os atributos necessários a cada tarefa de mineração de dados, para possibilitar a importação pela ferramenta de mineração de dados utilizada. Também foi feita uma breve análise estatística dos dados reunidos.

Após a preparação de dados a mineração de dados foi realizada baseada nos objetivos definidos no início do processo.

Para conhecer o perfil do cliente que compra na loja foi feita uma clusterização (ou agrupamento) de forma a definir grupos de clientes parecidos de acordo com suas características mais relevantes, conforme pesos pré-definidos ressaltando a importância de cada característica. Nessa tarefa foram necessárias

várias tentativas com muitas mudanças de parâmetros até conseguir resultados satisfatórios.

Depois o autor buscou determinar o perfil do cliente associado com as compras que o mesmo faz na loja, através de agrupamentos e também extração de regras de associação.

Em seguida, o objetivo era criar uma lista dos produtos mais rentáveis e verificar quais os clientes com mais tendência a realizar compras desses produtos, também através de clusterização. Os produtos mais rentáveis e considerados para análise foram 'roupa de luxo' e 'conjunto de louças para banheiro'. Dessa forma pode-se obter uma lista com os clientes com maiores probabilidade de comprar estes produtos. Os proprietários da loja consideraram válidos os resultados uma vez que a lista apresentou índices de probabilidade de compra e confiança das regras elevadas.

A mineração de dados seguiu e foi feita uma análise de cesta de mercado para conhecer quais os produtos são comprados associados uns com os outros através da descoberta de regras de associação. Nessa tarefa percebeu-se que os valores de suporte que permitiram a obtenção das regras eram muito baixos, devendo-se, segundo o autor, à variedade de produtos vendidos na loja.

Após essa tarefa o autor tentou criar um modelo para prever se o cliente que compra a primeira vez tende a não voltar mais na loja, através de predição neural. Os resultados surpreenderam os proprietários da loja, pois o número de clientes que compram uma única vez e não voltam foi maior do que o esperado. Uma questão interessante dessa tarefa foi a geração de código em linguagem de programação C pela ferramenta de mineração de dados com a lógica do modelo, o que possibilita a integração deste com outras aplicações.

A tarefa final consistia em conhecer o perfil dos clientes que compram pela primeira vez na loja. Nessa etapa também foi utilizada a clusterização.

Os resultados de todo o processo de mineração de dados é apresentado no tópico 4.3.3.5, abordando o pós-processamento das informações geradas, bem como as conclusões a que Schaeffer (2003) chegou.

#### 4.3.3.2 Tarefas, Técnicas e Algoritmos

Conforme já foi mencionado no tópico anterior, de acordo com os objetivos do processo de mineração de dados realizado pelo autor, as tarefas de mineração aplicadas foram as seguintes (os objetivos foram identificados com letras assim como dispostos no tópico anterior):

- Objetivo a: Análise de grupos (também denominado agrupamento ou clusterização);
- Objetivo b: Análise de grupos e análise de associação;
- Objetivo c: Análise de grupos;
- Objetivo d: Análise de associação;
- Objetivo e: Modelagem preditiva, nesse caso utilizando algoritmos de redes neurais;
- Objetivo f: Análise de grupos;

Dessa forma o autor utilizou principalmente a análise de grupos para a descoberta de conhecimento em seu trabalho.

#### 4.3.3.3 Hardware e Software

Schaeffer (2009) mostra em seu trabalho todo o *hardware* e *software* tanto utilizado pela empresa no dia a dia, como os utilizados por ele próprio para o processo de mineração de dados. Pode-se destacar, no caso da loja, que o banco de dados operacional era mantido em *Microsoft Access*.

No processo de mineração de dados, o *hardware* utilizado foi o um microcomputador com processador *Intel Pentium III*, com 256 *Megabytes* de memória e Disco Rígido com capacidade para 40 *Gigabytes*. Já os *softwares* utilizados no processo foram:

- *Microsoft Excel*, *Word* e *Access* para o processo de extração e estruturação dos dados para serem importados pela ferramenta de mineração de dados;
- *Oracle 9i Database Server* e *Oracle SQL Plus* como banco de dados e ferramenta de consulta ao banco de dados, respectivamente, para o tratamento e limpeza dos dados a serem minerados;

- *IBM DB2 Intelligent Miner for Data Version 6.1, IBM DB2 Universal Database Personal Edition* e *WEKA 3.2.3* como ferramentas para mineração dos dados em si;
- *Borland C++ 3.1*, utilizado para tratamento de código gerado pela ferramenta de mineração de dados no caso da predição neural, que possui a lógica do modelo, possibilitando uma integração com outras aplicações.

Schaeffer (2003) avaliou as ferramentas de mineração de dados como muito produtivas e com muito potencial de gerar informações que podem ser traduzidas em lucro para a empresa.

O *WEKA* é desenvolvido em linguagem *Java* e possui maior facilidade para gerar regras de associação do que o *IBM DB2 Intelligent Miner*. No entanto, é muito rigoroso no tratamento de arquivo texto, mas tem bibliografia de fácil acesso e ajuda na *Web*.

Já o destaque para *IBM DB2 Intelligent Miner* é a quantidade de telas de passo a passo para configuração dos algoritmos e vantagem de geração de código em linguagem de programação C com a lógica das predições neurais. Ele também possui bastantes gráficos que, no entanto, necessitam de um bom conhecimento para interpretá-los.

#### **4.3.3.4 Custos**

Schaeffer (2003) não especifica em seu trabalho os custos do processo de mineração de dados aplicado.

#### **4.3.3.5 Resultados**

A cada tarefa de mineração de dados realizada, o autor apresentou os resultados obtidos fazendo comentários e sugestões de como estas poderiam ser usadas, focando sempre em campanhas de *marketing*.

Em virtude da quantidade de informação gerada, principalmente pelas regras de associação, serão apresentadas a seguir aquelas que foram consideradas mais relevantes, porém não sendo as outras menos importantes:

- Quanto ao perfil dos clientes que compram pela primeira vez na loja, pode-se perceber que estes são na maioria homens, que moram em casas próprias;
- 76% dos clientes que compram na loja e moram em apartamentos, residem no centro da cidade.
- Os clientes tendem a fazer reformas quando o tempo de residência no imóvel é entre 1 e 2 anos, adicionando ainda que 38% das vendas em geral são destinadas a reformas.
- 68% dos que moram em casas alugadas, moram no bairro Bela Vista.
- A venda de materiais de construção ocorre em sua maior parte nas segundas-feiras para mulheres, e nas sextas-feiras para homens, exigindo atenção do controle de estoque.
- No bairro Vila Feliz 77% das vendas são de material hidráulico, com objetivo de construção de residência própria, casa ou apartamento, então, onde estes clientes estão comprando os outros materiais de construção?
- 54% das vendas de rolo de pintura e 48% das vendas de tinta para pintura são feitas para reformas em casas, verificando uma preocupação constante dos clientes com a pintura de suas casas.
- Constatou-se que certos funcionários estão propensos a vender um determinando tipo de produtos, para um público mais específico, exigindo troca de experiências entre eles para aumentar as vendas
- Dentre as muitas regras geradas pelas análises de cestas de compras, muitas permitiram a mudança da disposição física de produtos na loja. Destaca-se nesse caso, o 'joelho hidráulico' que é vendido em consequência da venda de:
  - 'tomada', 'te hidráulico' e 'fita veda rosca';
  - 'cano hidráulico' e 'parafuso';
  - 'fita isolante' e 'te hidráulico';
  - 'tomada' e 'te hidráulico';
  - 'piso', 'te hidráulico' e 'argamassa'
  - 'cola/adesivo', 'luva' e 'parafuso';
  - 'fio elétrico', 'tomada' e 'te hidráulico'.

- Contatou-se na base muitos clientes que compram pela primeira vez e não retornam, surpreendendo os proprietários da loja.
- Percebeu-se ainda que os gastos com materiais de construção tendem a acompanhar o tempo de casado e o tempo de residência no imóvel, diminuindo conforme aumenta o tempo de casamento e de residência. Além de que 85% dos clientes que compram pela primeira vez têm até 2 anos de casado.

Dessa forma, o autor reuniu vasto material a ser analisado enfatizando seu potencial no campo do *marketing*, o que, segundo ele, fez os proprietários da loja analisar a possibilidade de contratação de empresa especializada nesse ramo para indicar como melhor trabalhar com as informações obtidas. Schaeffer (2003) reforça essa idéia afirmando que é preciso avaliar qual o verdadeiro lucro que os resultados podem trazer à empresa.

O autor afirma ainda que os resultado precisam ser atualizados com o tempo, e que o trabalho foi muito útil para identificar pontos importantes a serem levados em conta em um trabalho de mineração de dados.

Uma metodologia deve ser seguida, pois é o guia fundamental de um processo de mineração de dados de sucesso. Outra questão que também pode ter grande influência nos resultados é a classificação de produtos em estoque, sua correta descrição e disposição em categorias e subcategorias, uma vez que, correções deste tipo são muito demoradas, mas possibilitam resultados mais interessantes.

Muitas das informações geradas, segundo o autor, têm grande potencial comercial, sendo que as regras de associação merecem especial atenção, o que mudou o modo com que os proprietários da loja se relacionam com seus clientes. Estes também perceberam a necessidade de uma melhor padronização dos itens do estoque.

Por fim, Schaeffer (2003), ressalta a importância da mineração de dados em pequenas e médias empresas para que fidelizem seus clientes além de conquistar novos diante das adversidades do mercado competitivo.

#### 4.4 Caso de Estudo 3: Empresa de abastecimento de água

O terceiro trabalho selecionado para estudo é abordado no decorrer deste tópico. Ele trata sobre a aplicação de técnicas de mineração de dados sobre um *Data Warehouse* de uma empresa de abastecimento de água.

##### 4.4.1 O trabalho

Este terceiro trabalho foi desenvolvido por Roberta Macêdo Marques Gouveia em 2009. Trata-se de uma dissertação de mestrado, com o título “Mineração de Dados em *Data Warehouse* para Sistema de Abastecimento de Água”, desenvolvido pela autora como requisito para obtenção do título de Mestre em Informática, pela Universidade Federal da Paraíba (GOUVEIA, 2009).

Os objetivos da autora eram colher informações para tomada de decisão e descobrir perfis de consumidores ou dados a respeito das instalações de água que tendessem a gerar perdas ou a terem fraudes.

Dessa forma Gouveia (2009) propôs a criação de um Sistema de Apoio à Decisão (SAD) para minimizar perdas aparentes no sistema de abastecimento de água urbano de modo a apoiar, contribuir e influenciar no processo de tomada de decisão.

O SAD proposto consistia na criação de um *Data Warehouse* departamental, isto é, um *Data Mart*, para o setor comercial da empresa de abastecimento de água, com o intuito de promover uma eficiência no armazenamento e gestão dos dados, fornecendo informações gerenciais do negócios. Em seguida seriam aplicadas técnicas de mineração de dados sobre esse armazém de dados, de modo a gerar conhecimento novo para um melhor controle sobre o comportamento dos consumidores.

O trabalho também é embasado na importância dos serviços de abastecimento de água e o correto racionamento dos recursos hídricos, pois a existência de perdas é preocupante frente à escassez de água em regiões como o nordeste brasileiro onde foi aplicado o estudo de caso do trabalho.

Segundo a autora, no Brasil as companhias de água perdem em média mais de 40% de água em decorrência de fraudes ou perdas físicas em tubulações e

reservatórios, o que ocasiona além de perdas econômicas para essas empresas, uma maior demanda de água retirada da natureza para atender os consumidores.

Portanto, Gouveia (2009) visava proporcionar com seu trabalho um melhor gerenciamento dos recursos hídricos e diminuir as perdas econômicas da empresa de abastecimento de água onde foi aplicada a mineração de dados.

Assim, ela guiou-se pelos seguintes objetivos:

- Determinar o perfil do setor e do consumidor, por meio da verificação dos consumos de água, valores faturados (conta de água) e pontos de utilização de água;
- Verificar e diagnosticar a situação dos medidores (hidrômetros) presentes nos imóveis;
- Encontrar respostas para as anormalidades e irregularidades praticadas pelos consumidores da qual a empresa de abastecimento de água desconhece;
- Avaliar as inadimplências dispostas no setor selecionado para o estudo de caso.

#### **4.4.2 A empresa**

A CAGEPA, Companhia de Água e Esgoto da Paraíba, foi a empresa foco do trabalho de Gouveia (2009). Com sede em João Pessoa, na Paraíba, a CAGEPA atende 185 dos 223 municípios do estado, que são divididos em 6 unidades de negócio.

O banco de dados da empresa, no entanto, fica localizado na cidade de Recife, estado de Pernambuco, na sede da empresa que administra seu sistema de informação.

A CAGEPA acompanha a média nacional no que diz respeito às perdas de água, com índice de perda aparente entre 40 e 50%. Entende-se por perda aparente toda a água que não é medida, isto é, perdas geradas por erros de medição, ligações clandestinas e fraudes. Por outro lado, perdas reais são perdas físicas no transporte da água, isto é, problemas em tubulações reservatórios.



Dessa forma, a aplicação de mineração de dados objetivando reduzir as perdas de água seria muito bem vinda, com a geração de conhecimento que permitisse combater ou prevenir esse problema.

#### **4.4.3 O Processo de Mineração de Dados**

Este tópico descreve como foi desenvolvido o processo de mineração de dados conduzido por Gouveia (2009). Nele serão descritos a metodologia, tarefas, técnicas e algoritmos de mineração de dados aplicados, *hardwares* e *softwares* utilizados, custos do processo, bem como os resultados obtidos pela autora.

##### **4.4.3.1 Metodologia**

Gouveia (2009) procurou primeiro selecionar as informações que seriam mineradas, tentando definir o escopo de sua análise. Decidiu-se efetuar a análise sobre as informações de um ano de consumo da localidade 001, setor 64, que corresponde à comunidade de Miramar, na cidade de João Pessoa, capital paraibana, que conta com 1.285 consumidores e 17.800 pontos de utilização de água, distribuídos em 79 quadras, totalizando 15.420 registros de faturas/contas.

Segundo a autora, a razão de se escolher esta região específica foi porque na mesma constam diversos tipos de consumidores com diferentes realidades sócio-econômicas, além de diferentes tipos de estabelecimentos, o que permitia uma boa análise, que poderia ser espelhada para outras regiões em trabalhos futuros.

Os atributos que seriam necessários ao processo foram definidos e os dados foram coletados, como por exemplo, registros de consumo dos imóveis, conta/faturas, informações sobre os hidrômetros entre outros, referentes a um período de 12 meses. Esses dados foram fornecidos pela CAGEPA à autora em arquivo texto.

Em seguida foi feita a limpeza, transformação e integração dos dados, isto é, exclusão de registros inconsistente, nulos, duplicados, fazendo-se ajustes e agrupamentos de faixas de valores, com o intuito de não se provocar erros na construção do *Data Warehouse* e nem na mineração de dados.

Após esta etapa partiu-se para a criação do sistema de apoio à decisão proposto, começando pela construção do *Data Warehouse* departamental do setor comercial.

Assim optou-se pela criação de um DW estruturado com o esquema de constelação de fatos, isto é, quando se define duas ou mais tabelas de fatos centrais a serem descritos compartilhando diversas tabelas de dimensões (visões) destes fatos. No caso do trabalho de Gouveia (2009) foram considerados dois fatos: perfil do setor, referente a dados dos consumidores, e perfil das perdas aparentes, com dados relacionados às irregularidades na rede de distribuição.

Com os cubos de dados construídos no DW, através de uma ferramenta OLAP (em inglês, *On-Line Analytical Processing*) já foi possível extrair muitas informações gerenciais relevantes. Entende-se por OLAP como sendo tecnologias que permitem a usuários explorar dados dos *Data Warehouses*, fornecendo funcionalidades para análise iterativa de dados em diferentes dimensões e granularidades.

Gouveia (2009) partiu então para a mineração dos dados contidos no DW departamental através de diversos algoritmos. A autora promoveu ainda uma comparação entre esses algoritmos avaliando sua eficiência.

Antes de começar a mineração de dados efetivamente, primeiro realizou-se uma pré-mineração, isto é, uma análise inicial dos dados, com o intuito de conhecer melhor seu contexto, através de gráficos gerados pela ferramenta de mineração.

Para o perfil do setor analisou-se a situação da ligação de água, tipos de consumidores, adimplências e inadimplência, bem como o cruzamento destas informações.

Para o perfil das perdas aparentes, verificou-se a situação dos hidrômetros, que acabou constatando que os instalados no setor 64, foco do estudo, eram um pouco velhos e com pouca capacidade de vazão, sendo os residenciais os mais comuns.

O processo de descoberta de conhecimento então seguiu e para a mineração de dados visando determinar o perfil do setor, o atributo classe definido foi 'inadimplência', ou seja, as regras geradas visavam determinar a tendência dos perfis de consumidores analisados serem ou não inadimplentes.

Logo depois foram minerados os dados para o perfil de perdas aparentes, que tinha como atributo classe ‘decisão’, isto é, decidir sobre a troca do hidrômetro, conforme as regras geradas, sendo possíveis valores para este atributo: ‘favorável à troca’, ‘desfavorável à troca’, fazer ‘análise mais detalhada’ ou então ‘indiferente’.

Em ambos os casos foram utilizados quatro algoritmos de mineração de dados, que implementavam árvores de decisão, classificação estatística e regras de associação. Estes algoritmos e suas tarefas, bem como os resultados obtidos pela mineração de dados são apresentados nos próximos tópicos.

#### **4.4.3.2 Tarefas, Técnicas e Algoritmos**

Gouveia (2009) utilizou basicamente duas tarefas de mineração de dados para a análise do perfil do setor e do perfil das perdas aparentes, sendo os mesmos algoritmos aplicados a ambos os casos, conforme segue:

- Modelagem de previsão:
  - Algoritmo ID-3, classificação através de árvores de decisão;
  - Algoritmo J4.8, que também resulta em um árvore de decisão;
  - Algoritmo *NaiveBayes*, que efetua uma classificação baseada em estatística;
- Tarefas de Associação:
  - Algoritmo *Apriori*, que gera regras de associação entre os itens minerados.

A autora efetuou uma comparação entre os algoritmos visando apontar aquele que melhor atendeu às suas necessidades no contexto específico do trabalho. Segundo ela, o algoritmo J4.8 destaca-se pela visualização gráfica da árvores de decisão gerada, altas taxas de acertos, baixas taxas de erros, e tempo de processamento baixo.

#### **4.4.3.3 Hardware e Software**

A autora não especifica o *hardware*, mas descreve bem detalhadamente o *software* utilizado no processo.

É importante destacar primeiro que o banco de dados da CAGEPA era mantido em um SGBD *IBM DB2* pela empresa responsável pelo sistema de informação desta. Os dados foram fornecidos por ela em arquivo texto para serem trabalhados.

Para a organização dos dados trabalhados Gouveia (2009) utilizou o SGBD *PostgreSQL Database 8.3.1* com *pgAdmin 1.8*, enquanto que para a criação do *Data Warehouse* e posterior mineração de dados a ferramenta utilizada foi *Pentaho Open BI Suite*, pacote de aplicativos de apoio à decisão gratuito e de código aberto, que possui ainda:

- *Pentaho Schema Workbench*, que possibilitou a criação dos cubos de dados do *Data Warehouse*;
- *Pentaho Analysis View*, ferramenta OLAP para a consulta de dados sobre os cubos do *Data Warehouse*, bastante fácil de usar e que possibilita ainda a geração de gráfico com os referidos dados, facilitando o entendimento destes;
- *WEKA* versão 3, ferramenta que permite a mineração de dados a partir do uso de diversos algoritmos pré-instalados.

Gouveia (2009) considerou essas ferramentas adequadas à proposta do trabalho, sendo que todas elas juntas formam o sistema de apoio à decisão proposto pela autora.

#### **4.4.3.4 Custos**

Gouveia (2009) não especifica em seu trabalho se teve algum custo com o processo de mineração de dados aplicado.

#### **4.4.3.5 Resultados**

O trabalho de Gouveia (2009) obteve várias regras em decorrência da aplicação da mineração de dados, podendo-se destacar as seguintes:

- Todos os consumidores da subcategoria ‘favela’, com situação do esgoto ‘ligado normal’, e situação da água ‘cortada’, estão inadimplentes;

- Todos os consumidores da subcategoria 'escritório/associação' com atividade 'comercial', e situação da água 'cortada' ou 'suprimida total', estão inadimplentes;
- A situação da ligação de água sendo 'ligada', implica em clientes adimplentes;
- A situação da ligação de água sendo 'suprimida\_total', implica em clientes inadimplentes;
- Ligação de água com situação 'cortada', subcategoria de consumo 'escritório/associação', com atividade 'comercial' e 'indústria que usa água no processo industrial', tendem a ser inadimplentes;
- Predominam consumidores inadimplentes em situação de água 'ligada', situação da ligação de esgoto 'ligado normal', categoria de consumo 'residencial' e subcategoria 'favela';
- Todas as ligações de esgoto com situação 'ligado normal' e água 'ligada' implicam em adimplência, sendo que adicionando-se à essa regra a categoria 'residencial', essa regra cai para 91% de certeza;
- Hidrômetros que têm consumo tarifário 'residencial' entre '10' e '20 m<sup>3</sup>', com anormalidade 'hidrômetro soterrado', tipo de hidrômetro 'mult magnético', e ano fabricação de '1994 a 1998', são 'favoráveis à troca';
- Hidrômetros que têm consumo tarifário 'residencial' 'até 10 m<sup>3</sup>', com ano fabricação de '2004 a 2008', e anormalidade igual a 'by pass', precisam de uma 'análise mais detalhada';
- Hidrômetros do tipo 'mult magnético', com consumo tarifário 'comercial' 'acima de 10 m<sup>3</sup>', 'sem anormalidade', com 'menos de 3 anos' de instalação, e fabricados de '1999 a 2003' e de '2004 a 2008' são 'desfavoráveis à troca', enquanto que para os com fabricação de '1994 a 1998' deve-se realizar uma 'análise mais detalhada';
- Todos os consumidores com hidrômetro com capacidade de até '3 m<sup>3</sup>' estão adimplentes, destes 96% de seus hidrômetros são 'desfavoráveis à troca';
- Todos os consumidores com hidrômetros 'sem anormalidade', do tipo 'mult magnético' estão adimplentes.

Segundo a autora os padrões descobertos foram satisfatórios, pois conseguiram extrair conhecimentos desconhecidos nos dados, que podem ser

usados para apoiar a tomada de decisão pelos gestores da CAGEPA visando um uso racional de água e proporcionando medidas corretivas e preventivas para minimizar as perdas de água. Porém, os resultados gerados precisam ser validados pelos especialistas da empresa para serem colocado em prática, avaliando seu potencial

O sistema de apoio à decisão também foi construído como proposto, sendo importante ressaltar a importância do *Data Warehouse* para o processo de mineração de dados, pois os dados são melhores organizados, facilitando o trabalho.

Uma dificuldade encontrada pela autora foi em relação à escolha dos *softwares* a serem utilizados no processo, pois são escassos aqueles disponíveis de forma gratuita ou que se destinem à pesquisa acadêmica. Outro ponto de dificuldade foi a interdisciplinaridade do trabalho, visto que foi necessário se estudar muito à respeito de como se comporta sistemas de abastecimento de água e suas peculiaridades, assunto que envolve a engenharia hidráulica.

Por fim, Gouveia (2009) destaca o fator humano do processo de mineração de dados, pois nem todas as regras geradas pela mineração de dados são consideradas relevantes. É necessário um especialista para saber interpretá-las de acordo com o contexto do negócio e só depois aplicá-las. Isso implica ainda na necessidade de se trabalhar com equipes interdisciplinares, agregando conhecimentos entre especialistas e analistas.

#### **4.5. Análise dos Casos de Estudo**

Com base nas descrições dos trabalhos selecionados para análise é possível sintetizar suas principais características no QUADRO 01 a seguir.

QUADRO 01:  
Comparativo entre os trabalhos analisados

	<b>Trabalho 1: Site de Comércio Eletrônico</b>	<b>Trabalho 2: Loja de Materiais de Construção</b>	<b>Trabalho 3: Empresa de Abastecimento de Água.</b>
<b>Porte da Empresa</b>	Pequena.	Pequena.	Grande.
<b>Objetivos da Mineração de Dados</b>	Segmentação dos clientes; Determinar o perfil das compras; Classificação de novos clientes; Prever o valor do pedido do cliente.	Determinar o perfil dos clientes; Estabelecer o perfil das compras; Conhecer a probabilidade de o cliente voltar à loja.	Traçar o perfil dos consumidores; Verificar quais os perfis de clientes tendem a ser inadimplentes; Determinar quais hidrômetros tem necessidade de troca; Conhecer as razões que levam os consumidores a fraudar as ligações de água.
<b>Tarefas de Mineração de Dados aplicadas</b>	Análise de grupos; Análise associativa; Modelagem de previsão.	Análise de grupos; Análise associativa; Modelagem de previsão.	Análise associativa; Modelagem de previsão.
<b>Custos do processo</b>	Não informado.	Não informado.	Não informado.
<b>Potencial das informações geradas</b>	Aplicação em campanhas de <i>marketing</i> e	Utilização em campanhas de <i>marketing</i> ,	Prevenção de fraudes nas ligações de água,

	fidelização de clientes.	fidelização de clientes, gerenciamento do estoque de mercadorias e disposição física dos produtos na loja.	bem como manutenção eficiente dos hidrômetros, com isso menos perdas de água e menos prejuízos econômicos.
--	--------------------------	--	--

Conforme o quadro acima é possível verificar que os Trabalhos 1 e 2 foram aplicados em empresas de pequeno porte, enquanto que no Trabalho 3, trata-se de uma grande empresa.

Ambos os trabalhos se assemelham nos objetivos no que diz respeito à determinar o perfil dos clientes, porém cada um têm suas peculiaridades. Os Trabalhos 1 e 2 também objetivam a análise de cesta de compras, isto é, traçar o perfil das compras feitas pelos clientes. Adicionalmente o Trabalho 1 tem o intuito de classificar seus novos clientes, prevendo também o valor de suas compras, e o Trabalho 2 propõe a verificação da probabilidade do cliente voltar à loja após a primeira compra.

Por outro lado, o Trabalho 3, objetiva determinar se os consumidores tendem a ser inadimplentes com base em seus perfis, de modo a entender também a razão que os levam a fraudar as ligações de água, e ainda, determinar o momento de trocar os hidrômetros evitando problemas nas redes de distribuição.

Em relação às tarefas utilizadas no processo de Mineração de Dados, em todos os trabalhos foram aplicadas a modelagem de previsão e análise de associação. Somente nos Trabalhos 1 e 2 foi utilizada a análise de grupos.

Nenhum dos trabalhos especificou os custos do processo, em relação a *hardware*, *software*, mão-de-obra, material bibliográfico ou quaisquer outros que tenham eventualmente sido gerados.

Quanto ao potencial das informações obtidas pelos trabalhos, a partir de seu estudo, pode-se chegar às conclusões relatadas a seguir.



No Trabalho 1, as informações têm como utilidade, conforme bastante frisado por Carvalho (2009), principalmente a elaboração de campanhas de *marketing* visando tornar os consumidores não-ativos em ativos, bem como conquistar novos clientes. Além disso, podem ser aplicadas também na fidelização dos clientes, isto é, com base nos perfis e tendências de compras dos clientes é possível dar um tratamento diferenciado aos consumidores, valendo também como mecanismo para conquistar de novos clientes.

O mesmo ocorre com o Trabalho 2, onde Schaeffer (2003) também sugere a aplicação das informações em *marketing* direcionado de acordo os perfis dos clientes, podendo tanto fidelizá-los, como também conquistar novos mercados consumidores.

No Trabalho 2 também é importante ressaltar a utilidade das informações para um melhor gerenciamento dos estoques, que determinaram quando o consumo de determinados produtos é maior. As regras de associação também deixaram clara a necessidade de uma melhor disposição das mercadorias no espaço físico da loja para alavancar as vendas. O autor inclusive cita que os proprietários da loja de materiais de construção cogitavam contratar uma equipe especializada em *marketing*, fazer uma melhor classificação dos produtos no seu sistema de informação, além de já terem providenciado uma melhor disposição dos produtos na loja.

Já no Trabalho 3 a utilidade das informações é mais específica, sendo direcionadas à prevenção de perdas de água de água, condizendo com os objetivos propostos por Gouveia (2009). Elas podem ser aplicadas na prevenção de fraudes nas ligações de água e uma melhor manutenção dos hidrômetros que, por consequência podem evitar perdas de água e perdas econômicas da empresa.

A partir do estudo dos trabalhos analisados até aqui, conclui-se que foi reunido um material de qualidade, que aborda a aplicação de técnicas de Mineração de Dados em diferentes situações, conforme os critérios estabelecidos no tópico 4.1.1 deste capítulo.

O objetivo não era em nenhum momento questionar a qualidade de um ou outro trabalho, mas sim, extrair o que de melhor foi exposto por estes, para permitir um estudo dos seus resultados visando determinar quais seus benefícios para a

inteligência da organização. Tais benefícios serão abordados no próximo tópico, que relata conclusões do estudo realizado até aqui.

#### **4.6 Benefícios da Mineração de Dados para a Inteligência Competitiva**

Este trabalho iniciou-se pela motivação de se conhecer os reais benefícios das informações geradas pela Mineração de Dados para a Inteligência Competitiva de uma organização. Foram feitos os seguintes questionamentos a serem respondidos pela pesquisa desenvolvida: Qual o valor agregado da informação gerada por ferramentas de mineração de dados para a inteligência competitiva da empresa? Em relação aos custos de tal processo, seria vantajoso aplicá-lo?

Primeiramente, em relação aos custos, os trabalhos pesquisados não especificam qual o investimento necessário para desenvolver os processos desenvolvidos por eles. Mesmo assim, é possível fazer algumas considerações.

Em relação à *hardware*, os custos são muito relativos, dependendo da quantidade de dados que serão minerados durante o processo de descoberta de conhecimento. Nos trabalhos estudados aqui, a quantidade de dados analisada é relativamente pequena, não exigindo máquinas com poder de processamento tão grande, nem além da capacidade financeira da maioria das empresas. Isso se deve ao fato de serem trabalhos acadêmicos aplicados em contextos bem definidos, não visando, nesse caso, a análise específica de equipamentos de *hardware*.

Contudo, quando o número de informações é muito grande, isto é, em casos que a quantidade seja bem além do que foi abordado nos trabalhos analisados, com certeza serão necessárias uma ou mais máquinas com maior poder de processamento, utilizando talvez técnicas de computação distribuída. Esse poder de processamento alto é exigido devido ao fato de os algoritmos de Mineração de Dados, conforme abordado no capítulo 3, utilizarem técnicas de inteligência artificial e estatísticas, atuando sobre dados em formato texto, gerando muitos cálculos.

Grandes volumes de dados são gerados por grandes corporações, ou por empresas que lidam com engenharia genética e previsão do tempo, por exemplo. Porém essa não é a realidade da maioria das empresas no Brasil.

Objetivando tornar este trabalho útil para a maioria das organizações, conclui-se que, em termos de *hardware* os investimentos são variáveis conforme o volume

de dados a ser minerado. No entanto, este investimento, é considerado baixo em relação aos outros custos do processo e ao potencial que o mesmo tem se bem conduzido por um profissional capacitado.

Em relação aos *softwares*, os trabalhos analisados apresentaram várias aplicações, tanto proprietárias como livres, incluindo SGBDs e as próprias ferramentas de mineração de dados. Schaeffer (2003) e Gouveia (2009), que utilizaram ferramentas de mineração de dados livres, afirmaram que as mesmas atenderam aos seus propósitos, sendo ainda relativamente fáceis de usar.

Dessa forma, em termos de custos com *softwares*, estes são próximos de zero, pois, havendo ferramentas livres disponíveis que apóiam o processo de descoberta de conhecimento com eficiência, a utilização de ferramentas proprietárias vai depender única e exclusivamente da preferência de quem está aplicando ou solicitando a aplicação da mineração de dados.

Os trabalhos também não abordam os custos com mão-de-obra especializada, uma vez que foram os próprios autores a realizarem o processo. Uma empresa que decide investir na Mineração de Dados poderá fazê-lo contratando empresas terceirizadas responsáveis por realizar o processo ou investir na capacitação dos seus funcionários, administradores de banco de dados, caso a própria empresa seja responsável pelo seu sistema de informação.

No caso de salários mensais, conforme consulta ao site da revista INFO (2011) em junho de 2011, os valores para administradores de banco de dados e analistas de dados podem ser, em média, em R\$ (Reais):

- Administrador de banco de dados:
  - Junior: 3.922,33;
  - Pleno: 5.255,67;
  - Sênior: 8.376,33;
- Analista de Dados
  - Junior: 3.200,00;
  - Pleno: 4.500,00;
  - Sênior: 6.000,00.

O custo com o profissional de Mineração de Dados é o principal de todo o processo, sendo uma das peças mais importantes, pois é responsável por conduzi-lo e, literalmente, descobrir o conhecimento.

Assim, as organizações devem estar atentas para contratar profissionais devidamente qualificados, definindo os objetivos que desejam alcançar com o projeto. Este deverá ser conduzido de forma eficiente, sem, no entanto, demandar muito tempo, pois o investimento não é barato.

O presente trabalho também deveria responder ao questionamento de qual o valor agregado das informações geradas pela Mineração de Dados para a Inteligência Competitiva. Entende-se por valor agregado o valor percebido pelo cliente que adquire um produto ou serviço. No contexto deste trabalho, o cliente é a empresa foco da aplicação do processo de descoberta de conhecimento, e o valor agregado está ligado ao valor percebido pela empresa no que diz respeito à natureza às informações geradas por este processo.

Os trabalhos analisados não apresentam de forma implícita essa questão, uma vez que todos eles somente ressaltam o potencial das informações geradas. Assim, são feitas apenas recomendações aos administradores das empresas foco dos estudos, não sendo relatado como realmente foram aplicados os resultados do processo. Então cabe a esses administradores tomarem as providências para a aplicação do conhecimento gerado ao negócio, para então, ser possível aferir em termos quantitativos o valor agregado obtido pela geração de conhecimento.

Isso se deve ao fato do conhecimento, bem como o termo Inteligência Competitiva serem muito abstratos. As Tecnologias da Informação dão suporte a ambos, mas, o conhecimento, a inteligência, o saber, são difíceis de serem medidos. Isso é possível somente após a aplicação destes às atividades da organização, que poderão então, ser monitoradas, fazendo transparecer seu valor.

Este estudo também mostrou que o analista de negócios é a peça fundamental para que haja retorno financeiro a partir da análise das informações obtidas na Mineração de Dados. O analista de dados é responsável por realizar o processo KDD, mas o analista de negócios é quem tem conhecimento do negócio da empresa e está em melhores condições de explorar o potencial das informações geradas. No entanto, mesmo diante da dificuldade da análise proposta, foi possível chegar a algumas importantes conclusões.

Conforme a revisão bibliográfica feita neste trabalho, a Tecnologia da Informação tem importante papel na implementação dos mecanismos de captura, armazenamento e disseminação de informações necessárias ao funcionamento de um Sistema de Inteligência Competitiva.

A Mineração de Dados é uma tecnologia que permite a criação de um Sistema de Apoio à Decisão, possibilitando o processamento de grandes volumes de dados e a geração de conhecimento útil. Assim, levando em conta os tipos de informações geradas nos trabalhos estudados, a Mineração de Dados trabalha principalmente sobre os dados internos da empresa, diretamente coletados dos seus bancos de dados, dando assim, suporte à gestão do conhecimento da organização, que por sua vez fornece grande apoio à Inteligência Competitiva. Portanto, uma organização capaz de aprender com suas próprias informações é uma organização inteligente, que possui vantagem sobre os concorrentes, conforme abordado no capítulo 2.

O processo KDD, possuindo grande potencial de utilização, também pode ser aplicado focando informações externas à organização, que, no entanto, demandará mais trabalhoso no que diz respeito à coleta dos dados.

Foi possível verificar ainda nos trabalhos estudados que a Mineração de Dados não é milagrosa. É preciso muito esforço em todas as etapas do processo KDD. Especial atenção é dada à interpretação dos resultados gerados para validá-los quanto à sua utilidade ao negócio da empresa. Daí a necessidade de investimentos em profissionais capacitados para a obtenção de resultados efetivos, principalmente na interpretação dos resultados para geração conhecimento. É muito importante ainda, o acompanhamento do processo por um analista de negócios, conforme recomendado pelos autores estudados. Portanto, só é possível extrair conhecimento se as informações geradas forem corretamente interpretadas.

Dessa forma, a partir do estudo realizado como um todo, principalmente dos trabalhos analisados, que deram uma ampla visão do processo de descoberta de conhecimento, é possível citar os seguintes benefícios da Mineração de Dados para a Inteligência Competitiva:

- As informações geradas pela Mineração de Dados têm grande potencial de utilização para pequenas e médias empresas, aplicadas principalmente à inteligência de *marketing* e detecção de fraudes;

- A Mineração de Dados possibilita melhor relacionamento com os clientes, pois permite traçar seu perfil através da análise de agrupamentos;
- A modelagem preditiva da Mineração de Dados permite à Inteligência Competitiva agir antecipadamente diante do mercado, uma das suas principais características;
- A Inteligência Competitiva necessita analisar informações e conhecimentos internos e externos à organização, sendo a Mineração de Dados um mecanismo muito eficiente no suprimento dessa necessidade;
- A Prospecção Informacional, tarefa da Inteligência Competitiva que identifica as informações relevantes para a organização e que devem ser controladas, está diretamente ligada ao Pré-Processamento do processo KDD, uma vez que este identifica os dados a serem minerados, fonte de provável conhecimento e, portanto, devem ser monitorados pela Inteligência Competitiva;
- Um processo de Mineração de Dados feito periodicamente permite a realização do Monitoramento Informacional, tarefa que monitora as informações identificadas na Prospecção Informacional;
- A Mineração de Dados apóia, principalmente, a Inteligência Tática da organização, isto é, a obtenção de vantagens de mercado, busca de melhores processos e estabelecimento de parcerias;
- A Mineração de Dados, mesmo diante de dificuldades como as encontradas nos trabalhos analisados (dados muito homogêneos, classificação e descrição dos produtos falhos, etc.), se bem conduzida, é capaz de trazer algum conhecimento benéfico à inteligência da empresa;
- A aplicabilidade das informações geradas é muito abrangente, sendo limitada apenas pelo contexto da análise feita sobre as mesmas;
- Algumas informações contidas na organização são benéficas à Inteligência Competitiva apenas quando confrontadas com outras fontes, sendo que as regras de associação da Mineração de Dados desempenham bem esta tarefa;

- A Mineração de Dados se mostrou como um insumo de informações e conhecimentos muito importante para a Gestão do Conhecimento e a Inteligência Competitiva;
- A Mineração de Dados promove uma mudança de paradigma nos administradores de pequenas e médias empresas, que passam a enxergar o seu negócio de maneira diferente, estimulando a geração de Inteligência Competitiva.

Além dos benefícios citados acima, a pesquisa realizada possibilitou, com base na experiência relatada pelos autores abordados, a definição de algumas recomendações para o desenvolvimento de um processo de descoberta de conhecimento em banco de dados de sucesso, a saber:

- Contar com profissionais capacitados para a realização do processo de Mineração de Dados;
- Conhecer o contexto da empresa e os processos de negócio, tentando perceber quais as informações mais importantes para a mesma;
- Definir o problema a ser resolvido, isto é, estabelecer objetivos e o escopo do projeto;
- Escolher uma metodologia de Mineração de Dados para guiar o processo;
- Selecionar com cautela os dados necessários à mineração para não comprometer os resultados;
- Caso não exista onde será aplicado o processo, e seja possível, desenvolver um *Data Warehouse*, que facilita o trabalho de mineração, devido a melhor organização dos dados, além de proporcionar a obtenção de muitas informações gerenciais;
- Fazer o conhecimento e contextualização com os dados antes de iniciar a mineração, o que pode auxiliar no entendimento dos resultados e geração de informações mais relevantes, bem como possibilitar o ajuste de parâmetros para o melhoramento dos mesmos;
- Na aplicação dos algoritmos de mineração de dados, é preciso refazer o processo diversas vezes para resultados expressivos, alterando parâmetros e atributos a cada ciclo;

- Realizar todo o processo de Mineração de Dados com o apoio de um analista de negócios, que será capaz de auxiliar na validação das regras geradas, avaliando se são relevantes para o negócio ou não e guiando o analista de dados na obtenção de resultados satisfatórios e na sua interpretação.

Diante disso, é possível se dizer que a Mineração de Dados é muito benéfica à Inteligência Competitiva da organização, pois permite, a geração de conhecimento útil a partir de informações internas e externas, identificação de informações relevantes e seu monitoramento, obtenção de vantagem de mercado, elaboração de mecanismos de *marketing* direcionado, prevenção à fraudes, entre outros. Mas, sobretudo instiga muitos administradores a verem seu negócio com outra perspectiva, estimulando a criação de um sistema de inteligência da organização.

Conclui-se, então que é vantajoso investir na Mineração de Dados, pois o processo de descoberta de conhecimento, se bem conduzido, e levando em conta as recomendações citadas acima, muito provavelmente trará retorno do investimento feito. Em muitos casos pode até mesmo ampliar a visão gerencial dos administradores e mudar a maneira de trabalhar e capturar informações em pró da alimentação de sistemas de inteligência, com o intuito de obter vantagem frente aos concorrentes.



## 5. CONCLUSÃO

Este capítulo apresenta as considerações finais do trabalho realizado, analisando a qualidade da pesquisa realizada, se houve sucesso no alcance dos objetivos definidos e quais as dificuldades encontradas. É apresentado também sugestões de trabalhos futuros para a continuidade desta pesquisa visando alcançar novas descobertas.

### 5.1 Considerações Finais

A pesquisa realizada pelo presente trabalho possibilitou reunir um material amplo e de qualidade, útil para o estudo da Inteligência Competitiva e a Mineração de Dados. Apesar de serem sintéticos, os capítulos 2 e 3 apresentam as principais características de ambos, respectivamente.

De modo geral, apesar das dificuldades encontradas, o trabalho obteve sucesso, pois conseguiu responder aos questionamentos propostos. A maior dificuldade enfrentada foi o fato de os trabalhos analisados não apresentarem muitas informações sobre a aplicação efetiva do conhecimento gerado pela Mineração de Dados aos processos da empresa. A investigação mais profunda dessa questão é, inclusive, sugerida para futuras pesquisas.

O investimento em técnicas de Mineração de Dados é benéfico à Inteligência Competitiva se o processo KDD for bem realizado, uma vez que traz conhecimento que pode ser potencializado em vantagens muitas vezes ímpares às empresas, conforme abordado com detalhes no final do capítulo 4.

Dentre essas vantagens pode-se destacar a geração de conhecimento útil a partir de informações internas e externas à organização, identificação de informações relevantes e seu monitoramento, obtenção de vantagem de mercado, elaboração de mecanismos de *marketing* direcionado, prevenção à fraudes, além de promover a ampliação da visão do negócio com informações antes desconhecidas, pois muitas vezes trazem resultados inesperados e surpreendentes.

Adicionalmente o trabalho permitiu elaborar, com base no estudo dos trabalhos analisados, uma lista de recomendações para a condução de um processo de descoberta de conhecimento de sucesso.

Ressalta-se, porém, que o retorno financeiro do investimento feito é obtido apenas se o processo KDD for conduzido por profissionais capacitados, e se os conhecimentos gerados forem bem aplicados aos processos organizacionais.

Nesse contexto, grande importância deve ser dada à interpretação dos resultados, pois só assim é possível gerar conhecimento relevante. Em seguida é necessária a elaboração de planos de ação para a aplicação do conhecimento gerado, permitindo a melhoria dos processos da organização, bem como seu posicionamento inteligente, que trarão vantagens diante dos concorrentes. Dessa forma é compensador investir em Mineração de Dados aos poucos, para constituir futuramente um sistema completo de Inteligência Competitiva.

Finalmente, diante da pesquisa realizada acerca dos temas, questões importantes foram descobertas, sem mesmo estarem diretamente relacionadas aos objetivos definidos inicialmente. Esse fato fez que o trabalho fosse muito proveitoso e espera-se que o conteúdo aqui apresentado tenha grande valia para a aquisição de conhecimento sobre os referidos temas e na decisão de adotar ou não a utilização de técnicas de Mineração e Dados.

## **5.1 Trabalhos Futuros**

Visando a continuidade desta pesquisa com o intuito de conseguir novas descobertas e ainda torná-la mais abrangente, sugere-se que sejam realizados os seguintes trabalhos:

- Entrar em contato com as empresas foco dos trabalhos estudados para efetuar uma pesquisa que demonstre os resultados financeiros advindos após a aplicação das informações geradas pelo processo de mineração de dados realizado;
- Realizar uma análise mais abrangente abordando mais trabalhos em contextos diferentes dos que já foram analisados;

- Elaborar uma metodologia de aplicação de Mineração de Dados com base nas experiências relatadas na literatura.

## REFERÊNCIAS

ABRAIC. Associação Brasileira de Analistas de Inteligência Competitiva. Disponível em <<http://www.abraic.org.br/v2/>>. Acesso em 02/06/2011.

ANTONIAZZI, Elisiane Ap.; STEFANO, Silvio Roberto. **Inteligência Competitiva: Estudo de Caso em uma Cooperativa Agro-Industrial do Estado do Paraná**. 2007. Disponível em <[http://web03.unicentro.br/especializacao/Revista\\_Pos/P%C3%A1ginas/6%20Edi%C3%A7%C3%A3o/Aplicadas/PDF/15-Ed6\\_CS-Intel.pdf](http://web03.unicentro.br/especializacao/Revista_Pos/P%C3%A1ginas/6%20Edi%C3%A7%C3%A3o/Aplicadas/PDF/15-Ed6_CS-Intel.pdf)>. Acesso em 02/06/2011.

AZARIAS, Priscila; MATOS, Simone Nasser; CARVALHO, Hélio Gomes. **Contribuições da Mineração de Dados para Gestão do Conhecimento**. Disponível em <[http://www.conged.pr.gov.br/arquivos/File/CONGED\\_Artigos/S1A6.pdf](http://www.conged.pr.gov.br/arquivos/File/CONGED_Artigos/S1A6.pdf)>. Acesso em 01/06/2011.

BARBOSA, Denise Chaves Carvalho. **Mineração de Dados usando o Software WizRule em Bases de Dados de Compras de TI**. 2007. 64 p. Dissertação (Mestrado) – Programa de Pós-Graduação e Pesquisa em Administração e Economia, Faculdade de Economia e Finanças IBMEC. Rio de Janeiro. 2007. Disponível em <[http://www.ibmecrj.br/sub/RJ/files/dissert\\_mestrado/ADM\\_denisebarbosa\\_fev.pdf](http://www.ibmecrj.br/sub/RJ/files/dissert_mestrado/ADM_denisebarbosa_fev.pdf)>. Acesso em 06/06/2011.

BARNEY, Jay B.; HESTERLY, William S. **Administração Estratégica e Vantagem Competitiva**. São Paulo: Pearson Prentice Hall, 2007.

CARVALHO, Hélio Gomes de. **A Estreita Relação entre Gestão do Conhecimento e Inteligência Competitiva**. Disponível em <[http://www.abraic.org.br/V2/periodicos\\_teses/ic\\_a25.pdf](http://www.abraic.org.br/V2/periodicos_teses/ic_a25.pdf)>. Acesso em 02/06/2011.

CARVALHO, Leandro da Silva. **Mineração de Dados Aplicada ao Entendimento do Comportamento do Consumidor para dar Suporte ao Processo de Tomada de Decisões**. 2009. 108 p. Dissertação (Mestrado) – Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, Universidade Federal do Rio de Janeiro. Rio de Janeiro. 2009. Disponível em <<http://www.dominiopublico.gov.br>>. Acesso em 15/06/2011.

CORRÊA, Ulisses. **Mineração de Dados de Help Desk Usando Rattle – O Caso Petrobras**. 2007. 92 p. Dissertação (Mestrado) – Programa de Pós-Graduação e Pesquisa em Administração e Economia, Faculdade de Economia e Finanças IBMEC. Rio de Janeiro. 2007. Disponível em <[http://www.ibmecrj.br/sub/RJ/files/dissert\\_mestrado/ADM\\_ulissescorrea\\_fev.pdf](http://www.ibmecrj.br/sub/RJ/files/dissert_mestrado/ADM_ulissescorrea_fev.pdf)>. Acesso em 06/06/2011.

DIAS, Maria Madalena. **Parâmetros na Escolha de Técnicas e Ferramentas de Mineração de Dados**. Disponível em <<http://periodicos.uem.br/ojs/index.php/ActaSciTechnol/article/viewFile/2549/1569>>. Acesso em 15/12/2010.

DZIEKANIAK, Gisele. **Tecnologias de descoberta de conhecimento na gestão do conhecimento: contextualizações com a sociedade do conhecimento**. 2010. Disponível em <[http://www.dgz.org.br/fev10/Art\\_05.htm](http://www.dgz.org.br/fev10/Art_05.htm)>. Acesso em 01/06/2010.

GONÇALVES, Loren Pinto Ferreira. **Avaliação de Ferramentas de Mineração de Dados como Fonte de Dados Relevantes para a Tomada de Decisão: Aplicação na Rede Unidão de Supermercados, São Leopoldo-RS**. 2001. 104 p. Dissertação (Mestrado) – Grupo de Estudos em Sistemas de Informação e de Apoio à Decisão, Escola de Administração, Programa de Pós-Graduação em Administração, Universidade Federal do Rio Grande do Sul. Porto Alegre. 2001. Disponível em <<http://www.lume.ufrgs.br/bitstream/handle/10183/1994/000313171.pdf?sequence=1>>. Acesso em 02/06/2010.

GOUVEIA, Roberta Macêdo Marques. **Mineração de Dados em Data Warehouse para Sistema de Abastecimento de Água**. 2009. 147 p. Dissertação (Mestrado) – Programa de Pós-Graduação em Informática, Departamento de Informática, Centro de Ciências Exatas e da Natureza, Universidade Federal da Paraíba. João Pessoa. 2009. Disponível em <[http://www.lenhs.ct.ufpb.br/html/downloads/serea/teses/teses/dissertacao\\_roberta.pdf](http://www.lenhs.ct.ufpb.br/html/downloads/serea/teses/teses/dissertacao_roberta.pdf)>. Acesso em 15/06/2011.

INFO. Revista Info (Web Site). Disponível em <[www.info.abril.com.br](http://www.info.abril.com.br)>. Acesso em 26/06/2011.

MACEDO, Dayana Carla de; REZENDE, Marcos Antônio. **Contribuições da Mineração de Dados para a Inteligência de Clientes utilizada na Análise de Mercado**. Disponível em <[http://www.pg.utfpr.edu.br/epege-2010/artigos/Contribuicoes\\_da\\_Mineracao\\_de\\_Dados\\_para\\_a\\_Inteligencia\\_de\\_C.pdf](http://www.pg.utfpr.edu.br/epege-2010/artigos/Contribuicoes_da_Mineracao_de_Dados_para_a_Inteligencia_de_C.pdf)>. Acesso em 02/06/2011.

MONTGOMERY, Cynthia A.; PORTER, Michael E. **Estratégia: A busca da vantagem competitiva**. Rio de Janeiro: Campus, 1998.

PORTER, Michael E. **Vantagem Competitiva: Criando e sustentando um desempenho superior**. Rio de Janeiro: Campus: 1992.

SCHAEFFER, André Gustavo. **Data Mining no Varejo: estudo de caso para loja de materiais de construção**. 2003. 86 p. Dissertação (Mestrado) – Programa de Pós-Graduação em Computação, Instituto de Informática, Universidade Federal do Rio Grande do Sul. Porto Alegre. 2003. Disponível em <<http://www.lume.ufrgs.br/bitstream/handle/10183/3430/000400435.pdf?sequence=1>>. Acesso em 15/06/2011.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao Data Mining Mineração de Dados**. Rio de Janeiro: Editora Ciência Moderna. 2009. 900 p.

TERRA, José Cláudio Cyrineu. **Gestão do Conhecimento e Inteligência Competitiva**. Disponível em  
<<http://www.terraforum.com.br/biblioteca/Documents/Gestao%20do%20Conhecimen%20e%20Inteligencia%20Competitiva.pdf>>. Acesso em 02/06/2011.

VALENTIM, Mara Lúcia Pomim. **Inteligência Competitiva em Organizações: Dado, Informação e Conhecimento**. Disponível em  
<[http://www.dgz.org.br/ago02/Art\\_02.htm](http://www.dgz.org.br/ago02/Art_02.htm)>. Acesso em 01/06/2011.

VALENTIM, Marta Lúcia Pomim; et. al. **O Processo de Inteligência Competitiva em Organizações**. 2003. Disponível em  
<[http://www.arquivar.com.br/espaco\\_profissional/sala\\_leitura/artigos/Processo\\_de\\_Inteligencia\\_Competitiva\\_em\\_Organizacoes.pdf](http://www.arquivar.com.br/espaco_profissional/sala_leitura/artigos/Processo_de_Inteligencia_Competitiva_em_Organizacoes.pdf)>. Acesso em 01/06/2011.