

EXTRAÇÃO DE CONHECIMENTO ATRAVÉS DA MINERAÇÃO DE DADOS

Dayana Carla de Macedo (UTFPR/UEPG) dayanamacedo@yahoo.com.br

Simone Nasser Matos (UTFPR) snasser@utfpr.edu.br

Resumo: Para garantir sua permanência no mercado e competitividade as organizações constantemente buscam conhecimentos para fundamentar a tomada de decisão. Como muitas vezes esse conhecimento está oculto em uma grande base de dados, é necessário recorrer a um sistema de tecnologia da informação para auxiliar esse processo de extração de conhecimento. Nesse sentido a Mineração de Dados é uma alternativa que visa extrair conhecimento de um grande volume de dados, descobrindo novas correlações, padrões e tendências entre as informações de uma empresa. Este artigo tem como objetivo abordar os principais conceitos através de uma breve revisão bibliográfica sobre a extração de conhecimento através da Mineração de Dados.

Palavras chave: Gestão do Conhecimento Organizacional, Mineração de Dados, extração de conhecimento.

KNOWLEDGE EXTRACTION THROUGH OF THE MINING OF DATA

Abstract: To guarantee its permanence in the market and competitiveness the organizations constantly look for knowledge to base the taking of decision. As a lot of times this knowledge is occult in a big base of data, it is necessary to appeal to a system of technology of the information to aid that process of knowledge extraction. In that sense the Mining of Data is an alternative that seeks to extract knowledge of a great volume of data, discovering new correlations, patterns and tendencies among the information of a company. This article has as objective to approach the main concepts through a brief bibliographical revision about the knowledge extraction through of the Mining of Data.

Key words: Management knowledge organizational, Mining of Data, knowledge extraction.

1. INTRODUÇÃO

Atualmente as organizações possuem em seus sistemas de informações grandes bases de dados, porém a análise destas, requer a utilização de ferramentas da tecnologia da informação que auxiliam esse processo. Com o intuito de tomar decisões de forma inteligente, segura e confiável as organizações analisam de forma precisa, grandes volumes de dados visando minimização de riscos e resultados não satisfatórios. Nesse sentido utilizar ferramentas que possam extrair informações de uma base de dados em forma de conhecimento voltado para a tomada de decisão resulta em otimização e flexibilização de processos gerenciais. Por meio de grandes avanços na área da tecnologia da informação há a possibilidade de armazenarem grandes e múltiplas bases de dados seja de natureza comercial, administrativa, governamental e científica. Mas por outro lado, a análise de um grande volume de dados pelo homem é inviável sem o auxílio de ferramentas computacionais apropriadas (GOLDSCHIMIDT & PASSOS 2005). Segundo Fayyad (1996) o homem não está preparado para interpretar uma grande quantidade de dados. Sendo uma das alternativas

para o gestor a utilização de técnicas e ferramentas, destacando o processo de descoberta de Conhecimento – KDD (Knowledge Discovery in Databases) onde uma das etapas é a Mineração de Dados. Como a informação e o conhecimento são os elementos essenciais de qualquer organização em um mercado altamente competitivo, através da mineração de dados há a possibilidade de descobrir novos e úteis conhecimentos para o processo de tomada de decisão.

O presente artigo tem como objetivo abordar os principais conceitos sobre Mineração de Dados através de uma pesquisa bibliográfica, onde a seção 2 apresenta os principais conceitos sobre o processo de KDD, bem como suas etapas. A seção 3 apresenta as principais ferramentas de mineração de dados. Na seção 4 aborda a extração de conhecimento através da mineração e na seção 5 reporta a conclusão do trabalho.

2. KDD – DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS

O termo KDD - Knowledge Discovery in Databases, foi formalizado em 1989 em referência ao amplo conceito de procurar conhecimento a partir de base de dados. Segundo Fayyad (1996) “KDD é um processo, de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados”. Segundo Goldschmidt & Passos (2005) o termo iterativo sugere a possibilidade de repetições integrais ou parciais do processo de KDD e a expressão não trivial alerta para a complexidade normalmente presente na execução de processos de KDD. Já com relação a expressão padrão válido indica que o conhecimento deve ser verdadeiro e adequado ao contexto da aplicação de KDD e o termo padrão novo deve acrescentar novos conhecimentos aos existentes, para que todo esse processo gere conhecimento útil que pode ser aplicado de forma a proporcionar benefícios ao contexto de aplicação de KDD. Porém, a extração de conhecimento de uma grande base de dados através da aplicação de um processo de KDD exige a melhor compreensão das diferenças entre dado, informação e conhecimento, conforme ilustra a figura 1.

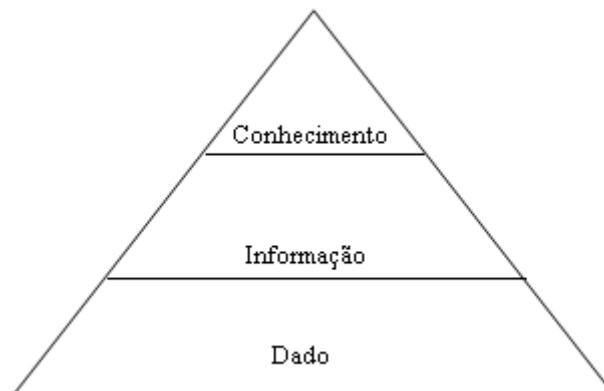


Figura 1: Hierarquia entre Dado, Informação e Conhecimento (ALESSIO, 2004).

Os dados, na base da pirâmide, podem ser interpretados como itens elementares, captados e armazenados por recursos da Tecnologia da informação (GOLDSHIMIDT & PASSOS, 2005). Ele é, basicamente, um evento fora de contexto e sem uma relação de significado com qualquer outra coisa (CARVALHO, 2000). Com relação a informação Drucker (2001) afirma que, “informação é dado investido de relevância e propósito”. Para Prusak e Mcgee (1994), não é simplesmente um conjunto de dados; “informação são dados coletados, organizados, ordenados, aos quais são atribuídos significados e contexto”. No topo da Pirâmide encontra-se o conhecimento que é o padrão ou conjunto de padrões, cuja

formulação pode envolver e relacionar dados e informações (GOLDSCHMIDT & PASSOS, 2005).

Porém, como os processos de descoberta de conhecimento em bases de dados são complexos deve-se realizar um planejamento das atividades, ou seja, estabelecer os objetivos de acordo com a aplicação de KDD com o objetivo da extração do conhecimento da base de dados, que devem ser planejados desde o início do processo. Para aplicar um processo de KDD deve-se primeiramente definir “o que fazer” diante da base de dados apresentada, estabelecendo os objetivos incluindo as seguintes atividades de acordo com (BOENTE, GOLDSCHMIDT & ESTRELA, 2006):

- _ Identificar pessoas e áreas envolvidas com a aplicação de KDD;
- _ Levantar o hardware e software existente;
- _ Inventar bases de dados disponíveis;
- _ Analisar bases de dados, procurando compreender o significado e a relevância dos atributos, avaliar a qualidade e a quantidade dos dados disponíveis;
- _ Esboçar uma lista inicial de necessidades e expectativas por parte das pessoas e áreas envolvidas;
- _ Identificar e documentar conhecimento prévio.

Para definição dos objetivos em um processo de KDD requer, primeiramente, um entendimento claro da situação vigente no ambiente onde será realizado o processo. A etapa de definição de objetivos requer uma forte interação entre o analista de KDD e os especialistas do domínio da aplicação. A Figura 1 ilustra as principais funções de um analista de KDD.

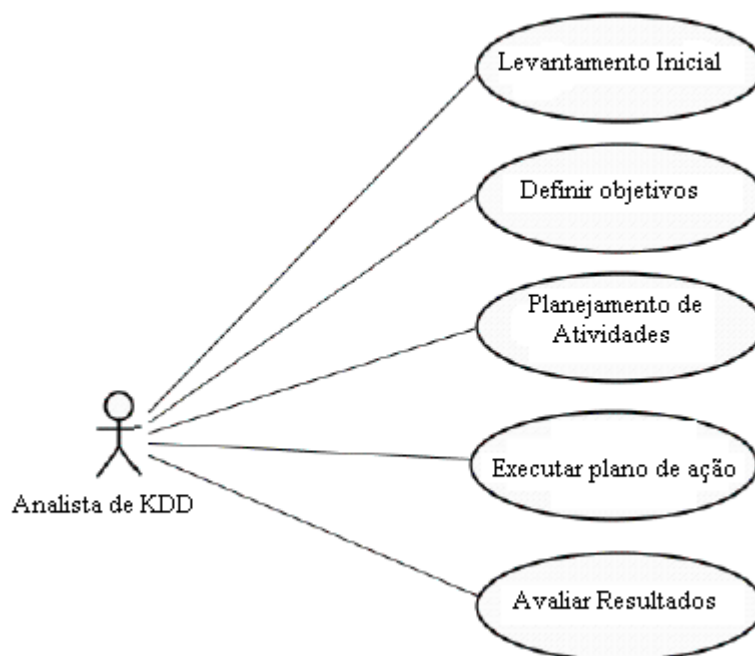


Figura 1: Diagrama de implementação da Metodologia do KDD (GOLDSCHMIDT & PASSOS, 2005).

Após toda a etapa de levantamento inicial, definição de objetivos, planejamento de atividades, execução do plano de ação e avaliação de resultados do processo de KDD, para que os dados sejam extraídos de forma eficiente, os dados são submetidos a esse processo, que possui o *Data Mining* ou mineração de dados como uma das etapas principais, sendo o

núcleo do processo, pois durante essa etapa é realizada a busca efetiva por conhecimentos úteis no contexto da aplicação de *KDD*. A etapa de descoberta de conhecimento possui várias etapas operacionais, composto pelas etapas de pré-processamento, mineração de dados e pós-processamento.

2.1. PRÉ-PROCESSAMENTO

Para iniciar a etapa de pré-processamento é necessário efetuar a seleção de dados considerada importante para a organização, ou seja, selecionar um conjunto de dados, pertencentes a um domínio, para que, a partir de um critério definido pelo especialista do domínio, possa ser analisado. Após seleção de dados, é necessário aplicar métodos de tratamento, pois na maioria das vezes os dados disponíveis para análise encontram-se em um formato inadequado para realização do processo de *KDD*, denominado de limpeza dos dados. Após o processo de limpeza dos dados é necessário realizar a codificação dos dados, com o intuito de que estejam na forma correta para serem usados como entrada dos algoritmos de mineração de dados, para posteriormente enriquecê-los de forma a agregar de alguma forma mais informação para o processo de extração de conhecimento. A seguir encontram-se descritas as principais funções de pré-processamento dos dados segundo Boente & Goldschmidt & Estrela (2006):

- _ Seleção de dados: nesta função é necessário efetuar a identificação de quais informações da base de dados existentes devem ser efetivamente consideradas durante o processo de *KDD*.
- _ Limpeza de dados: Com o intuito de assegurar a qualidade relacionada a completude, veracidade e integridade é realizado uma espécie de tratamento sobre os dados, ou seja, informações errôneas ou inconsistentes nas bases de dado devem ser corrigidos de forma a não comprometer a conhecimento a ser extraído no final do processo de *KDD*.
- _ Codificação dos Dados: Para a utilização dos dados como entrada dos algoritmos de Mineração de Dados na forma correta, estes devem ser codificados, podendo ser: Numérica – Categórica, que transforma valores reais em categorias ou intervalos; ou Categórica – Numérica, que representa numericamente valores de atributos categóricos.
- _ Enriquecimento dos dados: Têm como objetivo agregar mais informações aos registros existentes, enriquecendo os dados, para que estes forneçam mais informações para o processo de *KDD*, podendo ser realizadas as pesquisas para complementação dos dados, as consultas a bases de dados externas, entre outras técnicas.

É importante ressaltar que na etapa de limpeza de dados Goldschmidt & Passos (2005) identifica as seguintes funções que podem ser aplicadas para a limpeza de dados:

- _ Limpeza de informações ausentes: compreende a eliminação de valores ausentes em conjunto de dados;
- _ Limpeza de inconsistências: abrange a identificação e a eliminação de valores inconsistentes em conjunto de dados;
- _ Limpeza de valores não pertencentes ao domínio: compreende a identificação e a eliminação de valores que não pertençam ao domínio dos atributos do problema.

Outro ponto importante a ser considerado na etapa de pré-processamento é que em virtude das restrições de espaço em memória ou tempo de processamento com relação ao número de exemplos e de atributos disponíveis para a análise, pode inviabilizar a utilização de algoritmos de extração de padrões, sendo necessária a aplicação de métodos de redução de dados antes de

iniciar a busca pelos padrões (REZENDE, 2003).

2.2. MINERAÇÃO DE DADOS

Etapa do *KDD* que consiste na aplicação de algoritmos específicos, que extraem padrões a partir dos dados (FAYYAD et al., 1996). Nesse sentido de acordo com o objetivo da extração de conhecimento deve-se escolher o algoritmo apropriado para o processo de *KDD*. As principais tarefas de mineração de dados encontram-se são:

Descoberta de Associação: “Uma tarefa da *Data Mining* que tem por objetivo encontrar relacionamentos ou padrões freqüentes que ocorrem em um conjunto de dados” (BRUSSO, CERVO & GEYER, 2002); ou seja, é a busca por itens que ocorram de forma simultânea freqüentemente em transações do banco de dados. Algoritmos tais como o Apriori (AGRAWAL, 1993), GSP, DHP, entre outros, são exemplos de ferramentas que implementam a tarefa de descoberta de associações (GOLDSCHMIDT & PASSOS, 2005).

- _ Classificação: Tem como objetivo encontrar uma função para mapear um conjunto de registros em um conjunto de rótulos categóricos pré-definidos, denominados classes. Após descoberta essa função, esta pode ser aplicada a novos registros de forma a prever a classe que tais registros se enquadram. Exemplos de tecnologias que podem ser aplicados na tarefa de classificação como Redes Neurais (DIAS & PACHECO, 2005), Algoritmos Genéticos (DIAS & PACHECO, 2005) e Lógica Indutiva (DUARTE, 2001).
- _ Regressão: semelhante a tarefa de classificação, na qual analisam valores numéricos, tendo com principal objetivo apresentar uma previsão a partir de dados históricos contido em uma base de dados, ou seja, compreende a busca por uma função que mapeie os registros de um banco de dados em valores reais. Estatística, Redes Neurais, dentre outras áreas, oferecem ferramentas para implementação da tarefa de regressão (MICHIE et al. 1994).
- _ Clusterização ou análise de agrupamento: Por meio do agrupamento de dados os *clusters* são definidos baseados em medidas de similaridade ou modelos probabilísticos. Promove a associação de um item a uma ou várias classes categóricas em que as classes são determinadas pelos dados, diversamente da classificação em que as classes são pré-definidas. A análise de *cluster* tem como objetivo verificar a existência de diferentes grupos dentro de um determinado conjunto de dados, e em caso de sua existência, determinar quais são eles (SFERRA & CORRÊA, 2003). Para implantar esse tipo de tarefa são utilizados algoritmos tais como *K-Means*, *K-Modes*, *K-Prototypes*, *K-Medoids*, *Kohonen* (CARLANTONIO, 2001);
- _ Sumarização: consiste em procurar identificar e indicar características comuns entre conjuntos de dados (GOLDSCHMIDT & PASSOS, 2005). Exemplos de tecnologias que podem ser aplicadas para implantar a tarefa de sumarização são: Lógica indutiva e Algoritmos Genéticos.
- _ Detecção de Desvios: Na totalidade do contexto a identificação de registros que não são considerados normais, ou seja, que não atendem ao padrão considerado normal é o objetivo principal da tarefa da detecção de desvios (BOENTE, GOLDSCHMIDT & ESTRELA, 2006);
- _ Descoberta de Sequências: A análise de sequências é usada para encontrar padrões de dados escondidos numa sequência de estados temporais. É uma extensão da tarefa de descoberta de associações em que se busca itens frequentes considerando-se várias transações ocorridas ao longo de um período.

A Tabela 1 ilustra as principais tarefas da mineração de dados e os tipos de algoritmos que

utilizam.

Tabela 1: Tarefas da Mineração de dados e os Algoritmos

| Estratégia | Algoritmos |
|---------------|--|
| Classificação | Árvores de Decisão e Redes Neurais |
| Agregação | Métodos Estatísticos e Redes Neurais |
| Associação | Métodos Estatísticos e Teoria de conjuntos |
| Regressão | Métodos de Regressão e Redes Neurais |
| Predição | Métodos Estatísticos e Redes Neurais |

Fonte: BOENTE, GOLDSCHMIDT & ESTRELA, 2006.

Por meio das tarefas de Mineração de Dados é possível extrair conhecimento de um grande volume de dados, descobrindo novas correlações, padrões e tendências entre as informações de uma empresa.

2.3. PÓS-PROCESSAMENTO

A etapa de Pós-processamento compreende o processo de tratamento do conhecimento adquirido por meio da Mineração de Dados, com o intuito de facilitar a interpretação e avaliação deste, para priorizar a utilidade do conhecimento descoberto. Dentre as principais funções da etapa de Pós-processamento estão a elaboração e organização, podendo incluir a simplificação de gráficos, diagramas e outros tipos de relatórios demonstrativos, além da conversão da forma de representação do conhecimento extraído no processo de *KDD* (BOENTE, OLIVEIRA & ROSA, 2007). A seguir encontram-se comentadas os métodos e procedimentos utilizados na etapa de pós-processamento segundo Nicolai & Pelinski (2006):

- Avaliação: etapa onde o objetivo maior é a avaliação do conhecimento extraído da base de dados por meio de critérios, tais como precisão, compreensibilidade e interessabilidade.
- Interpretação e Explanação: consiste em tornar o conhecimento extraído compreensível ao usuário, ou seja, documentá-lo, visualizá-lo, modificá-lo e/ou compará-lo o conhecimento pré-existente com o intuito de compreender melhor o conhecimento descoberto no processo de *KDD*.
- Filtragem: consiste em filtrar o conhecimento extraído do conjunto de dados, realizado por meio de mecanismo que variam de acordo com a técnica utilizada, para que posteriormente a análise do conhecimento na etapa de Pós-processamento, este possa ser utilizado no processo de tomada de decisão;
- Interpretação: fase que inclui o processo de interpretação do modelo descoberto, onde pode requerer a repetição de vários passos, porém normalmente é encarada como uma simples visualização dos dados. Os padrões identificados pelo sistema são interpretados em conhecimento, que pode ser utilizado como ferramenta de apoio ao processo de tomada de decisão.

Após a etapa de pós-processamento, o conhecimento extraído depois de avaliado e validado é consolidado na fase de utilização do conhecimento, sendo incorporado a um sistema inteligente, que é utilizado pelo usuário final para o apoio a algum processo de tomada de decisão, ou seja, relatado às pessoas interessadas.

3. FERRAMENTAS PARA A MINERAÇÃO DE DADOS

Para Rezende (2005) atualmente há um progresso da área de extração de conhecimento de uma base dados nas organizações, onde ocorre um estímulo para o desenvolvimento de ferramentas capazes de facilitar a extração do conhecimento por meio de uma variedade de técnicas, com diferentes tipos de algoritmos voltados para a Mineração de Dados. Na Tabela 2 é descrita resumidamente as principais ferramentas, disponíveis no mercado, utilizadas na Mineração de Dados.

Tabela 2: Ferramentas para a Mineração de Dados

| | Tarefas Disponíveis | Fabricante e Site de Acesso |
|-------------------------------------|--|--|
| <i>PolyAnalyst</i> | Classificação, regressão, regras de associação, clustering, sumarização e modelagem de dependência | <i>Megaputer Intelligence</i> www.megaputer.com |
| <i>DataMite</i> | Regras de associação | Dr. <i>Phillip Vasey</i> através da LPA Prolog www.ipa.co.uk/dtm.htm |
| <i>Microsoft Data Analyser 2002</i> | Classificação e clustering | <i>Microsoft Corp.</i> www.microsoft.com |
| <i>Oracle 9i Data Mining</i> | Classificação e regras de associação | <i>Oracle Corp.</i> www.oracle.com |
| <i>Darwin</i> | Classificação, regressão e clustering | <i>Oracle Corp.</i> www.oracle.com |
| <i>MineSet</i> | Classificação, regressão e clustering | <i>Slicon Graphics inc.</i> www.sgi.com |
| <i>WEKA</i> | Classificação, regressão, regras de associação | <i>University Waikato</i> www.cs.waikato.ac.nz |
| <i>Intelligent Miner</i> | Regras de associaçã, padrões sequências, classificação, clustering, sumarização e modelagem de dependência | <i>IBM Corp.</i> www.ibm.com |

FONTE: Adaptado de Rezende (2005)

Após captação dos dados considerados importantes para a organização efetuar o processo de KDD, é necessário a escolha da ferramenta adequada e o tipo de tarefa a ser executada para a extração do conhecimento, fornecendo as informações necessárias para à de tomada de decisão.

4. EXTRAÇÃO DE CONHECIMENTO ATRAVÉS DA MINERAÇÃO DE DADOS

A extração do conhecimento por meio da utilização da mineração de dados possibilita descobrir novas correlações, padrões e tendências entre as informações de uma empresa, contidas em uma grande base de dados. Esse processo oferece um diferencial para a organização, pois em tempos de concorrência acirrada, exigência de qualidade por parte dos clientes e dificuldades econômicas ter um sistema de apoio a decisão é necessário. Através da Mineração de dados é possível extrair conhecimentos que estão ocultos em um banco de dados, podendo utilizá-los como estratégias no negócio.

Segundo KREMER (1999):

Prospecção de conhecimento em bases de dados (*Knowledge Discovery in Databases - KDD*) é um processo que envolve a automação da identificação e do reconhecimento de padrões em um banco de dados. Trata-se de uma pesquisa de fronteira, que começou a se expandir mais rapidamente nos últimos cinco anos. Sua

principal característica é a extração não-trivial de informações a partir de uma base de dados de grande porte. Essas informações são necessariamente implícitas, previamente desconhecidas, e potencialmente úteis.

Atualmente muitas organizações perdem negócios ou tomam o rumo errado por tomarem decisões infundadas, ou seja, decisões que não estão fundamentadas em dados reais, obtidos de fontes seguras. Nesse sentido o conhecimento extraído de uma base de dados pode fornecer as informações necessárias e imprescindíveis para a tomada de decisão. De acordo com Mardegan, Azevedo & Oliveira (2002), devido a grande concorrência, a utilização de informações imprecisas e dados incorretos em uma empresa podem prejudicá-la em muito no processo de tomada de decisão, em sua produtividade e, inclusive, na própria permanência no mercado e competitividade.

A mineração de dados como uma tecnologia que tem como objetivo extrair conhecimentos visa coletar dados de todas as formas e inter-relacioná-las para assim poder gerar informação, podendo ou não ser útil para o empresário, ou seja, é o conhecimento gerencial. Mesmo se esses dados não tiverem valor específico para a situação em questão, podem chamar a atenção para outras áreas de atuação da empresa, muitas vezes esquecidas, pela preocupação em manter-se focada na sua área foco (CARNIEL & STAHNKE, 2009).

Nesse sentido a utilização de Mineração de Dados nas organizações é um diferencial, pois em tempos de concorrência acirrada, exigência pela qualidade e dificuldades econômicas é necessário tomar decisões fundamentadas em dados reais para assegurar a competitividade e permanência no mercado

6. CONCLUSÃO

A mineração de dados é uma das etapas pertencentes ao processo de *KDD - Knowledge Discovery in Databases*, que é composto pelas etapas de onde visa a extração de conhecimento de uma grande base dados exigindo uma compreensão melhor das diferenças entre dado, informação e conhecimento. Por meio do conhecimento extraído este pode ser útil para os empresários, pois se trata de conhecimento gerencial fornecendo subsídios necessários para a tomada de decisão segura e confiável. Nesse sentido a utilização de Mineração de Dados nas organizações oferece um diferencial, pois em tempos de concorrência acirrada, exigência pela qualidade e dificuldades econômicas é necessário tomar decisões fundamentadas em dados reais para assegurar a competitividade e permanência no mercado.

REFERÊNCIAS

AGRAWAL, R.; SRIKANT, R. *Mining generalized association rules in large relational tables. Proc. Of 21 st Int`L Conference on Very Large Databases*. Zurique, Suíça, 1995. Disponível em: <[HTTP://rakesh.agrawal-family.com/pubs.html](http://rakesh.agrawal-family.com/pubs.html)> acesso em 07 de junho de 2009.

ALESSIO, Paulo Agostinho. *Informação e Conhecimento um modelo de gestão para potencializar a inovação e a cooperação Universidade-Empresa*. Tese (Doutorado)-Pós Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina. Florianópolis, Santa Catarina, 2004.

BOENTE, Alfredo Nazareno Pereira ; GOLDSCHMIDT ; ESTRELA, Vania Vieira . *Uma Metodologia de Suporte ao Processo de Descoberta de Conhecimento em Bases de Dados*. In: V Simpósio de Excelência em Gestão e Tecnologia, 2008, Resende - RJ. V SEGeT, 2008. v. 1, p. 4-5.

BOENTE, A. N. P. ; OLIVEIRA, F. S. G. ; ROSA, J. L. A. . *Utilização de Ferramenta de KDD para Integração de Aprendizagem e Tecnologia em Busca da Gestão Estratégica do Conhecimento na Empresa*. Anais do Simpósio de Excelência em Gestão e Tecnologia, v. 1, p. 123-132, 2007.

BRUSSO, M. J.; CERVO, L. V.; GEYER, C. F. R. Ferramenta para Descoberta de Regras de Associação em Banco de Dados Relacionadas a área da Saúde. Porto Alegre, 2002, 12 páginas. Monografia (Graduação). Universidade Federal do Rio Grande do Sul.

CARLANTONIO, Lando Mendonça di. Novas metodologias para clusterização de dados. Dissertação (Mestrado)-Engenharia Civil, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2001.

CARNIEL, Juliano D; STAHNKE, Aron M. Transformação de dados em conhecimento gerencial para apoio a decisões. Disponível em: <<http://www.dainf.cefetpr.br/~graeml/InfoDigital/artigos.htm>> acesso em 16 de junho de 2009.

CARVALHO, Hélio Gomes de. Inteligência Competitiva Tecnológica para PMEs Através da Cooperação Escola-Empresa. Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina - UFSC. Florianópolis, 2000.

DIAS, M. M. ; PACHECO, R. C. S. . Uma Metodologia para o Desenvolvimento de Sistemas de Descoberta de Conhecimento. Acta Scientiarum (UEM), Maringá - PR - Brasil, v. 27, n. 1, p. 61-72, 2005.

DRUCKER, Peter F. O Advento da Nova Organização. Série Harvard Business Review. Rio de Janeiro: Ed. Campus Ltda, 2001.

DUARTE, Denio. Utilizando Técnicas de Programação Lógica Indutiva para Mineração de Banco de Dados Relacional. Dissertação (Mestrado) – Pós Graduação em Engenharia de Produção, Universidade Federal do Paraná, Curitiba, 2001.

FAYYAD, Usama, PIATETSKY-SHAPIRO, Gregory e SMYTH, Padhraic. From Data Mining to Knowledge discovery. *American Association for Artificial Intelligence*. 1996.

GOLDSCHIMIDT, R e PASSOS, E. *Data mining: Um guia prático*. Rio de Janeiro: Campus, 2005.

KREMER, Ricardo. Sistema de apoio à decisão para previsões genéricas utilizando técnicas de Data Mining. TCC (Graduação) – Curso de Ciências da Computação, Bacharelado, Universidade Regional de Blumenau, Blumenau, Santa Catarina, 1999.

MARDEGAN, Ronaldo; AZEVEDO, Rodrigo C.; OLIVEIRA, João F. G. de. Os Benefícios da Coleta Automática de Dados no Chão-de-Fábrica para o Processo de Negócio e Gestão da Demanda. Curitiba, ENEGEP 2002.

MICHIE, D.; SPIEGELHALTER, D.; TAYLOR, C. Machine Learning, Neural and Statistical Classifications. Ellis Horwood, 1994.

NICOLAIO, Rodrigo Augusto e PELINSKI, Rogério. Estudo e aplicação da tarefa de associação de Data Mining em uma base de dados real. Trabalho de Conclusão de Curso. (Graduação em Superior de Tecnologia em Sistemas de Informação) - Universidade Tecnológica Federal do Paraná, 2007.

PRUSAK, Lawrence e McGEE, James. Gerenciamento Estratégico da Informação. Rio de Janeiro: Ed. Campus, 1994.

REZENDE, Solange Oliveira. Sistemas Inteligentes: Fundamentos e Aplicações. São Paulo: Barueri, 2003.

SFERRA, Heloisa Helena; CORREA, Ângela M. C. Jorge. Conceitos e Aplicações de *Data Mining*. Jul/Dez de 2003, Revista Ciência & Tecnologia, PP. 19-34.