

BUSINESS INTELLIGENCE (B.I.) : ANÁLISE PREDITIVA ATRAVÉS DA MINERAÇÃO DE DADOS.

Karine Rodrigues Coelho (FIAP)

karcoelho@hotmail.com

Agesandro Scarpioni (UMC)

agesandro@fiap.com.br

Edgard Riiti Massago Goi (FIAP)

goi.edgard@gmail.com

Pedro Zanela da Silva (FIAP)

pedro.zanela@hotmail.com



A nova geração de consumidores, totalmente conectada, define as diretrizes do comércio nacional e internacional. O futuro do varejo é antecipar as necessidades dos clientes, a fim de atendê-los da melhor forma possível. Os conceitos de análise preditiva Business Intelligence (B.I.), quando bem aplicados, podem impulsionar os negócios e maximizar o lucro das organizações, tendo em vista que a capacidade de extrair, analisar e interpretar grande quantidade de dados vem se tornando cada dia mais essencial para a tomada de decisões estratégicas nas corporações. Baseando-se na rotina do cliente, ou seja, analisando as buscas frequentes por ele realizadas, é possível traçar o perfil do consumidor e apresentá-lo anúncios focados com base em seus interesses. O objetivo dessa pesquisa é verificar o uso da análise preditiva podendo trazer impactos positivos para as organizações. Visando obter algum conhecimento relevante, foi realizada uma pesquisa descritiva, utilizando a pesquisa bibliográfica e o método Delphi.

Palavras-chave: ANÁLISE PREDITIVA; BUSINESS INTELLIGENCE; IMPACTOS POSITIVOS;

1. Introdução

A construção de um banco de dados que integre dados a respeito dos clientes, informações de mercado e dados operacionais têm como resultado um aumento de informações. Em meio à vasta quantidade de ofertas de produtos e serviços dos mais diversos setores, é necessário que haja tempo e investimentos adequados para a análise dos dados obtidos. Uma crescente lacuna entre a capacidade de armazenamento de dados, sistemas de restauração e habilidade do alto escalão das organizações em obter vantagens competitivas por meio da análise das informações obtidas pode ser verificada. As ferramentas de mineração de dados são uma excelente forma de estruturação e priorização de informações críticas de marketing para a resolução de problemas específicos encontrados por responsáveis pela tomada de decisões nas empresas. Ademais, a mineração de dados pode ser utilizada por uma vasta gama de setores, que serão expostos neste estudo.

Duas metodologias de pesquisa para elaboração deste trabalho foram utilizadas: a pesquisa bibliográfica, onde podem ser encontrados artigos, teses nacionais e internacionais, além da pesquisa utilizando o método Delphi.

O presente artigo possui quatro tópicos: o primeiro tópico, abordará a definição *Knowledge Discovery in Databases* (KDD), bem como as fases deste processo. No segundo tópico, são expostos conceitos sobre mineração de dados, expressão proveniente do inglês *Data Mining*, além de possíveis aplicações da mineração de dados, com enfoque nas áreas do varejo e no mercado financeiro. No terceiro tópico, será exposta a metodologia utilizada, e, por fim, no quarto tópico, serão mostrados os impactos da prospecção de dados, trazendo como resultado as vantagens da utilização da análise preditiva em grandes organizações.

2. *Knowledge Discovery in Databases* – KDD (Descoberta de Conhecimento em Base de Dados)

2.1. O que é *Knowledge Discovery in Databases*

Diversos métodos de extração de informações sobre determinadas quantidades de dados vêm sendo propostos, considerando que a utilização de tecnologias com a finalidade de armazenar

e analisar dados das mais diversas áreas do conhecimento ultrapassa a capacidade de processamento humano. Conforme Cunha (2006), a Descoberta de Conhecimento em Base de Dados (*Knowledge Discovery in Databases - KDD*) surge, neste contexto.

Tal processo utiliza conceitos de ferramentas de visualização, técnicas de inteligência artificial, métodos estatísticos e base de dados, dividindo-se em diversas etapas. Para Galvão e Marin (2009), entre tais etapas, a que se faz mais importante é a mineração de dados, que comprova o pressuposto de que dados transformam-se em informação, e posteriormente em conhecimento, o que leva à sabedoria, tornando a mineração de dados um processo imprescindível para as organizações.

A figura a seguir mostra os quatro níveis de abstração.

Figura 1 – Quatro níveis de abstração



Fonte: Autores (2015)

2.2. Fases da descoberta de conhecimento em bases de dados (KDD)

O processo de *Knowledge Discovery in Databases*, ou, em português, descoberta de conhecimento em banco de dados, pode ser utilizado em qualquer tipo de banco de dados desde que tenha sido realizada uma limpeza de forma a serem analisados apenas os dados mais importantes e necessários. Conforme Bueno e Viana (2012), este processo é composto por cinco etapas, sendo:

- a) Seleção: é a primeira fase do KDD. Nesta fase, ocorre a definição sobre quais dados são relevantes para que as informações obtidas sejam úteis;

- b) Pré-processamento: Conforme afirmação realizada por Batista (2003), podem existir diversos objetivos nesta fase, como, por exemplo, solucionar problemas nos dados. Desta forma, informações errôneas ou inconsistentes nas bases de dados devem ser corrigidas de forma a não comprometer a qualidade dos modelos de conhecimento a serem extraídos ao final do processo de KDD. Esta fase depende da capacidade da pessoa que a conduz em identificar os problemas presentes nos dados e utilizar métodos apropriados com a finalidade de solucionar cada um dos problemas;
- c) Transformação: Os dados retirados do processo anterior são modificados para que a etapa de mineração de dados possa ser realizada. Ocorre a reorganização dos dados de forma específica. Um software de mineração de dados será utilizado para interpretar os dados reorganizados;
- d) Mineração de dados: Após os dados terem sido transformados na fase anterior, serão lidos e interpretados nesta fase. Os dados serão transformados em informação. É a etapa mais importante do KDD. Segundo Castanheira (2008), o cérebro humano consegue realizar apenas oito comparações ao mesmo tempo. A prospecção de dados possui como função ampliar esta comparação, e tornar isso visível ao olho humano. Esta fase será tratada especificamente no próximo capítulo;
- e) Análise dos resultados: É a última fase do processo de KDD. Após a interpretação dos dados da fase anterior, poderão surgir padrões, relacionados a novas descobertas, podendo ser utilizados em otimização, pesquisa e outros.

3. A mineração de dados

Muitas definições de mineração de dados são encontradas na literatura. Podem ser citadas, como exemplo, as definições a seguir.

“Data Mining, ou Mineração de Dados, pode ser entendido como o processo de extração de informações, sem conhecimento prévio, de um grande banco de dados e seu uso para tomada de decisões. É uma metodologia aplicada em diversas áreas que usam o conhecimento, como empresas, indústrias e instituições de pesquisa.” (SFERRA; CORRÊA, 2003)

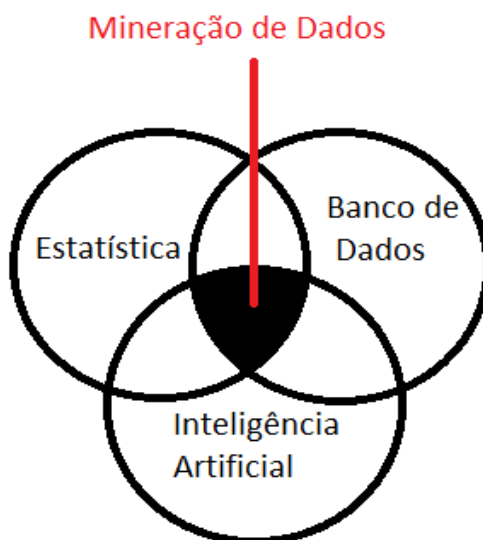
“A mineração de dados é uma das tecnologias mais promissoras da atualidade. Um dos fatores deste sucesso é o fato de dezenas, e muitas vezes centenas de milhões de reais serem

gastos pelas companhias na coleta de dados, e, no entanto, nenhuma informação útil é identificada.” (CAMILO; SILVA, 2009)

Segundo descrito por Fortulan (2006), a mineração de dados, também conhecida como prospecção de dados, envolve diferentes técnicas de abordagens, que extraem informações previamente desconhecidas, aplicáveis em áreas que apoiam as decisões organizacionais.

Informações confiáveis e consolidadas são cada vez mais necessárias para as organizações, considerando que proporcionam apoio ao gerenciamento das empresas, tornando, desta forma, a utilização da mineração de dados cada vez mais frequente. Para Bueno e Viana (2012), entre as técnicas utilizadas em mineração de dados, temos as técnicas estatísticas, técnicas de aprendizado de máquina e técnicas baseadas em crescimento-poda-validação. Entretanto, a união de inteligência artificial, banco de dados e estatística é o que torna a prospecção de dados possível, como pode ser visto na figura a seguir.

Figura 2 – O que torna a mineração de dados possível?



Fonte: Autores (2015)

4. Possíveis aplicações da mineração de dados

Para Galvão e Marin (2009), o conhecimento adquirido através da mineração de dados tem se mostrado bastante útil em diversas áreas, como medicina, finanças, varejo, marketing, telecomunicações, meteorologia, agropecuária, bioinformáticas, entre outras. Neste artigo, serão abordadas as aplicações na área do varejo e no mercado financeiro.

4.1. Varejo

Um dos grandes objetivos de uma organização é conhecer seus clientes. No âmbito comercial, a análise de dados através de sua mineração, permite a verificação da circulação de manufaturados, auxiliando nas decisões de compras de novos itens e atendimento ao público geral. Fornece uma perspectiva de padrões de comportamento observando as informações coletadas, fomentando decisões a curto e longo prazo, prevendo de forma segura prováveis comportamentos futuros. Comportamentos que fogem ao padrão (outliers), podem ser detectados, possibilitando verificar se esse é ou não um bom comportamento, se deveria ser seguido, formando um novo padrão, ou, deveria ser evitado. (NISBET et al. 2009)

A análise dos perfis de clientes, proporcionada pela interpretação dos dados, permite o agrupamento e disposição de produtos, de forma mais eficaz. Itens que inicialmente não estão diretamente ligados, podem ser dispostos de forma a alterar a disposição e organização dos itens no ambiente, a fim de melhorar o atendimento ao usuário.

O conhecimento deve ocorrer em vários níveis, desde o tipo de produto desejado até que tipo de ofertas estão dispostos a aceitar ainda que os produtos não sejam essenciais no momento. Se aplicado a análise de perfis através da mineração dos dados, a organização pode segmentar sua relação, alcançando um tratamento exclusivo a clientela.

Esta forma de mercado dirigido pode atingir uma relação individual com cada cliente à medida que a empresa deseje investir em segmentações (classificações) sucessivas de sua clientela.

4.2. Mercado financeiro

Prever o mercado financeiro é uma tarefa árdua pois a curto prazo os boatos e as especulações alteram preços muito rapidamente, a longo prazo ocorrem mudanças políticas nacionais e internacionais e outros fatos que alteram completamente as tendências. O volume de dados gerado através do mercado financeiro é exuberante, enquanto os prazos trabalhados e o tempo utilizado são curtos, requerendo uma aplicação de tratamento de dados eficaz.

Com a utilização da mineração de dados, é possível aumentar o retorno e minimizar os riscos de perda. Baseando em evidências atuais através da extração de informação via massas de

dados, as tendências do mercado após um determinado tempo podem ser previstas, auxiliando na compra e venda de papéis corretos, nos tempos mais apropriados. A organização tenta extrair dos dados passados e presentes alguma visão das variações futuras. (NISBET et al. 2009)

5. Metodologia

Com o objetivo de coletar os dados estatísticos necessários para comprovar os impactos positivos da utilização da mineração de dados, optou-se por uma pesquisa descritiva em que os meios utilizados foram a pesquisa bibliográfica e a aplicação do método Delphi.

O método de pesquisa Delphi é utilizado para obter um consenso de opiniões, através da análise de respostas de um grupo de especialistas, que, ao serem consultados, devem responder a um questionário. Este método tem por característica estruturar um processo de comunicação em grupos, para que as respostas sejam empregadas em pesquisas que serão realizadas utilizando-se destes dados. (WRIGHT; GIOVINAZZO, 2000).

O anonimato garantido aos respondentes no método Delphi mantém a integridade da opinião dos participantes, uma vez que não há influência de fatores externos, como grupos majoritários sob uma minoria.

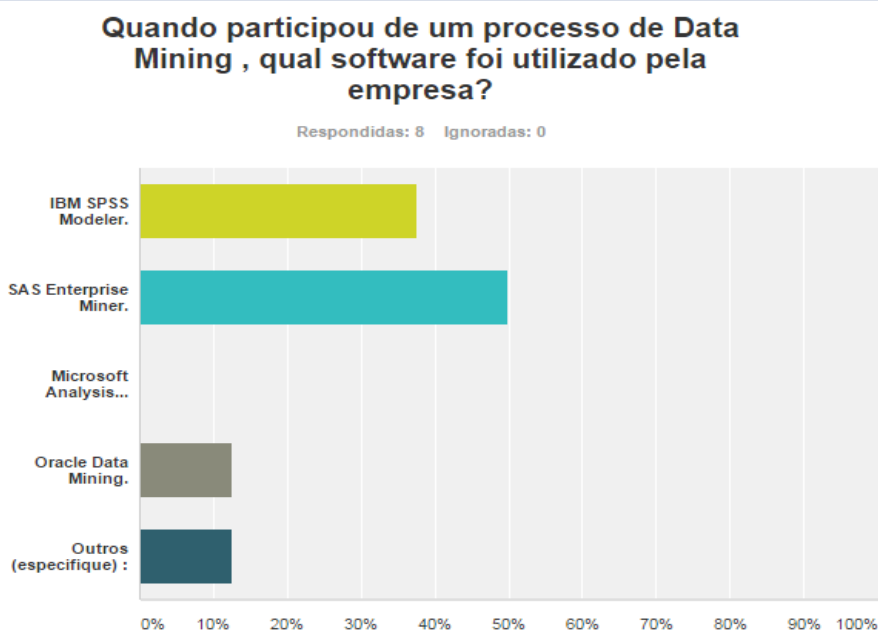
No presente artigo, o questionário foi disponibilizado ao grupo de especialistas com a utilização da ferramenta ‘SurveyMonkey’. A análise dos resultados pode ser visualizada no capítulo seguinte.

6. Análise e discussão de resultados

A amostragem foi obtida após a consulta de oito especialistas cuja empresa de atuação faz parte dos mais diversos setores, sendo que: 50% dos especialistas afirmam fazer parte do grupo de colaboradores de uma organização voltada ao setor de Tecnologia da Informação. 25% responderam que a organização é voltada à Engenharia, enquanto 12,5% dos especialistas afirmaram, respectivamente, que a empresa atua na área de Treinamentos e Educação.

Como pode ser observado na imagem abaixo, não foi obtido um consenso a respeito do *software* utilizado no processo de mineração de dados.

Figura 3 - Qual software foi utilizado pela empresa no processo de mineração de dados?



Opções de resposta	Respostas
IBM SPSS Modeler.	37,50% 3
SAS Enterprise Miner.	50,00% 4
Microsoft Analysis Services.	0,00% 0
Oracle Data Mining.	12,50% 1
Outros (especifique): Pentaho	Respostas 12,50% 1
Total de respondentes: 8	

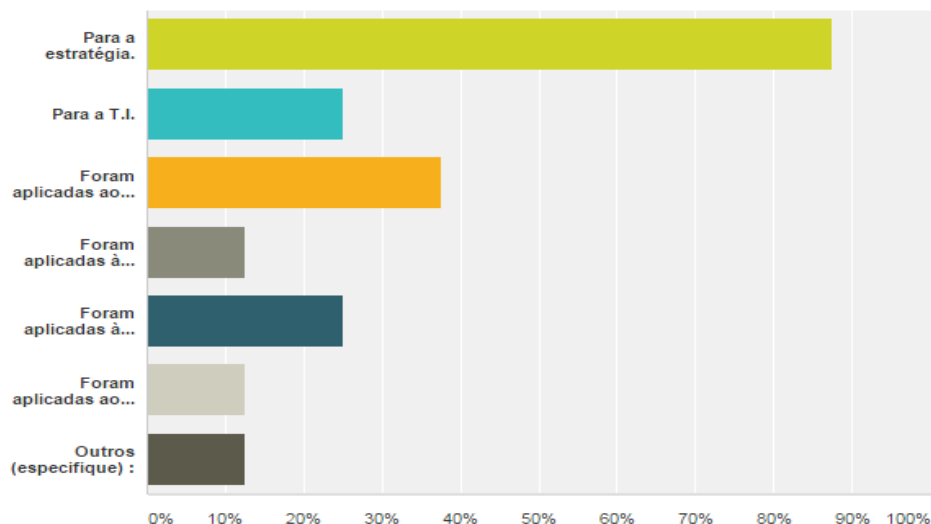
Fonte: Adaptado de SurveyMonkey (2015)

Os especialistas foram questionados a respeito da utilização das informações adquiridas pelo processo de *Data Mining*, que, conforme citado anteriormente, possui diversas aplicações. Para esta questão, foi possível a escolha de mais de uma alternativa como sendo válida. Como pode ser observado, no gráfico a seguir, 87,5% dos especialistas afirmaram que as informações foram utilizadas para a estratégia da organização. Em segundo lugar, com 37,5%, está a aplicação para o marketing.

Figura 4 - A utilização das informações adquiridas pelo processo de mineração de dados.

Como as informações adquiridas no processo de Data Mining foram/são utilizadas na organização que você trabalha/trabalhou?

Respondidas: 8 Ignoradas: 0



Opções de resposta	Respostas
Para a estratégia.	87,50% 7
Para a T.I.	25,00% 2
Foram aplicadas ao marketing.	37,50% 3
Foram aplicadas à pesquisas.	12,50% 1
Foram aplicadas à inovações.	25,00% 2
Foram aplicadas ao desenvolvimento de processos da empresa.	12,50% 1
Outros (especifique): Risco de crédito e prevenção a fraude	12,50% 1
Total de respondentes: 8	

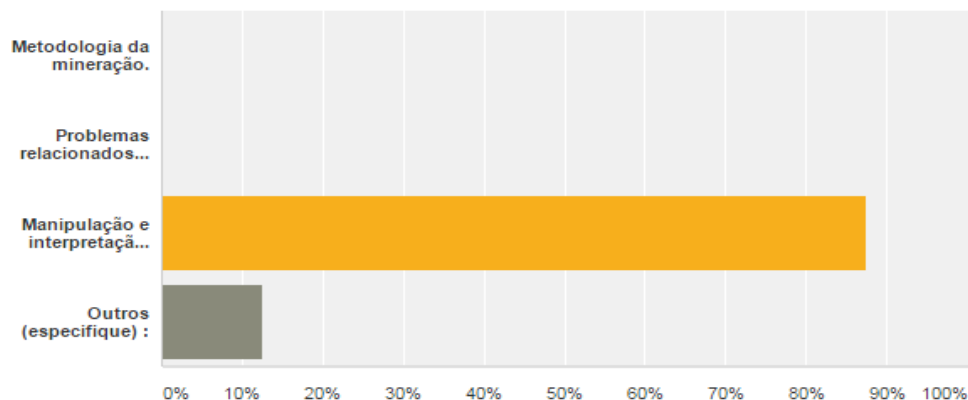
Fonte: Adaptado de SurveyMonkey (2015)

Algumas dificuldades foram apontadas pelos entrevistados a respeito da implementação do processo de *Data Mining*. 87,5% dos especialistas afirmaram que a manipulação e interpretação de diferentes tipos de dados foi a principal dificuldade no processo, enquanto 12,5% dos entrevistados citaram a disponibilidade de dados como principal dificuldade. O gráfico a seguir mostra a resposta dos especialistas.

Figura 5 - Dificuldades de implementação do *Data Mining*.

Quais as maiores dificuldades em se implementar o processo de Data Mining?

Respondidas: 8 Ignoradas: 0



Opções de resposta	Respostas
Metodologia da mineração.	0,00% 0
Problemas relacionados à performance da ferramenta.	0,00% 0
Manipulação e interpretação de diferentes tipos de dados.	87,50% 7
Outros (especifique) : Disponibilidade dos dados	12,50% 1
Total	8

Fonte: Adaptado de SurveyMonkey (2015)

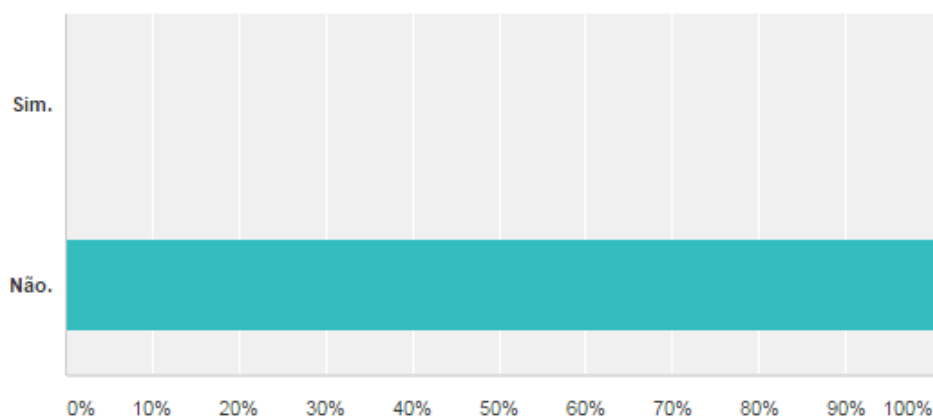
A busca de vantagens competitivas, sem as ferramentas de mineração de dados, pode ser realizada com a utilização de outros recursos. Na pesquisa realizada, 100% dos especialistas disseram obter vantagens competitivas com o aprendizado de experiências anteriores. 50% dos entrevistados afirmaram obtê-la através da indicação de clientes, enquanto 25% dos entrevistados dizem obtê-la através do endomarketing. É importante citar que, para esta questão, foi possível a escolha de mais de uma alternativa como sendo válida.

Para completa validação da pesquisa, uma pergunta a respeito da efetividade do uso da mineração de dados nas organizações foi realizada. Sabe-se que o uso da análise preditiva pelas organizações pode trazer inúmeras vantagens, em suas mais diversas formas de aplicação. Entretanto, foi constatado que a organização não obtém os resultados desejados sem o uso das ferramentas de *Data Mining*, como pode ser visualizado no gráfico a seguir.

Figura 6 - Resultados semelhantes teriam sido obtidos sem o uso da mineração de dados?

Você acredita que a organização teria obtido resultados semelhantes sem o uso da mineração de dados?

Respondidas: 8 Ignoradas: 0



Opções de resposta	Respostas	
Sim.	0,00%	0
Não.	100,00%	8
Total		8

Fonte: SurveyMonkey (2015)

7. Conclusão

Através da análise dos resultados obtidos, é possível verificar que a necessidade da aplicação da análise preditiva é cada vez mais presente para as organizações, haja vista que buscam obter vantagens competitivas, sendo elas no marketing, T.I, varejo, risco de fraude ou demais aplicações do *Data Mining*.

Apesar de dificuldades terem sido apontadas por especialistas na implementação do processo de mineração de dados, foi verificado que as organizações que não aderem a este conceito, ou seja, que ainda não utilizam de ferramentas específicas para uso em mineração de dados, não obtém os mesmos resultados obtidos com seu uso.

As corporações que optem pelo uso da prospecção de dados devem escolher entre diversas ferramentas disponíveis no mercado. A medição da eficiência de cada uma delas não está presente neste estudo, uma vez que a escolha da ferramenta a ser utilizada deve ser realizada

pela empresa em questão de acordo com o objetivo a ser atingido e área de atuação da organização.

REFERÊNCIAS

BATISTA, Gustavo Enrique de Almeida Prado Alves. **Pré-processamento de Dados em Aprendizado de Máquina Supervisionado**. São Carlos: USP, 2003. 204 p. Tese (Doutorado) - Instituto de Ciências Matemáticas e de Computação, São Carlos, 2003. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-06102003-160219/pt-br.php>>

BUENO, Michel Ferreira; VIANA, Maury Reis; Mineração de dados: aplicação, eficiência e resultados. In: Congresso de iniciação científica do INATEL – INCITEL, 2012, Santa Rita do Sapucaí. **Anais...** Minas Gerais: INATEL-INCITEL, 2012. p. 86-88.

CAMILO, Cássio Oliveira; SILVA, João Carlos da. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**. Goiás, 2009. 29 p. Relatório Técnico – Universidade Federal de Goiás, Goiás, 2009. Disponível em: <<http://inf.ufg.br/biblioteca-digital>>

CASTANHEIRA, Luciana Gomes. **Aplicação de Técnicas de Mineração de Dados em Problemas de Classificação de Padrões**. Belo Horizonte: UFMG, 2008. 95 p. Tese (Mestrado) – Departamento de Engenharia Elétrica, Belo Horizonte, 2008. Disponível em: <http://www.ppgee.ufmg.br/diss_defesas_detalhes.php?aluno=>

CUNHA, Márcio José da. **Descoberta do conhecimento em base de dados como ferramenta aplicada em processos sucroalcooleiros**. São Carlos: USP, 2011. 108 p. Tese (Doutorado) – Escola de Engenharia de São Carlos, São Carlos, 2011. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/18/18149/tde-20092011-180126/en.php>>

FORTULAN, Marcos Roberto. **O uso de business intelligence para gerar indicadores de desempenho no chão-de-fábrica: uma proposta de aplicação em uma empresa de manufatura**. São Carlos: USP, 2006. 179 p. Tese (Doutorado) – Escola de Engenharia de São Carlos, São Carlos, 2006. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/18/18145/tde-11062006-185813/pt-br.php>>

GALVÃO, Noemi Dreyer; MARIN, Heimar de Fátima. Técnica de mineração de dados: uma revisão da literatura. **Acta Paulista de Enfermagem**, São Paulo, vol. 22 no. 5, 2009. Disponível em: <http://www.scielo.br/scielo.php?pid=S0103-21002009000500014&script=sci_arttext>

NISBET, Robert; ELDER, John; MINER, Gary. **Handbook of Statistical Analysis and Data Mining Applications**. p. 420-453, Elsevier Inc, 2009.

SFERRA, Heloisa Helena; CORRÊA, Ângela M. C. Jorge. Conceitos e Aplicações de Data Mining, **Revista de Ciência e Tecnologia**, Piracicaba, v.11, n° 22, p. 19-34, 2003. Disponível em: <<http://www.unimep.br/phpg/editora/revistaspdf/rct22art02.pdf>>

WRIGHT, James Terence Coulter; GIOVINAZZO, Renata Alves. **Delphi – uma ferramenta de apoio ao planejamento prospectivo**. p. 2- 12. São Paulo, 2000. Caderno de pesquisas em administração, São Paulo, 2000.