

Referência completa para citação:

GONÇALVES (L. P. F.) e FREITAS (H.). **Ferramentas de mineração de dados: algo confiável?** Porto Alegre/RS: Anais do XXXVII CLADEA, Outubro de 2002, anais em CD-ROM.

FERRAMENTAS DE MINERAÇÃO DE DADOS: RESULTADO EFETIVO?

Lóren Pinto Ferreira Gonçalves – Professora da URCAMP, loren@aquiles.alternet.com.br
Henrique Freitas – Professor da EA/PPGA/UFRGS, pesquisador CNPq, hfreitas@ea.ufrgs.br

Palavras-chave: ferramentas, mineração de dados, cesta de supermercados.

Resumo

A mineração de dados é uma forma relativamente nova de exploração dos recursos de tecnologia de informação que vem disponibilizando diversas ferramentas com grandes promessas e exigência de altos investimentos, mas que, por outro lado, ainda não é amplamente utilizada pelos tomadores de decisão das organizações, nem detém toda sua confiança. A avaliação de algumas destas ferramentas pode auxiliar os decisores quanto à veracidade daquilo que é prometido sem ter de investir antes de estar seguro do cumprimento de tais promessas: pode-se de fato confiar nos resultados emitidos pela ‘caixa preta’ de uma tal ferramenta? O foco da pesquisa é avaliar a efetividade de algumas dessas caixas pretas, tendo como campo a análise de cesta de supermercado (market basket analysis) a partir de bases de dados reais de uma rede de supermercados na região metropolitana de Porto Alegre/RS. Analisando-se os resultados obtidos, e através da entrevista com os usuários potenciais, concluiu-se que as ferramentas em questão não auxiliaram os tomadores de decisão da empresa, conforme o prometido pela tecnologia. Os sistemas apresentaram vários problemas, tais como a falta de confiabilidade dos resultados gerados e a impossibilidade de trabalhar com grandes bases de dados. Ao mesmo tempo, verificou-se que todas as ferramentas utilizadas eram bastante amigáveis, mas pecavam principalmente em aspectos referentes à qualidade e utilidade.

1. Introdução

A mineração de dados (MD) é uma forma relativamente nova de exploração dos recursos de tecnologia de informação que vem disponibilizando diversas ferramentas com grandes promessas e exigência de altos investimentos. A MD é definida como uma etapa na descoberta do conhecimento em bancos de dados, que consiste no processo de analisar grandes volumes de dados sob diferentes perspectivas, a fim de descobrir informações úteis que normalmente não estão claramente visíveis. Assim, as ferramentas de MD, por definição, devem trabalhar com grandes bases de dados e retornar, como resultado, conhecimento novo e relevante (Niederman, 1997; Lesca, Freitas e Cunha, 1996). Porém, deve-se ser cético quanto a esta afirmação, pois este tipo de ferramenta irá criar inúmeras relações e equações, o que pode tornar impossível o processamento destes dados (Santos e Becker, 1990). De toda forma, tem sido anunciado como um meio importante de geração de informação para a tomada de decisão rumo a uma maior competitividade das organizações (Freitas e Lesca, 1992).

O mercado de mineração de dados tem crescido consideravelmente nos últimos anos (Krivda, *apud* Niederman, 1997), porém Fayad (*apud* Niederman, 1997) chama a atenção para o fato de que existem poucas ferramentas de MD bem desenvolvidas; ele salienta que a maioria delas não foi testada em uma variedade de ambientes, que a maioria não é robusta quanto à falta de dados e ao surgimento de erros, e que não está claro o quanto elas podem ser utilizadas por outras pessoas que não sejam os seus desenvolvedores, daí a importância da a-

valiação de ferramentas disponíveis no mercado utilizando-se bases de dados reais (Freitas e Ballaz, 1992).

A MD ainda não é amplamente utilizada pelos tomadores de decisão das organizações, nem detém toda sua confiança. A avaliação de algumas destas ferramentas pode auxiliar os decisores quanto à veracidade daquilo que é prometido sem ter de investir antes de estar seguro do cumprimento de tais promessas: pode-se de fato confiar nos resultados emitidos pela 'caixa preta' de uma tal ferramenta? O foco da pesquisa é avaliar a efetividade de algumas dessas caixas pretas, tendo como campo a análise de cesta de supermercado (market basket analysis) a partir de bases de dados reais de uma rede de supermercados na região metropolitana de Porto Alegre/RS. Devido ao número de tarefas e de ferramentas existentes, foram avaliadas neste trabalho, somente as ferramentas que utilizam a tarefa associação ou análise de cesta de supermercado.

A seguir, é abordada a descoberta do conhecimento em base de dados e mineração de dados (seção 2), e definido o método de pesquisa (seção 3), quando são analisados os principais resultados (seção 4) e apresentadas algumas reflexões finais (seção 5).

2. A descoberta do conhecimento em bases de dados (DCBD) e a MD

A pesquisa em descoberta do conhecimento em bases de dados tem crescido e atraído esforços, baseada na disseminação da tecnologia de bancos de dados e na premissa de que as grandes coleções de dados hoje existentes podem ser fontes de conhecimento útil, o qual está implicitamente representado e pode ser extraído (Feldens e Castilho, 1997).

O alto desenvolvimento das tecnologias de bancos de dados fez com que, hoje, a capacidade de coletar e armazenar informações supere os recursos (sobretudo nosso tempo, preparação e disposição) para efetivamente utilizar a informação armazenada. As tecnologias de coleta e armazenamento de informações têm evoluído muito nos últimos anos, porém o processamento desta informação tornou-se cada vez mais difícil de ser realizado, podendo-se dizer que as capacidades de coleta e armazenamento de dados atualmente superam os recursos para se utilizar efetivamente a informação.

A descoberta de conhecimento em bases de dados é o processo de extração de conhecimento novo, útil e interessante a partir de bases de dados. Esta é uma das áreas que investem no desenvolvimento de tecnologias mais eficientes para a recuperação de informações, ao projetar, implementar e validar sistemas para a extração de conhecimento útil a partir de bases de dados (Fayyad, *apud* Feldens e Castilho, 1997). As etapas do processo DCBD são: seleção dos dados, pré-processamento dos dados, transformação dos dados, aplicação de técnicas de *Data Mining*, interpretação e avaliação dos padrões, gerando, assim, conhecimento (Fayyad, *apud* Pilla *et al.*, 1998)

Devido à importância da etapa de mineração, o termo 'mineração de dados' tem sido utilizado para identificar todo o processo, como um sinônimo para o processo DCBD: a seguir, alguns detalhes sobre as promessas e o mercado da mineração de dados.

2.1. As promessas da mineração de dados

As promessas encontradas através da revisão de literatura são mostradas no quadro abaixo, no intuito de se poder, na seqüência do estudo, investigar se elas são ou não uma realidade disponível a partir das tecnologias e produtos disponíveis no mercado.

Quadro 1 – Promessas da mineração de dados encontradas na revisão da literatura

Promessa	Descrição	Fonte
P1	Analisar grandes volumes de dados sob diferentes perspectivas, a fim de descobrir informações úteis que normalmente não são imediatamente visíveis	Brusso, 1998 Berry e Linoff, 1997
P2	Trabalhar com grandes bases de dados	Niederman, 1997
P3	Retornar conhecimento novo e relevante	Niederman, 1997
P4	Gerar hipóteses	Figueira, 1998
P5	Encontrar padrões que não são encontrados por sistemas ditos tradicionais	Figueira, 1998 Moxon, 1996
P6	Os sistemas são capazes de aprender e apoiar a realização de descobertas a partir dos dados	Feldens <i>et al.</i> , 1997

2.2. O mercado de MD

A quantidade de dados armazenados pelas empresas está explodindo, e ultrapassa a habilidade técnica e a capacidade humana na sua interpretação (Figueira, 1998). Por isso, diversas ferramentas têm sido usadas para examinar tais dados. No entanto, a maioria dos analistas tem reconhecido que existem padrões escondidos nestes dados, os quais não podem ser encontrados utilizando-se dos métodos tradicionais.

A necessidade de transformar a ‘montanha’ de dados armazenados em informações significativas é óbvia. Entretanto, a sua análise ainda é demorada, dispendiosa, pouco automatizada e sujeita a erros, mal entendidos e falta de precisão (Newing, 1996). A automatização dos processos de análise de dados, com a utilização de softwares ligados diretamente à massa de informações, tornou-se uma necessidade (Figueira, 1998). Este motivo deve ser o responsável pelo crescimento do mercado de tecnologias de informação voltadas para a MD.

Conforme Fayad (*apud* Niederman, 1997), existem muitas ferramentas de MD disponíveis no mercado, porém, ainda não se sabe o quanto estas ferramentas podem ser utilizadas por outras pessoas que não sejam seus desenvolvedores.

3. Método de pesquisa

O método se caracteriza como um estudo de caso múltiplo: o caso de quatro ferramentas de mineração de dados, as quais são aplicadas às bases de dados de uma rede de supermercados.

Para esta pesquisa, foram usados dados primários - oriundos das entrevistas com os tomadores de decisão, e secundários – aqueles contidos nas bases de dados da empresa. Foi realizado um estudo aprofundado da literatura, ao mesmo tempo em que eram contatadas empresas com grandes massas diárias de dados (bancos, supermercados, etc), visando a seleção de um campo específico de aplicação. Os dados finalmente coletados foram preparados para análise e para funcionamento nos diferentes sistemas a que se teve acesso. Interações foram realizadas com os fornecedores dos sistemas, visando elucidar as dúvidas que surgiam, como mesmo melhorar os sistemas, quando estes mostravam disfunções.

3.1. O contexto de aplicação

O contexto de aplicação desta pesquisa, o setor de supermercados, foi escolhido devido à riqueza de seus dados e à forte competição existente. A empresa (Comercial Unida de Cereais – Rede Unidão de Supermercados) foi selecionada por conveniência, em função de: (1) disponibilidade de suas bases de dados e (2) porte da empresa, considerando o ranking da AGAS (Associação Gaúcha de Supermercados), onde a empresa encontrava-se na quinta posição (Agas, 1999). As lojas que fizeram parte desta pesquisa foram as seguintes filiais: Feitoria, Canela, Sapiranga e Viamópolis. As filiais foram escolhidas pela própria empresa, que as

considerou como representativas de núcleos diferenciados de mercado e com potencial de crescimento.

3.2. Dados utilizados

Os instrumentos de coleta de dados utilizados foram: a) literatura, para o conhecimento das promessas da tecnologia; b) arquivos com os dados da movimentação das lojas, durante dois meses; c) observação participante, durante a aplicação das ferramentas; e d) entrevistas com os tomadores de decisão da empresa.

Os dados começaram a ser armazenado para a realização deste trabalho, por isso não houve muito esforço para prepará-los para a mineração (pré-processamento). Foram utilizadas três bases de dados de tamanhos diferentes para a aplicação das ferramentas, pois, assim, seria possível verificar se o volume de dados influenciava na performance dos sistemas. O quadro 2 mostra a descrição das bases de dados.

Quadro 2 - Descrição das três bases de dados utilizadas

Base	Nº de registros	Nº de Entidades	Média de registros por entidade	Composição da base
B1	10.000	1.801	5,55	10.000 primeiros registros do movimento de maio 2000 na loja de Viamão
B2	287.935	38.842	7,41	Movimento completo da loja de Viamão no mês de maio 2000
B3	3.326.859	148.115	22,46	Movimento das quatro lojas nos meses de maio e junho 2000

A estrutura das bases de dados é mostrada no quadro 3.

Quadro 3 - estrutura das bases de dados

Nome do Campo	Descrição	Tipo	Tamanho	Casas Decimais
Código	Código EAN do Produto	String	13	-
Descrição	Descrição do Produto	String	30	-
Qtd	Quantidade Vendida	Numérico	10	3
Valor	Preço do Produto	Numérico	11	2
PDV	Ponto de Venda	String	3	-
Cupom	Número do Cupom Fiscal	Numérico	6	-
Data	Data da Venda	String	10	-
Diasemana	Dia da Semana	Numérico	1	-
Loja	Identificação da Loja do Unidão	String	3	-

As bases de dados encontravam-se no formato vertical, usando múltiplas linhas para armazenar cada entidade (venda). As linhas para uma entidade particular são ligadas através do campo chamado 'cupom'.

3.3 Obtenção e aplicação das ferramentas de mineração de dados

As ferramentas obtidas foram seis: Aira Data Mining®, CBA®, PolyAnalyst®, SuperQuery Discovery Edition®, SuperQuery Office Edition® e Xaffinity®. Todas foram escolhidas por conveniência, em função de: (1) realização da análise de cesta de supermercado, (2)

rodar em plataforma Windows® e, (3) sua disponibilização por parte de seus proprietários, sem ônus para a equipe de pesquisa. As ferramentas estão caracterizadas no quadro 4.

Somente quatro (**Aira Data Mining, Super Query Discovery Edition, Super Query Office Edition e Xaffinity**) das seis ferramentas obtidas puderam ser analisadas, pois duas (**Poly Analyst e CBA**) eram incompatíveis com os tipos de dados disponíveis (trabalhavam com dados no formato horizontal).

Todas as ferramentas foram executadas num computador com a seguinte configuração: Processador K6 550 MHz, 128 MB de memória RAM, 10 GB de disco rígido e Monitor SVGA. Cabe salientar que esta configuração está além de todos os requisitos mínimos exigidos pelos sistemas avaliados.

Quadro 4 - Caracterização das ferramentas utilizadas

Ferramentas Item	Aira Data Mining®	Xaffinity®	Poly Analyst®	Super Query Discovery Edition®	Super Query Office Edition®	CBA®
Desenvolvedor	Hycones IT	Exclusive Ore	Mega Puter	Azmy	Azmy	National University of Singapore
Ano	1996	2000	1994	*	*	*
Idioma	Português e Inglês	Inglês	Inglês	Inglês	Inglês	Inglês
Sistema Operacional	Windows	Windows	Windows	Windows	Windows	Windows
Configuração Mínima	Pentium, 32MB Ram, 20 MB HD	Pentium 266, 64 MB Ram	Pentium 266, 32 MB Ram, 20 MB HD	Pentium 266, 32 MB Ram, 20 MB HD	Pentium 266, 32 MB Ram, 20 MB HD	Pentium, 32 MB Ram, 5 MB HD
Preço (US\$)	5.000	5.000	*	449,95	149,95	Livre (a ferramenta CBA é um software freeware)
Bases de Dados	Oracle, Sybase, Informix, SQL Server, Paradox e Dbase	SQL Server, MS Access, Oracle, Red Brick e White Cross	MS Excel, ODBC, Oracle, IBM Visual Warehouse, arquivos .csv	SQL Server, MS Access, Oracle e dBase	MS Excel e MS Access	Texto
Regras Geradas	Associativas e hierárquicas	Associativas e sequenciais	Associativas	Associativas	Associativas	Associativas e hierárquicas
Formato dos Resultados	Regras em MS-Word	Regras em linguagem	Lista de regras,	Lista de regras e	Lista de regras e gráfico	Lista de regras

	e HTML	natural e em colunas	histograma, gráficos 3D e 2D	gráfico com os valores mais frequentes	com os valores mais frequentes	
--	--------	----------------------	------------------------------	--	--------------------------------	--

* Os valores não foram informados.

4. Análise dos principais resultados

A análise da pesquisa foi realizada através de: a) observação do pesquisador, como usuário das ferramentas, ou seja, a pessoa que utilizou as ferramentas e realizou as minerações; e b) avaliação dos resultados obtidos pelas ferramentas (informações geradas), efetuando-se uma entrevista com os tomadores de decisão da empresa.

Para a avaliação, utilizou-se um instrumento de avaliação de Sistemas de Informação desenvolvido por Freitas (Freitas, 1993; Freitas, Ballaz e Trahand, 1993), usado também por Stumpf (1998). Este modelo baseia-se em dois pilares básicos: a utilidade (ligada ao usuário) e a facilidade (ligada ao sistema). Cada um desses pilares do instrumento de coleta de dados gerou um quadro onde foram colocadas as ferramentas, as variáveis e as suas respectivas avaliações. Tais quadros são mostrados no quadro a seguir.

Quadro 5 - Avaliação da Utilidade da ferramenta

Parâmetro	Ferramentas			
	1	2	3	4
U1. Rapidez	Insatisfatório	Insatisfatório	Insatisfatório	Insatisfatório
U2. Desempenho e produtividade	Insatisfatório	Insatisfatório	Insatisfatório	Insatisfatório
U3. Eficácia	Nenhuma	Nenhuma	Nenhuma	Nenhuma
U4. Facilidade no trabalho	Nenhuma	Nenhuma	Nenhuma	Nenhuma
U5. Utilidade	Pouca	Pouca	Nenhuma	Nenhuma
U6. Tempo de resposta	Médio	Médio	Pouco	Pouco
U7. Aprendizagem	Nenhuma	Nenhuma	Nenhuma	Nenhuma
U8. Autonomia e independência	Nenhuma	Nenhuma	Média	Média

As quatro ferramentas não foram consideradas úteis ao desenvolvimento do trabalho dos diretores da empresa. O que justifica esta posição é o fato de as ferramentas não funcionarem da forma esperada e porque seus resultados não foram retornados num formato adequado.

Quadro 6 – Avaliação da facilidade

Parâmetro	Ferramentas			
	1	2	3	4
F1. Aprendizado de operação	Muito Fácil	Muito Fácil	Muito Fácil	Muito Fácil
F2. Domínio	Muito Fácil	Muito Fácil	Muito Fácil	Muito Fácil
F3. Interação	Média	Média	Alta	Alta
F4. Flexibilidade	Média	Média	Média	Média
F5. Habilidade	Muito Fácil	Muito Fácil	Muito Fácil	Muito Fácil
F6. Facilidade no uso	Muito Fácil	Média	Muito Fácil	Muito Fácil
F7. Funcionalidade	Pouca	Pouca	Nenhuma	Nenhuma
F8. Impacto da apresentação gráfica	Nenhum	Nenhum	Nenhum	Nenhum
F9. Qualidade da apresentação gráfica	Baixa	Baixa	Baixa	Baixa

Quanto à facilidade (quadro 6), percebeu-se que os sistemas não apresentaram problemas, ou seja, os quatro são fáceis de operar, utilizam padrão semelhante ao do Windows®,

o que facilita muito a sua utilização por pessoas que não dominam tais sistemas. As ferramentas utilizadas não apresentam informações através de recursos gráficos: esta foi uma das reclamações dos usuários, pois, no ponto de vista deles, se estes recursos fossem utilizados, talvez a interpretação das informações se tornasse mais fácil. Ainda, um problema apontado pelos usuários foi a funcionalidade dos sistemas, a qual pode ser verificada através do cumprimento das promessas da MD por parte das ferramentas.

Quadro 7 – Funcionalidade dos sistemas

Promessas	Ferramentas			
	1	2	3	4
Análise de grandes volumes de dados sob diferentes perspectivas, a fim de descobrir informações úteis que normalmente não estão sendo visíveis	Não	Não	Não	Não
Trabalha com grandes bases de dados	Não	Não	Não	Não
Retorna conhecimento novo e relevante	Pouco	Pouco	Não	Não
A mineração de dados é responsável pela geração de hipóteses	Pouco	Pouco	Não	Não
Encontra padrões que não são encontrados por sistemas ditos	Sim	Sim	Não	Não
É capaz de aprender e apoiar a realização de descobertas a partir dos dados	Pouco	Pouco	Não	Não

Percebe-se, através das respostas, que as ferramentas não cumprem as promessas da MD (quadro 7). Para os usuários, o maior problema encontrado refere-se ao tratamento de grandes bases de dados, já que esta é a grande promessa da tecnologia. Porém, trabalhou-se somente com dois meses de quatro lojas, o que para a empresa é uma base de dados bem pequena; mesmo assim, as quatro ferramentas ‘travaram’ devido ao volume de dados. No entendimento dos tomadores de decisão da empresa, dever-se-ia trabalhar com, no mínimo, dois anos para realmente verificarem-se padrões de comportamento dos consumidores sem que se caísse em questões sazonais.

Quadro 8 – Avaliação da qualidade

Parâmetro	Ferramentas			
	1	2	3	4
Q1. Precisão e confiabilidade	Média	Média	Nenhuma	Nenhuma
Q2. Oportunidade	Média	Média	Nenhuma	Nenhuma
Q3. Dificuldade para obter informação	Muita	Muita	Muita	Muita
Q4. Facilidade de interpretação	Pouca	Pouca	Pouca	Pouca
Q5. Fonte de Informação	Média	Média	Nenhuma	Nenhuma
Q6. Corretude	Nenhuma	Nenhuma	Nenhuma	Nenhuma
Q7. Medidas subjetivas de interesse	Pouca	Pouca	Pouca	Pouca
Q8. Medidas objetivas de interesse	Sim	Sim	Sim	Sim

Com relação ao aspecto precisão e confiabilidade (quadro 8), os usuários ficaram preocupados, pois não se conseguiu obter resultados significativos sem a intervenção dos desenvolvedores. Para os tomadores de decisão, seria impossível confiar em informações que foram geradas através da intervenção do desenvolvedor da ferramenta, no sentido de programar novas rotinas conforme os problemas iam surgindo. No caso de duas ferramentas (Super Query Discovery Edition e Super Query Office Edition), esta intervenção não foi necessária, porém, alguns valores apresentados não correspondiam à realidade da base de dados.

Quadro 9 - Avaliação do impacto e dos benefícios

Parâmetro	Ferramentas			
	1	2	3	4
IB1. Contribuição	Pouca	Pouca	Nenhuma	Nenhuma
IB2. Impacto	Nenhum	Nenhum	Nenhum	Nenhum
IB3. Entendimento do negócio	Pouco	Pouco	Nenhum	Nenhum
IB4. Discussões dentro da organização	Nenhuma	Nenhuma	Nenhuma	Nenhuma
IB5. Competitivade	Pouca	Pouca	Nenhuma	Nenhuma

A partir da análise realizada (quadro 9), concluiu-se que os sistemas não causaram impacto sobre a tomada de decisão da empresa, ou seja, não contribuíram para melhorar o processo decisório dentro da organização.

5. Conclusões

Atualmente, a necessidade de informações relevantes e desconhecidas parece importante para obter vantagem competitiva: pelo menos, dispor de todos esses dados, e de toda essa informação e conhecimento ‘adormecidos’, contidos nas bases que repousam no seio da empresa, parece um luxo um tanto grande. Por isso, muitas novas tecnologias de informação têm surgido nos últimos anos, algumas com grandes promessas. A mineração de dados é uma dessas tecnologias. Devido aos altos investimentos necessários à implantação destas tecnologias, é importante que se realize uma avaliação prévia do cumprimento das suas promessas. Em especial, deve-se ter certeza de como funcionam os ditos algoritmos ou mesmo as ‘caixas-pretas’ que geram informações e conhecimento. Senão, como confiar nos seus resultados? (não daria para se imaginar um sistema de pagamento de pessoal onde não domina-se as fórmulas internas, certo?).

Neste estudo, foi possível perceber quais eram as promessas da tecnologia de mineração de dados e o que algumas das ferramentas existentes no mercado realmente oferecem. Analisando-se os resultados obtidos, e através da entrevista com os usuários potenciais, concluiu-se que as ferramentas em questão não auxiliaram os tomadores de decisão da empresa, conforme o prometido pela tecnologia. Os sistemas apresentaram vários problemas desde aqueles mais simples, como incompatibilidade de formato de datas (americano com o brasileiro), o que trancava a licença de utilização do sistema a cada vez que se tentava rodá-lo novamente, até problemas relacionados ao retorno de informações com valores errados.

Os maiores problemas relatados pelos usuários foram: a) a falta de confiabilidade dos resultados gerados - devido ao fato dos desenvolvedores precisarem intervir para a sua geração e à existência de valores incorretos que foram retornados por duas das ferramentas utilizadas; e b) a impossibilidade de trabalhar com grandes bases de dados, o que torna difícil obtermos padrões de comportamento do consumidor que realmente reflitam o que acontece no dia a dia dos negócios. Ao mesmo tempo, foi verificado que as ferramentas não apresentaram problemas com relação à facilidade de utilização. Todas as ferramentas utilizadas eram bastante amigáveis, mas pecavam principalmente em aspectos referentes à qualidade e utilidade. Em relação a limitações desta pesquisa, destacam-se os seguintes: a) ter trabalhado somente com ferramentas de mineração de dados que realizassem a tarefa associação; b) a base de dados de uma rede de supermercados, com suas peculiaridades; e c) as ferramentas utilizadas, que foram obtidas por conveniência.

Pode-se concluir que as ferramentas de mineração de dados utilizadas nesta pesquisa ainda não estão prontas para a utilização plena no ambiente empresarial. Muitos testes e melhorias deveriam ser realizados por parte dos desenvolvedores destas ferramentas, no intuito de melhorá-las no que se refere à utilidade para a gestão. Fayad (*apud* Niederman, 1997) já levantava dúvidas quanto à utilização de ferramentas de mineração de dados por pessoas que

não fossem os seus desenvolvedores: o que se pode verificar neste trabalho é que este problema persiste e isto já é suficiente para impossibilitar estas ferramentas de se encontrarem disponíveis no mercado. Ao mesmo tempo, isto é um alerta aos gestores, de forma que, antes de adotarem soluções dessa natureza, façam uma clara e profunda avaliação da parte técnica envolvida, de forma a dar maior segurança aos usuários, em especial quanto aos resultados gerados.

Referências bibliográficas

- AGAS (Associação Gaúcha de Supermercados) [25 de novembro de 1999] Disponível na World Wide Web <<http://www.agas.com.br>>
- BERRY, Michael J. A., LINOFF, Gordon. **Data mining techniques: for marketing, sales and customer support**. USA: Wiley Computer Publishing, 1997.
- BRUSSO, Marcos José. **O paralelismo na mineração de regras de associação**. Porto Alegre: UFRGS, 1998 (Trabalho Individual I, Programa de Pós-Graduação em Computação).
- FELDENS, Miguel Artur, MORAES, Rodrigo Leal, PAVAN, Altino, CASTILHO, José Mauro Volkmer. **Mineração de dados na gestão hospitalar**. Porto Alegre: UFRGS, 1997. [20 de dezembro de 1998] disponível na WWW <<http://www.inf.ufrgs.br/~feldens/datamining.html>>.
- FELDENS, Miguel Artur, MORAES, Rodrigo Leal, PAVAN, Altino, CASTILHO, José Mauro Volkmer. **Engenharia da descoberta de conhecimento em bases de dados: estudo e aplicação na área de saúde**. Porto Alegre: UFRGS, 1997. [20 de dezembro de 1998] disponível na WWW <<http://www.inf.ufrgs.br/~feldens/dm2.html>>.
- FIGUEIRA, Rafael. **Mineração de dados e bancos de dados orientados a objetos**. Rio de Janeiro: UFRJ, Dissertação, Mestrado, Ciência da Computação, 1998.
- FREITAS, Henrique. **A informação como ferramenta gerencial: um telessistema de informação em marketing para o apoio à decisão**. Porto Alegre: Ortiz, 1993.
- FREITAS (H.) e LESCA (H.). Competitividade empresarial na era da informação. São Paulo/SP: **Revista de Administração da USP**, v. 27, n. 3, Julho/Setembro 1992. p.92-102.
- FREITAS (H.) e BALLAZ (B.). Um método implícito para avaliação de um SAD. Santiago-Chile: **Revista ACADEMIA, CLADEA**, n. 10, 1992 (2º semestre), p. 63-91.
- FREITAS (H.), BALLAZ (B.) e TRAHAND (J.). Sistema de informações em marketing e apoio à decisão. São Paulo - SP: **Revista de Administração da USP**, v. 28, n. 2, Abril-Junho 1993, p. 33-49.
- GONÇALVES, Lóren Pinto Ferreira. **Avaliação de ferramentas de mineração de dados como fonte de dados relevantes para a tomada de decisão: aplicação na Rede União de Supermercados – São Leopoldo-RS**. Porto Alegre: UFRGS, Dissertação de Mestrado, PPGA/EA, 2001.
- LESCA (H.), FREITAS (H.) e CUNHA Jr. (M. V. M.). Instrumentalizando a Decisão Gerencial. Rio de Janeiro - RJ: **Revista Decidir**, ano III, n. 25 Agosto 1996, p. 6 a 14.
- LESCA (H.), FREITAS (H.) e CUNHA Jr. (M. V. M.). Como dar um senso útil às informações dispersas para facilitar as decisões e ações dos dirigentes: o problema crucial da inteligência competitiva através da construção de um “PUZZLE” (“quebra cabeça”)®. Porto Alegre - RS: **Revista READ** (<http://read.adm.ufrgs.br>), v. 3, n. 1, Outubro de 1996.
- MOXON, Bruce. **Defining Data Mining**. DBMS, Data Warehouse Supplement, August, 1996.
- NEWING, Rod. **Mineração de dados**. Management Accounting. p. 34-35. oct. 1996.
- NIEDERMAN, Fred. **Data mining: a research framework**. Baltimore: Information Systems Research Center, 1997.

PILLA, A. D., CAPRETZ. **O processo KDD**. 1998. [08 de março de 2000] Disponível na WWW <<http://www.igce.unesp.br/igce/grad/computacao/~cintiab/datamine>>

SANTOS, Eduardo Ribas, BECKER, João Luiz. **BIACS – base de dados inteligente para aquisição de conhecimentos de sistemas**. Série Documentos para Estudo: PPGA, nº 13, 1990.

STUMPF, Evandro Carlos. **Concepção e desenvolvimento de um painel de controladoria em uma organização do setor de autopeças utilizando a tecnologia da informação**. Porto Alegre: UFRGS, Dissertação de Mestrado, PPGA/EA, 1998.