

DATA MINING, SEUS BENEFÍCIOS, UTILIZAÇÕES, METODOLOGIA, CAMPO DE ATUAÇÃO DENTRO DE GRANDES E PEQUENAS EMPRESAS.

Flávio Henrique da Silva Veloso¹

Iury Soprani Toledo Moreira²

Leonardo Lopes de Andrade Silva³

Regina Célia M. F. Silva³

Resumo: Neste Artigo é proposto o estudo das técnicas de Data Mining para auxiliar na tomada de decisões. A metodologia aqui apresentada permite com que, com a utilização de Data Mining (Mineração de dados) investiguem-se dados a procura de padrões que tenham valor para as empresas.

Palavras - chave: decisões; Data Mining; padrões; valor; empresas.

Abstract: In this Article is considered the study of the techniques of Mining Date (Mining of data) to assist in the taking of decisions. The methodology presented here allows that, with the use of Mining Date, it is investigated given the search of standards that have value for the companies.

Key - Words: decisions; Dates Mining; standards; value; companies.

INTRODUÇÃO

Na atualidade, com o advento da tecnologia de armazenamento digital, quase tudo que produzimos de informações passou a ser colocado diretamente no mundo de bits e bytes e o que existe no meio físico como em livros, revistas e jornais, para citar alguns exemplos, estão sendo transmutados para os discos rígidos e memórias digitais.

Com isso, na grande parte das empresas a manipulação de dados é muito importante para que não se perca informações causando assim imprevistos, falhas e custos desnecessários que podem ocorrer devido à má qualidade na hora de manipular e administrar as informações. É neste contexto que entra a mineração de dados (*Data Mining*), utilizado para trabalhar em grandes bancos de dados de forma a procurar padrões e encontrar correlações e tendências entre as informações, fazendo assim com que as empresas se antecipem aos concorrentes e que superem as expectativas de seus clientes.

Desta forma, este artigo tem como objetivo destacar e mostrar os desafios e os artifícios de empresas que utilizam Data Mining tende a enfrentar e superar diante de um mercado tão disputado e tão competitivo.

Fundamentação Teórica

Esta seção traz a base para se entender o tema do trabalho, apoiando-se em teorias relacionadas ao Data Mining e suas utilidades. Nas próximas subseções serão explorados cada um desses temas.

Data Mining: Metodologia

A metodologia é identificada na literatura em nove passos (Fayyad, Shapiro & Smith, 1996; existem variações em Adriaans & Zantinge, 1996; Bracman & Anand, 1996; Elmasri & Navathe, 2000, entre outros), como segue:

- 1.Desenvolver a compreensão do mínimo da aplicação, o conhecimento anterior relevante e os objetivos do usuário final;
- 2.Criar um conjunto de dados-alvo no qual a prospeção deverá ser efetuada;
- 3.Realizar a limpeza e o pré-processamento dos dados;
- 4.Realizar a redução e projeção de dados, reduzindo o número efetivo de variáveis consideradas, ou encontrar representações não variáveis para os dados;
- 5.Escolher as tarefas de mineração de dados, decidindo sobre se o objetivo do processo KDD é classificação, regressão, “clusterização” ou outro;
- 6.Escolher os algoritmos de mineração de dados, selecionando métodos para ser usados na busca de padrões nos dados;
- 7.Mineração de dados;
- 8.Interpretação dos padrões obtidos;
- 9.Consolidação do conhecimento obtido.

1 Data Mining: Campo de atuação e motivo de sua utilização.

O Data Mining tem um vasto campo de atuação no mercado.

Ele pode atuar no Governo, podemos tomar como exemplo os EUA, que utilizam-no há muito tempo, na identificação de padrões de transferências de fundos internacionais que se parecem com lavagem de dinheiro do narcotráfico, além disso eles também o utilizam para traçar perfis de pessoas, analisando suas atividades e deduzindo padrões de informações.

A Mineração de Dados também é aplicada em

bancos, como o Bank of America que se utiliza desta técnica para selecionar dentre seus clientes aqueles com menos riscos de dar calote em um empréstimo. A partir destes relatórios enviam cartas oferecendo linhas de crédito, com isso, em três anos o banco lucrou 30 milhões de dólares.

O Wall Mart também desfruta do uso do Data Mining, pois há cinco anos ao procurar relações entre o volume de vendas e os dias da semana um software de Data Mining mostrou que as sextas-feiras a venda de cerveja, aumentava na mesma proporção que a venda de fraldas, detalhando ainda mais notou-se que ao comprar fraldas para seus bebês, eles aproveitam para abastecer o estoque de cerveja para o final de semana.

No jornalismo o Data Mining pode ser utilizado para digitalizar publicações impressas, fazendo assim com que o leitor possa ver edições mais antigas com mais facilidade, isso foi usado no *New York Times*.

Como podemos observar nos exemplos anteriores a Mineração de Dados é bastante utilizada, os motivos de sua utilização são muitos, porém o principal motivo é porque o volume de informações aumentaram muito, então o Data Mining se tornou imprescindível para uma empresa que não que perder tempo e dinheiro com informações perdidas no sistema.

1.1 Data Mining utilizado nas empresas:

Tanto em grandes e pequenas empresas, a mineração de dados é amplamente aplicada, pois dados como estoques, vendas, pedidos e etc., são acumulados dia-a-dia, gerando um grande número de informações.

FIGURA 1
Esquema do Data Mining



O Data Mining atua coletando estes dados e gerando constatações onde é possível prever variações de mercado, lucros e prejuízos, e se houverem prejuízos ele aponta a melhor e mais rápida alternativa a se tomar.

A mineração de dados faz parte de um processo de manipulação de dados.

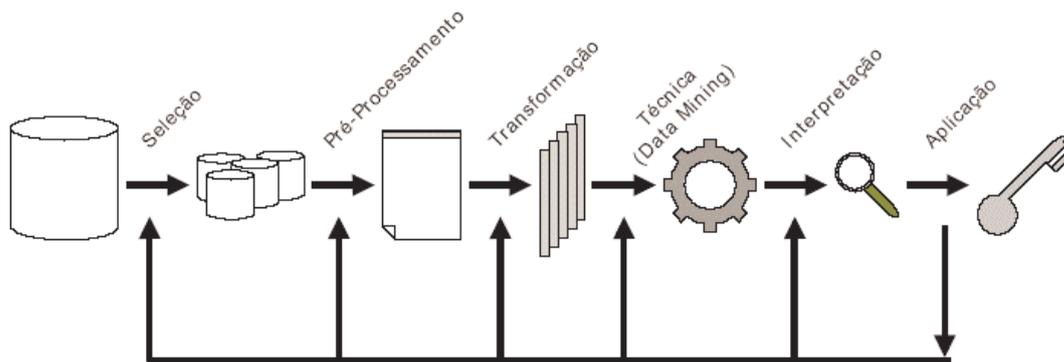


Figura 2 - Etapas do KDD [FAYYAD,96]

1.2 Benefícios do uso do DM(Data Mining)

O DM tem várias utilidades dentro de uma empresa, seus benefícios variam de ajudar tomar decisões corretas e mais rápidas, apontar falhas nos sistemas das empresas e também, dependendo do campo de atuação, traz mais lucros da empresa, melhorando o Marketing.

Como nos exemplos utilizados no 1.1 Campo de atuação e motivo de sua utilização. No Wall Mart.

Segundo pesquisas realizadas, empresas que utilizam o DM, tem menos chance de sofrerem abalos econômicos e mais chances de conseguirem lucros maiores.

1.3 Implantação do DM dentro das empresas:

O DM não pode ser apenas jogado dentro das empresas sem qualquer preparação ou treinamento, pois além do software que manipula os dados o ser humano também tem grande importância, pois o programa apenas apresenta as informações coletadas cabe ao funcionário cruzá-las de forma a apresentar os dados almejados.

Portanto antes da implantação do DM em uma empresa é necessário o treinamento dos funcionários, para que os mesmos se acostumem e aprendam a utilizar esse software com eficiência.

1.4 Possibilidades

O data warehouse é a tecnologia predominante no auxílio à obtenção dos dados necessários ao processo KDD, enquanto a mineração de dados é a tecnologia predominante no auxílio à interpretação desses dados. Isso não significa que essas tecnologias, materializadas nos produtos que oferecem essa funcionalidade, sejam a única maneira de realizar-se um processo KDD. As fontes de dados que alimentarão um processo KDD podem ser:

- arquivos de diversos formatos e procedências, até mesmo provenientes da Internet;
- sistemas de arquivos;
- bancos de dados populados em sistemas gerenciadores de banco de dados(SGBDs).

Da mesma forma, existem programas e sistemas, particularmente das áreas da estatística e de inteligência artificial, que podem ser utilizados, e muitas vezes são comercializados, com a funcionalidade atribuída aos sistemas de mineração de dados.

A principal desvantagem no uso desses sistemas está no fato de possuírem formatos proprietários em relação às fontes de dados, o que acarreta demora na importação e exportação de dados, bem como os respectivos pré e pós-processamentos para esses ambientes.

Como essa característica é bastante freqüente, torna-se desejável que sistemas de mineração de dados sejam integrados ao SGBD que contenha os dados a ser analisados. Segundo Elmasri & Navathe (2000), esses sistemas ainda não são bem integrados aos SGBDs, contudo Goebel & Gruenwald (1999) avaliam 43 sistemas de mineração de dados de prateleira, dos quais 26 aceitam fontes de dados provenientes de SGBDs relacionais.

A questão da integração de sistemas de mineração de dados aos sistemas gerenciadores aos bancos de dados (SGBDs) e aos sistemas de *data warehouse* (DWs) merece uma reflexão mais detalhada sobre a natureza dos ambientes de consulta disponíveis nessas ferramentas, bem como na adequação do uso de cada uma delas. À medida que tecnologias, tais como novas linguagens de programação, SGBDs e equipamentos foram sendo desenvolvidas, uma série de recursos tornou-se disponível, e um dos grandes problemas do até então chamado processamento de dados pode ser atacado e bem resolvido: o processamento de transações on-line (*On-Line Transaction Processing* ou OLTP).

Sistemas que possuem aplicações em bancos de dados operacionais, que atendem às demandas do dia-a-dia das empresas e contêm transações concorrentes ou OLTP são, na maior parte, SGBDs relacionais e suportam linguagem de consulta SQL (*Structured Query Language*). A SQL é uma linguagem declarativa com grande poder de expressão.

A urgência das corporações e da indústria de software na solução dos problemas OLTP teve como conseqüência um retardamento nas soluções para problemas mais complexos, normalmente atendidos por sistemas de apoio à decisão

(ver apêndice B). As aplicações que solucionam esta categoria de problemas foi revisitada e batizada por Codd de On-Line Analytical Processing, ou OLAP.

Considerações Finais

Em suma, podemos perceber que Data Mining vem sendo aplicado cada vez mais dentro das organizações, pois atua em suas bases de dados, que muitas vezes são imensas e repleta de informações úteis escondidas em um emaranhado de dados, sendo capaz de inter-relacionar informações antigas com as do presente e predeterminar as de futuro.

Desta forma, o interesse por Data Mining, em particular em dependências funcionais, tem aumentado recentemente, principalmente por causa da alta demanda por transformar grandes quantidades de dados em informações úteis. Aonde ele vem demonstrando ser uma ferramenta muito importante aprimorando e objetivando ações futuras.

Assim, podemos esperar das ferramentas do Data Mining agilidade, confiança, prevenção e comparação, e que extrairão informações úteis a organização que será tomada como estudo.

Referências Bibliográficas:

BASTOS, Helder. **Jornalismo Eletrônico: internet reconfiguração de práticas nas redações**. Coimbra: Livraria Minerva Editora, 2000.

BERRY, M. J. A., LINOFF, G. – **Data Mining techniques**. USA: John Wiley, 1997.

BUSSAB, A.; MIAZAKI, E. S.; ANDRADE, D. F. – **Introdução à análise de agrupamentos**. São Paulo: IX SINAPE, 1990.

FAYYAD, Usama; PIATETSKI-SHAPIRO, Gregory; SMYTHI, Padhraic. **The KDD Process for Extracting Useful Knowledge from Volumes of Data**. In: Communications of the ACMI, pp.27-34, Nov.1996.

HAN, J., KAMBER, M. – **Data Mining: concepts and techniques**. USA: Morgan Kaufmann, 2001.

HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. – **The elements of statistical learning**. USA: Springer, 2001.

JOHNSON, R. A., WICHERN, D. W. - **Applied multivariate statistical analysis**, 4 ed. USA: Prentice Hill, 1998.

LIMA JR. Walter Teixeira. **Mídia digital: o vigor das práticas jornalísticas em um novo espaço**. São Paulo, 2003. Tese (Doutorado em Jornalismo) – Escola de Comunicação e Artes da Universidade de São Paulo.

MAYFIELD, Kendra. **Read All About it**. In: Revista Wired, 29 Jul 2002 .
<www.wired.com/news/business/0,1367,54030,00.html>

NAVEGA, Sérgio. **Princípios Essenciais do Data Mining**.
<www.intelliwise.com/snavega>.

NETER, J.; KUTNER, M. H.; NACHTSHEIM, C. J.; WASSERMAM, W. – **Applied Linear Regression Models**. London: Richard D. Irwing, Inc., 3. ed., 1996.

ROZADOS, Helen Beatriz Frota. **O jornal e seu Banco de Dados: uma simbiose obrigatória**. In: DoIS (Documents in Information Science), Issue 1, Volume 26, Ano 1997.

<<http://dois.mimas.ac.uk/DoIS/data/Articles/juljqbfchy:1997:v:26:i:1:p:2805.html>>

TOGNOLLI, Júlio Cláudio. **Investigação na Internet: sonho dirigido ou delírio controlado**. <

<http://observatorio.ultimosegundo.ig.com.br/artigos/eno130220021.htm>>

ZETTER, Kim. **GAO: Fede Data Mining Extensive**. In: Wired Magazine.
<www.wired.com/news/privacy/0,1848,63623,00.htm>