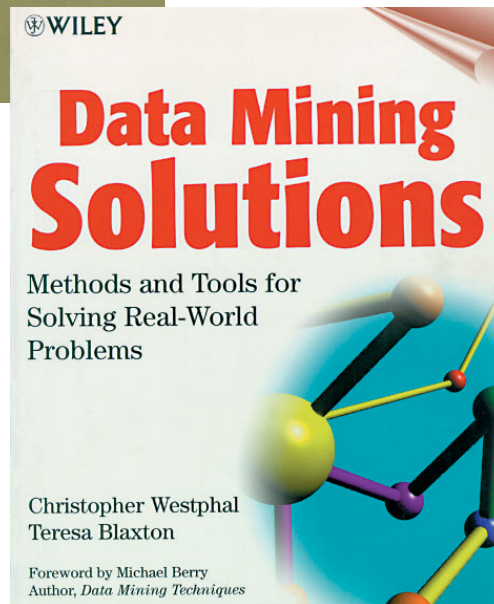


de **Michael J. A. Berry e Gordon Linoff**
New York: Wiley, 1997. 454 p.

de **Christopher Westphal e Teresa Blaxton**
New York: Wiley, 1998. 448 p.



por **Francisco Aranha**, Professor do Departamento
de Informática e Métodos Quantitativos e
Consultor de Empresas.
E-mail: faranha@fgvsp.br

APLICAÇÕES DIVERSIFICADAS

Detecção de *insider trading* nas bolsas de valores. Identificação de operações de lavagem de dinheiro. Seleção de seqüências de DNA para desenvolvimento e teste de produtos farmacêu-

ticos. Identificação do perfil de usuários de bibliotecas. Análise de padrões de compras de varejo em diferentes áreas geográficas. Avaliação de eficácia de campanhas de *marketing*. Seleção de clientes para realização de ofertas cruzadas.

Das mais policiais às mais empresariais, das mais esotéricas às mais mundanas, essas atividades têm em comum a análise estatística de grandes bases de dados para a descoberta de informação antes desconhecida e sobre a qual se pode agir de maneira eficaz. “Grandes bases de dados”, nesse contexto, quer dizer “GRANDES MESMO”. Dezenas ou centenas de milhões de registros. Montanhas de *bytes* que precisamos escavar em busca das pepitas de conhecimento desejadas. Daí o nome da disciplina, *Data Mining*.

EXAGERO E REALIDADE

As técnicas de mineração mais comentadas ultimamente realmente capturam a imaginação e evocam cenas de ficção científica como as do quase humano computador Hall, do filme “2001, Uma Odisséia no Espaço”.

As redes neurais, por exemplo, utilizam estruturas de processamento descentralizadas inspiradas no funcionamento do cérebro: apresentadas a uma série de estímulos (os dados a analisar), as redes os organizam e “aprendem” a reconhecer padrões e fazer previsões. Se recebem informações sobre a magnitude e direção dos erros que cometem, “aperfeiçoam seu conhecimento”, melhorando as respostas seguintes. Parecem providas de existência própria.

Os algoritmos genéticos tomam uma “população” de possíveis respostas a um problema e “selecionam” as mais bem adaptadas. Estas deixam mais “descendentes” que as demais. Sua “prole” passa por trocas “genéticas” ou mutações, incorporando novas características. As sucessivas descendências são novamente depuradas: respostas mais adaptadas deixam sempre mais descendentes. Ao final de várias “gerações”, a população de respostas “evolui” e, dentre elas, as melhores alternativas são escolhidas. Lembra a história das espécies.

Metáforas biológicas à parte, no entanto, as técnicas de *data mining* têm muito de reciclagem: são grandes fatias de estatística “velha” embrulhadas em celofane novo. As variações de regressão múltipla, por exemplo, continuam importantes como sempre. As novidades ficam por conta de mudanças de estratégia: em vez

de pequenas amostras, utilizam-se dados referentes à população toda (ou, pelo menos, milhões de observações); em vez de modelos bem desenvolvidos, aplica-se a força bruta do processamento digital; e, em vez de analistas com conhecimento de estatística, empregam-se ferramentas inteligentes, que “sabem o que estão fazendo”.

O celofane é provido pela informática: dentro do pacote, no entanto, os dados e a estatística continuam dando substância ao produto. Não encontramos ainda trabalhos publicados relatando comparações, mas arriscamos a opinião de que não há nada que se possa fazer com *data mining* que já não pudesse ser feito com as técnicas tradicionais. As aplicações, contudo, podem ser operacionalizadas bem mais facilmente, e isso não é pouco.

DATA MINING E MARKETING

Mesmo evitados modas e modismos, e descontado um certo exagero da mídia não especializada, o tema é realmente interessante e relevante, em particular no ambiente empresarial. Há possibilidades de aplicação em diversas áreas das organizações, mas as empresas que acumulam dados sobre clientes, produtos e concorrentes estão especialmente ansiosas para direcionar iniciativas de mineração para suas atividades de *marketing*. Principalmente agora que a adoção de estratégias de *marketing* de relacionamento está se tornando uma exigência inevitável, uma imposição competitiva.

Para a maioria das empresas de certo porte, e para quase todas as envolvidas com *e-commerce*, dados estão disponíveis em grandes volumes e muitas vezes até bem organizados em *data warehouses*. No entanto, não é exato nem automático que elas os estejam utilizando eficazmente. Dado não é informação.

As ferramentas de *data mining* prometem uma solução para a letargia de dados “encalhados” nos escaninhos. Agregando inteligência automática ou semi-automática aos *data warehouses* e *data marts*, permitem que se extraia deles informação que vale dinheiro. Uma estimativa do META Group (citada em GROTH, Robert, *Data mining: a hands-on approach for*

business professionals, Upper Saddle River: Prentice-Hall, 1998. 264 p.) prevê que as vendas de *software* para *data mining* praticamente dobrarão a cada ano, nos próximos anos, já atingindo 800 milhões de dólares no final do ano 2000, quando a mineração de dados será uma tecnologia crítica para a maioria das organizações internacionais. Não surpreende. De nada adianta colecionar dados e deixá-los adormecidos nas prateleiras. Quem já foi obrigado a investir em infra-estrutura tecnológica vai achar razoável investir um pouco mais para potencializar seu retorno.

Os livros *Data mining solutions* e *Data mining techniques* são, portanto, muito oportunos. Complementam-se. E, por oferecer uma visão ampla, clara, conceitual e ao mesmo tempo aprofundada e prática, destacam-se dentre as quase duas dezenas de títulos utilizados na disciplina de *Data Mining*, oferecida na EAESP a partir deste semestre.

Westphal é pesquisador e consultor na área de aplicações de mineração. Blaxton é especialista em representação, armazenagem e recuperação de conhecimento em sistemas biológicos (humanos) e artificiais (computadores). Berry e Linoff são diretores de uma empresa americana especializada em *data warehousing* e *data mining* para negócios, com atuação nas companhias listadas na *Fortune 100*. Em conjunto, os quatro autores cobrem os tópicos básicos para uma primeira visita ao tema da mineração: inteligência artificial, modelos quantitativos, aplicações práticas e mercadológicas e ferramentas específicas.

Data mining solutions orienta seus leitores quanto às etapas necessárias à coleta e ao processamento de dados na mineração, enfatizando sua modelagem como elemento determinante do sucesso da aplicação. Não podemos concordar mais completamente. Enfatiza, também, soluções visuais, reafirmando o poder das imagens na comunicação de conteúdos complexos. Westphal e Blaxton incluem, no livro e no CD-ROM, material sobre aplicativos comerciais, alguns em versão *demo* operacional, outros só com vídeos ou apresentações. Uma parte desse material sobre os aplicativos é extremamente interessante; outra, nem tanto, resvalan-

do para uma desavergonhada publicidade. No entanto, ainda que se descarte o que não interessa, o saldo é valioso – e deve ser conhecido logo, antes que se torne obsoleto, pois, na área de informática, o prazo de validade dos produtos é pequeno.

Data mining techniques aborda mais especificamente as aplicações mercadológicas. Os seis capítulos iniciais do livro introduzem conceitos básicos e colocam a mineração no contexto de negócios, retratando uma área em que as aplicações estão maduras. Berry e Linoff enfatizam a realização de estudos que levem a conclusões práticas e à implementação de estratégias que efetivamente conduzam ao aumento dos lucros. Os autores sugerem que, durante a leitura do texto, recitemos o mantra “analisar, agir, medir... analisar, agir, medir...”. Com essa proposta, conquistam imediatamente os adeptos do *marketing* quantitativo (embora talvez percam um pouco da estima dos fiéis aos Leões de Ouro) e os entusiastas do resgate das ferramentas estatísticas para a gestão empresarial. Após a leitura, muitos continuarão recitando o mantra durante suas atividades profissionais.

Os sete capítulos seguintes tratam de técnicas específicas: detecção de agrupamentos, modelagem por meio de exemplos, análise de cestas de produtos, algoritmos genéticos, análise de ligações, árvores de decisão e redes neurais. O texto é ilustrado com exemplos, e a matemática é mantida no plano conceitual. Além de informativo, *Data mining techniques* é de leitura agradável e acessível.

EM RESUMO

No conjunto, *Data mining solutions* e *Data mining techniques* abrem novos caminhos empresariais para profissionais de administração, *marketing*, informática e estatística: identificam oportunidades, explicam os conceitos, sugerem aplicações, relatam casos, recomendam leituras adicionais. Além disso, levantam várias oportunidades de pesquisa, interessando, portanto, também a mestrandos e doutorandos em fase de escolha de temas. Finalmente, em tempos recessivos como os de hoje, são, seguramente, um investimento em empregabilidade. ○