**cnet NEWS.COM**
TECH NEWS FIRST

# Will data mining revolutionize e-commerce?

By [Rachel Konrad](#)
Staff Writer, CNET News.com
February 8, 2001

**You like science fiction books, and Amazon.com wants to sell them to you. So why does the e-commerce giant peddle DVDs, Q-Tips and Valentine's Day chocolates when you click on its site?**

The answer is simple, scientists say: Amazon and most other e-tailers have yet to perfect a practice known as "data mining," the use of statistical analysis to uncover hidden patterns in otherwise random information.

Experts predict data mining will be one of the most revolutionary developments of the next decade, key to delivering a "personal Web," tailored to an individual's preferences, by identifying a useful structure in collected information and analyzing it in real time. The influential MIT Technology Review recently [hailed](#) data mining as one of the 10 emerging technologies that will "change the world."

[Data mining could dig gold for e-commerce](#)
Usama Fayyad, CEO, DigiMine

But some academics warn that mainstream mining merely "dumbs down" the sophisticated craft--and may result in screwy conclusions. Already, analysts are cautioning potential investors that the volatile segment may be unduly hyped.

"A lot of people think, 'I'm just going to put this in the hands of the marketer and we'll get the secret sauce,'" said Bob Moran, a managing vice president at the Boston-based Aberdeen Group. "But there's no such thing as 'secret sauce.' Data mining is all about pushing back the gray zone. It's never entirely uncovering the black and white."

But marketers who recognize its vast commercial potential see data mining as more than black and white. They also see green in the science's potential to create higher margins and inflate revenue.

Sophisticated or not, various forms of data-mining development are being undertaken by companies looking to make sense of the raw data that has been mounting relentlessly in recent years. A recent article in the Engineering News-Record noted that e-commerce has empowered companies to collect vast amounts of data on customers--everything from the number of Web surfers in a home to the value of the cars in their garage.

"DATA MINING IS ALL ABOUT PUSHING BACK THE

"Over the past few years, while (database) construction has gradually taken up digital information tools in pursuit of efficiency and profit, a by-product--mountains of recorded data-- has been gathering," Tom Sawyer wrote in a November edition of the industry trade publication. "Now, the realization is spreading that the mountains are filled with gold."

About a dozen small data-mining companies are jockeying to gain market share, and database and software companies such as Oracle and IBM are edging into the field. Others are creating more automated data-mining applications for nonstatisticians, making the science more tangible to marketers and other algorithm-ignorant users.

Through data mining, marketers can target customers with personalized stock quotes, news updates, special promotions and other information they are most likely to use, dramatically reducing advertising budgets and boosting revenue. It is also entirely automated, reacting instantly to changes in a customer's behavior, unlike the vast majority of personalized services on the Web today that require people to fill out questionnaires.

Perhaps the biggest challenge for data mining is one that many experts say cannot be solved--and one that may justify skepticism about the entire niche. Data mining is a good predictor of consumer behavior based on past behavior--what people are likely to purchase based on previous transactions, demographic information and other data points. But, critics say, it will never be able to predict what people really want to buy.

For example, data mining can determine that a 34-year-old, home-owning woman with two children is likely to purchase a detached microwave every three years for the next decade. Yet it cannot determine that this particular consumer would rather purchase a more expensive integrated microwave-convection oven combination if it came vaguely into her price range.

Kyle Johnstone, director of business intelligence for Emerald Solutions, said figuring out what people would rather purchase, as opposed to what they merely settle for, is the key to inflating profit margins--the ultimate goal of marketers. The only way to do that is to ask people what they really want, as opposed to relying on previous spending habits.

"People will tell you they like steak, but when they have parties for the Fourth of July, they buy hamburger. There's a

disconnect between what you buy and what you desire," Johnstone said. "You can figure out the behavior of performance metrics, but what you're missing--the biggest piece of the puzzle--is what it is that people really want…It's mathematically impossible to determine that."

## Skirting privacy concerns

Most data-mining companies get customer information from the corporate clients that hire them to build and host their databases for fees that usually start at about $10,000 per month. The data miners skirt privacy concerns by keeping the information in-house.

They then crunch the data and send it back to the client in the form of spreadsheets, graphics, bar charts and other visual documents. Some data-mining companies also act as consultants, recommending to clients how to tweak Web pages for maximum effectiveness.

Few data-mining companies are willing to discuss real-world examples of how the craft has boosted sales or customers. But Usama Fayyad, a former Microsoft executive, who left the company to create Kirkland, Wash.-based DigiMine, said he used data mining to help revamp Microsoft's MSNBC.com Web site and boost readership.

Fayyad found that a 22 percent slice of MSNBC readers had nearly identical online behavior, clicking on exactly the same reports. But these users didn't fit into any of the company's five reader categories, which included political news-hounds, sports junkies and weather buffs.

Fayyad, who holds a doctorate from the University of Michigan, said his company determined that the glue holding this mysterious group together was vaguely scandalous stories similar to those in gossip tabloids. MSNBC changed its format significantly to appeal to this large group, and now the home page is required to have at least one such feature per day. The research helped turn MSNBC's Living section into the site's most popular destination, Fayyad said.

"The lesson is that before data mining, they didn't know what was happening to a quarter of their database," Fayyad said. "If three or four shelves fall over in a brick-and-mortar store, the customers won't walk around them and the clerks will fix them. The equivalent is happening on the Web, but no one knows how to fix the bottlenecks."

> "PEOPLE WILL TELL YOU THEY LIKE STEAK, BUT WHEN THEY HAVE PARTIES FOR THE FOURTH OF JULY, THEY BUY HAMBURGER. THERE'S A DISCONNECT BETWEEN WHAT YOU BUY AND WHAT YOU DESIRE."
>
> –KYLE JOHNSTONE, DIRECTOR OF BUSINESS INTELLIGENCE AT EMERALD SOLUTIONS

## Applying data mining to e-commerce

For decades, utility companies have been using data mining to predict with some accuracy when generators are likely to fail. The technique started making more inroads into the corporate world in the 1990s, catching on as a means to detect fraud in the insurance, health care and credit card industries. By finding patterns and predicting likely behavior, companies can catch people who lie on applications or are likely to engage in dangerous or illegal activities.

Department stores, supermarkets and other brick-and-mortar retailers have used data mining to guess customer buying habits for years, but relatively few general consumer e-tailers and content producers have fully exploited the research technique. That's partly because the practice--involving algorithms, samplings and parallelisms--is complicated and poorly understood. But it's starting to find its way into the mainstream.

"E-commerce is the newest and hottest use," said Michael Gilman, president and chief executive of Data Mining Technologies of Bethpage, N.Y. "Anywhere you have historical data, you can use it to get patterns that you can't see with the human eye."

One of the oldest and largest data-mining companies is the 25-year-old SAS Institute, based in Cary, N.C., which says it had already been working with 98 percent of Fortune 500 companies and is now targeting e-commerce. Retailers that sell products via catalogs and Web sites routinely increase their return on investment by more than 1,000 percent by using data mining, according to SAS statisticians.

"A lot of catalog companies were doing a fine business before, thank you very much," said Anne Milley, analytical strategist for SAS. "Then we came in and they were amazed. You look at who they're targeting, what they're sending and how often, and the frequency of repeat purchasers. You look at marketing mix--who is buying through catalogs, who is buying online--and figure out what is the optimal way to contact customers."

Data mining is likely to penetrate society further as the technology becomes easier to use.

San Mateo, Calif.-based Epiphany is one of several Web-based customer relationship-management companies that is deeply involved in data mining and is well known for its relatively easy-to-use tools.

George John, who has a doctorate in statistics from Stanford University and is the self-declared "data mining guru" of Epiphany, said the company's controversial simplification of data mining was intentional. He considers it one of Epiphany's biggest attributes when vying for business against other data-mining companies--which feature software that may be more sophisticated but is usually vastly more elusive to the average business.

"In the first generation of data mining at Epiphany, we tried to step back and see what business users would use it for--we knew they'd be asking lighter questions, where you wouldn't need 10 Ph.Ds forecasting profitability down to the penny," said

**VARIOUS**
FORMS OF
DATA-MINING
DEVELOPMENT
ARE BEING
UNDERTAKEN
BY COMPANIES
LOOKING TO MAKE
SENSE OF THE RAW
DATA THAT HAS
BEEN MOUNTING
RELENTLESSLY IN
RECENT YEARS.

John, an IBM veteran who began the data-mining program at Epiphany. "Every time we tried to make the (user interface) cleaner, we thought, 'Now the marketers will use it.' It was just paying attention to what people wanted to do."

Though it seems logical, the practice of simplifying data-mining results has its detractors. Fayyad and other experts warn that excessive simplification can skew results and lead executives to make pricing or inventory decisions based on faulty reasoning.

A more fundamental controversy is also brewing as data mining moves out of academia and into the corporate world: Academic statisticians take pride in their complex analyses, and many snub fellow Ph.Ds who enter corporate environments, calling them turncoats pandering to marketers.

John, the Epiphany guru, says he must constantly correct people who use the term "dumbing down" to refer to the company's color charts and other simple statistical diagrams. He prefers to call it "deeper penetration" of data mining into the ranks of marketers and other nonstatisticians.

"We profile a set of customers with nice charting, drawing pictures of what customers are like," John said, almost apologetically. "The key was admitting that was OK. It was OK if the technology behind it wouldn't get you a Nobel Prize." ■

c|net