



# Data Mining: What Lies Beneath?

BY SARAH L. ROBERTS-WITT NOVEMBER 19, 2002

*Finding patterns in customer behavior can deliver profitable insights into your business.*

When you have well over half a million customers, it's almost impossible to understand the likes and dislikes of each or what motivates them to buy your products. But as the manager of Dow Jones Interactive's Wall Street Journal Online knows, if you can understand how customers behave, you can get much closer to delivering what they want.

When you have as many different types of subscribers as WSJ Online does, that's no small feat. Some go directly to the site to sign up; some receive the print version but also want the online version, which they get for a discount. Some are corporate subscribers, and some pick up subscription information at Barnes & Noble. "The incredible thing is, that's not even all of the ways people can subscribe," says Todd Larsen, general manager of consumer electronic publishing for WSJ Online. "Trying to understand how all these groups behave on the site and why is a major challenge."

The trick, he says, is to correlate the log and clickstream information generated by the Web site with the customer files. Though the online staff has been able to grab that information for some time, it wasn't easy. Doing so required manually matching files, which was time-consuming and prone to errors.

That's why in 2001, the online team started evaluating more automated approaches, such as data mining and analytics (or business intelligence). After considering outsourcing the project and building its own data warehouse, WSJ Online signed up with data-mining and analytics provider digiMine in December of last year. The goal was to have a data infrastructure ready for WSJ Online's relaunch on January 28, 2002.

digiMine delivered on time, and the company's services are now helping Larsen and his team better understand how the site is performing and devise improvements. WSJ Online's department heads and marketers now have access to analysis that helps them identify which customer groups are worth spending marketing dollars on. They can also determine the best pricing models. The point, of course, is to boost the bottom line.

"When we started this process, we were looking at the return-on-investment factor," says Larsen. "Making better marketing-spend and pricing decisions definitely justifies the investment here."

# Data Mining: What Lies Beneath?

BY SARAH L. ROBERTS-WITT NOVEMBER 19, 2002

*Finding patterns in customer behavior can deliver profitable insights into your business.*

## What's in the Data?

Many of the underlying technologies that WSJ Online is now benefiting from—data-mining algorithms, data modeling, data cleansing, and even analytics techniques like online analytical processing (OLAP)—have been around for decades, but only now have companies seriously put them to work on corporate data. Having spent much of the 1990s busily building Web sites and installing software for enterprise resource planning, customer relationship management, and sales-force automation, companies are finding themselves awash in more data than they ever dreamed of. The result is more data and a growing need to understand its implications for business, especially during tough economic times.



Many companies are folding formerly separate Web entities back into their cores, which is assisting in merging online and off-line data, especially customer-related data, says Anne Milley, director of analytical strategies at SAS, an analytics software firm. Businesses are also more comfortable with data mining and analytics today than they were five years ago. "Companies aren't as scared or skeptical of data mining as they once were," says Jeff Jones, director of strategy for IBM's data management solutions. "In the last three to four years, the levels of familiarity with these concepts and technologies have increased—as has the level of sophistication."

Some of the data-mining and business intelligence tools have gained a secure foothold in retail, telecom, finance, life sciences, pharmaceuticals, and government. These sectors are using the technology to assist with customer segmentation, correlation of disparate data, and fraud detection. Data mining is also being used in security. (For more on data mining and security, see the sidebar ["Mining for Counterterrorism."](#))

Take Ace Hardware Corp., the 78-year-old "Helpful Hardware Place" based in Oak Brook, Illinois. Ace Hardware is a wholesale cooperative made up of 5,100 independent retail stores. The buyers in Ace's corporate-merchandising department purchase products centrally, which store managers then purchase at wholesale. Ace offers 65,000 products from 3,000 vendors.

Though the merchandising staff has advanced data-mining tools that help them choose the best products, the same tools have not been available to the retail stores. "We've got a good handle on the analytics systems that serve our buyers and dealers, but now we're focusing on the retail stores," says Brian Smetana, senior business analyst at Ace.

The push to deliver better data to Ace's retail stores started in 1997, when the company began refining its data warehouse. The best solution so far for getting data in has been for the stores to dial into the corporate network and download data from their point-of-sale systems. Some 1,600 stores call in each night.

Getting their data into the warehouse lets corporate analysts suggest marketing strategies and decide which products to sell at which retail stores. Though stores can already get basic sales data, Smetana and his team are developing reports that will let stores compare their sales to those of other stores across their region and across the country.



# Data Mining: What Lies Beneath?

BY SARAH L. ROBERTS-WITT NOVEMBER 19, 2002

*Finding patterns in customer behavior can deliver profitable insights into your business.*

## Text Mining on the Horizon

These data-mining projects are encouraging, but even more exciting developments are on the horizon. Predictive analytics is one of the top five data-mining trends for 2002 and 2003, according to Aaron Zornes, executive vice president of research firm Meta Group. Nearly every developer we spoke with for this story either has or is developing predictive-analytics products. "This is about mixing history with predictions of what's going to happen next and then having the system present relevant recommendations for what to do," says Sanju Bansal, COO at MicroStrategy, an analytics software company.

Text mining is also on developers' radar screens. Though search technology developers have claimed text-mining capabilities for years, companies are now defining text mining as the ability to correlate trends found in text (e-mail, marketing documents, conversations conducted between customers and call centers, and threaded discussions) with the numbers-oriented or structured data typically mined in data warehouses. IBM, SAS, Insightful, and Intelligent Results are all working in this area.

Meanwhile, companies like Cisco Systems are starting to offer data-mining security products to automate the analysis of unusual network activity. And newcomer Netezza recently launched the Netezza Performance Server 8000 appliance to increase the speed of data mining.

Any data-mining project should be driven by extreme pragmatism. First, you should know the business questions you want your data to answer before spending one red cent. digiMine CEO Usama Fayyad urges that businesses should always say, "Sell me something that solves a business problem," when evaluating products and services.

Second, realize that leveraging your data is a project that will never end, but it is one that can deliver great rewards. "We're continually refining the way we segment and analyze our data," says WSJ Online's Larsen. "But that also means we're getting even better information out of it, which is part of the payback." And you don't need a data-mining tool to see the benefits of that.