The New York Times
nytimes.com

May 20, 2007

BRIGHT IDEAS

Reaping Results: Data-Mining Goes Mainstream

By **STEVE LOHR**

RODNEY MONROE, the police chief in Richmond, Va., describes himself as a lifelong cop whose expertise is in fighting street crime, not in software. His own Web browsing, he says, mostly involves checking golf scores.

But shortly after he became chief in 2005, a crime analyst who had retired from the force convinced him to try some clever software. The programs cull through information that the department already collects, like "911" and police reports, but add new streams of data — about neighborhood demographics and payday schedules, for example, or about weather, traffic patterns and sports events — to try to predict where crimes might occur.

"It sounded nutty at first," Mr. Monroe recalled, "but the more and more you get into it, the more sense it makes."

The technology, for example, pointed to a high rate of robberies on paydays in Hispanic neighborhoods, where fewer people use banks and where customers leaving check-cashing stores were easy targets for robbers. Elsewhere, there were clusters of random-gunfire incidents at certain times of night. So extra police were deployed in those areas when crimes were predicted.

The crime rate in Richmond declined about 20 percent last year, and it is down again this year.

The Richmond experience is part of a wave of sophisticated computing and mathematical analytics that is moving into the mainstream. Fueling the trend are the digitization of information, ever faster and cheaper computing, and the explosion of online networks and data collection.

The results, says Jon M. Kleinberg, a computer scientist at Cornell University, are a "revolution in measurement" and the "introduction of computing and algorithmic processes into the social sciences in a big way." The phenomenon is strikingly evident in economics, business and crime prevention.

Productivity research has traditionally focused on manufacturing, because the output of widgets and the headcount of factory workers were easy to measure, notes Erik Brynjolfsson, a professor at the Sloan School of Management at the Massachusetts Institute of Technology.

The productivity of information workers — much of the nation's work force — was shunted into a category that economists labeled "difficult to measure" and given short shrift.

But the digital age, he says, has opened the door to detailed measurement of the labor of professionals and office workers who handle ideas and information from customers, suppliers, colleagues and marketers.

"My thinking on productivity has completely changed," says Mr. Brynjolfsson, who is also a research associate at the National Bureau of Economic Research.

By tracking e-mail traffic, instant messages and other digital communications — stripped of personally identifiable information — he and other researchers are beginning to study the flow of work and ideas through the social networks inside companies — minute by minute, bit by bit.

"We're really on the cusp of being able to understand what goes on inside corporations in a much more scientific way than ever before," he said. "It's similar to the way that the microscope opened up biology in the 17th century, so that you could see blood cells. Now, we can start to see bits of information as they flow through the organism of the corporation."

The desire to exploit computing and mathematical analytics is by no means new. In the 1960s and '70s, "operations research" combined computing and math mainly to make factory production work more efficient. And in that period, "decision support" software was intended to help managers more intelligently use information in the big computing file cabinets — databases — that were becoming common in corporations.

But the earlier efforts were limited mainly to information access and reporting systems, says Thomas H. Davenport, a professor at Babson College. The quantity and quality of data were typically inadequate, he notes, and the software could not do the advanced optimization and predictive calculations of today's programs.

Faster and cheaper computing and ample sources of information in digital form — plucked from enterprise resource planning systems, point-of-sale devices and Web sites — mean that most companies now have the tools to do the kind of competitive analytics that only a relative handful of elite companies could do in the past. "It's really starting to become mainstream," says Mr. Davenport, co-author with Jeanne G. Harris of "Competing on Analytics: The New Science of Winning" (Harvard Business School Press, 2007). The entry barrier, he says, "is no longer technology, but whether you have executives who understand this."

There are plenty who do. Big retailers like Wal-Mart Stores and Kohl's use today's advanced computing and math to more accurately predict what sizes of clothes should go to what stores. Harrah's and other casinos decipher slot-machine results to optimize customer traffic and profits, and they use face-recognition software to identify people with criminal records. And Stockholm and other cities use traffic data and patterns to determine "congestion pricing."

In the financial industry, Capital One and other banks mine all kinds of transaction data to identify, and stop, fraudulent transactions. And Cemex, the big cement company, uses global positioning satellite locators and traffic and weather data to improve delivery-time performance in Mexico.

In the last year or so, Whirlpool, the appliance maker, has begun using new analytics software to automatically scan warranty reports as well as manufacturing, supplier, sales and service data to try to further trim its warranty costs and improve quality. That is no small task, since it sells an average of 25,000 washing machines a day, for example. "A human being cannot see and detect all those trends," says John Kerr, the general manager for global quality.

With the new computing tools, Whirlpool has trimmed by 30 to 90 days the time required to detect and fix parts or manufacturing problems that cause defects. "The math is astounding," Mr. Kerr says.

The results help explain why business-intelligence software is one of the hot markets in technology, supplied by companies like SAS, Business Objects, Cognos, MicroStrategy and Information Builders. In March, Oracle offered a hefty $3.3 billion for Hyperion, a maker of business intelligence software. Microsoft has entered the field as well.

But packaged software is not the only way to combine powerful computing with deep math tools. The major technology services companies, like I.B.M., Accenture and Hewlett-Packard, have researchers, programmers and industry specialists doing this kind of work for clients.

Internet marketing and advertising is a social market made for the use of heavy-duty computing and

sophisticated mathematics. Investment and start-up money is pouring into the market, and so are many high-powered computing brains.

Basem Nayfeh has a Ph.D. from Stanford, where he did his graduate research down the hall from one of Google's founders, Sergey Brin. Mr. Nayfeh's thesis was on multiprocessor chips, and he has worked in corporate labs in Silicon Valley on things as diverse as climate and computer design.

Today Mr. Nayfeh, 37, is the chief technology officer of Revenue Science, which tracks, analyzes and predicts online behavior to help advertisers find people most likely to buy their products. Many of his fellow computer wizards are in online marketing.

"If you asked any of us 5 or 10 years ago if we would be in advertising," he says, "none of us would have said yes."