An Executive's Guide – Digging up Dollars with Data Mining

By Tim Graettinger, President, Discovery Corps, Inc.

Introduction

Traditionally, organizations use data tactically - to manage operations. For a competitive edge, strong organizations use data strategically – to expand the business, to improve profitability, to reduce costs, and to market more effectively. Data mining creates *information assets* that an organization can leverage to achieve these strategic objectives.

In this article, we address some of the key questions executives have about data mining. These include:

- What is data mining?

- What can data mining do for my organization?

- How can my organization get started?

Business Definition of Data Mining

Data mining is a new component in an enterprise's decision support system (DSS) architecture. It complements and interlocks with other DSS capabilities such as query and reporting, on-line analytical processing (OLAP), data visualization, and traditional statistical analysis. Shown in Figure 1, these other DSS technologies are generally retrospective. They provide reports, tables, and graphs of what happened in the past. A user who knows what she's looking for can answer specific questions like: "How many new accounts were opened in the Midwest region last quarter," "Which stores had the largest change in revenues compared to the same month last year," or "Did we meet our goal of a ten-percent increase in holiday sales?"

We define data mining as "***the data-driven discovery and modeling of hidden patterns in large volumes of data***." Data mining differs from the retrospective technologies above because it produces models - models that capture and represent the hidden patterns in the data. Via data mining, a user can discover patterns and build models automatically, without knowing exactly what she's looking for. The models are both descriptive and prospective. They address why things happened and what is likely to happen next. A user can pose "what-if" questions to a data-mining model that can not be queried directly from the database or warehouse. Examples include: "What is the expected lifetime value of every customer account," "Which customers are likely to open a money market account," or "Will this customer cancel our service if we introduce fees?"
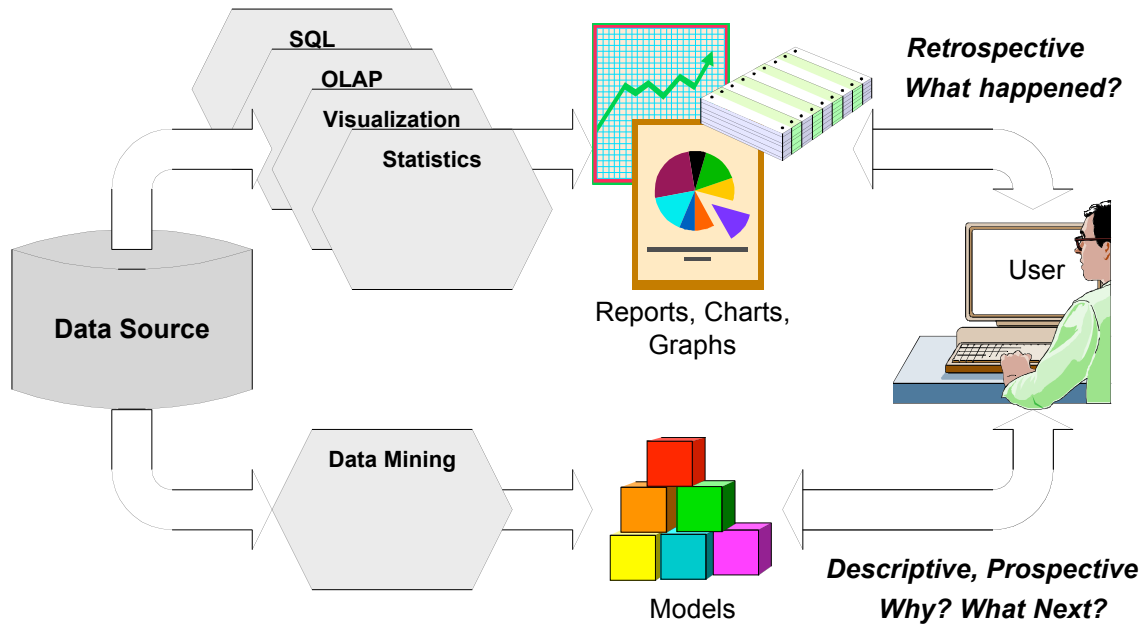
Figure 1 - A Decision Support System (DSS) architecture highlighting the descriptive, prospective nature of data mining

The information technologies associated with data mining are neural networks, genetic algorithms, fuzzy logic, and rule induction. It is outside the scope of this article to elaborate on all of these technologies. Instead, we will focus on business needs and how data mining solutions for these needs can translate into dollars.

Mapping Business Needs to Solutions and $$$$s

What can data mining do for your organization? In the introduction, we described several strategic opportunities for an organization to use data for advantage: business expansion, profitability, cost reduction, and sales and marketing. Let's consider these opportunities very concretely through several examples where companies successfully applied data mining.

Expanding your business: Keystone Financial of Williamsport, PA, wanted to expand their customer base and attract new accounts through a LoanCheck offer. To initiate a loan, a recipient just had to go to a Keystone branch and cash the LoanCheck. Keystone introduced the $5000 LoanCheck by mailing a promotion to existing customers.

The Keystone database tracks about 300 characteristics for each customer. These characteristics include whether the person had already opened loans in the past two years, the number of active credit cards, the balance levels on those cards, and finally whether or not they responded to the $5000 LoanCheck offer. Keystone used data mining to sift through the 300 customer characteristics, find the most significant ones, and build a model of response to the LoanCheck offer. Then, they applied the model to a list of 400,000 prospects obtained from a credit bureau.

By selectively mailing to the best-rated prospects determined by the data-mining model, Keystone generated $1.6M in additional net income from 12,000 new customers.

**Reducing costs**: Empire Blue Cross/Blue Shield is New York State's largest health insurer. To compete with other healthcare companies, Empire must provide quality service and minimize costs. Attacking costs in the form of fraud and abuse is a cornerstone of Empire's strategy, and it requires considerable investigative skill as well as sophisticated information technology.

The latter includes a data mining application that profiles each physician in the Empire network based on patient claim records in their database. From the profile, the application detects subtle deviations in physician behavior relative to her/his peer group. These deviations are reported to fraud investigators as a "suspicion index." A physician who performs a high number of procedures per visit, charges 40% more per patient, or sees many patients on the weekend would be flagged immediately from the suspicion index score.

What has this data mining effort returned to Empire? They realized fraud-and-abuse savings of $29M in 1995, $36M in 1996, and $39M in 1997.

**Improving sales effectiveness and profitability**: Pharmaceutical sales representatives have a broad assortment of tools for promoting products to physicians. These tools include clinical literature, product samples, dinner meetings, teleconferences, golf outings, and more. Knowing which promotions will be most effective with which doctors is extremely valuable since wrong decisions can cost the company hundreds of dollars for the sales call and even more in lost revenue.

The reps for a large pharmaceutical company collectively make tens of thousands of sales calls. One drug maker linked six months of promotional activity with corresponding sales figures in a database, which they then used to build a _predictive model for each doctor_. The data-mining models revealed, for instance, that among six different promotional alternatives, only two had a significant impact on the prescribing behavior of physicians. Using all the knowledge embedded in the data-mining models, the promotional mix for each doctor was customized to maximize ROI.

Although this new program was rolled out just recently, early responses indicate that the drug maker will exceed the $1.4M sales increase originally projected. Given that this increase is generated with no new promotional spending, profits are expected to increase by a similar amount.

Looking back at this set of examples, we must ask, "Why was data mining necessary?" For Keystone, response to the loan offer did not exist in the new credit bureau database of 400,000 potential customers. The model predicted the response given the other available customer characteristics. For Empire, the suspicion index quantified the differences between physician practices and peer (model) behavior. Appropriate physician behavior was a multi-variable aggregate produced by data mining - once again, not available in the

database. For the drug maker, the promotion and sales databases contained the historical record of activity. An automated data mining method was necessary to model each doctor and determine the best combination of promotions to increase future sales.

Getting Started

In each case presented above, data mining yielded significant benefits to the business. Some were top-line results that increased revenues or expanded the customer base. Others were bottom-line improvements resulting from cost-savings and enhanced productivity. The natural next question is, "How can my organization get started and begin to realize the competitive advantages of data mining?"

In our experience, *pilot projects* are the most successful vehicles for introducing data mining. A pilot project is a short, well-planned effort to bring data mining into an organization. Good pilot projects focus on one very specific business need, and they involve business users up front and throughout the project. The duration of a typical pilot project is one to three months, and it generally requires 4 to 10 people part-time.

The role of the executive in such pilot projects is two-pronged. At the outset, the executive participates in setting the strategic goals and objectives for the project. During the project and prior to roll out, the executive takes part by supervising the measurement and evaluation of results. Lack of executive sponsorship and failure to involve business users are two primary reasons data mining initiatives stall or fall short.

In reading this article, perhaps you've developed a vision and want to proceed - to address a pressing business problem by sponsoring a data mining pilot project. Twisting the old adage, we say "just because you should doesn't mean you can." Be aware that a capability assessment needs to be an integral component of a data mining pilot project. The assessment takes a critical look at data and data access, personnel and their skills, equipment, and software. Organizations typically underestimate the impact of data mining (and information technology in general) on their people, their processes, and their corporate culture. The pilot project provides a relatively high-reward, low-cost, and low-risk opportunity to quantify the potential impact of data mining.

Another stumbling block for an organization is deciding to defer any data mining activity until a data warehouse is built. Our experience indicates that, oftentimes, data mining could and should come first. The purpose of the data warehouse is to provide users the opportunity to study customer and market behavior both retrospectively and prospectively. A data mining pilot project can provide important insight into the fields and aggregates that need to be designed into the warehouse to make it really valuable. Further, the cost savings or revenue generation provided by data mining can provide bootstrap funding for a data warehouse or related initiatives.

Recapping, in this article we addressed the key questions executives have about data mining - what it is, what the benefits are, and how to get started. Armed with this knowledge, begin with a pilot project. From there, you can continue building the data mining capability in your organization; to expand your business, improve profitability, reduce costs, and market your products more effectively.

Tim Graettinger, Ph.D., is the founder and President of Discovery Corps, Inc. For nearly two decades, he has worked with clients to develop data mining solutions to significant business problems. Along with his consulting work, Dr. Graettinger provides data mining training courses and workshops to client organizations. Tim also has patented three data mining software inventions (US Patent Nos. 5839438, 5640491, and 5477444). For more info, visit http://www.discoverycorpsinc.com or contact:

Discovery Corps, Inc.
Phone: (724)-743-3642
E-mail: tgraettinger@discoverycorpsinc.com