

What your reports AREN'T telling you

By RDA Senior Consultant Jim Pletscher with RDA Project Manager Derek Strausbaugh

EVERY business wants to answer the same basic question. “What activities are helping to grow the business and what activities are wasting our valuable time?” This same question applies whether the focus is on marketing, sales, operations, or customer satisfaction. Initially, the primary tool that businesses could rely on was intuition – a ‘gut feeling’ about what worked and didn’t work. However, even intuition is based on information, and the more informed one is about their business, the better equipped they are to determine what works and doesn’t work.

As a result, for decades, collecting information was the goal and computer applications have been created to generate and gather mountains of data. But now what? You have lots of information, and you have reams of reports that answer the questions asked, but what AREN'T your reports telling you? Reports only answer the questions that you THINK to ask. What about the nuggets of value in your information that no one has identified or for which no report has been created? How do you get to this valuable, unearthed data? Answer: Data Mining. Data mining can point you to trends in your data that may have been obscured in reports and therefore may have previously been overlooked. As such, it can be a powerful tool for analysts to use in the pursuit of Business Intelligence.

What Data Mining IS (and is NOT)

Data mining is one of those concepts that has generated much interest, and as a result, some misconceptions have developed. So to help define what data mining *IS*, it is helpful to clarify what it is *NOT*.

- Misconception #1: Data mining is the next ‘Big Thing’ and will revolutionize our business. Truth: While recent advancements in hardware and software have made data mining more accessible to a wider range of businesses, the underlying concepts have been around for quite some time. And while the benefits from data mining can be significant, its impact should be viewed as more evolutionary to your business than revolutionary.
- Misconception #2: Data mining is a complex process only understandable by mathematicians. Truth: Data mining does make use of very advanced mathematics and statistical analysis in order to identify trends in your data. However, while these processes used to be very manual in nature and difficult to apply, recent advancements in data mining software have made these tools much more user friendly.
- Misconception #3: Data mining is only beneficial to very large organizations. Truth: While data mining’s roots are with very large organizations, like credit card processing companies, the idea behind data mining is to learn from the past, and that is a concept that is beneficial to businesses of all sizes.

Having addressed some misconceptions, data mining can be defined as follows:

Data mining is a process of applying known statistical models to historical data in order to reveal trends, patterns, and success factors which businesses can then apply to present and future business activities.

Now, to better illustrate the role that data mining can play in an organization, let's consider a *traditional* mining scenario and then see how these concepts translate to *data* mining.

Mining for Gold

Suppose we acquired a large plot of land during the gold rush and we wanted to mine it for gold. We could grab a pick and just start digging in one corner of the property, but results would probably not be what we want.

A better approach would be to first survey the entire area so that we can have a more accurate picture of our resources. We could then head to some target areas that seem most likely to yield results and start digging. While finding gold in this manner is possible, success is likely a hit-or-miss proposition.

What would help is if we had advice from someone who knew this land and could advise us as to the most likely places to dig. So, using our survey of the area, we decide to talk to our neighbors to see where they have already found gold on their properties. One neighbor tells us that he's found most of his gold in sections of streams that are 12" deep or less. Another neighbor tells us that he's found most of his gold in particular kinds of rock formations. Using the knowledge gained from the neighbor's experiences, we are now able to identify and dig in very targeted areas of our land and before long are gathering significant value from our investment.

Mining for Success

Let's relate this analogy to data mining. We have mountains of data hiding nuggets of information – like our plot of land. Most companies just run reports on this data as is, but that's grabbing a pick and just starting to dig. The data is scattered across multiple applications and isolated reports do not tell us the whole story. What we need is a complete picture of the business. This is the purpose of a data warehouse and OLAP cubes.

A data warehouse is intended to consolidate data from various sources in the business into a consistent and usable format for reporting and analysis. OLAP cubes are data structures that enable optimized ad-hoc analysis of targeted slices of data in the data warehouse. Together, the data warehouse and OLAP cubes are like the survey of the property from our illustration. However, while they give us a more complete view of the organization, stopping here is analogous to digging without the benefit of advice from someone knowledgeable about the land – we may have some success, but this can also be a hit-or-miss proposition. What we need is an indication of where to dig – that is where data mining comes in.

With data mining, we can apply well-known statistical methods to develop mining models. These models are similar to the experienced neighbors in our analogy. The model analyzes historical data to identify trends and patterns that can indicate where to look. To take the analogy one step further, when we asked our two neighbors for advice, we didn't get the same answer from both of them. That doesn't mean that one was right and one was wrong, just that there may be more than one answer to the question. Similarly, the use of different data mining models is not necessarily going to produce the same results, since there isn't necessarily a right and wrong answer. Instead, the different mining models will reveal patterns and trends of which we may not have been aware. For this reason, it's often best to process data through several models in order to gain a complete understanding of what is happening in the business.

Appreciating the value of data mining also means understanding its limitations. Data mining is most accurate when predicting the *probable* results of a *group*, not necessarily individual results. Just as, in our analogy, the advice we received helped us know where gold is LIKELY to be found, it did NOT tell us whether it would be under a specific

rock or not. In the same way, data mining results will help us identify the common attributes of our customers, but they can't tell us whether or not a given individual will buy our product or service.

Lastly, data mining tools are just that – tools! They don't eliminate the need for good business analysts. Rather, the data mining tools can help point the business analysts in the right direction so that they are digging in the best possible locations, and therefore able to produce more well-developed insight and better analysis.

Practical examples

How might data mining provide value in a real business scenario? Let's suppose we are in the publishing business, and our marketing efforts are very broad. We'd like to develop more targeted marketing campaigns in order to maximize our marketing budget. Presuming that we have identified and are collecting the appropriate data in our data warehouse, we may choose to implement some data mining models to analyze our sales data for the past year and see some interesting trends emerge. We notice that the fitness books that we publish sell best to men and women who are in the 25-35 age group and who have an income greater than \$25,000. We also notice that the diet books we publish sell primarily to women, typically women with two or more children, and sell mostly between the months of January and April. Now that we have uncovered very specific trends in our sales data, we may use this information to develop much more targeted marketing campaigns.

This is a simple example, but demonstrates the practical benefits that data mining can provide. Additional application areas could include the following scenarios:

- A credit card company can analyze its transaction data to better identify fraud indicators, allowing them to improve their customer qualification processes and minimize costly payouts.
- A manufacturer can apply data mining algorithms to product testing data in order to better understand the common attributes of products that fail. This can enable them to improve the manufacturing processes in order to minimize scrap and rework, which allows them to become more profitable.
- An e-commerce business can use market-basket analysis to determine up-sell and cross-sell recommendations based on the buying patterns of prior customers.

Conclusion

Data mining is NOT an obscure discipline that can only be understood by mathematicians nor is it a replacement for strong business analysts. It is also NOT a precise solution that will provide you with a direct formula for success. However, data mining IS a process of applying known statistical models to historical data in order to reveal trends, patterns, and success factors which businesses can then apply to present and future business activities. Considering the advancements in computer hardware and software over the past decade, data mining is no longer only an option for very large organizations, but small-to-midsize ones as well. In periods of strong growth or in challenging economic times, all businesses need to leverage the valuable nuggets of information that may be hidden in their data. Data mining can help to achieve that objective.