A BUSINESS APPROACH TO DATA MINING

The modern way to analyse data.

SDART Ltd. MINSHULL HOUSE, CHORLTON STREET, MANCHESTER M1 3FY, UNITED KINGDOM
PHONE: +44 (0) 161 88 302 40, EMAIL: INFO@SDART.CO.UK

Data Mining is a modern way to analyse data, a process revealing useful patterns that can be essential in the decision-making process. By using proper Data Mining techniques and tools we are able to find patterns hidden from the human eye. We can deal with large data sets beyond human capabilities as well as reduce the analysis costs compared to conventional analysis.

Let's look at how Data Mining can support various decisions:

Firstly we need to have data to analyse. In most of the cases we are surrounded by available data sets and if not we can always collect it. Some might be related to our problem and some completely meaningless. However, at this early stage it may not be straightforward to decide its usefulness e.g. we have transactional data from sale systems, web usage logs or stock market quotes.

The next step is to give meaning to the observations collected about the process and thus make them into pieces of information related to our problem. At this stage information starts to be useful and provide meaningful details of the situation analysed e.g. we may examine the sales for a particular client, common patterns of website usage or trend changes for stock market.

Now we have information about different aspects of the problem we can transform it into a generally applicable construction, or knowledge. Now we have the capability to analyse situations and take decisions in relation to their context. That's how we can decide about inventory levels, profiles of website users or when to buy and sell shares.

Finally, a proper knowledge model can be used to deliver more reliable and in-depth information and thus bring business success. This scheme isn't new and we all use it even if we haven't heard about Data Mining. The difference made possible by Data Mining is the scale of problem and success of decision-making we can achieve with use of technology. But the increasing complexity of the business environment means we cannot do it without support.

Here are a few typical uses of Data Mining in various types of decision:

ASSOCIATION	Identifies frequently appearing observations e.g. we want to know which products are often bought together and could be offered in packs to increase sales.
CLASSIFICATION	Allows us to attach the observations to existing groups or categories e.g. recognising fraudulent transactions or possible bankrupt companies.
CLUSTERING	Searches for similarities between observations and groups them without having previous knowledge of such similarities e.g. looking for website usage or customer behaviour to create profiles for website users or clients.
DESCRIPTION	Gives us more details about patterns, presenting them visually or performing exploratory analysis. We not only perform data analysis to achieve applicable results, but we also want to know more about the analysed process e.g. we want to know what relations there are between the financial ratios and stock prices.
ESTIMATION	Reveals features that we cannot observe directly, because of the cost of observation or technical problems e.g. we are looking for the value of affordable budget for a household knowing only its size, members' education or location.
PREDICTION	Tells us about the future using past and current observations e.g. what the sales will be in the next period.

How can i use data mining in my business?

There are many models of the Data Mining process available. Some are perfect for a strictly scientific approach to a problem, where we just want to find the unknown or undiscovered relations between the data.

Other models were designed to be more specific to business use. A typical case would be a business question leading to a business goal e.g. Who are our most valuable clients and how to improve their satisfaction? Or what are our most lucrative sales and how to achieve more of them? What is causing our best selling products to no longer be popular? What can we do to protect our business from financial crisis?

Other areas in which Data Mining could provide valuable information include: investments evaluations, insurance, energy saving, political influence, health of the population, prediction for the future of the product or a service etc.

To prepare for analysis, you need to consider the following questions:

1. What problem could be solved by the data analysis?
2. What do you expect from this analysis? Better customer understanding? Lower costs? Better forecasting ability?
3. Who will be using it in your company - which employees? Do they have the skills to understand charts or detailed information, or do they require an easily understood proposition?
4. Are there any procedures used in this task? Can you form their precise rules?
5. Are there any repositories that could be used as data sources?
6. Should the solution developed be integrated with any existing software?

The most popular model in business used for organising the Data Mining process is **CRISP-DM** (*Cross-Industry Standard Process for Data Mining*).

The process consists of the following stages:

1

Business understanding starts with the business problems we want to solve and a description of existing solutions. In other words we define what we have at the moment and what we would like to have at the end of the process. These analyses must be transformed into a plan operating on proper data mining techniques.

2

Data understanding involves analysis of the data source and data gathering. Once we have this we can create charts, summaries and reports explaining interdependences between features. At this stage the quality of gathered data should also be verified.

3

Data preparation transforms raw datasets into sets containing ordered, verified and significant data. At this stage we also filter errors, fill the gaps and transform the features to forms required by Data Mining techniques. It is also important to reduce the dimension (range?) of datasets and to leave only important ones. A common approach is to gather all possible data and to examine their significance later without assumptions which might lead to the incorrect rejection of features with high potential.

4

Modelling activates the knowledge acquired using previously generated information. To do so we must select the modelling techniques, quality evaluation criteria, and the form of modelling and model evaluation. Usually this phase has multiple steps and at each of them we are trying to adjust models and parameters to find the best matching knowledge.

5

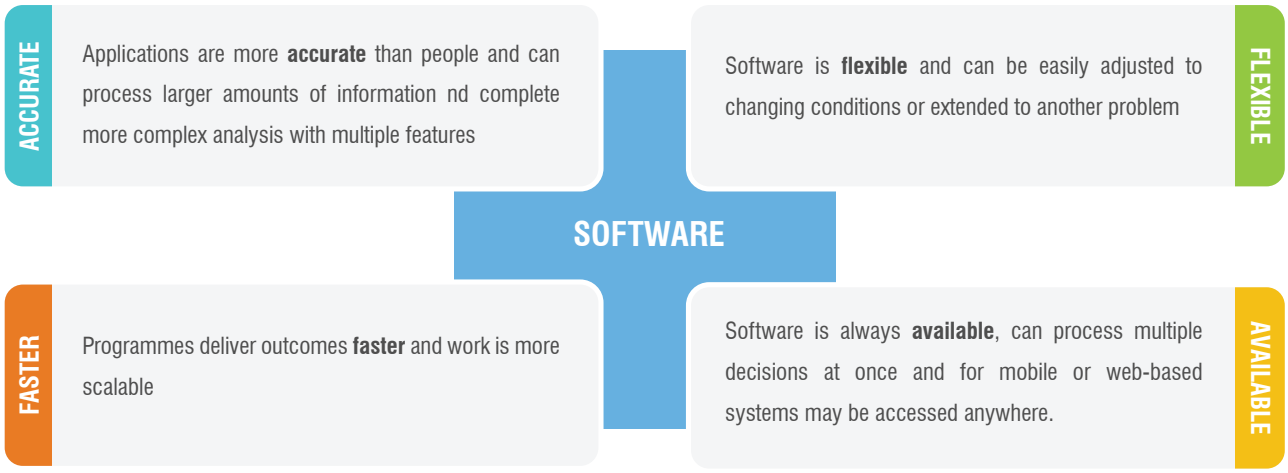
Evaluation gives an answer to the fundamental question “Have we achieved the results we expected?” Therefore, we analyse performed tasks and look for potential errors and any simplifications or improvements that could be introduced. Having the evaluated results we can prepare them for the final step, in which they will be applied.

6

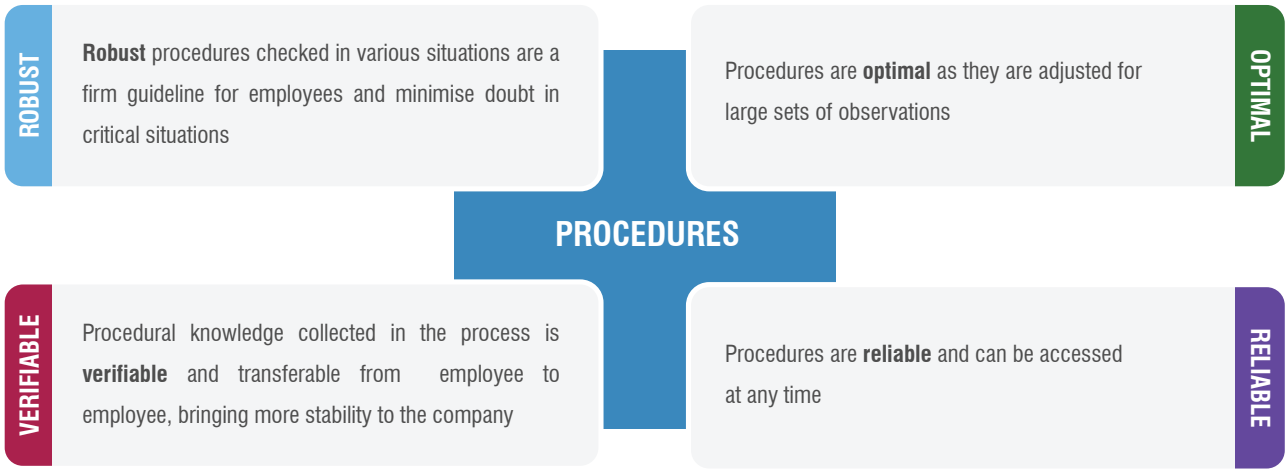
Deployment applies the results of previous steps and involves the creation of decision supporting software or reports allowing unskilled employees to use newly generated knowledge and defined processes.

How to implement data mining in your business?

First of all, we should identify the benefits of Data Mining we want to achieve in the business:



There are also organizational benefits influencing the procedures improved by the knowledge gained:



Case studies

To find out more about Data Mining let's take a look at three situations we have dealt with.

USER ANALYSIS

The increasing use of web-based systems management involves observation of users' behaviour. This activity is essential to maintain the high quality of services and identify potential problems caused by software or users. However, there are several problems with this task – size of web logs, unclear usage patterns and web log structures not supporting meaningful and easy observation.

The solution created used clustering techniques which were deployed to find similar patterns of usage and to create profiles for different groups of users. The method developed was easy to use, cost effective and provided clear signals to be used by managerial staff.

FINANCIAL EVALUATION

Financial evaluation is one of the fundamental financial activities. It constitutes a basis for credit scoring, long-term stock market investments or the verification of business partners. The main problem relates to the skills required to perform this task, so it can only be done by costly financial experts.

Data Mining was able to solve this problem by analysing financial data for various companies to create evaluation models. Using data from standard financial statements, they performed estimations of financial health and industry benchmarking. Incorporated in the software, these models could then be used by all users.

CUSTOMER PROFILING

All sellers try to divide their clients into various groups with particular requirements, each having expectations that must be dealt with differently. One example of such an application is in insurance. Operators must divide clients according to their customers' profile including their insurance risk. Having such capability, agents can propose appropriate insurance policies and calculate their costs more precisely.

The analysis done for this situation revealed different useful patterns that could be used to divide clients beyond the standard rules used by insurance companies. With more accurate methods of classification it is possible to offer better insurance products tailored to customers and to reduce potential losses generated by the inaccurate identification of risk groups.

IF YOU ARE NOW READY TO TAKE ADVANTAGE OF DATA MINING IN YOUR BUSINESS - CONTACT SDART LTD.