Data Mining and Marketing Intelligence

*Alberto Saccardi**

Abstract

The technological advance has made possible to create data bases designed for the marketing intelligence, with the availability of large information.

Data Mining is the process of selection, research and modeling of large amounts of data aimed to find out unknown business relationships in order to achieve a business edge.

The Data Mining activity has a searching approach in the data analysis: discover unknown relationships, new relationships without an 'a priori' hypotheses on the subject.

1. Data Mining: a Simple Neologism or an Efficient Approach for the Marketing Intelligence?

The streamlining of a marketing campaign, the creation of new products or services, the set-up of new branches, the design of a cross-selling activity are generally issues tackled by using the experience in the market as well as mistakes made in the past. By contrast, the technological advance has made possible to create Data Bases designed for the Marketing Intelligence, so the availability of large information raises a few questions, e.g.:

- which are the guidelines that enable us to carry out effectively quantitative analyses focused on business matters?
- Which are the characteristics and what is the expertise needed to infer relevant information from the large company Data Bases?
- In what way can an appropriate methodology, supported by an adequate "tool box", give rise to remarkable economic benefits?

Data Mining, the process of selection, research and modeling of large amounts of data aimed to find out unknown business relationships in order to achieve a business edge, is a satisfactory answer to the above issues.

---

* Professor of Marketing Quantitative Methods, L. Bocconi University-Milan (alberto.saccardi@nunatac.it)

2. Query, Reporting, Data Mining

In recent years, the expression Data Mining has been sometimes used for other types of analyses, like Data Retrieval or OLAP. These analyses, however, are different both for their implementation and their purposes. The Data Retrieval is a simple query activity that enables to get detailed answers with regard to detailed requests: who and how many are, for instance, the clients between 40 and 50 years of age who have subscribed the XY policy in the first quarter of 2005. The Data Mining activity, on the other hand, has the objective to discover unknown relationships aimed to access information like, for instance, who are the customers more likely to purchase the XY policy. We could say that the Data Mining activity has a searching approach in the data analysis: discover new relationships without an "a priori" hypotheses on the subject. The Data Retrieval has, instead, a simply assess-approach. In addition, the Data Mining activity cannot be confused with the multidimensional reporting. As a matter of fact, Data Mining enables to go well beyond the visualization of simple statistics provided by OLAP applications, by carrying out statistical models consistent with business activity.

3. How Appropriately to Implement a Data Mining Project

An element that typifies the Data Mining activity is putting the results into production: on the one hand the business expertise, on the other hand the creation of rules and their application in the operational systems are the necessary tools to set up a decision-making engine. The results achieved through the activity of analysis, therefore, are not just written reports on specific issues; but they are operative regulations designed to select a specific target which may be used by a marketing campaign manager or they can be rules used to compute the insolvency score of a bank loan or the churn rate of a telephone contact.

Such rules are the outputs of projects aimed at business activities. Let's now have a look at the main issues which we may come across in the implementation of these projects. Specifically, we are going to address the following issues:
- settlement of the business problem;
- modelling of Data Bases;
- choice of the instruments;
- line-up of the work team.

3.1 Settlement of the Business Problem

The good outcome of a Data Mining project consists of turning a business matter into an issue of quantitative analysis. For instance, one of the most common requests in marketing field is choosing the best recipients of a commercial promotion. The choice of such recipients cannot always be carried out by using a simple query led by experience or by product knowledge. The application of Data Retrieval criteria sometimes brings about the choice of too large a number of targets in considering the budget available for the campaign, especially if the contact with each client turns out to be quite expensive. In such cases, it is

necessary to sort out the clients so as to identify subsets with different degree of likelihood to join a commercial activity or to purchase specific products or services. Such a classification is implemented through a Scoring Model giving each client the probability to accept the commercial promotion or not, to buy the product-service or not. Therefore it is possible to sort out the population of potential recipients of a marketing campaign based on estimated degree of likelihood.

Once the population is classified, the problem arising is about how many clients should be contacted. The choice of an exact number of recipients is crucial to the objectives of the marketing campaign which may be the revenue maximization, the achievement of the break even point, as well as the maximization of profitable contacts. In order to make such a choice we need to take into consideration also the economic data such as:

- the costs sustained in relation to the campaign structure,
- the costs for each client contacted,
- the revenues expected from the product-service sold in the campaign.

By so doing we can simulate the outcome of the campaign and build up the "Campaign Chart": the line which matches the expected revenue with the number of the recipients to be contacted. Below we show an example of Campaign Chart in which the 12% of clients with the highest degree of purchase likelihood ensures the highest possible profit, the extension of the campaign to the 27% of the population would enable to reach the break even point, while a further extension of the 35% of population would lead us into an area of expected economic loss.

Figure 1: *Campaign Chart*
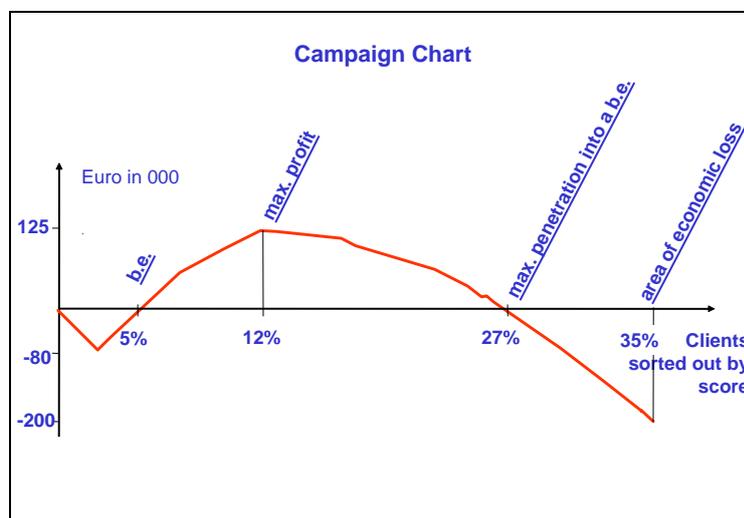
3.2 Modelling of Data Bases

Those who have experienced Data Mining projects are well aware of how crucial it is to have a data environment focused on the activity of quantitative analysis. In particular, the set-up and use of a Customer Database (CDB), regularly updated with information regarding customers and prospects, plays a major role in order to successfully carry out the Data Mining activity. The CDB is the place where the

clients are identified and where the significant data for the next stages of the analysis are stored. In other words the CDB is the necessary tool to count and assess the basic indicators which are put into the analysis processes of the Data Mining.

The logical and conceptual design of the CDB must reflect the complexity of the customers-company relationship. To make it simple, a CDB structure may be represented in two levels:

- Level 1: subject oriented Detail Tables (facts for dimensions) with the highest possible amount of information. The Query-Report tools can refer to these Detail Tables in order to select and/or visualize information multidimensionally. In addition, the processes of creation and updating of the Customer Tables refer to the above Detail Tables, too.
- Level 2: subject oriented table sorted out by customers (Customer Table: one record for each customer), which enable to measure all the dimensions relevant for the quantitative analyses. In other words they are the matrix of basic data used in the Data Mining processes.

The shift from the Detail tables to the Customer Table involves taking decisions on how to sum up the detailed data, and this is a highly significant aspect for the effectiveness of the forthcoming analyses.

**Figure 2:** *Operative Environment for Data Mining Activity*
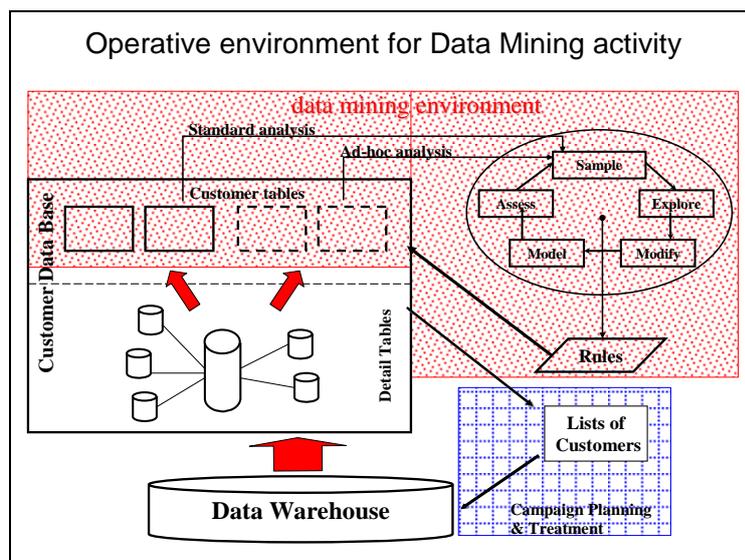
3.3 Choice of the Instruments

An instrument of Data Mining must meet specific requirements for its use:

- *Power of calculation*: it must be an instrument capable of operating on multi-platformed environments, with a client/server frame, to be able to quickly solve algorithms of calculus and implement patterns of analysis;
- *Exploring analysis*: it must enable the analyst to achieve descriptive statistics on the variables of analysis (i.e. mean, median, mode, quantiles of the distribution, frequency distribution, contingency tables) so as to find out in advance potential relationships and outliers in the data;

- *Preparation of data*: tools must be present for the calculation of new variables, for the replacement of the missing values, for the identification and possible treatment of the outliers to enable the analyst to build up the final data base on which the modeling should be carried out;
- *Variety of techniques*: it must enable the analyst to have a wide range of analysis instruments in order to choose the best solution both in terms of statistical performances and in terms of fulfilling the company objectives;
- *Comparison*: it must provide assessment indexes suitable for the different models created in order to enable the analyst to carry out a quick comparison of the various performances; it is especially useful to compare graphically the performances through lift chart, response threshold chart, roc chart;
- *Generalization*: the build-up of the models must be carried out according to validation approaches of the results such to ensure the generalization capacity of the models;
- *Production environments*: it must be capable of interacting with the company CDB and must enable the transfer of decisional rules in a production environment conducive to the implementation of the whole process time and again.

Figure 3: *An Example: SAS Enterprise Miner*
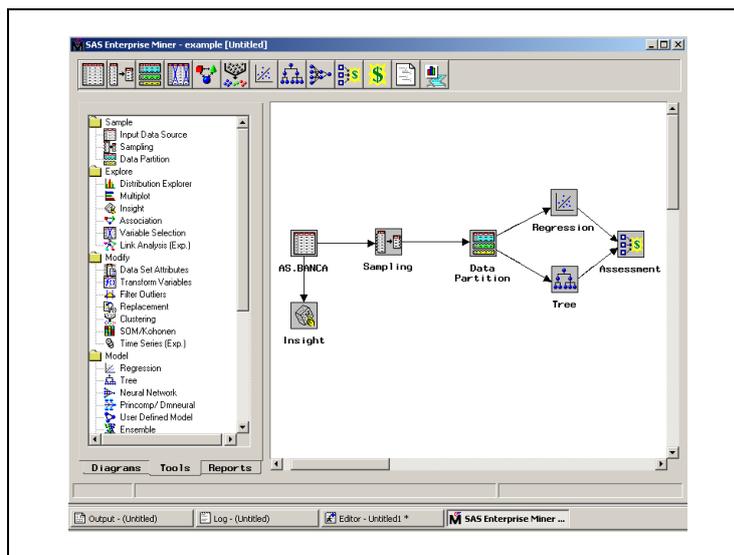
The software solutions nowadays in the market can meet the above mentioned features. In particular, the most effective solutions are the ones provided by SPSS with Clementine and by SAS with SAS Enterprise Miner.

3.4 Line-Up of the Work Team

A good outcome of the project implies the set-up of a work team made of people working in different areas which involves the following aspects:
- sponsor of the project;
- business expert;
- data environment expert;

- data analysis techniques expert.

The sponsor of the project must be of high standards: a Data Mining project may affect different areas of the company and may also have a significant impact on its strategies. Therefore the company sponsorship is needed to achieve the objective.

The other members of the team have a more operative task which can be described as follows:

- a proper use both of the IT instruments supporting the project and of the implementation of those rules produced by data analysis;
- set-up of the statistical models;
- interpretation of these models and their application to the business;
- marketing campaign planning;
- set-up of the operative processes to carry out the campaigns;
- set-up of the operative processes to assess the campaigns' outcome.

4. Data Mining Applications

We have described Data Mining as a process in which several statistical techniques may be used to produce through the data information supporting the business activity: the choice of which techniques during the analysis essentially depends on the nature of the problem being handled (dependence analysis – interdependence analysis) as well as on the type of data available for the analysis (quantitative data – qualitative data). Among the most common applications are:

*Customer Profiling (Segmentation)***:** application of clustering techniques in order to identify homogeneous groups in terms of behavior and social demographical features. The identification of the various typologies of customers enable to carry out direct marketing campaigns and assess their implications as well as to achieve valuable information as to how to modify the commercial offer. This will make it possible to monitor the evolution of one's customers in time and the possible presence of new typologies.

Market Basket Analysis (Affinity Analysis): application of statistical techniques to sales data in order to know which products are purchased along with others. This type of information allows to improve the offer of the products (shelf layout) as well as to increase the sales of some products through specific offers on other products related to them.

Scoring System (Predictive Modelling): integrated system of procedures aimed to set up a model which may link features of target population (list of prospects) with an objective variable. Such variable measures the acceptance of a commercial offer, while the explanatory variables measure the behavior of target population individuals with regard to similar initiatives in the past. The purpose of such technique is to be able to give a score to each individual of the target population. This score will be related to how likely the individual is to accept the commercial offer. By so doing we can sort out the customers based on their score and eventually select those who have turned out to be the most likely to accept the initiative.

Text Mining (Segmentation): Application of clustering techniques in order to identify homogeneous groups of documents and papers related to the

words/subjects found in them. This application enables to quickly assess the chosen subjects and identify possible links with others.

It is common wisdom that Data Mining uses statistical techniques related only to large numbers: this phrase is way too vague if you think that the above techniques have been developed in psychology, medicine and biology; in those fields the width of sample surveyed is in the order of some dozens or at most some hundreds. Therefore even though Data Mining is being widely used nowadays in those business sectors where companies have a large number of customers like the telecoms and big bank and insurance groups, this does not mean that the analyses carried out on some geographical areas being potentially more interested in what is offered could be quite helpful even for those companies of a smaller size.

In conclusion, we can again emphasize that Data Mining has a wide range of applications and that the development of the software instruments and the increasingly big availability of data will further expand the applications in potentially new business areas.

Bibliography

Berry M.J.A., Linoff G., *Data Mining Techniques*, Wiley & Sons, New York, 1997.

Camillo F., Tassinari G., *Data Mining, Web Mining e CRM : metodologie, soluzioni e prospettive*, Franco Angeli, Milan, 2002.

Del Ciello N., Dulli S., Saccardi A., *Metodi di Data Mining per il Customer Relationship Management*, Franco Angeli, Milan, 2001.

Dulli S., Polpettini P., Trotta M., *Text Mining: Teoria e Applicazioni*, Franco Angeli, Milan, 2004

Hand D., Mannila H., Smyth P., *Principles of Data Mining*, MIT Press, Massachusetts, Cambridge, 2001.

Inmon W.H., *Building the Data Warehouse*, Wiley & Sons, New York, 1996.

Molteni L., Troilo G., *Ricerche di Marketing*, McGraw-Hill, Milan, 2003.