Business Intelligence

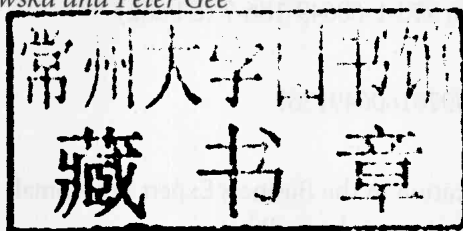## Making Decisions through Data Analytics

Jerzy Surma

Business Intelligence

Making Decisions
Through Data Analytics

Jerzy Surma

translated by

Magdalena Górniakowska and Peter Gee

businessexpert
Press

Cover design by Jonathan Pennell
Interior design by Scribe Inc.

Abstract

This book is about using business intelligence as a management information system for supporting managerial decision making. It concentrates mainly on practical business issues and demonstrates how to apply data warehousing and data analytics to support decision making. This book progresses through a logical sequence, starting with data model infrastructure, then data preparation, followed by data analysis, integration, knowledge discovery, and finally the actual use of discovered knowledge. All examples are based on the newest achievements in business intelligence (BI). Finally, this book outlines an overview of a methodology that takes into account the complexity of developing applications in an integrated BI environment.

Keywords

Contents

Introduction

At present, information technology (IT) is applied in so many business fields that it seems to be the "sine qua non" of development in the modern economy. It is fascinating that IT appeared in areas directly connected with the intellectual achievements of human beings. More surprisingly, it all happened when the most optimistic of research projects, artificial intelligence, spectacularly collapsed. The realistic approach, in which a computer program is designed not to replace a man but to support his activities based on intelligence, turned out to be extremely fruitful and was successfully employed in the business world. A few years later, this field was called business intelligence (BI), as it refers to the implementation of IT systems, which simulate intelligent behavior, to support business decisions.

This book is an account of my long-standing experience in building solutions that support intelligent systems in terms of both research and real business applications. When it comes to research, it all began during my studies at the Wroclaw Technical University, where at the Institute of Control Systems in the late 1980s the theory of expert systems was being formed. This was a turbulent yet creative moment in Poland's history—that is, a few years before the fall of the communism—and as a student I was lucky enough to meet on my road eminent professors and other people committed to an independent Poland. Let me mention my outstanding mathematics professors Hanna Pidek-Lopuszanska, Andrzej Kisielewicz, Roman Rozanski, and Andrzej Zarach, who as heirs of the Lwow-Warsaw school of logic shared this spirit with me.

Finally, I wrote my PhD dissertation on the application of expert systems in management, which I defended at the Wroclaw University of Economics, where one of the first teams dealing with the business aspects of artificial intelligence was created. Later, I spent a few years doing research into data-mining methods and case-based reasoning in business

in research institutes in France and Belgium. After returning to Poland, I worked for international consulting firms and I was in charge of a large number of BI implementations in leading Polish enterprises. Currently, I am a research worker at the Collegium of Business Administration at the Warsaw School of Economics and the director of postgraduate BI studies. In this book I give my original presentation of the subject. I have selected the most significant aspects and I have presented them in the context of real business applications and their real influence on the enhancement of a company's value.

In the first chapter the history of BI is presented, the terms are defined, and the representative applications in business are reviewed. The second chapter describes a warehouse, which is an elementary component of BI systems. I am convinced that insight into the technological dimension translates into a better understanding of both the limitations of those systems and business owners' realistic expectations. The third chapter introduces analytical data mining and describes basic analytical operations with a simple example. Section 3.1 deals with managerial information and its application in various areas in enterprise. The use of BI tools in the management of business processes and a company's strategy by means of a balanced scorecard are discussed separately. The subsequent section, 3.2, is devoted to the problem of strategic information management, or the analysis of information in an enterprise environment, which influences the strategic development of a given company. The whole of the fourth chapter deals with data exploration. Four classic tasks are discussed at length: classification, estimation, discovery of association rules, and cluster analysis. All of these are conducted by means of advanced data analysis algorithms. Limitations of data exploration methods are analyzed in detail, and standard applications are reviewed. The fifth chapter presents issues connected with the analysis of personal data and its application in direct marketing. It also attempts to show the development of this field and its revolutionary influence on the daily life of potential customers. The last chapter summarizes the previously discussed issues. It presents a model based on the concept of a strategy map, which shows relationships between the application of BI systems and potential enhancement of a company's value. The possible ways of connecting BI implementation with a company's business processes, conditions for the success of such projects, and potential threats and sources of failure are also discussed.

Finally, in the conclusion I reflect on the trends of BI development. The majority of the issues are illustrated with case studies of a hypothetical chain of ALFA stores. A chain store is a typical area of BI applications and gives a comprehensive picture of the subject. Each key issue is followed by a list of recommended reading and websites at the end of every chapter. I carefully selected every single recommendation so that readers could further explore the subjects of their interest. The recommendation list is of particular importance in the case of data-mining methods, where I purposefully refrain from mathematical formalism and make an attempt to support an intuitive understanding for the reader of the discussed methods. Still, I would like to encourage readers to familiarize themselves with the more formal attitude toward the analytical methods. All the examples of real BI tools that I give in this book are purely illustrative and as such should not be treated as recommendations of any kind.

This book is an introduction to the subject of BI and can potentially be used

- by students who study business administration,
- as supplementary reading for students in technical universities,
- by PhD and postgraduate students of economics and management,
- by researchers who specialize in business application of IT technologies.

It might also be useful for

- managers and members of the board of directors,
- consultants and business advisors,
- anyone who has a passion for practical applications of artificial intelligence methods.

Many thanks to my colleagues from the faculty of business administration in the Warsaw School of Economics for our discussions and their support, which contributed to the creation of this book. I also wish to express my gratitude toward the students who wrote their master's theses under my supervision for their assistance in designing the outline of a warehouse and demonstration reports for the case studies. I would like

to particularly thank my wife, Ewa, and my son, Grześ, for the warmth that is always there regardless of the weather. Many thanks for your understanding when I annexed our kitchen table for a few months while engrossed in my work on this book.

Enjoy the book and feel free to send your comments to jerzy.surma @gmail.com.

Warsaw-Cambridge-Tyniec, summer 2010

CHAPTER 1

An Introduction to Business Intelligence

1.1. The Origins of Business Intelligence

During the 1970s Herbert Simon, a Nobel Prize winner in economics, was developing his world-famous concept of bounded rationality. He was certainly inspired by an interest in examining the cognitive boundaries of a man following the disappointing results of his own trials concerning the computer stimulation of human decision-making processes.[1] However, the intensive development of computer technologies gave him great hope for building systems to support human activities related to thinking and rational behavior. Those hopes were at least partly fulfilled thanks to the development of business intelligence (BI)—that is, a system that supports managerial decision making in enterprise management in the broadest sense of this word. A distinctive feature of BI is its powerful pragmatism: Out of the broad spectrum of technologies, only those that can be applied to business are selected. The four fundamental sources of information for BI and its tools are as follows:

1. *Statistics and econometrics*, including inter alia statistical theories of pattern recognition, econometric methods, statistical reasoning, and forecasting techniques
2. *Operations research*, including inter alia linear programming, decision theory, and game theory
3. *Artificial intelligence*, including inter alia heuristic search methods, machine learning, expert systems, genetic algorithms, artificial neural networks, and case-based reasoning systems

4. *Database technologies,* including inter alia data modeling, query languages, query optimization, and indexing methods

The application of computers in statistics and operations research resulted in the creation of so-called decision support systems. These systems required the application of formal mathematical models and were mainly based on quantitative data. Simultaneously, they naturally reduced the areas of use and application of computers in modeling real decision-making problems.

Artificial intelligence (AI) faced a much more challenging task. Work on AI was initiated in the 1950s, but to date it has not been crowned with any spectacular success. It was then a lesson in humility for the academic environment and proved how complex and refined human intelligence is. Nevertheless, thanks to those attempts numerous algorithms for supporting real decision-making processes were worked out that go far beyond the capabilities of decision support systems.

Although both AI and decision support systems had a large influence on BI, it was the development of database technologies initiated in the 1960s that was the most important. The development of databases based on the relational data model that allows for the relatively simple interpretation of business data and structured query language (SQL), which was easy to use for those times, proved to be particularly important. Progress in database technologies led to a boom in business applications of enterprise resource planning (ERP), which allows the standard processes of an enterprise to be automatic and well arranged. These transaction data (e.g., expense entering, invoice registering, and recording of bank account transactions or phone call records in a billing system) were soon discovered to be a source of interesting insights into the activities of a business. First, transaction data was aggregated into various reports, which were generated by decision support systems usually by means of SQL. Then techniques derived from decision support systems and artificial intelligence, which could conduct more sophisticated analyses, started to be applied. Those activities, performed at the turn of the 1970s and 1980s, were occasional and originally developed in two areas—namely, in shopping malls[2] and telecommunications companies.

1.2. BI as an Autonomous Discipline

In the 1980s, business applications became so advanced that a separate discipline of designing and creating databases for business decision support emerged. So-called data warehouses[3] and specialist toolsets appeared. Although the term "business intelligence" was first used in 1958 in a paper for *IBM Journal*,[4] a new sense was imparted to it by Howard Dresner from the Gartner company in 1988. Having analyzed the information technology (IT) market, he referred to business intelligence as to a kind of "umbrella" that covers numerous methods, technologies, and applications oriented to real business decision support in an enterprise:[5]

> *Business Intelligence* is a user-oriented process of gathering, exploring, interpreting and analyzing of data, which leads to the streamlining and rationalization of the decision-making process. Those systems support managers in business decision-making in order to create economy value growth of an enterprise.

Such a definition explicitly points out that BI is an IT management system and, strictly speaking, a third-generation IT management system.[6] In light of such an understanding of decision support systems, they encompass a broad spectrum of technologies, including the following:

- *Online analytical processing (OLAP) tools.* Software for multidimensional analysis of business data by integration, aggregation, and adequate mode of presentation and visualization of different data
- *Data-mining tools.* Algorithms for automatic analysis of great volumes of data using statistical and econometric methods, as well as machine learning methods that can analyze not only quantitative but also qualitative data
- *Knowledge management tools.* Tools that allow for storage, indexing, and analysis of textual documents and their further linkage with other data

This class of technological systems is based on the data collected by data warehouses—that is, database systems that gather data from various sources and make it readily available to businesses.

In the 1990s, BI became a widely known term among specialists, and on the level of tools, it was a standard offered not only by specialist companies but also by the greatest software manufacturers to enterprises, such as IBM, Microsoft, Oracle, or SAP.

1.3. BI and Company Management

At the beginning of the 21st century, IT technologies were developing extremely rapidly due to sudden Internet developments. Despite the almost total computerization of fundamental business processes, managers still have a fragmentary knowledge of their own businesses and often make decisions intuitively. Simultaneously, confusion caused by the excess of available data and a lack of its organization can be observed. Moreover, errors in data, lack of cohesion, and having a few versions of "the truth" in an enterprise have also led to a lack of trust in the gathered data. All of these factors aided the increased use of BI by enterprises. There are some sectors that already cannot do without such solutions—for instance, telecommunications and banking. The belief that analytical technologies are key tools to gaining a competitive advantage is also clearly visible.[7] Generally speaking, the successful development of BI contributed to the fulfillment of Herbert Simon's vision. According to his theories, the tasks managers deal with fall into three categories:[8]

1. *Supervising* the standard activities connected with the management of business processes and subordinates

2. *Solving well-structured tasks (problems),* that is, programmable decisions that are routine and repeatable and for which strict procedures have been worked out: For these tasks (e.g., establishing a selling price when logistics costs and purchase price are familiar), every single case does not have to be considered individually

3. *Solving ill-structured tasks (problems)*—that is, nonprogrammable decisions that don't have a cut-and-dried answer—related to new cases in which no pattern of behavior is established, results are

unknown, and there is also no ready-made solution (e.g., a strategic decision about starting up manufacturing abroad)

It should be stressed that at the moment BI is only applied to the first and the second category. The third category is definitely the most interesting one. At present, trials are being conducted in the application of BI methods as solutions for this type of problem. In order to understand the complexity of this issue, let us look at Table 1.1, which describes well- and ill-structured problems.

In formal terms, managerial tasks are decision processes, by which a decision means the selection of one possibility from a set of possible solutions. A BI system can generally support managerial decisions in the following ways (see Figure 1.1):

1. *Providing a decision maker with some information.* This mode encompasses the preparation of adequate information: business reports and outcome from complex analysis.
2. *Proposing managerial decisions.* This approach also includes the possibility of a system making decisions itself.[9]

Making decisions in supervision activities and solving well-structured problems can be supported by providing information and by proposing managerial solutions, while solving ill-structured problems might be supported by BI by their rationalization, that is, by providing the management board of an enterprise with suitable information (see chapter 3). However, proposing managerial solutions for ill-structured problems is not the subject of BI systems' activities.

Table 1.1. Well- and Ill-Structured Problems

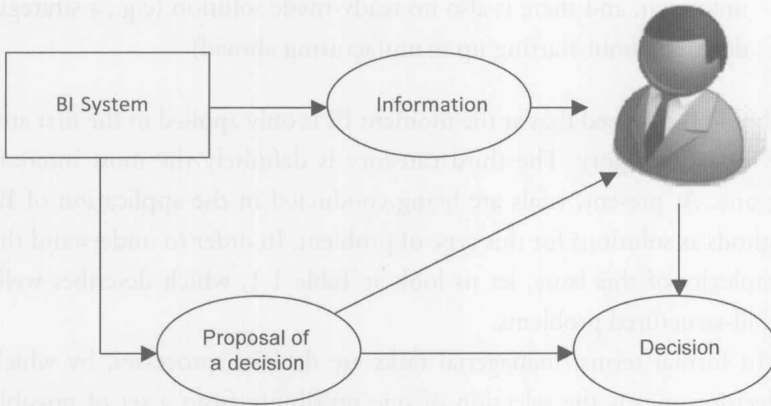|	**Well-structured problems**	**Ill-structured problems**
Data	Quantitative, specific	Qualitative, unspecific
Knowledge	Mathematic model, algorithm	Experience, heuristics, intuition
Problem solving	• Procedure • Independence from context • Clearly specified, objective goals	• Ad hoc • Dependence from context • Vague, subjective goals

Source: Based on Turner (1988).

Figure 1.1. Support of managerial decisions by BI system.

Source: author.

1.4. Review of BI Applications in Some Business Areas

Applications of BI in four selected business areas are reviewed in Table 1.2. Each example is backed up by managerial information, which is

Table 1.2. Review of Some BI Applications

Area	Information	Insight	Decision
Customer management	Customer's lifetime value determined by the services provided to the customer	Analysis of the impact of the product portfolio and extra services on a customer's value	Reduction of additional services offered to the least valuable customers in order to reduce their service costs
Sales	History of sales expressed quantitatively and qualitatively according to product groups	Research into the impact of change in product price on the volume of sales	Correction of the prices of specific groups of products
Finance	Unit costs of product manufacturing connected to sales results	Analysis of profitability to specify the least profitable products	Modification of manufacturing process in order to reduce manufacturing costs
Logistics	Information on shipping routes connected to transport costs	Optimization of the routes to minimize transport costs	Modification of the procedure of route and transport selection

Source: author.

given to a decision maker, and solutions, which can be suggested on the basis of the information. Every decision is justified by analyses, which can be conducted in order to obtain certain recommendations.

Case Study 1.1 shows the use of BI concepts to fulfill a company's strategic initiatives. Subsequently, the second chapter of the book discusses the technological aspects connected with building BI systems.

Case Study 1.1

The chain store ALFA (a fictional store) is an important player in the market of discount stores with an ambition to become the leader in its business line. At present, its chain consists of several dozen stores spread across the northeastern United States. ALFA stores have identical exteriors, similar usable floor space, and an almost identical assortment of products: a few thousand indexes of groceries, cosmetics, tobacco, alcoholic beverages, newspapers, and so on. These stores are stocked from a logistics center (central warehouse) located in the center of the sales area (see Figure 1.2). The average customer is a middle-class person who prefers buying high-quality products or a great variety of products at a low price. The ALFA chain has a very effective procurement system, which has built a stable group of suppliers offering relatively low prices, and a selected assortment of ALFA's own brand of products. The chain competes against a few similar chain discount stores and shopping malls by offering low prices and an almost all-day availability of standard product offerings.

ALFA's management board is completely aware that because there is a limited diversification of products, it must focus on low prices and tight cost management to hold on to its competitive advantage. The fundamental strategic aim is to develop more quickly than its competitors (increase in the market share) and to improve its cost position. To realize its strategy, the board has formulated the following strategic initiatives (projects):

1. An intensive growth in market share by the rational selection of new locations and the fast opening of new stores in the chain
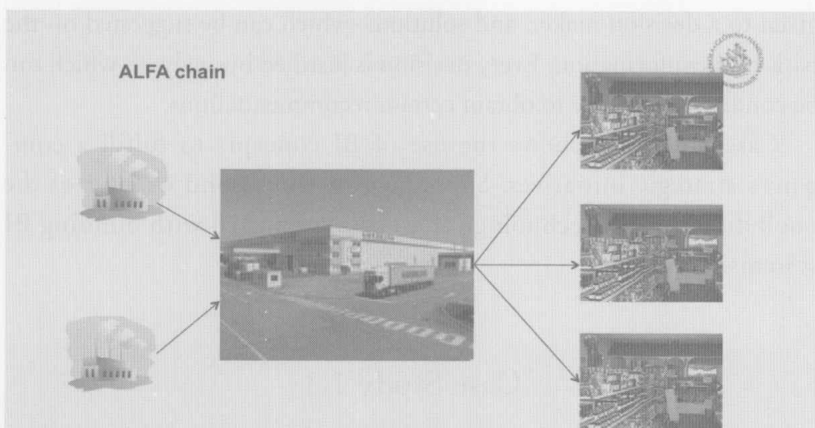
Figure 1.2. The ALFA store chain network.

Source: author.

2. An increase in the average sales per store in comparison with competitors (benchmarking) by the selection of suitable product offerings, proper arrangement of products in the stores, and a quick response to changes in demand for individual products
3. Optimal cost management, mainly by a reduction of logistics costs

In taking steps to implement these strategic initiatives, the board was made aware that its management is based on intuition ("following one's nose") and that there is no suitable information about the chain store's performance, which results in an inability to manage it rationally. The following points were discovered:

1. The new store locations were selected in most cases intuitively without taking into consideration knowledge of how chain stores function in similar locations and available demographic data, environmental data, and so on.
2. The board and regional managers get the sales results from all the stores for individual products weekly. A weekly interval in the observation of sales of certain products makes it impossible to respond effectively to radical change in demand in comparison with projections.
3. The managers of the individual stores independently and intuitively arrange products on shelves and use floor space.