Preface

Business Intelligence (BI) is one area of the Decision Support Systems (DSS) discipline and refers to information systems aimed at integrating structured and unstructured data in order to convert it into useful information and knowledge, upon which business managers can make more informed and consequently better decisions. Being rooted in the DSS discipline, BI has gone through a considerable evolution over the last years and is, nowadays, an area of DSS that attracts a great deal of interest from both the industry and researchers. A BI system is a particular type of system. One of the main aspects is that of user-friendly tools, which make systems truly available to the final business user.

The term Knowledge Discovery in Databases (KDD) was coined in 1989 to refer to the broad process of finding knowledge in data, and to emphasize the “high-level” application of particular Data Mining (DM) methods (Fayyad, Piatetski-Shapiro, & Smyth, 1996). The DM phase concerns, mainly, the means by which patterns are extracted and enumerated from data. DM has several successful applications in many diversified fields.

DM is being implemented with success in BI and several examples of applications can be found. Despite efforts being made for the integration of DM in BI systems, DM has not yet reached non-specialized users, and thus, it is not yet completely integrated with BI. Powerful analytical tools, such as DM, remain too complex and sophisticated for the average consumer of BI systems. McKnight supports that bringing DM to the front line business personnel will increase their potential to attaining BI’s high potential business value (McKnight, 2002). Another fundamental issue that is pointed out by McKnight is the capability of DM tools to be interactive, visual, and understandable, to work directly on the data, and to be used by frontline workers for intermediate and lasting business benefits. Currently, DM systems are functioning as separate isles, and hereby it is considered that only the full integration of the KDD process on BI can conduct to an effective usage of DM in BI (Azevedo & Santos, 2011). This book presents several interesting applications, for instance in healthcare systems, in finance, as well as Web and text-mining applications.

Three main reasons can be pointed out for DM to be not completely integrated with BI, each one leading to a specific problem that constraints DM usage in BI. Firstly, the models/patterns obtained from DM are complex and there is the need of an analysis from a DM specialist. This fact can lead to a non-effective adoption of DM in BI, being that DM is not really integrated on most of the implemented BI systems, nowadays. Secondly, the problem with DM is that there is not a user-friendly tool that can be used by decision makers to analyze DM models. Usually, BI systems have user-friendly analytical tools that help decision makers in order to obtain insights on the available data and allow them to take better decisions. Examples of such tools are On-Line Analytical Processing (OLAP) tools, which are widely used. There are not equivalent tools for DM that allow business users to obtain insights in DM models. Finally, but
extremely importantly, sufficient emphasis to the development of solutions that allow the specification of DM problems through business-oriented languages and that are also oriented for BI activities has not been given. With the expansion that has occurred in the application of DM solutions in BI, this is, currently, of increasing importance. BI systems are, usually, built on top of relational databases and diverse types of languages are involved. As a consequence, DM integration with relational databases is an important issue to consider when studying DM integration with BI. Codd’s relational model for database systems (Codd, 1970, 1982) was adopted long ago in organizations. One of the reasons for the great success of relational databases is related with the existence of a standard language: Structured Query Language (SQL). SQL allows business users to obtain quick answers to ad-hoc business questions through queries on the data stored in databases. SQL is nowadays included in all the Relational Database Management Systems (RDBMS). SQL serves as the core above which are constructed the various Graphical User Interfaces (GUI) and user-friendly languages, such as Query-By-Example (QBE), included in RDBMS. It is also necessary to define a standard language, which can operate likewise for DM. Several approaches have been proposed for the definition of DM languages. In the literature there can be found some language specifications, namely DMQL (Han, Fu, Wang, Koperski, & Zaiane, 1996), MINE RULE (Meo, Psaila, & Ceri, 1998), MSQL (Imielinski & Virmani, 1999), SPQL (Bonchi, Giannotti, Lucchesse, Orlando, Perego, & Trasarti, 2007), KDDML (Romei, Ruggieri, & Turini, 2006), XDM (Meo & Psaila, 2006), RDM (De Raedt, 2002), and QMBE (Azevedo & Santos, 2012), among others.

DM integration with BI systems can be tackled from different perspectives. On the one hand, it can be considered that the effective integration of DM with BI systems must involve final business users’ access to DM models. This access is crucial in order to business users to develop an understanding of the models, to help them in decision-making (Azevedo & Santos, 2011, 2012). On the other hand, a different approach can be considered through the outgrowth of new strategies that allow business users and DM specialists developing new communication strategies. Wang and Wang (2008) introduce a model that allows knowledge sharing among business insiders and DM specialists. It is argued that this model can make DM more relevant to BI. This book presents both perspectives.

Some efforts are being made seeking the establishment of standards in the DM area, both by academic and by people in the industry field. Above all, the academic efforts towards a theory for DM and KDD follow closely the theory developed by Codd for the Relational Model. The main goal is to integrate DM with relational databases, thus allowing an easier application of DM to business systems and making it more available to decision making. This also represents an important aspect to the integration of DM with BI systems. An important issue in this domain concerns DM languages, which were already discussed in this preface. Some of the efforts in the industrial field concern the definition of processes/methodologies that can guide the implementation of DM applications. For instance, SEMMA and CRISP-DM can be pointed out as such examples. Other efforts in the industrial field focus on the development of software suites for implementing some selected DM algorithms. There are also some efforts being made that intend to develop standards that will allow model representation to be platform independent. This book includes both academic and industry perspectives.

The primary objective of this book is to provide insights concerning the topic of the integration of DM in business BI. This is a cutting-edge and important topic that deserves a significant approach, and this book is an excellent opportunity to do it. The book aims to provide the opportunity for a reflection on this important issue, increasing the understanding of using DM in the context of BI, providing relevant academic work, empirical research findings, and an overview of this relevant field of study.
Professionals in the area of DM and BI, managers, researchers, academicians, practitioners, and graduate students, are the target of this book.

STRUCTURE OF THE BOOK

The book is constituted by 13 chapters divided into 5 sections. Contributions come from several countries and from academics and people in the industry.

Section 1, “Fundamentals and Literature Review,” includes three chapters. This first section aims to introduce some of the main facets of the area of the book.

In Chapter 1, Ana Azevedo presents a historical comparative perspective of BI and DM. From the analysis of the evolution of the scientific publications in DM, BI, and DM+BI, she establishes the starting point for the main topic of this book, which is the integration of DM in BI systems.

In Chapter 2, Kijpokin Kasenmap presents a useful literature review on the relations between Knowledge Management (KM), DM, and BI. Starting with the introduction of the concepts of DM, BI, and KM, he develops the relations between the three, practical applications, and emergent trends. KM is presented as a key to DM integration in BI systems. In addition, the references and the additional readings lists can be used as a good starting point for those who intend to explore more deeply these subjects, since they constitute a comprehensive list of the main references.

In Chapter 3, Arun Thotapalli presents a literature review of the important topic of Data Quality (DQ) for DM in BI systems. The presentation of the main DQ measures that can be found in the literature, divided in three main categories (objective measures, subjective measures, and semantic measures), functions as the starting point to present the gaps that can be identified in this field of research. A framework for measuring DQ for DM is presented by the author, and the steps for implementation of the framework are recommended.

Section 2, “Approaches and Methodologies,” includes three chapters. This section presents some useful approaches and methodologies that aim to benefit the integration of DM in BI systems.

Mouhib Alnoukari presents, in Chapter 4, a methodology named ASD-BI. That methodology implements the philosophy of the agile methodologies for software development in the context of DM in BI systems. The implementation of BI systems is difficult and takes a long time and effort to be successful. Integrating DM introduces even more challenges and difficulties to this process. Thus, ASD-BI can constitute a step towards an effective integration of DM in BI systems, since it allows a quicker access of the users to useful results.

In Chapter 5, Stanley Lo defines the concept of proactive BI, compared with reactive BI. In this proactive approach, initial hypotheses are not defined, and analysts do not know exactly what to search for. DM plays an important role in this proactive approach to BI, as is stated by the author. Thus, this approach can also favor the integration of DM in BI systems.

Thanachart Ritbumroong, in Chapter 6, presents the philosophy of the OLAM approach to BI with a specific application, which is analyzing customer behavior. OLAM consists in combining OLAP, a very popular and useful BI tool, with DM. OLAM constitutes a very promising tool to the integration of DM in BI systems, since it allows the business users to access DM models at the same level as data available through traditional OLAP tools, which are very successful due to its user-friendliness.
Section 3, “Web and Text Mining Applications,” includes two very interesting chapters of Web and text mining applications, respectively, integrated in BI systems. This section aims to attract attention to these “new” forms of mining, namely Web and text mining.

In Chapter 7, Marcos Aurélio Domingues, Alípio Mário Jorge, Carlos Soares, and Solange Oliveira Rezende present their work in the development of an advanced decision support system, employed to analyze a Website metadata quality, using Web mining techniques in the context of BI systems. An example of an application to an e-news Website is explored, and results are presented and discussed. In addition, a very useful literature review of Web mining for BI is presented, which can help those commencing in this topic.

In Chapter 8, Lipika Dey and Ishan Verma introduce a framework for the integration of structured and unstructured data in BI systems. In this framework, text mining is applied, and the results are included in BI dashboards, conjunctly with other data. An example with practical implications to business is included. Usually, BI systems only deal with structured data, losing important insights that can be brought by the inclusion of unstructured data. Thus, this chapter represents an important topic in this context.

Section 4, “Applications to Specific Domains,” includes three interesting applications of the utilization of DM in BI systems. This section aims to present important applications of DM in the context of BI.

In Chapter 9, Eliana Pereira, Andreia Brandão, Maria Salazar, Carlos Filipe Portela, Manuel Filipe Santos, José Machado, António Abelha, and Jorge Braga present an application of DM in a BI system whose goal is to improve the triage of patients in a maternity/gynecology hospital. Traditional triage systems used in hospitals, like Manchester, do not adjust to this particular case. The decision support system presented in this chapter uses DM in the context of a BI system and helps in the decision-making process of pre-triage in maternity/gynecology, which in this specific case has special characteristics. The results are promising.

In Chapter 10, Eva Silva, Ana Alpuim, Luciana Cardoso, Fernando Marins, César Quintas, Carlos Filipe Portela, Manuel Filipe Santos, José Machado, and António Abelha describe an application whose aim is to help prevent the occurrence of nosocomial infections in an intensive care unit. This is a very critical situation in hospital environments, which leads to several deaths in patients. The authors present the system and show the obtained improvements in a specific hospital. All the five phases of the KDD process are explained in the perspective of a BI system.

Rui Sarmento, Luís Trigo, and Liliana Fonseca, in Chapter 11, apply several DM techniques to predict business bankruptcy, a common application of BI systems. KDD process is used, the best technique is selected, and the complete process is contextualized with BI.

Section 5, “Software Issues,” includes a valuable participation from the industry players. This section aims to include the important point of view of people who are not academic but are concerned with the integration of DM in BI systems. We consider this an important contribution of this book.

Gábor Bakos, in Chapter 12, discusses some issues encountered with the integration of two open source tools, namely RapidMiner with KNIME. The main problems are explained and solutions are presented from a technological perspective.

The main idea of Keith McCormick, Richard Creeth, and Scott Mutchler’s Chapter 13 is “placing predictions at the point of decision.” A case study with suggestive applications is presented.
CONCLUSION

This book presents a reflection about crucial issues concerning the integration of DM in BI systems. In this book, approaches, methodologies, and applications are introduced, presenting important insights. Viewpoints from both the academic and the industry fields are included, which brings a deeper perceptive. All the chapters included represent important contributions to the development of the subject matter.

Nevertheless, we are aware that this work represents just the beginning of a long path. Several research questions are still open, and several important business problems still remain to be resolved by the industry players. The research included in this book brings fundamental issues, which are very important to the development of the field, and represents an introduction to this important area of research.

Ana Azevedo
Algoritmi R&D Center/University of Minho, Portugal & Polytechnic Institute of Porto/ISCAP, Portugal

Manuel Filipe Santos
Algoritmi R&D Center/University of Minho, Portugal

REFERENCES


