

Preface

The main task of business intelligence (BI) is providing decision support for business activities based on empirical information. The term business is understood in a rather broad sense covering activities in different domain applications, for example, an enterprise, a university, or a hospital. In the context of the business under consideration, decision support can be at different levels ranging from the operational support for a specific business activity up to strategic support at the top level of an organization. Consequently, the term BI summarizes a huge set of models and analytical methods such as reporting, data warehousing, data mining, process mining, predictive analytics, organizational mining, or text mining.

In this book, we present fundamental ideas for a unified approach towards BI activities with an emphasis on analytical methods developed in the areas of process analysis and business analytics.

The general framework is developed in Chap. 1, which also gives an overview on the structure of the book. One underlying idea is that all kinds of business activities are understood as a process in time and the analysis of this process can emphasize different perspectives of the process. Three perspectives are distinguished: (1) the production perspective, which relates to the supplier of the business; (2) the customer perspective, which relates to users/consumers of the offered business; and (3) the organizational perspective, which considers issues such as operations in the production perspective or social networks in the customer perspective.

Core elements of BI are data about the business, which refer either to the description of the process or to instances of the process. These data may take different views on the process defined by the following structural characteristics: (1) an event view, which records detailed documentation of certain events; (2) a state view, which monitors the development of certain attributes of process instances over time; and (3) a cross-sectional view, which gives summary information of characteristic attributes for process instances recorded within a certain period of time.

The issues for which decision support is needed are often related to so-called key performance indicators (KPIs) and to the understanding of how they depend on certain influential factors, i.e., specificities of the business. For analytical purposes,

it is necessary to reformulate a KPI in a number of analytical goals. These goals correspond to well-known methods of analysis and can be summarized under the headings business description goals, business prediction goals, and business understanding goals. Typical business description goals are reporting, segmentation (unsupervised learning), and the identification of interesting behavior. Business prediction goals encompass estimation and classification and are known as supervised learning in the context of machine learning. Business understanding goals support stakeholders in understanding their business processes and may consist in process identification and process analysis.

Based on this framework, we develop a method format for BI activities oriented towards ideas of the L* format for process mining and CRISP for business analytics. The main tasks of the format are the business and data understanding task, the data task, the modeling task, the analysis task, and the evaluation and reporting task. These tasks define the structure of the following chapters.

Chapter 2 deals with questions of modeling. A broad range of models occur in BI corresponding to the different business perspectives, a number of possible views on the processes, and manifold analysis goals. Starting from possible ways of understanding the term model, the most frequently used model structures in BI are identified, such as logic-algebraic structures, graph structures, and probabilistic/statistical structures. Each structure is described in terms of its basic properties and notation as well as algorithmic techniques for solving questions within these structures. Background knowledge is assumed about these structures at the level of introductory courses in programs for applied computer science. Additionally, basic considerations about data generation, data quality, and handling temporal aspects are presented.

Chapter 3 elaborates on the data provisioning process, ranging from data collection and extraction to a solid description of concepts and methods for transforming data into analytical data formats necessary for using the data as input for the models in the analysis. The analytical data formats also cover temporal data as used in process analysis.

In Chap. 4, we present basic methods for data description and data visualization that are used in the business and data understanding task as well as in the evaluation and reporting task. Methods for process-oriented data and cross-sectional data are considered. Based on these fundamental techniques, we sketch aspects of interactive and dynamic visualization and reporting.

Chapters 5–8 explain different analytical techniques used for the main analysis goals of supervised learning (prediction and classification), unsupervised learning (clustering), as well as process identification and process analysis. Each chapter is organized in such a way that we first present first an overview of the used terminology and general methodological considerations. Thereafter, frequently used analytical techniques are discussed.

Chapter 5 is devoted to analysis techniques for cross-sectional data, basically traditional data mining techniques. For prediction, different regression techniques are presented. For classification, we consider techniques based on statistical principles, techniques based on trees, and support vector machines. For unsupervised

learning, we consider hierarchical clustering, partitioning methods, and model-based clustering.

Chapter 6 focuses on analysis techniques for data with temporal structure. We start with probabilistic-oriented models in particular, Markov chains and regression-based techniques (event history analysis). The remainder of the chapter considers analysis techniques useful for detecting interesting behavior in processes such as association analysis, sequence mining, and episode mining.

Chapter 7 treats methods for process identification, process performance management, process mining, and process compliance. In Chap. 8, various analysis techniques for problems are elaborated, which look at a business process from different perspectives. The basics of social network analysis, organizational mining, decision point analysis, and text mining are presented. The analysis of these problems combines techniques from the previous chapters.

For explanation of a method, we use demonstration examples on the one hand and more realistic examples based on use cases on the other hand. The latter include the areas of medical applications, higher education, and customer relationship management. These use cases are introduced in Chap. 1. For software solutions, we focus on open source software, mainly R for cross-sectional analysis and ProM for process analysis. A detailed code for the solutions together with instructions on how to install the software can be found on the accompanying website:

www.businessintelligence-fundamentals.com

The presentation tries to avoid too much mathematical formalism. For the derivation of properties of various algorithms, we refer to the corresponding literature. Throughout the text, you will find different types of boxes. Light grey boxes are used for the presentation of the use cases, dark grey boxes for templates that outline the main activities in the different tasks, and white boxes for overview summaries of important facts and basic structures of procedures.

The material presented in the book was used by the authors in a 4-h course on Business Intelligence running for two semesters. In case of shorter courses, one could start with Chaps. 1 and 2, followed by selected topics of Chaps. 3, 5, and 7.

Vienna, Austria
Vienna, Austria

Wilfried Grossmann
Stefanie Rinderle-Ma

<http://www.springer.com/978-3-662-46530-1>

Fundamentals of Business Intelligence

Grossmann, W.; Rinderle-Ma, S.

2015, XVIII, 348 p. 116 illus., 81 illus. in color.,

Hardcover

ISBN: 978-3-662-46530-1