

Chapter 1

Introduction

Abstract In this chapter, we provide definitions of Business Intelligence (BI) and outline the development of BI over time, particularly carving out current questions of BI. Different scenarios of BI applications are considered and business perspectives and views of BI on the business process are identified. Further, the goals and tasks of BI are discussed from a management and analysis point of view and a method format for BI applications is proposed. This format also gives an outline of the book's contents. Finally, examples from different domain areas are introduced which are used for demonstration in later chapters of the book.

1.1 Definition of Business Intelligence

If one looks for a definition of the term Business Intelligence (BI) one will find the first reference already in 1958 in a paper of H.P. Luhn (cf. [14]). Starting from the definition of the terms “Intelligence” as “the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal” and “Business” as “a collection of activities carried on for whatever purpose, be it science, technology, commerce, industry, law, government, defense, et cetera”, he specifies a business intelligence system as “[an] automatic system [that] is being developed to disseminate information to the various sections of any industrial, scientific or government organization.” The main task of Luhn's system was automatic abstracting of documents and delivering this information to appropriate so-called *action points*.

This definition did not come into effect for 30 years, and in 1989 Howard Dresner coined the term Business Intelligence (BI) again. He introduced it as an umbrella term for a set of concepts and methods to improve business decision making, using systems based on facts. Many similar definitions have been given since. In Negash [18], important aspects of BI are emphasized by stating that “...business intelligence systems provide actionable information delivered at the right time, at the right location, and in the right form to assist decision makers.”

Today one can find many different definitions which show that at the top level the intention of BI has not changed so much. For example, in [20] BI is defined as “an integrated, company-specific, IT-based total approach for managerial decision

support” and Wikipedia coins the term BI as “a set of theories, methodologies, processes, architectures, and technologies that transform raw data into meaningful and useful information for business purposes.”

Summarizing the different definitions, BI can be characterized by the following features:

Features of BI

- *Task of BI*: The main task of BI is providing decision support for specific goals defined in the context of business activities in different domain areas taking into account the organizational and institutional framework.
- *Foundation of BI*: BI decision support mainly relies on empirical information based on data. Besides this empirical background, BI also uses different types of knowledge and theories for information generation.
- *Realization of BI*: The decision support has to be realized as a system using the actual capabilities in information and communication technologies (ICT).
- *Delivery of BI*: A BI system has to deliver information at the right time to the right people in an appropriate form.

Corresponding to the development in ICT and availability of data, we can distinguish different epochs in BI. The prehistory of BI mainly runs under the heading decision support systems (DSS) and is documented, for example, in [19]. The review covers the era from the 1960s up to the beginning of the twenty-first century and considers theory development in computer science, optimization, and application domains, as well as systems development like model-driven DSS (planning models or simulation), data-driven DSS (from data bases up to OLAP systems), communication-driven DSS (collaboration networks), document-driven DSS (document retrieval and analysis), and knowledge-driven DSS (expert systems).

According to Howard Dresner’s definition in 1989, the term BI became popular in the 1990s and was understood mostly as data-driven decision support closely connected to the development of data warehouses, the usage of online analytical processing (OLAP), and reporting tools. In parallel to the developments in the area of data management, other analysis tools such as data mining or predictive analytics became popular. Sometimes, these were summarized under the heading *business analytics*, and one got the impression that BI is a collection of a loosely related heterogeneous set of tools supporting different tasks within a business. Hence, it was necessary to consolidate the different lines of development and to focus again on the decision support perspective.

One influential approach putting the data warehouse into the center is the Kimball methodology (cf. [12]). This methodology defines a life cycle for data warehouse solutions with dimensional modeling as the core element. The design of appropriate technical architectures supports the realization of a data warehouse. Applications like reporting and analytical models provide decision makers with the necessary information.

The software life-cycle model as a framework for integration of different aspects of BI is used in [17]. Other approaches like CRISP [4] start from the analysis process in knowledge discovery from databases. Besides such conceptual ideas, one can also frequently find pragmatic definitions, for example, in [6] it is argued that BI should be divided into querying, reporting, OLAP, alert tools, and business analytics. In this definition; business analytics is a subset of BI based on statistics, prediction, and optimization. In the book, we will follow this idea and understand BI in such a broad sense.

In the last years, data availability and analysis capabilities have increased tremendously, and new research areas for BI have emerged. In [22], a number of topics are listed under the heading *Business Intelligence 2.0*. Looking at these topics from the perspective of the four main BI characteristics stated above, one can organize these new challenges as shown in the overview box.

Actual Challenges of BI

- *Tasks of BI:* Nowadays we can find a well-structured understanding of the business logic in almost all domain areas. This new understanding has also led to a process-oriented conceptual view, which integrates workflow considerations and process mining into BI [23]. Another aspect is that new organizational structures like decentralized organizations want to apply decision support within their environment, and, hence, ideas from collective intelligence or crowd sourcing are applied in BI.
- *Foundations of BI:* Besides the traditional data warehouse, we also have to take into account data on the Web. Such data is often not well-structured, but only semistructured such as text data. The need to integrate different data useful for decision support in a coherent way has led to models for linking data in BI. In connection with such new data, the scope of analytical methods has broadened and new tools such as visual mining, text mining, opinion mining, or social network analysis have emerged.
- *Realization of BI systems:* Today's software architectures allow interesting new realizations of BI systems. From a user perspective, Software as a Service (SaaS) constitutes an interesting development for BI systems. From a computational point of view, we have to deal with large and complex data sets nowadays. Moreover, cloud computing and distributed computing are important concepts opening new opportunities for BI applications.
- *Delivery of BI:* Mobile devices offer a new dimension for delivering information to users in real-time. However, these developments have to take into account that quality of real-time information is a new challenge for BI.

Obviously, many of the mentioned new developments cover more than one aspect of the aforementioned BI characteristics, but this classification should support the understanding that the basic definition and characteristics of BI are still valid.

Due to the importance of BI for business applications, there is a big market, and many companies offer BI solutions. These vendors create a lot of terms and acronyms and propose integrated formats for BI applications, but precise and generally accepted definitions of terms are frequently missing in the BI context. For an overview on vendors and tools, we refer to [21].

1.2 Putting Business Intelligence into Context

In the previous section, we characterized BI and stated its goals in a rather general way. In order to make this more precise, we want to discuss first the connection between business and BI from a management point of view. An interesting reference in this context that is worth reading is [13].

We understand the term *business* in a rather broad sense, i.e., as “any kind of activities of an organization for delivering goods or services to consumers.” These organizations may be active in different application domains, for example, an enterprise, an administrative body, a hospital, or an educational institution such as a university. Besides the different application domains, we have to be aware that decision support is needed for businesses of different *size* and *scope*. By size we understand a classification of the organization with respect to criteria such as number of employees (e.g., SMEs or big enterprises), regional dispersion (from local up to global players), number of customers, or revenues. Scope refers to the number of activities of the organization for which we look for decision support. For example in business administration, we may be interested in decision support at the global level for the enterprise or at a specific functional level (e.g., production or marketing). In medical applications, our focus may be decision support for the treatment of a specific disease or for the management of a hospital. In the administrative context, we can look for decision support for efficient organization of services or for improving customer satisfaction with the services.

1.2.1 Business Intelligence Scenarios

For development of a general framework of such diverse problems, we will follow ideas as outlined in [13] which organize BI activities according to principles used in business enterprises. A management level, an organizational level, a functional analytical level, and levels for data organization and acquisition are distinguished, and the role of BI in connection with *business models* is discussed. As in the case of BI, there are many definitions of the term business model (cf. [1]), but for our purpose the following rather naive understanding seems sufficient: A *business*

model reflects the strategy of an enterprise for creating value. There are four different scenarios that link BI to the business context, ranging from rather simple applications of decision support for a specific problem up to BI as an essential part of strategic planning [13].

BI Scenarios

1. *Business intelligence separated from strategic management:* In this case BI is mainly concerned with the achievement of short-term targets in a division of an organization, for example, a department of an enterprise or a clinic in a hospital. Typically, results of the BI application are more or less standardized reports for a dedicated part of the business.
2. *BI supports monitoring of strategy performance:* Such a BI application is motivated by overall strategic goals and formulated in accordance with these goals. Monitoring of the performance is done by defining measurable targets. A data warehouse allowing a unified view onto the business is usually a prerequisite for such an application scenario.
3. *BI feedback on strategy formulation:* This application goes one step beyond the previous strategy and aims at an evaluation of the performance using analytical methods. In the best case, such an application can be used for the optimization of a strategy. A typical end-product in this scenario may be a balanced scorecard.
4. *BI as strategic resource:* This strategy uses the information generated by BI not only for optimization but also as an essential input for the definition of the strategy at the management level. Typical examples are customer-based marketing or development of standard operation procedures for patient treatment.

Obviously, this classification depends on the size of the organization and the scope of the business under consideration. For example, a BI application at a university department may be used as feedback on strategy formulation at the level of the department but also as a tool for monitoring the performance at the university level.

At first glance, the third and fourth strategies seem to be favorable, but in general, we have to take into account specificities of the application, how many resources can be attributed to BI, and the availability of information. For large production-oriented enterprises, the third option may be a good choice, and in service-oriented businesses the fourth strategy has yielded many success stories. But sometimes decision problems occur ad hoc, are hard to formalize, and it is not clear whether implementation of a high-level strategy is worth it in the long run. Moreover, results of such ad hoc applications may lead to standardized new BI activities at a higher strategic level.

1.2.2 Perspectives in Business Intelligence

After the determination of the overall BI strategy, we have to think about the structure of business activities. The description of the structure is frequently done by formulating a *business process*. We understand the term business process as a *collection of related and structured activities necessary for delivering a certain good or service to customers together with possible response activities of customers*.

Note that most definitions of business processes such as [5] omit the last part of the definition. However, we think that understanding the customer as an active decision maker inside the business process is more suited for BI. In the book, generally speaking, we will take the position that all kinds of business activities are processes, which means that activities take place within a period of time and follow some rules such as the partial ordering or the exclusion of an activity under certain conditions. However, we have to be aware that, to some extent, the incorporation of customer activities into the business process limits the application of the idea that business activities resemble the structure of purely rule-based activities. Instead of such a mechanistic consideration of business processes, BI is more concerned with the empirical realization of business process defined by *process instances*. In order to scrutinize these instances, we introduce the following three *BI perspectives* for the business process.

Perspectives in BI

- *Production perspective*: This perspective considers decision support for answering questions such as what kind of products should be offered to the customers and how the production should be operated. This perspective plays an important role for product development and for internal organization of the business.
- *Customer perspective*: This perspective focuses on customer behavior and aims at understanding how customers perceive products or services and how they react to this offer. The customer perspective plays an essential role in service-oriented businesses.
- *Organizational perspective*: This perspective examines the organizational background of the business process. It may refer to the organizational background for the operations in connection with the production perspective or to the influence of social networks on customer behavior.

Obviously, such perspectives depend on the application domain, the size, and the scope of the business. Practical applications usually encompass all three perspectives, but for BI applications such a division is useful for choosing appropriate information and analysis models. To some extent, this division also reflects the historical development of models and analytical methods nowadays applied in BI.

The production perspective usually requires detailed information and data about the internal organization of the business. Typically, the organizational structure of enterprises is specified and maintained in terms of organizational models that consist of organizational entities such as roles, organizational units, and actors. The organizational units are typically linked by different relations. For BI applications, the following roles are of interest¹

Roles in Context of BI

- The first role is the *process owner*, defined as the entity setting the rules governing the process. Traditionally, the process owner is defined from a production perspective, but in service-oriented businesses customers also may be process owners. Think, for example, of patients who decide about their treatment.
- The next step is to identify the *process subjects* as the entities that identify the process instances. In most cases, these process subjects are defined by the customers, but specific products or networks of people involved in the business process are also possible candidates. One can understand the process subjects as the entities triggering the initialization of a process instance by some event. For example, a patient with a certain health problem shows up at the hospital triggering a certain treatment process.
- Besides the process subjects, other people or in organizations can generate events in the process as well. We will denote these entities as *process actors*, or short as *actors*. In business administration; the actors are usually the part of the organization responsible for the production of goods or services.

The customer perspective frequently needs a much simpler view on the internal processes, because customers are usually not aware of the internal organization and only react according to their personal view on the business. On the other hand, for understanding customer behavior we need a more detailed description of personal customer characteristics like sex, age, or social status, together with their organizational embedding into the business process.

An important issue is the interaction between production and customer perspective. This interaction may be rather simple, for example, when a customer decides to buy goods in a shop for some time and quits the business relation afterwards. Other processes may be rather complex and comprise many interactions like negotiations

¹In the business process management literature, different roles or stakeholders within the process life cycle are mentioned. See, for example, Dumas et al. [8] or Weske [25]. In this book, we will explain a selection of these roles and augment them with the roles of the process subjects.

or the usage of multiple services. A typical example from health care is the treatment of patients. Depending on the complexity of the interaction, a specific combination of the different perspectives might be required.

In summary, business processes can be analyzed along three goal perspectives: customer, production, and organization. We aim at discussing BI using all three perspectives throughout the book. In Sect. 1.4, we will show by examples how these perspectives may be realized in various application contexts.

1.2.3 *Business Intelligence Views on Business Processes*

Structural analysis of business processes, i.e., the analysis based on a model describing the business process, is, in many cases, an interesting and useful task. However, as already mentioned, BI applications are more interested in understanding the real-world process behavior.

This real-world behavior is reflected by the execution of, possibly, a multitude of instances that are created, initiated, and executed according to an often not explicitly stated process model. As a consequence, we have to think about how to exploit empirical information collected during the execution of process instances.

Depending on the effort spent on data collection, a broad spectrum of data might be available about the execution of process instances. Ideally, there exists a log of *events*, observed and stored during the execution of a process instance. For each activity, the log of a process will record its beginning by a *start event*, the completion of the activity by an *end event*, and, if necessary, also *interruption* or *resumption events*. Additionally, the time of occurrence for all these events is known (usually reflected by a time stamp), including additional attributes characterizing the activity associated with the events. Necessary for subsequent analysis is an attribute that reflects the activity label or id. Further attributes include the outcome of the activity, the people involved in the activity (mostly working on the activity), the cost of the activity, and the resources required for activity execution.

This description of data collection resembles the idea of a fully automated process which is hardly realized in practice. In particular, this is the case if events are triggered by customers. Moreover, the use of all available data for decision making is not recommendable, because one may get lost in too much detail. Hence, in BI we need specific views on the data about the instances of the business process. The overview box summarizes the different views on business processes.

BI Views on Business Processes

- *Event view*: The main emphasis is on the events in the business process characterized by a time stamp for the start, a time stamp for the end, and, if necessary, also a time stamp for the resumption of the activity execution after an interruption.

- *State view*: Besides the occurrence of events the state view also considers the values of attributes, the so-called state variables, measured in connection with the events.
- *Cross-sectional view*: In this case, we investigate the history of many process instances at a certain reference time. Usually, this view considers information about events as well as the values of state variables and summarizes the information about process instances for decision making.

The *event view* puts the main emphasis on the rules defining the partial ordering of the business process events according to the production perspective. This ordering of the events defines the *control flow perspective* of the business process. Figure 1.1 a outlines the recording of four events $e_1, \dots, e_4$ at the corresponding time stamps $t_1, \dots, t_4$ which defines a partial order between these events. Let us mention that, in some cases, it is rather difficult to exactly record the start and end events for the activity. In medical applications, for example, the start event for an illness is often hard to define and we have only information about the time of diagnosis.

The last remark leads us to the second view on the business process, which emphasizes the outcome of the process activities. These outcomes switch the focus from the business process to the corresponding process subjects. These subjects may be customers, delivered goods or services, or a network of business partners. The understanding of the behavior of these process subjects is based on measured quantities, so-called *state variables*. This notion suggests that the business process is treated as a dynamic system. Obviously, the values of the state variables change over time, either due to certain business process activities or due to some kind of inherent

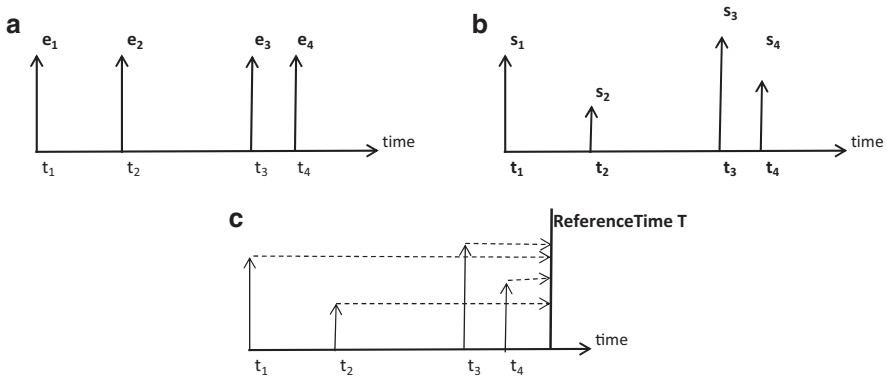


Fig. 1.1 BI views on business process instances. (a) Event view, (b) state view, (c) cross-sectional view

variability of the process subjects themselves. Think, for example, about the state of health of a person measured by some medical parameters or the change in quality of a product due to ageing. Understanding the instances of the business process as trajectories of such state variables will henceforth be called the *state view* on the business process. Figure 1.1b shows four states of one state variable, $s_1, \dots, s_4$, that are taken by the system at four points in time, i.e., $t_1, \dots, t_4$. The different heights of the bars represent the different values of the state variable, e.g., if a patient is in the state *running a fever* the height of the corresponding bar would record temperature at a certain time. The state view is sometimes blurred with the event view, because usually we need a business process activity for recording or changing the values of the state variables. For example, in business applications, the activity of examination of a customer's account is necessary for obtaining the financial state of the customer. Event view and state view result in a data structure for observed instances which is frequently called *temporal data* or *time stamped data*.

The third view starts from the observation that in many applications our main interest is on the aggregated quantities of instances of business processes at a certain point in time. Such aggregations may be obtained from the event view or the state view. Concerning the event view, we may consider the counts of the instances of the different activities occurring within a certain observation period, the total processing time of these activities, or the consumed resources for these instances. In the case of the state view, the aggregation may refer to an average value of the state for each instance over the observation period or the actual state of the instances. We will call this view the *cross-sectional view* on the business process. The idea behind the cross-sectional view on business processes is depicted in Fig. 1.1c. Note that the view on the states is taken at some reference time T .

Event, state, and cross-sectional views on business processes can be used in combination with the three perspectives: customer, production, and organization as defined above. In the production perspective, one can put the main emphasis on the sequence of events corresponding to activities for the production of goods or services. Another possibility is to focus on state variables describing the production (e.g., the utilization of resources over a period of time), or one can look at the summary of characteristics for production in a certain period of time. In the case of the customer perspective, either the cross-sectional or state view is dominant. Events play an important role in connection with the classification of customers. A credit default, for example, is an event that defines two classes of customers. In medical applications, the state view describing the behavior of patients with respect to a certain parameter is often of main interest. In the organizational perspective, a cross-sectional view is important for many applications, in particular for the analysis of networks of customers, but the organizational structure behind specific events may also be of interest.

1.2.4 Goals of Business Intelligence

The starting point for BI applications are analysis goals. Such goals range from the acquisition of information about some aspects of the business process over improving the performance of the process up to understanding the implications of the process for achieving strategic goals. The goals can be formulated in two different ways. The first one is based on so-called *key performance indicators (KPIs)*. KPIs allow measuring the performance of the business with respect to some goals in any perspective of the business. In the economic context, a key performance indicator may refer to the acquisition of new customers or the improvement of the produced goods or services measured with respect to customer satisfaction. A systematic overview and classifications of KPIs can be found in [13, 21]. In the educational context, KPIs may be the drop-out of students, the costs per degree, or measurements showing the position of the institution in the community. In medical applications, KPIs may refer to the efficiency of the treatment process or to the well-being of patients. Besides such quantitative indicators, one can also define other types of indicators that are more difficult to measure. Identification of KPIs is based on the identification of a predefined business process, on the definition of requirements for the business process, and on a measurement of the results of the business in comparison with set goals. In Sect. 1.4, we will define a number of KPIs for the use cases.

KPIs as Goals for BI

- *Key performance indicator*: A KPI links the activities of the business to objectives by defining a measurable quantity. KPIs may refer to some aspects of the performance of the business process or to the business as a whole. One can distinguish between quantitative indicators presented as numbers, practical indicators interfacing with processes, directional indicators showing whether the organization is getting better or not, actionable indicators for controlling effects of change, or financial indicators.
- *Influential factors*: Attributes that may influence the behavior of the KPI in any BI perspective.

KPIs describe the business at a top level. However, many a time BI goes one step beyond goal formulation and measurement of performance and aims to understand how business performance can be affected by so-called *influential factors* that determine and explain the values of KPIs. The relation between KPIs and influential factors can be used later on for decision support. For example, in the case of customer relationship, we are not only interested in the number of

customers who quit the business relation but want to understand the reasons for their behavior. Possible influential factors have to be investigated from all three business perspectives. The production view of the business, for example, looks at influential factors in connection with the production of goods or services. From the customer perspective, possible influential factors are frequently based on customer attributes defining sociodemographic characteristics and attributes referring to the perception of the offered products or services. The organizational perspective can help in the identification of influential factors connected with the internal organization of the business or the influence of social networks on customer behavior.

For understanding the relation between KPIs and influential factors, we use a second formulation of goals in BI called *analytical goals*. This formulation is based on a typology of the questions with respect to possible approaches in the analysis. One can distinguish three broad types of analytical goals summarized in the overview box. The first type, the descriptive goals, occurs in all perspectives and can be based on all three views on the business process. The basic descriptive goal is reporting which is frequently a supplementary goal for achieving other analytical goals. The descriptive goals segmentation and detection of interesting behavior are frequently summarized under the heading *unsupervised learning*. Predictive goals are more ambitious; they are of main interest in the case of the customer perspective using the cross-sectional view. For predictive goals, the term *supervised learning* is frequently used. Even more ambitious are understanding goals, which are usually closely related to the production perspective using either the event view or the state view. In the subsequent chapters, we will cover various models and analysis methods for these analytical goals from the different business perspectives.

Typology of Analytical Goals

- *Descriptive goals* generate a summary description for the instances of the business process from the different BI perspectives. Three main goals can be summarized under this heading:
 1. *Reporting*: Summarize the instances in such a way that one can use the information for decisions.
 2. *Segmentation*: Group the instances according to a similarity measure and find representative instances for these groups.
 3. *Detect interesting behavior*: Identify events during business process execution that allow the identification of important aspects of the process.
- *Predictive goals* predict the behavior of instances of the business process. Two different kinds of prediction may be distinguished:
 1. *Regression*: Find a function that allows the prediction of the output (usually a KPI) from a number of input variables (influential factors).
 2. *Classification*: Given a partition for observed instances into disjoint classes, assign a new instance to one of the classes.

- *Understanding goals* support stakeholders in understanding their business processes. Two main goals can be formulated:
 1. *Process identification*: Identify the rules that determine the relationships between the events of the process.
 2. *Process analysis*: Investigate the performance of the instances with respect to their conformance with a defined business process.

Note that the goal orientation is complementary to the life-cycle analysis of the business process along three phases: design time, run-time, and change time. At *design time*, the business process is described in terms of process models (cf. Sect. 2.3.2) defined in agreement with a KPI for certain business needs. Many a time, the background for the design is formulated as a *business plan*, and detailed formulation requires the investigation of different analytical goals. For example, if we are interested in launching a new product, analytical goals like prediction of market opportunities or detailed product description have to be achieved. The analysis of the process at *runtime* refers to data from process execution, i.e., data from process instances. Also in this case the analysis requires a precise formulation of analytical goals. Similarly, the analysis at *change time* corresponds to a specific formulation of a KPI. However, attention should be paid to the fact that traditionally this analysis along the lifetime of a process is mainly understood in connection with the production perspective. Our goal-oriented formulation of analytical goals seems more open to the other perspectives.

1.2.5 Summary: Putting Business Intelligence in Context

For the development of a unified umbrella for BI, we use a process-oriented definition of the term business applicable in many different domains. One can look at such a business process from different perspectives, in particular the production perspective, the customer perspective, and the organizational perspective are identified. In connection with the perspective, it is often important to identify the roles of actors within the business process; in particular, process subjects as the actors that generated instances of the business are of utmost importance in BI.

The main input for all BI activities are data about the instances of business processes. These data are generated according to a specific view on the business process. Three views are identified: the event view, the state view, and the cross-sectional view. In the production perspective, the event view is of utmost importance and in the customer perspective the cross-sectional view is dominant.

Using data as input, any BI activity starts from a certain goal. For the goal measurable quantities, so-called key performance indicators (KPIs) are defined. The KPIs have to be seen in connection with the strategic use of BI inside the business. This strategic use ranges from application of BI for achieving short-term targets with no connection to the management strategy over use of BI as a feedback for the overall management strategy up to understanding BI as a strategic resource for management decisions.

Many a time, BI applications aim for understanding the dependence of a KPI from other quantities called influential factors. This leads to the formulation of analytical goals for BI. Different analytical goals can be identified: descriptive goals, predictive goals, and business understanding goals. These analytical goals allow a formal analysis, and the results of the analysis can be used later on for decision support.

1.3 Business Intelligence: Tasks and Analysis Formats

Achieving the different analytical goals requires the completion of a number of tasks, including an analysis format for the execution of these tasks. In this section, we briefly describe the tasks and propose an analysis format.

1.3.1 Data Task

The data task is a prerequisite for all BI activities. The main goal is organization of available information about the business and its environment. Typically, the information are data about the structural properties of the enterprise and the registered customers, the transactional data from business process instances, the data describing production activities, or traces of activities in social networks. These data are collected under different data-capturing regimes and stored in different data sources using multifarious structures ranging from data with diverse temporal and spatial granularity up to semistructured text data. The major challenge is organizing the data in such a way that they can be utilized in various BI activities.

Many a time, one can start with an existing organization of the data in a data warehouse, which offers coherent data of high quality and thus supports diverse BI activities, in particular standard reporting. This is the reason why BI is often understood as an endeavor of data modeling and retrieval. However, due to the changes of the business and its environment over time, even a well-designed data warehouse cannot answer all questions. Decision support for new challenges may require a reorganization of the data or collecting additional data for special purposes. Consequently, it is necessary to have knowledge about the methods for data collection and for the augmentation of existing data with new data.

The data task relies on *data modeling techniques* encompassing different data models like ER models, UML, or semistructured data models, including methods on how to apply the models, and an IT infrastructure for data provisioning. Chapter 3 discusses topics of data provision and introduces analytical data formats useful for the different business perspectives. Moreover, issues of data integration are discussed in Chap. 3.

1.3.2 Business and Data Understanding Task

The starting point of business and data understanding is an initial formulation of a goal, in the best case formulated as KPIs. The business and data understanding task considers the business regarding this intended goal and develops first ideas about what part of the business is of interest in connection with the goal and what data from the repository can be used. The results of the task are a formulation of analytical goals, an excerpt of the overall business relevant for the analytical goal, and data needed for achieving the analytical goal. Moreover, a first outline of the work schedule for further activities in the BI project is defined. This needs a number of interrelated activities that are summarized in the overview box.

The first two activities are more oriented towards business understanding and specify the application environment and the business perspectives of interest. Next we have to decide about the view on the business and the data as a first step in data understanding. In BI, data traces of past instances of the business process are the main source of knowledge used for analyzing the business with respect to the goals. Although BI often has an exploratory nature, some prestructuring according to domain knowledge is necessary. Consider a medical treatment process as an illustrative example:

It is not reasonable to use all possible parameters informing about the health status of a patient for monitoring a specific treatment process. Instead, a number of potential influential factors are selected according to expert knowledge, belief, and interest. In other words, knowledge about the process is mapped to a number of variables also taking into account factors that are probably not part of the established knowledge.

Issues in Business and Data Understanding

- *Application environment:* This topic explores the size and the scope of the analysis goal within the overall business and determines the BI scenario (Sect. 1.2.1) for the application. Furthermore, the resources and time horizon of the project are determined.
- *Business perspective:* This point covers the investigation of the analysis goal from the different business perspectives comprising the identification of process owners, process subjects, and actors in the business process (Sect. 1.2.2).

- *BI views*: That part of existing data relevant for the goal is identified and an appropriate view on the business and the data of interest for the analysis is specified (Sect. 1.2.3).
- *Analytical goals*: A precise definition of the envisaged KPIs and influential factors is given and the intended types of analytical goals (Sect. 1.2.4) are formulated.
- *Assessment of data*: Screening of data with respect to properties of the variables and data quality. Furthermore, a number of data transformations may be necessary for editing the data in such a way that they can be used in the envisaged models necessary for achieving the analytical goal.

The choice of view on the business depends not only on knowledge and goal but also on the availability of the data. Many a time, data in the cross-sectional view can be easily accessed, whereas detailed data about process instances in the event view are only available for parts of the process. In the cases where data in the desired view are not available, one has to decide whether the use of existing data is feasible for the goal. In the worst case, it may happen that the availability of the data is the limiting factor and a new data collection may be necessary.

Using the information gathered up to now, one can give a precise formulation of the goal in terms of analytical goals. This requires a combination of domain knowledge, i.e., business understanding, and knowledge about properties of the data, i.e., data understanding.

Besides the general considerations about data, the feasibility of the envisaged analysis depends often on data peculiarities. These peculiarities are found in data assessment. Investigation of properties of the data using data description and visualization techniques gives quantitative information about individual variables together with fundamental relations between the variables. Such an analysis supports often the selection of possible influential factors. Issues of data quality refer to data generation, for example, the completeness of data or the coherence of data from different sources. These aspects will be discussed in Chaps. 2 and 3.

Often data assessment includes data transformations for obtaining the data in a form needed as input for the modeling and analysis task. One type of transformations are those necessary for obtaining a unified view on the data of the business process. For example, if some of the influential factors are recorded as data in the event view and others only in the cross-sectional view, we have to transform the data in such a way that we can use them in one model. Another type of transformation is the computation of new variables out of existing ones, for example, scores.

The *business and data understanding task* uses *business understanding techniques* and *data understanding techniques*. Business understanding techniques answer the questions about the application environment and require domain knowledge about the business and experience in project management. In the literature about project management, one can find techniques for structuring this process,

but an open mind, discussions of the problem from different perspectives, and experience are probably the most important requirements. With respect to data understanding, one can rely on techniques for data description and data visualization, which will be treated in Chap. 4. We use the term technique to emphasize that we have to rely on models for the description and visualization, methods for using such models, as well as tools that support the realization of the task.

1.3.3 Modeling Task

The modeling task aims at setting up an *analytical business model*, i.e., a formal model that allows precise answers for the analytical goals. Depending on the BI perspectives, views, and goals of interest, we use formal structures to build a model that enables the transformation of the analytical goals into formal questions about the properties of a model. Sometimes, the model may be rather simple and is not more than a specific query on the available data. At other times, choosing a model may be rather intricate and is by no means evident. Consider, for example, the model formulation for analyzing a KPI in connection with customer acquisition:

- One can start with a marketing-oriented approach, take the customer perspective, and identify factors that attract new customers. After the identification of these factors, one can think about the necessary internal processes meeting the customer requirements.
- Another approach is to start with the production perspective, scrutinize the production process, develop possible scenarios for changes of the production processes. Afterwards, the different scenarios are analyzed with respect to the attractiveness for customers.

This shows that different model formulations are possible depending on the business perspective and the formulation of the analytical goal. Figure 1.2 illustrates the interrelations between goals, perspectives, views, and analytical models.

The circles in the center of the hexagon represent the different BI perspectives. The intersection of the circles illustrate that we often have to cope with analytical goals that need multiple BI perspectives. The inner labels at the sides of the hexagon describe the BI views on the perspectives, for example, taking an event view to analyze the production perspective or a combination of production and organizational perspective. For example, one may take the event view to analyze the production perspective or a combination of production and organizational perspective.

Above the hexagon, we denote the BI goals, i.e., understanding goals, descriptive goals, or predictive goals as discussed in Sect. 1.2.2. The lower part of Fig. 1.2 introduces the formal structures used for transforming BI goals into properties of the model. Basically, one can distinguish between models with an algebraic

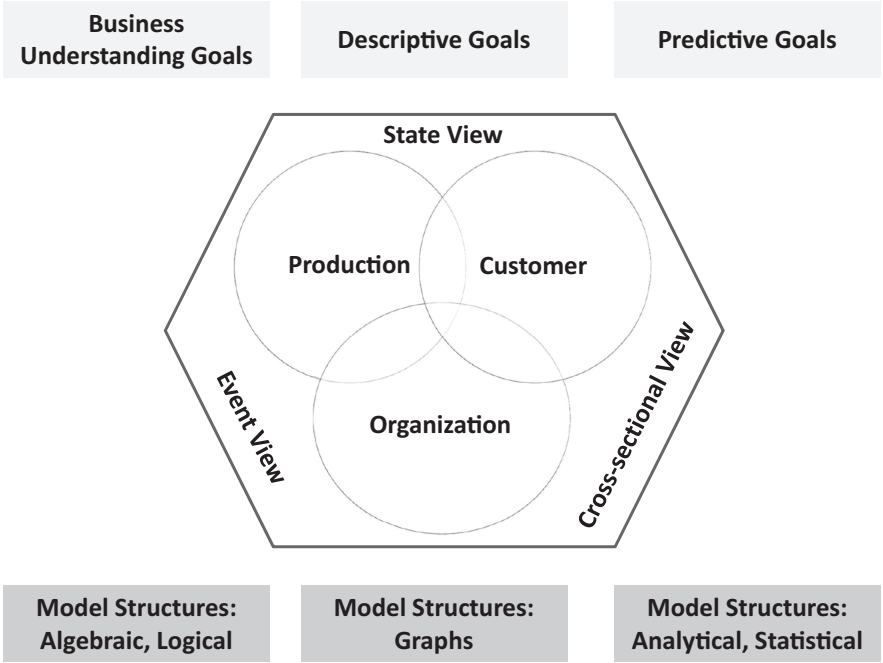


Fig. 1.2 Overview on modeling activities

and a logical structure, models with a graph structure, and models with analytical structure. Among the algebraic-logic structures, business process models are the most prominent ones. Among the analytical structures, probability and statistics are most important and frequently used. Graph structures combine analytical and algebraic elements and play an important role in BI modeling.

Corresponding to their many facets, successful BI applications rely on a rather large repository of possible *modeling techniques*. As we will discuss in Chap. 2, a modeling technique is based on a model structure, a method of using these structures, and tools supporting the formulation of a model. Generally speaking, Fig. 1.2 can help in structuring a model repository in such a way that one can define or find a model that fits into the different perspectives and views and tackles a certain goal. It can also help to think about frequently used approaches in different scenarios. However, leaving the standard path is always an option worth pursuing, provided available data allow for doing so.

Besides modeling techniques, the modeling task requires specific data preparation techniques. For example, in the case of analysis goals referring to text data, different techniques can be used for transforming such unstructured data to structured data which allow the application of algorithms. Such transformations will be considered in connection with the analysis methods.

1.3.4 Analysis Task

Having defined a model, one needs algorithms to compute a solution for the analytical goal within the model. In BI, these algorithms are usually denoted by the term *mining*, stressing that we are searching for a solution concerning a frequently not very well-defined problem. Another frequently used term is *Machine Learning* which has its origin in Artificial Intelligence and was formally defined in [16] as computer programs with the ability to learn to solve a task. Learning is understood as improving the performance of a program using experience from past executions. This experience is obtained from examples but the number of examples is not necessarily large. The term data mining is originally defined as the analysis step in the process of the knowledge discovery in data bases which has a more exploratory nature. Today, the two terms are often used synonymously, but in BI applications mining is predominant and we will use it throughout this book.

Different types of mining have been proposed; Fig. 1.3 shows an overview on these types in connection with the different BI perspectives. Note that this includes mining algorithms that frequently occur in connection with overlapping BI perspectives. At the intersection of the perspectives production and customer, for example, decision mining has been suggested in the literature.

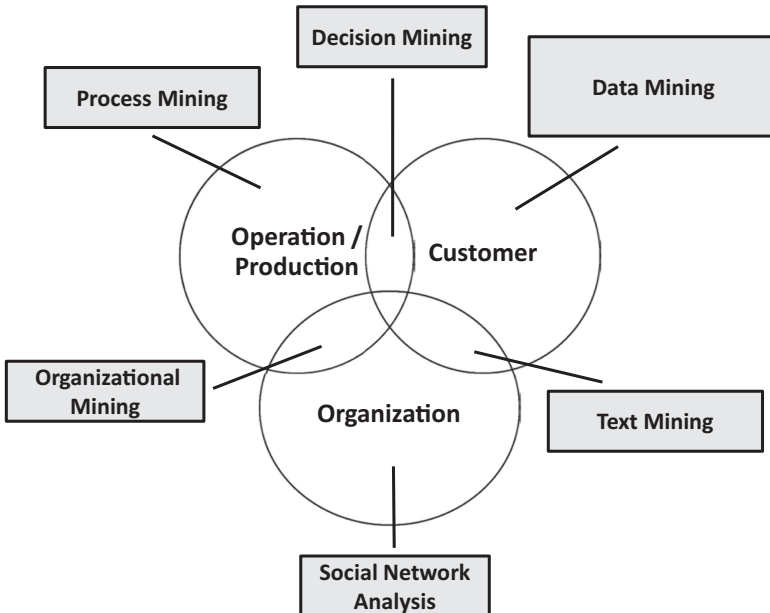


Fig. 1.3 Overview on mining algorithms along the BI perspectives

Besides the BI perspective, the choice of the algorithms depends on the view and the envisaged analytical goal. This combination provides us with an organization of the book chapters that deal with the analysis task in BI, i.e., Chaps. 5–8. Chapter 5 is devoted to data mining algorithms and tools for the cross-sectional view on different BI perspectives, mainly the customer. Chapter 6 features data mining for temporal data that are particularly suited for analyzing the state view applied to analytical goals occurring in the customer and the production perspective. Chapter 7 presents process analysis techniques and specifically takes the event view on the production perspective. Chapter 8 considers techniques that address questions of organizational mining and questions at the intersections of two perspectives taking different views. For example, in the case of text mining, analytical goals typically arise at the intersection of the customer and the organizational perspective, using a cross-sectional view and focusing on text data. For this, we could be interested in learning the opinion of customers about a certain person or product expressed at some social platform like Twitter.

Similar to the case of modeling, we need for analysis a repository of *analysis techniques* that stores the knowledge about the algorithms, the methods of using these algorithms, and the tools that support the application. These methods represent the knowledge about the required view, the analysis goal, and the data structure for the algorithm.

1.3.5 Evaluation and Reporting Task

The evaluation and reporting task has to view the analysis results from two different perspectives. The first one is evaluation of the results in context of the analytical goal and the second one is evaluation from a global business perspective, i.e., understanding the results of the analysis in the context of the business. The main goals are the interpretation of the results in reference to domain knowledge and coming to a decision of how to proceed further. Usually, the evaluation task employs reporting techniques that are similar to data description and visualization techniques. Depending on the intended audience of the report, different types of reporting can be distinguished. We will sketch some ideas in Chap. 4.

1.3.6 Analysis Formats

In the same way as we understand all business activities as a process, we can look at BI activities as a process and define a structure for organizing the different tasks. Such structures are subsumed under the term *analysis format* using ideas from life-cycle models for software development or from knowledge discovery in data bases. There is a basic distinction between cyclical and linear formats. We think that

cyclical formats are more useful in BI, because, in practice, covering the different perspectives often requires a sequence of models and a combined evaluation of the results. Another argument in favor of a cyclic model is that a first analysis frequently detects new and unexpected features that need further investigation.

An influential and widely used method format is the Kimball method [12]. It emphasizes the data task and starts with the definition of analysis goals and business understanding. The main focus is on the deployment activity, i.e., the integration of the results of BI activities into a data warehouse. The business analytics aspect is only treated as a secondary aspect, and the integration of unstructured data sources needs some extensions.

A number of analysis formats for the knowledge discovery process have been proposed as data mining formats. One can distinguish between academic-oriented efforts, cf. [10], and application-oriented approaches like CRISP [4], which is nowadays some kind of standard for data mining applications. Other formats are SEMMA or KDD [2]. CRISP mainly focuses on the cross-sectional view on the business process but adaptations to the state view are possible. Our definition of tasks is closely oriented towards CRISP. The main difference is that we formulate a closer connection between business understanding and data understanding. In addition, we approach the modeling and analysis task in a way that allows coping with the event view on the business process and the application of process mining techniques.

Recently, the L^* format [24] has been proposed as a method for business process analysis and mining, which emphasizes explicitly the idea of the application of different models and analysis techniques. The L^* method starts with the event view of a business process, mainly seen from the production perspective. After planning and justifying, which corresponds to business and data understanding, the data (in an event log format) is extracted and a process model is formulated and analyzed. Based on this model, the organizational perspective and the customer perspective are investigated using different analytical methods. Such investigations use a state view and a cross-sectional view on the business process. In a deployment phase, the derived process models can be implemented and operationally supported.

Combining the ideas of CRISP and L^* , we propose the *iMine* analysis format that supports the integrated application of data and process mining. *iMine* hereby stands for integrated mining.

The *iMine* workflow is depicted in Fig. 1.4. The format allows different types of analysis cycles. If the analysis goal is, for example, to discover the real-world processes from data sources, i.e., to conduct process mining, the data used will be in the event view and the modeling techniques and the *analytical techniques* will be process-oriented. If the analysis goal is obtaining knowledge about customers' preferences, the data formats will be rather found within classical, multidimensional table structures corresponding to the data of a warehouse application. In addition, the modeling techniques will be oriented towards statistics and the analysis techniques will provide algorithms for cross-sectional analysis, for example, clustering or association techniques (data mining).

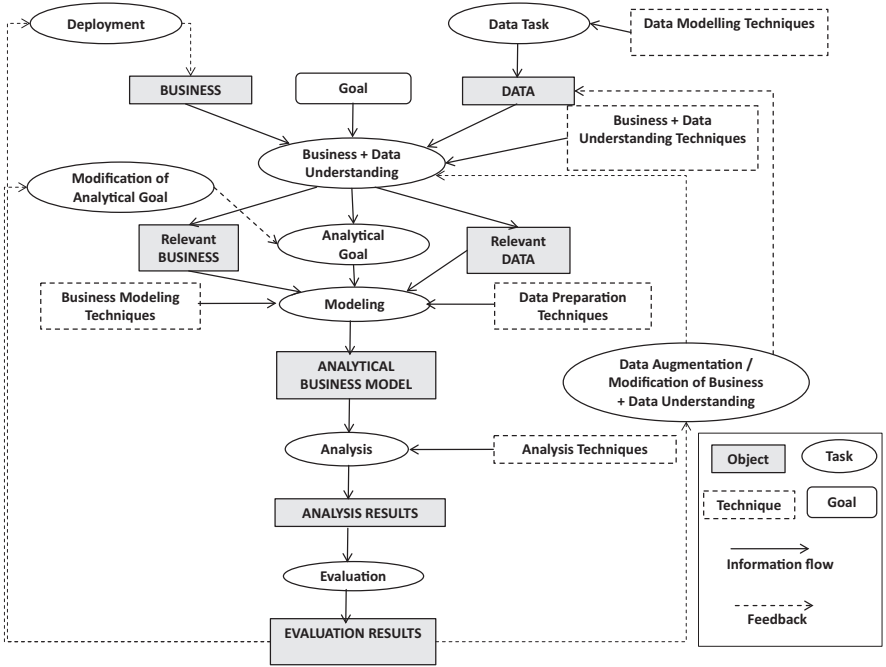


Fig. 1.4 The iMine method

The iMine Method: Basic Ingredients

- *Objects*: Rectangles with full borders represent the objects that are analyzed or produced during the BI process along the three goal perspectives.
- *Tasks*: Ellipses represent the BI tasks that have to be conducted by the BI analyst.
- *Techniques*: Rectangles with dashed border represent techniques for the different BI tasks. Techniques consist of repositories for procedures, methods for using the procedures, and tools for implementation.
- *Information flow*: Solid arrows represent the flow of information.
- *Feedback*: Dashed arrows refer to possible feedback loops.
- *Analysis goals*: Analysis goals are depicted as rectangles with rounded corners.

The iMine method starts from the available *data* that has been provided by the *data task* employing *data modeling techniques*. *Data* and *business* together with a certain *analysis goal* provide the input for the task of *business and data understanding*. Thereby, *business* and *goal* together represent that part of reality for which we aim for decision support. The output of this task are *analytical goals* together with *relevant business* and *relevant data*, possibly as a certain fragment or view on the original *business* and *data* along the different BI perspectives customer,

production, and organization. Note that frequently the initially formulated goal needs a number of different analytical goals. For example, a first analytical goal may be a descriptive goal, and afterwards a predictive goal is formulated. Based on the results from *business and data understanding*, an *analytical business model* is built in the *modeling task* using *business modeling techniques* and *data preparation techniques*. Data preparation techniques mainly use transformations mentioned in Sect. 1.3.2, either for unification of the business process views or for generating new data like scores. The *analytical business model* is then analyzed by different *analysis techniques*, leading to *analysis results*. These results are then evaluated yielding *evaluation results*.

In the best case, the evaluation indicates that the analysis results are satisfactory for deployment. This means that we have obtained a result which should be implemented for further business decisions (see the feedback loop at the very left in Fig. 1.4 from *evaluation results* over *deployment task* to the *business*). Such an implementation obviously depends on the chosen strategy about using BI within the business model as discussed in Sect. 1.2.1. The decision might be realized by means of a standardized report, a procedural rule for certain customers, or a reorganization of the business process that was subject to analysis. Corresponding to the strategy, we have to adapt the data and the business itself. Nowadays, it is common to provide a dedicated area for analysis tasks inside a data warehouse which is called *analytical sandbox*. If the results derived in the analytical sandbox are ready for deployment, the operative data might be augmented for usage in the business.

Many a time, further investigations are necessary as defined by the other feedback loops. The feedback loop on the left side of Fig. 1.4 coming from *evaluation results* over a *modification of analytical goal* to the *analytical goal* initiates a new analysis cycle with a new or modified analytical goal and a new understanding of business and data.

Such a new goal may be the consequence of the evaluation of the results. For example, we start with a descriptive goal and learn from the analysis results that the analysis of a predictive model may be reasonable, leading to a respective modification of the analytical goal. In more detail, starting from a scatter plot of a number of variables (descriptive), it can be concluded that certain regression models as predictive models may be feasible.

Another possibility is that the new analysis cycle has been planned already in advance. This frequently occurs for goals at the intersection of the perspectives. If we want to understand, for example, a certain business process, we start with process mining techniques for obtaining the structure of the process. In a second round of analysis, decision trees are applied for understanding the decision in the process. In Fig. 1.3, this corresponds to the intersection of the perspectives production and customer referred to as *Decision Mining*.

The feedback loops *data augmentation* and *modification of business and data understanding* on the right side in Fig. 1.4 refers to further investigations. It takes the *evaluation results* as input for the task of *data augmentation and adaptation of business and data understanding* which provides the input for the adapted *data* and the task of *business and data understanding*. This means that we define a new

analytical (sub)-goal with a different view on the data and the business processes. In some cases, this may imply new data collection activities and a new analytical business model.

1.3.7 *Summary: Tasks and Analysis Formats*

This section presented an introduction into the different BI tasks. The starting point for any BI activity is a business goal that is frequently formulated in terms of KPIs. In the *business and data understanding task*, the framework of further analysis is defined. This requires background knowledge of the application domain and knowledge about possible models and analysis techniques. As a result of this task, one obtains a problem-oriented view on the available data and the business. Analytical goals of interest are defined and a first model for the analysis is formulated. Depending on the goal and the business perspectives of interest, the formulation of the model uses different analytical structures. These structures may be logical oriented, graph oriented, or statistical oriented. Each structure supports different techniques for the analysis task. The term *mining* is used as an umbrella term for the different analysis techniques. The last step is the evaluation and reporting task, which summarizes the findings and gives an interpretation in terms of the original goal.

Frequently, one analysis sequence is not sufficient for achieving the goal and a cyclic analysis format is defined. The format combines ideas of the CRISP format for data mining and the L* format for process mining. In that sense, it is open for application of analysis methods for data in the different views of the business.

1.4 Use Cases

In this section, we introduce five use cases presented throughout the book. They stem from four different application domains, i.e., patient treatment, higher education, logistics, and customer relationship management (CRM). A description of the use cases is given by the following template, which defines the input for an analysis cycle of the *iMine* format outlined in Sect. 1.3. Each use case has a number of goals, which implies the execution of a number of cycles.

Use Case Description Template

- **Business Case:**
- **Goals:**
- **Data Task:**

Afterwards, the business and data understanding task for the use cases is summarized in a template along the lines of the overview box in Sect. 1.3.2.

Business and Data Understanding Template

- **Application Environment:**
- **Business Perspectives:**
- **BI Views:**
- **Analytical Goals:**
- **Assessment of Data:**

For the different analytical goals templates for the modeling task, the analysis task and the evaluation and reporting task are presented in Chaps. 4–8.

1.4.1 *Application in Patient Treatment*

Health care is a very demanding application area and has been researched by various fields of computer science during the last years. We present two use cases from health care, i.e., skin cancer treatment and pre-eclampsia in pregnancy. The first one is based on a real-life data whereas the second use case is based on a realistic setting, but with simulated data. We opted for these two settings in order to demonstrate different aspects of analysis.

The EBMC² Project: Description

- **Business Case:** The Evidence-based Medical Compliance Cluster project (EBMC²) (see <http://ebmc2.univie.ac.at/>) aims at the analysis of skin cancer treatment processes. The project has been conducted as joint funding and effort between the Medical University of Vienna and the University of Vienna, more precisely, the Department of Dermatology, the Center of Medical Statistics, Informatics, and Intelligent Systems, and the research groups Data Analytics and Computing, Knowledge Engineering, and Workflow Systems and Technology.
- **Goals:** Analysis goals refer to the treatment of patients as well as to the performance of the institution (hospital). With respect to the treatment of patients the following KPIs are of interest:
 - Survival time of patients
 - Compliance of patients with preventive medical check-ups

For performance of the institution, we will consider the following KPIs:

- Compliance of the institution with the international melanoma guidelines
- Organization of internal work processes

- **Data Task:** The data sources are data repositories at the local department, the Austrian Cancer Registry, and data from the Austrian Social Insurance System. Data are different with respect to temporal granularity, the quality of information and its completeness, as well as its structure (structured and semistructured data). Hence, we need a flexible data model that allows the integration (i.e. linking of data) according to the analysis goals of interest. Such linking also allows additional quality considerations.

The following template shows the details for the business and data understanding task.

EBMC² Use Case: Business and Data Understanding

- **Application Environment:** With respect to size, we look at a certain department of a hospital and a specific illness (melanoma) over a longer period of time. This involves many different activities by the hospital and different reactions of the patients. The reference to the global environment is of utmost importance such as the international comparison of comparable institutions and their embedding into the public health-care system. In this case, the BI scenario mainly monitors the strategy performance of the department.
- **Business Perspective:** Corresponding to the different goals, we consider all business perspectives. The process owners are the Department of Dermatology, the process subjects are patients, and the process actors are defined by the staff of the hospital.
- **BI Views:** All three BI views are used depending on the analytical goals. For measuring the effectiveness of the check-up and the treatment, we use the cross-sectional view, for issues of survival time the state view, and for the process-oriented analytical goals the event view.
- **Analytical Goals:** Corresponding to the initial formulation of KPIs a number of analytical goals can be formulated.
 - Effectiveness of preventive medical check-ups
 - Regression model for survival time of patients
 - Compliance analysis for the treatment process
 - Organizational mining for hospital staff
- **Assessment of Data:** Each analytical goal needs a specific data excerpt, which is obtained by data integration and transformations of the data, for example, cross-sectional summaries of the event view. Data description has to be done for each data excerpt, and data quality has to be checked as well. Important issues are completing missing information and improving temporal resolution.

The second use case in medical context considers a specific problem of complication in pregnancy.

Pre-eclampsia Use case: Description

- **Business Case:** Pre-eclampsia is a complication in pregnancy caused by multiple factors. In order to detect pre-eclampsia, weight, blood pressure and proteinuria of women are monitored during pregnancy.
- **Goals:** In this use case, we consider the following goals:
 - Representation of the monitoring process by quantitative measures
 - Rules for deciding about the need for hospitalization of a persons
- **Data Task:** This use case was mainly considered for evaluating the *DPA^{TimeSeries}* methodology as presented in [9] for decision point analysis. Time series of measurements for weight, systolic and diastolic blood pressure, and proteinuria were generated for 300 cases containing 8 % of cases with pre-eclampsia.

For the business and data understanding task, the specification is given in the following template.

Pre-eclampsia Use case: Business and Data Understanding

- **Application Environment:** This application is general without reference to a specific business environment. The scope of the process is rather simple and defined by six types of events: proteinuria check, blood pressure check, weight check, hospitalize patient, homecare, giving birth. The checks are repeated from the 20th week of pregnancy onwards. The BI scenario is BI separated of strategy performance, but the results of the analysis can be used later on for defining new business strategies for monitoring pregnancy.
- **Business Perspective:** The perspective in this use case is on the customer who is owner of the process and process subject. Furthermore, medical practitioners are passive actors who decide about hospitalization.
- **BI View:** This is a typical example of data in the state view, i.e. we have time series about the medical parameters.
- **Analytical Goal:** The KPI is defined by hospitalization in dependence on the temporal behavior of the medical parameters. Correspondingly the analytical goals are descriptive goals (reporting) and predictive goals (classification rules allowing a decision about hospitalization).
- **Assessment of Data:** Due to the control through simulation, the data is complete allowing the evaluation of algorithms. For reporting purposes, summaries for the instances are calculated.

1.4.2 *Application in Higher Education*

The Higher Education Processes (HEP) project features a realistic data set on business processes that reflects the realization of different undergraduate courses at a university. These courses incorporate different stakeholders and a fair amount of process subjects (i.e., students) and participants (e.g., tutors, lecturers).

HEP Use Case: Description

- **Business Case:** As part of the Higher Education Processes (HEP) project (see <http://www.wst.univie.ac.at/communities/hep/index.php?t=main>, [15]) undergraduate teaching courses at the Faculty of Computer Science, University of Vienna, were observed based on the usage of the teaching platform CeWebs [7].
- **Goals:** The following goals are of main interest:
 - Derivation of a reference process model exploiting additional compliance constraints, e.g., setting out the relation of milestones and submissions
 - Conformance between real-life teaching processes and a given reference model
 - Utilization of the forum by students and the relation of students' performance to activities in the forum
- **Data Task:** Data are collected from four distinct services, i.e., forum, submission, registration, and code evaluation, on the service-oriented learning platform CeWebs. Logs from annually offered undergraduate courses over a period of 3 years (course conducted every year) are available. In total, there were 330 students and 18,511 events. For the use case, data were collected in the .csv format and anonymized.

The specifications for the business and data are shown in the following template.

HEP Use Case: Business and Data Understanding

- **Application Environment:** The business case under consideration is defined for a faculty. The scope of the processes is complex due to constraints that hold for an undergraduate teaching process. An example of a constraint with high enforcement level is “For each milestone, no upload must take place after the corresponding milestone deadline” [15]. Finally, we acquired the reference teaching process based on interviews with process participants, e.g., lecturers. The business scenario is monitoring the process and provides feedback on the business strategy of the faculty. In the long run, the results can also be used as a strategic resource for improving the teaching process.
- **Business Perspectives:** All three business perspectives occur in this use case. The process owner is the faculty and the customers and process subjects are the students. Other actors are the tutors and lecturers.

- **BI Views:** The event view will be used for the business understanding goals and the cross-sectional view for the goals in connection with student performance.
- **Analytical Goals:** Corresponding to the goals, the following analytical goals can be formulated:
 - Understanding of the business process
 - Conformance analysis for the process instances
 - Segmentation of students into groups with different utilization of the forum
- **Assessment of Data:** Data are extracted from the service-oriented blended learning environment and integrated along the analysis goals. After data purging and cleaning, data are represented in a log-oriented format. Using data transformations, additional attributes are generated in the cross-sectional view.

1.4.3 *Application in Logistics*

Contrary to the above examples taken from patient treatment and higher education, the logistics use case targets the transportation of containers, i.e., the subjects are not human. The use case is based on a realistic setting but uses synthetic data that was produced by simulation.

Logistics Use Case: Description

- **Business Case:** The container transportation case has been adapted from a realistic process that was described and implemented in [3]. It describes the process of loading a vehicle at the origin and starting to move towards its destination. During the movement of the container, temperature is constantly monitored. If the temperature exceeds a certain threshold for some time, the vehicle has to move back to its origin. Otherwise, it continues to the destination where the containers are unloaded.
- **Goals:** Two main goals are identified:
 - Derivation of a model for the transportation process
 - Derivation of a decision rule on when to return the container to their origin before they were delivered
- **Data Task:** Simulated data of the transportation process and the time series measurements of container transportation are used.

Logistics Use Case: Business and Data Understanding

- **Application Environment:** Similar to the pre-eclampsia use case, there is no detailed specification of the business environment defined. We consider a general problem for logistics and the scope of the business only refers to a small

number of activities. The application scenario uses BI separated from business strategy for a specific subprocess of a possibly larger process, but results can be used later on.

- **Business Perspective:** The logistic company is the process owner, the containers are the process subjects, and further actors are personnel of the company. Since the containers are understood as customers involved in the transportation process, the main perspective is the customer perspective.
- **BI View:** We use the event view for the first goal and the state view for the second goal.
- **Analytical Goals:** For achieving the goals, we formulate two analytical goals: The first one is process identification and the second one is classification of the different process instances in such a way that an economically favorable decision strategy for returning to the origin can be formulated. Influential factors is the temperature of the container.
- **Assessment of Data:** Due to the fact that the data were simulated, data are already in an adequate format and are of good quality.

1.4.4 Application in Customer Relationship Management

This use case is a typical application in customer relationship management (CRM). It is based on a real use case but data and attributes have been modified.

CRM Use Case: Description

- **Business Case:** A company with outlets in different cities offers customers a variety of services. Customers are registered by a loyalty card. In order to improve services according to customer needs and learning about the image of the company, a survey was carried out. About 2,300 customers participated in the survey. For each customer, age, sex, duration of the business relationship, and usage type (either private or business user) is known. A user profile for long-term customers is defined by three indicators: an indicator for sales, an indicator for the intensity of usage, and an indicator for the duration of customer relationship. Usage of the services is known from the transaction database.
- **Goals:** The following KPIs are of interest for the company:
 - Sales of customers in dependence of their usage profile
 - Usage of the different services
 - Customer profiles with respect to usage of the different services
- **Data Task:** Three data sources are used: a customer data base (loyalty card), a transactional data base, and a survey data base.

Business and data understanding is summarized in the following template.

CRM Use Case: Business and Data Understanding

- **Application Environment:** The use case considers a medium-sized company with a large number of activities in the different services, which do not have many dependencies. The application scenario uses BI as feedback for strategy formulation (different services) and probably also as strategic resource (bundling and redefining services).
- **Business Perspective:** Corresponding to the sales orientation, the main perspective is the customer perspective. The owner of the business processes is the company and the process subjects are the customers. Additionally personnel in the outlets are actors in the sales process.
- **BI Views:** All available data are cross-sectional data.
- **Analytical Goals:** In that case, the KPIs define the analysis goal estimation, segmentation, and classification. Further analytical goals are description of customers and detecting interesting behavior.
- **Assessment of Data:** For obtaining cross-sectional data, all transactional data were aggregated at different temporal resolutions (monthly, quarterly, annual). The description of the different sales variables and the identification of outliers are done as well as the description of the survey about user satisfaction.

1.5 Structure and Outline of the Book

Chapter 2 discusses different approaches towards the modeling task in BI applications. Chapter 3 presents an overview and details on the data provisioning task, including an excursion on big data. In Chap. 4, the tasks data description, visualization, and reporting are discussed. Chapter 5 introduces data mining techniques for cross-sectional data. Different techniques for the analysis of temporal data are presented in Chap. 6. Chapter 7 discusses techniques for the analysis of process data followed by the introduction of analysis techniques for multiple BI perspectives in Chap. 8. The book closes with a summary and discussion in Chap. 9.

Throughout all chapters, tools are recommended, described, and applied. The focus is on open-source tools in order to be able to set up BI applications at no cost. A survey on tools can be found in the Appendix.

Further on, the concepts presented in this book will be illustrated and deepened by means of selected exercises. These exercises are available on the book's website:

www.businessintelligence-fundamentals.com

1.6 Recommended Reading (Selection)

As it was argued in the historical overview, there are many different approaches towards business intelligence, and an abundance of literature is available ranging from an elementary introduction up to textbooks at an advanced formal level. Hence, any recommendation is a matter of taste and the understanding of the term business intelligence. Starting from the business context, we aim in this book for an understanding of BI that encompasses data warehousing business analytics and process analysis. For an introduction into BI from a management perspective, we recommend Laursen and Thorlund (2010). If one is mainly interested in the data warehouse perspective, Kimball and Ross (2010) is a standard reference. An extensive exposition of the data mining approach towards BI can be found in Han and Kamber (2011). This book covers in depth the different data structures summarized as different BI views. An excellent reference from a process-oriented point of view is van der Aalst (2012), which covers the different perspectives of BI.

- Laursen G, Thorlund J (2010) Business analytics for managers: taking business intelligence beyond reporting. J. Wiley & SAS Business, New York
- Kimball R, Ross M (2010) The Kimball Group reader: relentlessly practical tools for data warehousing and business intelligence (1). J. Wiley, New York
- Han J, Kamber M (2011) Data mining: concepts and techniques. Morgan Kaufmann Series in Data Management Systems, Waltham MA
- van der Aalst WMP (2011) Process mining—discovery, conformance and enhancement of business processes. Springer, Heidelberg

References

1. Amit R, Zott Ch (2012) Strategy in changing markets: new business models—creating values through business model innovation. MIT Sloan Manag Rev 53(3):41–49
2. Azevedo A, Santos MF (2008) KDD, SEMMA and CRISP-DM: a parallel overview. In: Weghorn H, Abraham AP (eds) IADIS'08: European conference data mining. IADIS Publications, pp 182–185, Pedreira, Portugal
3. Bassil S, Keller RK, Kropf P (2004) A workflow-oriented system architecture for the management of container transportation. In: Desel J, Pernici B, Weske M (eds) BPM'04: international conference on business process management. Lecture notes in computer science, vol 3080. Springer, Heidelberg, pp 116–131
4. Chapman PJ, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, Wirth R (2000) CRISP-DM 1.0 Step-by-step data mining guide. <http://www.the-modeling-agency.com/crisp-dm.pdf>. Accessed 20 December 2014
5. Davenport TH (1992) Process innovation: engineering work through information technology. Reissued edition. Harvard Business Press, Boston
6. Davenport TH (2006) Competing on analytics. Harv Bus Rev 84(1):98–107
7. Derntl M, Mangler J (2004) Web services for blended learning patterns. In: Looi K, Sutinen E, Sampson DG, Aedo I, Uden L, Kähkönen E (eds) ICALT'04: international conference on advanced learning technologies. IEEE, New York, pp 614–618

8. Dumas M, La Rosa M, Mendling J, Reijers HA (2013) Fundamentals of business process management. Springer, Berlin/Heidelberg
9. Dunkl R, Rinderle-Ma S, Grossmann W, Fröschl KA (2014) Decision point analysis of time series data in process-aware information systems. In: Nurcan S, Pimenidis E, Pastor O, Vassiliou Y (eds) *CaISE Forum: joint proceedings of the CAiSE 2014 Forum and CAiSE 2014 Doctoral Consortium*, CEUR workshop proceedings 1164, CEUR-WS.org, pp 33–40
10. Džeroski S (2007) Towards a general framework for data mining. In: Džeroski S, Struyf J (eds) *KDID'07: knowledge discovery in inductive databases. Lecture notes in computer science*, vol 4747. Springer, Heidelberg, pp 259–300
11. Han J, Kamber M (2011) Data mining: concepts and techniques. Morgan Kaufmann series in data management systems. Morgan Kaufmann, Waltham, MA
12. Kimball R, Ross M (2010) The Kimball Group Reader: relentlessly practical tools for data warehousing and business intelligence, vol 1. Wiley, New York
13. Laursen G, Thorlund J (2010) Business analytics for managers: taking business intelligence beyond reporting. Wiley & SAS Business, New Jersey
14. Luhn HP (1958) A business intelligence system. *IBM J Res Dev* 2(4):314–319
15. Ly LT, Indiono C, Mangler J, Rinderle-Ma S (2012) Data transformation and semantic log purging for process mining. In: Ralyté J, Franch F, Brinkkemper S, Wrycza S (eds) *CaISE'12: international conference on advanced information systems engineering. Lecture notes in computer science*, vol 7328. Springer, Heidelberg, pp 238–253
16. Mitchell T (1997) Machine learning. McGraw Hill, New York
17. Moss LT, Atre S (2003) Business intelligence roadmap: the complete project lifecycle for decision-support applications. Addison-Wesley Professional, Boston
18. Negash S (2004) Business intelligence. *Commun Assoc Inf Syst* 13(1):177–195
19. Power DJ (2007) A brief history of decision support systems. *DSSResources. COM*, <http://dssresources.com/history/dsshistory.html>. Accessed 20 Dec 2014
20. Rausch P, Sheta AF, Ayesh A (eds) (2013) Business intelligence and performance management: theory, systems and industrial applications. Springer, Berlin/Heidelberg
21. Roebuck K (2011) Business intelligence (BI): high-impact strategies—what you need to know: definitions, adoptions, impact, benefits, maturity, vendors. Emereo, ISBN: 9781743046289
22. Trujillo J, Maté A (2012) Business intelligence 2.0: a general overview. In: Aufaure M-A, Zimanyi E (eds) *Business intelligence. Lecture notes in business information processing*, vol 96. Springer, Heidelberg, pp 98–116
23. van der Aalst WMP (2011) Process mining—discovery, conformance and enhancement of business processes. Springer, New York/Berlin
24. van der Aalst WMP et al (2012) Process mining manifesto. In: Daniel F, Barkaoui K, Dustdar S (eds) *Business process management workshops. Lecture notes in business information processing*, vol 99. Springer, Heidelberg, pp 169–194
25. Weske M (2012) Business process management: concepts, languages, architectures. Springer, Berlin/Heidelberg

<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